Decision Support Systems

Decision Support Systems

Edited by

Chiang S. Jao

Intech

Published by Intech

Intech
Olajnica 19/2, 32000 Vukovar, Croatia

Free online edition of this book you can find under www.sciyo.com
Additional copies can be obtained from:
publication@sciyo.com

First published January 2010
Printed in India

Technical Editor: Teodora Smiljanic
Cover designed by Dino Smrekar

Decision Support Systems, Edited by Chiang S. Jao
p. cm.
ISBN 978-953-7619-64-0

Preface

Decision support systems (DSS) have evolved over the past four decades from theoretical concepts into real world computerized applications. DSS architecture contains three key components: a knowledge base, a computerized model, and a user interface. DSS simulate cognitive decision-making functions of humans based on artificial intelligence methodologies (including expert systems, data mining, machine learning, connectionism, logistical reasoning, etc.) in order to perform decision support functions. The applications of DSS cover many domains, ranging from aviation monitoring, transportation safety, clinical diagnosis, weather forecast, business management, to internet search strategy. By combining knowledge bases with inference rules, DSS are able to provide suggestions to end users to improve decisions and outcomes.

At the dawn of the 21st century, more sophisticated applications of computer-based DSS have been evolved and they have been adopted in diverse areas to assist in decision making and problem solving. Empirical evidence suggests that the adoption of DSS results in positive behavioral changes, significant error reduction, and the saving of cost and time. This book provides an updated view of the state-of-art computerized DSS applications. The book seeks to identify solutions that address current issues, to explore how feasible solutions can be obtained from DSS, and to consider the future challenges to adopting DSS.

Overview and Guide to Use This Book

This book is written as a textbook so that it can be used in formal courses examining decision support systems. It may be used by both undergraduate and graduate students from diverse computer-related fields. It will also be of value to established professionals as a text for self-study or for reference. Our goals in writing this text were to introduce the basic concepts of DSS and to illustrate their use in a variety of fields.

Chapter 1 first discusses the motivational framework that highlights the significance of motivational factor, a psychological construct, in explaining and facilitating the comprehension of DSS use and decision performance. The motivational framework translates several user-related factors (task motivation, user perception of a DSS, motivation to use DSS, DSS adoption, and decision performance) to the driving force in using DSS to improve task processing effectively and efficiently. To understand thoroughly the motivation framework will assist system designers and end users in reducing the barriers of system design and adoption.

Chapters 2 and 3 introduce complicated decision support processes. Chapter 2 explores how to apply intelligent multi-agent paradigm architecture in DSS for the distributed environment. Intelligent multi-agent technology is adopted to develop DSS in enhancing the system operation in a dynamic environment and in supporting the adaptability of the system under complicated system requirements. An intelligent agent is capable to adapt DSS on the new situation through effective learning and reasoning. Employing multi-agents will simplify the complex decision making process and expedite the operation more efficiently.

Chapter 3 applies a hybrid decision model using generic algorithm and fuzzy logic theory to provide decision makers the ability to formulate nearly optimal sets of knowledge base and to improve the efficiency of warehouse management. This model incorporates error measurement to reduce the complexity of process change during the development and selection of the best warehouse design for a given application.

Chapter 4 reviews connectionist models of decision support in a clinical environment. These models connect the implementation and the adoption of DSS to establish effective medical management, maintenance and quality assurance and to predict potential clinical errors. These models aim to provide clinicians effective drug prescribing actions and to ensure prescription safety. The implementation of DSS accompanies the advantages of staff education and training to promote user acceptance and system performance.

Chapter 5 integrates DSS with data mining (DM) methodology for customer relationship management (CRM) and global system for mobile communication (GSM) for the business service requirements. Data mining is appropriate for analyzing massive data to uncover hypothetical patterns in the data. A data mining DSS (DMDSS) offers an easy-to-use tool to enable business users to exploit data with fundamental knowledge, and assists users in decision making and continual data analysis.

Chapter 6 highlights the importance of DSS evaluation using various testing methods. Integrating several testing methods would help detect primary errors generally found in the DSS adoption. A gold standard knowledge source is critical in choosing DSS testing methods. Correct use of these testing methods can detect significant errors in DSS. At this point, you are able to understand how to design and evaluate DSS for general purposes.

Chapter 7 adopts artificial neural network (ANN) model in developing DSS for pharmaceutical formulation development. The use of ANNs provides the predictive "black-box" model function that supports the decision difficult to explain and justify because numerous system parameters are under consideration. Integrating DSS with ANNs applies data mining methodology and fuzzy logic algorithm, mentioned in Chapter 3, for decision making under multiple influential factors after performing statistical sensitivity analysis on feasible decision making mechanisms. The ANN in DSS is especially useful in improving drug substance original characteristics for optimized pharmaceutical formulation.

Chapters 8 and 9 introduce the application of DSS in the clinical domain. Chapter 8 investigates the characteristics of clinical DSS (CDSS) and illustrates the architecture of a CDSS. An example of embedding CDSS implementation within computerized physician order entry (CPOE) and electronic medical record (EMR) is demonstrated. A CDSS aims to assist clinicians making clinical errors visible, augmenting medical error prevention and promoting patient safety.

Chapter 9 introduces the importance of knowledge bases that provide useful contents for clinical decision support in drug prescribing. Knowledge bases are critical for any DSS in providing the contents. Knowledge bases aim to fulfill and be tailored timely to meet specific needs of end users. Standards are vital to communicate knowledge bases across different DSS so that different EMRs can share and exchange patient data on different clinical settings. Knowledge bases and CDSS have been proved to be helpful in daily decision making process for clinicians when instituting and evaluating the drug therapy of a patient.

Chapters 10 and 11 introduce the concepts of spatial DSS. Chapter 10 introduces the framework of a web service-based spatial DSS (SDSS) that assists decision makers to generate and evaluate alternative solutions to semi-structured spatial problems through integrating analytical models, spatial data and geo-processing resources. This framework aims to provide an environment of resource sharing and interoperability technically through web services and standard interfaces so as to alleviate duplication problems remotely and to reduce related costs.

Chapter 11 introduces another SDSS for banking industry by use of geographic information systems (GIS) and expert systems (ES) to decide the best place for locating a new commence unit in the banking industry. This SDSS aims to improve the decision making process in solving issues of choosing a new commence location for the banking industry, expanding possibilities through spatial analysis, and assisting domain experts in managing subjective tasks.

Chapters 12 and 13 introduce DSS adoption in monitoring the environment. Chapter 12 introduces a web-based DSS for monitoring and reducing carbon dioxide (CO_2) emissions to the environment using an intelligent data management and analysis model to incorporate human expert heuristics and captured CO_2 emission data. Using object-linking and embedding (OLE) technology, this DSS aims to automatically filter and process massive raw data in reducing significant operating time.

Chapter 13 illustrates case studies of Canadian environmental DSS (EDSS). The EDSS makes informed resource management decisions available to users after integrating scientific data, information, models and knowledge across multimedia, multiple-disciplines and diverse landscapes. The EDSS is also using GIS mentioned in Chapter 11 to deal with temporal and spatial consistency among different component models. The EDSS can solve complex environmental issues by providing informed resource and perform data analysis effectively. The schematic EDSS concepts of an EDSS can assist in developing a good EDSS with required functions to achieve the goals of environmental monitoring.

Chapters 14 to 21 illustrate several examples of DSS adoption in diverse areas (including business partnership, internet search, wine management, agribusiness, internet data dependencies, customer order enquiry, construction industry, and disaster management) to solve problems in the current world.

Chapter 14 presents a set of different DSS that extend the decision support process outside a single company. An automatic speech synthesis interface is adopted in the web-based DSS for the operational management of virtual organizations. Incorporating different business partners can provide decision support in multiple useful scenarios and extend the interoperability in a centralized cooperative and distributed environment. This trend is very useful to meet decision support requirements for global business in the 21st century.

Chapter 15 introduces a DSS for analyzing prominent ranking auction markets for internet search services. This strategy has been broadly adopted by the internet search service provider like Google. This DSS aims to analyze ranking auction by the bidding behavior of a set of business firms to display the searched information based on the ranking by bids strategy. You will be able to understand how the searching information being displayed on the internet by the searching engine, just like what you have seen by using Google Search.

Chapter 16 introduces a DSS for evaluating and diagnosing unstructured wine management in the wine industry. This DSS offers effective performance assessment of a given winery and ranks the resource at the different levels of aggregation using statistical data. It aims in improved resource utilization and significant operational cost and time reduction. Fuzzy logic theory is adopted in the decision support process to compute a give winery performance in term of several dependent factors.

Chapter 17 introduces a DSS adopted in agribusiness (hop industry) concerning issues related to personnel safety, environmental protection and energy saving. This DSS aims to monitor all functions of an agricultural process and to satisfy specific performance criteria and restrictions. Automation Agents DSS (AADSS) is adopted to support decision making in the range of the agribusiness operation, production, marketing and education. The AADSS facilitates the support to farmers in e-commence activities and benefits effective labor and time management, environmental protection, better exploitation of natural sources and energy saving.

Chapter 18 introduces a framework for automating the building of diagnostic Bayesian Network (BN) model from online data sources using numerical probabilities. An example of a web-based online data analysis tool is demonstrated that allows users to analyze data obtained from the World Wide Web (WWW) for multivariate probabilistic dependencies and to infer certain type of causal dependencies from the data. You will be able to understand the concept in designing the user interface of DSS.

Chapter 19 introduces a DSS based on knowledge management framework to process customer order enquiry. This DSS is provided for enquiry management to minimize cost, achieve quality assurance and enhance product development time to the market. Effective and robust knowledge management is vital to support decision making at the customer order enquiry stage during product development. This DSS highlights the influence of negotiation on customer due dates in order to achieve forward or backward planning to maximize the profit.

Chapter 20 introduces a web-based DSS for tendering processes in construction industry. This DSS is used to benefit the security of tender documents and to reduce administrative workload and paperwork so as to enhance productivity and efficiency in daily responsibilities. This DSS is used in reducing the possibility of tender collusion.

Chapter 21 introduces the concept of DSS used in disaster management based on principles derived from ecology, including preservation of ecological balance, biodiversity, reduction of natural pollution in air, soil and water, and exploitation of natural resources. This DSS provides complex environment management and public dissemination of environment-related information.

The book concludes in Chapter 22 with the introduction of a theoretical DSS framework to secure a computer system. This CDSS framework adopts an accurate game-theoretic model to identify security primitives of a given network and assesses its security enhancement. Through the set-up of a game matrix, the DSS provides the capability of analysis, optimization and prediction of potential network vulnerability for security assessment. Five examples are provided to assist you in comprehending the concept of how to construct networks with optimal security settings for your computer system.

It is exciting to work in the development of DSS that is increasingly maturing and benefits our society to some degree. There is still ample opportunity remaining for performance enhancement and user acceptance as new computer technologies evolve and more modern problems in the current world are being faced. In light of the increasing sophistication and specialization required in decision support, it is no doubt that the development of practical DSS needs to integrate multi-disciplined knowledge and expertise in diverse areas. This book is dedicated to providing useful DSS resources that produce useful application tools in decision making, problem solving, outcome improvement, and error reduction. The ultimate goals aim to promote the safety of beneficial subjects.

Editor

Chiang S. Jao
National Library of Medicine
United States

Contents

1

Motivational Framework: Insights into Decision Support System Use and Decision Performance

Siew H. Chan and Qian Song
Washington State University
United States

1. Introduction

The purpose of this chapter is to discuss how characteristics of a decision support system (DSS) interact with characteristics of a task to affect DSS use and decision performance. This discussion is based on the motivational framework developed by Chan (2005) and the studies conducted by Chan (2009) and Chan et al. (2009). The key constructs in the motivational framework include task motivation, user perception of DSS, motivation to use a DSS, DSS use, and decision performance. This framework highlights the significant role of the motivation factor, an important psychological construct, in explaining DSS use and decision performance. While DSS use is an event where users place a high value on decision performance, the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT) do not explicitly establish a connection between system use and decision performance. Thus, Chan (2005) includes decision performance as a construct in the motivational framework rather than rely on the assumption that DSS use will necessarily result in positive outcomes (Lucas & Spitler, 1999; Venkatesh et al., 2003). This is an important facet of the framework because the ultimate purpose of DSS use is enhanced decision performance.

Chan (2009) tests some of the constructs in the motivational framework. Specifically, the author examines how task motivation interacts with DSS effectiveness and efficiency to affect DSS use. As predicted, the findings indicate that individuals using a more effective DSS to work on a high motivation task increase usage of the DSS, while DSS use does not differ between individuals using either a more or less effective DSS to complete a low motivation task. The results also show significant differences for individuals using either a more or less efficient DSS to complete a low motivation task, but no significant differences between individuals using either a more or less efficient DSS to perform a high motivation task only when the extent of DSS use is measured dichotomously (i.e., use versus non-use). These findings suggest the importance of task motivation and corroborate the findings of prior research in the context of objective (i.e., computer recorded) rather than subjective (self-reported) DSS use. A contribution of Chan's (2009) study is use of a rich measure of DSS use based on Burton-Jones and Straub's (2006) definition of DSS use as an activity that includes a user, a DSS, and a task.

Chan et al. (2009) extends the motivational framework by investigating the alternative paths among the constructs proposed in the framework. Specifically, the authors test the direct effects of feedback (a DSS characteristic) and reward (a decision environment factor), and examine these effects on decision performance. The results indicate that individuals using a DSS with the feedback characteristic perform better than those using a DSS without the feedback characteristic. The findings also show that individuals receiving positive feedback, regardless of the nature (i.e., informational or controlling) of its administration perform better than the no-feedback group. These results provide some evidence supporting the call by Johnson et al. (2004) for designers to incorporate positive feedback in their design of DSS. Positive feedback is posited to lead to favorable user perception of a DSS which in turn leads to improved decision performance. The findings also suggest that task-contingent reward undermine decision performance compared to the no reward condition, and performance-contingent reward enhance decision performance relative to the task-contingent reward group. The study by Chan et al. (2009) demonstrates the need for designers to be cognizant of the types of feedback and reward structures that exist in a DSS environment and their impact on decision performance.

The next section presents Chan's (2005) motivational framework. Sections 3 and 4 discuss the studies by Chan (2009) and Chan et al. (2009) respectively. The concluding section proposes potential research opportunities for enhancing understanding of DSS use and decision performance.

2. Motivational framework

The motivational framework (Chan, 2005) provides a foundation for facilitating understanding of DSS use and decision performance. A stream of research is presented based on a review of the literature on motivation, information processing, systems, and decision performance. The framework illustrates the factors that affect task motivation, and the DSS characteristics that influence user perception of a DSS which in turn impacts motivation to use the DSS. Task motivation and motivation to use the DSS are posited to influence DSS use. The framework also depicts a link between DSS use and decision performance. Figure 1 shows the adapted motivational framework developed by Chan (2005). The constructs in the framework are discussed below.

2.1 DSS characteristics

The characteristics of a DSS include ease of use (Davis, 1989), presentation format (Amer, 1991; Hard & Vanecek, 1991; Umanath et al., 1990), system restrictiveness (Silver, 1990), decisional guidance (Silver, 1990), feedback (Eining & Dorr, 1991; Gibson, 1994; Stone, 1995), and interaction support (Butler, 1985; Eining et al., 1997).

2.1.1 Ease of use

DSS use is expected to occur if users perceive a DSS to be easy to use and that using it enhances their performance and productivity (Igbaria et al., 1997). Less cognitive effort is needed to use a DSS that is easy to use, operate, or interact with. The extent of ease of use of a DSS is dependent on features in the DSS that support the dimensions of speed, memory, effort, and comfort (Thomas, 1996). A DSS is easy to use if it reduces user performance time (i.e., the DSS is efficient), decreases memory load with the nature of assistance provided (memory), reduces mental effort with simple operations (effort), and promotes user comfort (comfort). An objective of developers is to reduce the effort that users need to expend on a

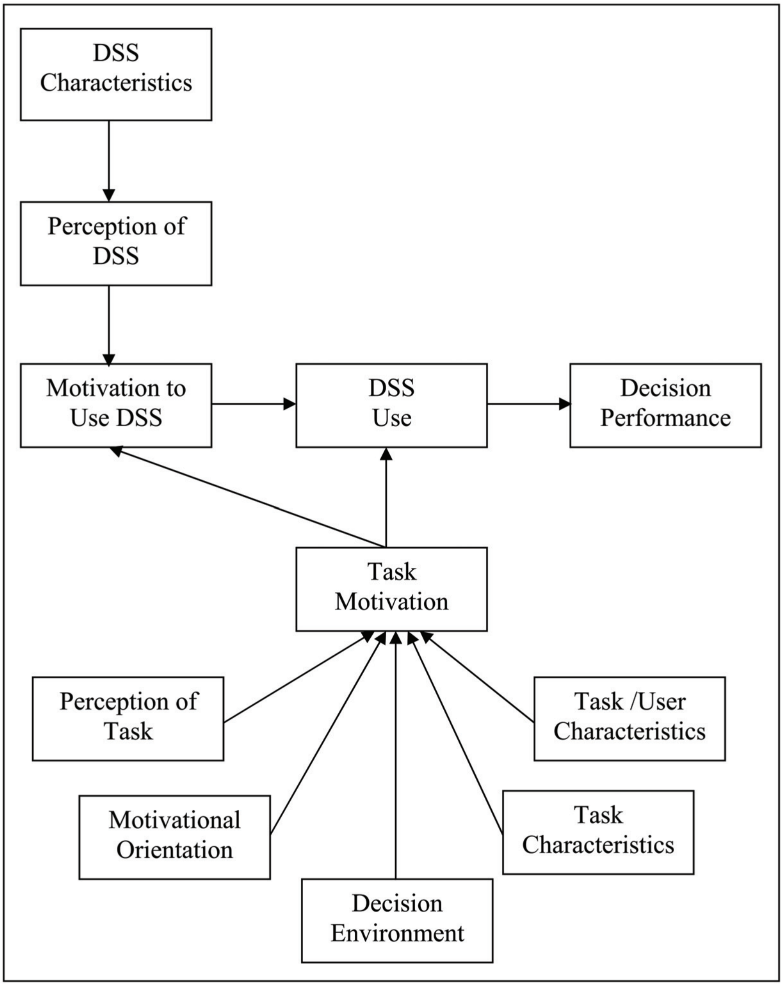


Fig. 1. A Motivational Framework for Understanding DSS Use and Decision Performance (Adapted from Chan (2005))

task by incorporating the ease of use characteristic into a DSS so that more effort can be allocated to other activities to improve decision performance. DSS use may decline if increased cognitive effort is needed to use a DSS because of lack of ease of use.

2.1.2 Presentation format

Presentation of a problem can be modified based on the assumption that information is correctly processed when it is presented in a form that evokes appropriate mental procedures (Roy & Lerch, 1996). The prospect theory (Kahneman & Tversky, 1979) suggests that presentation (framing) of alternatives can affect the riskiness of decision outcomes. This theory suggests that the way information is presented may influence a user's judgment or decision. In addition, the cognitive fit theory (Vessey, 1991; Vessey & Galletta, 1991) indicates that the level of complexity in a given task is reduced effectively when the problem-solving tools or techniques support the methods or processes required for doing the task. Thus, problem solving with cognitive fit results in effectiveness and efficiency gains.

2.1.3 System restrictiveness and decisional guidance

Two DSS attributes, system restrictiveness and decisional guidance, have been examined to show what users can and will do with a DSS (Silver, 1990). System restrictiveness refers to the degree to which a DSS limits the options available to the users, and decisional guidance refers to a DSS assisting the users to select and use its features during the decision-making process. If a decision-making process encompasses the execution of a sequence of information processing activities to reach a decision, then both the structure and execution of the process can be restricted by a DSS. The structure of the process can be restricted in two ways: limit the set of information processing activities by providing only a particular subset of all possible capabilities, and restrict the order of activities by imposing constraints on the sequence in which the permitted information processing activities can be carried out. User involvement is often essential during the execution of information processing activities after the structure of the process has been determined. The structure in the decision-making process is also promoted with the use of a restrictive DSS; in this respect, users are not overwhelmed with choices among many competing DSS. In certain cases, additional structure may actually enhance DSS use when ease of use is facilitated. However, lesser system restrictiveness may be preferred to enhance learning and creativity. Users may not use a DSS that is too restrictive because they may consider DSS use to be discretionary (Silver, 1988).

2.1.4 Feedback

Several researchers have undertaken exploration of the impact of various types of message presentation on users' behavior (Fogg & Nass, 1997; Johnson et al., 2004; Johnson et al., 2006; Tzeng, 2004). Fogg and Nass (1997) focus on the use of "sincere" praise, "flattery" (i.e., insincere praise) and generic feedback, and report that the sincere and flattery forms are perceived to be more positive. The authors suggest that incorporating positive feedback into training and tutorial software increases user enjoyment, task persistence, and self-efficacy. The positive feelings provided by the positive feedback engage the users and lead to greater success in system use (Fogg & Nass, 1997).

Tzeng (2004) uses a similar type of strategy to alleviate the negative reactions to system use arising from debilitated use of the system. The feedback from the system is examined in the context of "apologetic" versus "non-apologetic" presentation. As anticipated, the apologetic feedback provided by the system creates a favorable experience for the users (Tzeng, 2004). The results add to the body of research suggesting that system interface designers should be conscious of the need to create favorable user perception of systems to increase positive user experience to obtain increased system use and enhanced decision performance.

2.1.5 Interaction support

Interaction support is present when users are allowed a certain level of interactivity with a DSS. The design of a DSS has a determining effect on the degree of interaction between a user and a DSS (Silver, 1990). Individuals may perceive control over a DSS when some level of interaction support is provided by the DSS. Perceived control over the use of a DSS may have positive effects on motivation to use the DSS. Indeed, motivation is enhanced by the provision of information choice (Becker, 1997). Individuals using a DSS that allows user input (choice) in determining the DSS contents are more motivated than those using a DSS that does not allow this input (Roth et al., 1987). The effectiveness and acceptance of a DSS increase when users are provided with some control over the DSS (Roth et al., 1987). In a study where DSS with different levels of interaction support are designed, expert system users are reported to be in more frequent agreement with the DSS than the statistical model and checklist users (Eining et al., 1997). Specifically, individuals using a DSS with increased interaction support place more reliance on the DSS than those using the DSS with limited interaction support. Hence, the interaction support provided by the DSS has a positive impact on DSS use (Brown & Eining, 1996).

2.2 User perception of a DSS

User perception of a DSS (i.e., effectiveness, efficiency, and effort) is one of the two significant constructs that affects motivation to use a DSS. The relationship between user perception of a DSS and motivation to use the DSS is expected to be positive. That is, motivation to use a DSS is expected to increase when the DSS is perceived to be more effective or efficient, or less effortful to use.

2.2.1 Effectiveness

Prior research (e.g., Amer 1991; Eining & Dorr, 1991; Hard & Vanecek, 1991) has measured effectiveness in the context of DSS use. However, limited research has examined how the characteristics of a DSS influence DSS use. Factors, including the importance of a decision, may cause individuals to place more emphasis on effectiveness (Payne et al., 1993). Users may also place more weight on effectiveness and exert more effort to attain their goals when they realize the benefits of improved decisions; consequently, user considerations of decision performance lead to increased DSS use (Chenoweth et al., 2003). As individuals increase their focus on decision performance, DSS effectiveness becomes a positive factor affecting DSS use.

2.2.2 Efficiency

A DSS is efficient if it assists users in their decision-making in a timely manner. Rapid advances in computing technology, especially processing speed, result in less user tolerance for any delay in Internet applications (Piramuthu, 2003). Slow speed and time delays debilitate ease of use and have a negative impact on system use (Lederer et al., 2000; Lightner et al., 1996; Pitkow & Kehoe, 1996). Previous research has shown that system response time has an impact on the extent of system use. For example, download speed has been identified as one of the technology attributes that significantly influences intention to shop and actual purchase behavior in online consumer behavior research (Limayem et al., 2000). Download speed is also one of the key factors underlying user perception about the quality of a system (Saeed et al., 2003). Users may become anxious and less satisfied with a website or DSS when they experience delay in their processing requests (Tarafdar & Zhang, 2005). A delay that exceeds 10 seconds can cause users to lose concentration on the contents of a website (Nielsen, 2000). Novak et al. (2000) develop a speed of interaction scale and find that higher interaction speed has a positive impact on users' experience in system use.

2.2.3 Effort

Individuals experience a certain degree of effort in doing a task (Eisenberger & Cameron, 1996) and they tend to minimize effort when they engage in the task (Todd & Benbasat, 1992). The extent of effort-sensitive cognitive processes required by a specific activity must be taken into consideration when establishing a relationship between increases in effort and changes in performance. The decision strategies that individuals employ to process information vary in terms of the amount of effort involved in using these strategies. For example, the additive compensatory strategy is considered to be an effortful decision strategy (Payne et al., 1993) because individuals are required to examine all the attributes for two alternatives at a given time. In contrast, the elimination-by-aspects strategy is viewed to be a less effortful decision strategy (Payne et al., 1993) because the size of the alternative set is reduced each time an attribute is selected. The reduced alternative set decreases the amount of information processing.

Previous research demonstrates that DSS use increases when a DSS decreases the effort required for implementing an effortful strategy (Todd & Benbasat, 1992), and when use of the DSS leads to increased decision quality or accuracy (Todd & Benbasat, 1996). Todd and Benbasat (1994) extend and complement previous studies on the role of effort and accuracy in choice tasks by examining the role of DSS in reducing cognitive effort and, therefore, influencing strategy selection. They stress the importance of understanding the role of cognitive effort because it provides valuable insight into how a DSS influences the selection of problem-solving strategies by changing the effort relationships among the component processes that make up these strategies. Specific features can be incorporated into a DSS to change the relative effort required to implement different choice strategies; this can in turn affect strategy selection by a decision maker. Therefore, choice processes can be engineered to influence users to adopt strategies that maximize their value or utility (Todd & Benbasat, 1994).

2.3 Task motivation

Task (intrinsic) motivation is an important psychological construct in the motivational framework. Task motivation arises from one's propensity to engage in activities of interest and the resultant promotion in learning and development and expansion of the individual's capacities (Ryan & Deci, 2000). Task motivation entails "positively valued experiences that individuals derive directly from a task" and conditions specific to the task that produce motivation and satisfaction (Thomas & Velthouse, 1990, p. 668). People are motivated to perform a task when they engage in an activity simply for the satisfaction inherent in the behavior. This satisfaction can arise from positive feelings of being effective (White, 1959) or being the origin of behavior (deCharms, 1968). Task motivation is critical for high quality performance (Utman, 1997). The literature on the impact of task characteristics on work performance (e.g., Aldag & Brief, 1979; Hackman & Oldham, 1980; Lawler, 1973; Thomas & Velthouse, 1990) indicates a need for identifying factors that affect task motivation.

Task motivation (Amabile, 1983, 1988) is influenced by the following five factors: user perception of a task, users' motivational orientation, decision environment, task characteristics, and task/user characteristics (ability, knowledge, and experience).

2.3.1 Perception of task

The four components of the Perception of Task Value scale (Eccles et al., 1983) are interest, importance, utility, and cost. The motivation theory suggests that task motivation is high when a task is perceived to be high in interest, importance or utility, or the cost of engaging in the task is low, and vice versa.

Individuals experience interest when their needs and desires are integrated with the activity. From this perspective, interest is the driving mechanism for all actions, including cognitive activity (Piaget, 1981). A person is said to be experientially interested when a certain quality of attention and sense of delight is present. Interest leads to the performance of intrinsically motivated behaviors (Deci, 1998). In this respect, interest and intrinsic motivation are considered to be synonymous (Tobias, 1994). Consistent with the definition offered by Sansone and Smith (2000), this chapter defines task (intrinsic) motivation as a person's experience of interest in an activity.

The importance component pertains to the importance of performing well in an activity (Eccles et al., 1983). Importance is also related to the relevance of engaging in an activity to either confirm or disconfirm salient features of a person's actual or ideal self-schema (Wigfield & Eccles, 1992). A task is deemed to be high in importance if it allows individuals to confirm salient attributes of their self-schemata (e.g., competence in the domains of sports or arts) (Wigfield & Eccles, 1992). When users perceive a task to be personally important, they become motivated by the task, leading to increased task motivation.

The utility component refers to the importance of a task for the pursuance of a variety of long-term or short-term goals without any regard for a person's interest in the task (Wigfield & Eccles, 1992). The utility factor relates to a person's extrinsic reasons for engaging in an activity; that is, a person may engage in a task not for its own sake but to obtain desired goals (Wigfield & Eccles, 1992). Utility can also be viewed as perceived usefulness of the task for goal attainment (e.g., individuals' belief about how the task can assist them to attain specific goals such as career prospects or outperforming others) (Pintrich & Schrauben, 1992).

The cost of engaging in a task is affected by the (1) amount of effort necessary for succeeding, (2) opportunity cost of engaging in the activity, and (2) anticipated emotional states such as performance anxiety, fear of failure, or fear of the negative consequences of success (Wigfield & Eccles, 1992). A negative relationship is proposed to exist between the value of a task and the cost/benefit ratio in terms of the amount of effort required for doing well in the task (Eccles et al., 1983). The opportunity cost of a task refers to the time lost for engaging in other valued alternatives (Eccles et al., 1983). Further, a person may experience anxiety, fear of failure, or fear of the negative consequences of success in the course of a task engagement (Eccles, 1987).

2.3.2 Motivational orientation

Individuals may be intrinsically motivated (i.e., perform a task for the sake of interest), extrinsically motivated (i.e., complete a task for the sake of extrinsic incentives) or have no motivation for doing a task (Amabile, 1988). Individuals have a desire to perform well either for internal (e.g., interest or enjoyment) or external (e.g., to impress others or to attain goals) reasons. A person's baseline attitude toward an activity can be considered as a trait (Amabile, 1983). Researchers (deCharms, 1968; Deci & Ryan, 1985; Harter, 1981) have treated the intrinsic-extrinsic motivational orientation as a stable individual difference variable. This means that an individual can walk into a situation with a specific motivational orientation. The type of motivational orientation (i.e., intrinsic, extrinsic, or both) determines a person's initial task motivation. Motivational orientation has an impact on the final and type of motivation in a specific task. The Work Preference Inventory (WPI) has been developed to assess the intrinsic and extrinsic motivation of individuals (Amabile et al., 1994). This scale directly assesses the intrinsic and extrinsic motivation of individuals, assumes the coexistence of intrinsic and extrinsic motivation, and incorporates a wide range of cognitions and emotions proposed to be part of intrinsic and extrinsic motivation. Chan's (2005) motivational framework suggests examination of the impact of motivational orientation (a trait variable) on task motivation (a state variable).

2.3.3 Decision environment

The decision-making process is frequently influenced by factors in the environment. These factors have an impact on the behaviors of decision makers. Factors in the decision environment (i.e., reward, justification, accountability, and time constraint) have an effect on task motivation. Task motivation is expected to be high when individuals are (a) provided with rewards that do not undermine their interest in a task (b) required to justify their performance in the task, (c) held accountable for the outcome of their decision performance, or (d) required to complete the task in a specific time frame. Task motivation is predicted to be low when the above decision environmental factors are absent.

(a) Rewards

Factors affecting motivation, and thus effort and performance, are difficult to consider without also considering the reward structures that are in place for effort and performance. While rewards are primarily viewed as necessary to provide extrinsic motivation, a meta-analysis of 128 well-controlled experiments examining the relationship between rewards and intrinsic motivation reveals significant and consistent negative impact of rewards on intrinsic motivation for interesting activities (Deci et al., 1999). This effect may be due to reward-oriented individuals being more directed toward goal-relevant stimuli, and the rewards actually divert such individuals' attention away from the task and environmental stimuli that might promote more creative performance (Amabile, 1983). Indeed, rewarded individuals "work harder and produce more activity, but the activity is of a lower quality, contains more errors, and is more stereotyped and less creative than the work of comparable non-rewarded subjects working on the same problems" (Condry, 1977, p. 471-472). On the other hand, there are many positive effects on performance derived generally from the introduction of rewards. Rewards can be used to motivate individuals to spend more time on a task (Awasthi & Pratt, 1990) and influence their focus on the task (Klein et al., 1997).

(b) Justification

The impact of justification and accountability on the decision makers' behaviors has been studied extensively in the judgment and decision making literature (e.g., Cuccia et al., 1995; Hackenbrack & Nelson, 1996; Johnson & Kaplan, 1991; Lord, 1992). Existing studies have used justification and accountability interchangeably. One explanation for the lack of distinction between these two constructs is the expectation of similar effects of justification and accountability on behaviors. Justification is defined as the need to justify one's decisions (Arnold, 1997); this definition is very similar to the definition of accountability offered by Kennedy (1993). Thus, the distinction between justification and accountability is unclear (Johnson & Kaplan, 1991).

Decision makers are constantly faced with the need to justify their decisions or to account to their sources for their decisions. Justification refers to the process that individuals experience to provide support or reasons for their behavior. Since individuals only need to provide justification for their behavior, they are not held responsible for the outcome as long as they are able to provide reasonable justification for their behavior. In contrast, when individuals are held accountable for their behavior, they are responsible for the outcome; that is, they will either be rewarded for a positive outcome or punished for a negative outcome. In this respect, two definitions of justification offered in the literature can promote understanding of the distinction between justification and accountability; that is, justification is "the act of providing evidence to support one's judgments or decisions" (Peecher, 1996, p. 126), or "the actual physical and/or mental process of explaining a judgment" (Johnson & Kaplan, 1991, p. 98).

(c) Accountability

Accountability is a "pre-existing expectation that an individual may be called on to justify his/her judgments to a significant other" (Johnson & Kaplan, 1991, p. 98). This implies that an important element of accountability is a person's responsibility for an outcome. In most business contexts, individuals are frequently expected to account for their decisions both to themselves and to others (Arnold, 1997). Some research evidence suggests that accountability can have an effect on decisions (Arnold, 1997). For example, MBA students show significant recency effect (i.e., they place more weight on evidence received later in a sequence) while this behavior is not observed with the auditor participants; however, the recency effect is absent when accountability is imposed on the MBA students (Kennedy, 1993).

(d) Time constraint

Time has frequently been used as a surrogate measure for cognitive effort or decision performance (Brown & Eining, 1996). For example, individuals in the highest time constraint condition exhibit more consistent performance than other groups when information load and presentation format in the context of a simple audit task are examined (Davis, 1994). The more consistent results obtained in this study can be attributed to the use of relatively simple strategies by the participants to reduce the effects of time constraint in the decision environment (Brown & Eining, 1996). Time constraint has also been reported to exert a negative impact on a judgment task relative to a choice task (Smith et al., 1997). Research can promote understanding of the effect of time constraint on task motivation.

2.3.4 Task characteristics

Task motivation is affected by characteristics of a task such as complexity, difficulty, structure, ambiguity, and novelty. Task motivation is expected to be high when a task is less complex, difficult, or ambiguous or has more structure or novelty, and vice versa.

(a) Complexity

Task complexity can occur at the stages of input, processing, or output and may relate to either the amount or clarity of information (Bonner, 1994). At the input stage, the amount of information can vary in terms of the number of alternatives, the number of attributes on which each alternative is compared, and attribute redundancy. Clarity of input may be reduced by relevant cues that are not specified or measured well, inconsistency between presented and stored cues, and presentation format. Processing can be complex when the amount of input increases, the number of procedures increases, procedures are not well specified, and the procedures are dependent on one another. Internally inconsistent cues or low or negative cue validities in nonlinear functions may reduce clarity and increase processing complexity. Complexity may also increase with the number of goals or solutions per alternative (i.e., the amount of output), and indefinite or unspecified goals (i.e., lack of clarity in output) created by the environment or by a person's lack of familiarity with the goals (Bonner, 1994).

(b) Difficulty

Difficulty can be defined as the amount of attentional capacity or mental processing required for doing a task (Kahneman, 1973). Task difficulty increases with increased similarity of the alternatives and this hampers a person's ability in discriminating the alternatives from one another (Stone & Kadous, 1997). A task is high in difficulty when a person perceives a tremendous amount of cognitive effort in information processing. The level of difficulty of a specific task has an effect on task motivation. Individuals are unlikely to be motivated by a task when they perceive the task to be difficult and vice versa. It is important to distinguish task complexity from task difficulty because these two constructs are not synonymous. That is, a complex task may involve an increased number of steps but it may not require increased cognitive effort to process the information (i.e., the task can be low in difficulty).

(c) Structure

Structure refers to the specification level of what is to be accomplished in a given task (Simon, 1973). A task can be classified on a continuum that indicates the degree of structure. A highly structured task requires a person to follow a predefined procedure for completing an activity. A task is highly unstructured when a predefined procedure for performing an activity is absent.

(d) Ambiguity

DSS use is reported to be influenced by task ambiguity (Brown & Jones, 1998). Although no significant difference in decision performance is found for both the DSS and non-DSS groups in relatively unambiguous decision situations, the DSS group outperforms the no-DSS group in relatively ambiguous decision contexts (Brown & Eining, 1996). Research is needed to provide insight into the impact of task ambiguity on task motivation and the resultant effect on motivation to use a DSS and DSS use.

(e) Novelty

Most conceptual definitions of creativity include the novelty characteristic (Hennessey & Amabile, 1988). Creativity is enhanced when novelty is present in a task. Individuals are most creative when they are motivated by a task and task motivation is further increased when the task entails a certain degree of novelty. Future work can facilitate understanding of the long- or short-term effects of the novelty characteristic on task motivation.

2.3.5 Task/User characteristics

Task/user characteristics refer to the users' ability, knowledge, and experience in a given task. These characteristics are discussed in the context of Libby's model. Ability relates to the users' capacity to engage in information processing activities that lead to problem solving; knowledge pertains to the information stored in memory; and experiences refer broadly to the task-related encounters that provide users with an opportunity to learn (Libby, 1992). Chan's (2005) motivational framework suggests that the users' ability, knowledge, and experience in a task have a positive effect on task motivation. That is, users with high ability are expected to be high in task motivation because their increased capacity in information processing results in effective and efficient problem solving. Users with low ability are predicted to be low in task motivation because of their limited capacity in information processing which in turn impairs their ability to solve problems. Users who are knowledgeable may possess essential information in memory that allows them to do a task effectively and efficiently; consequently, their task motivation is expected to be high. Less knowledgeable users may be low in task motivation because they do not have the necessary information stored in memory that permits them to carry out the task effectively and efficiently. Experienced users with task-related encounters are stimulated by the opportunities to learn and this increases their task motivation. Since less experienced users tend to have fewer task-related encounters and fewer opportunities to learn, their task motivation may be low.

2.4 Motivation to use a DSS

Researchers have conducted studies to enhance understanding of why and when users may become motivated to use a DSS. Use of an expert system is found to enhance the engagement of users and increase DSS use (Eining et al., 1997). In contrast, passive DSS use leads to deficient user behavior (Glover et al., 1997). This effect can be attributed to lack of motivation to use a DSS. The Perceptions of Task Value scale (Eccles et al., 1983) can be modified to obtain the Perception of DSS scale to measure a user's motivation to use a DSS. The four components in the scale include interest, importance, utility, and cost. Although these components can be differentiated, it is not easy to distinguish their relations (Jacobs & Eccles, 2000). Motivation to use a DSS is predicted to be high when the DSS is perceived to be high in interest, importance or utility, or the opportunity cost of using the DSS is low, and vice versa.

2.5 DSS use

A review of 22 articles published in MIS Quarterly, Decision Sciences, Management Science, Journal of Management Information Systems, Information Systems Research, and Information and Management indicates that self-reported system use is measured in 11 of the 22 studies (Legris et al., 2003). The method frequently comprised two or three questions pertaining to the frequency of use and the amount of time spent using the system. Ten studies do not measure use; that is, use is either mandatory or ignored. Many studies using TAM do not measure system use directly. Instead, these studies measure the variance in self-reported use (Legris et al., 2003). It is important to recognize that self-reported use is not a precise measure of system use (Davis, 1993; Legris et al., 2003; Subramanian, 1994). Use of omnibus measures such as perceived use/nonuse, duration of use or extent of use to measure the content of an activity may not be effective if a respondent is unclear about the specific part of the usage activity actually being measured. Thus, these perception measures may not be appropriate for measuring system use when the content of the activity is absent. In contrast, rich measures incorporate the nature of the usage activity that involves the three elements of system use -- a user, a system, and use of the system to do a task (Burton-Jones & Straub, 2006).

2.6 Decision performance

In general, a DSS is used to make better decisions or to make a decision with less effort. DSS use increases when the DSS decreases the effort required for implementing an effortful strategy (Todd & Benbasat, 1992), and when use of the DSS leads to increased decision quality or accuracy (Todd & Benbasat, 1996). Individual-level decision performance measures include objective outcomes, better understanding of the decision problem, or user perception of the system's usefulness (Lilien et al., 2004). Previous research on decision support has also used decision performance as a means of comparing systems (e.g., Lilien et al., 2004; Todd & Benbasat, 1994) and comparing other facets of decision support, such as data representations (e.g., Vessey, 1991). When a DSS extends the capabilities of users, it enables them to overcome limited resources and assists them in making better decisions (Todd & Benbasat, 1999). Empirical research indicates that improved decision performance results if a DSS is a good fit for a task and supports the user through reduced effort (Todd & Benbasat, 1999).

Additionally, a meta-analysis conducted by Fried and Ferris (1987) supports the relationship between task motivation and decision performance. Task motivation has been reported to be a strong predictor of performance (Kuvaas, 2006). The impact of task motivation on performance has been supported in the context of sports (e.g., Callahan et al., 2003; Catley & Duda, 1997) and education (Lin et al., 2001; Vansteenkiste et al., 2004; Wang & Guthrie, 2004). Research on the job characteristics model (Hackman & Oldham 1976) also reports that variables with job motivating features have a positive impact on performance (Fried & Ferris, 1987).

Chan's (2005) motivational framework provides a stream of research for investigating the impact of various variables on DSS use and decision performance. It is important to recognize the existence of alternative relationships among the constructs in the framework. For example, Chan (2009) proposes and tests a model that examines how task motivation interacts with DSS effectiveness and efficiency to affect DSS use. Chan et al. (2009) also present a model that examines how feedback and rewards influence decision performance.

The next section discusses a study by Chan (2009) that tests some of the constructs in the motivational framework.

3. The effects of task motivation, and DSS effectiveness and efficiency on DSS use

Task motivation and DSS effectiveness and efficiency are constructs in the motivational framework for understanding DSS use and decision performance. Task motivation is an important variable that influences DSS use (e.g., Davis et al., 1992; Hackbarth et al., 2003; Venkatesh, 2000; Venkatesh & Speier, 1999). Since TAM does not model task (intrinsic) motivation explicitly, Venkatesh (1999, 2000) attempts to fill this gap by conceptualizing intrinsic motivation as computer playfulness. To augment these efforts, Chan (2009) proposes a research framework that links DSS effectiveness and efficiency with task motivation. In this framework, the effects of DSS effectiveness and efficiency are moderated by task motivation while task motivation has a direct effect on DSS use. In particular, the author examines whether task motivation affects use of a DSS to do a task and whether task motivation interacts with DSS effectiveness and efficiency to affect DSS use.

Chan (2009) conducts an experiment where the participants use a DSS to do one of two choice tasks that induces different levels of task motivation. The total number of iterations of the participants' use of the DSS and the total time taken on each choice task are captured and used as dependent variables. The results show that participants in the high task motivation condition use the DSS more (i.e., they have more iterations and spend more time on the task) than those in the low task motivation condition. Individuals performing a high motivation task also use a DSS more when it is more effective while DSS effectiveness does not affect the level of usage for individuals doing a low motivation task. In addition, the findings indicate that DSS efficiency has a significant impact on DSS use for individuals working on a high or low motivation task when DSS use is measured as the extent of use (i.e, the number of iterations or total time spent on a task). However, DSS efficiency does not have a significant impact on DSS use in the high task motivation condition when the DSS use construct is dichotomized as use or non-use rather than the extent of use. This result is consistent with the author's expectation that individuals performing a high motivation task are less concerned with the efficiency of a DSS.

In summary, DSS use increases (decreases) for individuals using a more (less) effective DSS to work on a high motivation task. As expected, DSS effectiveness is not a concern when individuals perform a low motivation task. The findings suggest that the strong negative impact of lack of task motivation undermines DSS use, regardless of the level of its effectiveness. The efficiency of a DSS is found to interact with task motivation to affect DSS use. That is, individuals completing a high motivation task exhibit higher tolerance for a DSS that is low in efficiency. In contrast, lack of task motivation exacerbates the users' low tolerance for a DSS that is low in efficiency.

An interesting design of the DSS in Chan's (2009) study is the built-in feature of an effortful but accurate decision strategy -- additive difference (AD). AD processing compares two alternatives simultaneously by comparing each attribute, finding the difference, and summing the differences. It requires some method for weighting each attribute, some transformation to put all the attributes into compensatory units, and a way to sum the weighted values of the attributes. After a series of alternative comparisons, the alternative with the greatest sum is chosen. AD processing is compensatory in that values on one attribute necessarily offset the values on another attribute. It makes more complete use of the available information and is normatively more accurate than non-compensatory strategies such as elimination-by-aspects (Tversky, 1972). Use of the more accurate and more effortful AD strategy relative to other less accurate and less effortful strategies (e.g., elimination by aspects) may be encouraged if users are provided with a DSS that reduces the cognitive effort for using the AD strategy to complete a task. The effort required for completing a task is minimal when the DSS provides high support for the AD strategy (Todd & Benbasat, 2000). In the study by Chan (2009), individuals use a DSS to select two alternatives for comparison and the DSS provides the results of how the selected alternatives differed on the attributes. Thus, the DSS in the study provides enhanced automation that reduces the effort that a user may otherwise have to expend to process information manually.

The next section describes a study by Chan et al. (2009) that examines the effects of feedback and reward on decision performance.

4. The effects of feedback and reward on decision performance

Chan et al. (2009) extend the findings of Ryan et al. (1983) on the use of informational versus controlling feedback and rewards in the context of a DSS and the interface design. While Ryan et al. examine the effects of verbal feedback on intrinsic motivation, Chan et al. focus on the impact of text-based feedback from a DSS on decision performance. The authors also explore the effect of task-contingent versus performance-contingent rewards on decision performance. The results reveal a differential effect from that of Ryan et al. (1983) when feedback is provided through a DSS and the focus is on decision performance rather than the precursor condition of intrinsic motivation.

4.1 Informational feedback versus controlling feedback

Chan et al. (2009) use cognitive evaluation theory to examine feedback as a DSS characteristic. Cognitive evaluation theory suggests that events can be categorized as either informational or controlling. Informational feedback occurs when individuals receive information about their competency at a task in a self-determined performance context. When controlling feedback is administered, individuals experience pressure toward the achievement of specific outcomes such as attaining a specified level of performance (Ryan et al., 1983). Informational feedback facilitates an autonomy-supportive context that promotes autonomy, making individuals more inwardly focused and thus increasing task (intrinsic) motivation (Deci & Ryan, 1987). Controlling feedback debilitates autonomy, creativity (Amabile, 1983) and cognitive flexibility (McGraw & McCullers, 1979), leading individuals to perform in a specific manner in which they believe they "should" (Deci & Ryan, 1987). While individuals are more intrinsically motivated when they expect an informational rather than a controlling evaluation (Shalley & Perry-Smith, 2001), task (intrinsic) motivation is undermined by controlling feedback (Rigby et al., 1992). Previous studies (e.g. Ryan, 1982; Ryan et al., 1983) examine feedback in an informational or controlling manner and report that individuals exhibit higher task motivation in the informational feedback than controlling feedback condition.

While getting a user to accept and use a DSS is critical and the nature of the supportiveness of the feedback is important, some form of positive feedback assists individuals in performance improvement. In a DSS environment, the focus is on providing useful feedback for improving decision performance. Greater task motivation generated by informational feedback as opposed to controlling feedback leads to enhanced decision performance (Chan et.al, 2009). Individuals' level of interest in an activity increases when they receive feedback on their competence in the activity; consequently, they exert more effort to improve performance (Harackiewicz & Sansone, 2000).

4.2 Task-contingent versus performance-contingent reward

Cognitive evaluation theory also provides insight into the effect of rewards on individuals' behavior. In essence, rewards can be viewed as one type of feedback mechanism and classified as task noncontingent, task-contingent or performance-contingent rewards (Ryan et al., 1983).

Task noncontingent rewards occur when individuals receive rewards for doing a task, without requirement of engagement in the task (Deci et al., 1999). For example, providing a gift for participation without regard for how the participants perform during the experiment is a task noncontingent reward (Deci, 1972). Task noncontingent rewards are unlikely to affect task motivation because individuals are not required to perform well in the task, complete the task, or even engage in the task (Deci et al., 1999). Three meta-analyses performed by Deci et al. (1999), Tang and Hall (1995), and Cameron and Pierce (1994) do not suggest any significant impact of task noncontingent rewards on task motivation.

Task-contingent rewards require individuals to actually perform a task and can be classified as completion-contingent or engagement-contingent rewards (Deci et al., 1999). Completion-contingent rewards are provided only upon explicit completion of the target activity. For example, individuals work on four variations of a three-dimensional puzzle and receive $1 for each puzzle completed in the required time (Deci, 1971). Engagement-contingent rewards are offered simply for engagement in the task, without consideration of completion of the task. For instance, participants receive a reward for engaging in a series of hidden-figures puzzles (Ryan et al., 1983). These individuals are not aware of their performance in the task or the extent of their completion of the activity because they do not know the number of hidden figures in each drawing (Deci et al., 1999). Both completion-contingent and engagement-contingent rewards have about the same level of undermining effect (i.e. negative effect) on free-choice behavior and self-reported interest (Deci et al., 1999).

Performance-contingent rewards are administered for superior performance in an activity. Such rewards are either a direct function of actual performance success (e.g. an 80% accuracy rate on a task that leads to 80% of the maximum possible reward) or achievement of a specific standard (e.g., perform better than 80% of the other participants or achieve at least an 80% accuracy rate on a task). Performance-contingent rewards can have a facilitating or debilitating effect on task motivation, depending on the saliency of the informational or controlling aspect of the reward (Ryan et al., 1983). In particular, informational (controlling) administration of performance-contingent rewards leads to increased (decreased) task motivation (Harackiewicz, 1979; Ryan et al., 1983). Task motivation is maintained or increased if the performance-contingent reward is perceived to provide competence information; in contrast, task motivation is impaired if the reward is used to control how well a person does in a task (Ryan & Deci, 2000). The context in which performance-contingent rewards are administered can convey either competency or pressure to do well in an activity (Ryan et al., 1983).

Individuals using a DSS based on different reward structures are expected to exhibit different performance effects. Relative to the no reward condition, task-contingent rewards may be perceived as overjustification which undermines task motivation (e.g., Deci, 1972; Lepper et al., 1973; Ryan & Deci, 1996; Sansone & Harackiewicz, 1998). This undermining effect occurs when individuals are rewarded for doing an interesting task. The response to the reward is generally for individuals to exhibit less interest in, and willingness to, work on a task (Deci & Ryan, 1987). Performance-contingent rewards have also been shown to debilitate task motivation and decision performance (e.g., Boggiano & Ruble, 1979; Daniel & Esser, 1980; Ryan et al., 1983). Additionally, performance-contingent rewards can be more controlling, demanding, and constraining than task-contingent rewards because a specific standard of performance is expected. This leads to greater pressure and subsequent larger decrements in task motivation than in conditions where task-contingent rewards are administered (Harackiewicz & Sansone, 2000). In contrast, performance-contingent rewards may lead to better performance when individuals are motivated to work harder and put in more effort than they otherwise would (Harackiewicz & Sansone, 2000); therefore, performance-contingent rewards may be effective for improving decision performance (Lepper, 1981).

4.3 Interactive effect of feedback and reward on decision performance

It is imperative for researchers to consider the combined effects of feedback and reward on individuals' behavior (Ryan et al., 1983). Reward structures have informational and controlling attributes perceived by the individuals subject to the reward, and these informational and controlling attributes commingle with the informational and controlling nature of the feedback characteristic of a DSS. Perception of reward structures can be significantly influenced by the nature of feedback, with informational (controlling) feedback highlighting the informational (controlling) aspect of a reward structure.

Reward is an example of a controlling event that in itself may work against the positive effect of the information contained in the performance-contingent reward (Ryan & Deci, 2000). Although task motivation may be undermined by the prospect of reward during task performance, this effect may be offset by enhanced performance motivated by the expectation of reward (Deci & Ryan, 1985). Decision performance may not be undermined in the presence of informational feedback and performance-contingent rewards because cue values (Harackiewicz et al., 1984) may highlight the informational aspect of performance-contingent rewards and offset their controlling aspect. This sheds light on Chan et al.'s (2009) findings on insignificant decision performance effects for individuals provided with either an informational or controlling feedback when performance-contingent reward is administered. Consistent with Ryan et al.'s (1983) findings for their intrinsic motivation variable, Chan et al. (2009) report that the informational feedback/performance-contingent reward group marginally outperforms the no-feedback/task-contingent reward group. However, contrary to Ryan et al.'s (1983) finding of no significant difference for their intrinsic motivation measure, Chan et al. (2009) demonstrate that the controlling feedback/performance-contingent reward group performs better than the no-feedback/task-contingent reward group. This alternative finding is not surprising considering the combined effects of the participants' positive response to the controlling feedback in a DSS environment and the positive effect theorized for performance-contingent rewards on decision performance (as opposed to the negative effect on intrinsic motivation in Ryan et al.'s study).

5. Conclusion

Chan's (2005) motivational framework provides a foundation for facilitating understanding of DSS use and decision performance. Instead of relying on the assumption that DSS use necessarily results in improved decision performance, the motivational framework proposes a link between DSS use and decision performance. Chan (2005) also identifies the significant role of the motivation factor in explaining DSS use and decision performance. The author proposes examination of motivation as two separate components; namely, task motivation and motivation to use a DSS. Separation of these two effects assists researchers in identifying the underlying reasons for lack of DSS use.

Additionally, the motivational framework developed by Chan (2005) presents abundant future research possibilities. Future work can examine factors that affect task motivation, a key construct in the motivational framework. Task-related factors such as interest, utility, importance or the opportunity cost of engaging in a task can be manipulated to obtain a measure of self-reported task motivation to provide additional insight into future research findings. It might be interesting to investigate factors (e.g., the users' motivational orientation, decision environmental factors and task characteristics) that influence task motivation.

The motivation theory may provide insight into the findings by Todd and Benbasat (1992) on why users do not translate the effort savings from use of a DSS to perform a task into increased information processing. An examination of task motivation also helps us consider ways for increasing DSS use. DSS use is posited to occur when the benefits (i.e., effectiveness and efficiency) outweigh the costs (i.e., cognitive effort) associated with usage (Todd & Benbasat, 1996). For example, features can be incorporated into a DSS to reduce the cognitive effort involved in the use of a strategy (Todd & Benbasat, 1994a, 1994b) and to encourage DSS use (Todd & Benbasat, 1996).

A rich measure of DSS use consistent with Burton-Jones and Straub's (2006) definition of a DSS (that includes a user, a DSS, and use of the DSS to complete a task) is a more relevant construct than behavioral intention (Chan, 2009). Caution should be exercised to avoid the misleading assumption that behavior would follow intention (Limayem et al., 2000). For example, one might intend to lose 20 pounds; however, the individual might not engage in actual behavior (i.e., exercise or cut down on calories) to lose the intended weight. TAM posits that behavioral intention leads to system use (Davis et al., 1989); however, prior research findings on the relationship between intention to use systems and system use are mixed. Lack of a strong correlation between self-reported and objective usage data (Szajna, 1996) and the low correlation between intention and system use (Kim & Malhotra, 2005) present a challenge to the use of intention as a proxy for system use. Further, many TAM studies have used the intention (i.e., self-reported) measure as a proxy for system use although the focus of these studies is on system use (Kim & Malhotra, 2005). Since most TAM studies measure the variance in self-reported use, future research should measure system use rather than usage intention (Davis, 1993; Legris et al., 2003; Lucas & Spitler, 1999; Subramanian, 1994).

Further, empirical evidence in the behavioral decision-making literature suggests that decision makers make tradeoff between accuracy and effort in their formulation and subsequent use of DSS (Bettman et al., 1990; Creyer et al., 1990; Jarvenpaa, 1989; Johnson & Payne, 1985; Johnson et al., 1988; Payne, 1982; Payne et al., 1988, 1993; Stone & Schkade, 1991). Although accurate decision strategies such as additive difference (AD) can lead to improved decision performance, the effort required for using these strategies may discourage use of such strategies. Use of the more accurate AD strategy is expected to increase when the effort required for using the strategy is reduced; that is, when a DSS provides high support for the strategy (Todd & Benbasat, 2000).

Insights can also be gained from future work on whether user perception of a DSS might affect motivation to use a DSS, and whether task motivation interacts with DSS characteristics (e.g., ease of use, presentation format, system restrictiveness, decisional guidance, feedback or interaction support) to affect DSS use. Research can assist system developers in understanding the types of characteristics that can be incorporated into a DSS to create favorable user perception of the DSS to increase motivation to use the DSS, DSS use, and decision performance.

Finally, alternative paths among the constructs are implicit in the motivational framework developed by Chan (2005). Chan (2009) conducts a study to examine how task motivation interacts with DSS effectiveness and efficiency to affect DSS use. Chan et al. (2009) also examine the effects of feedback (a characteristic of a DSS) and reward (a characteristic of the decision environment) on decision performance. These studies demonstrate the existence of alternative paths in the motivational framework. Future work can explore other possible alternative models from the framework.

6. Reference

Aldag, R. & Brief, A. (1979). *Task design and employee motivation.* Scott. Foresman, Glenview, IL

Amabile, T. M. (1983). *The social psychology of creativity*, Springer-Verlag, New York, NY

Amabile, T. M. (1988). A model of creativity and innovation in organizations. *Research in Organizational Behavior*, 10, 123-167

Amabile, T. M.; Hill, K. G.; Hennessey, B. A. & Tighe, E. M. (1994). The Work Preference

Inventory: Assessing intrinsic and extrinsic motivational orientations. *Journal of Personality and Social Psychology*, 66, 5, 950-967

Amer, T. (1991). An experimental investigation of multi-cue financial information display and decision making. *Journal of Information Systems*, 5, 18-34

Arnold, V. (1997). Judgment and decision making, Part I: The impact of environmental factors. In: *Behavioral Accounting Research Foundations and Frontiers*, V. Arnold & S. G. Sutton (Ed.), 164-187, American Accounting Association, Sarasota, FL

Awasthi, V. & Pratt, J. (1990). The effects of monetary incentives on effort and decision performance: The role of cognitive characteristics. *Accounting Review*, 65, 4, 797-811

Becker, D. A. (1997). The effects of choice on auditors' intrinsic motivation and performance. *Behavioral Research in Accounting*, 9, 1-19

Bettman, J. R.; Johnson, E. J. & Payne, J. W. (1990). A componential analysis of cognitive effort in choice. *Organizational Behavior and Human Decision Processes,* 45, 111-139

Boggiano, A. K. & Ruble, D. N. (1979). Competence and the overjustification effect: A developmental study. *Journal of Personality and Social Psychology*, 37, 1462-1468

Bonner, S. E. (1994). A model of the effects of audit task complexity. *Accounting, Organizations and Society*, 19, 3, 213-214

Brown, D. L. & Eining, M. M. (1996). The role of decision aids in accounting: A synthesis of prior research. *Advances Accounting Information Systems*, 4, 305-332

Brown, D. L. & Jones, D. R. (1998). Factors that influence reliance on decision aids: A model and an experiment. *Journal of Information Systems*, 12, 75-94

Burton-Jones, A., & Straub, D. W., Jr. (2006). Reconceptualizing system usage: An approach and empirical Test. *Information Systems Research*, 17, 3, 228-246

Butler, S. A. (1985). Application of a decision aid in a judgmental evaluation of substantive test of details samples. *Journal of Accounting Research*, 23, 2

Callahan, J. S.; Brownlee, A. L.; Brtek, M. D. & Tosi, H. L. (2003). Examining the unique effects of multiple motivational sources on task performance. *Journal of Applied Social Psychology,* 33, 2515-2535

Cameron, J. & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research*, 64, 363-423

Catley, D. & Duda, J. L. (1997). Psychological antecedents of the frequency and intensity of flow in golfers. *International Journal of Sports Psychology,* 28, 309-322

Chan, S. H. (2005). A motivational framework for understanding IS use and decision performance. *Review of Business Information Systems*, 9, 4, 101-117

Chan, S. H. (2009). The roles of user motivation to perform a task and decision support system (DSS) effectiveness and efficiency in DSS use. *Computers in Human Behavior*, 25, 1, 217-228

Chan, S. H.; Sutton, S. G. & Yao, L. J. (2009). The paradoxical effects of feedback and reward on decision performance. *Advances in Accounting Behavioral Research*, 12, 109-143

Chenoweth, T.; Dowling, K. L. & St. Louis, R. D. (2003). Convincing DSS users that complex models are worth the effort. *Decision Support Systems*, 1050, 1-12

Condry, J. (1977). Enemies of exploration: Self-initiated versus other-initiated learning. *Journal of Personality and Social Psychology*, 35, 7, 459-477

Creyer, E. H.; Bettman, J. R. & Payne, J. W. (1990). The impact of accuracy and effort feedback and goals on adaptive decision behavior. *Journal of Behavioral Decision Making,* 3, 1-16

Cuccia, A. D.; Hackenbrack, K. & Nelson, M. W. (1995). The ability of professional standards to mitigate aggressive reporting. *Accounting Review*, 70, 227-248

Daniel, T. L. & Esser, J. K. (1980). Intrinsic motivation as influenced by rewards, task interest, and task structure.*Journal of Applied Psychology*, 65, 5, 566-573

Davis, C. E. (1994). Presentation format, information load, and time pressure effects on the consistent application of a decision rule. Working paper, Baylor University, Waco, TX

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 3, 319-339

Davis, F. D.; Bagozzi, R. P. & Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of Applied Social Psychology*, 22, 14, 1111-1132

Davis, F. D. (1993). User acceptance of information technology: System characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies*, 38, 3, 475-487

deCharms, R. (1968). *Personal causation: The internal affective determinants of behavior*. New York Academic Press, New York, NY

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18, 105-115

Deci, E. L. (1972). Effects of contingent and non-contingent rewards and controls on intrinsic motivation. *Organizational Behavior and Human Performance*, 8, 217-229

Deci, E. L. (1998). The relation of interest to motivation and human needs – The self-determination theory viewpoint. In: *Interest and learning: Proceedings of the Seeon Conference on Interest and Gender*, Hoffmann, L.; Krapp, A.; Renninger, K. & Baumert, J. (Ed.), 146-163, Kiel, Germany

Deci, E. L.; Koestner, R. & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125,6, 627-668

Deci, E. L. & Ryan, R. M. (1985). The General Causality Orientations Scale: Self-determination in personality. *Journal of Research in Personality*, 19, 109-134

Deci, E. L. & Ryan, R. M. (1987). The support of autonomy and the control of behavior. *Journal of Personality and Social Psychology*, 53, 6, 1024-1037

Deci, E. L. & Ryan, R. M. (1985). The General Causality Orientations Scale: Self-determination in personality. *Journal of Research in Personality*, 19, 109-134

Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly*, 11, 135-172

Eccles, J. S.; Adler, T. F.; Futterman, R.; Goff, S. B.; Kaczala, C. M.; Meece, J. L. & Midgley, C. (1983). Expectancies, values, and academic behaviors, In: *Achievement and Achievement Motives*, J. T. Spence (Ed.), 75-146, W. H. Freeman and Company, New York, NY

Eining, M. M. & Dorr, P. B. (1991). The impact of expert system usage on experiential learning in an auditing setting. *Journal of Information Systems*, 5, 1-16

Eining, M. M.; Jones, D. R. & Loebbecke, J, K. (1997). Reliance on decision aids: An examination of auditors' assessment of management fraud, *Auditing: A Journal of Practice and Theory*, 16, 2, 1-19

Eisenberger, R. & Cameron, J. (1996). Detrimental effects of reward: Reality or myth? *American Psychologist*, 51, 11, 1153-1166

Fogg, B. J. & Nass, C. (1997). Silicon sycophants: The effects of computers that flatter. *International Journal of Human Computer Studies*, 46, 551-561

Fried, Y. & Ferris, G. R. (1987). The validity of the job characteristics model: A review and a meta-analysis, *Personnel Psychology*, 40, 287-322

Gibson, D. L. (1994). The effects of screen layout and feedback type on productivity and satisfaction of occasional users. *Journal of Information Systems*, 8, 2, 105-114

Glover, S. M.; Prawitt, D. F. & Spilker, B. C. (1997). The influence of decision aids on user behavior: Implications for knowledge acquisition and inappropriate reliance. *Organizational Behavior and Human Decision Processes*, 72, 2, 232-255

Hackbarth, G.; Grover, V. & Yi, M. Y. (2003). Computer playfulness and anxiety: Positive and negative mediators of the system experience effect on perceived ease of use. *Information and Management,* 40, 221-232

Hackenbrack, K., & Nelson, M. W. (1996). Auditors incentives and their application of financial accounting standards. *Accounting Review*, 71, 43-59

Hackman, J. R. & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory, *Organizational Behavior and Human Performance*, 16, 250-279

Hackman, J. R. & Oldham, G. R. (1980). *Work redesign reading*, MA Addison-Wesley

Harackiewicz, J. M. (1979). The effects of reward contingency and performance feedback on intrinsic motivation. *Journal of Personality and Social Psychology*, 37, 1352-1363

Harackiewicz, J. M.; Manderlink, G. & Sansone, C. (1984). Rewarding pinball wizardry: The effects of evaluation on intrinsic interest. *Journal of Personality and Social Psychology*, 47, 287-300

Harackiewicz, J. M. & Sansone C. (2000). Rewarding competence: The importance of goals in the study of intrinsic motivation. In: *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance*, C. Sansone & J. M. Harackiewicz (Ed.), 79-103, Academic Press, San Diego, CA

Hard, N. J. & Vanecek, M. T. (1991). The implications of tasks and format on the use of financial information. *Journal of Information Systems*, 5, 35-49

Harter, S. (1981). A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components. *Developmental Psychology*, 17, 3, 300-312

Hennessey, B. A. & Amabile, T. M. (1988). In: *The Nature of Creativity: Contemporary Psychological Perspectives*, R. J. Sternberg (Ed.), 11-38, Cambridge University Press New York, NY

Igbaria, M.; Zinatelli, N.; Cragg, P. & Cavaye, A. L. M. (1997). Personal computing acceptance factors in small firms: A structural equation model. *MIS Quarterly*, 21, 3, 279-305

Jacobs, J. E. & Eccles, J. S. (2000). Parents, task values, and real-life achievement-related choices. In: *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance*, Sansone, C. & Harackiewicz, J. M. (Ed.), 408-439, Academic Press, San Diego, CA

Jarvenpaa, S. L. (1989). The effect of task demands and graphical format on information processing strategies. *Management Science*, 35, 285-303

Johnson, E. & Payne, J. (1985). Effort and accuracy in choice. *Management Science,* 31, 395-415

Johnson, E.; Payne, J. & Bettman, J. (1988). Information displays and preference reversals. *Organizational Behavior and Human Decision Processes*, 42, 1-21

Johnson, D.; Gardner, J. & Wiles, J. (2004). Experience as a moderator of the media equation: The impact of flattery and praise. *International Journal of Human-Computer Studies*, 61, 3, 237-258

Johnson, R. D.; Marakas, G. M. & Palmer, J. W. (2006). Differential social attributions toward computing technology: An empirical investigation. *International Journal of Human-Computer Studies*, 64, 5, 446-460

Johnson, V. E. & Kaplan, S. E. (1991). Experimental evidence on the effects of accountability on auditor judgments. *Auditing: A Journal of Practice & Theory*, 10, 98-107

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 2, 263-292

Kennedy, J. (1993). Debiasing audit judgment with accountability: A framework and experimental results. *Journal of Accounting Research*, 31, 2, 231-245

Kim, S. S. & Malhotra, N. K. (2005). Predicting system usage from intention and past use: Scale issues in the predictors. *Decision Sciences*, 36, 1, 187-196

Klein, B. D.; Goodhue, D. L. & Davis, G. B. (1997). Can humans detect errors in data? Impact of base rates, incentives, and goals. *MIS Quarterly*, 21, 2, 169-194

Kruglanski, A. W.; Friedman, I. & Zeevi, G. (1971). The effects of extrinsic incentive on some qualitative aspects of task performance. *Journal of Personality*, 39, 606-617

Kuvaas, B. (2006). Performance appraisal satisfaction and employee outcomes: Mediating and moderating roles of work motivation. *International Journal of Human Resource Management*, 17, 3, 504-522

Lawler, E. E. (1973). *Motivation in Work Organizations*. Brooks Cole, Monterey, CA

Lederer, A. L.; Maupin, D. J.; Sena, M.P. & Zhuang, Y. (2000). The technology acceptance model and the World Wide Web. *Decision Support Systems*, 29, 269-282

Legris, P.; Ingham, J. & Collerette, P. (2003). Why do people use information technology? A critical review of the technology acceptance model. *Information and Management*, 40, 191-204

Lepper, M. R.; Greene, D. & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology*, 28, 1, 129-137

Lepper, M. R. (1981). Intrinsic and extrinsic motivation in children: Detrimental effects of superflous social controls. In: *Aspects of the Development of Competence: The Minnesota Symposium on Child Psychology*, W. A. Collins, (Ed.), 55-214, Erlbaum, Hillsdale, NJ

Libby, R. (1992). The role of knowledge and memory in audit judgment. In: *Judgment and Decision-making Research in Accounting and Auditing*, Ashton, R. H. & Ashton, A. H. (Ed.), 176-206, Cambridge University Press, New York, NY

Lightner, N.; Bose, I. & Salvendy, G. (1996). What is wrong with the World Wide Web? A diagnosis of some problems and prescription of some remedies. *Ergonomics,* 39, 8, 995-1004

Lilien, G. L.; Rangaswamy, A.; Van Bruggen, G.H. & Starke, K. (2004). DSS effectiveness in marketing resource allocation decisions: Reality vs. perception, *Information System Research*, 15, 3, 216-235

Limayem, M.; Khalifa, M. & Frini, A. (2000). What makes consumers buy from Internet? A longitudinal study of online shopping, *IEEE Transactions on Systems, and Cybernetics*, 30, 4, 421-432

Lin, Y. G.; McKeachie, W. J. & Kim, Y.C. (2001). College student intrinsic and/or extrinsic motivation and learning, *Learning and Individual Differences,* 13, 3, 251-258

Lord, A. T. (1992). Pressure: A methodological consideration for behavioral research in auditing. *Auditing: A Journal of Practice and Theory*, 90-108

Lucas, H. C. Jr. & Spitler, V. K. (1999). Technology use and performance: A field study of broker workstations. *Decision Sciences*, 30, 2, 291-311

McGraw, K. O. & McCullers, J.C. (1979). Evidence of a detrimental effect of extrinsic incentives on breaking a mental set. *Journal of Experimental Social Psychology*, 15, 285-294

Nielsen, J. (2000). *Designing Web Usability,* New Riders Publishing, Indianapolis, IN

Novak, T. P.; Hoffman, D. L. & Yung, Y. F. (2000). Measuring the customer experience in online environments: A structural modeling approach

Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin,* 92, 382-402

Payne, J. W.; Bettman, J. R. & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Human Learning, Memory, and Cognition,* 14, 534-552

Payne, J. W.; Bettman, J. R. & Johnson, E. J. (1993). *The Adaptive Decision Maker*. Cambridge University Press

Peecher, M. E. (1996). The influence of auditors' justification processes on their decisions: A cognitive model and experimental evidence. *Journal of Accounting Research*, 34, 1, 125-140

Piaget, J. (1981). *Intelligence and Affectivity: Their Relationship during Child Development*. Palo Alto: Annual Reviews

Pintrich, P. R. & Schrauben, B. (1992). Students' motivational beliefs and their cognitive engagement in classroom academic tasks. In: *Student Perceptions in the Classroom*, Schunk, D. H. & Meece, J. L. (Ed.), 149-183, Erlbaum, Hillsdale, NJ

Piramuthu, S. (2003). On learning to predict Web traffic. *Decision Support Systems*, 35, 213-229

Pitkow, J. E. & Kehoe, C. M. (1996). Emerging trends in the WWW user population. *Communications of the ACM*, 39, 6, 106-108

Rigby, C. S.; Deci, E. L.; Patrick, B. C. & Ryan, R. M. (1992). Beyond the intrinsic-extrinsic dichotomy: Self-determination in motivation and learning. *Motivation and Emotion*, 16, 3, 165-185

Roth, E. M.; Bennett, K. B. & Woods, D. D. (1987). Human interaction with an "intelligent" machine. *International Journal of Man-Machine Studies*, 27, 479-525

Roy, M. C. & Lerch, J. F. (1996). Overcoming ineffective mental representations in base-rate problems. *Information Systems Research*, 7, 2, 233-247

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43, 450-461

Ryan, R. M.; Mims, V. & Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology,* 45, 736-750

Ryan, R. M. & Deci, E. L. (1996). When paradigms clash: Comments on Cameron and Pierce's claim that rewards do not undermine intrinsic motivation. *Review of Educational Research*, 66, 33-38

Ryan, R. M. & Deci, E. L. (2000). When rewards compete with nature: The undermining of intrinsic motivation and self-regulation. In: *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance*, Sansone, C. & Harackiewicz, J. M. (Ed.), 257-307, Academic Press, San Diego, CA

Saeed, K. A.; Hwang, Y. & Yi, M. Y. (2003). Toward an integrative framework for online consumer behavior research: A meta-analysis approach. *Journal of End User Computing,* 15, 4, 1-26

Sansone, C. & Harackiewicz, J. M. (1998). "Reality" is complicated: Comment on Eisenberger & Cameron. *American Psychologist*, 53, 673-674

Sansone, C. & Smith, J. L. (2000). Interest and self-regulation: The relation between having to and wanting to. In: *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance*, Sansone, C. & Harackiewicz, J. M. (Ed.), 341-372, Academic Press, San Diego, CA

Shalley, C. E. & Perry-Smith, J. E. (2001). Effects of social-psychological factors on creative performance: The role of informational and controlling expected evaluation and modeling experience. *Organizational Behavior and Human Decision Processes*, 84, 1, 1-22

Silver, M. S. (1988). On the restrictiveness of decision support systems, *Proceeding of IFIP WG 8.3 Working Conf*, pp. 259-270, North Holland, Como, Italy, Elsevier Science Publishers B. V.

Silver, M. S. (1990). Decision support systems: Directed and nondirected change. *Information Systems Research*, 1, 1, 47-70

Simon, H. A. (1973). The structure of ill structured problems. *Artificial Intelligence*, 4, 181-201

Smith, C. A. P.; Arnold, V. & Sutton, S. G. (1997). The impact of time pressure on decision-making for choice and judgment tasks. *Accounting and Business Review*, 365-383

Stone, D. N. & Schkade, D. (1991). Numeric and linguistic information representation in multiattribute choice. *Organizational Behavior and Human Decision Processes,* 49, 42-59

Stone, D. N. (1995). The joint effects of DSS feedback and users' expectations on decision processes and performance. *Journal of Information Systems*, 9, 1, 23-41

Stone, D. N. & Kadous, K. (1997). The joint effects of task-related negative affect and task difficulty in multiattribute choice. *Organizational Behavior and Human Decision Processes*, 70, 2, 159-174

Subramanian, G. H. (1994). A replication of perceived usefulness and perceived ease of use measurement. *Decision Sciences*, 25, 863-874

Szajna, B. (1993). Determining information systems usage: Some issues and examples. *Information Management*, 25, 3, 147-154

Tang, S-H. & Hall, V. C. (1995). The overjustification effect: A meta-analysis. Applied *Cognitive Psychology*, 9, 365-404

Tarafdar, M. & Zhang, J. (2005). Analyzing the Influence of Website Design Parameters on Website Usability, *Information Resources Management Journal*, 18, 4, 62-80

Thomas, K. W. & Velthouse, B. A. (1990). Cognitive elements of empowerment: An interpretive model of intrinsic task motivation. *Academy of Management Review,* 15, 4, 666-681

Thomas, J. D. E. (1996). The importance of package features and learning factors for ease of use. *International Journal of Human-Computer Interaction*, 8, 2, 165-187

Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research*, 64, 37-54

Todd, P. & Benbasat, I. (1992). The use of information in decision making: An experimental investigation of the impact of computer-based decision aids. *MIS Quarterly*, 16, 373-393

Todd, P., & Benbasat, I. (1994). The influence of decision aids on choice strategies: An experimental analysis of the role of cognitive effort. *Organizational Behavior and Human Decision Processes*, 60, 36-74

Todd, P. & Benbasat, I. (1994a). The influence of decision aids on choice strategies. *Organizational Behavior and Human Decision Processes,* 60, 36-74

Todd, P. & Benbasat, I. (1994b). The influence of decision aids on choice strategies under conditions of high cognitive load. *IEEE Transactions on Systems, Man, and Cybernetics,* 24, 4, 537-547

Todd, P. & Benbasat, I. (1996). The effects of decision support and task contingencies on model formulation: A cognitive perspective. *Decision Support Systems*, 17, 241-252

Todd , P. & Benbasat, I. (1999). Evaluating the impact of DSS, cognitive effort, and incentives on strategy selection, *Information System Research*, 10, 4, 356-374

Todd, P. & Benbasat, I. (2000). Inducing compensatory information processing through decision aids that facilitate effort reduction: An experimental assessment. *Journal of Behavioral Decision Making,* 13, 91-106

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review,* 79, 4, 281-299

Tzeng, J. (2004). Toward a more civilized design: Studying the effects of computers that apologize. *International Journal of Human-Computer Studies*, 61, 3, 319-345

Umanath, N. S.; Scamell, R. W. & Das, S. R. (1990). An examination of two screen/report design variables in an information recall context. *Decision Sciences*, 21, 216-240

Utman, C. H. (1997). Performance effects of motivational state: A meta-analysis. *Personality and Social Psychology Review,* 1, 170-182

Vansteenkiste, M.; Simons, J.; Lens, W.; Sheldon, K. M. & Deci, E. L. (2004). Motivating learning performance, and persistence: The synergistic effects of intrinsic goal contents and autonomy-supportive contexts, *Journal of Personality and Social Psychology*, 87, 246-260

Venkatesh, V. & Speier, C. (1999). Computer technology training in the workplace: A longitudinal investigation of the effect of mood. *Organizational Behavior and Human Decision Processes*, 79, 1, 1-28

Venkatesh, V. & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46, 2, 186-204

Venkatesh, V.; Morris, M. G.; Davis, G. B. & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27, 3, 425-478

Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, 22, 219-240

Vessey, I. & Galletta, D. (1991). Cognitive fit: An empirical study of information acquisition. *Information Systems Research*, 2, 1, 63-84

Wang, J. H. Y. & Guthrie, J. T. (2004). Modeling the effects of intrinsic motivation, extrinsic motivation, amount of reading achievement on text comprehension between US and Chinese student. *Reading Research Quarterly*, 39, 162-186

White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review,* 66, 297-333

Wigfield, A. & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review*, 12, 365-510

2

New Architecture for Intelligent Multi-Agents Paradigm in Decision Support System

Noor Maizura Mohamad Noor and Rosmayati Mohemad
Universiti Malaysia Terengganu
Malaysia

1. Introduction

In recent years, intelligent agent concepts have been applied in decision support systems (DSS) for business users (Bose & Sugumaran, 1999). Another research was done by Rahwan et al. (2004) where they used intelligent agents in one to many e-commerce negotiation to automate decisions making processes. Computer technology is increasingly being used to support executive decision-making in DSS environment (Moynihan et al., 2002). DSS are computer programs that aid users in a problem solving or decision-making environment. These systems employ data models, algorithms, knowledge bases, user interfaces, and control mechanisms to support a specific decision problem (Barkhi et al., 2005). Various researches have shown the uses of DSS in order to handle complex decision modeling and management process. Based on my readings so far, there are no literature discussing and applying intelligent multi-agent architecture in DSS especially for distributed environment.

In this research, intelligent multi-agent technology is proposed in developing DSS to enhance the system to be able to work in dynamic environments and support the adaptability of the system. Agent is defined as a software abstraction and logical model. The idea is that agents are not strictly invoked for a task, but activate themselves. Related and derived concepts include intelligent agents where they have the ability to adapt on the new situation with some aspect of learning and reasoning. Another derived concept is multi-agent systems that involve distributed agents that do not have the capabilities to achieve an objective alone and thus must communicate. In the environment of distributed system, agents play a major role in assisting a real user in making decisions where these agents are given the authority to communicate to each other in order to achieve the objective.

2. Motivation

Research that has been done by Parunak (2006) proposed multi-agent support system that can adapt to a user's resource constraints, resource priorities, and content priorities in a dynamic environment. Here, multi-agents that cooperates each other will consider the preferences and constraints of a user while gathering and assembling information.

The construction industry involves multiple parties such as clients, consultants and contractors. Project success relies heavily on the timely transfer of information among these parties (Kashiwagi, 2002). Projects involve a large number of organisations that may be geographically dispersed.

The planning of a construction project is among the most challenging tasks faced by a project team (clients, consultants and constructors). Decisions made during this stage have a tremendous impact on the successful execution of the project from its early conceptual phases, through to project construction and completion. The construction industry is seen by many as being backward in its deployment and use of IT (Huang, 2001). Application of IT has been comparatively slow and only very few construction companies have a comprehensive and integrated IS to support its core business.

Many businesses use the Internet as a new technology platform to build a range of new products and services, and even to redesign their communication system and services (Ngai et al., 2003). With great advances in Internet and WWW technologies, various attempts have been made to implement Web-based DSSs for different applications in the areas of sales, design and manufacturing (Smith & Wright, 1996). However, very little has been reported on the use of the Web-based technologies for selecting the best alternative in the construction businesses.

The causes of time and cost problems in construction management projects can be traced back to poor coordination caused by inadequate, inappropriate, inaccurate, inconsistent, or late information, or a combination of them all (Deng et al., 2001). However, the physical distance between the parties further contributes to the communication barrier and thus a key factor in inhibiting information transfer. The availability of timely and accurate information is also important for all parties as it forms the basis on which decisions are made and such that concrete progress could be achieved.

The above issues arise despite the remarkable advancement in information management, handling, storage and exchange techniques. Improving communication among parties is thus a key factor that could lead to the success or failure of decision making processes. Furthermore, IT can be seen as having a mediating effect on communication, leading to new patterns of communication or changes in the content and quantity of existing kinds of communications. A coordination technology can encourage participants to control the way a decision is made, monitor results, improve productivity of meetings and support the team's decision-making process (Yazici, 2002).

With the advent of the Internet, collaborative e-business has been attracting more and more attention from both the academia and industry. Various technologies are being developed to support collaborative e-business, such as customer relationship management (Siebel, 2001), supply chain management (Indent, 2001, Commonce One, 2001), electronic market (Ariba, 2001), automated negotiation and auction systems (Huang, 2001, Kumar & Feldman, 1998, Su et al., 2001) and DSS in various applications (Mysiak et al., 2005).

3. Decision support system

Computer technology is increasingly being used to support executive decision-making (Moynihan et al., 2002). Nemati et al. (2002) explains that decision-making is the ability to make the 'right' decisions. The tendency is to focus on decision makers (DM's) moment of choice even though the process is complex (Simon, 1977). This focus is however not limited to a DM's preceding and subsequent decision-making processes as some DMs bear the responsibilities for decisions that were made even by their subordinates and groups. Simon (1977) identified three steps in a decision process: (a) *intelligence* as searching the environment for conditions requiring decision, (b) *design* as inventing, developing, and analysing possible course of action, and (c) *choice* as selecting a particular course of action.

A DM receives information in various formats such as printed materials, graphics, verbal and visual observation. While a computer can extend the memory of a human being, humans are often not very good information processors (Chen & Lee, 2003). The condition when information overload occurs can cause a rapid and severe degradation in the performance of decision-making. Good information is, therefore necessary but not sufficient for a good decision. A DSS can also be designed to support the creative and intuitive aspects of decision-making (Pearson & Shim, 1994, Phillips-Wren et al., 2004).

DSSs can also increase managerial effectiveness by improving personal efficiency, expediting problem solving, facilitating interpersonal communication, promoting learning, especially about how the system works, and increasing organisational control (Alter, 1980). These potential benefits lead many researchers and practitioners to believe that DSSs can be a powerful strategic weapon for organisations.

Bharati & Chauhury (2004) projected the organisational and individual impacts of DSSs and they categorised the impact and benefit into organisational and individual impacts as well. The former reflects on the structure, centralisation of authority, power and status. The latter reflects on productivity and job satisfaction. This factor constitutes the manager's satisfaction with the uses and benefits of its DSS for example in supporting decision making processes.

4. Intelligent agent technology

Intelligent agents are software or hardware entities that perform a set of tasks on behalf of a user with some degree of autonomy (Barley & Kasabov, 2005). Several applications in a variety of domains including: Internet-based information systems, adaptive (customizable) software systems, autonomous mobile and immobile robots, data mining and knowledge discovery, smart systems (smart homes, smart automobiles, etc.), decision support systems, intelligent design and manufacturing systems. Current research on intelligent agents and multi-agent systems builds on developments in several areas of computer science including: artificial intelligence (especially agent architectures, machine learning, planning, distributed problem-solving), information retrieval, database and knowledge-base systems, and distributed computing (Godoy et al., 2004).

Distributed intelligent multi-agent systems offer modular, flexible, scalable and generalisable algorithms and systems solutions for information retrieval, extraction, fusion, and data-driven knowledge discovery using heterogeneous, distributed data and knowledge sources in information rich, open environments (Parunak, 2006). Such systems consist of multiple interacting intelligent software agents. Such as reactive, proactive, anticipatory, goal-driven, adaptive, reflective, introspective, knowledge-seeking, autonomous, interactive, communicative, collaborative agents and multi-agent systems find applications in Internet-based information systems, adaptive (customizable) software systems, large scale data-driven knowledge discovery from heterogeneous, distributed data and knowledge sources, collaborative scientific discovery (e.g., in bioinformatics), intelligent decision support systems (e.g., monitoring and control of complex, distributed, dynamic systems) (Jennings et al., 1998).

Intelligent agents play the role of assistants by letting managers delegate work that they could have done to these agents. Agent technology is finding its way into many new systems, including DSS, where it performs many of the necessary decision support tasks formerly considered a uniquely human activity. Intelligent agents are useful in automating repetitive tasks, finding and filtering information, intelligently summarizing complex data, and so on; but more importantly, just like their human counterparts, they have the capability to learn from the managers and even make recommendations to them regarding a particular course of action. Agents utilize several artificial intelligence techniques such as machine learning, and inductive and deductive reasoning to exhibit the "intelligent" behaviours.

There are many definitions of intelligent agents found in the literature. The definition, by Maes (1994), that "intelligent agents are software entities that carry out some set of operations on behalf of a user or another program with some degree of independence or autonomy, and in doing so, employ some knowledge or representation of the user's goals or desires," is the most appropriate for our purposes. Agents possess several common characteristics, such as their ability to communicate, cooperate, and coordinate with other agents in a multiple-agent system. Each agent is capable of acting autonomously, cooperatively, and collectively to achieve the collective goal of a system.

The coordination capability helps manage problem solving so that cooperating agents work together as a coherent team. The coordination is achieved, for example, by exchanging data, providing partial solution plans, and handling constraints among agents.

5. Objectives

The aim of the research is to propose a new architecture for a multi-agent based DSS in distributed environment.

The objectives of this research are:

1. To determine the characteristic of conventional DSS system.
2. To construct new architecture of an intelligent multi-agent of DSS.
3. To test and evaluate the applicability of the approach using a real-world scenario by using proof-of-concept and operational prototype system.

6. DSS design and functions

DSS consists of the following agents: a) contractors interface agent, b) client interface agent, c) coordinator agent, d) report agent, and d) database agent. The whole architecture is depicted in Figure 1. The overall DSS agent architecture consists of three high-level modules: a) interface module, b) process module, and c) knowledge module. The interface module deals with is publicly visible to other agents and users (consultants and clients). It provides mechanisms for interacting with the agent and supports inter-agent communication and collaboration. The process module and knowledge module are restricted only to the agent that is, other agents or users cannot directly manipulate the contents of these modules without access privileges.

The process module contains methods and heuristics that implement a variety of functions and processes using which the agent can respond to requests from other agents or users. Thus, the process module basically provides the services and computations that may be necessary in solving a particular problem. The knowledge module contains domain-specific and domain-independent knowledge relevant to problem solving. The detailed design of the three above-mentioned DSS agents, in terms of the three high level modules, is described section 8.

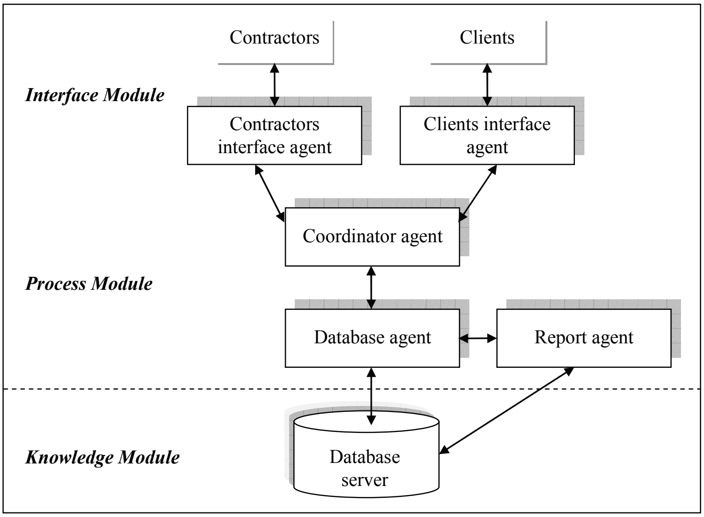


Fig. 1. DSS Based on Intelligent Multi-Agents Paradigm

7. Functions of each agent

There are several agents involved in this research such as user interface agent, coordinator agent, and database agent.

a. User interface agent

The user interface agent is divided into two types of user; a) contractors and, b) clients. Each of different users has their own interface agent. Generally, user interface agent provides a Web interface for the users to interact with DSS and to help him/her deal with several online forms, perform uploading/downloading related documents, and do data analysis activities. The user can provide a general description of the problem at hand in terms of high-level goals and objectives, or provide specific details about the data analysis or mining task to be performed. The user interface agent is responsible for receiving user specifications and delivering results. It also keeps track of user preferences.

The interface module of the user interface agent contains methods for inter-agent communication as well as getting input from the user. Its process module contains scripts and methods for capturing the user input and communicating it to the DSS coordinator agent.

The functions of the user interface agent are providing Web interface for user interaction, Web page for user input and problem description, provide parameters to use, Web page for status information - feedback providing states of various processes, Web page containing final results, dynamically creating HTML documents with special formatting needs, communicating the user input to the DSS coordinator agent, capturing user activities and preferences and create user profiles

b. Coordinator agent

The coordinator agent is responsible for coordinating the various tasks that need to be performed in cooperative problem solving. After receiving the user input from the interface agent, the coordinator agent identifies the relevant criteria, determines the alternative that need to be evaluated and generates a plan of action such as ranking of the alternatives. These alternatives may include identifying the relevant data sources, requesting services from other agents, and generating reports. The interface module of the coordinator agent is responsible for inter-agent communication.

The process module contains methods for control and coordination of the various tasks as well as generating the task sequence. The sequence of tasks to be executed is created utilizing specific formula stored in the knowledge module using a rule-based approach. The knowledge module also contains meta-knowledge about the capabilities of other agents in the federation, available data sources and databases. The coordinator agent may seek the services of a group of agents and synthesize the final result.

The functions of the coordinator agent are from user input, identify high-level objectives based on these objectives, identify tasks, generate "task sequence" and delegate actions to corresponding agents, provide intermediate feedback to user, synthesize and generate final result, perform the calculation/evaluation of problem-specific information.

c. Database agent

The database agent is responsible for keeping track of what data are stored in database. It provides predefined and ad hoc retrieval capabilities. It is also responsible for retrieving the necessary data requested by the data mining agent in preparation for a specific data mining operation. The database agent takes into account the heterogeneity of the databases that may exist within the organization, and resolves conflicts in data definition and representation.

The interface module of the database agent provides not only the public interface for inter-agent communication, but also to existing databases. This improves inter-operability and enables users to gain access to a variety of data sources which otherwise might be inaccessible. The process module provides facilities for ad hoc and predefined data retrieval. Based on the user request, appropriate queries are generated and executed against the data warehouse. The results of these queries are communicated back to the user or other agents. The knowledge module contains meta-data information, including the local schemas and a global schema. These schemas are used in generating the necessary queries for data retrieval.

The functions of the database agent are inter-agent message communication, provide interface to databases, application program interface (API) to commercial database products, Ad hoc and predefined data retrieval, maintain local and global schema and formatting query outputs based on user needs.

8. Importance of our methodology

The proposed architecture for a multi-agent based DSS in distributed environment is organised around the decision-making model. Unlike other approaches, this architecture uses the multi-discipline concepts in computer science such as agent-programming, decision support system, distributed system and employs tendering processes as a case study in order to test the performance and usability of the framework. There are three development phases such as:

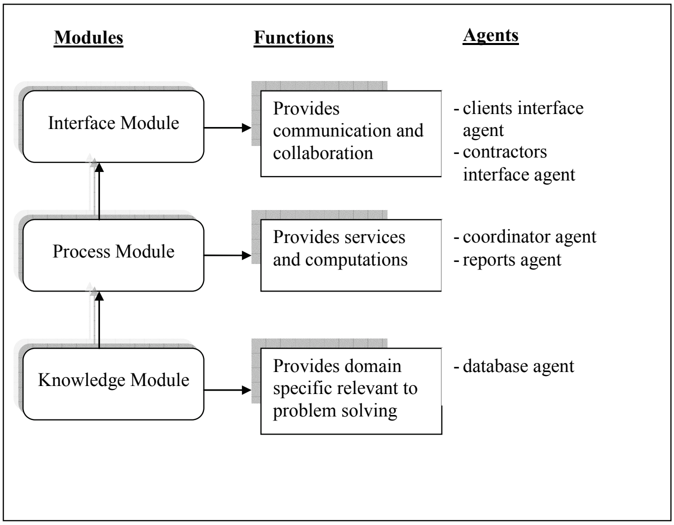

Modules
Functions
Agents
Interface Module
Provides communication and collaboration
- clients interface agent
- contractors interface agent
Process Module
Provides services and computations
- coordinator agent
- reports agent
Knowledge Module
Provides domain specific relevant to problem solving
- database agent

Fig. 2. Functions on Interface, Process and Knowledge Module

Phase 1: Construct architecture of a multi-agent based DSS in distributed environment. The architecture consists of two parts: the object (passive) and the agents (active). The major components of the architecture include: interface, data, models and agents. Interface provides interaction between users and the proposed application. The data and model are components containing data (collected online from other databases, or the users) and models relevant to decision making.

Phase 2: Design agent-programming in distributed environment by developing the three main agents: a) user interface agent, b) coordinator agent, c) database agent.

Phase 3: Evaluate the performance and usability of the approach, Web-based management system for the decision making process will be embedded as a research prototype.

9. Advantages

The government, construction industry and clients are all seeking to bring about some changes in the construction industry in order to improve quality, competitiveness and profitability and to increase value to clients and contractors.

The Web-based DSS has shown to have tremendous potential not only in adding value to registered clients and contractors in construction companies but also to the whole construction industry. The Web-based DSS is very much concerned with the display of information about a tender project and also the process of placing the tender. Many advantages can be achieved such as it enables a DM to accomplish a task more effectively, it reduces costs and time taken, and it can be customised to other applications.

The several Web-based DSS applications have been used in some countries and have proven to be effective in facilitating a faster and more efficient use of processes. However it is still a new research area in Malaysia where there are still many questions to be answered and challenges to be addressed. With the fast expansion of the Internet developing the Web-based DSS has gained more and more interest from many researchers all the over the world.

The Web-based DSS technologies will become an important management tool to enhance the performance of decision making processes. Although the development, deployment and thus the use of the Web-based DSS is still very much in its infancy in the construction industry, we expect to see a rapid increase in its functionality that will assist both clients and contractors in the electronic environments over the next few years.

The motivation for this architecture stems from the need to support the DMs in their task of solving complex problems such as those that exist in decision making processes. It emphasises ease of use, user friendliness, flexibility and adaptability. The DSS can be used in other contexts for example to be commercialised as a Web-based application for trading any goods or services in the world where a structured approach to supporting decision-making remain as the main concepts. Or in other words it can also be used as an e-commerce tool.

10. Expected outcomes

One of the major aims of this research is to provide quality decision support and performance assessments tools and services through the use of cutting edge techniques in decision sciences and computer technologies. We wish to see that DMs and decision-making processes in any field of human activities can be supported in a way that is systematic, rational, transparent, efficient and reliable.

i. A new documented architecture for a multi-agent based DSS in distributed environments.

ii. Robust prototype with a user-friendly environment. Users can come up with his/her own solutions, or modify solutions generated by the agents.

iii. The proper combination of agent-based technology for DSS in distributed environment will provide a powerful tool to support decision-making in distributed environment.

11. Conclusion

Decision-making for choosing the right alternatives for the right activities is immensely complex, involving various processes and communication mechanisms among them. We feel that the significant conclusions that can be drawn from our research are: (1) Building decision-support is much more complicated than merely applying a technical tool to solve a well-defined decision problem. The DSS is the application of the most potent forces in contributing to Web-based problem domain. Decision making problem that is based on the use of IT to achieve the primary principle of perfect competition for example to increase transparency of selecting alternatives. (2) Real time decision making process evaluation together with employing intelligent multi-agent will expected to provide a better way in order to make the processes more efficient, i.e. faster and lower cost.

12. References

Alter, S. L. (1980) *Decision Support Systems: Current Practices And Continuing Challenges*, Massachusetts, Addison Wesley.

Ariba (2001) Ariba Marketplace.

Barkhi, R., Rolland, E., Butler, J. & Fan, W. (2005) Decision Support System Induced Guidance For Model Formulation And Solution. *Decision Support Systems,* 40, 269-281.

Barley, M. & Kasabov, N. (2005) *Intelligent Agents And Multi-Agent Systems: 6th Pacific Rim International Workshop On Multi-Agent*, Springer

Bharati, P. & Chaudhury, A. (2004) An Empirical Investigation Of Decision-Making Satisfaction In Web-Based Decision Support Systems. *Decision Support Systems,* 37, 187-197.

Bose, R. & Sugumaran, V. (1999) Application Of Intelligent Agent Technology For Managerial Data Analysis And Mining. *Data Base,* 30, 77-94.

Chen, J. Q. & Lee, S. M. (2003) An Exploratory Cognitive Dss For Strategic Decision Making. *Decision Support Systems,* 36, 147-160.

Commonce One (2001) Commerce One E-Business Solutions.

Deng, Z. M., Li, H., Tam, C. M., Shen, Q. P. & Love, P. E. D. (2001) An Application Of The Internet-Based Project Management System. *Automation In Construction,* 10, 239-246.

Godoy, D., Schiaffino, S. & Amanda, A. (2004) Interface Agents Personalizing Web-Based Tasks. *Cognitive Systems Research,* 5, 207-222.

Huang, C. (2001) A Web-Based Negotiation Server For Supporting Electronic Commerce. *Department Of Computer And Information Science And Engineering.* Usa, University Of Florida.

Indent (2001) Cost-Benefit Analysis Of Supply Chain Alternatives Through Simulation. *Innovative Technologies.*

Jennings, N. R., Sycara, K. & Wooldridge, M. A. (1998) Roadmap Of Agent Research And Development. *Autonomous Agents And Multi-Agent Systems Journal,* 1, 7-38.

Kashiwagi, D. T. (2002) Project Ma Case Study Of Potential Impact Of Subjective Decision Making On Construction Performance. *The Journal Of Construction Procurement,* 8, 32-41.

Kumar, M. & Feldman, S. (1998) Business Negotiation On The Internet. *Ibm Institute For Advanced Commerce (Iac) Report.*

Moynihan, G. P., Purushothaman, P., Mcleod, R. W. & Nichols, W. G. (2002) Dssalm: A Decision Support System For Asset And Liability Management. *Decision Support Systems,* 33, 23-38.

Mysiak, J., Giupponi, C. & Rosato, P. (2005) Towards The Development Of A Decision Support System For Water Resource Management. *Environmental Modelling & Software,* 20, 203-214.

Nemati, H. R., Steiger, D. M., Iyer, L. S. & Herschel, R. T. (2002) Knowledge Warehouse: An Architectural Integration Of Knowledge Management, Decision Support, Artificial Intelligence And Data Warehousing. *Decision Support Systems,* 33, 143-161.

Ngai, E. W. T., Cheng, T. C. E. & Lee, C. M. Y. (2003) Development Of A Web-Based System For Supporting Sales In A Mineral Water Manufacturing Firm: A Case Study. *International Journal Of Production Economics,* 83, 153-167.

Parunak, H. V. D. (2006) *Evolving Swarming Agents In Real Time Book Series Genetic Programming* vol.9. US:Springer.

Pearson, M. J. & Shim, J. P. (1994) An Empirical Investigation Into Decision Support Systems Capabilities: A Proposed Taxonomy. *Information & Management,* 27, 45-57.

Phillips-Wren, G. E., Hahn, E. D. & Forgionne, G. A. (2004) A Multiple-Criteria Framework For Evaluation Of Decision Support Systems. *Omega,* 32, 323-332.

Rahwan, T., Rahwan, T., Rahwan, I. & Ashri, R. (Eds.) (2004) *Agent-Based Support For Mobile Users Using Agentspeak(L),* Germany, Lncs. Springer.

Siebel, S. (2001) E-Business Architecture.

Simon, H. A. (1977) *The New Science Of Management Decision,* Englewood Cliffs, New Jersey, Prentice Hall, Inc.

Smith, C. S. & Wright, P. K. (1996) Cybercut: A World Wide Web Based Design To Fabrication Tool. *Journal Of Manufacturing Systems,* 15, 432-442.

Su, S. Y. W., Huang, C., Hammer, J., Huang, Y., Li, H., Wang, L., Lui, Y., Pluempitiwiriyawej, C., Lee, M. & Lam, H. (2001) An Internet-Based Server For E-Commerce. *The Very Large Database Journal,* 10, 72-90.

Yazici, H. J. (2002) The Role Of Communication In Organizational Change: An Empirical Investigation. Information & Management, 39, 539-552.

3

A Hybrid Decision Model for Improving Warehouse Efficiency in a Process-oriented View

Cassandra X.H. Tang and Henry C.W. Lau
The Hong Kong Polytechnic University
P.R.China

1. Introduction

The Concept of Supply Chain Management (SCM) has been paid much more attention over the past decades. As one of the essential components of a supply chain, warehousing is valued because of the following major functions: smoothening the material flow; accomadating variability influenced by factors such as product seasonality or transportation scheduling; ensuring proper inventory level by product consolidation; guaranteeing the operation within high tolerances of speed, accuracy and lack of damage (Frazelle, 2002; Christopher, 2005; Harrison & van Hoek, 2005; Baker, 2007; Gu et al., 2007).

According to (Bernardy & Scherff, 1998), all the activities involved in a warehouse can be described by processes and are characterized by entailing a large number of differing, but interdependent sub-processes and many complex influential factors. Since there are diverse functional processes within which different combinations of influencing factors exist, the throughput capacity of the warehouse may be strongly affected, especially when the staffs at the operation level always keep different views upon process parameter settings based on their personal experiences. Hence it is essential to find out the optimal factor settings for the compound functional processes regarding the experts' knowledge so as to make the right strategy, and finally obtain satisfying warehouse operation.

World has witnessed the soaring use of Artificial Intelligence (AI) for operations management (OM) with the purpose of decision support (Kobbacy et al., 2007). Hybrid architecture has become a new field of AI research, in light of the development of the next generation of intelligent systems. Current research in this field mainly concentrates on the marriage of Genetic Algorithms (GA) and Fuzzy Logic (Feng & Huang, 2005; Lau et al., 2009). Exploring the similarities of the essential structures of these two knowledge manipulation methods is where intelligent decision support systems can possibly play an important role. However, such hybrid systems have not shown great significance in the warehousing sector.

This chapter aims to develop a Fuzzy-GA capacity decision model (FGCDM) to enhance rack efficiency in a One-Warehouse, N-Supplier warehouse by taking into consideration the performance metrics and various driving factors of the related processes. The hybrid framework is proposed to enable decision makers to formulate nearly optimal sets of knowledge-based fuzzy rules so as to identify better solutions for fully utilizing the warehouse capacity.

2. Research background

2.1 Performance measurement

The supply chain encompasses a complex set of activities which require a collection of metrics to adequately measure performance (Caplice & Sheffi, 1995; Tompkins & Smith, 1998). (Bowersox & Closs, 1996) states three objectives for developing and implementing performance measurement systems: to monitor historical system performance for reporting, to control ongoing performance so that abnormal processes may be prevented, and to direct the personnel's activities. A conceptual framework for measuring the strategic, tactical and operational level performance in a supply chain is proposed in (Gunasekaran et al., 2001), in which performance measures on warehousing and inventory in a SCM was emphasized. An activity-based approach for mapping and analyzing the practically complex supply chain network is identified in (Chan & Qi, 2003), which can be regarded as a primary step on measuring the performance of processes. (Lohman et al., 2004) points out that by means of local key performance indicators (KPIs), The measurement scheme should be developing at a organization-wide scale. The interplay between organizational experiences and new performance measurement initiatives is highlighted (Wouters & Sportel, 2005). Furthermore, the research work in (Angerhofer & Angelides, 2006) shows how the key parameters and performance indicators are modelled through a case study which illustrates how the decision support environment could be used to improve the performance of a collaborative supply chain. (Niemi, 2009) optimizes the warehousing processes and assesses the related management attributes, realizing the objective of improving the warehousing practices and adopting more sophisticated warehousing techniques supported by knowledge sharing.

In addition, trade-off phenomenon on variable settings is a crucial aspect in the process-oriented supply chain. Leung and Spiring (Leung & Spiring, 2002) have introduced the concept of the Inverted Beta Loss Function (IBLF), which is a further deduction of the Taguchi Loss Function (Taguchi, 1986) in the industrial domain, helping to balance the possible loss resulting from trade-offs generated from different combinations of performance measures involved.

2.2 AI-based decision support system

Much work has been conducted in machine learning for classification, whereas the motivation is to attain a discovery of high-level prediction. Artificial intelligence (AI) has been widely used in knowledge discovery by considering both cognitive and psychological factors. Genetic Algorithm (GA), one of the significant AI search algorithms, is widely used to perform a global search in the problem space based on the mechanics of natural selection and natural genetics (Holland, 1992; Gen & Cheng, 2000; Freitas, 2001).

GA is regarded as a genetic optimization technique for global optimization, constrained optimization, combinatorial optimization and multi-objective optimization. GA has been used to enhance industrial engineering for achieving high throughput with quality guaranteed (Santos et al., 2002; Li et al., 2003; Al-Kuzee et al., 2004). There is a variety of evolutionary techniques and approaches of GA optimization, discussed in the research work by (Lopes et al., 1999; Ishibuchi & Yamamoto, 2002; Golez et al., 2002; de la Iglesia et al., 2003; Zhu & Guan, 2004; Goplan et al., 2006). Recently GA is also considered to be an essential tool in optimizing the inventory management (Radhakrishnan et al., 2009).

On the other hand, the fundamental concept of fuzzy logic is that it is characterized by a qualitative, subjective nature and linguistically expressed values (Milfelner et al., 2005).

Fuzzy rule sets, together with the associated membership functions, have been proven of great potential in their integration into GA to formulate a compound knowledge processing decision support system (Mendes et al., 2001; Leung et al., 2003; Ishibuchi & Yamamoto, 2004). Studies on applying fuzzy logics to systems for different sectors have been extensively undertaken (Cordon et al., 1998; Teng et al., 2004; Hasanzadeh et al., 2004; Chen & Linkens, 2004; Chiang et al., 2007; Tang & Lau, 2008).

2.3 Summary

Inspiring from all above, a Fuzzy-GA Decision Capacity Model is proposed for decision-makers to better select the proper warehousing strategies in terms of the corresponding performance metrics. The capacity will be evaluated by the rack utilization of the designated warehouse.

3. The proposed hybrid decision model

The proposed decision-supoort approach consists of two major processes: knowledge representation and knowledge assimilation, which are shown in Fig.1. In the first stage, the

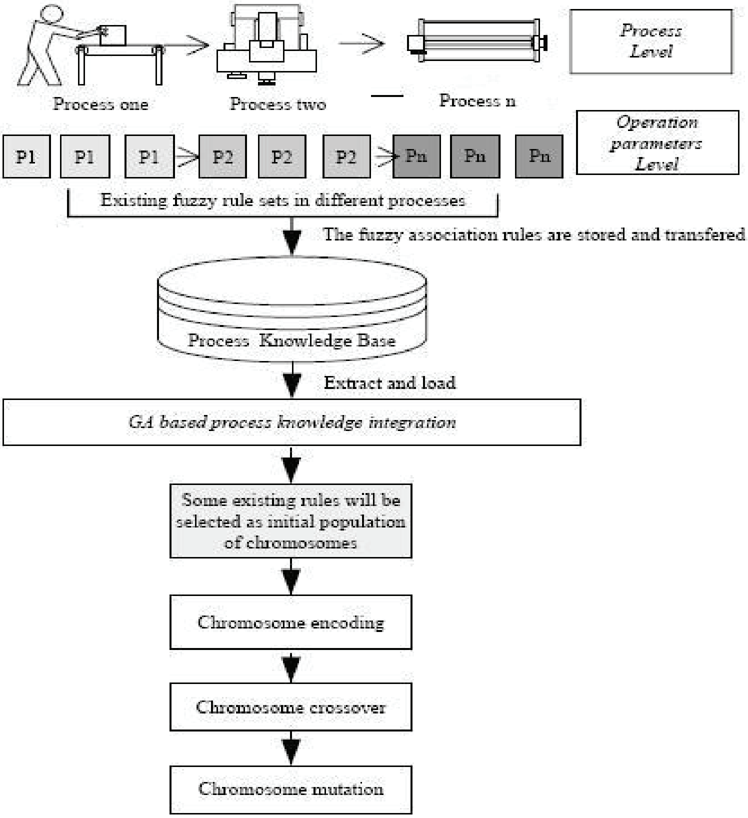

Process one
Process two
Process n
Process Level
P1
P1
P1
P2
P2
P2
Pn
Pn
Pn
Operation parameters Level
Existing fuzzy rule sets in different processes
The fuzzy association rules are stored and transfered
Process Knowledge Base
Extract and load
GA based process knowledge integration
Some existing rules will be selected as initial population of chromosomes
Chromosome encoding
Chromosome crossover
Chromosome mutation

Fig. 1. The proposed decision-support framework

expertise of factor setting, which is represented by IF-THEN rules, is encoded as a string with fuzzy rule sets and the associated fuzzy membership function. The historical process data are also included into the strings mentioned above, contributing to the formulation of an initial knowledge population. Then in knowledge assimilation, GA is used to generate an optimal or nearly optimal fuzzy set and membership functions for the entitled performance indicators. Accordingly, it is necessary to set relative weights for them to aggregate the measurement results since there naturally contains essential fuzziness and ambiguity in human judgments.

Fig. 1 depicts the overview of the entire proposed knowledge-based framework, while the initial rules extracted from process knowledge base are used to form the initial population of the GA. Fig. 2 illustrates the data flow of the proposed capacity-optimizing model, indicating how the iterations envelop fuzzy rule mining, improving the quality of generated rule sets and streamlining the various functional processes in a single warehouse.

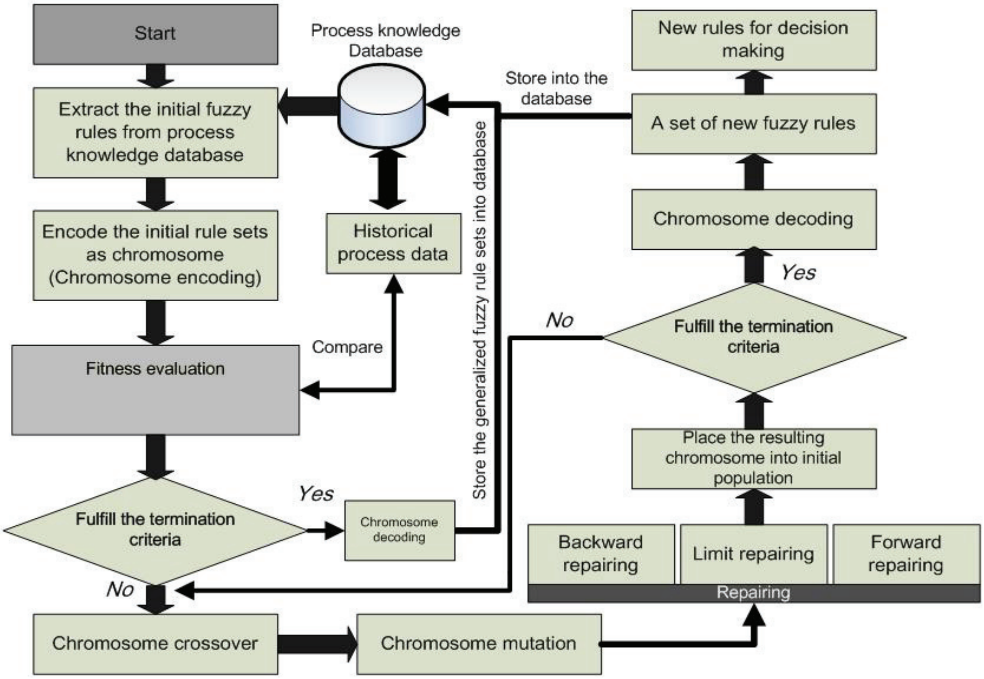

Start
Process knowledge Database
Extract the initial fuzzy rules from process knowledge database
Encode the initial rule sets as chromosome (Chromosome encoding)
Historical process data
Compare
Fitness evaluation
Fulfill the termination criteria
Yes
No
Chromosome decoding
Store the generalized fuzzy rule sets into database
Store into the database
New rules for decision making
A set of new fuzzy rules
Chromosome decoding
Yes
No
Fulfill the termination criteria
Place the resulting chromosome into initial population
Backward repairing
Limit repairing
Forward repairing
Repairing
Chromosome crossover
Chromosome mutation

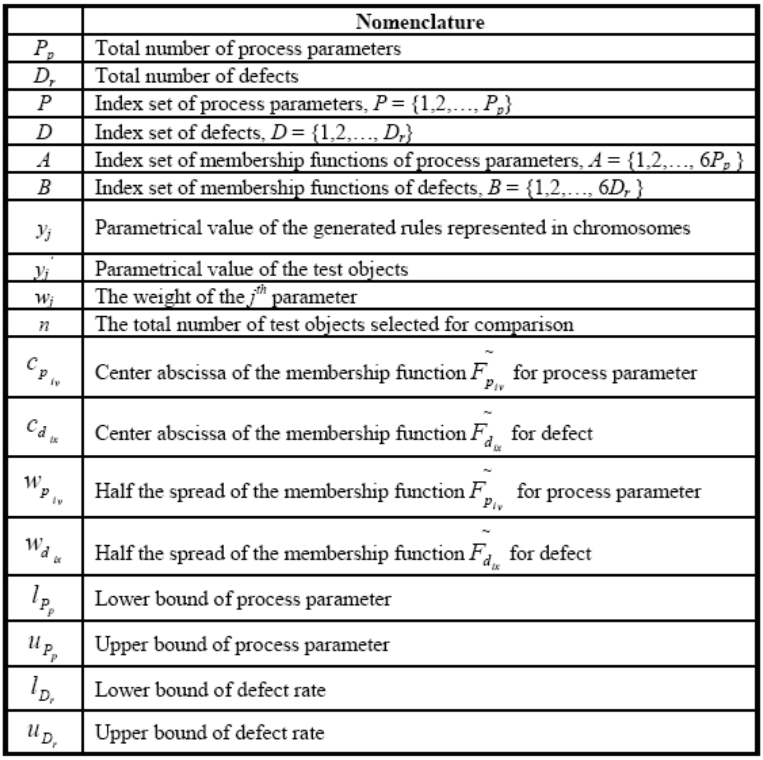
Fig. 2. Information flow of the proposed algorithm (Reference: Ho et al., 2008)

3.1 Problem formulation

Fuzzy-encorporated GA is proposed for capturing domain knowledge from an enormous amount of data. The proposed approach is to represent the knowledge with a fuzzy rule set and encode those rules together with the associated membership into a chromosome. A population of chromosomes comes from the past historical data and an individual chromosome represents the fuzzy rule and the related problem. A binary tournament, using roulette wheel selection, is used for picking out the best chromosome when a pair of chromosomes is drawn. The fitness value of each individual is calculated using the fitness function by considering the accuracy and the trade-off of the resulting performance measure setting, where the fitter one will remain in the population pool for further mating. After crossover and mutation, the offspring will be evaluated by the fitness function and the optimized solution will then be obtained.

The practitioners could freely select the specifically influential performance measures from a large pool of the candidate performance metrics based on the unique condition of the warehouse, leading to the optimized warehousing rack efficiency amongst all by comparing the weights.

3.2 Nomenclature



	Nomenclature
P_p	Total number of process parameters
D_r	Total number of defects
P	Index set of process parameters, $P = \{1,2,\ldots, P_p\}$
D	Index set of defects, $D = \{1,2,\ldots, D_r\}$
A	Index set of membership functions of process parameters, $A = \{1,2,\ldots, 6P_p\}$
B	Index set of membership functions of defects, $B = \{1,2,\ldots, 6D_r\}$
y_j	Parametrical value of the generated rules represented in chromosomes
y_j'	Parametrical value of the test objects
w_j	The weight of the j^{th} parameter
n	The total number of test objects selected for comparison
$c_{p_{iv}}$	Center abscissa of the membership function $\tilde{F}_{p_{iv}}$ for process parameter
$c_{d_{ix}}$	Center abscissa of the membership function $\tilde{F}_{d_{ix}}$ for defect
$w_{p_{iv}}$	Half the spread of the membership function $\tilde{F}_{p_{iv}}$ for process parameter
$w_{d_{ix}}$	Half the spread of the membership function $\tilde{F}_{d_{ix}}$ for defect
l_{P_p}	Lower bound of process parameter
u_{P_p}	Upper bound of process parameter
l_{D_r}	Lower bound of defect rate
u_{D_r}	Upper bound of defect rate

Table 1. Nomenclature of the proposed algorithm

Table. 1 above indicates the notations of the mathematical expressions involved in the proposed decision-support algorithm.

3.3 Chromosome encoding

Fuzzy concept is used to map the above linguistic decision rules into genes for GA optimization.

Definition 1: $C_h = \{1,2,\ldots,M\}$ represents the index set of chromosomes where M is the total number of chromosomes in the population.

Definition 2: $G_{m\times t}$ represents a gene matrix generated for the population where

$$G_{m\times w}=\left[\begin{array}{cccc|cccc|cccc|cccc} p_{11} & p_{12} & \cdots & p_{1a} & d_{11} & d_{12} & \cdots & d_{1b} & k_{11} & k_{12} & \cdots & k_{1c} & q_{11} & q_{12} & \cdots & q_{1d} \\ p_{21} & p_{22} & \cdots & p_{2a} & d_{21} & d_{22} & \cdots & d_{2b} & k_{21} & k_{22} & \cdots & k_{2c} & q_{21} & q_{22} & \cdots & q_{2d} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{ma} & d_{m1} & d_{m2} & \cdots & d_{mb} & k_{m1} & k_{m2} & \cdots & k_{mc} & q_{m1} & q_{m2} & \cdots & q_{md} \end{array}\right]$$

$$=\left((p_{iu})_{m\times a}(d_{ix})_{m\times b}(k_{iy})_{m\times c}(q_{iv})_{m\times z}\right)$$

$$p_{iv}=random\left[l_{P_p},u_{P_p}\right],d_{ix}=random\left[l_{D_r},u_{D_r}\right],$$
$$k_{i,\tau}=c_{piv},k_{i,\lambda}=w_{p_{iv}},q_{i,\tau}=c_{dix},q_{i,\lambda}=w_{dix}$$

$$\forall i\in C_h,\forall v\in P,\forall x\in D,\forall y\in A,\forall z\in B,\tau=1,3,5,......;$$
$$\lambda=2,4,6,......;m=M,a=P_p,b=D_r,c=6P_p,d=6D_r$$

Note that the decoding method of an element in the first sub-matrix $(p_{iv})_{m\times b}$ or second sub-matrix $(d_{ix})_{m\times s}$ of $G_{m\times w}$ to a linguistic variable is given by:
(i) 0: ignore, (ii) 1: low, (iii) 2: medium, and (iv) 3: high. For any row of the third sub-matrix $(k_{iy})_{m\times e}$ of $G_{m\times w}$, a group of six consecutive values $k_{i(6\rho-5)},k_{i(6\rho-4)},k_{i(6\rho-3)},k_{i(6\rho-2)},k_{i(6\rho-1)},k_{i(6\rho)}$ in the matrix forms a single set $\tilde{F}_{p_{iv}}=\left\{c_{p_{iv}}-w_{p_{iv}},w_{p_{iv}},c_{p_{iv}},w_{p_{iv}},c_{p_{iv}}+w_{p_{iv}},w_{p_{iv}}\right\}$ for process parameter pv where $\rho=1,2,3,.......$. Also, for any row of the fourth sub-matrix $(q_{iz})_{m\times n}$ of $G_{m\times w}$, a group of six consecutive values $q_{i(6\rho-5)},q_{i(6\rho-4)},q_{i(6\rho-3)},q_{i(6\rho-2)},q_{i(6\rho-1)},q_{i(6\rho)}$ in the matrix forms a single set $\tilde{F}_{d_{ix}}=\left\{c_{d_{ix}}-w_{d_{ix}},w_{d_{ix}},c_{d_{ix}},w_{d_{ix}},c_{d_{ix}}+w_{d_{ix}},w_{d_{ix}}\right\}$ for defect rate dx where $\rho=1,2,3,.......$. For both two cases, there are totally 6 genes in the sets of membership functions shown in Fig. 3.

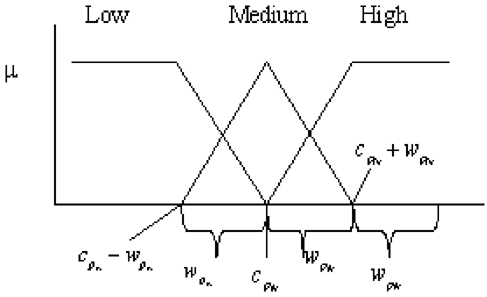


Fig. 3. Fuzzy membership functions of the influencing factors

$\tilde{F}_{p_{iv}}$ consists of aggregated membership functions which relate to a fuzzy rule set is assumed to be isosceles-triangle functions.

$c_{p_{iv}}$ is the center abscissa of $\tilde{F}_{p_{iv}}$; $w_{p_{iv}}$ represents half the spread of $\tilde{F}_{p_{iv}}$.

In " $c_{p_{iv}}$ ", "p_{iv}" indicates that the v-th feature test is included, while i specifies the order of all the condition levels of each feature test. For instance, $c_{p_{i1}}$ stands for the center abscissa of the 1st process test, within the whole membership function matrix.

Definition 3: $B_{m\times 1}$ denotes a random number matrix generated for selection and crossover where

$$Bm\times 1=(bi)_{m\times 1}$$

$$b_i = random[0,1], \forall i \in C_h, m = M\ .$$

Definition 4: $C_{h\_c} = \{1,2,.....,S\}$ denotes the index set of the chosen chromosomes in the crossover where S is the total number of chosen chromosomes

Definition 5: $G'_{m\times w}$ indicates the gene matrix in which the Q chromosomes chosen in crossover are stored where

$$G'_{m\times w} = \left((p'_{iu})_{m\times a}(d'_{ix})_{m\times b}(k'_{iy})_{m\times c}(q'_{iv})_{m\times z}\right)$$

3.4 Fitness evaluation

To have a good set of process parameters, the genetic algorithm selects the best chromosome for mating according to the fitness function suggested below.

$$\text{Fitness Funtion} = \text{accuracy with error rate}$$

$$Accuracy = \frac{\text{objects correctly matched within error range}}{\text{total number of objects}}$$

$$\text{Error rate } (\varepsilon) = \sum_{j=1}^{m} w_j \frac{(y_j - y_j')^2}{2n}$$

Each chromosome is evaluated by calculating its mean-square error for the error measurement. As each chromosome is represented as the fuzzy rule, the quality of the chromosome is then validated by comparing its defuzzified output with the actual output of the test samples. The centre of gravity (COG) is used as the defuzzification method to obtain the crisp values of the finished quality level.

3.5 Chromosome crossover

Crossover is a genetic operation aiming at producing new and better offspring from the selected parents, while the selection is determined by a crossover rate. The current crossover methods include single-point crossover, two-point crossover, multi-point crossover, uniform crossover, random crossover, etc. Uniform crossover is selected in this research.

3.6 Chromosome mutation

Mutation is intended to prevent all solutions in the population from falling into the local minima. It does this by preventing the population of chromosomes from becoming too similar to each other, which might slow down or even stop evolution. Mutation operation randomly changes the offspring resulting from crossover, given that the value of the mutation rate must range within 0 and 1. In our paper a bit-flip mutation is used.

3.7 Chromosome repairing

After the mutation and crossover in the two regions, some violations in the chromosome may occur. If the membership function is not in ascending order, the new offspring should be modified by exchanging the gene order in accordance with the definition of $\tilde{F}_{p_{iv}} = \{c_{p_{iv}} - w_{p_{iv}}, w_{p_{iv}}, c_{p_{iv}}, w_{p_{iv}}, c_{p_{iv}} + w_{p_{iv}}, w_{p_{iv}}\}$.

The repairing is divided into two categories which are: the forward and backward repairing as illustrated in Fig.4(a) and Fig. 4(b).

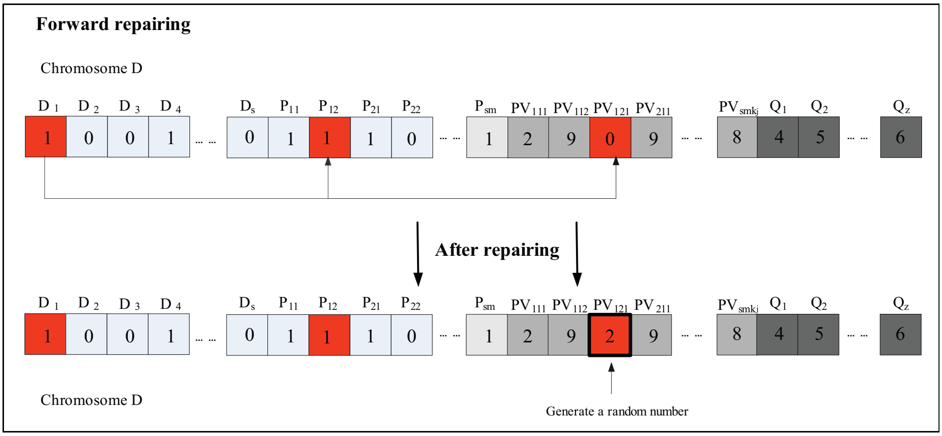


Fig. 4(a). Sample chromosome of forward repairing

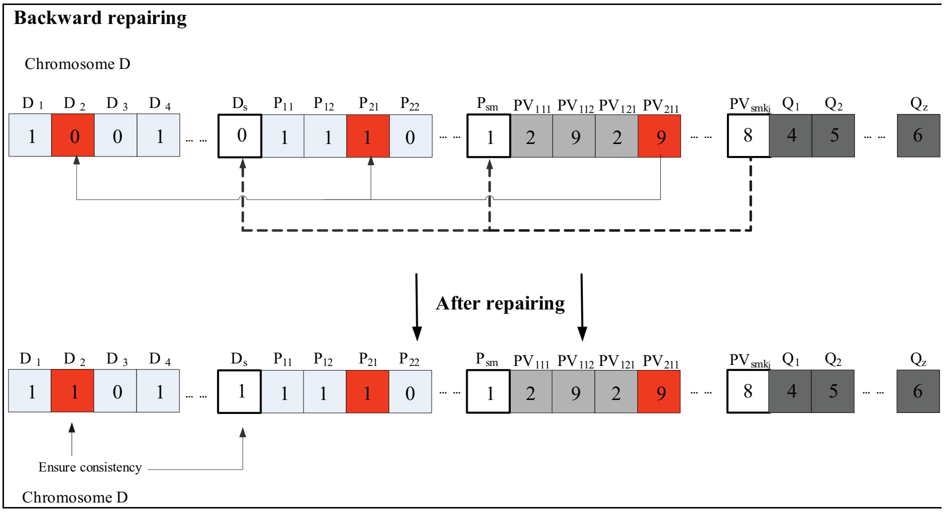


Fig. 4(b). Sample chromosome of backward repairing

3.8 Chromosome decoding

Once the termination criterion is fulfilled, the decoding process will be implemented on the whole set of optimum chromosomes (Fig. 5). The optimum chromosomes decode into a series of linguistic fuzzy rule sets as shown in Table 2 and their associated membership functions which are stored in the repository for further investigation.

Condition part <IF> *(Warehousing Influencing Factors)*	***Consequent part <THEN>*** *(Rack Utilization)*
Rule 1: *Process1*.Inventory cost is adjusted to medium AND *Process2*.Backorder cost is adjusted to medium AND *Process3*.Maintenance cost is adjusted to high … … AND *ProcessN*.	Rack utilization of Drive-in is extremely low AND Rack utilization of APR is high AND Rack utilization of Double-deep is medium AND ...
Rule 2: *Process1*.Inventory cost is adjusted to low AND *Process4*.Backorder cost is adjusted to high AND … … AND *ProcessN*+1.	Rack utilization of Drive-in is medium AND Rack utilization of Double-deep is extremely high

Table 2. Sample of generalized fuzzy rules obtained in the FGCDM

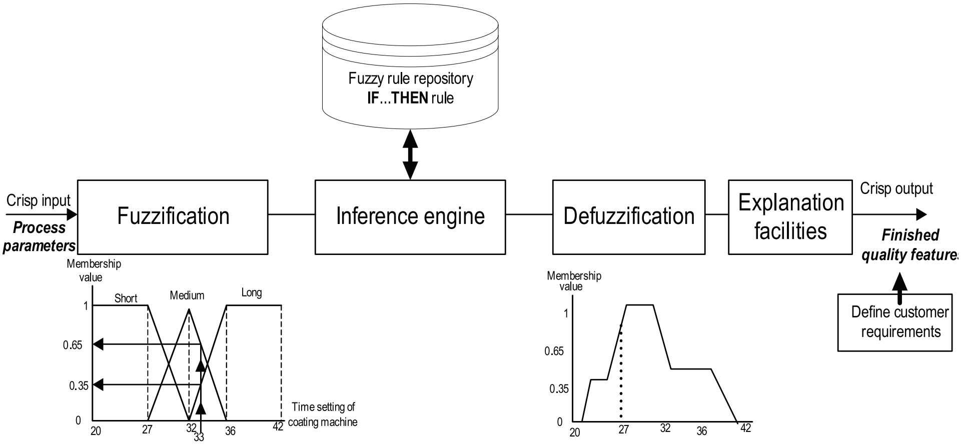


Fig. 5. Sample of generalized fuzzy rules obtained in the FGCDM

3.9 De-fuzzification

Once the termination criterion is fulfilled, the decoding process will be implemented on the whole set of optimum chromosomes. The optimum chromosomes decode into a series of fuzzy rule sets and their associated membership functions which are stored in the repository for further investigation.

4. Discussion and experiment results

The warehousing background for the simulation is of medium volumes (300 pallets/day throughput) and with 90 SKUs to be placed into the storage. The existing rack system include Block-stack, Drive-in, APR, Double deep and VNA. The evaluation criterion of the warehouse performance is mainly based on the utilization of the above racks.

In order to verify the proposed Fuzzy-GA capacity decision model (FGCDM), simulations on searching ability were carried out. Two different stochastic-based search methods, Simulated Annealing (SA) and Tabu Search (TS), were used for comparison with the proposed FGCDM approach. In this experiment, the historical data for supporting the warehousing operation and 30 performance indicators were used for the simulation. The results reported are all averaged over 10 independent runs. In each data set, the best (minimum) fitness value among the 10 simulation runs was documented for the comparison of each search technique mentioned above.

Number of runs	**SA**	**TS**	**FGCDM**
1	0.822	0.89	0.913
2	0.87	0.923	0.892
3	0.91	0.887	0.93
4	0.762	0.781	0.795
5	0.863	0.871	0.88
6	0.836	0.82	0.933
7	0.816	0.848	0.853
8	0.902	0.833	0.892
9	0.827	0.911	0.958
10	0.842	0.892	0.884
Average	0.845	0.866	0.893

Table 3. Best (Minimum) fitness values obtained by FGCDM, SA & TS

	Rack Utilizations (%)	
Warehouse Rack Type	Model Result	Observed
Block-stack	91.8%	88.2%
Drive-in	91.2%	75.7%
APR	95.5%	96.1%
Double deep	89.7%	77.3%
VNA	93.1%	92.8%

Table 4. Rack Utilization of Observed and Model Results

Table 3 presents that ten independent runs of fitness values acquired by various search techniques using 30 performance indicators. According to the experiment, SA was the worst performer in all 10 independent runs and the proposed FGA approach achieved the smallest average object value at 0.893 in the maximization of rack utilization over the interval 0 to 1. Compared with the observed test data which are half-extracted from the historical records, our approach shows an overall better result in Table 4.

5. Conclusion

In this research, the design and implementation of a GA based process knowledge model, which embraces the fuzzy theory and genetic algorithm to achieve warehouse capacity improvement, has been introduced. Implementing the proposed methodology in the aspect of warehouse management through simulation has been successful. By incorporating the error measurement and complexity of process change into the fitness evaluation, the generalized fuzzy rule sets can be of less complexity and higher accuracy. An extension of different measures can also be included in order to improve the generalized rules. In the matter of generation of new fuzzy rules, the membership functions are assumed to be static and known. The proposed intelligent model can help the decision makers in the development and selection of the best warehouse design for the given application.

Other fuzzy learning methods should be considered to dynamically adjust the membership functions of various parameters to enhance the model accuracy. Future contribution of this endeavour goes to validation of the decision model to be launched in case companies.

6. Acknowledgements

The authors wish to thank the Research Committee of The Hong Kong Polytechnic University for the support of this research.

7. References

Al-Kuzee, J.; Matsuura, T.; Goodyear, A.; Nolle, L.; Hopgood, A.A.; Picton, P.D. & Braithwaite, N.St.J. (2004). Optimization of plasma etch processes using evolutionary search methods with in situ diagnostics, *Plasma Sources Science Technology*, vol. 13, no. 4, pp. 612 – 622.

Angerhofer, B.J. & Angelides, M.C. (2006). IA model and a performance measurement system for collaborative supply chains, *Decision Support Systems*, vol. 42, Issue 1, pp. 283-301.

Baker, P. (2007). An exploratory framework of the role of inventory and warehousing in international supply chains, *International Journal of Logistics Management*, Vol. 18, Issue 1, pp. 64–80.

Bernardy, G. & Scherff, B. (1998). SPOC - Process Modelling Provides On-line Quality Control and Predictive Process Control in Particle and Fibreboard Production. *Proceedings of the 24th Annual Conference of IEEE Industrial Electronics Society*, IECON '98, 31.08.-04.09., Aachen.

Bowersox, D.J. & Closs, D.J. (1996). *Logistical Management: the Integrated Supply Chain Process*, Macmillan, New York, NY.

Caplice, C. & Sheffi, Y. (1995). A review and evaluation of logistics performance measurement systems, *The International Journal of Logistics Management*, Vol. 6, Issue 1, pp. 61-74.

Chan, F.T.S. & Qi H.J. (2003). Feasibility of performance measurement system for supply chain: a process-based approach and measures, *Integrated Manufacturing Systems*, Vol.14, Issue 3, pp. 179-190.

Chen, M.Y., Linkens, D.A. (2004). Rule-base self-generation and simplification for data-driven fuzzy models, *Fuzzy Sets and Systems*, Vol. 142, Issue 2, pp. 243-265.

Chiang, T.C.; Huang, A.C. & Fu, L.C. (2007). Modeling, scheduling, and performance evaluation for wafer fabrication: a queueing colored Petri-net and GA-based approach, *IEEE Transactions on Automation Science and Engineering*, vol. 3, no. 3, pp. 912-918.

Christopher, M. (2005). *Logistics and Supply Chain Management*, third ed. Pearson, Harlow.

Cordon O., Del Jesus M.J., Herrera F. (1998). Genetic learning of fuzzy rule-based classification systems cooperating with fuzzy reasoning methods, *International Journal of Intelligent Systems*, Vol. 13, Issue 10–11, pp. 1025–1053.

De la Iglesia, B.; Philpott, M.S.; Bagnall, A.J. & Rayward-Smith, V.J. (2003). Data Mining Rules Using Multi-Objective Evolutionary Algorithms, In: *Proceedings of IEEE Congress on Evolutionary Computations*, Vol. 3, pp. 1552-1559.

Feng, X. and Huang, H. (2005). A fuzzy-set-based Reconstructed Phase Space method for Identification of Temporal Patterns in Complex Time Series, *IEEE Transactions on Knowledge and Data Engineering*, Vol.17, No.5, pp. 601-613.

Frazelle, E. (2002). *World-class Warehousing and Material Handling*. McGraw-Hill, New York.

Freitas, A.(2001). A survey of evolutionary algorithms for data mining and knowledge discovery. In: *Advances in Evolutionary Computation*. Springer- Verlag.

Gen, M. & Cheng, R. (2000). *Genetic algorithms and engineering optimization*. New York: Wiley.

Gomez, J.; Gonzalez, F. & Dasgupta, D. (2002). Complete Expression Trees for Evolving Fuzzy Classifier Systems with Genetic Algorithms, In: *Proceedings of the Evolutionary Computation Conference GECCO'02*, 2002.

Gopalan, J.; Alhajj, R. & Barker, J. (2006). Discovering Accurate and Interesting Classification Rules Using Genetic Algorithm, In: *Proceedings of the 2006 International Conference on Data Mining*, pp. 389-395. June 26-29, 2006.

Gu, J.; Goetschalckx, M. & McGinnis, L.F. (2007). Research on warehouse operation: A comprehensive review. *European Journal of Operational Research*, vol. 177, Issue 1, pp. 1–21.

Gunasekaran, A.; Patel, C. & Tirtiroglu, E. (2001). Performance measures and metrics in a supply chain environment, *International Journal of Operations & Production Management*, vol.21, Issue 1/2, pp. 71-87.

Harrison, A. & van Hoek, R. (2005). *Logistics Management and Strategy*. second ed. Pearson, Harlow.

Hasanzade, M., Bagheri, S., Lucas, C. (2004). Discovering Fuzzy Classifiers by Genetic Algorithms, In: *Proceedings of 4th international ICSC Symposium on Engineering of Intelligent Systems (EIS2004)*, 2004, Island of Madeira, Portugal.

Higginson, J.K. & Bookbinder, J.H. (2005). Distribution centres in supply chain operations. In: Langevin, A.L. & Riopel, D. (2005), *Logistics Systems: Design and Optimization*. Springer, New York, pp. 67–91.

Ho, G.T.S.; Lau, H.C.W.; Chung S.H.; Fung R.Y.K.; Chan, T.M. & Lee, C.K.M (2008). Development of an intelligent quality management system using fuzzy association rules, *Industrial Management & Data Systems*, vol. 108, no. 7, pp. 947-972.

Holland, J.H. (1992). *Adaptation in Natural and Artificial Systems*. Cambridge, MA: MIT Press.

Ishibuchi, H. & Yamamoto, T. (2002). Fuzzy rule selection by data mining criteria and genetic algorithms, In: *Proceedings of Genetic and Evolutionary Computation Conference (GECCO 2002)*, pp. 399-406, New York, July 9-13.

Ishibuchi, H. & Yamamoto, T. (2004). Fuzzy Rule Selection by Multi-Objective Genetic Local Search Algorithms and Rule Evaluation Measures in Data Mining, *Fuzzy Sets and Systems*, Vol. 141, no. 1, pp. 59-88.

Kobbacy, K.; Vadera, S. & Rasmy, M.H. (2007). AI and OR in management of operations: history and trends, *Journal of the Operational Research Society*, vol.58, pp. 10-28.

Lau, H.C.W.; Tang, C.X.H.; Leung, B.P.K.; Lee, C.K.M. & Ho, G.T.S. (2009). A Performance Tradeoff Function for Evaluating Suggested Parameters in the Reactive Ion Etching Process, *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 37, no. 5, pp. 758–769.

Leung, R.W.K.; Lau, H.C.W. & Kwong, C.K. (2003). An expert system to support the optimization of ion plating process: an OLAP-based fuzzy-cum-GA approach, *Expert Systems with Applications*, vol. 25, no. 3, pp. 313 – 330.

Leung, B.P.K. & Spiring, F.A. (2002). The inverted beta loss function: properties and applications, *IIE Transactions*, vol. 34, no. 12, pp. 1101 – 1109.

Li, T.S.; Su, C.T. & Chiang, T.L. (2003). Applying robust multi-response quality engineering for parameter selection using a novel neural–genetic algorithm, *Computers in Industry*, vol. 50, no. 1, pp. 113 – 122.

Lohman, C.; Fortuin, L. & Wouters, M. (2004). Designing a performance measurement system: A case study, *European Journal of Operational Research*, vol.156, Issue 2, pp.267-286.

Lopes, C.; Pacheco, M.; Vellasco, M. & Passos, E. (1999). Rule-Evolver: An Evolutionary Approach For Data Mining, In: *Proceedings of the 7th International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*, pp. 458-462.

Mendes, R.R.F.; Voznika, F.; de B Freitas, A.A. & Nievola, J.C. (2001). Discovering Fuzzy Classification Rules with Genetic Programming and Co-Evolution, In: *Principles of Data Mining and Knowledge Discovery* (Proceedings of the 5th European Conference PKDD 2001) –Lecture Notes in Artificial Intelligence, Springer-Verlag.

Milfelner, M; Kopac, J.; Cus, F. & Zuperl, U. (2005). Genetic equation for the cutting force in ball-end milling, *Journal of Materials Processing Technology*, vol. 164/165, pp. 1554 – 1560.

Niemi, P.; Huiskonen, J. & Karkkainen, H. (2009). Understanding the knowledge accumulation process—Implications for the adoption of inventory management techniques, *International Journal of Production Economics*, vol.118, Issue 1, pp.160-167.

Radhakrishnan, P.; Prasad, V.M. & Gopalan, M.R. (2009). Inventory Optimization in Supply Chain Management using Genetic Algorithm, *International Journal of Computer Science and Network Security*, Vol.9 No.1, pp.33-40.

Santos, C.A.; Spim, J.A.; Ierardi, M.C.F. & Garcia, A. (2002). The use of artificial intelligence technique for the optimisation of process parameters used in the continuous casting of steel, *Applied Mathematical Modelling*, vol. 26, no. 11, pp. 1077 – 1092.

Tang C.X.H. & Lau H.C.W (2008). A Fuzzy-GA Decision Support System for Enhancing Postponement Strategies in Supply Chain Management, In: *Lecture Notes in Computer Science*, vol. 5361, Springer-Verlag Berlin Heidelberg, pp.141-150.

Taguchi, G. (1986). *Introduction to Quality engineering: Designing Quality into Products and processes*. NY: Kraus, White Plains.

Teng M., Xiong F., Wang R. & Wu Z. (2004). Using genetic algorithm for weighted fuzzy rule-based system, In: *Proceedings of Fifth World Congress on Intelligent Control and Automation*, 2004, Hangzhou, China.

Tompkins, J.A. & Smith, J.D. (1998). *The Warehouse management Handbook*. Tompkins Press.

Wouters, M. & Sportel, M. (2005). The role of existing measures in developing and implementing performance measurement systems, *International Journal of Operations & Production Management*, vol.25, Issue 11, pp.1062-1082.

Zhu, F. & Guan, S.U. (2004). Ordered Incremental Training with Genetic Algorithms, *International Journal of Intelligent Systems*, Vol. 19, Issue 12, pp. 1239-1256.

4

Connectionist Models of Decision Making

Angel Iglesias, M. Dolores del Castillo, J. Ignacio Serrano and Jesus Oliva
Instituto de Automatica Industrial, CSIC
Spain

1. Introduction

This chapter discusses some approaches to computational modelling of decision making. Concretely, it concerns with connectionist models of decision making and it contributes to the categorization of such models. The models presented in this chapter are algorithmic and structural descriptions of the mental process of decision making.

First of all, there are some definitions that must be stated in this chapter. A decision occurs when a person faces several options (alternatives), then evaluates the consequences of choosing each alternative and, finally, selects one depending on her/his preferences (Rustichini, 2009). Preference is an abstract relation between two options: when presented with options A_1 and A_2, it is assumed that a subject either prefers A_1 to A_2 or the subject prefers A_2 to A_1 (or is indifferent between them). The decision is guided by the subjective experience and preference of the subject. It depends on internal factors of the subject and external factors of the environment. Due to these considerations, the goodness of a decision is subjective and it should be considered only within the context of the subject and the environment. The parameters that characterized each alternative are called criteria.

A model is a simpler and more abstract version of a system that keeps its key features while omitting less relevant details. It is used to investigate a system or phenomenon that is too complex to deal with directly. An important class of models is represented by computational models (Fum et al., 2007), which are implemented as algorithms. While statistical and mathematical models describe a phenomenon without reproducing it, computational models do. Therefore, computational models make easier the observation and measurement of a phenomenon's behaviour.

There are two opposite points of view concerning how the human mind works (Chown, 2004). One considers the basis of human mind as a symbol processing system and the other assumes that the brain must be modelled in neural terms. This chapter is focused on connectionism, which is a theoretical framework for cognition that assumes two main statements. The first one is that cognitive phenomena arise from the propagation of activation among nodes. The second is that such propagation is mediated by weighted connections between nodes. So, computational models built on connectionism principles are composed by nodes and connections. A node represents an entity (idea, concept, etc.) which has an associated value (activation). A node can transmit its activation through its connections. One node can either excite or inhibit another node depending on the strength of the connection that links them. Thus, a connectionist model must specify, among other things, the way nodes transmit activation.

There are neuropsychological evidences that suggest that the human mechanism for making a decision is divided into two stages (Glimcher, 2009). The first stage lies in the assessment of all alternatives and the second is concerned with choosing one of them depending of the value assigned in the previous stage. The assessment mechanism is associated with the representation of values, while the choice mechanism is associated with a process that takes these values and returns the best alternative. These mechanisms are the core of the models presented in this chapter. These evidences confirm the Prospect theory (Kahneman & Tversky, 1979), which distinguishes two phases in the choice process: an early phase of editing and a subsequent phase of evaluation. The editing phase consists of a preliminary analysis of the offered options, which often yields a simpler representation of these options. In the second phase, the edited options are evaluated and the option of highest value is chosen.

There are several models that describe the decision process. Depending on the application, different constraints may be enforced on the computational modelling task. This chapter deals with the following categories of models. Threshold Models make a decision when there is sufficient evidence accumulated favouring one alternative over the others. Ranking models rank the alternatives according to their estimated consequences and then choose the best one. Rule-based models apply rules to choose the best alternative. Emotional models take into account emotion in the decision process. Physiologically motivated models aim to describe the decision process using interconnected modules which represent different brain areas. An important remark is that some of the models presented in this chapter fall into more than one category.

This chapter is organized as follows. Section 2 presents the category of Threshold Models and explains different models within this category. In section 3 there are some Ranking models. Section 4 describes Rule-based models. Section 5 contains different models that include emotions in the process of decision making. Some physiologically motivated models are compiled in section 6. Finally, section 7 presents the conclusions.

2. Threshold Models

Threshold Models assume that decisions are based on accumulated information about the alternatives. Therefore, a decision is the result of continuously accumulating information until a threshold is reached. Threshold Models emulate the decision process as a race between alternatives, with the response determined by the first alternative to reach a threshold. The threshold that determines the amount of information needed for a response is under the control of the needs of the decision maker (the subject that makes the decision). For instance, the threshold is reduced with the necessity to respond quickly and increased with the necessity to respond accurately. Two main features of these models are the starting value of the accumulation process and the thresholds. The interest of these models is that they provide a description of the relationship between time and accuracy, and hence they are suitable for modelling speed-accuracy decision effects. In the following sections different Threshold Models are discussed.

Within the category of Threshold Models there are differences in how the accumulation is assumed to occur. These models contain a node, which is called accumulator, for each alternative, i.e. the information favouring each alternative is accumulated separately. Threshold Models gather information through other kind of nodes that represents environmental and subject's features. In Dependent Accumulators models, information in favour of one alternative is the evidence against the others. Thus, the accumulators are mutually inhibitory. There is another class in which accumulators are independent and there is no inhibition: Independent Accumulators models.

2.1 Decision field theory

Decision Field Theory (Busemeyer & Townsend, 1993) is a dynamic model of decision making that has been used to explain different aspects of the decision process such as the similarity effect, the attraction effect, the compromise effect and preference reversals (Roe et al., 2001; Johnson & Busemeyer, 2005). This model assumes that the decision process can be formulated as a connectionist model as shown in figure 1.

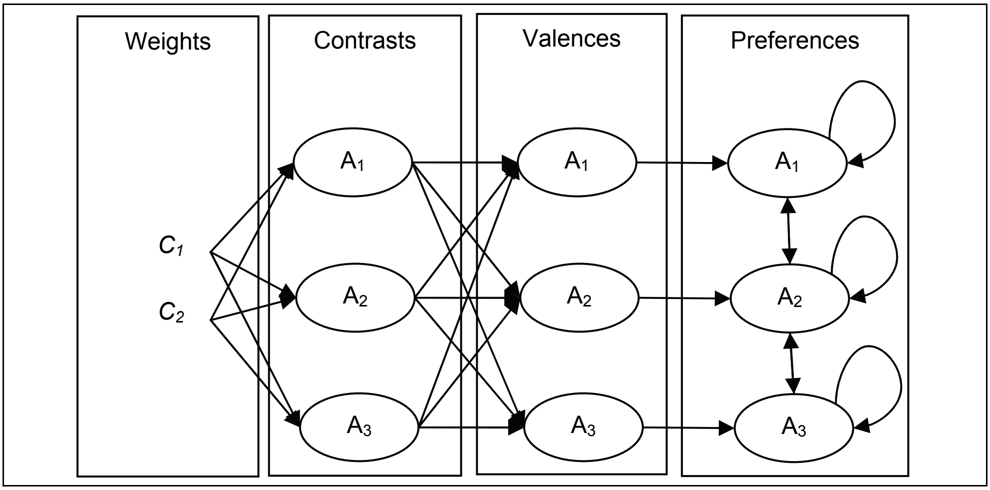


Fig. 1. Diagram showing the connectionist model of Decision Field Theory (DFT) which deals with a decision making problem consisting in three alternatives: A_1, A_2 and A_3

Information about the various possible consequences of each alternative represents the inputs into the model (C_i), i.e. criteria. The input criteria are filtered by an attention process, which maps the evaluations of consequences into a momentary evaluation for each alternative, represented by the second layer of nodes. Then the momentary evaluations are transformed into valences, one for each alternative, represented by the third layer of nodes. The valence of an alternative represents the momentary advantage or disadvantage of that alternative compared to the other alternatives. Finally, the valences are input to a recursive system at the final layer which generates the accumulation of information favouring each alternative at each moment in time. Decision Field Theory calls preferences to the values accumulated in the last layer. In this model, as attention switches across criteria over time, the accumulation of information also varies until the preference for one alternative exceeds the threshold and the winner is chosen.

The assessment mechanism lies in the first layer where information is gathered and criteria are weighted. The choice mechanism lies in the accumulation process through the last three layers of the model. This model is an example of a Dependent Accumulators model due to the connections between accumulators in the last layer. Notice that there exists inhibition.

2.2 Leaky, competing accumulators

The model presented in this section is based on the Leaky, Competing Accumulator model (Usher & McClelland, 2001) and operates as follows. At each time step, one criterion (*C_i*) of the consequences of the alternatives is chosen randomly to be the focus of attention. Therefore, the selected criterion is the only one that transmits its activation. The input to each of the Leaky, Competing Accumulator nodes (accumulators) is determined by an input pre-processing stage. This stage calculates the differences (dij) between all pairs of alternatives over the chosen criterion and then, converts these differences before transmitting them to the accumulators. This stage is performed in the second and third layer. The nodes in the second layer represent each alternative according to its weights over the criteria and transmit their activation to the third layer. The nodes in the third layer compute and transform the differences between the alternatives and, finally, transmit them to the last layer, which contains the accumulators.

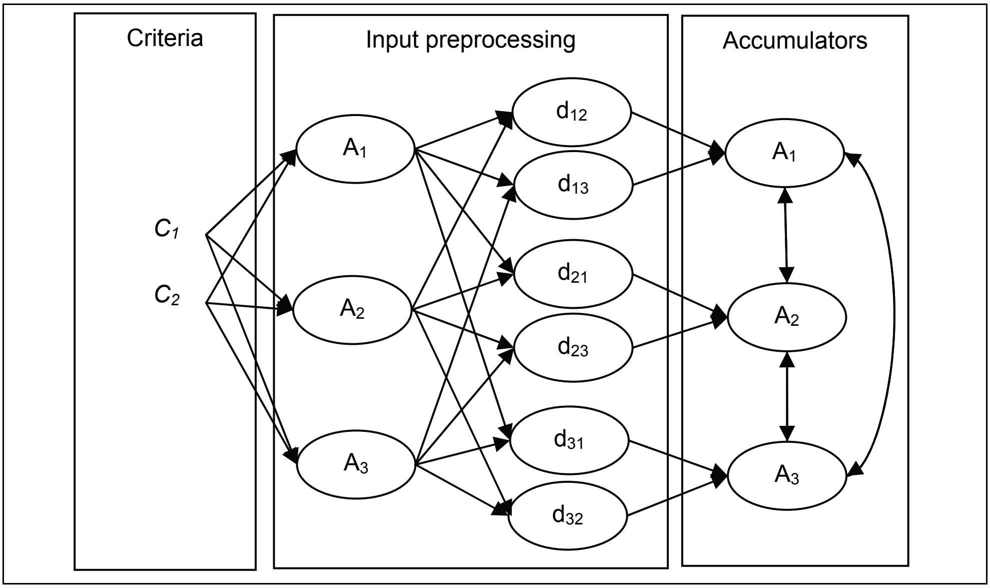


Fig. 2. Scheme illustrating the leaky competing accumulator model incorporating switching of attention for a choice between three alternatives (*A_1*, A_2 and *A_3*) and two criteria (*C_1* and *C_2*)

Figure 2 illustrates this model for a situation with three alternatives and two criteria. The model presented in this section (Usher & McClelland, 2004) extended the Leaky, Competing Accumulator model of perceptual choice incorporating switching of attention between criteria. As in the Decision Field Theory, the assessment mechanism lies in the first layer and the choice mechanism consists in accumulating information through the different layers of the model. This model is another instance of a Dependent Accumulators model.

2.3 Accumulator Model

The accumulator Model (Vickers 1970; Smith & Vickers, 1988) deals with decisions that have two alternatives. In this model there is no inhibition between the two accumulators. Each criterion (C_i) characterizing the alternatives (A_i) has associated a reference value (R_i). At each time step, one criterion is selected randomly and if its value is greater than the reference value, then the model adds the difference between the reference and the value to one accumulator. If the value is lower than the reference, then the model adds the difference to the other accumulator. The decision is determined by the first accumulator to reach the threshold. Figure 3 shows a connectionist interpretation of this model that contains two accumulators, one for each alternative, and two criteria.

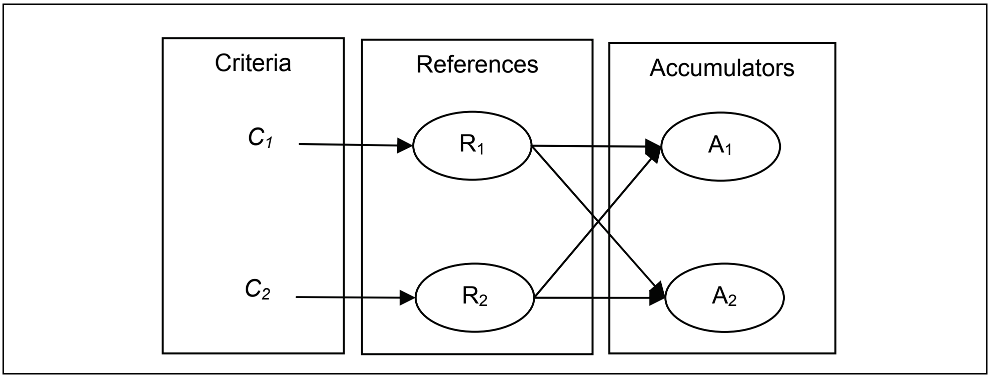


Fig. 3. Scheme showing the accumulator model for a choice consisting in two alternatives (A_1 and A_2) and two criteria (C_1 and C_2)

In this model, the choice mechanism is represented in the accumulation process and the assessment mechanism is based on the representation of the criteria. Notice that there is no inhibition between accumulators.

3. Ranking models

This approach makes the assumption that there is a global comparison of the alternatives. These models lie in the evaluation of the alternatives over each criterion and the determination of a score for each alternative. After this process of assessment, they determine a global ranking on the alternatives based on the scores. The decision is stated by the alternative with the best score. One of the most difficult tasks is to normalize the original values of the criteria, i.e. the assessment mechanism.

Within the category of Ranking models there are differences in how the global score is computed. For instance, a model built on these principles can lie in computing the weighted sum of some partial scores given by the criteria that characterize the alternatives. This chapter presents different Ranking models in the following sections.

3.1 Fuzzy cognitive map

A fuzzy cognitive map (Kosko, 1986) is a technique for modelling complex systems that consists of a great number of highly related and interconnected elements. Fuzzy cognitive maps represent knowledge capturing the cause and effect relationships among elements in order to model the behaviour of a system. The first fuzzy cognitive maps used numbers to describe causality and introduced positive and negative concepts. Fuzzy cognitive maps have been extended in order to be applied to decision making (Montibeller & Belton, 2009).

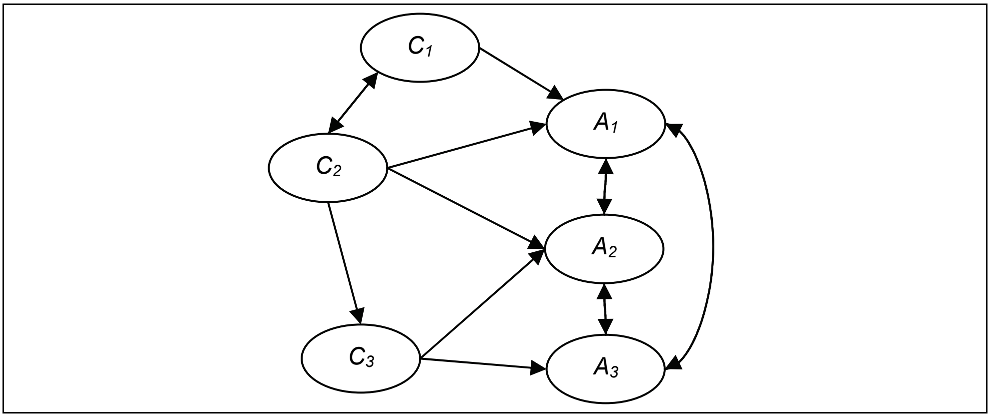


Fig. 4. Competitive fuzzy cognitive map for a choice consisting in three alternatives (A_1, A_2 and A_3) and three criteria (C_1, C_2 and C_3)

A kind of fuzzy cognitive map developed for decision making is called competitive fuzzy cognitive map (Stylios et al., 2008). The competitive fuzzy cognitive map introduced the distinction of two main kinds of concepts: decision concepts and factor concepts. Figure 4 illustrates a competitive fuzzy cognitive map that includes three alternatives (A_1, A_2 and A_3) and several criteria (C_1, C_2 and C_3). All the nodes can interact with each other and determine the value of the alternatives, which are mutually exclusive, in order to indicate always a single decision. The connections between concepts represent the causal relations among them. This model operates as follows. The model assigns the activation of factor nodes according to the decision making problem. These nodes are the input of the model. Then, the model begins a simulation divided in time steps. At each simulation step, the activation of a node is calculated by computing the influence of the interconnected nodes on the specific one following a sigmoid threshold function. The simulation finishes when there are no variations in the activation of every node. In such situation, when the competitive fuzzy cognitive map has converged, the decision node that has the greatest activation represents the best alternative.

In this model the assessment mechanism is represented by the propagation of activation through the nodes of the map representing criteria and the choice mechanism consists in selecting the alternative associated to the best scored decision node. Notice that there is an inhibition between alternatives. This is not a Threshold model because the decision is not made when a decision node reaches a threshold. Instead of it, the decision is made when the map has converged.

3.2 Hybrid model

The model presented in (Iglesias et al., 2008a) is composed of an artificial weighted net of concepts, an evolution module, a transformation module and a assessment module. The net of concepts represent the environment and the expert knowledge about the criteria involved in a decision. A net concept stands for a criterion (C_i) or an event (E_i) whose value may depend on the values of other different events. The association weights between net concepts, like in the rest of the presented models, are considered as a level of influence of the source concept on the destination concept.

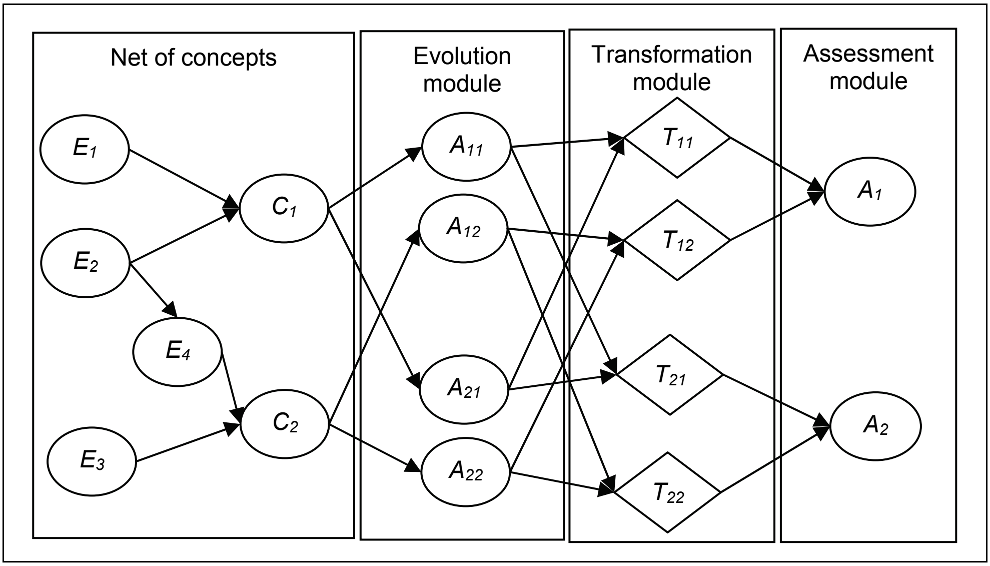


Fig. 5. Connections between the net of concepts, the evolution module, the transformation module and the assessment module. The different nodes represent events (*E_1*, *E_2* and *E_3*), criteria (*C_1* and *C_2*), predicted criteria (*A_{11}*, *A_{12}*, *A_{21}* and *A_{22}*), discrete values (*T_{11}*, *T_{12}*, *T_{21}* and *T_{22}*) and alternatives (*A_1* and *A_2*).

The evolution module takes the values of the criteria stored in the net of concepts and modifies them depending on each alternative. This module predicts the environment evolution, i.e. the consequences a decision would produce over the environment and hence, over the values of the criteria. The transformation module applies a fuzzy transformation (T_{ij}) to obtain three discrete values (maximum, most possible and minimum value) from each evolved criterion. Finally, the assessment module takes the discrete values and scored each alternative using one of three evaluation methods: the general fuzzy method, a fuzzy method based on eigenvectors or influence diagrams. The assessment module ranks the alternatives depending on the score computed by the selected evaluation method and chooses the best alternative.

As figure 5 shows, the assessment mechanism of this model lies in the first three layers. The choice mechanism is represented by the evaluation method applied in the assessment layer.

4. Rule-based models

These models assume that multiple decision rules coexist in the brain. Some rules are based on heuristics and other rules involve deliberative calculation. On the one hand, heuristics rules allow decision makers to avoid irrelevant information and make fast decisions. On the other hand, deliberative rules allow decision makers to evaluate complex situations in order to extract relevant information.

An example of this category of models is DECIDER (Levine, 2009). This model is composed of a module that represents the decision maker's needs and a module of decision rules. The state of the needs module determines which rule must be applied through the orienting module. Depending on the pattern identified by the network contained in each rule module, the model applies the corresponding rule. Figure 6 shows a simplified version of DECIDER.

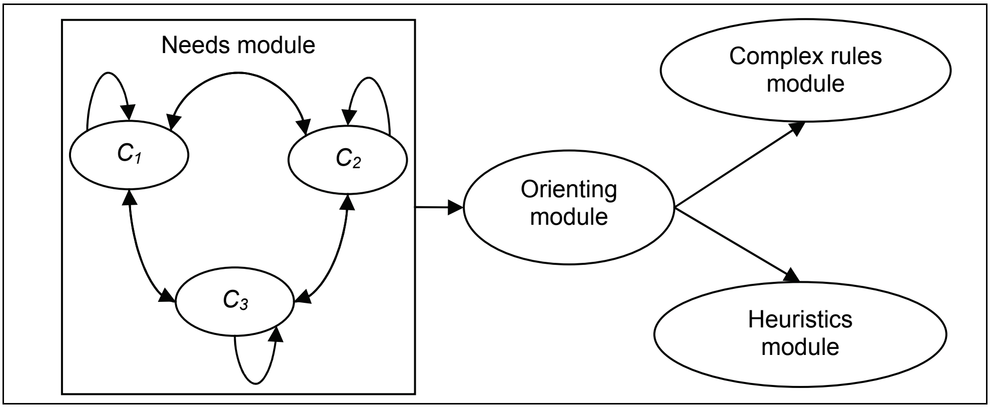


Fig. 6. Diagram that illustrates a simplify version of DECIDER

The assessment mechanism is compiled in the needs module that establishes which rule must be applied. The choice mechanism lies in the rule that chooses the best alternative.

5. Emotional models

Neurophysiological and neuropsychological evidences demonstrate that emotions are an indispensable requirement for deciding properly (Damasio, 1994; Simón, 1998; Pessoa, 2008). Therefore, it is necessary to take into account the participation of emotions in decision processes. There are several models that include emotions in the decision process in order to describe better the decision mechanisms. The following section presents a representative instance of these models.

5.1 Integrated cognitive-motivational network

This section introduces a model for integrating cognition and emotion into a single decision process (Busemeyer et al. 2007). This model is an extension of the Decision Field Theory (Busemeyer & Townsend, 1993). The momentary evaluations of the Decision Field Theory are affected here in this model by emotions. The effect of a criterion on a momentary evaluation of a consequence depends on two factors: the quality and the need for the criterion. The quality represents the amount of satisfaction that a consequence can deliver with respect to a criterion. This model assumes that a subject has an ideal reference on each criterion as well as a current level of achievement for a criterion. The need is the difference between the reference and the current level of achievement for a criterion. The need for a criterion varies across time. These two factors, the quality and the need, are combined to produce a motivational value for a criterion. Then, motivational values are associated with the corresponding decision weights to compute the momentary evaluation.

A scheme of the integrated cognitive-motivational network that shows the influence of emotions in the decision process is shown in figure 7. This model also belongs to the category of Threshold Models because the node called preferences accumulates information about each alternative as in the Decision Field Theory.

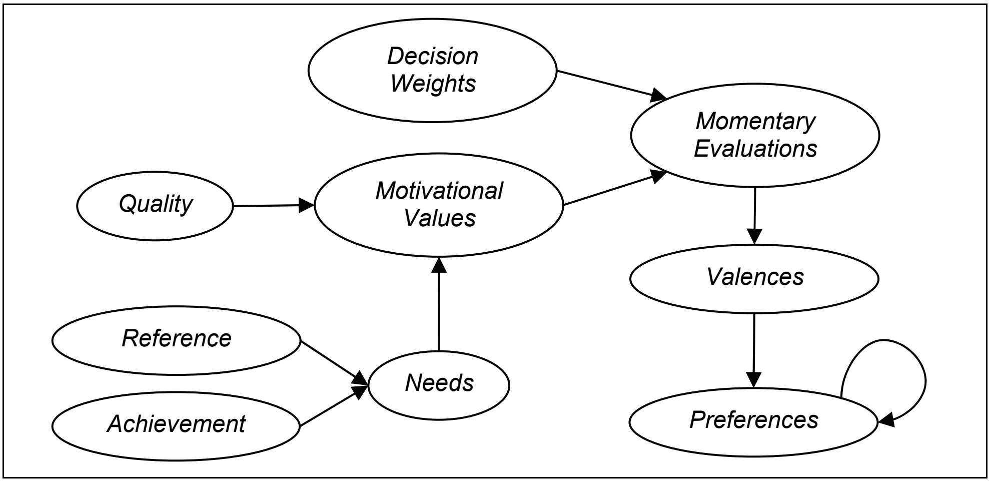


Fig. 7. Cognitive-motivational network

6. Physiologically motivated models

Physiologically motivated models aim to describe the decision process using several interconnected modules which represent different brain areas (e.g. orbitofrontal cortex, amygdala, etc.). There are some brain areas closely related with decision making. One of them is the amygdala, which is capable of assigning emotional meaning to environmental stimuli. The following sections describe some of those models.

6.1 Neural affective decision theory

This theory specifies brain mechanisms underlying decision making. It lies in four principles: affect, brain, assessment and framing. Affect means that decision making is a cognitive affective process dependent on emotional evaluation of alternatives. The brain principle represents that decision making is a neural process driven by coordinated interactions among several brain areas. Assessment suggests that the brain computes preferences via two distinct mechanisms for positive and negative outcomes. Framing defines that judgments and decisions vary depending on the presentation of information.

A representative model of this theory is ANDREA (Litt et al., 2008). ANDREA is divided into seven different modules that represent major brain areas that contribute to the assessment and choice mechanisms: the amygdala, orbitofrontal cortex, anterior cingulated cortex, dorsolateral prefrontal cortex, the ventral striatum, midbrain dopaminergic neurons and serotonergic neurons centered in the dorsal raphe nucleus of the brainstem. Figure 8 illustrates the connectivity scheme between the different modules.

This model describes a biological mechanism for decision making. The assessment mechanism lies in the input to the modules representing the amygdale and the orbitofrontal cortex. The choice mechanism is based on the recurrent connections between all the modules. This model is also in the category of ranking models.

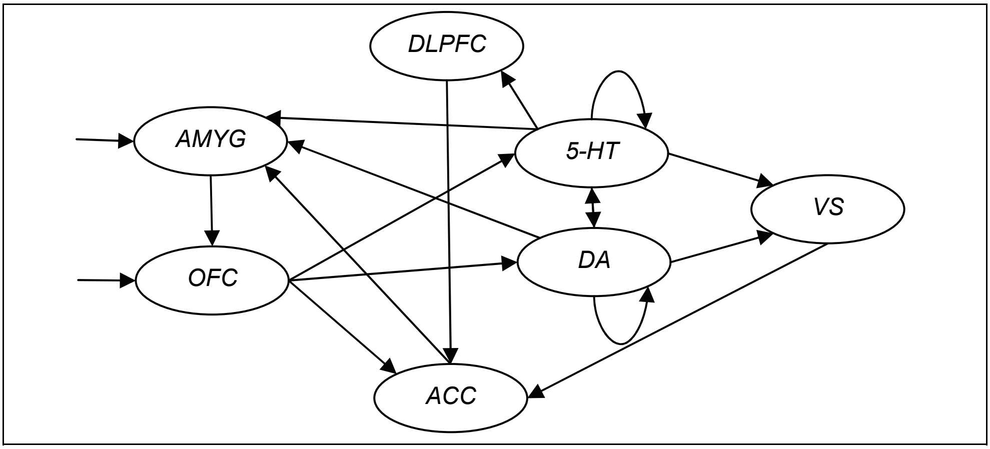


Fig. 8. Scheme of the ANDREA model (5-HT: dorsal raphe serotonergic neurons, ACC: anterior cingulated cortex, AMYG: amygdala, DA: midbrain dopaminergic neurons, DLPFC: dorsolateral prefrontal cortex, OFC: orbitofrontal cortex, VS: the ventral striatum)

6.2 GAGE

Another physiologically motivated model presented in (Wagar & Thagard, 2004) is GAGE. The individual neurons in GAGE are more realistic than those used in most artificial neural network models because they exhibit the spiking behaviour found in real neurons. GAGE organizes neurons into populations related to brain areas, including the ventromedialprefrontal cortex (VMPFC), the hippocampus, the amygdala, the nucleus accumbens and the ventral tegmental area. Figure 9 illustrates a diagram of the neuronal mechanism developed in GAGE.

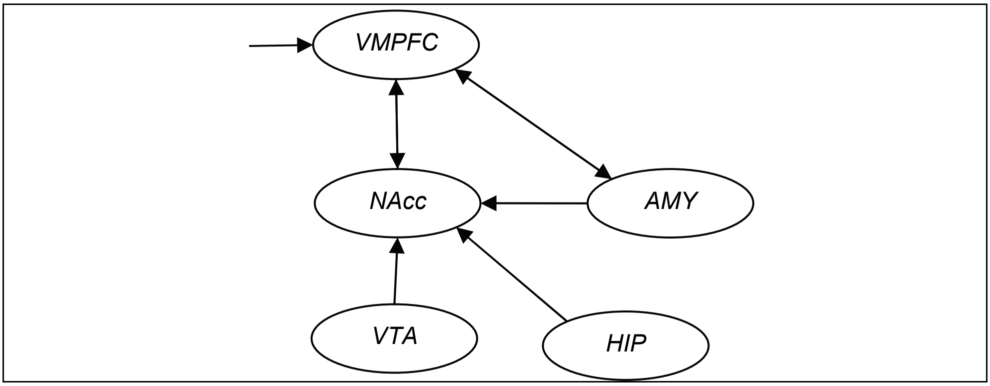


Fig. 9. Scheme of the GAGE model (VMPFC: ventromedial prefrontal cortex, NAcc: nucleus accumbens, VTA: ventral tegmental area, HIP: hippocampus, AMY: amygdala)

The node representing the ventromedial prefrontal cortex receives the inputs to the model so it contains the assessment mechanism. The choice mechanism is finally set by the nucleus accumbens, which is the node with the output of the model. This model is an instance of a ranking model.

6.3 Recurrent network model

The model detailed in (Lo & Wang, 2006) consists of three brain areas: the cortex, the superior colliculus and the basal ganglia. These brain areas are represented as neural networks. Each of the three networks contains competing neural populations for each alternative. Neural populations compete with each other through local recurrent synaptic inhibition. The cortical neurons slowly accumulate information about criteria. The neural population receiving a stronger input has a higher probability of reaching a threshold and winning the competition.

Figure 10 shows the model architecture. Neural pools in the cortical network integrate sensory information about criteria and also compete against each other. They propagate activation to both the superior colliculus and the basal ganglia. The superior colliculus provides the output that represents the winner alternative.

This model is an example of a Threshold model that is also physiologically motivated.

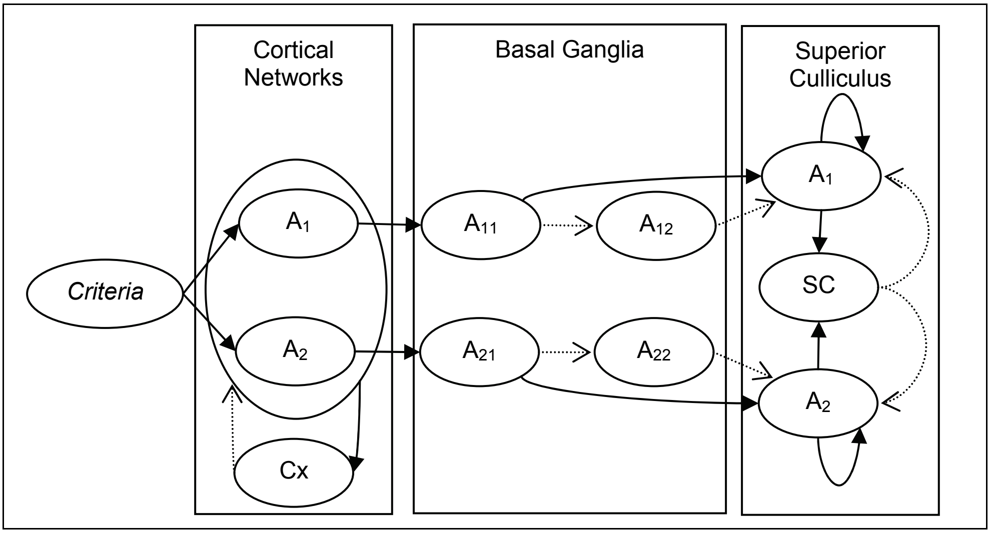


Fig. 10. Schematic model architecture (continuous lines represent positive or excitatory connections while doted lines represent negative or inhibitory connections)

7. Conclusion

The categorization just presented here compiles existing connectionist models related to decision making. The identification of the best model depends on the task to which it will be applied. The evaluation of decision making models is typically conducted experimentally, rather than analytically. There are two main ways to evaluate a model experimentally. The first way lies in computing the ability of the model to take the right decision, i.e. the decision that produces the best outcome. However, there is another interesting way of evaluation which seeks to calculate the ability of the model to take the same decisions as a human being does on a well-defined task. Therefore, if the goal is to find the model that best describes the decision process of human beings, then it might be used the second kind of evaluation. This second way of evaluation is often used to validate a model, suggesting that a good fit to human performance is good evidence for the theory implemented by the model.

In this chapter there are several models that have been applied on different tasks, so it is very difficult to compare them. Threshold Models are often used in decision making problems where time and accuracy are the two most important features. A representative task of this kind is the two-alternative forced choice task (Bogacz et al., 2006). The competitive fuzzy cognitive map is used as a medical decision support system in differential diagnosis. The hybrid model presented in (Iglesias et al., 2008a) is applied in fire emergencies in order to choose the best action to mitigate a crisis. DECIDER is used to model preference reversals by consumers deciding among competing soft drinks. In (Busemeyer et al., 2007), the model is applied in a situation where a motorcyclist is driving towards an obstacle and she/he must decide what to do. ANDREA is used to predict decisions in problems where a human being has to choose between two different lotteries regarding the possible gains and loses and their probabilities. GAGE is used to simulate experimental results concerning the Iowa gambling task (Bechara et al., 1994). Finally, the model detailed in (Lo & Wang, 2006) has been applied in reaction time tasks similar to the two-alternative forced choice task. It would be very useful to present every model within the context of the same decision making problem. One possibility is the use of a simple problem like the one presented in (Iglesias et al., 2008b) that lies in choosing the best means of transport. With this decision making problem, the comparison of the models would be easier and the differences on the theories that they implemented could be more notable.

It seems that in neuroscience and psychology is growing the use of physiologically motivated models as a tool to both test and develop theories. This kind of models explicitly contains modules representing different brain areas. This feature is very suitable in imaging studies that measures brain activity such as functional magnetic resonance imaging (De Martino et al., 2006). It is frequent to find correlations between psychological measures and measures of brain activity (Kahneman, 2009). Therefore, the similarity between brain activities and the values of the model parameters can be interpreted as a clue in the validation of the model.

An important characteristic of a decision making model must be its ability to explain the decisions that it makes. None of the models presented in this chapter seeks the explanation of its decisions. A model developed by the authors of this chapter which aims to explain its decisions while making the same decisions of human beings obtained the second position in the Dynamic Stocks and Flows challenge (Lebiere et al., 2009). The model presented at the challenge is a connectionist model of decision making and it belongs to the Ranking models category. This result confirms that this connectionist model can both explain its decisions and simulate human performance. The explanations of the decisions may lead to better understanding of decision making and they will soon play an important role in the process of studying how people decide.

8. References

Bechara, A., Damasio, A. R., Damasio, H. & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, Vol. 50, No. 1-3, (April-June 1994) 7-15, ISSN 0010-0277

Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen J. D. (2006). The Physics of Optimal Decision Making: A Formal Analysis of Models of Performance in Two-Alternative Forced-Choice Tasks. *Psychological Review*, Vol. 113, No. 4, (October 2006) 700-765, ISSN 0033-295X

Busemeyer, J. R. & Townsend, J. T. (1993). Decision filed theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, Vol. 100, No. 3, (July 1993) 432-459, ISSN 0033-295X

Busemeyer, J. R., Dimperio, E. & Jessup, R. K. (2007). Integrating Emotional Processes Into Decision-Making Models, In: *Integrated Models of Cognitive Systems*, Gray, W. D. (Ed.), 213-229, Oxford University Press, ISBN 9780195189193, New York

Chown, E. (2004). Cognitive Modeling. In: *Computer Science Handbook*, Tucker, A. (Ed.), 69.1-69.13, CRC Press, ISBN 1-58488-360-X, Florida

Damasio, A. R. (1994). Descartes' Error: Emotion, Reason and the Human Brain, Putnam, ISBN 978-0- 3991-3894-2, New York

De Martino, B., Kumaran, D., Seymour, B. & Dolan, R. J. (2006). Frames, Biases, and Rational Decision-Making in the Human Brain. *Science*, Vol. 313, No. 5787, (August 2006) 684-687, ISSN 0036-8075

Fum, D., Del Missier, F. & Stocco, A. (2007). The cognitive modeling of human behavior: Why a model is (sometimes) better than 10,000 words. *Cognitive Systems Research*, Vol. 8, No. 3, (September 2007) 135-142, ISSN 1389-0417

Glimcher, P. W. (2009). Choice: Towards a Standard Back-pocket Model, In: *Neuroeconomics: Decision Making and the Brain*, P. W. Glimcher, C. F. Camerer, E. Fehr, R. A. Poldrack (Ed.), 503-521, Academic Press, ISBN 978-0-12-374176-9, London

Iglesias, A., del Castillo, M. D., Serrano, J. I. & Oliva, J. (2008a). A Comparison of Hybrid Decision Making Methods for Emergency Support, *Proceedings of the Eighth International Conference on Hybrid Intelligent Systems*, pp. 162-167, ISBN 978-0-7695-3326-1, Barcelona, Spain, September 2008, IEEE Computer Society, California

Iglesias, A., del Castillo, M. D., Santos, M., Serrano, J. I. & Oliva, J. (2008b). A comparison between possibility and probability in multiple criteria decision making, In: *Computational Intelligence in Decision and Control*, D. Ruan, J. Montero, J. Lu, L. Martinez, P. D'hondt, E. E. Kerre (Ed.), 307-312, World Scientific, ISBN 978-981-279-946-3, London

Johnson, J. G. & Busemeyer, J. R. (2005). A dynamic, stochastic, computational model of preference reversal phenomena. *Psychological Review*, Vol. 112, No. 4, (October 2005) 841-861, ISSN 0033-295X

Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, Vol. 47, No. 2, (March 1979) 263-291, ISSN 0012-9682

Kahneman, D. (2009). Remarks on Neuroeconomics, In: *Neuroeconomics: Decision Making and the Brain*, P. W. Glimcher, C. F. Camerer, E. Fehr, R. A. Poldrack (Ed.), 503-521, Academic Press, ISBN 978-0-12-374176-9, London

Kosko, B. (1986) Fuzzy Cognitive Maps. *International Journal of Man-Machine Studies*, Vol. 24, No. 1, (January 1986) 65-75, ISSN 0020-7373

Lebiere, C., Gonzalez, C., Dutt, V. & Warwick, W. (2009). Predicting cognitive performance in open-ended dynamic tasks: A modelling comparison challenge. *Proceedings of the 9th International Conference on Cognitive Modeling*, Manchester, UK, July 2009, University of Manchester, Manchester

Levine, D. S. (2009). Brain pathways for cognitive-emotional decision making in the human animal. *Neural Networks*, Vol. 22, No. 3, (April 2009) 286-293, ISSN 0893-6080

Litt, A., Eliasmith, C. & Thagard, P. (2008). Neural affective decision theory: Choices, brains, and emotions. *Cognitive Systems Research*, Vol. 9, No. 4, (October 2008) 252-273, ISSN 1389-0417

Lo, C. C. & Wang, X. J. (2006). Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nature Neuroscience*, Vol. 9, No. 7, (July 2006) 956-963, ISSN 1097-6256

Montibeller, G. & Belton, V. (2009). Qualitative operators for reasoning maps: Evaluating multi-criteria options with networks of reasons. *European Journal of Operational Research*, Vol. 195, No. 3, (June 2009) 829-840, ISSN 0377-2217

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, Vol. 9, No. 2, (February 2008) 148-158, ISSN 1471-003X

Roe, R. M., Busemeyer, J. R. & Townsend, J. T. (2001). Multi-alternative Decision Field Theory: A dynamic connectionist model of decision-making. *Psychological Review*, Vol. 108, No. 2, (April 2001) 370-392, ISSN 0033-295X

Rustichini, A. (2009). Neuroeconomics: Formal Models of Decision Making and Cognitive Neuroscience, In: *Neuroeconomics: Decision Making and the Brain*, P. W. Glimcher, C. F. Camerer, E. Fehr, R. A. Poldrack (Ed.), 33-46, Academic Press, ISBN 978-0-12-374176-9, London

Simón, V. (1998). Emotional participation in decision-making. *Psychology in Spain*, Vol. 2, No. 1, (January 1998) 100-107, ISSN 1137-9685

Smith, P. L. & Vickers, D. (1988). The Accumulator Model of Two Choice Discrimination. *Journal of Mathematical Psychology*, Vol. 32, No. 2, (June 1988) 135-168, ISSN 0022-2496

Stylios, C. D., Georgopoulos, V. C., Malandraki, G. A. & Chouliara, S. (2008). Fuzzy cognitive map architectures for medical decision support systems. *Applied Soft Computing*, Vol. 8, No. 3., (June 2008) 1243-1251, ISSN 1568-4946

Usher, M. & McClelland, J. L. (2001). The Time Course of Perceptual Choice: The Leaky Competing Accumulator Model. *Psychological Review*, Vol. 108, No. 3, (July 2001) 550-592, ISSN 0033-295X

Usher, M. & McClelland, J. L. (2004). Loss Aversion and Inhibition in Dynamical Models of Multialternative Choice. *Psychological Review*, Vol. 111, No. 3, (July 2004) 757-769, ISSN 0033-295X

Vickers, D. (1970). Evidence for an Accumulator Model of Psychophysical Discrimination. *Ergonomics*, Vol. 13, No. 1, (January 1970) 37-58, ISSN 1366-5847

Wagar, B., M. & Thagard, P. (2004). Spiking Phineas Gage: A Neurocomputational Theory of Cognitive-Affective Integration in Decision Making. *Psychological Review*, Vol. 111, No. 1, (January 2004) 67-79, ISSN 0033-295X

5

Data Mining and Decision Support: An Integrative Approach

Rok Rupnik and Matjaž Kukar
University of Ljubljana, Faculty of Computer and Information Science
Slovenia

1. Introduction

Modern organizations use several types of application systems to facilitate knowledge discovery and decision support. Transactional application systems usually have sophisticated reports presenting data by using concepts like sorting, grouping and data aggregation. OLAP systems, also referred to as management information systems, use a data warehouse as a data source and represent a higher-level tool enabling decision support. In such a data warehouse, data are periodically extracted in an aggregated form from transactional information systems and other external sources by data warehouse tools. Both, transactional information systems and OLAP systems, are generally based on concepts of sorting, grouping and data aggregation, where with data aggregation one of the aggregating functions like sum, minimum, maximum, count and average are used. Both, transactional application systems reports and OLAP systems enable the presentation of different viewpoints on aggregated data in different dimensions, the latter however presenting more dimensions than the former (Bose & Sugumaran, 1999).

However, the advancement of strategies for information analyses, business prediction and knowledge extraction have lagged the corresponding developments in data storage and representation, especially in real world applications. Among the reasons for this, there are the inherent complexity of certain aspects of real world problems and the lack of data analysis expertise among business planners, which are compounded by confusing marketing literature produced by a few vendors about the capabilities of their analyses and prediction tools. These factors are combined to reduce the business planners' degree of belief in many of these tools, which in turn leads to these tools being given less importance in key decisions, hence causing the IT vendors to focus less on the development and implementation of the cutting edge techniques.

Traditionally, statistical and OLAP tools have been used for an advanced data analysis. It is often assumed that the business planners would know the specific question to ask, or the exact definition of the problem that they want to solve. Both methods follow what is in essence a deductive approach (Jsr73, 2004), which has several drawbacks. They put a significant strain on the data analyst, who must take care of inventing a hypothesis and storing it in an appropriate way (Hirji, 2001). They lack algorithmic approach and depend on the analysts' insight, coincidence or even luck for acquiring the most valuable information, trends and patterns from data. And finally, even for the best analyst there is a limitation to a number of attributes he can simultaneously consider in order to acquire accurate and valuable information, trends and patterns (Goebel & Gruenwald, 1999). It seems that with the increase in data volume, traditional manual data analysis has become insufficient, and methods for efficient computer-based analyses indispensable. From this need, a new interdisciplinary field of data mining arose. Data mining encompasses statistical, pattern recognition, and machine learning tools to support the analysis of data and discovery of principles that lie within data given.

In the past decades several data mining methods have emerged, showing high potentials for knowledge discovery from data and decision support. Performing analysis through data mining follows an inductive approach of analyzing data where machine learning algorithms are applied to extract non-obvious knowledge from data (Jsr73, 2004). Data mining reduces or even eliminates the above mentioned disadvantages. As opposed to classical data analysis techniques, data mining strategies often take a slightly different view, i.e. the nature of the data itself could dictate the problem definition and lead to discovery of previously unknown but interesting patterns in diverse business applications, for example sales forecasting during promotion events, inventory optimization, and customer profiling and segmentation.

Data mining methods also extend the possibilities of discovering information, trends and patterns by using richer model representations (e.g. decision rules, trees, tables , ...) than the usual statistical methods, and are therefore well-suited for making the results more comprehensible to the non-technically oriented business users. It may well be that by the introduction of data mining to information systems the knowledge discovery process and decision process will move to a higher quality level.

The mission of information systems is, among other things, to facilitate decision support and knowledge discovery. Both, the decision process and the knowledge discovery process are dependent on each other. Knowledge discovery, on one hand, enables accumulation of knowledge and as a result facilitates better decision process. On the other hand, decisions set rules and directions which influence objectives for knowledge discovery. The use of data mining within information systems consequently means the semantic integration of data mining and decision support.

1.1 Motivation

The motivation for our pioneering work in integration of decision support system with data mining methods originates from a real-world problem. Recently we have performed an extensive CRM survey for a leading local GSM operator.

One of the aims of the survey was to explore and demonstrate various approaches and methods for the area of analytical CRM. The survey clearly revealed the benefits of the use of data mining for analytical CRM. The survey was performed between 2002 and 2003 by the authors of the chapter, who had primary roles in the survey project.

As the survey progressed, it turned out that immense quantities of raw data had been collected and needed to be assessed. Immediately after the survey had been conducted, the development project for data mining application system was initiated and managed by the group executing survey. The application system is called DMDSS (Data Mining Decision Support System). DMDSS application system will be introduced later on in the chapter.

1.2 Structure of the chapter

The chapter is organized as follows. In Section 2 we are introducing the basic concepts, underlying ideas of data mining in general and in particular. We are focusing on data mining subfields and particular methods that were used in our study. We are going to outline crucial similarities and differences between data mining on one side and on-line analytical processing and statistical approaches on the other side.

In Section 3 we are introducing motivation for the use of data mining within information systems, and highlight some historical examples of case studies. We are presenting historical generations of data mining, data mining standards, data mining process model (CRISP-DM) and provide motivation for its use in practice.

In Section 4 we are going to proceed to our case study of using data mining in a decision support system developed for a leading local GSM operator. We are describing the development process, review some practical considerations, introduce functionalities of DMDSS and present end-users' experiences after one year of practical use. At the end we are introducing the semantic contribution of the use of DMDSS in the information system.

In Section 5 we are going to summarize our work and provide some conclusions as well as directions for future work.

2. An introduction to data mining

"*Now that we have gathered so much data, what do we do with it?*"

This is the famous opening statement of the editorial by Usama Fayyad and Ramasamy Uthurusamy in the Communications of the ACM, Special issue on Data Mining (Fayyad & Uthurusamy, 1996). Recently, many statements of this kind have appeared in journals, conference proceedings, and other materials that deal with data analysis, knowledge discovery, and machine learning. They all express concern about how to "make sense" from the large volumes of data being generated and stored in almost all fields of human activity.

Especially in the last few years, the digital revolution has provided relatively inexpensive and available means to collect and store the data. The increase in data volume causes greater difficulties in extracting useful information for decision support. The traditional manual data analysis has become insufficient, and methods for efficient computer-based analysis indispensable. From this need, a new interdisciplinary field of data mining was born. Data mining encompasses statistical, pattern recognition, and machine learning tools to support the analysis of data and discovery of principles that lie within the data.

Results of computer-based analysis have to be communicated to humans in an understandable way. In this respect, the analysis tools have to deliver transparent results and most often facilitate human intervention in the analysis process. A good example of such methods are symbolic machine learning algorithms which, as a result of data analysis, aim to derive a symbolic model (e.g., a decision tree or a set of rules) of preferably low complexity but high transparency and accuracy. Being in the core of data mining, the interest and research efforts in machine learning have been largely increased.

Data mining is about automated extraction of hidden information from large databases. They consist of digitised information which is easy to capture and fairly inexpensive to store. But why do people store so much data? Besides the fact that it is easy and convenient to do so, people store data because they think some valuable assets are implicitly coded within it. In scientific endeavours, data represents observations carefully collected about some phenomenon under study. In business, data captures information about critical markets, competitors, and customers. In manufacturing, data captures performance and optimisation opportunities, as well as the keys to improving processes and troubleshooting problems.

Raw data is rarely of direct benefit (Witten & Frank, 2000). Its true value is predicated on the ability to extract information useful for decision support or exploration and understanding the phenomena governing the data source. Traditionally, the analysis was strictly a manual process. One or more analysts would become familiar with the data and—with the help of statistical techniques—provide summaries and generate reports. In fact, the analysts acted as sophisticated query processors. However, such an approach is rapidly breaking down as the quantity of data grows and the number of dimensions increases. Who could be expected to "understand" millions of cases, each having hundreds of fields (attributes)? Further complicating this situation, the amount of data is growing so fast that manual analysis (even if possible) simply cannot keep pace.

The allure of data mining is that it promises to improve the communication between users and their large volumes of data and allows them to ask of the data complex questions such as: "What has been going on?" or "What are the characteristics of our best customers?" The answer to the first question can be provided by the data warehouse and multidimensional database technology (OLAP) that allow the user to easily navigate and visualize the data. The answer to the second question can be provided by data mining tools built on variety of machine learning algorithms: decision trees and rules, neural networks, nearest neighbour, support vector machines, and many others (Chen et al., 1996).

Data mining takes advantage of advances in the fields of artificial intelligence (AI) and statistics. The challenges in data mining and learning from data have led to a revolution in the statistical sciences (Hasti & Tibisharani, 2002). Both disciplines have been working on problems of pattern recognition and classification. Both communities have made great contributions to the understanding and application of different paradigms, such as neural nets and decision trees.

Since statistical discovery is essentially a hypothesis-driven deductive process (Hirji, 2001), an analyst generates a series of hypotheses for patterns and relationships, and uses statistical tests against the data to verify or disprove them. But what happens when the number of variables being analyzed is in the dozens or even hundreds? It becomes much more difficult and time-consuming to find a good hypothesis (let alone be confident that there is not a better explanation than the one found), and analyze the database with statistical tools to verify or disprove it.

Data mining is different from statistical discovery because rather than to verify hypothetical patterns, it uses the data itself to uncover such patterns. It is essentially a discovery-driven inductive process, where a data mining tool is used to formulate (induce) hypotheses from data completely by itself or with moderate guidance from the analyst (Glymour et al., 1996).

Data mining does not replace traditional statistical techniques. On the contrary; it is an extension of statistical methods that is in part the result of a major change in the statistics community. The development of most statistical techniques was, until recently, based on elegant theory and analytical methods that worked quite well on the modest amounts of data being analyzed. This is especially the case in sensitive application areas such as finance or medicine (Kukar et al., 1999; Kononenko, 2001) where useful results are required regardless of the amount of data.

The increased power of computers and their lower cost, coupled with the need to analyze enormous data sets, have allowed the development of new techniques based on a brute-force exploration of possible solutions.

New techniques include relatively recent algorithms like neural nets, decision rules and trees, and new approaches to older algorithms such as discriminant analysis. By virtue of bringing to bear the increased computer power on the huge volumes of available data, these techniques can approximate almost any functional form or interaction on their own, as well as provide additional information on the functional solution (Kukar, 2003). Traditional statistical techniques rely on the modeler to specify the functional form and interactions.

The key point is that data mining is the application of these and other AI and statistical techniques to common business problems in a fashion that makes these techniques available to the skilled knowledge worker as well as the trained statistics professional. Data mining is a tool for increasing the productivity of people trying to build predictive models.

3. Introduction of data mining to information systems

Modern organizations use several types of application systems to facilitate knowledge discovery and decision support. Transactional application systems have sophisticated reports presenting data using concepts like sorting, grouping and data aggregation. Data is aggregated on the group level using one of the following aggregating functions: sum, minimum, maximum, count, average, standard deviation and variance. For example: a report can be grouped by customers by regions, showing average profits won in descending order. Such reports represent the basic level of facilitating decision support on tactical level.

OLAP systems, also referred to as management information systems, use data warehouse as data source and represent higher level tool enabling decision support on tactical level. In a data warehouse data are periodically extracted in aggregated form from transactional information systems and other external sources by data warehouse tools. A data warehouse is typically a multidimensional structure where dimensions represent attributes. OLAP systems enable drill-down concept, i.e. digging through a data warehouse from several viewpoints to acquire the information the decision maker is interested in (Bose & Sugumaran, 1999).

Both, transactional information systems reports and OLAP systems, in general base on concepts of sorting, grouping and data aggregation, where with data aggregation one of the aforementioned aggregating functions are used. They both enable the presenting of different viewpoints on aggregated data by different dimensions. Through the use of sorting they also enable the observing of better/worse relations among elements within individual dimension, i.e. attribute. For example: the form in an OLAP system contains a three-dimensional view showing the profit won by products by regions by departments. If the profit is sorted in descending order, the analyst can observe better/worse relations within the profit won by products within region and department.

Transactional information systems reports and OLAP systems both support the knowledge discovery and analysis process, where the analysts are supposed to look for information, trends and patterns. They do it by running numerous reports with several parameter sets and by viewing OLAP forms swapping dimensions and drilling-down through them (Goebel & Gruenwald, 1999). Performing analysis through OLAP and reports running follows a deductive approach of analyzing data (Jsr73, 2004). Such an approach has the following disadvantages:

- This way analysts can discover information, trends and patterns, but they must take care of storing it themselves to a shared source in a form and structure that will enable the knowledge assimilation and the exploiting of knowledge (Hirji, 2001),

- It lacks algorithm-based approach and depends on coincidence or even luck of choosing the right parameter set or the right dimensions to acquire the most valuable information, trends and patterns,
- There is a limitation to a number of attributes a human being can simultaneously consider in order to acquire accurate and valuable information, trends and patterns (Goebel & Gruenwald, 1999).

The introduction of data mining in previous section indicates high potentials of the use of data mining to facilitate knowledge discovery and decision support. Performing analysis through data mining follows an inductive approach of analyzing data where machine learning algorithms are applied to extract non-obvious knowledge from data (Jsr73, 2004). Data mining, on one hand, reduces or even eliminates before mentioned disadvantages, because:

- It enables the creation of models explaining trends and patterns, which can be stored in a standardized form to a shared source,
- It is an algorithm-based approach, because the models are acquired by data mining methods and algorithms,
- The majority of data mining algorithms are capable of creating models based on a large number of attributes.

3.1 Some examples of the use of data mining in information systems

There are a number of examples of successful application of data mining in information systems. In this section we are introducing some of them. Web usage mining is the application of data mining methods to discover patterns from Web data. Web Data can be classified into the following types: content, structure, usage and user profile. The purpose of Web usage mining is to understand better how users use e-commerce applications in order to improve them. The main application areas of Web mining are personalization, system improvement, site modification, business intelligence and usage characterization (Srivastava et al., 2000). Organizations conducting electronic commerce can, without any doubt, benefit from the use of data mining (Kohavi et al., 2002).

Lee and Stolfo (1998) have developed general and systematic methods for network intrusion detection. They use data mining techniques to discover consistent and useful patterns of system features that describe program and user behaviour, and use the set of relevant system features to compute inductively learned classifiers that can recognize anomalies and known intrusions. The discovered patterns help guide the audit data gathering process and facilitate feature selection (Lee & Stolfo, 1998).

A customer retention analysis represents an important type of analysis in the area of analytical CRM (Customer Relationship Management). It represents the area where data mining can be effectively used. A customer retention analysis is extremely important for sales and service related businesses, because it costs significantly more money to acquire a new customer than to retain existing customers. The key application in the area of customer retention is the detection of potential defectors. The possibility to predict defectors is often the key to the eventual retention of the customer (Ng & Liu, 2000; Han et al., 2002).

The education domain offers a fertile ground for many interesting data mining applications. Selecting the right students for a particular course is one of the application areas within the education domain. The aim of the application mentioned is to help students select courses in which they have high prospects to perform well (Ma et al., 2000). There are plenty of other application areas within education domain appropriate for analysis purposes. For example: predictive model for excellent, good and bad students using classification method.

Several authors have indicated insurance as one of the areas with the highest potentials in the use of data mining (Grossman et al., 2002). Fraud detection is the key application area for data mining in insurance companies. A U.S. health insurance company, for example, used data mining methods to detect submitting of false bills. The analysis identified several geographical areas where claims exceeded the norm and the investigation confirmed and detected physicians who were submitting false bills (Furlan & Bajec, 2008).

We believe that in mobile telecommunication industry there is also a high potential in the use of data mining methods. Besides customer retention there are a number of interesting data mining application areas within analytical CRM:

- **Customer segmentation**: the idea of customer segmentation is to acquire typical customer groups in order to perform group targeted marketing.
- **Subsidized mobile phone value analysis**: subsidized mobile phones represent a substantial investment by mobile operator. The idea of the analysis is to acquire a classification model for the customers spending significantly more money for GSM services after purchasing subsidized mobile phone.
- **Various CDR (call data record) analyses**: in a CDR database there are billions of records that represent a good foundation for various areas of analysis, which among others represent a basis for tariff plans adjustments and wireless network planning.
- **Various association analyses**: there are several options for a valuable analysis of relations. For example: the observation of consequences of the level of the use of services after switching a tariff plan.

There are many other areas where data mining was successfully used (Fayyad et al., 1996; Han et al., 2002; Kohavi et al., 2002). Generally we can say that data mining can be effectively used when sufficient amount of data described with sufficient amount of high quality attributes is available.

3.2 Generations of data mining

Examples introduced in previous sections show that companies can use data mining within their information systems in various areas with different objectives. The use of data mining is increasing, but has still not reached the level appropriate to the potential benefits of its use (Kohavi & Sahami, 2000). One of the reasons for that is with no doubt the lack of awareness and understanding of potentials of data mining. We are presenting other reasons in the course of introducing two generations of data mining, which we introduce as two different approaches.

3.2.1 First generation: data mining software tool approach

Data mining today and in the past has been typically used through ad hoc data mining projects (Goebel & Gruenwald, 1999; Kohavi & Sahami, 2000; Holsheimer, 1999). Ad hoc data mining projects are initiated by a particular objective on a chosen area which means defining of the domain. They are performed using data mining software tools. It is the first generation of data mining (Holsheimer, 1999).

Data mining software tools require a significant expertise in data mining methods, modelling methods, databases and/or statistics (Kohavi & Sahami, 2000). They usually operate separately from the data source, requiring a significant amount of additional time spent with data export from various sources, data import, pre-processing (extraction, filtering, manipulation), post-processing (validation, reporting) and data transformation (Holsheimer, 1999; Goebel & Gruenwald, 1999). The result of a project is usually a report explaining the models acquired during the project using various data mining methods.

Data mining software tool approach has a disadvantage in a number of various experts needed to collaborate in a project and in transferability of results and models (Srivastava et al., 2000; Hirji, 2001). The latter indicates that results and models acquired by the project can be used for reporting, but cannot be directly utilized in other application systems.

The disadvantages mentioned can be explored by the disadvantages of data mining software tools, which are implicitly also disadvantages of data mining software tool approach. Data mining software tools are very complex, they often offer a variety of methods that a user must understand in order to use them effectively. Some of the tools do not enable on-line access to a database or they enable on-line access only to a few database systems. On the other hand, some tools do not allow an access to any database and in this case data must first be extracted from a database to a file before used by a tool. Only a few data mining tools support pre-processing activities (Goebel & Gruenwald, 1999; Kohavi & Sahami, 2000).

In our opinion some of the disadvantages mentioned are in a way also advantages. Complexity of tools and a number of various experts needed are in most cases viewed as disadvantages due to the fact that there are not many experts providing the cutting edge of data mining methods and software tools. The involvement of various experts and a data mining expert providing expertise in data mining methods and in complex tool can only positively contribute to the results of a data mining project.

3.2.2 Second generation: a data mining application system approach

The data mining software tool approach has revealed some disadvantages, which point to the following demands for a different approach:

- The end users of the models and results acquired by data mining projects are business users. For that reason we need applications which will enable them to use data mining models effectively (Kohavi & Sahami, 2000; Goebel & Gruenwald, 1999).
- Business users are not interested in using advanced powerful software tools, but only in getting clear and fast answers by using simple-to-use applications (Goebel & Gruenwald, 1999; Holsheimer, 1999; Bose & Sugumaran, 1999; Fayyad &Uthurusamy, 2002).
- In order to achieve the highest level of the use of data mining models and results it must be possible to deploy them to other business applications in order to use them there (Holsheimer, 1999; Geist, 2002).
- The models and results acquired are dependent on data which is not stationary, but is constantly changing and evolving. For that reason ad hoc projects and a data mining software tool approach need to be enhanced by repeating the same model creation (analysis) process in periodic time intervals or at particular milestones (Goebel & Gruenwald, 1999).

The list of demands indicates the characteristics of application system approach for the use of data mining. It is an approach which focuses on business users, enabling them to view data mining models and results in their business domains. Models and results are presented in a user-understandable manner by means of a user friendly and intuitive GUI using standard and graphical presentation techniques (Aggarwal, 2002). Business users can focus on specific business problems covered by domains with the possibility of repeated analysis in periodic time intervals or at particular milestones. Through the use of data mining application systems approach, data mining becomes better integrated in business environments (Goebel & Gruenwald, 1999; Holsheimer, 1999; Kohavi & Sahami, 2000; Bayardo & Gehrke, 2001; Fayyad & Uthurusamy, 2002).

The data mining application system approach is enabled by application systems which use data mining methods. Those application systems can be divided into two categories. The first category are application systems which support the whole knowledge discovery process, where one set of functionalities is used by data mining experts and the rest of the functionalities by business users as described before. We call them data mining application systems. The second category is other business application systems which can utilize data mining models for various purposes. They can either directly access data mining models, or data mining models can be deployed to them. An example of data mining application will be introduced later on in the chapter. The discussion will reflect advantages and disadvantages of data mining software tool approach.

3.3 Data mining standards

Data mining standards undoubtedly represent an important issue for data mining application systems approach and data mining application systems (Holsheimer, 1999). Employing common standards simplifies the development of data mining application systems and business application systems utilizing data mining models (Grossman et al., 2002; Grossman, 2003). With the maturity of data mining standards, a variety of standards-based data mining applications and data mining platforms can be much easily developed (Grossman, 2003). Other fields such as data grids, web services and the semantic web have also developed standards relevant to knowledge discovery (2003; Chu, 2003; Kumar & Kantardzic, 2003). These new standards have the potential for further changing the way the data mining is used (Grossman, 2003).

A considerable effort in the area of data mining standards has already been done within the data mining community. Established and emerging data mining standards address several aspects of data mining (Grossman et al., 2002):

- Models: for representing data mining models,
- Attributes: for representing the cleaning, transforming and aggregating of the attributes used as input for model creation,
- Settings: for representing the algorithm parameters which affect the model creation,
- Process: for creating, deploying and utilizing the models,
- APIs: for unified access to all methods enabling models creating, deploying and utilizing,
- Remote and distributed data: For analyzing and mining remote and distributed data.

In the following subsections we introduce the most important data mining standards.

3.3.1 PMML

The Predictive Model Markup Language (PMML) is an XML standard being developed by the Data Mining Group, a vendor led consortium established to develop data mining standards (Grossman et al., 2002; Grossman, 2003; Clifton & Thuraisingham, 2001; Dmg). It consists of the following components: data dictionary, mining schema, transformation dictionary, model statistics and models. PMML describes data mining and statistical models in addition to some of the operations required for data cleaning and transforming prior to modelling. The aim of PMML is to provide infrastructure for an application to create a model and another application to utilize it (Grossman et al., 2002). PMML has been evolving from version 1.0 to version 2.1., changes and improvements for version 3.0 are being considered as well (Meyer, 2003). PMML directly covers the aspects of models, attributes and settings. Implicitly it also covers the aspect of API, because it provides the standard for infrastructure for manipulating with models: creating, deploying and utilizing.

3.3.2 JDM: the standardised data mining API

The standardized data mining API represents the most important issue for data mining application systems approach having the following advantages:

- Data mining algorithms are not coded by each team of application developers, but by teams which are specialized on data mining algorithms. Consequently, more reliable and efficient algorithms are developed,
- The possibility to leverage data mining functionality using standard API shared by all application systems within information systems reduces risk and cost,
- Standardized API facilitates API of one vendor to be replaced with API of another vendor.

Java technology and scalable J2EE architecture facilitate integration of various application systems within information system. For that reason many business application systems have been developed on J2EE platform in recent time. Consequently Java is probably the best option for standard data mining API, because it enables the integration of data mining application systems with other business applications within information systems (Hornick, 2003). Java based data mining API in a very effective way enables the implementation of data mining application systems approach and the development of data mining application systems and the integration of data mining in other business application systems.

JDM (Java Data Mining) specification has reached final release status in 2004 (JsrHp73). JDM specifies a pure Java API to facilitate development of Java-based data mining application systems and business application systems utilizing data mining models. As existing data mining APIs are vendor-proprietary, this is the first standardised API for data mining. The JDM expert group consists of representatives of several key software companies (including data mining pioneers IBM, SPSS and Oracle), what gives a certain guarantee for an exploitable standard. Detailed introduction of JDM can be found in (Jsr73, 2004).

3.3.3 CRISP-DM: data mining process model

A data mining process model defines the approach for the use of data mining, i.e. phases, activities and tasks that have to be performed. Data mining represents a rather complex and specialised field. A generic and standardized approach is needed for the use of data mining in order to help organizations use the data mining.

CRISP-DM (CRoss-Industry Standard Process for Data Mining) is a non-proprietary, documented and freely available data mining process model. It was developed by the industry leaders and the collaboration of experienced data mining users, data mining software tool providers and data mining service providers. CRISP-DM is an industry-, tool-, and application-neutral model created in 1996 (Shearer, 2000; Clifton & Thuraisingham, 2001; Grossman et al., 2002). Special Interest Group (CRISP-DM SIG) was formed in order to further develop and refine CRISP-DM process model to service the data mining community well. CRISP-DM version 1.0 was presented in 2000 and it is being accepted by business users (Shearer, 2000).

CRISP-DM process model breaks down the life cycle of data mining project into the following six phases which all include a variety of tasks (Shearer, 2000; Clifton & Thuraisingham, 2001):

- **Business understanding**: focuses on understanding the project objectives form business perspective and transforming it into a data mining problem (domain) definition. At the end of the phase the project plan is produced.
- **Data understanding**: starts with an initial data collection and proceeds with activities in order to get familiar with data, to discover first insights into the data and to identify data quality problems.
- **Data preparation**: covers all activities to construct the final data set from the initial raw data including selection of data, cleaning of data, the construction of data, the integration of data and the formatting of data.
- **Modelling**: covers the creation of various data mining models. The phase starts with the selection of data mining methods, proceeds with the creation of data mining models and finishes with the assessment of models. Some data mining methods have specific requirements on the form of data and to step back to data preparation phase is often necessary.
- **Evaluation**: evaluates the data mining models created in the modelling phase. The aim of model evaluation is to confirm that the models are of high quality to achieve the business objectives.
- **Deployment**: covers the activities to organize knowledge gained through data mining models and present it in a way users can use it within decision making.

There are some other data mining process models found in the literature. They use slightly different terminology, but they are semantically equivalent to CRISP-DM (Goebel & Gruenwald, 1999; Li et al., 2002).

4. DMDSS – Data Mining Decision Support System for GSM operator

We have developed a data mining application system for a GSM operator who was the first on the market. In the following part of the chapter it will be called simply a GSM operator. One of their measures to keep prevailing position on the market was the initiation of the survey of CRM implementation.

One of the aims of the survey was to explore and demonstrate various approaches and methods for the area of analytical CRM. The survey clearly revealed the benefits of the use of data mining for analytical CRM. The important statement of the survey was that our GSM operator intolerably needs data mining for performing analysis for CRM purposes. It was stated that the application system approach is more suitable for the introduction of data mining. That statement was outlined after the research introduced in the previous section was conducted. The main reason for choosing the application system approach was the fact that CRM in our GSM operator represent a rather dynamic environment with continual need for repeated analyses based on data mining methods. The results of the survey were several reports and some prototypes. The survey was performed between 2002 and 2003 by the authors of the chapter, who had primary roles at the survey project.

Right after the survey, the development project for data mining application system was initiated and managed by the group executing the survey. The application system is called

DMDSS (Data Mining Decision Support System). DMDSS application system will be introduced in the following part of the chapter.

4.1 Related work

Several data mining application systems have been developed in recent years and introduced in scientific literature. In this section we introduce some of them.

Geist (2002) introduced a framework for data mining application system. He proposes a three view architecture consisting of process view, model view and data view. The process view covers the user interaction with the application system supporting management and controlling of data mining process. It supports the analysis process by supporting different presentation modes. The process view uses the methods offered by the data view and the model view. The data view covers raw data sets providing methods for manipulating the data objects and describing structure of the data. A relational data model is proposed as architecture for a data view. The model view consists of a set of data mining models and methods manipulating models. It supports model creation and the access to all information about the data mining model.

Bose and Sugumaran (1999) introduced Intelligent Data Miner (IDM) application system in his paper. IDM is a Web-based application system intended to provide organization-wide decision support capability for business users. Besides data mining it also supports some other function categories to enable decision support: data inquiry, data interpretation and multidimensional analysis. In the data mining part it supports the creation of models of the following data mining methods: association rules, clustering and classification. Through the use of various visualization methods it supports the presentation of data mining models. On the top-level it consists of the following five agents: user interface agent, IDM coordinator agent, data mining agent, data-set agent and report/visualization agent. The user interface agent provides interface for the user to interact with IDM to perform analysis. It is responsible for receiving user specifications, inputs, commands and delivering results. The IDM coordinator agent is responsible for coordinating tasks between the user interface agent and other three of before mentioned agents. Based on the user specifications, input and commands, it identifies tasks that need to be done, define the task sequence and delegates them to corresponding agents. It also synthesises and generates the final result. The data-set agent is responsible for communication with data sources. It provides interface to data warehouses, data marts and databases. The data mining agent is responsible for creating and manipulating of data mining models. It performs data cleansing and data preparation, provides necessary parameters for data mining algorithms and creates data mining models through executing data mining algorithms. The report/visualization agent is responsible for generation of the final report to the user. It assimilates the results from data mining agent, generates a report based on predefined templates and performs output customization.

Holsheimer introduced the Data Surveyor application system in his papers (Holsheimer, 1999; Holsheimer et al., 1995). He did not emphasize the functionalities it offers to the user. Instead, he put emphasis on the implementation of data mining methods and the interaction between Data Surveyor and database systems. Author describes Data Surveyor as a system designed for the discovery of rules. It is a 3-tier application system providing customized GUI for two organization roles: the data analyst and the end user. It enables the use of data in several RDBMS systems. Important characteristic of Data Surveyor is the ability to store data mining models in RDBMS system.

Heindrichs and Lim (2003) have done research on the impact of the use of web-based data mining tools and business models on strategic performance capabilities. His paper reveals web-based data mining tools to be a synonym for data mining application system. The author states that the main disadvantage of data mining software tool approach is the fact that it provides results on a request basis on static and potentially outdated data. He emphasizes the importance of the data mining application system approach, because it provides ease-of-use and results on real-time data. The author also discusses the importance of data mining application systems through arguing that sustaining a competitive advantage in the companies demands a combination of the following three prerequisites: skilled and capable people, organizational culture focused on learning, and the use of leading-edge information technology tools for effective knowledge management. Data mining application systems with no doubt contribute to the latter. In the paper the author also introduces the empirical test which proves positive effect on dependent variable “Strategic performance capabilities” by independent variables “Web-based data mining tools” and “Business models”.

4.2 Pre-development activities

The development of DMDSS started with several pre-development activities. We are going to introduce them in the following sections.

4.2.1 Platform selection

The first step within the pre-development activities was the platform selection. Platform selection was highly influenced by two important factors. The first factor was the dominant presence of Oracle platform in our GSM operator. More than 80% of data needed for data mining was available in Oracle databases. The second one was the fact that Oracle RDBMS 9i introduced ODM (Oracle Data Mining) option. ODM has two important components. The first component is a data mining engine (DME), which provides the infrastructure that offers a set of data mining services to data mining API (JSR73, 2004). The second component is Java-based data mining API (ODM API), which enables access to services provided by DME.

Before finally accepting Oracle 9i and ODM, an evaluation sub-project was initiated. The aim of the project was to evaluate ODM, i.e. to verify the quality of its algorithms and results. It was the first version of ODM and evaluation was simply necessary to reduce the risk of using an immature product. The evaluation was performed through recreating data mining models on domains of some past projects which were performed by our research group using data mining software tools (Kukar, 2003; Kukar et al., 1999; Kononenko, 2001). The models acquired by past projects and the results acquired by ODM were compared and evaluation gave positive results for the verification of ODM. Another advantage of Oracle 9i is the security issue. As opposed to many other data mining platforms, in case of Oracle 9i data mining, data does not leave the database. Data mining models and their rules are stored in the database, which means that database security provides the control over access to data mining data, i.e. models and rules.

The introduced factors and the result of the sub-project implied the selection of Oracle 9i RDBMS. In order to develop DMDSS in J2EE architecture the JDeveloper development platform and Oracle OC4J were chosen.

4.2.2 Functional and other demands for DMDSS

The analysis of functional and other demands for DMDSS was done simultaneously with the design of data mining process model for DMDSS. Both activities are extremely interrelated, because the process model implies the functionality of an application to a great extent. The design of data mining process model for DMDSS is introduced in the next section.

It turned out that DMDSS directly or indirectly needs three roles: a data administrator, a data mining administrator and a business user. A data mining administrator role should be granted only to users with advanced or at least above-average knowledge of data mining methods and concepts. Business users are business analysts responsible for performing analysis in various business areas.

The analysis of functional and other demands led to the following important conclusions:

- The roles of the data mining administrator and the business user must be supported. Data administrator role and data preparation phase will not be supported by DMDSS, they should be supported by other tools.
- The access to the modules and functionalities should be dependent on user's role. This would prevent business users from using functionalities which demand advanced knowledge of data mining.
- The data mining administrator should have the possibility to create, evaluate and delete models. A set of model statuses should be defined in order to enable the administrator to make only good and useful models available for the business users.
- The data mining administrator should have the possibility to comment on the models and insert them in the database. Business users should have the possibility to see them, which would help them understand and interpret the models better.
- There should be various visualization and representation techniques available in order to enable various methods for model presentation for the business users.
- Before the training of business users there should be a data mining tutorial organized, where they could learn the concepts of data mining, which would enable them to use and truly exploit DMDSS.
- The key issue for the success of DMDSS is to define its functionalities in the way that will enable data mining administrator create and evaluate models. On the other hand, business users should be able to use it effectively with as little data mining knowledge as possible.

4.2.3 Data mining process model for DMDSS

The key pre-development activity was to determine the data mining process model for DMDSS, which would be appropriate for analysts in marketing department of our GSM operator. According to the level of their knowledge of data mining concepts it was obvious that DMDSS process model should enable analysts incorporate it in their decision process.

The analysis of CRISP-DM and other previously introduced data mining process models revealed that they are more appropriate for ad-hoc projects and a data mining software tool approach than for a data mining application system approach. The consequence was that none of them could be directly used for DMDSS and data mining application system approach. The analysis of data mining process models confirmed CRISP-DM as the most appropriate process model.

CRISP-DM was adapted to the needs of DMDSS as a three stage model where the last stage represents the final process model gained through first two stages. The first stage was the execution of business understanding phase, where the aim was to discover the domains with continual need for repeated analysis based on data mining methods. They are referred to as the areas of analysis.

The second stage was the execution of a data mining project for each area of analysis using a data mining software tool approach. The second stage was actually performed through development process of DMDSS, where multiple iterations of all CRISP-DM project phases were supported by iterations of development process and increments of DMDSS. The development process is introduced later on in the chapter.

The aim of executing multiple iterations of all CRISP-DM phases for every project was to achieve improvements in the areas of data preparation and to do the fine-tuning of data mining algorithms used in ODM API through finding proper parameter values for algorithms. Data sets were re-created automatically every night, based on the current state of the data warehouse and transactional databases. After the re-creation of data sets, data mining models were created and evaluated. It was essential to do iterations over longer period of time in order to implement automated procedures for data preparation and monitor the level of changes in data sets and data mining models acquired. One of the demands for DMDSS was the ability for daily creation of models for every area of analysis and for that reason the degree of changes in data sets and data mining models acquired were monitored.

The third stage represents the production phase of DMDSS and the final process model. Multiple iterations performed in the second stage assure the stability of data preparation phase and proper parameter value sets for data mining algorithms for modelling phase. Modelling and evaluation are performed by data mining administrator and deployment by business users. Some other details of DMDSS regarding process model will be introduced later on in the chapter.

4.3 Development of DMDSS

DMDSS was developed by using several diagramming techniques. UML use case diagrams and class diagrams were used for the process modelling of DMDSS. Entity relationship diagrams were used for data modelling. For several reasons we decided to use iterative incremental process model. As already mentioned, one of the reasons for multiple iterations was to achieve improvements and stability in the areas of data preparation and modelling. Iterations were also needed for incremental changes and improvements of functionalities of DMDSS. After the iteration had been finished, the functional testing was performed. Only then the analysis of functionalities were conducted done and based on that, the list of changes and improvements. The list of changes and improvements was used as the list of demands for the next iteration of development.

During development process we developed our data mining API (DMDSS API) based on ODM API. We did it because the interface of ODM API was badly documented and rather inconsistent, especially method naming. A considerable level of knowledge of Java and data mining algorithms was needed to understand ODM API interface and fully exploit it. For that reason we decided to develop DMDSS API, which would have the intuitive interface and method naming. The structure of DMDSS API interface was constructed in order to obscure the frequently changing details of the ODM API implementation, as well as to provide a consistent platform for both supervised and unsupervised data mining. DMDSS API implied the division of development team into two groups: a team developing DMDSS GUI and a team developing DMDSS API. Such a division of the development team was efficient, because the development process could be carried out consequently to a certain extent and developers could be grouped according to their areas of specialization and skills.

4.4 The introduction of DMDSS

In this section we are going to introduce DMDSS. First we are going to introduce its role-aware architecture and roles using DMDSS. Then we are going to introduce concepts of the use and functionalities of DMDSS through some example forms for data mining administrator and business user. The introduction of concepts of use and functionalities is done for classification data mining method supported by DMDSS. We are going to finish the introduction of DMDSS by presenting the experience of the use of DMDSS in our GSM operator.

4.4.1 Role aware architecture of DMDSS

DMDSS supports role-aware menus. Every role has its own role-aware menu which enables the access only to its dedicated modules. Every DMSDSS user is granted one of the following roles: the data mining administrator, the business user and the developer. The last one was introduced for administrative and maintenance purposes. DMDSS allows the developer to maintain the catalogue of areas of analysis. The catalogue of areas of analysis is a group of database tables having the following advantages:

- The lists of areas of analysis are built dynamically, based on the current catalogue contents. This is used in the building of menus and lists of values.
- The name of the training set, attribute names and the name of classification attribute (only for classification) are stored in the database. This enables changes in data sources and its structure without changing of DMDSS program code.
- The translations of keywords used in models are stored in the database. One of the ODM API methods enables the access to the rules of the model and returns the rules as a string. Through the use of translating of the keywords (if, then, in, …) the model presentation can be adopted and changed without changing of DMDSS program code. In order to achieve higher flexibility, every area of analysis has its own keyword translations.

These advantages clearly reveal the flexibility of DMDSS for the introduction of new areas of analysis without changing the program code. The approach with the catalogue of the areas of analysis stored in the database ensures efficient maintenance process. In order to enable more flexible and environment-independent deployment, DMDSS also enables the developer to maintain usernames/passwords for ODM and other ODM parameters. All these parameters are also stored in the database.

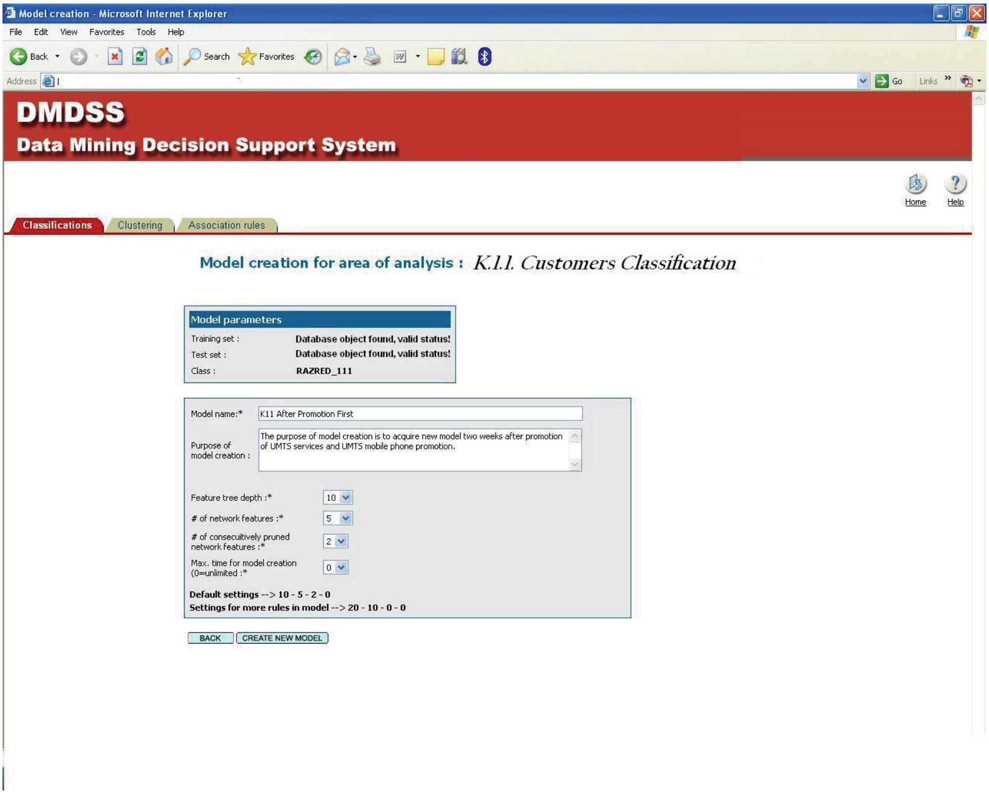
The information support provided for the roles of data mining administrator and business user will be reflected in the following part of the chapter.

4.4.2 Classification

Classification-based areas of analysis were first supported by DMDSS. The classification method was chosen to be a test area for the concepts of GUI and the use of DMDSS and the first four development iterations were dedicated only to classification.

The example of the area of analysis for classification method is called "Customers classification". For the purpose of area of analysis customers are ranked into three categories: a good customer, an average customer and a bad customer. The aim of the area of analysis is to acquire the customer model for each customer category. This information enables business users to monitor characteristics of a particular customer category and plan better marketing campaigns for acquiring new customers. Within the DMDSS application additional areas of analysis for the purposes of mobile phone sales analysis, customer analysis and vendor analysis were also investigated.

The data mining administrator can create classification models by using model creation form (Figure 1). When creating the model they input a unique model name and a purpose of model creation. Beside that there are four algorithm parameters to be set before the model creation. The user can choose the value for each parameter from the interval which was defined as proper in the second stage of process model. At the bottom of the form there are recommended values for parameters to acquire a model with fewer or more rules: default settings for fewer rules in a model, and settings for more rules in a model. The examples of forms in the figures shown below are for the area of analysis called "Customers classification".


Model creation - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address
Go Links
DMDSS
Data Mining Decision Support System
Home
Help
Classifications
Clustering
Association rules
BACK
CREATE NEW MODEL

Model creation for area of analysis : *K.1.1. Customers Classification*

Model parameters

Training set :	**Database object found, valid status!**
Test set :	**Database object found, valid status!**
Class :	**RAZRED_111**

Model name:* K11 After Promotion First

Purpose of model creation : The purpose of model creation is to acquire new model two weeks after promotion of UMTS services and UMTS mobile phone promotion.

Feature tree depth :* 10

# of network features :* 5

# of consecuitively pruned network features :* 2

Max. time for model creation (0=unlimited :* 0

Default settings --> 10 - 5 - 2 - 0

Settings for more rules in model --> 20 - 10 - 0 - 0

Fig. 1. A model creation form for classification

Model testing is performed automatically as the last phase of the model creation. Model testing is an evaluation process to perceive the quality of the model through using machine learning methods.

After the model creation, a data mining administrator can view and inspect the model. Model viewing is supported by two visualization techniques. The first technique is a table where classification rules are presented in a simple IF-THEN form. As already mentioned, keywords used in rules are translated in order to present the rules in a language more appropriate for the users. The second technique is decision trees, where classification rules are converted into decision trees showing equivalent information as rules. The decision trees technique is a graphical technique, which enables visual presentation of rules and for that reason it is very appropriate. While viewing and inspecting, the administrator can input comments for the model. As already mentioned, the role of the comments is to help the business users to understand and interpret the models better.

A data mining administrator can change the status of a model to a published status if the model quality reaches a certain level, and if the model is different from the previously created model of particular area of analysis. Business users can view only the models with published status.

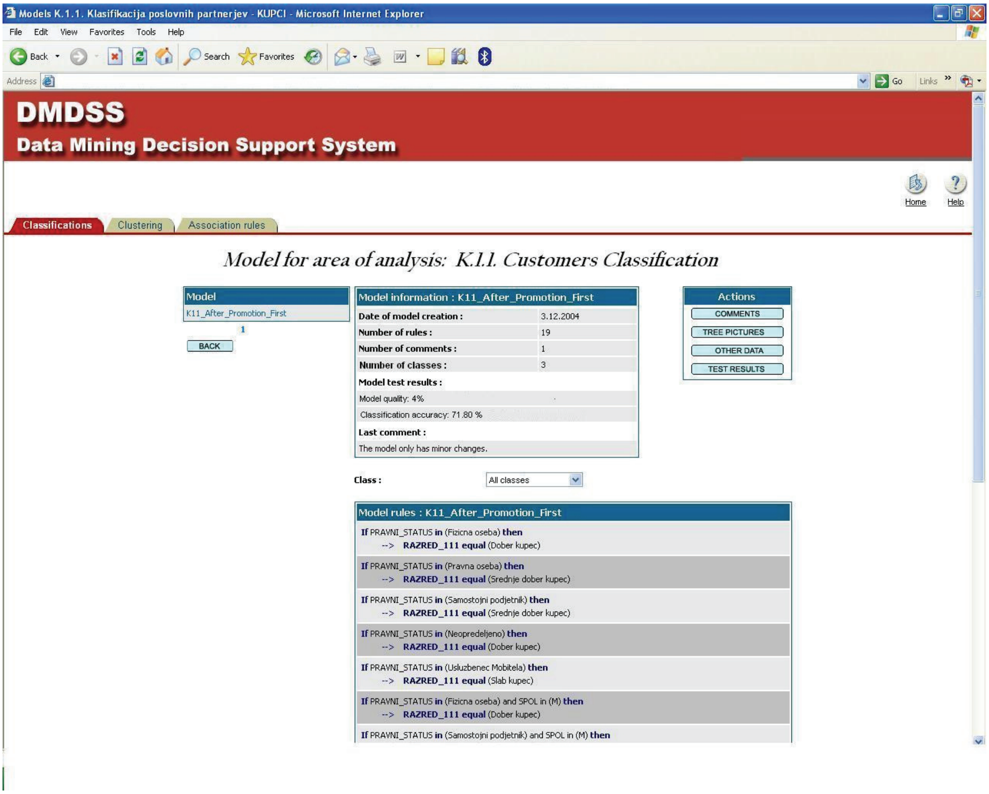
Business users have access to a fewer functionalities than the data mining administrator. The form for model viewing for a business user (Figure 2) is slightly different from the form for model viewing for data mining administrator, but has similar general characteristics. Business users can also view rules in both visualization techniques as the data mining administrator. On the other hand, the form also enables access to some general information about the model: creation date, purpose of model creation, etc. The form also enables business users to view comments of a model written by the data mining administrator.

4.5 The experience of the use of DMDSS

DMDSS has now been in production for several months. During the first year of production there will be supervising and consultancy provided by the development team. Supervising and consultancy have the following goals:

- The role of data mining administrator will be supervised by the data mining consultant form development team, having expertise and experience in data mining. The employee responsible for that role has enough knowledge, but not enough experience yet. Supervising will mainly cover support at model evaluation and model interpretation for data mining administrator and business users;
- Support at defining and introducing new areas of analysis;
- Support at all stages of DMDSS process model before the introduction of new areas of analysis.

Business users use DMDSS at their daily work. They use patterns and rules identified in models as the new knowledge, which they use for analysis and decision process at their work. It is becoming apparent that they are getting used to DMDSS. According to their words they have already become aware of the advantages of continual use of data mining for analysis purposes. Based on the models acquired they have already prepared some changes in marketing approach and they are planning a special customer group focused campaign, based on the knowledge acquired in data mining models. The most important achievement after several months of usage is the fact that business users have really started to understand the potentials of data mining. Suddenly they have got many new ideas for


Models K.1.1. Klasifikacija poslovnih partnerjev - KUPCI - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address
Go Links
DMDSS
Data Mining Decision Support System
Home
Help
Classifications
Clustering
Association rules

Model for area of analysis: K.1.1. Customers Classification

Model
K11_After_Promotion_First

1

BACK

Model information : K11_After_Promotion_First	
Date of model creation :	3.12.2004
Number of rules :	19
Number of comments :	1
Number of classes :	3
Model test results :	
Model quality: 4%	
Classification accuracy: 71.80 %	
Last comment :	
The model only has minor changes.	

Actions
COMMENTS
TREE PICTURES
OTHER DATA
TEST RESULTS

Class : All classes

Model rules : K11_After_Promotion_First

```
If PRAVNI_STATUS in (Fizicna oseba) then
    -->  RAZRED_111 equal (Dober kupec)
If PRAVNI_STATUS in (Pravna oseba) then
    -->  RAZRED_111 equal (Srednje dober kupec)
If PRAVNI_STATUS in (Samostojni podjetnik) then
    -->  RAZRED_111 equal (Srednje dober kupec)
If PRAVNI_STATUS in (Neopredeljeno) then
    -->  RAZRED_111 equal (Dober kupec)
If PRAVNI_STATUS in (Usluzbenec Mobitela) then
    -->  RAZRED_111 equal (Slab kupec)
If PRAVNI_STATUS in (Fizicna oseba) and SPOL in (M) then
    -->  RAZRED_111 equal (Dober kupec)
If PRAVNI_STATUS in (Samostojni podjetnik) and SPOL in (M) then
```

Fig. 2. A model viewing form for business users for classification

new areas of analysis, because they have started to realize how to define areas of analysis to acquire valuable results. The list of new areas of analysis will be made in several months, and after that it will be discussed and evaluated. Selected areas of analysis will then be implemented and introduced to DMDSS according to methodology introduced in the chapter.

The experience of the use of DMDSS has also revealed that business users need the possibility to make their own archive of classification rules. They also need to have an option to make their own comments to archived rules in order to record the ideas implied and gained by the rules. The future plan for classification model utilization is also to apply the model on new customers in order to predict the category a new customer potentially belongs to. These enhancements are planned to be implemented in the future.

4.6 Semantic contribution of the use of DMDSS

While designing and developing DMDSS and monitoring its use by the business users we have been considering and exploring the semantic contribution of the use of a data mining application system like DMDSS in a decision process and performing any kind of business analysis. For that reason one of our goals of the project was also to illustrate the semantic contribution of the use of DMDSS in decision processes. We decided to use the concept of data-model for that purpose.

A data-model is a concept which can be, among other things, used for describing a particular domain on a conceptual level (Bajec, 2001; Lavbic & Krisper, 2009; Sasa et al., 2008; Vavpotic et al., 2009). A meta-model shows domain concepts and relations between them. In this case the meta-model describes a decision process on the conceptual level with emphasis on demonstrating the contribution and the role of the use of DMDSS as data mining application system (Figure 3). UML class diagrams were used as technique for the meta-model. Decision support concepts are represented as classes and relations between them are represented as associations and aggregations. Concepts and relations, which in our opinion represent a contribution of the use of DMDSS in the decision process, are represented in a dotted line style.

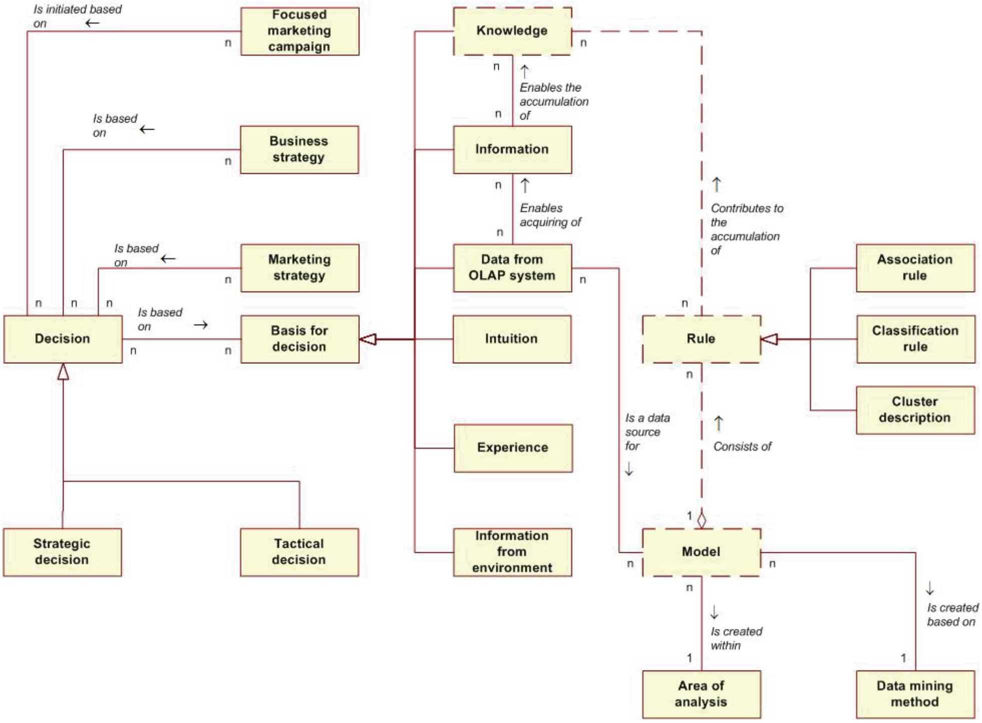


Fig. 3. A meta-model

The meta-model shows various concepts that influence the decision process and represent a basis for a decision. Information technology engineers often believe that decisions mostly depend on data from OLAP systems and other information acquired from information systems. It is true that they represent a very important basis for the decision, although in more than a few cases decisions mostly depend on factors like intuition and experience (Bohanec, 2001).

Knowledge is in our opinion probably the most important basis for the decision, because it enables the correct interpretation of data, i.e. acquiring of information. The contribution of the use of DMDSS and models and rules it creates is in contribution to the accumulation of the knowledge acquired by models and their rules. A detailed description of decision process and creation of a detailed meta-model is beyond the scope of the chapter.

5. Summary and conclusions

DMDSS is a data mining application system which enables a decision support, based on the knowledge acquired from data mining models and their rules. The mission of DMDSS is to offer an easy-to-use tool which will enable business users to exploit data mining with only a basic level of understanding of the data mining concepts. DMDSS enables the integration of data mining into daily business processes and decision processes through supporting several areas of analyses.

The experience of the use of DMDSS has revealed that "traditional" data mining expert role is different, according to the data mining software tool approach. A DMDSDS process model divides the traditional data mining expert role into a data mining administrator role and a data mining consultant. The data mining consultant provides support at defining and introducing of new areas of analysis. The data mining administrator executes daily model creation and provides support for business users. The former must have expertise in data mining; the latter must have enough knowledge of data mining to evaluate models acquired and detect problems at the model creation.

The experience of the use of DMDSS has also revealed that it has become a tool regularly used by business users at decision process and performing various kinds of analyses. They are getting used to DMDSS and they have become aware of the advantages of the continual use of data mining for analysis purposes. After several months of usage, business users have started to realize how to define the areas of analysis to acquire valuable results. We believe that we have succeeded in achieving the optimal data mining process organization and infusion of data mining into decision processes with DMDSS.

The first results of the use of DMDSS are some changes planned in the marketing approach and a special customer group focused campaign based on the knowledge acquired in data mining models. Another result is the revealing of bad data quality, which is a typical side-effect result for the use of data mining. For some areas of analysis bad data quality has been detected in the development of DMDSS and measures at the sources of data have been taken to improve data quality.

Although DMDSS is a rather new application system, there exists a plan for future development of DMDSS. On one hand, there is a list of new areas of analysis being built up by business users, on the other hand there are also enhancements planned in the area of functionalities of DMDSS. There are several directions we intend to explore in the future. We intensively follow the development of the application platform of choice, Oracle Data Mining and accompanying tools, which have already gained certain level of maturity. We intend to provide our users with more data mining methods (e.g. decision trees, rules, …) when they become available. We believe that both satisfying the user requirements as well as providing them with a choice of new data mining methods will contribute to better results of the use of DMDSS.

6. References

Aggarwal, C.C. (2002). Towards-Effective and Interpretable Data Mining by Visual Interaction, *SIGKDD Explorations*, 3, 11-22

Bajec, M. & Krisper, M. (2005). A Methodology and Tool Support for Managing Business Rules in Organisations, *Information Systems*, 30, 423-443

Bayardo, R. & Gehrke, J.E. (2001). Report on the Workshop on Research Issues in Data Mining and Knowledge Discovery Workshop (DMKD 2001), *SIGKDD Explorations*, 3, 43-44

Bohanec, M. (2001). What is Decision Support?. *Proceedings Information Society IS-2001: Data Mining and Decision Support in Action!* (pp. 86-89), Ljubljana, Slovenia

Bose, R. & Sugumaran, V. (1999). Application of Intelligent Agent Technology for Managerial Data Analysis and Mining, *The DATABASE for Advances in Information Systems*, 30, 77-94

Chen, M.S., Han, J. & Yu, P.S. (1996). Data Mining: An Overview from a Database prespective, *IEEE Transactions on Knowledge and Data Engineering*

Chu, R. (2003). XML for Analysis. *KDD-2003 Workshop on Data Mining Standards, Services and Platforms (DM-SSP 03), http://www.ncdm.uic.edu/workshops/dm-ssp03.htm*, Washington DC, USA: ACM

Clifton, C. & Thuraisingham (2001). Emerging Standards for Data Mining, *Computer Standards & Interfaces*, 23, 187-193

Farnstrom, F., Lewis, J. & Elkan, C. (2000). Scalability for Clustering Algorithms Revisited, *SIGKDD Explorations*, 2, 51-57

Fayyad, U.M. & Uthurusamy, R. (1996). Data Mining Knowledge Discovery in Databases (editorial), *Communications of the ACM*, 39, 24-26

Fayyad, U. & Uthurusamy, R. (2002). Evolving Data Mining into Solutions for Insight, *Communications of ACM*, 45, 28-31

Fayyad, U., Haussler, D. & Stolorz, P. (1996). Mining Scientific Data, *Communications of ACM*, 39, 51-57

Furlan, S. & Bajec, M. (2008). Holistic approach to fraud management in health insurance, *Journal of Information and Organizational Sciences*, 32, 99-114

Geist, I. (2002). A Framework for Data Mining and KDD. *Proceedings ACM Symposium on Applied Computing* (pp. 508-513), Madrid, Spain: ACM

Glaymour, C., Madigan, D., Pregibon, D. & Smyth, P. (1996). Statistical Inferrence and Data Mining, *Communications of the ACM*, 39, 35-41

Goebel, M. & Gruenwald, L. (1999). A Survey of Data Mining Knowledge Discovery Software Tools, *SIGKDD Explorations*, 1, 20-33

Grossman, R. (2003). KDD-2003 Workshop on Data Mining Standards, Services and Platforms (DM-SSP 03), *SIGKDD Explorations*, 5, 197

Grossman, L.G., Hornick, M.F. & Meyer, G. (2002). Data Mining Standard Initiatives, *Communications of ACM*, 45, 59-61

Han, J., Altman, R.B., Kumar, V., Mannila, H. & Pregibon, D. (2002). Emerging Scientific Applications in Data Mining, *Communications of ACM*, 45, 54-58

Hasti, T. & Tibisharani, R. (2001). The Elements of Statistical Learning, Springer

Heinrichs, J. & Lim, J.S. (2003). Integrating Web-based Data Mining Tools with Business Models for Knowledge Management, *Decision Support Systems*, 35, 103-112

Hirji, K.K. (2001). Exploring Data Mining Implementation, *Communications of ACM*, 44, 87-93

Hofmann, T. & Buhmann, J. (1997). Active Data Clustering. *Proceedings Advances in Neural Information Processing Systems* (pp. 528-534), Denver, USA

Holsheimer, M. (1999). Data Mining by Business Users: Integrating Data Mining in Business Process. *Proceedings International Conference on Knowledge Discovery and Data Mining KDD-99* (pp. 266-291), San Diego, USA: ACM

Holsheimer, M., Kersten, M. & Toivonen, H. (1995). A Perspective on Databases and Data Mining. *Proceedings International Conference on Knowledge Discovery and Data Mining KDD-1995*, Montreal, Canada: ACM

Hornick, M. (2003). Java™ Data Mining (JSR-73): Overview and Status. *KDD-2003 Workshop on Data Mining Standards, Services and Platforms (DM-SSP 03), http://www.ncdm.uic.edu/workshops/dm-ssp03.htm*, Washington DC, USA: ACM

JSR-73 Expert Group (2004). *Java™ Specification Request 73: Java™ Data Mining (JDM)*, Oracle Corporation and Sun Microsystems, Inc.

Kaufmann, L. & Rousseeuw, P.J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*, New York: John Wiley & Sons

Kohavi, R. & Sahami, M. (2000). KDD-99 Panel Report: Data Mining into Vertical Solutions, *SIGKDD Explorations*, 1, 55-58

Kohavi, R., Rothleder, N.J. & Simoudis, E. (2002). Emerging Trends in Business Analytics, *Communications of ACM*, 45, 45-48

Kononenko, I. (2001). Machine Learning for Medical Diagnosis: History, State of the Art and Perspective, *Artificial Intelligence in Medicine,* 23, 89-109

Kukar, M. (2003). Transductive Reliability Estimation for Medical Diagnosis, *Artificial Intelligence in Medicine,* 29, 81-106

Kukar, M., Kononenko, I., Groselj, C., Kralj, K. & Fettich, J. (1999). Analysing and Improving the Diagnosis of Ischaemic Heart Disease with Machine Learning, *Artificial Intelligence in Medicine,* 16, 25-50

Kumar, A. & Kantardzic, M. (2003). Grid Application Protocols and Services for Distributed Data Mining. *KDD-2003 Workshop on Data Mining Standards, Services and Platforms (DM-SSP 03), http://www.ncdm.uic.edu/workshops/dm-ssp03.htm*, Washington DC, USA: ACM

Lee, W. & Stolfo, S. (1998). Data Mining Approcahes for Intrusion Detection. *Proceedings USENIX Security Symposium*, San Antonio, USA

Lavbic, D. & Krisper, M. (2009). Rapid Ontology Development. *Proceedings of the 19th European-Japanese Conference on Information Modeling and Knowledge Bases*, Maribor, Slovenia

Li, T., Li, Q., Zhu., S. & Ogihara, M. (2002). A Survey on Wavelet Applications in Data Mining, *SIGKDD Explorations*, 4, 49-68

Ma, Y., Liu, B., Wong, C.K., Yu, P.S. & Lee, S.M. (2000). Targeting the Right Students Using Data Mining. *Proceedings International Conference on Knowledge Discovery and Data Mining KDD-2000* (pp. 457-464), Boston, USA: ACM

Meyer, G. (2003). PMML Version 3 – Overview and Status. *KDD-2003 Workshop on Data Mining Standards, Services and Platforms (DM-SSP 03), http://www.ncdm.uic.edu/workshops/dm-ssp03.htm*, Washington DC, USA: ACM

Ng, K. & Liu, H. (2000). Customer Retention via Data Mining, *Artificial Intelligence Review*, 14, 569-590

Sasa, A., Juric, M. & Krisper, H. (2008). Service Oriented Framework for Human Task Support and Automation, *IEEE Transactions on Industrial Informatics*, 4, 292-302

Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining, *Journal of Data Warehousing*, 5, 13-22

Srivastava, J., Cooley, R., Deshpande, M. & Tan, P.N. (2000). Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data, *SIGKDD Explorations*, 1, 12-23

Vavpotic, D. & Bajec, M. (2009). An approach for concurrent evaluation of technical and social aspects of software development methodologies, *Information and Software Technology*, 51, 528-545

Witten, I.H. & Frank, E. (2000). *Data Mining*, Morgan-Kaufmann, 978-1558605527, New York

Yarmus, J.S. (2002). *ABN: A Fast, Greedy Bayesian Network Classifier*, Oracle White Paper, Burlington, MA: Oracle Coorporation

6

Testing Methods for Decision Support Systems

Jean-Baptiste Lamy[1], Anis Ellini[1], Jérôme Nobécourt[1],
Alain Venot[1] and Jean-Daniel Zucker[2]
[1]Laboratoire d'Informatique Médicale et Bioinformatique (LIM&BIO), UFR SMBH, Université Paris 13
[2]LIM&BIO, UFR SMBH, Université Paris 13; Institut de Recherche pour le Développement
France

1. Introduction

Decision support systems (DSS) have proved to be efficient for helping humans to make a decision in various domains such as health (Dorr et al., 2007). However, before being used in practice, these systems need to be extensively evaluated to ensure their validity and their efficiency. DSS evaluation usually includes two steps: first, testing the DSS under controlled conditions, and second, evaluating the DSS in real use, during a randomised trial. In this chapter, we will focus on the first step.

The test of decision support systems uses various methods aimed at detecting errors in a DSS without having to use the DSS under real use conditions; several of these methods were initially developed in the field of expert systems, or software testing (Meyer, 2008). DSS testing methods are usually classified in two categories (Preece, 1994):

- *static methods* do not require to use the DSS. They usually consist in the review of the DSS' knowledge base (Duftschmid & Miksch, 2001), either manually by human experts, or automatically, using programs that search for syntactic, logical or semantic errors in the knowledge base. Static methods are sometimes called *verification*, as they consist in checking whether the DSS meets the requirements specified by the users (are you building the system right?) (Preece, 1998).
- *dynamic methods* do require the use of the DSS. They consist in using the DSS to solve a set of test cases. Various methods have been proposed for (a) choosing test cases that are meaningful for testing purpose, and then (b) for determining whether the DSS outputs are considered as erroneous or not, generally by asking human experts to solve the test cases by hand. Dynamic methods are sometimes called *validation*, as they aim at verifying whether the DSS satisfies the actual users' requirement (are you building the right system?) (Preece, 1998).

  Recently, we have proposed a dynamic method for testing almost exhaustively a DSS (Lamy et al., 2008); it involves a very large set of test cases, including potentially all possible cases. Consequently, the DSS outputs are very numerous and cannot be reviewed directly by a human expert. Thus, the method relies on learning or visualization algorithms (Andrews, 2002) to help reviewing the DSS outputs.

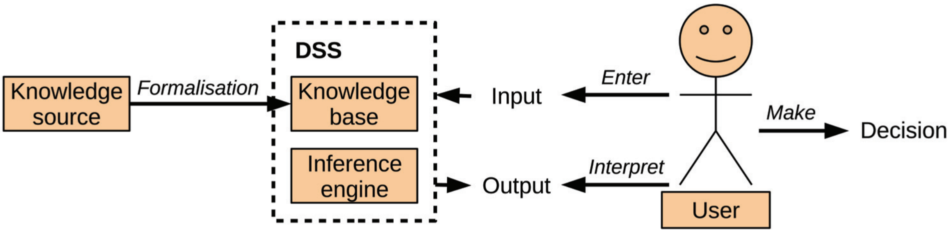


Fig. 1. General schema of a DSS.

In this chapter, we will first propose a classification of the errors that can be found in DSSs. Then, we will describe the various DSS testing methods that have been proposed, and finally we'll conclude by giving advice for choosing DSS testing methods.

2. Classification of errors in DSS

DSSs are usually built from a non-structured knowledge source, for instance a clinical practice guideline (a textual guide that provides recommendations to the physicians about the diagnosis or the therapy for a given disease), a set of cases (for a system using case-based reasoning) or a group of domain experts; this knowledge source is then structured into a knowledge base, for instance a set of rules or a case database, and finally, an inference engine applies the knowledge base to the system's input and determines the output (Fig. 1). Consequently, we can distinguish four main types of error:

- **Errors in the knowledge source**, *e.g.* the French clinical practice guideline for arterial hypertension (HAS, 2005) says "For diabetic patient, angiotensin converting enzyme inhibitors or angiotensin II antagonists are recommended, from the stage of microalbuminuria. For diabetic type 2 patient with other risk factors, angiotensin converting enzyme inhibitors are recommended". The recommendation is ambiguous because it is not clear whether it is "other risk factors than diabetes type 2" or "other risk factors than microalbuminuria", and this can lead to interpretation errors.
- **Errors in the knowledge base**, *i.e.* the structured knowledge base does not exactly translate the knowledge source, *e.g.* the following rule "if patient is diabetic and patient has microalbuminuria, then recommend angiotensin converting enzyme inhibitors or angiotensin II antagonists" does not correspond exactly to the first sentence of the previously cited guideline for hypertension. In fact, the guideline says "from the stage of microalbuminuria", and thus also includes the stages above (such as proteinuria), whereas the rule does not.
- **Errors in the inference engine**, which include both errors in the strategy for executing the knowledge base, *e.g.* in a rule based-system, an engine that does not apply the rules in the desired order, and software bugs in the inference engine.
- **Errors in the use of the DSS**, *i.e.* errors when the user enters the system's input, and errors when the user reads and interprets the system's output. These errors are not located in the DSS itself. However, as DSS are expected to help a human user to make a decision, it sometimes make sense to evaluate the user-DSS couple. Moreover, a badly-designed DSS can mislead the user, for instance, by providing uncommon or incoherent default values for some input. E Coiera *et al.* have studied these errors in the medical context (Coiera et al., 2006); in particular, errors during data entry seem to be quite frequent, and represent an important cause of medication errors.

Errors in the knowledge source are the most problematic, but as they can only be detected and fixed by referring to another source of knowledge, typically human experts, there is few works on them. Errors in the inference engine are less problematic, as the inference engine is normally domain-independent, it can be tested as any other software. As a consequence, most works focus on the errors in the knowledge base.

The errors in the knowledge base are divided in several categories:

- **Syntax errors** occur when the knowledge base does not respect the expected grammar, *e.g.* unbalanced parentheses in a rule database.
- **Logical anomalies**; we speak of "anomaly" and not "error", because a logical anomaly does not always lead to an error in the output of a DSS (Preece & Shinghal, 1994), for instance duplicating a rule in a rule-based system is an anomaly, but it has no influence in the system behavior. However, logical anomalies are often clues of other errors in the knowledge base, such as knowledge errors (see below), for instance a duplicated rule can actually be the same that another rule because a part of the rule has been forgotten. Four types of logical anomalies are considered (Santos et al., 1999; Preece & Shinghal, 1994):
  - **Inconsistency** (also called ambivalence) occurs when the knowledge base can lead to incompatible conclusions for a given input. For instance, a rule-based DSS having the following rules: "if the patient's diastolic blood pressure is inferior to 90 mmHg, the patient does not suffer from hypertension" and "if the patient is diabetic and his diastolic blood pressure is superior to 80 mmHg, the patient suffer from hypertension", because, for a diabetic patient with diastolic blood pressure between 80 and 90 mmHg, the rules conclude that the patient both suffers from hypertension and does not.
  - **Deficiency** occurs when there is missing knowledge in the knowledge base, *i.e.* there are some situations for which the knowledge base leads to no conclusion.
  - **Redundancy** occurs when there is useless elements in the knowledge base, *i.e.* removing these elements from the knowledge base does not affect the DSS's behavior at all. In particular, redundancy includes (but is not limited to) duplicated elements and unsatisfiable conditions, *e.g.* a rule that can never be triggered.
  - **Circularity** occurs when the knowledge base includes some statements that depend only on themselves. For example, the following rules define a circular dependency: "if patient is treated by insulin, then patient's glycemia should be monitored" and "if patient's glycemia is monitored, then patient should be treated by insulin".

The importance of the various types of anomaly depends of the application domain (Preece & Shinghal, 1994).

- **Semantic errors** occur when the knowledge base includes elements that are correct from the logic point of view, but conflicting with domain-specific knowledge. For example, it is a semantic error to conclude that a male patient is pregnant, or to consider a human body temperature of 60°C.
- **Knowledge errors** occur when the knowledge base does not correspond to the knowledge source, although it is syntactically, logically and semantically correct. For example, a clinical guideline says "For diabetic type 2 patient, then it is recommended to start the treatment by a diet", and the associated rule-based knowledge base states that "If patient is diabetic type 2, then start the treatment by prescribing metformin". The example given in the "Errors located in the knowledge base" paragraph is also a knowledge error.

Errors in a DSS can have a more or less important impact, both in term of frequency and gravity. However, the importance of errors is domain dependent. For example, when computerizing a clinical guideline, the guideline is assumed to be the "gold standard", and therefore errors in the knowledge source are not considered, since the source is considered as being the truth. On the contrary, when using a patient database in case-based reasoning, the patient database may be biased and not representative of the new patients for which the DSS is used.

3. Static methods

Static methods test a DSS without requiring to use the DSS. They usually consist in the inspection of the DSS' knowledge base (Duftschmid & Miksch, 2001), either manually by human experts, or automatically. By definition, static methods cannot detect errors in the inference engine or in the use of the DSS.

3.1 Manual approaches

Manual static methods consist in the inspection of the knowledge base by one or more domain experts. Expert inspections can detect errors in the knowledge source or in the knowledge base, however, since humans are not error-proof, they do not guarantee to detect *all* errors of these categories.

Usual recommendations for expert inspections of the knowledge base are the following (Wentworth et al., 1995):

- If the knowledge base has been designed with the support of some domain experts, the inspection should not be done by the same experts, for detecting errors in the knowledge source.
- If possible, the inspection should be done by more than one expert. To obtain an error rate of about 5%, it is required to have at least 4 experts that all agree on the knowledge base content.
- The knowledge base content should not be presented to the experts by someone they know well, for instance a well-known expert of their field (because it could bias their opinion on the knowledge base).

When setting up an expert inspection, several choices must be done. First, a way of presenting the knowledge base should be chosen. The formal representation of the knowledge base is usually not understandable by domain experts. Thus, the knowledge base should be translated into a more human-readable form, such as a set of "if-then" rules expressed in natural language or a decision tree. A more original presentation consists in a set of machine-generated examples for verifying intention-based definitions in the knowledge base (Mittal & Moore, 1996). When the knowledge base is complex, it is often possible to split it into several parts, for instance several decision trees corresponding to various situations. However, one should also verify that the knowledge base and its human-readable translation are really equivalent.

Second, the "gold standard" the knowledge base is compared to, can be either the expert's own knowledge, or the knowledge source used to build the DSS. In the first case, both errors in the knowledge source and in the knowledge base are detected, usually with a stress on the first ones, whereas in the second case, only errors in the knowledge base are detected. For instance, when computerizing a clinical guideline, the experts can be asked to check the knowledge base against their own knowledge, or against the paper guideline.

Finally, if several experts are inspecting the knowledge base, one must decide how to deal with expert disagreements. Disagreements are usually treated by searching a consensus between all the experts, however other methods such as voting have also been proposed.

In conclusion, expert inspections are very interesting for detecting errors in the knowledge source. However, the main drawback of these methods are the difficulties to express the knowledge base in a human-readable way, and to find the experts, since experts are often more motivated by the testing of the complete DSS (especially if the DSS is potentially useful for the expert), than the tedious reading of the DSS's knowledge base.

3.2 Automatic approaches

Automatic static methods rely on programs that search the knowledge base for syntactic, logical or semantic errors. They are sometimes called *verification*. Many of these methods have been proposed for verifying rule-based knowledge bases in expert systems, in the 1980 decade. More recently, some of these methods have been adapted for the verification of other forms of knowledge, such as ontologies (Gómez-Pérez, 1999) or structured clinical guidelines (Duftschmid & Miksch, 2001).

3.2.1 Check for syntax errors

Syntax errors can be found using traditional grammars, such as BNF (Backus-Naur Forms). Pre-formatting tools can also be used when writing the knowledge base, to help preventing syntax errors.

3.2.2 Check for logical anomalies

For rule-based systems, three algorithms have been proposed for detecting logical anomalies (Preece & Shinghal, 1994).

- **Integrity check** considers each rule individually, and checks its validity. It can detect only a few anomalies, such as unsatisfiable conditions.
- **Rule pair check** considers each pair of rules separately. It can detect all anomalies that involve only two rules, such as two inconsistent rules. However, some inconsistency may involve more than two rules, and are not detected.
- **Extension check** considers all the possible paths in the rule of the knowledge base. It can detect all logical anomalies.

Rule pair check requires more computation time than integrity check, and extension check requires even more time. However, it has been shown that even extension check can be achieved in an acceptable computation time on real-world knowledge bases for medical diagnosis, fault diagnosis, and product selection (Preece & Shinghal, 1994), and for power system control centers (Santos et al., 1999). Specific methods have also been proposed for verifying temporal constraints (Duftschmid et al., 2002).

3.2.3 Check for semantic errors

Checking semantic errors requires that the testing program includes some domain-specific knowledge. This additional knowledge typically consists in parameters' possible values (*e.g.* the human body temperature should be within 36°C and 43°C), and combinations of incompatible parameters values (*e.g.* the following combination sex=male and pregnant=true is incompatible) (Duftschmid & Miksch, 2001).

The detection of semantic errors is usually performed at the same time than the detection of logical anomalies (Preece & Shinghal, 1994), for instance integrity check can verify the parameters' values, and rule pair check or extension check can detect sets of rules that lead to conclusions that are semantically inconsistent. The additional knowledge considered for semantic error can also be taken into account in the detection of logical anomalies; for example it is not a deficiency anomaly to consider only three possible cases, a female pregnant patient, a female non-pregnant patient and a male non-pregnant patient, because the fourth case, a male pregnant patient, is not semantically correct.

In conclusion, automatic static verification methods are very interesting for detecting syntactic, logical or semantic errors and anomalies. Their main advantage is their automatic nature: it is easy to perform the test again when the DSS has been modified, and they ensure to detect all anomalies of a given type in the knowledge base (whereas an expert that manually reviews a knowledge base might not see an error). However, they also have several drawbacks. First, they cannot detect errors in the knowledge sources, the inference engine, and knowledge errors in the knowledge base. In many situations, such as the implementation of clinical practice guidelines, the main difficulty is to structure the knowledge source, and therefore knowledge errors are the more problematic ones. In these situations, automatic static methods are not helpful. Second, these methods work only on declarative knowledge, but not on procedural knowledge, and it is not always easy to transform a procedural knowledge into a declarative one. Finally, the detection of semantic errors requires the addition of domain-specific knowledge, which makes these methods less automatic (a domain expert may be necessary) and raise the question of the verification of this additional knowledge. For all these reasons, automatic static methods are no longer the more active field in DSS verification.

4. Dynamic methods

Dynamic methods test the DSS by running it over some test cases, and they often require the intervention of domain experts for checking the results obtained in the test cases.

4.1 Test bases

Traditional dynamic methods, sometimes called empirical testing, involve the use of a test base that includes a limited number of test cases (compared to the usually very high number of possible cases). To set up such a study, the first step is to build the test base; several methods have been proposed for choosing the cases in the test base.

First, when they are available, real cases can be used, *e.g.* from a patient database for a medical DSS, or a server log for a network monitoring DSS. However, it may be difficult to obtain all the input values required by the DSS, and it is sometimes required to complement the cases.

Second, the test cases can be arbitrarily chosen. A group of final users or domain experts can be asked to write a set of test cases, or, during evaluation, each evaluator can be asked to enter test cases of his choice. DSS designers can also choose and propose test cases that correspond to the difficulties they encountered during the design of the DSS, such as test cases for ambiguous situations or for situations that previously lead to an error in the DSS's outputs (to ensure that these errors have not been reintroduced).

Third, test cases can be generated at random. A basic method consists in creating a case by randomly choosing a value for each DSS input variable. More sophisticated methods can involve semantic constraints (*e.g.* to avoid generating test cases involving a male pregnant patient).

Fourth, various methods have been proposed for the automatic generation of “optimal” test cases, using heuristics. A first approach is to partition the input domain in several subdomains, each of them associated to a sub-domain of the output domain, and then to generate one or more test cases for each sub-domain. A. Preece reviewed the methods for partitioning (Preece, 1994): in *equivalent class partitioning*, the input domain is partitioned in subdomains that lead to the same output; in *risk-based partitioning*, input and output domains are partitioned in ten partitions according to the level of risk they can cause in real life, in particular, various metrics can be used to determine which test cases are the more complex to deal with; in *structure-based partitioning*, one partition is created for each *path* in the DSS (the definition of *path* being DSS-dependent, and potentially subject to discussion). Vignollet *et al.* (Vignollet & Lelouche, 1993) proposed another “optimal” approach for rulebased systems, which take into account the inference engine strategy, and generate a test base that triggers every rules of the knowledge base at least once during the test. Sensitivity analysis (Sojda, 2007) is a third approach, which considers cases that test the behaviour of the DSS for extreme input values, and ensure that the output evolves as expected when an input value increases or decreases. For instance, in a DSS for diabetes type 2 therapy, glycosilated haemoglobin is a marker of the gravity of the disease, and therefore, when glycosilated haemoglobin increases, the recommended treatment should not be weaker.

Finally, it is possible to build a test base by mixing several methods, *e.g.* by including both real and random cases. For validating a DSS, a good test base should typically include (a) realistic test cases (either real cases or cases written by final users or domain experts), (b) test cases for ambiguous or problematic situations, written by DSS designers, (c) randomly generated test cases, and (d) possibly “optimal” test cases.

Depending on how the right DSS output for each test case is determined and who runs the DSS, there are several possible protocols for the evaluation:

1. For real test cases corresponding to past data, the expected DSS outputs can be observed in the real life. In this case, the DSS designers can run the DSS and compare the DSS outputs to the observed ones. For instance, a DSS for predicting the evolution of bird populations have been tested on real past data (Sojda, 2007).
2. A group of experts is asked to determine the right output for each test case, according to their own expertise. In case of disagreement between experts, a consensus should be obtained. Then the DSS designers run the DSS and compare the DSS outputs to the expert’s ones.
3. When a gold standard is available, such as a clinical guideline in the medical domain, a group of experts is asked to interpret the gold standard and determine the right output for each test case according to the gold standard (even if the experts disagree with it). In case of disagreement between experts, a consensus should be obtained. Then the DSS designers run the DSS and compare the outputs to the expert’s ones.
4. Each expert runs the DSS and compares the DSS outputs to his personal opinion.

In the three first protocols, the right outputs for each test case are determined first, and then the DSS is run by the DSS designers (or a technician). In the last protocol, the DSS is run by the experts and there is no absolute “right” output for each test case, since each expert is free to compare the DSS output to his own opinion, possibly different from the ones of the other experts. To avoid bias, the experts involved in the testing should not have been involved in the DSS design.

Several measures have been proposed for quantifying the effectiveness of a DSS over a test base (Guijarro-Berdiñas & Alonso-Betanzos, 2002): contingency tables (including false positive and false negative rates), percentage agreement and the Kappa statistic for pair tests (*i.e.* comparing the DSS to a gold standard or a single expert), and Williams' index, cluster analysis and Multi-Dimensional Scaling (MDS) for group tests (*i.e.* comparing the DSS with a group of several experts).

Depending on the choice done for generating the test cases, and the evaluation protocol, validation over a test base can be used in various situations, and it can potentially discover all types of errors listed in section 2. The main drawback of this method is that the number of test cases is necessarily limited, and therefore it cannot ensure the absence of errors in the DSS for other cases.

4.2 "Exhaustive" method

Recently, we have proposed a new dynamic testing method that runs the DSS over a very high number of test cases, allowing an almost exhaustive testing (Lamy et al., 2008). As test cases are far too numerous to let human experts review the DSS outputs for each test case, the method relies on the use of learning algorithms or visualisation techniques to help verifying the DSS's outputs. The method includes three steps:

1. Generate an exhaustive (or almost exhaustive) set of the DSS input vectors, and run the DSS on each input vector to obtain the associated output. It is possible to generate an exhaustive set of input vectors by considering a set of variables expressing the various elements of input for the DSS, and generating all possible combinations of the variables' values. If the input vector includes continuous variables, they should be limited to a few values. Semantic constraints can be added to exclude impossible or infrequent cases.
2. Extract knowledge from the set of (input vector, output result) pairs by applying learning or generalization algorithms, or generate a graphical representation of the (input vector, output result) pairs.
3. Let an expert review the knowledge or the graphical representation produced at step 2, and compare them to the original knowledge source used to design the DSS, or to his own opinion.

We applied this method for testing the ASTI critiquing module, a medical DSS implementing therapeutical recommendations from clinical guidelines, and aimed at raising alerts whenever a physician's prescription does not follow the recommendations. The ASTI critiquing module includes knowledge bases for six diseases: type two diabetes, hypertension, tobacco addiction, dyslipaemia, atrial fibrillation and thrombo-embolic risk. In a first study (Lamy et al., 2008), we used Quinlan C4.5 algorithm (Quinlan, 1993) to generate a decision tree from an almost exhaustive set including hundreds of thousands (input vector, output result) pairs for each disease. To ensure 0% of error in the decision tree, pruning was disabled. However, for hypertension, the extracted decision tree was too huge for being human reviewed, and thus this testing method has not been applied to this disease.

To evaluate this approach, errors were introduced in the DSS. All the errors introduced were clearly visible on the decision tree.

In a second time, we built tables from more limited set of about thousands (input vector, output result) pairs. We divided the input vectors in two parts: the clinical profile (including comorbidities, and various patient characteristics such as age or sex; each clinical guideline lead to about ten profiles), and the treatments (including the current treatment, and the

Current treatment													Prescribed treatment
no treatment	diet	gemfibrozil	anionic resin	ezetimib	nicotinic acid	pravastatin	simvastatin	atorvastatin	rosuvastatin	statin + ezetimib	statin + resin	statin + nicotinic acid	
○	○	○	○	○	○	○	○	○	○	○	○	○	no treatment
●	○	○	○	○	○	○	○	○	○	○	○	○	diet
○	○	●	●	●	●	◑	◑	◑	◑	○	○	○	gemfibrozil
○	○	●	●	●	●	◑	◑	◑	◑	○	○	○	anionic resin
○	○	●	●	●	●	◑	◑	◑	◑	○	○	○	ezetimib
○	○	●	●	●	●	◑	◑	◑	◑	○	○	○	nicotinic acid
○	●	●	●	●	●	●	●	●	●	●	●	●	pravastatin
○	●	●	●	●	●	●	●	●	●	●	●	●	simvastatin
○	●	●	●	●	●	●	●	●	●	●	●	●	atorvastatin
○	○	●	●	●	●	●	●	●	●	○	○	○	rosuvastatin
○	○	○	○	○	○	○	○	○	○	○	○	○	statin + ezetimib
○	○	○	○	○	○	○	○	○	○	○	○	○	statin + resin
○	○	○	○	○	○	○	○	○	○	○	○	○	statin + nicotinic acid

Table 1. Example of the use of table for the graphical visualisation of the inputs and outputs of the ASTI critiquing module for patient with hypercholesterolaemia (a type of dyslipaemia) with no other risk factors. The current treatment of the patient is shown horizontally, and the treatment prescribed by the physician vertically. The DSS output is indicated by the following symbols: ● means that the DSS considers the prescribed treatment as conform to the recommendations, ○ that the DSS considers the prescribed treatment as non conform, and ◑ that the DSS considers the prescribed treatment as conform only if the current treatment is poorly tolerated by the patient (but not if it is inefficient). This table summarizes 338 test cases.

treatment prescribed by the physician). Then, for each clinical profile, we built a table displaying the current treatment horizontally, the prescribed treatment vertically, and at the intersections the corresponding DSS outputs, *i.e.* the conformity of prescribed treatment to the recommendations. Table 1 show an example of such a table.

Although not as exhaustive as the decision trees of the previous approach, these tables provided an interesting overview of the DSS behaviour. In particular, it is easy to visually detect some patterns on the graphical presentation, for instance, it is easy to see in Table 1 that the DSS behaviour is the same if the current treatment is a statin (pravastatin, simvastatin, atorvastatin or rosuvastatin). All the six diseases supported by the ASTI critiquing module were tested using this approach, and it allowed us to find several errors that were present but not discovered in the decision trees.

Compared to the standard dynamic methods that rely on small test bases, the “exhaustive” method tests the system over a much larger number of test cases. However, it is more complex to set up, and may not be suitable for all DSS.

The first approach, based on learning algorithms, should be quite easy to adapt to other DSSs. However, when the generated decision tree is too huge for being human-reviewed, the method cannot be applied. In addition, if the DSS knowledge base includes rules that are more complex that basic "if-then" rules with and / or operators, such as "if x out of y statements are true, then...", it might be necessary to use more sophisticated learning algorithms than C4.5. Several knowledge representations can also be used as alternative to decision trees, such as production rules or flowcharts; Wentworth *et al.* review them in the chapter 6 of their book (Wentworth et al., 1995).

The second, graphical, approach can only be used if it is possible to represent the inputs and outputs of the DSS in one or a few tables; this point is highly domain-dependent. Other visualisation technics could be used as alternatives to tables, such as 2D or 3D bar charts, or star-plot glyphs (Lanzenberger, 2003). Our experiments shew that both approaches are complementary, as they allowed to find different errors.

5. Conclusion: How to choose a testing method?

In the preceding sections, we have presented four main categories of testing methods: manual static methods, automatic static methods, test bases, and "exhaustive" dynamic methods. We have seen that all these methods have their own advantages and drawbacks: there is no perfect or ideal DSS testing method. In addition, it has been shown that the various methods do not detect the same errors (Preece, 1998). Therefore we recommend to apply several methods.

Table 2 shows the types of errors that can be found by the various testing methods. One should combine several testing methods so that the combination of methods covers all types of errors. In addition, one should typically:

- combine both static and dynamic methods,
- combine both automatic methods and methods relying on domain experts, and
- in a test base, mix test cases chosen randomly or by the system developers and test cases as close as possible to real cases (either real test cases or test cases written by final users).

In table 2, automatic static methods cannot detect knowledge errors in the knowledge base. As a consequence, these methods are not very useful when knowledge errors are frequent, which typically occurs when the knowledge source used to create the DSS is complex and difficult to interpret.

Another important element to take into account when choosing DSS testing methods is whether there is a "gold standard" knowledge source in the field covered by the DSS, or not. For instance, when designing a DSS to implement a clinical practice guideline, the decisions recommended by the guideline are assumed to be the best possible decisions, and therefore it is a "gold standard" knowledge source. In this case, errors in the knowledge source are not to be considered, and consequently one should favor protocol #3 for test base evaluation.

Finally, another question is related to the order in which the various testing methods should be applied. It is usually admitted that verification and static methods should be performed before validation and dynamic methods. Another advice is to perform automatic methods before methods relying on experts, because, if the DSS was heavily modified consequently to the first test, it is usually easier to perform again automatic testing rather than the work with the experts.

	Static methods		Dynamic methods	
Error types	Manual	Automatic	Test base	Exhaustive
Errors in the knowledge source	++[1]/-	-	+[2]/-	++[1]/-
Errors in the knowledge base				
Syntax errors	++	++	+	++
Logical anomalies				
Inconsistencies	++	++	+	++
Deficiencies	++	++	+	++
Redundancies	++	++	-	-
Circularities	++	++	+	++
Semantic errors	++	++	+	++
Knowledge errors	++	-	+	++
Errors in the inference engine				
Strategy errors	-	-	+	++
Software bugs	-	-	+	++
Errors in the use of the DSS				
Input entry errors	-	-	+[3]/-	-
Output interpretation errors	-	-	+[3]/-	-

Table 2. The various types of error in a DSS, and the test methods that can be used to detect them. "-" indicates that the method cannot detect error of this type. "+" indicates that the method can detect the errors of this type and covers only a part of the knowledge source, knowledge base or inference engine functionalities. "++" indicates that the method can detect the errors of this type and covers the whole knowledge source, knowledge base or inference engine functionalities.

In conclusion, many methods have been proposed for testing DSS, each of them having its own advantages and weaknesses. Correctly used, these methods can detect a lot of errors in a DSS. After testing the DSS thoroughly, the next step in the DSS evaluation is to set up a randomized trial in real use conditions, in order to ensure that final users really perform significantly better with the DSS than without, but also that the use of the DSS does not introduce other sources of errors (Coiera et al., 2006), such as automation bias, *i.e.* the user follows the DSS recommendations without question at all, or on the contrary errors of dismissal, *i.e.* the user totally ignore the DSS recommendations (or deactivate the system, if the user is allowed to).

6. References

Andrews, K. (2002). *Information visualisation: tutorial notes*, Graz University of Technology.

Coiera, E., Westbrook, J. & Wyatt, J. (2006). The safety and quality of decision support systems., *Yearbook of medical informatics* pp. 20–25.

[1] only if the gold standard is the expert knowledge, and not the knowledge source.

[2] protocols #1, #2 and #4 only (see section 4.1).

[3] protocol #4 only.

Dorr, D., Bonner, L., Cohen, A., Shoai, R., Perrin, R., Chaney, E. & Young, A. (2007). Informatics systems to promote improved care for chronic illness: a literature review, *J Am Med Inform Assoc* 14(2): 156–163.

Duftschmid, G. & Miksch, S. (2001). Knowledge-based verification of clinical guidelines by detection of anomalies, *Artif Intell Med* 22: 23–41.

Duftschmid, G., Miksch, S. & Gall,W. (2002). Verification of temporal scheduling constraints in clinical practice guidelines, *Artif Intell Med* 25(2): 93–121.

Gómez-Pérez, A. (1999). Evaluation of taxonomic knowledge in ontologies and knowledge bases, *Proceedings of the North American Workshop on Knowledge Acquisition, Modeling, and Management (KAW)*, Vol. 2, Banff, Alberta, Canada.

Guijarro-Berdiñas, B. & Alonso-Betanzos, A. (2002). Empirical evaluation of a hybrid intelligent monitoring system using different measures of effectiveness, *Artif Intell Med* 24(1): 71–96.

HAS (2005). Prise en charge des patients adultes atteints d'hypertension art´erielle essentielle, Available at http://www.has-sante.fr/portail/display.jsp?id=c 269118.

Lamy, J.-B., Ellini, A., Ebrahiminia, V., Zucker, J.-D., Falcoff, H. & Venot, A. (2008). Use of the C4.5 machine learning algorithm to test a clinical guideline-based decision support system, *Stud Health Technol Inform* 136: 223–228.

Lanzenberger, M. (2003). The interactive stardinates - design considerations, *Proceeding of Human-Computer Interaction (INTERACT'03)*, IOS Press, Zurich, Switzerland, pp. 688–693.

Meyer, B. (2008). Seven principles of software testing, *IEEE Computer* 41(10): 99–101.

Mittal, V. & Moore, J. (1996). Detecting knowledge base inconsistencies using automated generation of text and examples, *Proceeding of the 16th conference on Artificial Intelligence (AAAI-96)*, Portland, Oregon, pp. 483–488.

Preece, A. (1994). Validation of knowledge-based systems: The state-of-the-art in north america, *Journal of communication and cognition - Artificial intelligence* 11: 381– 413.

Preece, A. (1998). Building the right system right - Evaluating V&V methods in knowledge engineering, *Proceedings of the eleventh workshop on Knowledge Acquisition, Modeling and Management (KAW'98)*, Voyager Inn, Banff, Alberta, Canada.

Preece, A. D. & Shinghal, R. (1994). Foundation and application of knowledge base verification, *Int J Intell Syst* 22(8): 23–41.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*, Morgan Kaufmann.

Santos, J., Faria, L., Ramos, C., Vale, Z. & Marques, A. (1999). *Multiple approaches to intelligent systems*, Vol. 1611/2004 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, chapter Verification of knowledge based-systems for power system control centres, pp. 316–325.

Sojda, R. (2007). Empirical evaluation of decision support systems: Needs, definitions, potential methods, and an example pertaining to waterfowl management, *Environmental Modelling & Software* 22: 269–277.

Vignollet, L. & Lelouche, R. (1993). Test case generation using KBS strategy, *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'93)*, Chamb´ery, France, pp. 483–488.

Wentworth, J., Knaus, R. & Aougab, H. (1995). *Verification, validation, and evaluation of expert systems*, Vol. 1, A FHWA Handbook.

7

Decision Support Systems for Pharmaceutical Formulation Development Based on Artificial Neural Networks

Aleksander Mendyk and Renata Jachowicz
Dept. of Pharmaceutical Technology and Biopharmaceutics Jagiellonian University Medical College, Medyczna 9 Str, 30-688 Kraków, Poland

1. Introduction

Once discovered and established as therapeutic agent, the drug substance is used for pharmacotherapy of various diseases. The drug substance itself has unique properties, which in certain cases do not allow for effective therapy. This is the area, where pharmaceutical technology allows to improve drug substance original characteristics by optimization of pharmaceutical formulation. The latter is a complicated process involving many variables concerning formulation qualitative and quantitative composition as well as technology parameters. This chapter will be dedicated to the computer systems based on artificial neural networks allowing for guided pharmaceutical formulation optimization.

2. Artificial neural networks (ANN) foundations

The artificial neural networks (ANNs) are non-linear, information-processing systems designed in a manner similar to the biological neural structures, which is expressed in the structural and the functional composition of ANNs. The latter is based on so-called connectionist model of neural systems. It assumes that topology and electrophysiology of synapses (connections) in the brain or other biological neural systems are the key factors of neural systems ability to process information (Hertz et al. 1991; Wikipedia, 2009c, Żurada 1992).

One of the several definitions of ANNs is that they are dispersed knowledge processing systems built from so-called “nodes” hierarchically organized into the layers. This definition does not implement the most important feature of ANNs which is their ability to learn on the available data. Thus, ANNs are representatives of Computational Intelligence paradigm in contrast to classical Artificial Intelligence systems, where all the knowledge of the system must be implemented from the scratch by the programmer.

Typical ANN of the most common Multi Layer Perceptron type (MLP) is built on four main elements (Fig. 1):

1. input layer
2. hidden layer(s)
3. output layer
4. connections (weights)

Each layer consists of few "nodes" which in fact are artificial neurons connected between layers via “weights” – artificial synapses. The information flow is unidirectional from the input to the output.

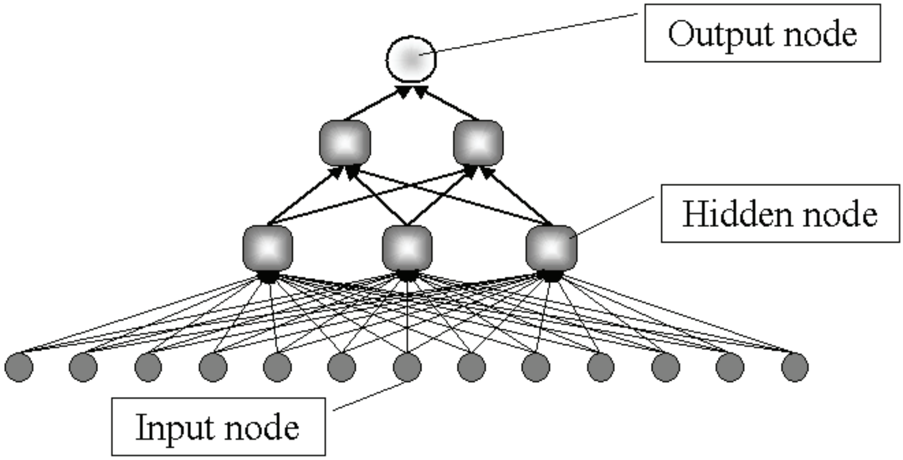


Fig. 1. Typical structure of MLP ANN.

MLP ANN works in two phases:

1. training
2. testing

The training phase is based on the iterative presentations of the available data patterns in order to teach ANN to perform designated task. Since MLP ANNs are supervised training systems, they have to be presented with data on the input and output as well. This allows for adjusting weights values in such a manner that ANN becomes competent in the designated task. Adjusting of the weights is performed automatically with use of special algorithm designed for this purpose. One of the most common training algorithms for ANNs is back propagation (BP), where the teaching signal is the difference between current output and the desired one and is propagated backwards from the output layer to the input layer in order to modify weights values (Fig. 2). The whole procedure is automatic and once started does not require any intervention from the user.

According to the connectionist model of the neural systems, ANNs topology is the most important factor influencing their modeling abilities. The topology of ANNs, called also architecture, is expressed in terms of number of layers and nodes in each layer. However, it is not the nodes themselves but number, signs and values of connections between the particular nodes, which encode the knowledge of the system. Since all the BP procedure is automatic, user does not have to put any assumptions about a model shape a priori to the system, thus ANNs represent empirical modeling approach. Automatic training procedure and model identification by ANNs are the most commonly known advantages of these systems. Another advantage is their superior ability to identify non-linear systems. It is because ANNs are usually built on non-linear activation functions, therefore being non-linear systems themselves. Next distinguishing feature of ANNs is their relative ease of dealing with large number of data cases and features. However, so-called curse of

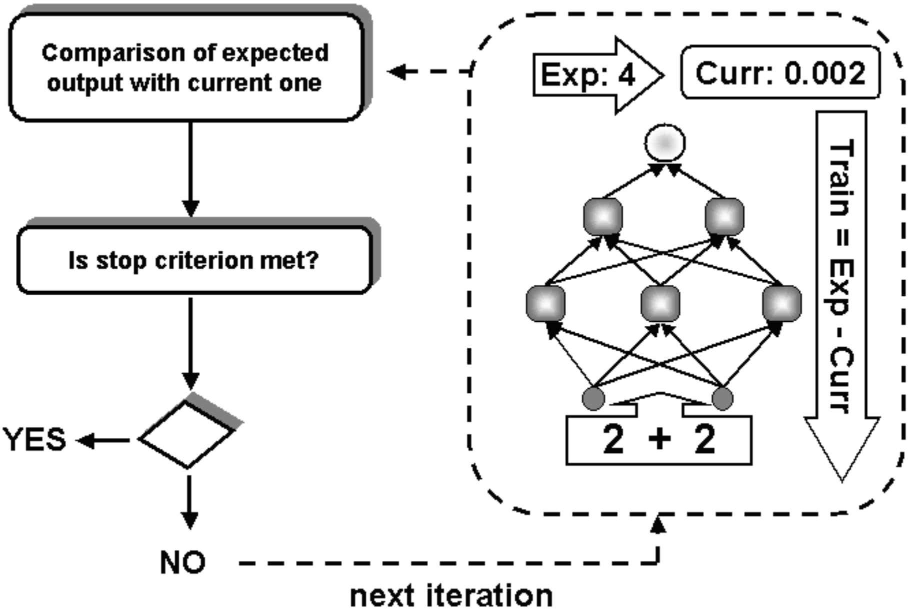


Fig. 2. Scheme of the back propagation algorithm.

dimensionality is also applicable to the ANNs, nevertheless it is less pronounced than for classical statistical systems. Moreover, ANNs are able to decide on inputs importance, thus providing sensitivity analysis feature, which is a way to reduce unnecessary inputs. It improves system performance but also provides knowledge about analyzed problem derived from ANNs behavior. Therefore, ANNs are also used as data mining tools allowing for automated knowledge extraction.

All the features of ANNs described above, allow using them as generic, empirical modeling tools in vast areas of science and technology:

- economy
- engineering
- chemistry
- neurobiology
- medicine and pharmacy

Although, it is impossible to present all applications of neural networks, there might be named major areas of their usage:

- signal processing (noise reduction, compression)
- pattern recognition and features extraction (handwriting, facial recognition, medical imaging, fraud detection)
- forecasting (financial, medical, environmental).
- data mining

Pharmaceutical applications of ANNs are still far from being routine, however ANNs are gradually coming into the focus in different pharmacy areas: pharmacokinetics (Brier & Aronoff, 1996; Brier & Żurada, 1995; Chow et al., 1997; Gobburu & Chen, 1996; Veng-Pedersen & Modi, 1992), drug discovery and structure-activity relationships (Huuskonen. et al, 1997; Polański, 2003; Taskinen & Yliruusi, 2003), pharmacoeconomics and epidemiology (Polak & Mendyk, 2004; Kolarzyk et al, 2006), in vitro in vivo correlation (Dowell et al., 1999) and pharmaceutical technology (Behzadia et al. 2009; Hussain et al., 1991; Bourquin et al., 1998a, 1998b, 1998c; Chen et al., 1999; Gašperlin et al., 2000; Kandimalla et al., 1999; Mendyk & Jachowicz, 2005, 2006, 2007; Rocksloh et al., 1999; Takahara et al., 1997; Takayama et al., 2003; Türkoğlu et al., 1995).

3. Empirical modeling as decision support systems (DSS):

3.1 General remarks

Decision support systems (DSS) are usually computer information processing tools that support decision-making activities in the field of particular interest (Wikipedia, 2009c). As computer tools, they are generally understood as an extension of commonly known expert systems – the systems derived from artificial intelligence field (AI). The expert systems' definition "enhancement" allows, among other differences, to use "black box" models in contrast to the classical hard AI systems, where the system behavior is algorithmic, thus understandable on the every level of its action. DSS exploit every available techniques of data processing in the benefit of accuracy of decision making support. This includes ANNs as well, which will be advertised in this chapter as very suitable tools for DSS in the pharmaceutical technology.

Every DSS has to include basic set of elements:

a. knowledge base
b. model or so-called inference machine
c. user interface (Hand et al., 2001)

A knowledge base is usually consisting a set of all available information gathered in the strictest organizational way that is possible to achieve. This includes data-formatting and preprocessing in order to make it easier to be processed by any numerical analysis tools to be employed in the future. It is a very tedious and complicated task and in the same time is crucial to the future system accuracy.

The knowledge sources might be categorized into two main classes:

a. empirical results
b. theoretical background

If available, both sources might be combined in the benefit of the DSS. In pharmaceutical technology there is a lot of strong physicochemical background, which allows for describing pharmaceutical formulations in terms of their components properties. However, pharmaceutical formulations are very complicated structures, where many factors play, sometimes not very well defined, role in their behavior. Complexity of the pharmaceutical formulations, including their preparation technology, make them very difficult to classical analytical description. Hundreds of well defined physicochemical factors are becoming well defined description only, without practical meaning for prospective decision support. Regarding this it is noteworthy, that so far in pharmaceutical technology empirical knowledge plays still most important role in particular problem description. It is that's why in this field, when numerical analysis of the data is employed, empirical modeling becomes the tool of the choice to create appropriate model (the inference machine). It allows to create the model based on the data only, without a priori assumptions and therefore without a need of a priori knowledge. The model is created based on the data only, which reflects current state of knowledge about the problem. With lack of the well established theories present, partially verified hypotheses or theories from different fields could be even misleading, therefore the model based on the data only has the advantage of lack of bias. Typical examples of empirical modeling tools are ANNs, which became very handy tools for empirical modeling implementation. Specifically, ANNs can work in two main modes:

a. predictive modeling
b. data mining

As it would be shown below, both modes are complementary to each other, which is another example of smooth and effective work of ANNs.

The user interface is a final part of DSS to be prepared and is strictly dependent on the particular problem specifics.

Complete algorithm of DSS preparation with emphasis on ANNs use could be described as follows:

1. Definition of the model function
2. Preparation of the knowledge database
    a. data acquisition
    b. data preprocessing
        - definition of input and output vector
        - scaling, normalization, noise addition, classes balancing
    c. splitting original dataset to two nonequal datasets according to k-fold cross-validation scheme
3. Construction of inference engine as ANN model
    a. ANN training and search for optimal (or suboptimal) architecture
    b. validation by k-fold cross-validation scheme
    c. sensitivity analysis and input vector reduction if applicable
    d. preparation of the higher order models – expert committees (ensembles)
4. User interface preparation

The above scheme depicts main steps to be performed in order to create DSS with use of ANNs. After preparatory phase including points 1 & 2, the modeling procedures have to be employed (p. 3). ANNs are used as tools to model relationships of interest in particular problem. This is usually done by creation of the predictive models designed to answer the question what would be the action of the new component introduction or modification of qualitative/qualitative composition. This would help to decide whether to use or not the composition tested in silico in the prospective laboratory experiments. The search for the most promising formulations-candidates could be realized in the most simplistic way as a combinatorial approach where there are set boundary conditions (i.e. the set of available excipients) and criteria of optimal formulation acceptance (Fig. 3). In case of the DSS total failure, i.e. all predictions were falsified by laboratory experiments, it is possible to enter interactive mode, (Fig. 3 dotted line) where the results of final (unsuccessful) laboratory experiments are added to the initial database and used for subsequent modeling procedure. Re-training of the neural models is usually much easier than the original step of optimal ANN model search, thus the interactive mode could be of choice when very little information is available at the beginning of the analysis.

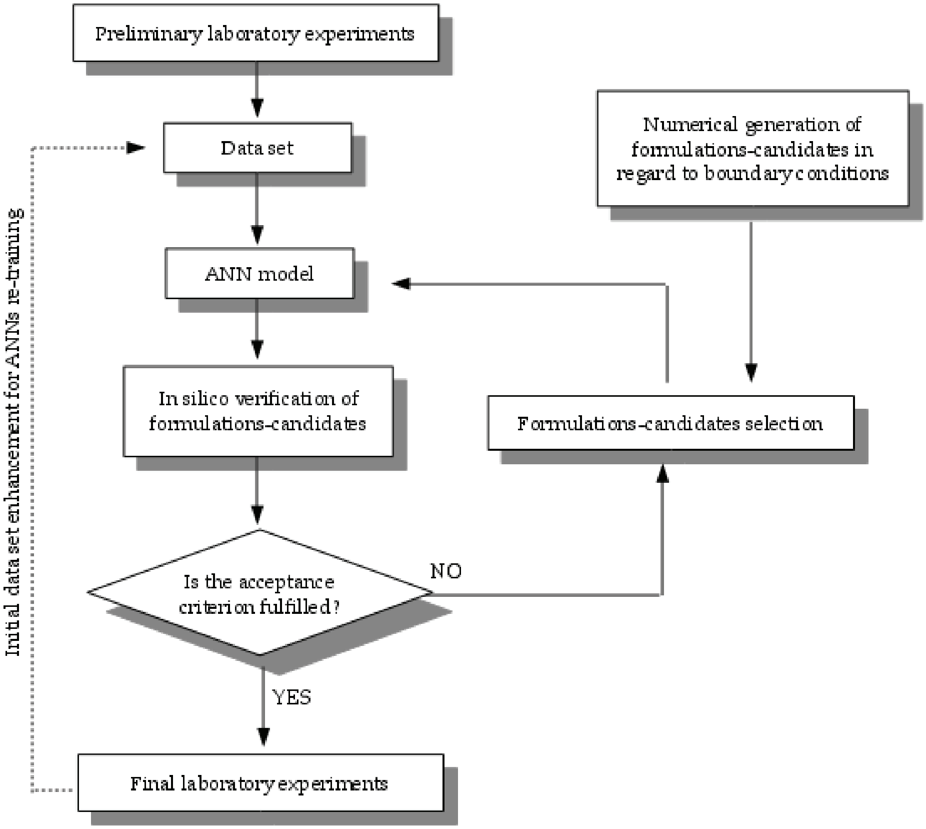

Preliminary laboratory experiments
Data set
ANN model
In silico verification of formulations-candidates
Is the acceptance criterion fulfilled?
NO
YES
Final laboratory experiments
Numerical generation of formulations-candidates in regard to boundary conditions
Formulations-candidates selection
Initial data set enhancement for ANNs re-training

Fig. 3. The algorithm of ANN used as a tool for computed-aided formulation procedure

The use of ANNs in the predictive models function supports the decision based on the "black box" model. This means that no decision explanation and justification is available from the system. Such an approach is acceptable in the DSS, however it could be sometimes unsatisfactory for the user. Therefore, ANNs could be also used in the data mining function in order to provide an insight into the data and some means to formulate hypotheses about the analyzed problem.

ANNs unique features allow them to perform following operations in the data mining approach:

a. select crucial variables for the problem
b. extract logical rules (neuro-fuzzy systems)
c. provide response surfaces for a single input variable or their set

The latter is especially interesting as it allows to switch from "black box" modeling to classical statistical analysis when the problem dimensions reduction was carried out to the sufficient level (i.e. less than 10 input variables). Therefore, it could be created an ordinary mathematical equation quantifying analyzed relationship. Selection of the crucial variables and logical rules extraction form neuro-fuzzy systems are another ANNs powerful features, which would be described further in this chapter. At this moment it is worthy to present only an interesting feature of ANNs employed as data mining tools. In order to obtain the most reliable results it is necessary to find the most competent ANN model. Since ANNs are empirical "black box" models, it is natural that their competence is assessed as the ability to solve unknown cases. This is nothing else but generalization error assessment, which is performed by predictive modeling. Based on the above statements it could be concluded, that data mining procedures include predictive modeling as well. This could be demonstrated by the analysis of the crucial variables iterative procedure with use of ANNs (Fig. 4).

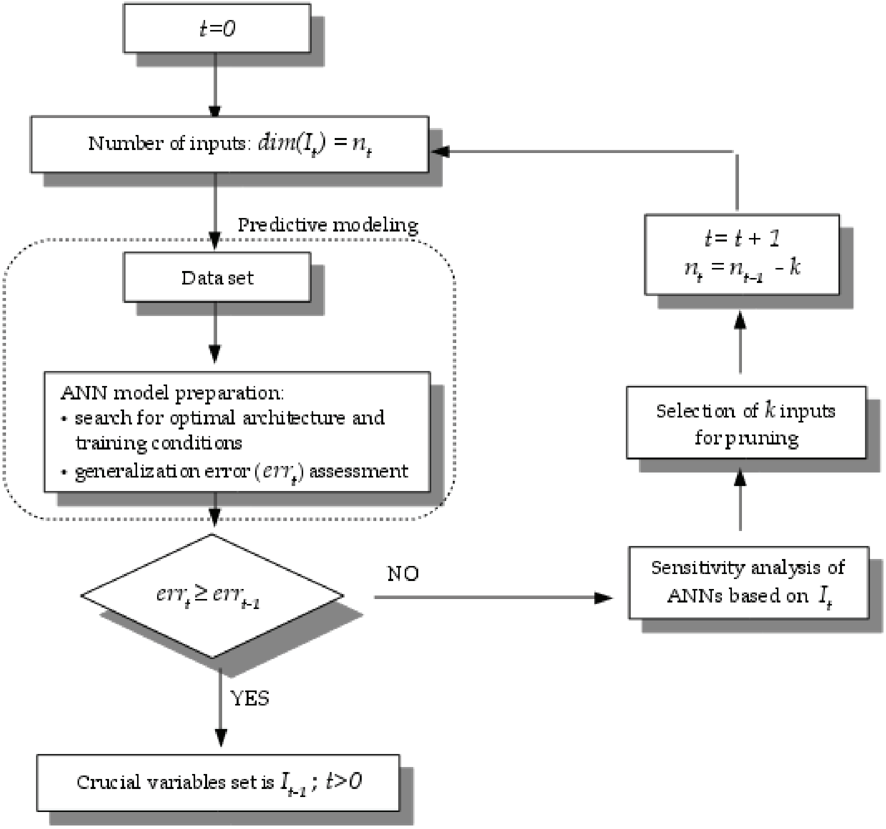

t=0
Number of inputs: $dim(I_t) = n_t$
Predictive modeling
Data set
ANN model preparation:
• search for optimal architecture and training conditions
• generalization error (err_t) assessment
$err_t \geq err_{t-1}$
NO
YES
Sensitivity analysis of ANNs based on I_t
Selection of k inputs for pruning
$t = t + 1$
$n_t = n_{t-1} - k$
Crucial variables set is I_{t-1} ; $t>0$

Fig. 4. The algorithm of inputs reduction with use of sensitivity analysis. *t* – time step; *I* – inputs vector; *n* – number of inputs; *k* – number of inputs for pruning; *err* – generalization error;

The algorithm presented in Fig. 4 allows the smallest number of input variables estimation with regard to the ANN model predictive competence. In other words, the final model is the most general of the best predictive models. This allows to decide, which variables are absolutely necessary to provide competent model, and which could be excluded without performance loss. This results in the very valuable information about the character of the analyzed problem and in the same time an inference machine for DSS is provided.

4. Predictive modeling

Predictive modeling is focused on the generalization abilities of the system, which is usually commonly understood as the extrapolation beyond available database. It is the most difficult task to be performed during the DSS construction.

4.1 Data preparation and preprocessing

Since ANNs are numerical analysis tools they require numerical representation of the whole data available for the problem. This statement is not as trivial as it seems, when the real life data, i.e. pharmaceutical technology, are at the focus. It's challenging to develop numerical representation of pharmaceutical formulation qualitative composition or its preparation technology. So far there is no universal solution of this problem, therefore several methods are used to deal with this task. Among them two main groups of numerical representations could be named:

a. topological
b. physical

In the topological representation input vector is usually binary and the presence of particular formulation compound is denoted by position of its non-zero element. The same could be adapted for formulation technology or other abstract information. The advantage of this approach is its simplicity. One of the disadvantages is a large number of inputs causing problems with high dimensionality of created model. Even if ANNs are working relatively well with multidimensional problems, it should be avoided if possible. More serious drawback of topological encoding is its lack of physical meaning as it is used as completely abstract and subjective design (Fig. 5). Therefore, it could be possible that by use of different encoding scheme (i.e. shifted arbitrary positions of particular components), there would be achieved different modeling results.

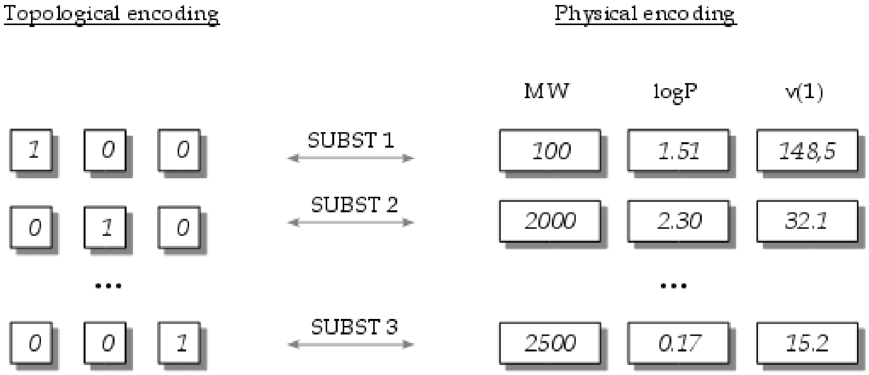


Fig. 5. A comparison between topological and physical representation of pharmaceutical formulations. SUBST – chemical substance, MW – molecular weight, logP – water/oil partition coefficient, v(1) – connectivity index.

The most important disadvantage is that ANN model is restricted only to the established set of substances available at the beginning of the modeling procedure, therefore it has no generalization abilities in terms of qualitative composition. Of course it could be possible to add some additional "dummy" inputs for unknown substances, however regarding previous remarks about arbitrary design of inputs topology without physical meaning, it could be achieved only prediction for some "unknown" substance but not for a specified, particular structure. This is the main reason why topological encoding is treated as the last resort. In contrast, physical encoding has no such drawbacks. It is based on available characteristics of particular excipient (i.e. molecular weight, melting point) or technological process (i.e. compression force). It looks straightforward and perfect approach. Unfortunately, there is one but major drawback of physical encoding – availability of ready-to-use information. Various manufacturers provide different sets of features of their products. Moreover, various substances cannot be characterized in the same manner due to their native character as i.e. being in solid or liquid state. Unification of substances description is required when ANN model has to be built on all available examples. The more data examples, the more competent is the model, thus it is advisable to include every information describing analyzed problem. This is however contradictory with above described problems with unified knowledge representation of the chemical substances. An effective solution could be application of chemical informatics tools, which generally are computer programs able to compute chemical substances properties (so-called molecular descriptors) based on their molecular structure. Chemical informatics has long history and many different applications (Agrafiotis et al., 2007). It is beyond the scope of this chapter to provide complete description of this vast discipline. In pharmaceutical applications, cheminformatics is mostly known at the very early stage of active pharmaceutical ingredient (API) search regarding its desired pharmacological activity. QSAR methods are now routinely applied as tools reducing laboratory experiments number in order to find new promising API, which could become valuable drug in the future. Prediction of toxicological properties of drugs is also at the scope. Cheminformatics is not so popular yet in pharmaceutical technology, however currently it is drawing more attention due to its advantages:

a. unified description of all substances
b. vast number of molecular descriptors counted in thousands
c. prediction of real physical properties (i.e. logP, logD, pKa, etc)

There are disadvantages of cheminformatics use as well:

a. requirements of high computational power for *ab initio* modeling
b. accuracy of physical parameters prediction
c. restrictions of maximum atoms numbers in the analyzed molecule

Unified numerical description of substances is the result of algorithms, on which cheminformatics software is based, thus all molecules are processed in the same reproducible manner. This is crucial for maintaining methodology of ANN model preparation. The large number of molecular descriptors available allows to choose the most representative ones for analyzed problem, which is the most important in data mining procedures, but improves predictability of the model as well. Moreover, in predictive modeling molecular descriptors could be treated as a numerical representation of the molecule without the need of complete understanding of their physical meaning. In fact many of the molecular descriptors are nothing else like numerical representation of 2D (sometimes 3D) structure of analyzed molecule with regard to number of atoms, its geometry, topology and other constitutional features involved. Since the procedure of computations is algorithmic, it allows to use molecular descriptors empirically, based on the

ANN selection of what is the most suitable to achieve maximum predictability of the model. Combining this approach with large number of molecular descriptors available, results in the powerful tool for creating numerical representation of pharmaceutical formulations. Specifically, in predictive modeling the accuracy of physical parameters prediction by cheminformatics software is not an issue as long as ANN model is used as a "black box" in the DSS and the same software is used to encode all substances in the database. The cheminformatics software will be commented in the next section of this chapter.

Overcoming all the problems with pharmaceutical formulation encoding results in the database or so-called "knowledge base" – a source of knowledge for ANN model. In order to be used effectively, the database must be preprocessed. First and obligatory preprocessing procedure is scaling according to the ANNs activation functions domains. Usually the scaling is performed in range (-1;1) but other ranges are also applied, like i.e. (0;1). The latter is sometimes realized as normalization procedure, however more frequently linear scaling is carried out.

4.2 ANNs training

ANNs need to be trained on the data in order to create competent model. Training of ANNs is a serious task and it is impossible to cover all aspects of this issue in this chapter. Following there will be described only the issues, which in authors' opininon are the most relevant to the neural modeling for DSS. Generally, training of ANNs requires several issues to be solved:

a. software and hardware environment
b. training algorithm and scheme
c. topology of ANN (architecture)
d. error measure and model accuracy criterion

Since for the software and hardware environment there will be dedicated further section of this chapter, it is only worthy noting in this place that there is plenty of software available either as free of charge or as commercial packages. The next issue is the subject of many research ongoing, as the universal and perfect ANNs training algorithm does not exist. This is confusing especially when the ANNs simulator provides many algorithms of the choice. Regarding applications of ANNs in pharmacy, the most common and robust ANNs training algorithms could be named as follows:

a. backpropagation with modifications
b. conjugated gradient and scaled conjugated gradient
c. Kalman filter and its extensions
d. genetic algorithms and particle swarm optimization

The above chosen algorithms are mostly associated with so-called supervised learning, where the knowledge base consists of known outputs associated with the inputs. This type of learning is the most suitable for building ANN-based DSS in pharmaceutical technology. Authors are using software with backpropagation (BP) learning algorithm including *momentum*, *delta-bar-delta* and *jog-of-weights* modifications. Backpropagation is a very old and therefore well-established algorithm, which is relatively slow-converging comparing to the newest ones, however is very robust and versatile: i.e. it is suitable for neuro-fuzzy systems as well. The above and BP mathematical simplicity makes it a good choice for implementation in DSS preparation with ANNs. BP with momentum modification has two parameters (learning rate and momentum coefficient), which are chosen arbitrary by user. However, *delta-bar-delta* and *extended delta-bar-delta* modifications allow ANN to modify these parameters during the training process – this improves learning dramatically. *Jog-of-weights* technique is a stochastic search of optimal solution, which is carried-out by simple addition of noise to the ANN weights values when no more training improvement is found during previously set number of iterations. Setting the architecture of ANN is another difficult task, which affects the model performance. Unfortunately, there is no algorithmic solution here. It is usually realized by trial and error experiments carried-out with large number of architectures-candidates in order to select the best one for particular problem. Some improvement is promised by use of hybrid ANNs systems with genetic algorithms (GA). In this evolutionary approach GA is responsible for ANNs architecture adjustment and ANN itself is trained by BP. However, there are still contradictory opinions about suitability of such hybrid systems. In order to decide, which architecture is the most suitable for becoming the core of DSS, it is necessary to apply some quality criterion. Predictive performance is in this case the most applicable criterion expressed as generalization error. The most commonly known method to measure ANNs generalization is k-fold *cross-validation*, where "k" is integer number in the range (0; ∞). The procedure is designed to assess generalization error on the whole available data set. The latter is divided into the two non-equal data-sets: the larger one as training data set and the smaller one as validation (test) set. The ANN is trained on the larger data set and after the training phase the validation set is presented – the error encountered on this set is the generalization error. After that, the validation set is returned to the training set and the new pair of training-validation sets is created, however no previously chosen validation data is included in the new validation set. Again, the ANN is trained on the training set and validated on the smaller one. This algorithm is repeated with respect to the "k" value. The most common "k" value is 10 and each time 10% of original database is excluded from the database to become validation set. After 10 iterations for each architecture the generalization error is assessed for the whole original database (10 x 10% = 100%). Although computationally expensive, this procedure is a standard when the database is small, which is almost an omnipresent situation in real-life examples. A modification of this procedure is *leave-one-out*, where "k" value is equal to the data records number, thus in the validation set there is always only one data record. This is even more computationally expensive, yet from the statistical point of view it provides the most unbiased estimation of ANNs generalization abilities. There are several error measures applicable to express the generalization error of ANNs. Among them, dependig on the analyzed problem type, the most commonly applied are:

a. linear correlation coefficient (R) of predicted vs. observed values
b. *mean squared error* (MSE) or *root mean squared error* (RMSE)
c. classification rate or other classification measures (specificity, sensitivity, etc.)
d. problem-specific measures, i.e.: *similarity factor* (f2) for drug dissolution tests (FDA, 2000)

Each of the error measures allows generalization error quantification, yet it is not absolute – there is no modeling success criteria available. This means that no error measure allows to prove mathematically, that on its specific level the model is competent and reliable. This situation is not only the domain of ANNs. There are present some rules of thumb that beyond some borderline value the model is acceptable. An example of such rule is correlation coefficient where the value over 0.95 is usually acceptable as the indication of good linear correlation between variables, however some authors are more restrictive and demand the value to be over 0.99. Therefore, every generalization error estimation should be regarded with care and related to the problem analyzed.

After the search phase of ANNs best architecture there is provided the ranking of ANNs generalization abilities. The best architecture of ANN is chosen as the final DSS inference machine. However, to improve performance of the model there are built so-called ensemble ANNs consisting of several neural models, which outputs are combined to provide final system output (Maqsood, 2004). The outputs combination is the key factor of ensemble performance. There are many methods for outputs combination, namely:

a. simple average
b. weighted average
c. non-linear regression
d. ANN of second order

The latter method with second order ANN is used very rarely due to the computational burden, yet seems very interesting as the method of non-linear estimation of each ensemble element influence on the final output of the system.

4.3 Modeling example

Preparation of ANN model for DSS in pharmaceutical technology could be illustrated by the example of neural modeling for optimization of so-called solid dispersions systems. Solid dispersions are usually defined as systems consisting of a poorly soluble drug and at least one carrier characterized by good water solubility. The purpose to formulate solid dispersions is to increase water solubility of poorly soluble drugs and in consequence to improve drugs pharmaceutical and biological availability. Unfortunately, there is no clear theory how to adjust quantitative and qualitative compositions of solid dispersions in order to achieve drug solubility enhancement. This could be the domain to DSS – to help in the right choice of the carrier and drug/carrier ratio in order to improve particular drug solubility in water. The neural model was constructed to predict dissolution profile of various drugs, in regard to the solid dispersion (SD) quantitative and qualitative composition as well as SD preparation technology. There were 17 inputs and one output of ANN. The inputs encoded following parameters in physical encoding system:

a. SDs' compositions
b. dissolution test conditions

There was also abstract classification of the methods of SDs preparation added to the input vector as well as the single input expressing the time-point after which the amount of dissolved drug was to be predicted by ANN and presented at the single output. The number of data records was around 3000. Totally, there were around 6 000 ANNs trained and tested in this experiment. The best ANN architecture derived generalization error RMSE = 14,2 vs. maximum output value 100. It was complex ANN with 4 hidden layers and hyperbolic tangent activation function. By introduction of ANNs ensemble with 10 ANNs included and simple average of their outputs, it was possible to achieve generalization error RMSE = 13.4.

The whole neural system was tested as DSS on the following possible scenario: what would be optimal ratio of papaverine (spasmolytic drug) and Macrogol (water-soluble polymer) in SD in order to achieve designated papaverine dissolution profile? This is a typical task to solve in pharmaceutical technology, where the formulation is a tool for modification of the drug course of action. The data were derived from publications, therefore the papaverine's dissolution profiles from various SDs were known and presented to DSS as a task to solve. The above mentioned data was of course unknown to ANNs, which means that the data was not included in the training data set. The system was working according to the

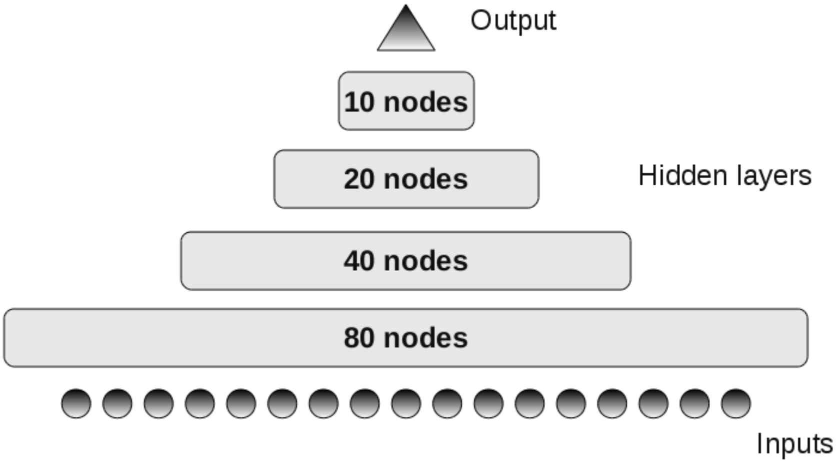


Fig. 6. Best ANN architecture for prediction of drugs dissolution from SDs.

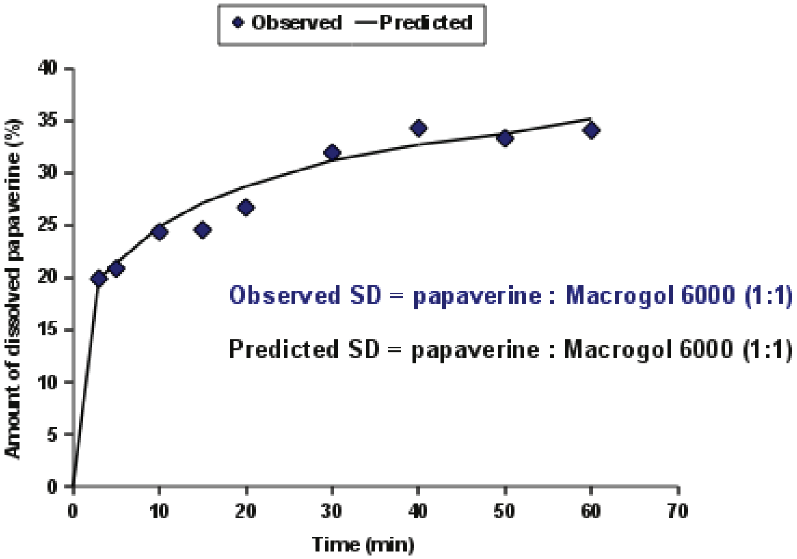


Fig. 7. Appropriate prediction of SD papaverine : Macrogol 6000 1:1 ratio. Prediction error RMSE = 1.3.

algorithm described previously (Fig. 3) wit boundaries selected for qualitative and quantitative composition. Iterative procedure based on the presentation of around 2 000 formulation-candidates with papaverine dissolution profiles as the acceptance criterion. There were 8 profiles presented to the system. As a result in 6 cases qualitative and quantitative compositions of SDs were predicted by the system accurately (Fig. 7). This meant that DSS recommended the same SD composition to achieve particular drug dissolution profile, which was in fact a true source of this profile described in the publication. In conclusion, it was confirmed that DSS based on the ANN could be competent and useful in assisting in the pharmaceutical formulation optimization according to the specified criteria.

5. Data mining

Data mining is a process of knowledge extraction from the database usually associated with discovery of hidden patterns in the data (Wikipedia, 2009b). Empirical modeling with ANNs is one of the standard tools applied in the data mining.

5.1 Sensitivity analysis

Sensitivity analysis is regarded as one of the data mining tools. As a result of this procedure the ranking of relative importance of inputs over the output is provided. It allows to select crucial variables set (Fig. 4). Detailed review of crucial variables characteristics leads to the deeper insight into the analyzed problem. The ranking created by ANNs is the result of observation of data made by machine learning system of empirical modeling. It is quite common, that machine observes data in a different manner than human, and thus the results of such observations are also different. That is exactly what is expected from ANNs at this moment – the unbiased observation of the data conceiving the results, which might be sometimes even contradictory with so-called "common knowledge". These contradictions, or at least unexpected outcomes, are supposed to direct researchers' reasoning to other paths, which could be successful in preparation of the optimal pharmaceutical formulation, when conventional approach fails.

There are many methods of a sensitivity analysis, but two of them are worth mentioning here, since they are commonly used for ANNs. First method is based on the simple assumption that inputs importance could be measured by ANN prediction error changes when particular input is excluded from ANN. The procedure is usually carried out by setting value of input of interest to "0" and assessment of prediction error on the data test set. The bigger error increase, the more important is the selected input. An advantage of this method is its simplicity and versatility – it could be used to every modeling system, not only ANNs. However, this method has some major drawbacks. The most important is that the outcome depends on the data test set used. This makes the procedure difficult to be reproducible. Another issue is the fact that sometimes the "0" value of the variable denotes some information to the system, therefore it creates confusion when all values of particular variable are set to "0". Last but not least is the fact that this method works on the ANN model in its non-natural state, when one of the inputs is in fact nonfunctional. The error increase is the reflection of how badly ANN was destructed by pruning one input. The criticism here is also augmented by unidimensional type of analysis performed. In contrast, second method is much more complicated mathematically but in the same time more sophisticated. Żurada (Żurada et al., 1997) developed method for pruning redundant features based on the analysis of derivative of outputs over ANN inputs (Eq.1).

$$S_{ki} = \frac{\delta y_k}{\delta x_i} \tag{1}$$

where:
S_{ki} – sensitivity of k-th output over i-th input
y – output
x – input
k/*i* – output/input indexes
The derivatives are computed according to the chain rule through the whole ANN for every training pattern. It results in the matrix, which after additional processing provides ranking of inputs. This procedure is reproducible as it works on the training dataset by default. ANN is not altered in any way – it is processed after the training phase in its natural, the most competent state. There is also one drawback of this method – so far it has been developed for MLP ANNs only.
In order to decide, which inputs to prune there must be applied some criterion of how to find a cut-off point in the inputs ranking. Unfortunately, regardless of the method used for ranking creation, there is no universal method of decision where would be the borderline. Usually, the cut-off point is chosen at the largest difference between sensitivity values of adjacent variables in the ranking – this is the borderline between pruned and remaining variables (Fig. 8).

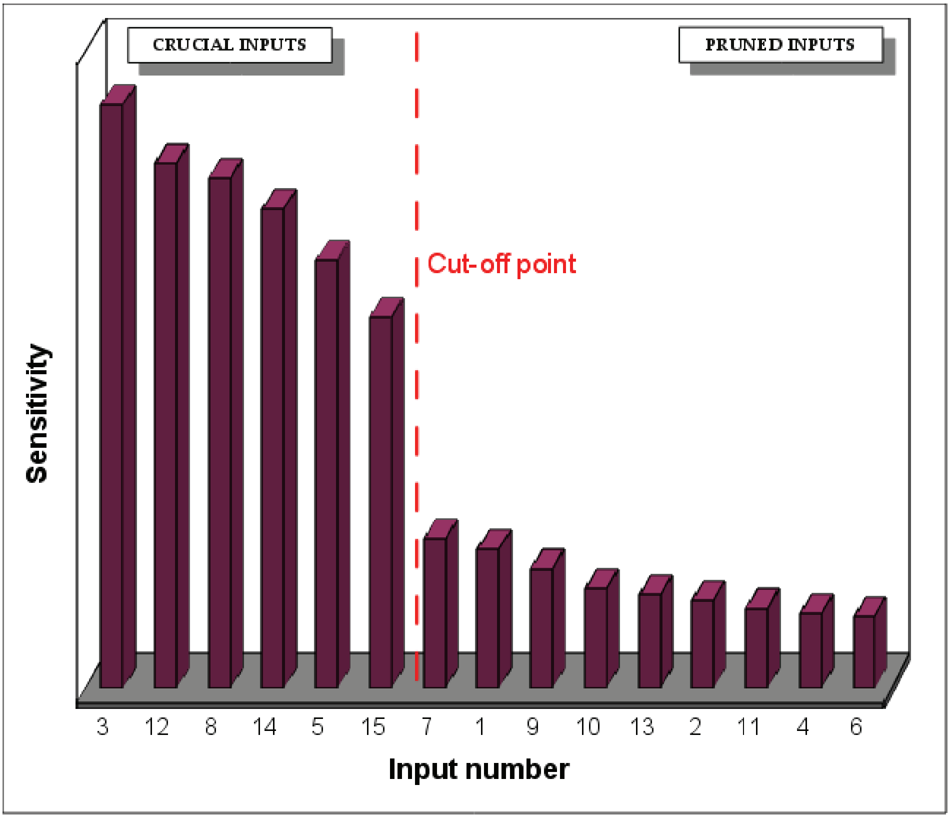


Fig. 8. Sensitivity analysis example with cut-off point selection.

5.2 Fuzzy logic and neuro-fuzzy systems

Fuzzy logic was defined in 1965 when Lotfi Zadeh proposed theory of fuzzy sets. In summary, fuzzy reasoning is based on the probabilistic approach, where every value could be expressed as probability of being a member of some values sets. This is another type of commonly known reasoning based on the classical, crisp numbers. In the simple example a value 0.1 could be a member of set "0" but in the same time be a member of set "1". Probabilities of the memberships to particular sets are designated by so-called membership functions.

Fuzzy reasoning could be encoded in rules tables (Eq. 2).

$$\text{IF } a = A \text{ AND } b = B \text{ AND} \ldots z = Z \text{ THEN } y = Y \tag{2}$$

The above example of simple logical rule could be extended in terms of number of variables and rules as well. Moreover, fuzzy reasoning allows to introduce so-called linguistic variables produced by human experts as non-numerical description of their professional experience expressed in qualitative terms like: "high", "low", "moderate", etc. However, for the improvement of DSS construction it is important to mention hybrid neuro-fuzzy systems: ANNs coupled with fuzzy logic. The neuro-fuzzy system exploits both approaches advantages, namely fuzzy rule-based problem description with self-learning empirical modeling abilities of ANNs. This creates powerful data analysis tool, which is able to observe presented data and to provide self-generated logical rules (Mansa et al. 2008). The latter could be easy decoded to the human-readable form like presented in Eq. 2. In the simplest Mamdani model (Yager & Filev, 1994) neuro-fuzzy system consists of only one hidden layer with specially augmented nodes representing "IF" part of the logical rule. Thus, the number of nodes determines the number of rules – their adjustment might be

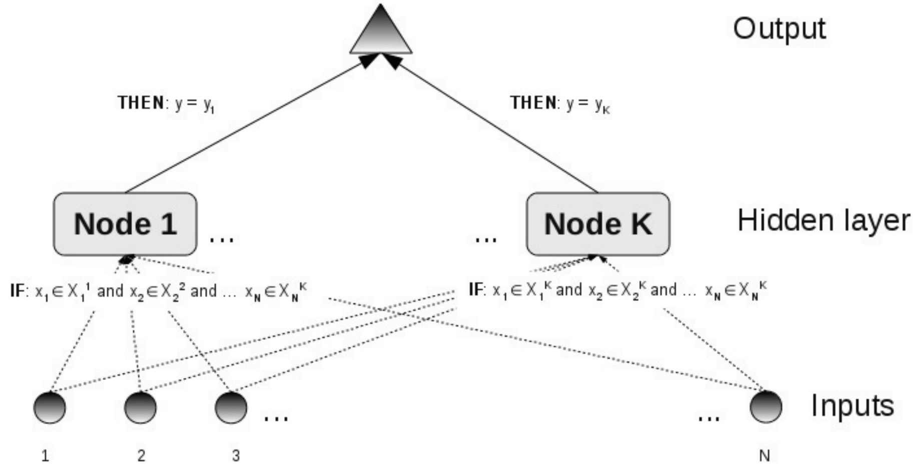

Output
THEN: $y = y_1$
THEN: $y = y_K$
Node 1
...
...
Node K
Hidden layer
IF: $x_1 \in X_1^1$ and $x_2 \in X_2^2$ and ... $x_N \in X_N^K$
IF: $x_1 \in X_1^K$ and $x_2 \in X_2^K$ and ... $x_N \in X_N^K$
...
...
Inputs
1
2
3
N

Fig. 9. A simplified scheme of neuro-fuzzy system of Mamdani multiple input single output (MISO) type; x – input, y – output, N – number of inputs, K – number of hidden units, capital letters – membership functions, small letters – crisp numbers.

made manually or automatically by specific algorithms. The outcome of the rule (THEN) is encoded in the synaptic weight connecting particular hidden node with the output node. The whole system could be trained with classical, well-established BP algorithm.

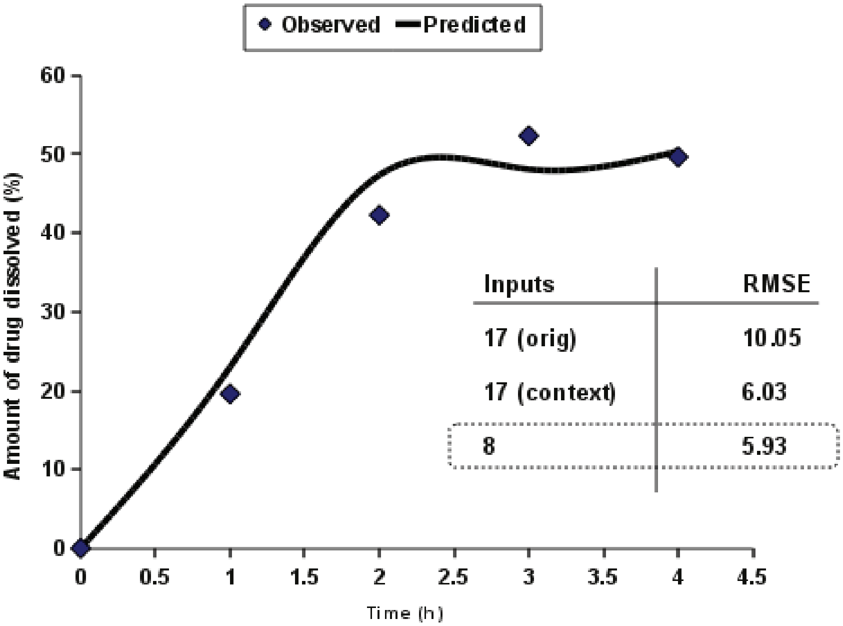
As for every tool, there are also drawbacks of the neuro-fuzzy systems. They are not so versatile like MLP ANNs. This means that not all the problems could be covered by neuro-fuzzy systems, since in fact they are classification-based tools. Their approximating abilities are far below MLP ANNs. In personal experience of authors, neuro-fuzzy systems provide sometimes contradictory or "dummy" logical rules, which from the professional, pharmaceutical point of view are useless and have to be reviewed with utmost care and criticism. In complex problems, like i.e. in pharmaceutical technology, the number of hidden nodes tends to become large, thus making logical rules harder for direct human interpretation. All the above criticism refers to the simplest Mamdani neuro-fuzzy systems. Perhaps the use of Takagi-Sugeno models or more sophisticated architectures optimization algorithms would solve abovementioned problems. This would be the task for the future research. The last, empirical remark about neuro-fuzzy systems would be in favor of their use as members of ensemble ANNs. It was observed several times that when neuro-fuzzy system was added, it improved ensemble performance significantly. This was found even when neuro-fuzzy system was far less competent than several MLPs in the ranking of ANNs generalization abilities. A working hypothesis is that coupling MLP with neuro-fuzzy system allows to exploit both tools different approaches for data analysis. However, for now it is too early research phase to conclude this hypothesis.

5.1 Modeling example

An example of successful sensitivity analysis would be the research about possible mechanisms of drugs release from solid dosage forms. The objective of this study was to identify the mechanisms of model drugs release from hydrodynamically balanced systems (HBS). HBS are prepared in a form of capsule filled with drug substance and mixture of polymers.

Ketoprofen (KT), a poorly soluble non-steroidal anti-inflammatory drug was chosen as a model active substance. Several polymers were used as matrices alone or in binary mixtures: cellulose derivatives (hypromelose), carrageens and alginates. ANNs models were constructed to predict drug release profile from HBS formulations based on their quantitative and qualitative composition. For qualitative composition encoding cheminformatics software was used in order to provide appropriate numerical representation. An initial number of input variables was around 2700. It was the result of cheminformatics encoding of HBS matrices. Data mining methodology was based on the crucial variables set analysis. Search for crucial variables set was performed according to the algorithm depicted in Fig. 4. However, classical sensitivity analysis method was altered due to difficulties with finding significant differences in the ranking of input variables, which made difficult to establish cut-off point. The altered procedure was "context-based" search for the minimum number of variables within original ranking of variables provided by sensitivity analysis. The final choice of variables was performed according to the information about chemical descriptors class, where only one representative of each class was chosen as crucial variable. Numerical experiments with comparison of generalization error between models based on the original and altered variables choice procedure confirmed that application of context based search is beneficial to the model performance

(Fig. 10). In result, it was possible to achieve substantial reduction from 2700 to 8 inputs finally. Final ANNs model confirmed its performance with generalization RMSE = 5.93. The successful generalization examples for unknown formulations were found (Fig. 10). Analysis of 8 inputs meaning allowed to formulate hypothesis about importance of the polymer geometry to the drug release profile.



Inputs	RMSE
17 (orig)	10.05
17 (context)	6.03
8	5.93

Fig. 10. Graph: results of prediction of HBS formulation with carrageen. A comparison between various ANNs with inputs selected by original sensitivity analysis (orig) and altered procedure (context).

6. Software and hardware requirements for DSS with ANNs

6.1 Software

Software environment is crucial for every IT project development. Apart from data processing software like spreadsheets and word processors for documentation preparation, the most important software for DSS preparation with ANNs is ANNs simulator. The term “simulator” is used because there are specialized hardware realizations of ANNs available even as PCI extension cards for PC computers, not mentioning specialized neurocomputers. Hardware ANNs have one advantage over software simulators: they perform parallel computations exploiting this ANNs feature. However, these specialized solutions are very expensive and regarding fast increase of computational power of PC computers, the use of software ANNs simulators seems to be justified. During last 20 years ANNs became so popular that to name all ANNs resources available is impossible for now. Therefore, let us present some examples based on the authors' experience with this type of software. There are several well established commercial packages available:

- NeuralWorks - Professional II/PLUS
- Matlab Neural Networks Toolbox
- statistics software: SPSS, Statistica
- NeuroSolutions

There is also a lot of free software for Windows and Linux/Unix/ MacOS:

- Stuttgart Neural Network Simulator (SNNS)
- GENESIS
- Emergent (former PDP++)
- WEKA

An important issue, when the software choice is to be decided, would be the work mode. If it would be only for data mining, then usually less computational power is required than for the predictive modeling. However, when strictly following previously described algorithm of inputs reduction (Fig. 4) then computational power requirements are high. It was roughly estimated before, that predictive modeling requires usually thousands of ANNs to be trained and tested in order to find the most optimal solution. The task of ANNs training is computationally expensive, therefore it is realized with use of distributed computing on so-called "grids" or "server farms", where several computers are working simultaneously and processing different ANNs. It is the simplest parallelization system, which is in the same time very effective when using ANNs. However, it requires as many licenses of the software as there will be the number of parallel processes running out simultaneously. Regarding the commercial packages, it becomes very expensive to buy separate licenses for each of running processes. Moreover, most of the commercial software is dedicated to MS Windows environment. The simulators are usually standalone packages with point-and-click GUI, without batch mode option. On the contrary, free software is at no cost with as many running instances as it is needed. Many of free packages are built for console mode, thus the batch processing mode is the default option. This is especially characteristic for Open Source software released under various versions of GPL (Gnu Public License). Authors are working with in-house ANN simulator written in Pascal and compiled with use of FreePascal and Lazarus. All computers are working under control of various Linux distributions and there is also developed in-house software for automatic control and distribution of computational tasks. In conclusion, it is worthy to consider Open Source software solutions and Linux environment for ANNs models preparation for DSS, because of a good cost-effectiveness ratio, availability of software and its stability.

Apart from ANNs simulator, cheminformatics software was mentioned as an important element of DSS preparation for pharmaceutical technology. It is a very similar situation in this field like in ANNs – there is plenty of the software available with even more Open Source or Free Software present (Linux4chemistry).

Commercial packages:

- Gaussian
- Gamess-UK
- Sybyl
- Dragon
- Molecular Modeling Pro

Open Source/Free packages:

- Gamess-US
- MarvinBeans (free for academia and non-profit activities)

- RDKit
- ABINIT
- AMMP
- Gromacs
- MOPAC

There is even a special Linux Live CD distribution dedicated to cheminformatics: Vigyaan.

6.2 Hardware

ANNs foundations were noted early 50's of the last century. After some disappointment in their abilities they were forgotten for some time, but 80's was the time of ANNs renaissance. It happened partially because of rapid growth of the computational power of PCs. Internet revolution and development of distributed computing was another factor of increasing interest in the neural modeling. Today, CPUs manufacturers developed new strategy of computational power increase and provide multicore CPUs for desktop computers. It allows for real multi-tasking in the work of modern computers. In order to build the mini-grid, all the infrastructure needed is a set of workstations, some LAN cables and switches. Coupled with Open Source software it provides low-cost, effective tool for ANNs development. There is no means to estimate minimum number of the workstations required. Regarding ANNs, an obvious truth is that the more computers available, the better. A very subjective estimation would be that a good start for the hardware environment is 10 workstations, each one based on 4-core CPU. The system is scalable. An enhancement of such structure with new workstations, even of different type, is very easy and does not generate additional costs beyond hardware price, assuming Open Source software use. In conclusion, building ANNs-based DSS is much easier and cheaper now, when there are present such interesting trends in the PC computers development.

7. References

Agrafiotis D.A. Bandyopadhyay D., Wegner J, & van Vlijmen H., (2007) Recent Advances in Chemoinformatics, *J. Chem. Inf. Model.*, Vol. 47, No 4, pp 1279–1293, ISSN: 1549-9596

Behzadia S.S., Prakasvudhisarnb C., Klockerc J., Wolschannc P. & Viernsteina H., (2009) Comparison between two types of Artificial Neural Networks used for validation of pharmaceutical processes, *Powder Technology*, Vol 195, No 2, 150-157, ISSN: 0032-5910.

Bourquin J., Shmidli H., van Hoogevest P. & Leuenberger H. (1998 a) Comparison of artificial neural networks (ANN) with classical modeling techniques using different experimental designs and data from a galenical study on a solid dosage form, *Eur. J. Pharm.Sci.*, 1998, Vol. 6, No 4, 287-300, ISSN: 0928-0987.

Bourquin J., Shmidli H., van Hoogevest P. & Leuenberger H. (1998 b) Advantages of Artificial Neural Networks (ANNs) as alternative modeling technique for data sets showing non-linear relationship using data from a galenical study on a solid dosage form. *Eur. J. Pharm. Sci.*, Vol. 7, No 1, 5-16, ISSN: 0928-0987.

Bourquin J., Shmidli H., van Hoogevest P. & Leuenberger H. (1998 c) Pitfalls of artificial neural networks (ANN) modeling technique for data sets containing outlier measurements using a study on mixture properties of a direct compressed dosage form. *Eur. J. Pharm.Sci.*, Vol. 7, No 1, 17-28, ISSN: 0928-0987.

Brier M. E. & Aronoff G. R. (1996), Application of artificial neural networks to clinical pharmacology. *Int. Jour. Clin. Pharm. Ther.*, Vol. 34, No 510-514, ISSN: 0174-4879.

Brier M. E. & Smith B. P. (1996), Statistical Approach to Neural Network Model Building for Gentamycin Peak Predictions., *J. Pharm. Sci.*, Vol. 85, No 1, 65-69, ISSN: 0022-3549.

Brier M. E. & Żurada J. M. (1995), Neural Network Predicted Peak and Trough Gentamicin Concentrations. *Pharm. Res.*, Vol. 12, No 3, 406-412, ISSN: 0724-8741.

Chen Y., McCall T.W., Baichwal A.R. & Meyer M.C. (1999), The application of an artificial neural network and pharmacokinetic simulations in the design of controlled-release dosage form., *J Contr. Release*, Vol. 59, No 1, 33-41, ISSN: 0168-3659.

Chow H-H., Tolle K.M., Roe D.J., Elsberry V. & Chen H. (1997), Application of Neural Networks to Population Pharmacokinetic Data Analysis. *J. Pharm. Sci.*, Vol. 86, No 7, 840-845, ISSN: 0022-3549.

Dowell J., Hussain A., Devane J. & Young D. (1999), Artificial Neural Networks Applied to the In Vitro - In Vivo Correlation of an Extended-Release Formulation: Initial Trials and Experience. *J. Pharm. Sci.*, Vol. 88, No 1, 154-160, ISSN: 0022-3549.

FDA (2000), *Guidance for industry. Waiver of In Vivo Bioavailability and Bioequivalence Studies for Immediate-Release Solid Oral Dosage Forms Based on Biopharmaceutics Classification System*, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), USA.

Gašperlin, M., Tušar, L., Tušar, M., Šmid-Korbar, J., Zupan, J. & Kristl, J. (2000) Viscosity prediction of lipophilic semisolid emulsion systems by neural network modeling. *Int. J. Pharm.*, Vol. 196, No 1, 37-50, ISSN: 0378-5173.

Gobburu V.S. & Chen E.P. (1996), Artificial Neural Networks As a Novel Approach to Integrated Pharmacokinetic - Pharmacodynamic Analysis. *J. Pharm. Sci.* Vol. 85, No 5, 505-510, ISSN: 0022-3549.

Hand D., Mannila H. & Smyth P. (2001), Principles of Data Mining, MIT Press, ISBN: 0-262-08290-X, USA.

Hertz J.; Krogh A. & Palmer R., (1991). *Introduction to the Theory of Neural Computation,* Addison-Wesley, ISBN-10: 0201515601, USA.

Hussain A.S., Yu X. & Johnson R.D. (1991) Application of Neural Computing in Pharmaceutical Product Development. *Pharm. Res.*, Vol. 8, No 10, 1248-1252, ISSN: 0724-8741.

Huuskonen J., Salo M. & Taskinen J. (1997), Neural Network Modeling for Estimation of the Aqueous Solubility of Structurally Related Drugs. *J. Pharm. Sci.*, Vol. 86, No 4, 450-454, ISSN: 0022-3549.

Kandimalla K. K., Kanikkannan N. & Singh M. (1999), Optimization of a vehicle mixture for the transdermal delivery of melatonin using artificial neural networks and response surface method. *J. Contr. Release.* Vol. 61, No 1-2, 71-82, ISSN: 0168-3659.

Kolarzyk E, Stepniewski M, Mendyk A, Kitlinski M & Pietrzycka A. (2006), The usefulness of artificial neural networks in the evaluation of pulmonary efficiency and antioxidant capacity of welders. *Int J Hyg Environ Health*, Vol. 209, No 4, 385-392, ISSN: 1438-4639 .

Linux4chemistry, http://www.redbrick.dcu.ie/~noel/linux4chemistry/

Mansa R.F., Bridson R.H., Greenwood R.W., Barker H. & Seville J.P.K., (2008), Using intelligent software to predict the effects of formulation and processing parameters on roller compaction. *Powder Technology*, Vol. 181, No 2, 217-225, ISSN: 0032-5910

Maqsood I., Khan M. R. & Abraham A. (2004), An ensemble of neural networks for weather forecasting. *Neural Comput & Applic,* Vol. 13, No 2, 112–122, ISSN 0941-0643

Mendyk A. & Jachowicz R. (2006), ME_expert - a Neural Decision Support System as a Tool in the Formulation of Microemulsions. *Biocybernetics and Biomedical Engineering,* Vol. 26, No 4, 25-32, ISSN: 0208-5216.

Mendyk A. & Jachowicz R. (2007), Unified methodology of neural analysis in decision support systems built for pharmaceutical technology. *Expert Systems with Applications,* Vol. 32, No 4, 1124–1131, ISSN: 0957-4174.

Mendyk A. & Jachowicz R., (2005) Neural network as a decision support system in the development of pharmaceutical formulation – focus on solid dispersions *Expert Systems With Applications*, Vol. 28, No 2, 285-294, ISSN: 0957-4174..

Polak S. & Mendyk A. (2004) Artificial Intelligence Technology as a Tool for Initial GDM Screening. *Expert Systems with Applications*, Vol. 26, No 4, 455-460, ISSN: 0957-4174.

Polański J. (2003), Self-organizing neural networks for pharmacophore mapping, *Adv. Drug Delivery Rev.*, Vol. 55, No 9, 1149-1162, ISSN: 0169-409X.

Rocksloh K., Rapp F.R., Abed Abu S., Müller W., Reher M., Gauglitz G. & Schmidt P.C. (1999) Optimization of Crushing Strength and Disintegration Time of a High-Dose Plant Extract Tablet by Neural Networks. *Drug Dev Ind Pharm*, Vol. 25, No 9, 1015-1025, ISSN 0363-9045.

Takahara J., Takayama K. & Nagai T. (1997), Multi–objective optimization technique based on an artificial neural network in sustained release formulations. *J. Control. Release,* Vol. 49, No 1, 11-20, ISSN: 0168-3659.

Takayama K., Fujikawa M., Obata Y. & Morishita M. (2003) Neural network based optimization of drug formulations. *Adv. Drug Delivery Rev.* Vol. 55, No 5, 1217-1231, ISSN: 0169-409X

Taskinen J. & Yliruusi J. (2003) Prediction of physicochemical properties based on neural network modelling. *Adv. Drug Delivery Rev.,* Vol. 55, No 5, 1163-1183, ISSN: 0169-409X.

Türkoğlu M., Özarslan R. & Sakr A. (1995) Artificial Neural Network Analysis of a Direct Compression Tabletting Study, *Eur. J .Pharm. Biopharm.,* Vol. 41, No 5, 315-322, ISSN: 0939-6411.

Veng-Pedersen & P. Modi N.B. (1992), Neural Networks in Pharmacodynamic Modeling. Is Current Modeling Practice of Complex Kinetic Systems at a Dead End? *J. Pharm. Biopharm.*, Vol. 20, No 4, 397-412, ISSN: 1567-567X.

Wikipedia (2009 a), http://en.wikipedia.org/wiki/Artificial_neural_network

Wikipedia (2009 b), http://en.wikipedia.org/wiki/Data_mining

Wikipedia (2009 c), http://en.wikipedia.org/wiki/Decision_support_system

Yager, R.R. & Filev, D.P., (1994), *Essentials of fuzzy modeling and control*. John Wiley & Sons, Inc., USA

Żurada J.M. (1992). *Introduction to Artificial Neural Systems,* West Publishing Company, ISBN-10: 053495460X, USA.

Żurada J.M., Malinowski A. & Usui S. (1997) Perturbation Method for Deleting Redundant Inputs of Perceptron Networks. *Neurocomputing*, Vol. 14, No 5, 177-193, ISSN: 0925-2312.

Clinical Decision Support Systems: An Effective Pathway to Reduce Medical Errors and Improve Patient Safety

Chiang S. Jao[1,2] and Daniel B. Hier[1]
[1]University of Illinois at Chicago,
[2]National Library of Medicine
United States

1. Introduction

1.1 Background

Medical errors are both costly and harmful (Hall, 2009). Medical errors cause tens of thousands of deaths in U.S. hospitals each year, more than from highway accidents, breast cancer, and AIDS combined (SoRelle, 2000). A phone survey by the National Patient Safety Foundation found that 42 percent of over 100 million Americans believed that they had personally experienced a medical mistake (Louis & Harris Associates, 2007). The 1999 Institute of Medicine report stated that medical errors were the eighth leading cause of death in the U.S., killing between 44,000 and 98,000 people each year (Kohn et al., 2000). Another study indicated 225,000 deaths annually from medical errors, including 105,000 deaths due to “non-error adverse events of medications” (Starfield, 2000). Medical errors threaten the quality of health care, increased healthcare costs, and add to the medical malpractice crisis (Studdert et al., 2005). According to the Patient Safety in American Hospitals Study Survey by HealthGrades (HealthGrades, 2004; HealthGrades, 2007; HealthGrades, 2008; HealthGrades, 2009), the number of deaths in U.S. hospitals each year that are reportedly due to medical errors has been disturbingly high since 2000:

1. Based on a study of 37 million patient records, an average of 195,000 people in the U.S. died due to potentially preventable, in-hospital medical errors in each of the years from 2000 through 2002.
2. Approximately 1.16 million patient safety incidents occurred in over 40 million hospitalizations for the Medicare population yielding a three-percent incident rate. These incidents were associated with $8.6 billion of excessive costs during 2003 through 2005. Although the average mortality rate in Medicare patients from 2003 through 2005 was approximate 21.35 percent and overall rates have been declining, medical errors may still have contributed to 247,662 deaths.
3. Patient safety incidents cost the federal Medicare program $8.8 billion and resulted in 238,337 potentially preventable deaths from 2004 through 2006.
4. Approximately 211,697 patient safety events and 22,771 Medicare deaths could have been avoided with a savings of $2.0 billion from 2005 through 2007.

These numbers indicate the magnitude of savings in both lives and dollars from improved patient safety.

1.2 Health information technology adoption in the U.S.

Health information technology (HIT) offers an opportunity to transform healthcare and make it safer (Bates & Gawande, 2003; Parente & McCullough, 2009). With the advent of electronic medical records (EMRs) and computerized physician order entry (CPOE), the maintenance of patient information has become easier. The EMR provides the clinician with a longitudinal source of patient information including diagnostic history, previous encounter history, drug allergies, and other relevant information. A computer-assisted decision support system can be designed to help clinicians collect critical information from raw clinical data and medical documents in order to solve problems and to make clinical decisions. A clinical decision support system (CDSS) links health observations with medical knowledge in order to assist clinicians in decision making. The embedding of a CDSS into patient care workflow offers opportunities to reduce medical errors as well as to improve patient safety, to enhance drug selection and dosing, and to improve preventive care. It is less certain whether a CDSS can enhance diagnostic accuracy (Bakken et al., 2008; Bates et al., 1998; Bates et al., 2001; Bates et al., 2003; Hunt et al., 1998; Kaushal et al., 2001a; Kaushal et al., 2001b). A CDSS can assist clinicians in reducing some errors and costs (ActiveHealth Management, 2005; Bates et al., 2001; Bates & Gawande, 2003; Bates et al., 2003; Berner, 2007; Chaudhry, 2008).

2. Significance

2.1 Clinician approach to health information technology

The U.S. national healthcare expenditures are projected to reach $2.6 trillion in 2010 and $4.7 trillion in 2019 (Foster & Heffler, 2009). President Barack Obama has called for wider use of HIT to help control rising healthcare costs. In February of 2009, Congress passed the HITECH Act (Health Information Technology for Economic and Clinical Health) which provided financial incentives to physicians and hospitals to adopt HIT. However, clinician acceptance of HIT remains critical to the success of efforts to use electronic medical records (EMRs) to reduce healthcare costs. Clinicians often view EMRs as costly, awkward, and disruptive of their workflow. Many clinicians remain reluctant to adopt EMRs. A recent survey of 423 physicians by the Massachusetts Medical Society (Chin, 2004) found that while 85% believe that doctors should adopt electronic prescribing, 49% say they do not intend to do so. Further, although 89% believe that doctors should record patient summaries electronically, 48.5% do not intend to do so. Another survey of 500 health care providers found that 52% thought the stimulus package would have little or no success in encouraging HIT adoption in the U.S. (IVANS, 2009).

2.2 Characteristics of clinical decision support systems

A CDSS is a computerized system that uses case-based reasoning to assist clinicians in assessing disease status, in making a diagnosis, in selecting appropriate therapy or in making other clinical decisions. There are three key elements of a successful CDSS (Musen et al., 2001):

1. *Access to accurate clinical data,*
2. *Access to pertinent medical knowledge*
3. *Ability to use appropriate problem solving skills.*

An effective CDSS involves six levels of decision making: alerting, interpreting, critiquing, assisting, diagnosing and managing (Pryor, 1990). Alerts are a vital component of a CDSS.

Automated clinical alerts remain an important part of current error reduction strategies that seek to affect the cost, quality, and safety of health care delivery (Kuperman et al., 2007; Raschke et al., 1998; Smith et al., 2006). The embedded knowledge component in a CDSS combines patient data and generates meaningful interpretations that aid clinical decision making (Liu et al., 2006). An effective CDSS also summarizes the outcomes, appraises and criticizes the caring plans, assists clinicians in ordering necessary medications or diagnostic tests, and initiates a disease management plan after a specific disease is identified (Colombet et al., 2005; Friedlin et al., 2007; Garg et al., 2005; Wadhwa, 2008; Wright et al., 2009).

2.3 The architecture of a clinical decision support system

Several practical factors contribute to the success of a CDSS. These factors include (1) considering the potential impact on clinical workflow, (2) creating an intuitive and configurable user interface, (3) delivering decision support in real time at the point of care, and (4) providing actionable alerts/reminders/recommendations that are succinct and relevant to patient care (Friedlin et al., 2007; Kawamoto et al., 2005). The minimum required technical architecture for a CDSS is identified as (1) a skilled communication engine to access disparate data, (2) a mandatory clinical vocabulary engine to perform semantic interoperability, (3) an optimized patient database to facilitate disease management, (4) a modular knowledge base to mine adequate diagnostic and therapeutic information, and (5) an effective inference engine to expedite decision making by relating embedded knowledge to ongoing problems (Pestotnik, 2005). How to best use a CDSS to influence clinician behavior is still a challenge in the clinical domain to provide high-quality care at lower cost (Bates & Gawande, 2003; Jao et al., 2008b).

The development of an effective CDSS has a significant impact on clinician's practice plans. The introduction of such a system will provide clinicians a useful guideline through which they can replicate their decisions on similar clinical cases. Furthermore, an effective CDSS can reduce the variation of clinician's practice plans that plagues the process of healthcare delivery. The dynamic environment surrounding patient diagnosis complicates its diagnostic process due to numerous variables in play; for example, individual patient circumstances, the location, time and physician's prior experiences. An effective CDSS reduces variation by reducing the impacts of these variables on the quality of patient care.

3. Major issues

3.1 Medical errors

Reducing medical errors requires an environment of continuous disclosure and analysis; an environment which is in conflict with the current medical liability climate (Clinton & Obama, 2006). Five common types of medical errors include (1) prescribing erroneous medications, (2) inappropriately ordering laboratory tests for the wrong patient at the wrong time, (3) filing system errors, (4) dispensing the wrong medications, and (5) failing to promptly respond to abnormal laboratory test results (Dovey et al., 2003). Accessing the EMR is the first step to controlling medical errors (Hillestad et al., 2005; Wang et al., 2003). Studies show improved patient safety from the use of EMR in hospitals and ambulatory care that primarily relies on alerts, reminders, and other components of CPOE in reducing adverse drug events (Bates et al., 1998; Bates et al., 2001). The concept of the Problem-Oriented Medical Record advocated by Weed builds a sound structure for medical decision-making that can lead to error reduction (Bayegan & Tu, 2002; Weed, 1968a; Weed, 1968b; Weed, 1968c).

3.2 The significance of accurate medication and problem lists

The need for a problem list (or diagnosis list) is clear. The problem list and medication list (list of prescribed drugs) provide an essential overview of diagnoses and treatment. The problem list is a critical part of the medical record because it contains the patient's active and resolved medical problems while the medication list contains the prescribed drugs for each diagnostic problem.

Optimal medication and problem lists accurately reflect ordered medications and ongoing problems. The problem list helps physicians check against potential prescribing errors, reminds them of issues often forgotten, and improves communication among health care providers (Simborg et al., 1976; Starfield et al., 1979). An accurate problem list facilitates automated decision support, clinical research, data mining and patient disease management (Hartung et al., 2005; Jao et al., 2004; Johnston et al., 1994; Rothschild et al., 2000). An accurate computerized medication list is a direct outgrowth of computerized physician order entry (CPOE) and e-prescribing, while an inaccurate medication list creates risks and adversely affects quality of health care (Kaboli et al., 2004; Rooney, 2003). Proper management of the medication and problem lists reduces the potential for medication and diagnostic errors.

3.3 Current state of problem list compliance

Since 2006, the maintenance of the diagnosed problem list has been mandated as a patient safety feature by the Joint Commission of Accreditation Health Organization. A computerized problem list in the EMR is more readily accessible than the paper chart, and codified terms in the medication and problem lists create an opportunity to implement clinical decision support features, including knowledge retrieval, error detection, and links to clinical guidelines (Wasserman & Wang, 2003). Nonetheless, accurate maintenance of the problem list and medication list is difficult in practice.

Despite previous research confirming that the problem list is vital to the evidence-based practice of medicine, physician compliance in creating an accurate medication and problem list remains unsatisfactory (Brown et al., 1999; Rowe et al., 2001). A recent case report ascribed the death of a female patient to the failure to maintain her ongoing problem list by her primary care physician (Nelson, 2002). According to another medical report, one in every 10 patients admitted to six Massachusetts community hospitals suffered serious and avoidable medication errors (Wen, 2008). In a review of 110 discharge medication lists in the Augusta Mental Health Institute of Maine, 22% contained errors (Grasso et al., 2002).

3.4 Clinician attitudes toward and knowledge of CDSS

A survey of physician attitudes showed that the perceived threat to professional autonomy was greater for CDSS than for an EMR (Walter & Lopez, 2008). Other results indicate that the degree of clinician acceptance of a CDSS seems to be correlated with their attitudes about their professional role and their attitudes towards the computer's role in disease management and decision-making (Toth-Pal et al., 2008). Other significant barriers to CDSS adoption have been ascribed to insufficient level of computer skills among clinicians and time constraints on clinicians. Studies have shown that lacking a useful CDSS at the point of care hinders informed clinical decision making and coordination of patient care (Kaushal et al., 2003; Sittig et al., 2006).

Clinicians are typically challenged by the complex interplay of multiple disease parameters with surrounding factors (e.g., the disease agent, the environment, the patient's description of self symptoms, the results from laboratory testing, the physician's capability of observation, etc.) that determine how a disease will present itself and how it will be perceived by a clinician. Lack of awareness of relevant scientific evidence and time constraints were the most often cited physician barriers to implementing effective decision-making in clinical practice (Cabana et al., 1999; Edwards & Elwyn, 2004; Graham et al., 2003).

3.5 Assessment of physician compliance on clinical documentation

Surveys and audits of medical records reveal that the diagnosed problem list and prescribed medication list are often inaccurate, out of date, or incomplete. Previous audits of patient charts at the University of Illinois Hospital (UIH) showed that problem list maintenance is haphazard (Galanter et al., 2008; Hier, 2002; Jao et al., 2008a). In many patient charts multiple versions of the problem list coexist; some lists lack critical problems (clinical diagnoses); other lists have many resolved or inactive problems. Similarly, many medical records contain numerous and inconsistent medication lists, which do not reflect the actual medications taken by a specific patient. Medication lists are often obsolete (containing medications no longer prescribed) or incomplete (lacking medications that are prescribed), while multiple reconciled versions of the medication list coexist in the same medical record. Most medical records make no attempt to establish medication-to-problem relationships or ordering by indication.

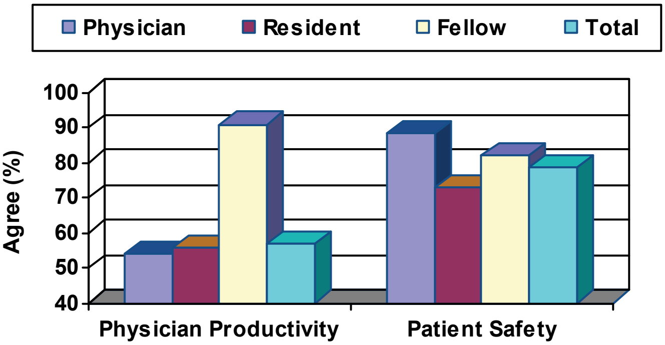


Fig. 1. Survey assessment of physician's knowledge regarding benefit contribution from the improvement of clinical documentation by the CDSS

To assess physician knowledge, attitudes, and practice patterns related to issues in problem list documentation, an online survey was distributed to more than 800 health care practitioners at the UIH. Among the 97 respondents, 30% were attending physicians, 68% were residents, and 12% were fellows (Jao et al., 2008b). The majority of respondents were reluctant to diligently maintain medication and problem lists, indicating a continuing gap in quality of documentation. According to the results of this survey, approximately 50 percent of surveyed providers said that (1) the problem lists were not well maintained in their own clinical units, (2) approximately 55 percent said that they audit and maintain the medication and problem list on their own behalf, and (3) approximately 42 percent said that they failed to update the centralized medication and problem lists after including a problem list in each of their own progress notes. Respondents felt that the CDSS could improve problem list documentation and would benefit patient safety more than physician productivity (as shown in Fig. 1).

4. Current trends

4.1 Barriers of CDSS implementation

Human knowledge and inspection is used to detect and correct errors in medical records. However, Bates et al. ascribed weak error-reduction strategies to the use of human knowledge and inspection in medical error discovery (Bates et al., 2001). The World Health Organization mandates reducing medical errors, providing high-quality disease-centred evidence/information, and lowering cost in health care by full adoption of e-Health strategies through full development of HIT, especially adopting a patient-centred EMR (WHO, 2005). A recent study suggested that the aggressive integration of clinical evidence from health care research into diagnostic decisions could influence patient outcomes by improving clinical diagnosis, reducing unnecessary testing, and minimizing diagnostic errors. However, significant barriers must be overcome to achieve this goal (Garg et al., 2005; Richardson, 2007). There are several potential impacts to clinical practice due to these common barriers (see Table 1).

4.2 Embedding CDSS implementation within CPOE and EMR

Recent studies indicate that an evidence-based CDSS works best when it is embedded within a CPOE system (Gross & Bates, 2007; Trivedi et al., 2009; Wolfstadt et al., 2008). It is critical to design a useful CDSS so that it improves a clinician's workflow, it provides satisfactory system performance, and results in acceptable system reliability. Moreover, organizational factors, such as the leadership support, strong clinician champions and financial support, play a role in the success of CDSS implementation.

A useable CDSS typically requires multifaceted domain knowledge that is expressed as inference rules in a computable, explicit and unambiguous form (Kuperman et al., 2006). Characteristics of individual patients are matched to a computerized knowledge base, and software algorithms in the CDSS generate patient-specific recommendations that are delivered to clinician-users of the EMR. (Garg et al., 2005). A recent study has identified three key elements for fully realizing the potential of a CDSS (Osheroff et al., 2007):

1. The best available clinical knowledge is well organized, accessible to clinicians, and encapsulated in a format that facilitates effective support for the decision making process
2. A useful CDSS is extensively adopted, and generates significant clinical value that contributes financial and operational benefits to its stakeholders.
3. Both clinical interventions and knowledge undergo constant improvement through user feedback, experience, and data analysis that are easy to aggregate, assess, and apply.

Categorized Barriers	Potential Impacts to clinical practice
Evidence-Related	
• Lack of supportive research evidence	• Decision may not be able to draw an acceptable conclusion or judgment
• Incomplete or contradictory evidence	• Decision may be infeasible to the clinical case
• Inaccessible evidence at the point of care	• Evidence could be not be reached to assist practitioners in decision making
Clinician-Related	
• Lack of in-depth knowledge in the specific nature of evidence	• Could not make full use of evidence to the specific type of a diagnostic problem
• Failure to use the CDSS or non-acceptance of computerized recommendations	• Could not efficiently manipulate evidence or adapt recommendations to accommodate the variance of diagnoses
• Obedience to others' diagnostic decision	• Will not employ independent analytic thought and reasoning on evidence
System-Related	
• Multiple requirements (e.g., billing and EMR) converge to stress clinicians for coding patient's disease with accurate diagnoses	• Throughput-oriented concerns may discourage the deliberate processes of analytic diagnostic thinking
• External incentives (e.g., reimbursement, patient satisfaction, quality demerits, malpractice) through the use of research evidence	• Desire for rewards or fear of punishments may influence diagnostic strategies more strongly than analytic thought using research evidence
• Poor usability or integration into practitioner's workflow	• Good system performance depends on the motivational effect of the developer's enthusiasm, creation of more usable and integrated software, better access to technical support and training, and improved on-site promotion and tailoring

Table 1. Common barriers to integrate research evidence into clinical practice

4.3 CDSS and patient safety

The quality and safety of health care leaves much to be desired (Leape & Berwick, 2005; McGlynn et al., 2003). Enhanced patient safety encompasses three complementary activities: preventing errors, making errors visible, and mitigating the effects of errors. Improvement and automation in a CDSS can assist clinicians making errors visible and augmenting error prevention. A CDSS provides several modes of decision support, including alerts, reminders, advice, critiques, and suggestions for improved care. In this way, CDSSs are able to decrease error rates by influencing physician behaviour, improving clinical therapy, and improving patient outcome (survival rate, length of patient stay, and cost). Computerized alerts can also allow rapid data collection from a large number of practices over a wide population (Johnson et al., 1991).

4.4 A CDSS example

4.4.1 The goal

To assist physicians in maintaining the accuracy and completeness of the problem and medications lists within the EMR, the Problem List Expert (PLE©) was developed at the University of Illinois Hospital (UIH) (Jao et al., 2008). This system was designed to test the hypothesis that a CDSS can assist in effectively identifying and maintaining problem-medication matches in the EMR. When medication and problem list mismatches were detected by the CDSS, expert clinicians examined the EMR to identify the nature of mismatches and causes for the mismatches including missing problems, inactive or resolved problems, missing medications, or duplicate prescribing.

4.4.2 The core of CDSS

The core of the PLE© is three linked database tables: the medication data dictionary, the problem data dictionary, and medication-problem relationship table. There were approximately 1,250 medication items in the UIH drug formulary added to the medication data dictionary. There were over 15,000 problem items (derived primarily from ICD-9-CM) added to the problem data dictionary. The database model is constructed as a network in which medications and the problems are associated by many-to-many relationships. Fig. 2 illustrates the structural model of the knowledge base. To simplify data query, each item in the medication data dictionary and each item in the problem data dictionary are connected by a common key attribute, an indication. In medicine, an indication is defined by the National Cancer Institute as "a sign, symptom, or medical condition that leads to the recommendation of a clinical treatment, a laboratory test, or a treating procedure" (http://cancernet.nci.nih.gov/Templates/db_alpha.aspx?CdrID=348991). Each medication can be linked with its associated indications that can be represented as a group of relevant clinical problems. Fig. 3 represents the hierarchical network model of the working database structure. Each normalized problem item in the problem data dictionary can be mapped to a unique ICD-9-CM (the International Classification of Diseases, Ninth Revision, Clinical

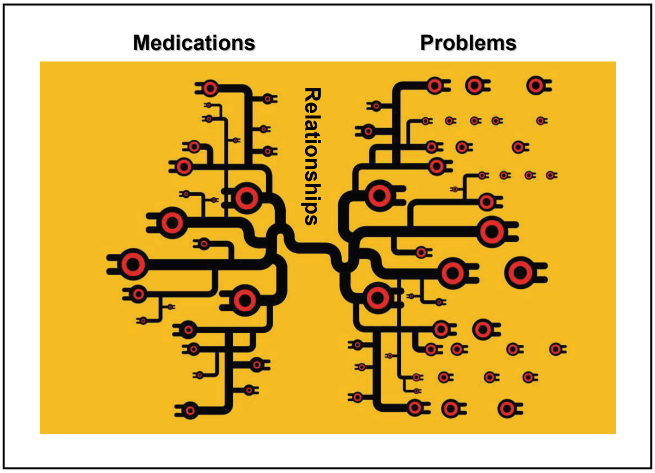


Fig. 2. The complex relationships are to connect prescribed medications to ongoing problems in the EMR.

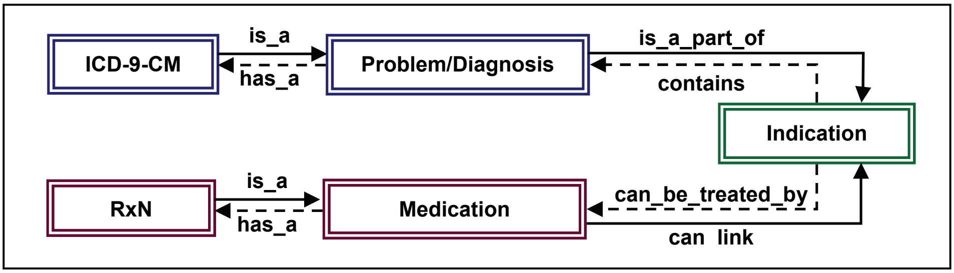


Fig. 3. Link medication orders to problems and diagnoses through the associated indications in a network model. All relationships in a hierarchical database are either one-to-one (1:1) or one-to-many (1:N). For example, each diagnostic problem item **has a** unique ICD-9-CM code when each ICD-9-CM code **is a** diagnostic problem. It is a one-to-one relationship between each problem and its associated ICD-9-CM code, and between each medication and its associated RxN (drug number). It is a one-to-many relationship between each indication and its related problems and medications.

Modification) code as defined by the Center for Disease Control and Prevention (CDC) (http://www.cdc.gov/nchs/about/otheract/icd9/abticd9.htm). Each normalized medication item in the medication data dictionary can be mapped to a unique self-defined drug number. Therefore, each ordered medication can be easily mapped by computer algorithm to one or more clinical problem(s) using established medication prescribing standards. This mapping methodology facilitates knowledge management and expedites clinical decision support.

4.4.3 Methodology of decision support

The PLE© was designed to simulate both a CPOE for ordering medication and an EMR for recording medication and problem lists. The PLE© assisted clinician experts in reviewing 140 patient records in three clinical units (general internal medicine, neurology, and rehabilitation) and discovering medication-problem mismatches (instances in which a medication was prescribed but had no indication on the problem list). Natural language processing assists in screening and matching the medications to problems. The matching algorithm in PLE© examines each medication on the Medication List by linking its indications to the indications for those problems on the Audited Problem List through the defined association in the Medication-Problem Relationship Table of the PLE©. A machine-learning algorithm is employed to correctly distinguish and classify the medications and problems entered in the CPOE. A data-mining algorithm is employed to discover the pattern and the relationship between the prescribed medications and the ongoing problems in the EMR. The data-mining algorithm facilitates the medication-problem matching and database management within a large set of data. Several common types of medication list errors (for example, unnecessary medications, inadvertently added medications, and missing medications) and problem list errors (for example, failure to remove inactive or resolved problems and failure to add active problems) may risk patient safety and can be fixed by physicians during chart audits.

Other key components of the PLE© are a patient data repository and a user interface. Through the enhanced user interface, physicians are able to create new patient records, create problem lists, and order medications. When a new medication is ordered through the CPOE, the PLE© assists in checking if an appropriate problem is on the active problem list that is an indication for the medication ordered. Fig. 4 shows the infrastructure and workflow of the PLE© implementation, where the problem list obtained from UIH's EMR is termed the Reported Problem List; the medication list obtained from UIH's EMR is termed the Medication List, the list for medication-problem relationships based upon clinician expert review is termed the Audited Problem List. The order of data entry was the patient's Reported Problem List, Audited Problem List, and Medication List, which were saved in the Patient Data Repository without patient identities. The PLE© first examined the existence of entered items in the Medication Data Dictionary and the Problem Data Dictionary. The PLE© adopted computer algorithms for knowledge updating and discovery. New data will be automatically added in the corresponding data dictionaries accordingly.

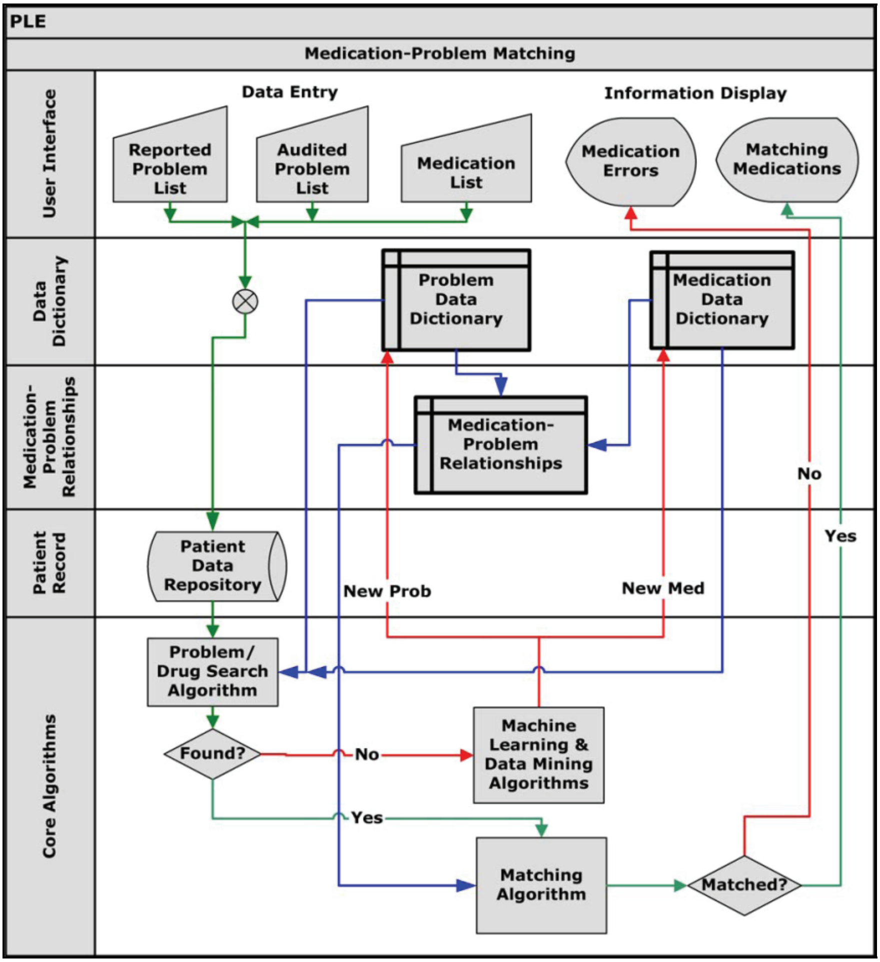


Fig. 4. The infrastructure and workflow of the PLE© implementation.

4.4.4 Results

The PLE© automates the maintenance of the medication and problem lists and detects likely medication-problem mismatches as visible medication and diagnostic errors on the screen of the EMR. With regard to the problem list, The PLE© found that approximately 11% of patient records had no problems listed on the Reported Problem List. Approximately 11% of patient records were perfectly matched (i.e., the count on the Reported Problem List equalled the count on the Audited Problem List). The remaining 78% of patient records showed various levels of problem deficiency on the Reported Problem Lists (i.e. the audit showed that problems were missing from the Reported Problem List).

The PLE© was programmed so that is was able to suggest the addition of non-specific problems that corresponded to common medications orders for treating problems which are generally unlisted on the problem list: for example, the medication "bisacodyl" for treating the problem "constipation," the medication "famotidine" for treating the problem "gastric acid," and the medications "acetaminophen" and "ibuprofen" for treating the problem "pain," etc. Most of these common medications are related to nursing diagnoses that are commonly not added to the problem list by physicians (e.g. fever, pain, constipation, etc.) This feature in the PLE© (matching common medications to minor non-recurrent problems) reduces the likelihood of finding medication-problem mismatches. The improvement rate of medication-problem matches on the problem lists was equal to the variance of the percentages of matched medications on the Medication List in the individual inpatient unit before and after expert chart review.

One approach to improve poor physician compliance with maintenance of the problem list is to link the ordering of medications to the problem lists by using a CDSS to automate the process of maintaining the EMR. In other words, when a medication is either ordered by CPOE or ePrescribing, the CDSS automates the process of adding the appropriate problem (the indication for the medication) to the problem list. The PLE©, an innovative CDSS, automates the maintenance of both medication and problem lists in the EMR. It exploits advanced decision support strategies to yield higher patient safety by improving the accuracy of the medication and problem lists. It effectively identifies potential medical errors to some degree and improves problem list documentation in the EMR.

5. Future challenges

The potential to develop more sophisticated computerized alerts and other types of CDSS will grow as more clinical data becomes accessible electronically. Automated computerized-based applications utilize the accurate and structured clinical information available in the EMR to improve patient care and lower costs. Preliminary studies have shown that the CDSS is an essential cornerstone of efforts to reduce medical errors and improve patient safety. Future challenges to implementing a CDSS that automates the maintenance of the medication and problem lists include: (1) it may not work at an acceptance level of accuracy to make it clinical useful; (2) it may be too cumbersome to use so that clinicians are resistant to using it; and (3) the decision support algorithms may fail to work in some specific cases because of the complexity of medical decision-making.

CDSSs can assist in preventing adverse drug reactions, reducing inappropriate drug dosing, and reinforcing the use of effective prophylactic measures, (Trowbridge & Weingarten, 2001). Sittig et al. listed ten grand challenges in clinical decision support, including improving the user interface to facilitate data entry and clinical workflow; disseminating best practice evidences in the CDSS design, development, and implementation; summarizing precise patient-level information in the real time performance; prioritizing and filtering useful recommendations to the clinician for decision making; creating a reusable system architecture for sharing executable CDSS modules and services among different health care providers; combining feasible recommendations for patients with comorbidities; prioritizing CDSS content development and implementation; creating an internet-accessible CDSS and data repositories for widespread adoption; using free text information to drive decision support in the clinical domain; and mining large set of accurate clinical data to create an innovative CDSS (Sittig et al., 2008).

An electronic ordering (e-Ordering) of diagnostic imaging services has been proposed by the newly formed Imaging e-Ordering Coalition (The Coalition, Washington). This e-Ordering system will be supported by a CDSS that will guide clinicians to order the most appropriate diagnostic tests. The e-Ordering system will electronically document the appropriateness of each order and provide value-assurance to the patient and measurable, comparable data to the payer (insurer).

6. Conclusion

The preponderance of evidence indicates that CDSSs are effective to some degree in the preventing medical errors and in improving patient safety, especially when embedded within an EMR and directly intercalated into the care process. CDSSs are generally able to alter physician behaviour and influence the process of care. Although the results of support CDSSs have been far less positive when applied to the problem of improving clinical diagnosis, or improving ongoing care of patients with chronic diseases, advances can be expected in the future.

An effective CDSS can assist users of an EMR to significantly reduce medical errors and thus making healthcare more efficient and promoting the quality of health care. Despite the federal government's recent unveiling of grants and incentives for the adoption of HIT, health care providers still face numerous challenges in transitioning to the full adoption of EMR systems (Hart, 2009). Nonetheless, CDSS remains a critical factor in reaping benefits from the adoption of EMRs.

7. References

ActiveHealth Management. (2005*). Computerized Decision Support System Reduces Medical Errors, Cuts Costs.* Retrieved July 31, 2009, from http://www.medicalnewstoday.com/articles/19959.php.

Bakken, S.; L. M. Currie; N. J. Lee; W. D. Roberts; S. A. Collins & J. J. Cimino (2008). Integrating evidence into clinical information systems for nursing decision support. *Int J Med Inform* 77(6): 413-420.

Bates, D. W.; M. Cohen; L. L. Leape; J. M. Overhage; M. M. Shabot & T. Sheridan (2001). Reducing the frequency of errors in medicine using information technology. *J Am Med Inform Assoc* 8(4): 299-308.

Bates, D. W. & A. A. Gawande (2003). Improving safety with information technology. *N Engl J Med* 348(25): 2526-2534.

Bates, D. W.; G. J. Kuperman; S. Wang; T. Gandhi; A. Kittler; L. Volk; C. Spurr; R. Khorasani; M. Tanasijevic & B. Middleton (2003). Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. *J Am Med Inform Assoc* 10(6): 523-530.

Bates, D. W.; L. L. Leape; D. J. Cullen; N. Laird; L. A. Petersen; J. M. Teich; E. Burdick; M. Hickey; S. Kleefield; B. Shea; M. Vander Vliet & D. L. Seger (1998). Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA* 280(15): 1311-1316.

Bayegan, E. & S. Tu (2002). The helpful patient record system: problem oriented and knowledge based. *Proc AMIA Symp*: 36-40.

Berner, E. S. (2007). *Clinical decision support systems: Theory and practice*. New York, Springer.

Brown, S. H.; R. A. Miller; H. N. Camp; D. A. Guise & H. K. Walker (1999). Empirical derivation of an electronic clinically useful problem statement system. *Ann Intern Med* 131(2): 117-26.

Cabana, M. D.; C. S. Rand; N. R. Powe; A. W. Wu; M. H. Wilson; P. A. Abboud & H. R. Rubin (1999). Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 282(15): 1458-1465.

Chaudhry, B. (2008). Computerized clinical decision support: will it transform healthcare? *J Gen Intern Med* 23 Suppl 1: 85-87.

Chin, T. (2004). Technology valued, but implementing it in practice is slow. *American Medical News*.

Clinton, H. R. & B. Obama (2006). Making patient safety the centerpiece of medical liability reform. *N Engl J Med* 354(21): 2205-2208.

Colombet, I.; A. R. Aguirre-Junco; S. Zunino; M. C. Jaulent; L. Leneveut & G. Chatellier (2005). Electronic implementation of guidelines in the EsPeR system: a knowledge specification method. *Int J Med Inform* 74(7-8): 597-604.

Dovey, S. M.; R. L. Phillips; L. A. Green & G. E. Fryer (2003). Types of medical errors commonly reported by family physicians. *Am Fam Physician* 67(4): 697.

Edwards, A. & G. Elwyn (2004). Involving patients in decision making and communicating risk: a longitudinal evaluation of doctors' attitudes and confidence during a randomized trial. *J Eval Clin Pract* 10(3): 431-437.

Foster, R. S. & S. K. Heffler. *(2009). Updated and extended national health expenditure projections, 2010-2019.* Retrieved July 30, 2009, from http://www.cms.hhs.gov/NationalHealthExpendData/Downloads/NHE_Extended_Projections.pdf.

Friedlin, J.; P. R. Dexter & J. M. Overhage (2007). Details of a successful clinical decision support system. *AMIA Annu Symp Proc*: 254-258.

Galanter, W. L.; D. B. Hier; C. Jao & D. Sarne (2008). Computerized physician order entry of medications and clinical decision support can improve problem list documentation compliance. *Int J Med Inform*.

Garg, A. X.; N. K. Adhikari; H. McDonald; M. P. Rosas-Arellano; P. J. Devereaux; J. Beyene; J. Sam & R. B. Haynes (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA* 293(10): 1223-1238.

Graham, I. D.; J. Logan; A. O'Connor; K. E. Weeks; S. Aaron; A. Cranney; R. Dales; T. Elmslie; P. Hebert; E. Jolly; A. Laupacis; S. Mitchell & P. Tugwell (2003). A qualitative study of physicians' perceptions of three decision aids. *Patient Educ Couns* 50(3): 279-283.

Grasso, B. C.; R. Genest; K. Yung & C. Arnold (2002). Reducing errors in discharge medication lists by using personal digital assistants. *Psychiatr Serv* 53(10): 1325-1326.

Gross, P. A. & D. W. Bates (2007). A pragmatic approach to implementing best practices for clinical decision support systems in computerized provider order entry systems. *J Am Med Inform Assoc* 14(1): 25-28.

Hall, J. (2009). First, make no mistakes. *The New York Times*. New York.

Hart, K. (2009*). Electronic medical records grants face challenges. The Hill* Retrieved August 24, 2009, from http://thehill.com/leading-the-news/electronic-medical-records-grants-pose-challenges-2009-08-24.html.

Hartung, D. M.; J. Hunt; J. Siemienczuk; H. Miller & D. R. Touchette (2005). Clinical implications of an accurate problem list on heart failure treatment. *J Gen Intern Med* 20(2): 143-147.

HealthGrades (2004). Quality Study: Patient safety in American hospitals.

HealthGrades (2007). Fourth Annual Patient safety in American hospitals Study.

HealthGrades (2008). Fifth Annual Patient safety in American hospitals Study

HealthGrades (2009). Sixth Annual Patient safety in American hospitals Study.

Hier, D. B. (2002). Audit of medical records at the University of Illinois Hospital, University of Illinois Medical Center at Chicago.

Hillestad, R.; J. Bigelow; A. Bower; F. Girosi; R. Meili; R. Scoville & R. Taylor (2005). Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health Aff (Millwood)* 24(5): 1103-1117.

Hunt, D. L.; R. B. Haynes; S. E. Hanna & K. Smith (1998). Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA* 280(15): 1339-1346.

IVANS. (2009*). White paper: Survey results show long-term care providers wary about healthcare reform.* June 11, 2009. from http://www.ivans.com/news/news_detail.aspx?id=39.

Jao, C. S.; D. B. Hier & W. Wei (2004). *Simulating a problem list decision support system: can CPOE help maintain the problem list?* Proceedings of the 11th World Congress on Medical Informatics, San Francisco, CA: 1666.

Jao, C.; D. Hier & W. Galanter (2008a). Automating the maintenance of problem list documentation using a clinical decision support system. *AMIA Annu Symp Proc*: 989.

Jao, C.; D. Hier; W. Galanter & A. Valenta (2008b). Assessing Physician Comprehension of and Attitudes toward Problem List Documentation. *AMIA Annu Symp Proc*: 990.

Jao, C. S.; D. B. Hier & W. L. Galanter (2008). *Using clinical decision support to maintain medication and problem lists*. 2008 IEEE International Conference on Systems, Man and Cybernetics (SMC 2008), Singapore: 739-743.

Johnson, N.; D. Mant; L. Jones & T. Randall (1991). Use of computerised general practice data for population surveillance: comparative study of influenza data. *BMJ* 302(6779): 763-765.

Johnston, M. E.; K. B. Langton; R. B. Haynes & A. Mathieu (1994). Effects of computer-based clinical decision support systems on clinician performance and patient outcome. A critical appraisal of research. *Ann Intern Med* 120(2): 135-142.

Kaboli, P. J.; B. J. McClimon; A. B. Hoth & M. J. Barnett (2004). Assessing the accuracy of computerized medication histories. *Am J Manag Care* 10(11 Pt 2): 872-877.

Kaushal, R.; K. N. Barker & D. W. Bates (2001a). How can information technology improve patient safety and reduce medication errors in children's health care? *Arch Pediatr Adolesc Med* 155(9): 1002-1007.

Kaushal, R.; D. W. Bates; C. Landrigan; K. J. McKenna; M. D. Clapp; F. Federico & D. A. Goldmann (2001b). Medication errors and adverse drug events in pediatric inpatients. *JAMA* 285(16): 2114-2120.

Kaushal, R.; K. G. Shojania & D. W. Bates (2003). Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. *Arch Intern Med* 163(12): 1409-1416.

Kawamoto, K.; C. A. Houlihan; E. A. Balas & D. F. Lobach (2005). Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ* 330(7494): 765.

Kohn, L. T.; J. M. Corrigan & M. S. Donaldson (2000). *To err is human: Building a safer health system*. Washington, Institute of Medicine (U.S.) Committee on Quality of Health Care in America, National Academies Press.

Kuperman, G. J.; A. Bobb; T. H. Payne; A. J. Avery; T. K. Gandhi; G. Burns; D. C. Classen & D. W. Bates (2007). Medication-related clinical decision support in computerized provider order entry systems: a review. *J Am Med Inform Assoc* 14(1): 29-40.

Kuperman, G. J.; R. M. Reichley & T. C. Bailey (2006). Using commercial knowledge bases for clinical decision support: opportunities, hurdles, and recommendations. *J Am Med Inform Assoc* 13(4): 369-371.

Leape, L. L. & D. M. Berwick (2005). Five years after To Err Is Human: what have we learned? *JAMA* 293(19): 2384-2390.

Liu, J.; J. C. Wyatt & D. G. Altman (2006). Decision tools in health care: focus on the problem, not the solution. *BMC Med Inform Decis Mak* 6: 4.

Louis and Harris Associates (2007). 100 million Americans see medical mistakes drectly touching them as patients, friends, relatives, National Patient Safety Foundation.

McGlynn, E. A.; S. M. Asch; J. Adams; J. Keesey; J. Hicks; A. DeCristofaro & E. A. Kerr (2003). The quality of health care delivered to adults in the United States. *N Engl J Med* 348(26): 2635-2645.

Musen, M. A.; Y. Shahar & E. H. Shortliffe (2001). Clinical Decision-Support Systems. *Medical Informatics*. E. H. Shortliffe and L. E. Perreault. New Yrok, Springer-Verlag.

Nelson, K. (2002). Close case abstract: follow-up, document, communication. *Failure to diagnose cancer: reducing the risks*. M. Schaefer, Forum at risk Management Foundation of Harvard Medical Institutes. 22**:** 25.

Osheroff, J. A.; J. M. Teich; B. Middleton; E. B. Steen; A. Wright & D. E. Detmer (2007). A roadmap for national action on clinical decision support. *J Am Med Inform Assoc* 14(2): 141-145.

Parente, S. T. & J. S. McCullough (2009). Health information technology and patient safety: Evidence from panel data. *Health Affairs* 28(2): 357-360.

Pestotnik, S. L. (2005). Expert clinical decision support systems to enhance antimicrobial stewardship programs: insights from the society of infectious diseases pharmacists. *Pharmacotherapy* 25(8): 1116-1125.

Pryor, T. A. (1990). Development of decision support systems. *Int J Clin Monit Comput* 7(3): 137-146.

Raschke, R. A.; B. Gollihare; T. A. Wunderlich; J. R. Guidry; A. I. Leibowitz; J. C. Peirce; L. Lemelson; M. A. Heisler & C. Susong (1998). A computer alert system to prevent injury from adverse drug events: development and evaluation in a community teaching hospital. *JAMA* 280(15): 1317-1320.

Richardson, W. S. (2007). We should overcome the barriers to evidence-based clinical diagnosis! *J Clin Epidemiol* 60(3): 217-227.

Rooney, W. R. (2003). Maintaining a medication list in the chart. *Fam Pract Manag* 10(3): 52-54.

Rothschild, A.; D. B. Hier; L. A. & J. Keeler (2000). Enthusiastic adapter and reluncant users: faculty physician attitude towards an electronic health record one year after implementation. *Internal Document*, University of Illinois Hospital

Rowe, A. K.; F. Onikpo; M. Lama; F. Cokou & M. S. Deming (2001). Management of childhood illness at health facilities in Benin: problems and their causes. *Am J Public Health* 91(10): 1625-1635.

Simborg, P. W.; B. H. Starfield & S. D. Horn (1976). Information factors affecting problem followup in ambulatory care. *Medical Care* 14: 848-856.

Sittig, D. F.; M. A. Krall; R. H. Dykstra; A. Russell & H. L. Chin (2006). A survey of factors affecting clinician acceptance of clinical decision support. *BMC Med Inform Decis Mak* 6: 6.

Sittig, D. F.; A. Wright; J. A. Osheroff; B. Middleton; J. M. Teich; J. S. Ash; E. Campbell & D. W. Bates (2008). Grand challenges in clinical decision support. *J Biomed Inform* 41(2): 387-392.

Smith, D. H.; N. Perrin; A. Feldstein; X. Yang; D. Kuang; S. R. Simon; D. F. Sittig; R. Platt & S. B. Soumerai (2006). The impact of prescribing safety alerts for elderly persons in an electronic medical record: an interrupted time series evaluation. *Arch Intern Med* 166(10): 1098-1104.

SoRelle, R. (2000). Reducing the rate of medical errors in the United States. *Circulation* 101(3): E39-40.

Starfield, B. (2000). Deficiencies in US medical care. *JAMA* 284(17): 2184-2185.

Starfield, B. D.; I. Steubwachs; G. Morris; G. Bause; S. Siebest & C. Westin (1979). Concordance between medical records and observations regarding information on coordination of care. *Medical Care* 17: 758-766.

Studdert, D. M.; M. M. Mello; W. M. Sage; C. M. DesRoches; J. Peugh; K. Zapert & T. A. Brennan (2005). Defensive medicine among high-risk specialist physicians in a volatile malpractice environment. *JAMA* 293(21): 2609-2617.

Toth-Pal, E.; I. Wardh; L. E. Strender & G. Nilsson (2008). Implementing a clinical decision-support system in practice: a qualitative analysis of influencing attitudes and characteristics among general practitioners. *Inform Health Soc Care* 33(1): 39-54.

Trivedi, M. H.; E. J. Daly; J. K. Kern; B. D. Grannemann; P. Sunderajan & C. A. Claassen (2009). Barriers to implementation of a computerized decision support system for depression: an observational report on lessons learned in "real world" clinical settings. *BMC Med Inform Decis Mak* 9: 6.

Trowbridge, R. & S. Weingarten (2001). Clinical Decision Support Systems. *Making Health Care Safer: A Critical Analysis of Patient Safety Practice*. K. G. Shojania, B. W. Duncan, K. M. McDonald and R. M. Wachter. Rockville, MD, Agency for Healthcare Research and Quality.

Wadhwa, R. (2008). Analysis of a failed clinical decision support system for management of congestive heart failure. *AMIA Annu Symp Proc*: 773-777.

Walter, Z. & M. S. Lopez (2008). Physician acceptance of information technologies: Role of perceived threat to professional autonomy. *Decision Support Systems* 408(1): 206-215.

Wang, S. J.; B. Middleton; L. A. Prosser; C. G. Bardon; C. D. Spurr; P. J. Carchidi; A. F. Kittler; R. C. Goldszer; D. G. Fairchild; A. J. Sussman; G. J. Kuperman & D. W. Bates (2003). A cost-benefit analysis of electronic medical records in primary care. *Am J Med* 114(5): 397-403.

Wasserman, H. & J. Wang (2003). An applied evaluation of SNOMED CT as a clinical vocabulary for the computerized diagnosis and problem list. *AMIA Annu Symp Proc*: 699-703.

Weed, L. L. (1968a). What physicians worry about: how to organize care of multiple-problem patients. *Mod Hosp* 110(6): 90-94.

Weed, L. L. (1968b). Medical records that guide and teach. *N Engl J Med* 278(11): 593-600.

Weed, L. L. (1968c). Medical records that guide and teach. *N Engl J Med* 278(12): 652-657 concl.

Wen, P. (2008). 1 in 10 patients gets drug errors: Study examines six community hospitals in Mass. *The Boston Globe*. Boston, MA.

WHO (2005). Global eHealth Survey 2005. Geneva, World Health Organization.

Wolfstadt, J. I.; J. H. Gurwitz; T. S. Field; M. Lee; S. Kalkar; W. Wu & P. A. Rochon (2008). The effect of computerized physician order entry with clinical decision support on the rates of adverse drug events: a systematic review. *J Gen Intern Med* 23(4): 451-458.

Wright, A.; D. F. Sittig; J. S. Ash; S. Sharma; J. E. Pang & B. Middleton (2009). Clinical Decision Support Capabilities of Commercially-available Clinical Information Systems. *J Am Med Inform Assoc*: 2009 Jun 30. [Epub ahead of print].

9

Knowledge Bases for Clinical Decision Support in Drug Prescribing – Development, Quality Assurance, Management, Integration, Implementation and Evaluation of Clinical Value

Birgit Eiermann[1,2], Pia Bastholm Rahmner[1], Seher Korkmaz[1],
Carina Landberg[1], Birgitta Lilja[1], Tero Shemeikka[1], Aniko Veg[1],
Björn Wettermark[1,2] and Lars L Gustafsson[1,2]
[1]Departmentof Drug Management and Informatics, Stockholm County Council
[2]Division of Clinical Pharmacology, Department of Laboratory Medicine, Karolinska Institutet, Stockholm,
Sweden

1. Introduction

The access and use of information technology is increasing in all parts of society and in particular in the health care sector in developed and developing countries (Bates & Gawanda, 2003; Lucas, 2008). The integration of health information technology into health care institutions governs the agenda in most countries presently (Lucas, 2008; Gustafsson et al. 2003). The US has recently enacted a $ 19 billion program to promote the use and adoption of health information technology (Blumenthal, 2009) and information systems including electronic health records (EHR). This program is seen as an essential component to improve the health of every American. Challenges discussed span over the whole area of installing electronic health records, supporting and updating the systems, assistance with the interoperability, training the personal, and implementation of the systems as well as medical education (Blumenthal, 2009). Information technology, in particular computerized decision support systems, is also seen in the recent report by the Institute of Medicine in the US as a key way to address the identified great risk of medication errors in American health care institutions (Aspen, 2006).

A recent European report published by the Swedish government analyses health care in 6 European member states. The report describes the impact of health technology on several political goals such as increasing the availability of health care, continuity, empowerment of patients, patient safety and quality of care. It states that in the 6 European member states studied, 100 000 yearly inpatient adverse drug events (ADE´s) could be avoided through usage of computerised physician order entry systems (CPOEs) with clinical decision support (CDS), which would correspond to a yearly saving of 300 million € (Gartner, 2009). This report combined with other studies and reviews (Sjöborg et al, 2007; Kelly et al., 2006) underlines the complexity of integration and implementation, including local conditions, the involvement of stakeholders and adoption and measurements of changes, all of which have to be tackled for a beneficial usage of the technology. The Gartner report envisages that increasing costs within the health care sector will accelerate the efforts to develop new technologies as well as lead to a beneficial usage of existing systems.

Computerised physician order entry systems (CPOEs) are one step towards increased safety in patient safety. They allow physicians and other health care staff to prescribe patient medication directly by using a computer, replacing hand-written orders, and thereby eliminating possible interpretation and transcription faults. Transcription and/or interpretation errors have been shown to cause 11% of all medication errors resulting in adverse drug events in hospitals (Krahenbuhl-Melcher et al., 2007). An additional step to improving patient safety and efficacy in the prescribing process is the integration of clinical decision support systems (CDS systems) within the CPOEs. This allows physicians to retrieve up-to-date medical knowledge of the optimal/recommended management of the diseases and drugs, thereby improving patient care through enhancing compliance with recent guidelines and recommendations. CDS systems deliver their information through knowledge bases (e.g. drug-drug or drug-food interactions, drugs & pregnancy, drugs & lactation, drug dosage according to kidney function and genotype and in risk groups), which are integrated through software algorithms, that will generate alerts, warnings and recommendations during drug prescribing. For optimisation of the effect of CDS systems, they should be integrated into EHRs´ resulting in patient specific recommendations and alerts using patient characteristics available in the EHRs.

Numerous studies have demonstrated positive effects of CDS systems in various settings including hospital or ambulatory care, intensive care units and in pediatric care (Ammenwerth et al. 2008, Eslami et al. 2008, Wolfstadt et al. 2008, van Rosse et al. 2009). Areas of improvement identified include costs, safety, adherence, alerts, user satisfaction and time. Reduction of medication errors have been demonstrated with the introduction of CDS systems as well as reduction of ADE´s. However, further studies are needed which focus directly on patient outcomes rather than the surrogate outcome such as "practitioners' performance" to further accelerate their introduction (Garg et al., 2006). Many studies pinpoint improvements of the knowledge bases or CDS systems including optimization of the content (Luna et al. 2007), introduction of classification systems to knowledge bases (Böttiger et al., 2009), and tiering alerts through introduction of severity levels (Paterno et al., 2009). Their introduction is likely to further improve CDS systems.

Evidence is growing though that CDS systems might not only lead to improved quality in health care but they can themselves create unintended errors jeopardizing patient safety (Ash et al., 2004). Introduction of CDS systems might cause diminished medical judgement, letting the computer overrule physicians´ own professional knowledge. Additional work tasks might create disturbances in the already burdened physicians work flow resulting in inefficiencies of the systems. Also the complexity of the systems increases the potential in design flaws thereby actually introducing new errors rather than preventing them (Bates et al. 2001). Therefore, the implementation and use of any CDS system should be linked to the establishment of a medical management, maintenance and quality assurance system, which leads to discovering, analysing and foreseeing possible errors.

Being responsible for paying the drug bill Stockholm County Council, the largest health care provider in Sweden, implemented a health care strategy in 1997 including the development of an IT architecture (Sjöborg et al., 2007). The aim was to provide numerous services to the prescribers to ensure safe and effective drug prescribing. Additional initiatives where started, like the formation of drug expert groups providing a Wise Drug List, which contains a list of about 240 first line drugs for common diseases incorporating therapeutic ladders or guidelines. Recommendations from 23 expert groups and 5 local drug and therapeutic committees are used to produce and refine the guidance. On the IT site Stockholm County Council in collaboration with Karolinska Institutet and other academic partners has designed, developed and implemented a prescribing tool (Eliasson et al., 2006) and the content for medical knowledge bases for drug-drug interactions, Sfinx (**S**wedish **F**innish **In**teraction **X**-Referencing), drugs & pregnancy, and drugs & breast feeding (Nörby et al., 2006, Böttiger et al., 2009). The knowledge bases are integrated into clinical decision support tools (Janus toolbar described below) or are accessible through the web (www.janusinfo.se). This strategy has been combined with a range of initiatives to promote rational use of drugs as described by Godman et al., (2009). The different knowledge bases and their life cycle from development to evaluation are used as examples for our own experiences in the following parts.

The review is based on more than 10 years experiences from joint efforts to develop, implement and evaluate user friendly and effective decision support systems for drug prescribing in Stockholm. It summarizes state-of-the art knowledge on development, integration, maintenance, implementation and evaluation of knowledge bases and CDS systems used for rational drug prescribing. Consequently, we see this review as a first step in the process of creating robust future models and international standards for the retrieval of medical and pharmacological knowledge, its conversion and organisation into knowledge bases, as well as their integration into CDS systems, their management and evaluation of user satisfaction and treatment outcome.

2. Development of knowledge databases

Why are knowledge bases needed and what advantages do they offer compared to other sources like e.g. the official product SPC (summary of products characteristics) issued as part of the registration of a drug product? One advantage with knowledge bases is the standardisation of information for all drugs containing the same substance. For instance the content of individual SPCs or physician desk references may vary considerably between drugs containing the same substance and with identical drug formulations produced by different pharmaceutical companies. This can cause confusion for the prescribing physician. For example information about drug-drug interactions can be found in the SPC for drug A from provider 1 but is missing in the SPC for the same product from provider 2. Alternatively, the drug-drug interaction between drug A and B can be found in the SPC text for drug A but not for drug B (Bergk et al., 2005). Another example for inconsistencies is the classification for drug and pregnancy alerts for pharmaceutical products. One provider may state, that the drug should be avoided during pregnancy, but another drug company may state, that the drug can be used without any problems. Other examples are variations in dosing information (maximum recommended therapeutic dose) between SPCs´ from different providers or in information published by the US Food and Drug Administration (Seidling et al., 2007). Consequently, knowledge bases should help by providing more consistent information about the substances and drugs related to that substance.

The starting point for development of any knowledge base is the analysis of the perceived needs of the potential users in the health care system (Revere et al., 2007). Likewise it is important to assess the potential of a new knowledge base to improve efficacy and safety in drug prescribing (Gustafsson et al., 2003; Schiff & Rucker, 1998). We believe the formation of user groups should be mandatory to explore the functional and content needs for a knowledge base and decision support system before other activities are undertaken (Eliasson et al., 2006). Consequently, a multidisciplinary group of clinical experts within the medical field the knowledge base should be aimed for (e.g. nephrologists for a database about drug dosage in patients with reduced kidney function) together with drug experts (e.g. clinical pharmacologists specialised on drug dosage, drug-drug interactions or drugs & lactation depending on the knowledge base), future users, experts within existing drug registries and software developers, should be convened to discuss the potential and obstacles for the knowledge base (Ash et al., 2004). Our own experience is that there often is a mismatch between users' expectations and the clinical and medical research basis or the availability of certain parameters or features within existing registries. For example, the clinical specialists will focus on one specific recommendation for the most common indications for both drugs for a certain drug-drug interaction. However, this recommendation might not fit all patient cases for which this pertinent drug-drug interaction alert will be shown, leading sometimes to suboptimal recommendations.

Another example is that the recommendation to achieve a certain therapeutic drug concentration interval can only be given, if there is scientific evidence. In addition, even though the potential user of a knowledge base (the general practitioner or any other physician) and the drug expert have the same basic medical education, they do not "talk the same language" and medical and clinical expertise differs, with clinical experts having more knowledge about patient treatment while drug experts possess more information about the properties of the drugs used. Medical advice given by the drug expert might not suit the practical needs of the physician and on the other hand the specialist physicians´ needs may not be fulfilled due to missing medical evidence.

It is also very important to clarify when CDS systems or knowledge bases can help and when they can't. For example, during the development of the drug-drug interaction database Sfinx one of the future users mentioned, that he now finally can detect all the drug interactions for herbal drugs his patients always take. But since this physician never enters herbal drugs to the patient's drug list, because he is not prescribing them, he will never get a warning for these drugs.

Prior to the development of the content of a knowledge database the multidisciplinary group needs to define its structure. For example developing the drug-drug interaction database (DDI db), Sfinx, physicians wished not only to receive warnings on certain drug interactions, but also recommendations on how to avoid and handle this interaction (Böttiger et al., 2009). The recommendation part is extremely important, since physicians do not only want warnings on avoiding certain drug combinations, but would like a recommendation how to handle the situation. In a survey among prescribers and pharmacists in the US both groups demanded that drug-drug interaction alerts should be accompanied by management options of the DDI (Ko et al., 2007). In a recent Australian study (Sweidan et al., 2009) recommendations for handling of drug-drug interactions were seen as a quality measure for the DDI databases. However, comparing 9 drug interaction systems used in primary care only 1 out of 9 systems provided useful management advices.

A number of studies have demonstrated, that physicians need timely, easy to digest and up-to-date information, which is filtered, summarized, and synthesized from reliable sources by clinical respected experts (Revere et al., 2007; Grol & Grimshaw, 2003; Schiff & Rucker, 1998). The expert group needs to define the relevant sources to be used for the knowledge base, which might consist of recent research publications, legal documents, information from pharmaceutical companies, textbooks, and other databases. It is critical for the integrity of the knowledge base to use scientifically rigorous methods for evaluation of scientific data by applying critical drug evaluation principles (Godman et al., 2009). Search strategies have to be developed and documented in standard operation procedure protocols (SOP´s) to assure reproducibility of the search results (Böttiger et al., 2009). This is critical in all cases but especially if different people are executing the same task or if expert groups are located in different places and can not communicate with each other on a daily basis.

Figure 1 describes the process from filling a knowledge base with data to providing it to the end user. Literature searched will be evaluated by different experts regarding their clinical relevance and their level of documentation according to standardised rules. It will then be synthesized into short text messages, according to a predefined structure (Böttiger et al., 2009). Different content providers have to use the same tool for data entrance. It is advisable

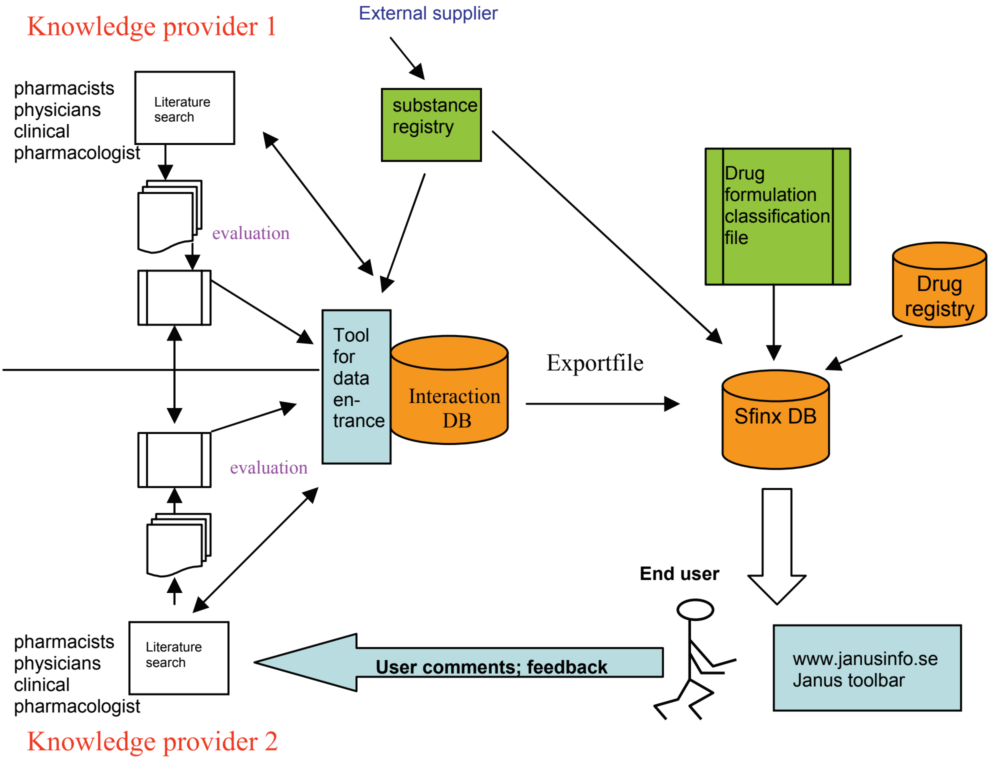


Fig. 1. The design and process of building and maintaining a knowledge base to be integrated into CDS systems for drug prescribing at point of care or accessible through the web. User feed back triggers new literature searches and improves the quality of the knowledge base. This figure outlines the data management process for the drug-drug interaction database Sfinx (Böttiger et al., 2009).

to define in advance certain standard terms for the text messages to avoid heterogenicity in the text content. It is easier for the end user to recognise standard phrases for certain conditions or recommendations.

Data are entered into the knowledge base and connected to various other registries or databases to assure optimal usage in the CDS system used by the prescribers. We have experienced, that access to experts is always the bottleneck in the production of knowledge bases, which agrees with the experiences by Kuperman et al. (2006). Consequently, data entrance into the knowledge base has to be simplified to save valuable expert time. For example through an easy text sharing function the experts should be able to reuse the same texts in different documents (e.g. interactions), which follow the same interaction mechanisms and rules, as a result reducing both the time for entering data and the size of the database. We have developed a terminology model for substances so that substances within a class such as different salts of the same substance belong to the same mother substance, if they react in the same way (Böttiger et al., 2009). The grouping of substances and text sharing function results in an effective and easy way to use the tool which simplifies data entrance and database updates.

Knowledge bases need to be connected to certain registries through software algorithms for their optimal use (see Fig. 1). Because texts in knowledge bases usually are written on a substance basis they need to be linked to specific drugs which contain the substance. This linkage can be done using substance registries, which contain substance names and drugs connected to the substance. Key fields for the linkage can be:

- ATC (Anatomical, therapeutic, chemical classification) codes
- CAS (Chemical Abstract Service) numbers
- other nationally available unique identifiers (Böttiger et al. 2009).

All systems have advantages and disadvantages. The ATC code system is valuable since it takes indications of drugs into account. This can be used to link or exclude drugs containing the same substance, but with different formulations or used in different strengths. A disadvantage of the ATC code system is the handling of combinational drugs, where the content of the drug most often is not specifically defined by the code. CAS numbers identify each substance in a unique way, which allows correct linkage. Problems within the system are its complexity. For example a substance which appears to be the same might have a different CAS number due to its varying content of crystal water, which is not obvious from the description of the drug. Another disadvantage is the limited use of CAS numbers in national registries. National identifiers might be the optimal way for linkage of knowledge bases to drug registries. However, substance based national identifiers do not take drug dosages into account. A substance can have a different interaction profile due to variations in dose (for example: high-dose versus low-dose acetylsalicylic acid). Consequently linkage just by a substance identifier would lead to wrong interactions alerts. Another disadvantage of national identifiers is, that the national identifier can't be used across nations, a problem we faced for the Sfinx database distributed in Sweden and Finland. Whatever registry or system is used for linkage it is of great importance to ensure correct update and maintenance of the registries as described in the next chapter.

If drug formulations are relevant for the triggered alert, these should be taken into account. Even here it is important to simplify matters for the knowledge expert and create drug formulation groups (e.g. all sorts of tablets, capsules or oral solutions should be grouped under the term "peroral"). The Swedish drug registry contains about 650 different drug formulations, which we have grouped into 5 different groups in the "drug formulation classification file" (Fig.1) to support data entrance. International standard terms are needed to reduce the work load for a single country and to facilitate the integration with other drug registries from other nations. To our knowledge there is no European or worldwide registry with standardised drug formulations available, which could facilitate integration of knowledge bases across countries.

Finally database updates have to follow the same procedures and rules as defined for the starting phase. Ideally, they should include incorporating the handling of end user comments and feedback for further improvement and refinement (Böttiger et al., 2009). Each specific update should be tested, documented and saved in order to be able to trace back incorrect alerts reported by the end users.

3. Combined quality assurance for knowledge bases and linked registries

Quality assurance of the knowledge bases and their linkage to other registries is an essential task often forgotten as it is time consuming, labour-intensive and requires significant effort and expertise. Quality assurance does not only refer to the medical content in question but stretches over the whole procedure from literature searches, the evaluation process to the linkage of the knowledge base to local or national registries and thereby requires experienced multidisciplinary staff.

Literature about quality assurance processes within clinical databases is limited. Quality assurance papers in medicine mainly deal with securing the quality for a certain medical treatment or procedure, but are not extended to databases and CDS systems. It is amazing that still today EHRs or CDS systems do not need to be certified by health or medical agencies. However, due to the increasing awareness of the possibility that information technology implemented into health care can actually increase the error rate even with risk for higher mortality rate (Han et al. 2005), changes are on their way both in the US (Blumenthal, 2009) and Europe (EU directive; 2007/47/EC;2007). Certification should cover not only the technical part of these systems, but should include even the medical content of knowledge bases integrated into CDS systems and implementation of the systems.

Quality assurance is mostly self evolving during the development phase of any database system including the handling of external registries for linkage purposes. Baorto et al. (2009) describe the experiences they made with the maintenance of a large medical ontology at one of the larger hospitals in New York. They state that the methods described even though developed specifically for their system can be used for carrying out similar tasks at other institutions. Many of the problems and procedures mentioned mirror exactly the situation with the development of our knowledge base systems. In our mind resources and expertise for quality assurance processes are needed for integration and maintenance of high quality knowledge bases. Standards need to be developed in this area.

Combining different registries or other knowledge sources has to be performed using "key fields" like ATC codes, specific identifiers, or CAS-numbers. We were surprised though when comparing different registries that the information in key fields could vary. For example a drug could be assigned to a specific ATC code in one registry and this could vary from the code in another registry. This could be due to simple typing mistakes, system requirements of the registry owner, delays in the update process or other possibilities. We, like Baorto and colleagues (2009), used the "diff approach" for detection of theses variations, where you compare two registries regarding the information in predetermined fields using the information in key fields to link the registries. For example we assume that a drug with a specific registration number has to have the same name, ATC code and drug formulation in all registries. Another approach to detect variations is to compare an older version of the same file with the newer one discovering changes for already existing fields, and new posts entered to the file. Logs are produced during the comparison process, which mainly have to be evaluated manually. Possible mistakes are corrected in the registries and reported to the source owners for correction in the original source.

Over the years we have discovered many mistakes at the point of acquisition of the data including missing information in essential fields (they were either completely empty or omitted), changes in the meaning of existing codes, existence of wrong characters in the master file or creation of redundant terms. The "diff approach" is also used for updating the knowledge base system, e.g. to identify new substances on the market, which will then be added and grouped into the mother child terminology, to discover new drug formulations, which have to be included into the drug formulation file, or to seek for new drugs on the market, which have to be linked to certain knowledge bases. The linkage has to be correct both technically and content wise else care will subsequently be compromised. For example, you can't link some new ear drops, containing a substance you already have in your knowledge base, to that base, if the text document is irrelevant for this new drug.

Auditing terminology and data structure of the registries linked to the knowledge base is mainly performed manually through reviewing log files created during the import process, which flag for changes and differences. These processes are labour-intensive and time consuming. Some of these audit processes though can be automated or at least semi-automated to save time and resources. For example, if you want to add a new substance child to the registry the hierarchy principle within your database requires the existence of the mother substance to be consistent with previously existing structures. Other examples are rules you create for maintenance purposes, like no two medications with the same registry number are allowed with different names.

As Baorto et al. (2009) stated quality assurance and maintenance of the knowledge base and its linked registries is a "mission critical" task that cannot tolerant errors. If we do not add one specific, new ATC code to a document the new drug assigned to that code will fail to be considered by the alerting system. It must be recognised though that all quality assurance processes rely at least partly on human surveillance so they are inevitably prone for mistakes. One can never be sure, that the knowledge base is completely correct. However, we can increase our confidence in the database through implementation of audits, rules and log files. This will help to create a system, which is detecting and minimising a large percentage of potential errors.

Any errors that occur through the usage of the knowledge base have to be handled by the medical management and maintenance system in a systematic way to enhance the utility of the database. This is described below.

4. Medical management and maintenance system for knowledge bases and CDS systems

The development and implementation of several knowledge bases, CDS systems and other IT applications within health care required the introduction of a surveillance system for possible errors introduced by its applications as an essential pre-requisite for the management of these systems. Our department has implemented a maintenance system, which allowed smooth handling of all procedures linked to its databases and depending registries e.g. regular update processes of the medical content, improvements or changes in the graphical interface and IT structure or adapting to new external registries. At the same time the EU Directive, 2007/47/EC, amending among others the Directive 93/42/EEC (http://Eur-Lex.europa.eu) concerning medical devices is under implementation in Sweden, and supports the process by raising the requirements for software and information systems used for clinical decisions regarding individual patients.

Important parts for the function of knowledge bases and CDS systems are management, maintenance and quality assurance of these applications after their implementation, together with handling of possible errors introduced by the systems, which could be of either technical or medical nature. Clinical, medical, and pharmaceutical competences, as well as competences in various IT-areas and in implementation are needed. Additionally, complete technical documentation of the knowledge base and the CDS system as well as guidelines (standard operation procedures = SOPs´) for producing their content and its distribution have to be part of the management plan to secure standardized procedures and avoid occurrence of mistakes.

Documented incidents include all kinds of subjects i.e. requests for further information, e-services, training, as well as reporting of major or minor errors. Minor or major errors include discussions of diverse opinions about recommendations or conclusions in the knowledge base, wishes for changes in classification levels or inclusion criteria. Technically it could be problems in the applications, or its documentation, or errors regarding the technical integration including design and functionality of user interfaces. All incidents are documented in the management system. Within the management and maintenance system the experiences we have made through real incidents and errors enable us to perform risk analysis on a regular basis. This helps us to foresee and judge possible incidents, which might occur through changes in the content, the graphical interface or the technical solution of our systems.

4.1 Management of errors using root cause analysis

The medical management of incidents or errors involves the processes of discovering the incidents, collecting documentation, performing event analysis and, if required, reporting of the error as a medical event - named Lex Maria - to the authorities in Sweden (Shemeikka et al., 2008). Root cause analysis (RCA) is a technique originally developed in psychology and systems engineering to identify “the basic and casual factors that underlie variation in performance”. We use RCA to investigate errors after they are discovered. It involves critical incident reporting followed by self-managed investigation of the event involving all staff in charge. It should answer three basic questions:

- what happened?
- why did it happen?
- what could be done to prevent it from happening again?

The investigation team consists of colleagues from the department and includes everybody involved in the processes related to the incident.

In the US root cause analysis for investigations of medical errors became mandatory in 1997 for hospitals accredited by the US Joint Commission on Health Care Safety. Models used for RCA were further developed and adopted for by health care systems in other countries like Australia (Iedema et al., 2005). Though RCA used in the US and other countries only included medical procedures and not handling of errors introduced by decision support systems, we have applied the technique for handling of our knowledge bases and decision support tools as well as the e-prescribing system. We have relied on experiences by team members from using the method in pharmaceutical companies to handle reports on adverse drug reactions, and from the health care system reporting events when harm or risk of harm for the patient has occurred during medical treatment.

Once an error of the CDS system is reported an initial rapid assessment is performed of the potential immediate and long term clinical consequences. If there is any risk for the safety of the patient or other patients due to the error, a decision is taken to shut down the e-services or keep it going whilst performing immediate changes. In these cases a report is sent to the national authorities in charge of monitoring and guarding patient safety during clinical care. The error is documented in detail often by requesting additional information from the reporter. The next step is to perform RCA to investigate the reason for the error (Iedema et al., 2005) and to suggests changes in for example the system, content, procedures or technical and user interfaces.

Incidents can be due to medical (e.g. wrong medical recommendation), pharmacological (e.g. wrong pharmacological mechanism thought to be cause for a DDI) or pharmaceutical (e.g. drugs with wrong formulations can be linked to a text) errors in the content of the knowledge base, or due to an unclear text, leading to misinterpretation. Errors in drug linkage can result in wrong alerts for a certain drug or missing alerts. The reason for the error could also be of technical nature. RCA may lead to organisational changes like education of the personal or policy changes, though they have a lower probability of reducing risk (Wu et al. 2008). It may also lead to changes of the content or processes for producing the knowledge bases or CDS systems or in redesign of the product or processes linked to knowledge base or CDS system, which are actions with a high probability of reducing risk (Wu et al., 2008). Procedural changes may lead to updates in the documentation or SOPs´ for the knowledge bases. Any changes in the device will be followed by extensive tests of the modified application before reintegration into the work environment. If the incident does not depend on one's own systems but on the EHR the CDS system is implemented in, the health record system owner has to solve the problem and document and proof changes. These changes are performed in close contact with vendors and producers of electronic health record systems.

Other incidents like inappropriate handling of the CDS system by the user may lead to a modification of the system and if necessary, user training must be performed. An example of a RCA is shown in figure 2 and 3. It describes an incident, where an ATC code was connected to a medical document by mistake. Drug name and ATC code was incorrectly send to the authors of the knowledge base. This led to the addition of the code to the document by the authors and a wrong linkage of drugs to the document. Processes for quality assurance of linkage of drugs to documents failed due to various reasons (technical equipment; frequent change of personal involved in the process). Consequently, users searching on the web for one of these drugs in one of our knowledge bases ended up in a document which had nothing to do with the drug searched for. Even if RCA has some benefits, including increased awareness of faulty processes and fixes to specific problems, more emphasis should be placed on drawing lessons across investigations rather than to approach each RCA independently. Most important, follow-up for implementation and outcome of each RCA and its actions should become a standard element of the process (Wu et al. 2008)

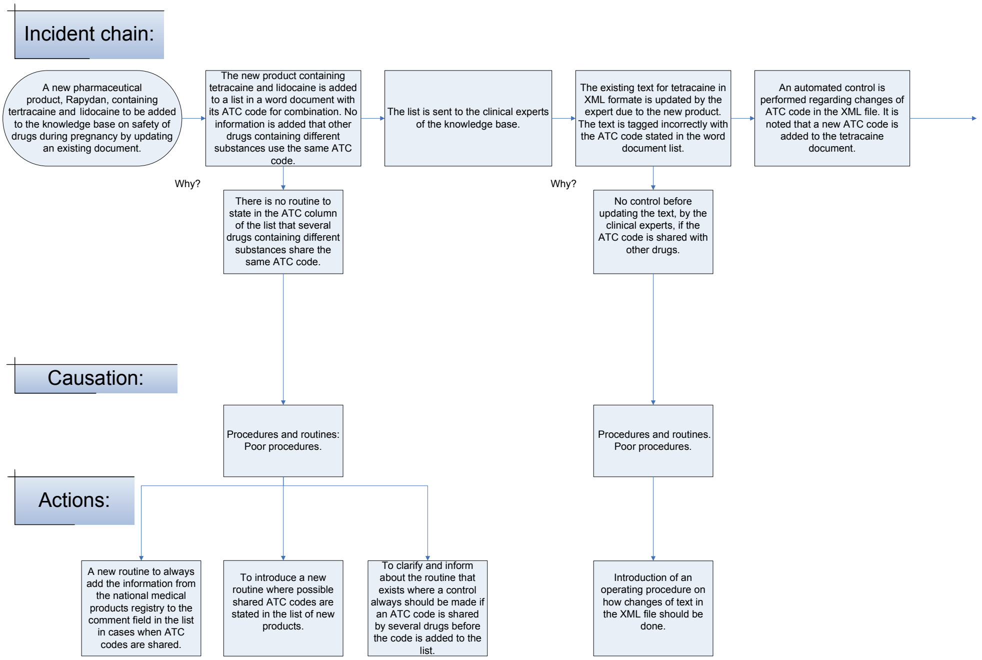

Incident chain:
A new pharmaceutical product, Rapydan, containing tertracaine and lidocaine to be added to the knowledge base on safety of drugs during pregnancy by updating an existing document.
The new product containing tetracaine and lidocaine is added to a list in a word document with its ATC code for combination. No information is added that other drugs containing different substances use the same ATC code.
The list is sent to the clinical experts of the knowledge base.
The existing text for tetracaine in XML formate is updated by the expert due to the new product. The text is tagged incorrectly with the ATC code stated in the word document list.
An automated control is performed regarding changes of ATC code in the XML file. It is noted that a new ATC code is added to the tetracaine document.
Why?
There is no routine to state in the ATC column of the list that several drugs containing different substances share the same ATC code.
Why?
No control before updating the text, by the clinical experts, if the ATC code is shared with other drugs.
Causation:
Procedures and routines: Poor procedures.
Procedures and routines. Poor procedures.
Actions:
A new routine to always add the information from the national medical products registry to the comment field in the list in cases when ATC codes are shared.
To introduce a new routine where possible shared ATC codes are stated in the list of new products.
To clarify and inform about the routine that exists where a control always should be made if an ATC code is shared by several drugs before the code is added to the list.
Introduction of an operating procedure on how changes of text in the XML file should be done.

Fig. 2. RCA part 1: On the top of each RCA the incidents following each other and leading to the mistake are stated. Next line gives the reasoning for each incident. These are followed by the causes, grouping the reasons into categories. Each reason is followed by one or several actions suggested.

4.2 Analyses of risks

Risk analyses are also included into the management and maintenance system. Using the experiences made with existing systems we apply this knowledge to other parts of the knowledge base and the CDS system to foresee possible risks. On a regular base we perform preventive risk analysis to identify and classify different kinds of risks. The method has been adapted and is now used even during development of new CDS systems or knowledge bases in our setting in Stockholm. It improves our possibilities to evaluate the costs, risks, and improvements made with the implementation of new knowledge bases or decision support tools. For example when the graphical interface of the decision support system is changed risk analysis can be performed on possible effects for end user performance.

5. Providing medical knowledge bases at point of care

The knowledge base can either be provided:

- as a website solution
- integrated into EHR systems
- used in learning tools.

The integration into EHR systems facilitate the exchange of patient-specific data with the knowledge base, thereby creating patient-specific alerts or reminders during the process of

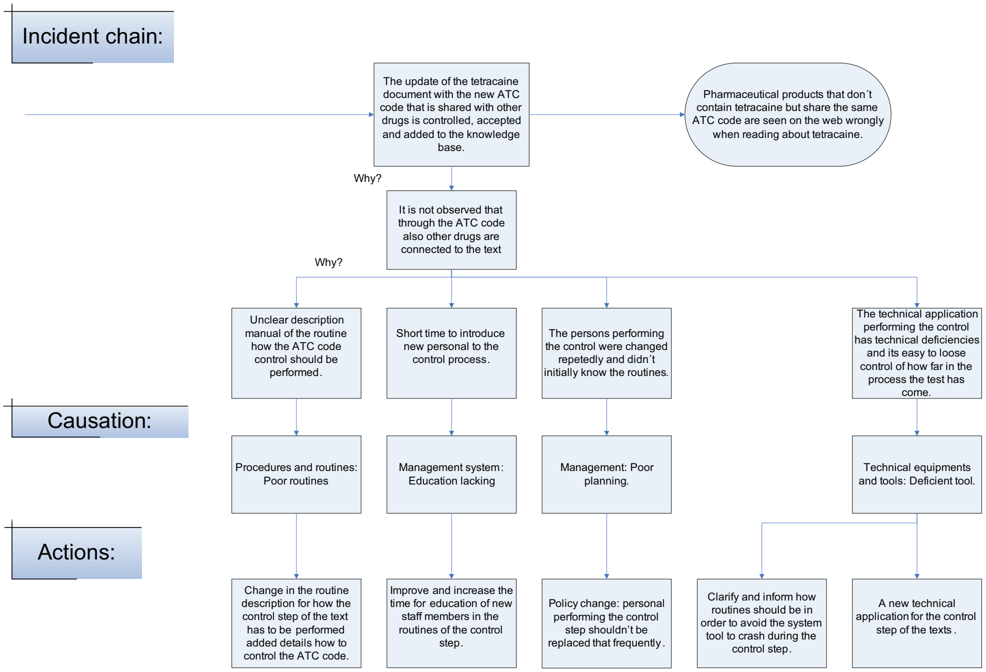

Incident chain:
The update of the tetracaine document with the new ATC code that is shared with other drugs is controlled, accepted and added to the knowledge base.
Pharmaceutical products that don´t contain tetracaine but share the same ATC code are seen on the web wrongly when reading about tetracaine.
Why?
It is not observed that through the ATC code also other drugs are connected to the text
Why?
Unclear description manual of the routine how the ATC code control should be performed.
Short time to introduce new personal to the control process.
The persons performing the control were changed repetedly and didn´t initially know the routines.
The technical application performing the control has technical deficiencies and its easy to loose control of how far in the process the test has come.
Causation:
Procedures and routines: Poor routines
Management system: Education lacking
Management: Poor planning.
Technical equipments and tools: Deficient tool.
Actions:
Change in the routine description for how the control step of the text has to be performed added details how to control the ATC code.
Improve and increase the time for education of new staff members in the routines of the control step.
Policy change: personal performing the control step shouldn't be replaced that frequently.
Clarify and inform how routines should be in order to avoid the system tool to crash during the control step.
A new technical application for the control step of the texts.

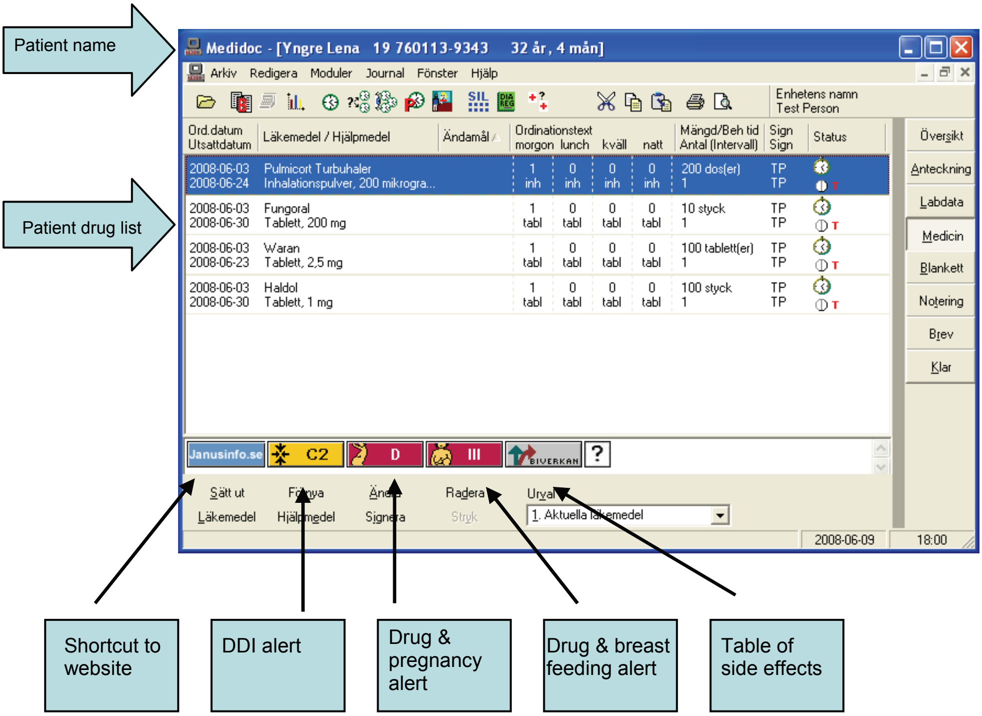
Fig. 3. RCA part 2: the second part of the incident chain explains, how the document with the wrong ATC code was added to the database without proper controls of the document and the effect it had on the linkage of drugs to the document. Actions suggested include changes in routines and policies, education and even changes in the technical tools used.

drug prescribing. The integration into an EHR system should be performed in collaboration between the providers of the knowledge base and the owners of the EHR systems. Contracts should specify the implementation of the database and how it is to be used and presented to the end user. The organizations implementing CDS systems must have detailed knowledge of the structure of the knowledge base and the architecture of the CDS system so that it is clear, how the systems interact (Kuperman et al.2006). Intensive testing of its integration following predefined protocols should be required to avoid unintended errors or mistakes due to lack of experiences and knowledge of the product. One must be sure that the knowledge base is behaving as intended (Kuperman et al. 2006).

The knowledge bases for drug-drug interactions, Sfinx, drugs & pregnancy and drugs & lactation produced by Stockholm County Council are provided free of charge through the county website on www.janusinfo.se. The website is aimed at health care personal. Physicians or nurses can search various knowledge bases by typing in the patient's medication and receive advice, whether specific drugs can be used during pregnancy or breast feeding or should be avoided (Norby et al., 2006). Drug-drug interactions can be searched for in Sfinx by either substance or drug names.

However, for optimal use knowledge bases should be implemented into a CDS system linked to an EHR, which will send patient specific data such as age, sex, height, weight, parameters for kidney function and the current drugs a patient is being prescribed to the knowledge base. Through certain software algorithms an alert or reminder could then be triggered or not, providing patient specific warnings for e.g. drug-drug interactions, drugs & lactation or drugs & breast feeding.

The DDI database Sfinx is integrated into the CDS system Janus toolbar, providing patient specific automatic alerts during drug prescribing (Sjöborg et al., 2007). In figure 4 we describe an example of the decision support system provided through Janus toolbar integrated into one EHR system in Stockholm County Council. The patient's name, sex and age can be seen at the top of the screen. The prescribing module within the EHR contains the current drug list, consisting of 4 different drugs. Sending those data to the knowledge base for DDIs´, pregnancy and breast feeding the alert buttons will be illuminated, if there is any information to be retrieved (Eliasson et al., 2006, Sjöborg et al., 2007). It is of great importance that the EHR and the knowledge base interact in an optimal and correct way. For example in a survey among ambulatory care clinicians in Massachusetts it was observed, that the local CDS system often delivered alerts with out-of-date medications, which led to scepticism towards the system among users (Weingart et al. 2009).



Medidoc - [Yngre Lena 19 760113-9343 32 år, 4 mån]

Enhetens namn
Test Person

Ord.datum Utsattdatum	Läkemedel / Hjälpmedel	Ändamål	Ordinationstext				Mängd/Beh tid Antal (Intervall)	Sign Sign	Status
			morgon	lunch	kväll	natt			
2008-06-03 2008-06-24	Pulmicort Turbuhaler Inhalationspulver, 200 mikrogra...		1 inh	0 inh	0 inh	0 inh	200 dos(er) 1	TP TP	T
2008-06-03 2008-06-30	Fungoral Tablett, 200 mg		1 tabl	0 tabl	0 tabl	0 tabl	10 styck 1	TP TP	T
2008-06-03 2008-06-23	Waran Tablett, 2,5 mg		1 tabl	0 tabl	0 tabl	0 tabl	100 tablett(er) 1	TP TP	T
2008-06-03 2008-06-30	Haldol Tablett, 1 mg		1 tabl	0 tabl	0 tabl	0 tabl	100 styck 1	TP TP	T

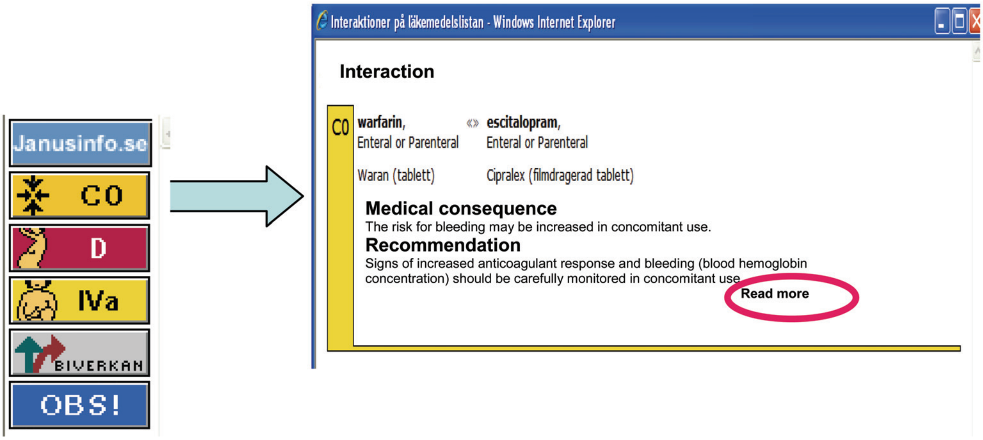
Urval: 1. Aktuella läkemedel

2008-06-09 18:00

Fig. 4. Implementation of Janus toolbar into an EHR. Patient specific alerts are illuminated related to the patient's age, sex and current list of drugs. Several different knowledge bases are the basis of the decision support system. For every new order of medication a new drug list will be send to the knowledge base, evaluated and may lead to changes in the alerts.

To further improve user friendliness, accessibility, and speed of the CDS system the most important information of the knowledge base should be short and concise only one click away. This principle is implemented into the Janus toolbar with the most important message being provided immediately, while for information about possible mechanisms, background studies for the statement and references users have to click further (Eliasson et al, 2007, Böttiger et al. 2009). We believe this quick access to pertinent information enhances the utilisation of the support tool. Even other surveys have shown that important information should be easily accessible and speed of use is a critical factor for the successful use of medical information systems (Dawes & Sampson, 2003; Bates et al., 2001).

Figure 5 shows the information provided by the knowledge base for drug-drug interactions, - Sfinx. Sfinx was developed by us together with partners from clinical pharmacology in Finland and in Sweden (Böttiger et al., 2009). Clicking on the yellow alert button, which is illuminated according to the patients' drug list, short and concise information about the medical consequence and recommendations can be seen immediately. Additional more educational information is available through clicking on the "read more" button.


Janusinfo.se
CO
D
IVa
BIVERKAN
OBS!
Interaktioner på läkemedelslistan - Windows Internet Explorer
CO
Read more

Interaction

warfarin, «» escitalopram,
Enteral or Parenteral Enteral or Parenteral
Waran (tablett) Cipralex (filmdragerad tablett)

Medical consequence

The risk for bleeding may be increased in concomitant use.

Recommendation

Signs of increased anticoagulant response and bleeding (blood hemoglobin concentration) should be carefully monitored in concomitant use.

Fig. 5. Warning texts of the drug-drug interaction database, Sfinx. The yellow colour code is used for interactions classified as "C" which means, that the interaction is clinical relevant but the drug combination can be handled by for example dose adjustment (Böttiger et al., 2009). A short consequence text describes, what can be expected medically. This is followed by a recommendation part, stating how to handle the interaction.

The Janus toolbar alert system delivers non-intrusive reminders. This means that the illuminated warnings are optional not forcing the physician to take any action and not disturbing the workflow for the practitioners. Shah et al. (2006) showed that acceptance of drug alerts was improved by minimizing workflow disruptions, designating only high severity alerts to be interruptive to clinicians work. Disadvantages with intrusive alerts are disruption of physicians' workflow and increased tendency to ignore, work around or override these warnings. In a survey by Krall & Sittig (2001) physicians indicated that intrusive or active alerts might be more useful but less easy to use. It was also stated that another important factor for increased compliance and effectiveness of a CDS system is the interface design in relation to the workflow process. Alerts showing up too early or too late in the workflow process might lead to decreased compliance and reliability of the users in the system or even worse, lead to errors and harm for the patient (Krall & Sittig, 2001; Khajouei & Jaspers, 2008).

Studies on the effectiveness of non-intrusive versus intrusive alerts are contradictory. One study (Palen et al., 2006) showed no significant difference between control and intervention groups in the overall rate of compliance to ordering certain laboratory monitoring values when prescribing certain medications. They used non-intrusive alerts in their intervention group. Another study (Tamblyn et al., 2008) compared the effectiveness of on-demand versus computer triggered decision supports regarding dosing information, drug-drug, drug-age, -allergy and -disease interactions. They found that physicians in the computer-triggered group saw more alerts, and made more changes. However, they also ignored more of the alerts shown (87.8%). The on-demand group requested less than 1 % of all alerts provided by the CDS, but ignored only 24.4%. There was no difference in the overall result of existing prescribing problems after intervention between both groups.

We believe that CDS systems need to keep a balance between producing too many alerts and reminders and delivering the message in a straight-forward manner. Too many alerts are likely to be overridden and cause "alert-fatigue", which leads to underestimation of the CDS systems as useful tools in the daily practice (Shah et al. 2006). To avoid too many uncritical alerts classification of the content of knowledge bases regarding clinical significance is of great importance. Numerous studies have shown that compliance to CDS systems and user satisfaction is related to the balance between useful alerting and overalerting (Paterno et al. 2009; Shah et al. 2006). Therefore, we have implemented classification systems in all our knowledge bases. Classification is performed regarding the clinical significance of the content and the level of documentation for the alerts. Colour codes are provided additionally to knowledge base specific classifications (letter or number codes) thereby supporting the prescriber, to identify the urgency of the information retrieved from the knowledge base. The red colour signalises very important messages (e.g. for drug- drug interactions it means: avoid combination) (Böttiger et al., 2009). A yellow colour code indicates information, which should be retrieved and could influence the prescribing (e.g. dose adjustment for a DDI warning). White colour means that information of more theoretical value is available but it has no clinical relevance which has to be considered during prescribing.

Isaac and colleagues (2009) recently showed, that physician's tendency to override alerts was less pronounced for the alerts with high-severity / high risk compared to medium or low severity alerts. Tiered alerting for severity for drug-drug interaction information, like in Sfinx, is one possibility to increase compliance rates for interaction warnings. That was confirmed in a study by Paterno et al (2009), where compliance in the tiered DDI alert group was significantly higher than in the non-tiered group (29% vs. 10%). Additionally, the most severe alerts were accepted to 100% in the tiered group while only 34% in the non-tiered group.

Commercially available DDI databases tend to put more emphasis on covering the whole medical domain rather than differentiating between clinical important and non-important messages. So there is a need for increased specificity to reduce extraneous workload and reduce "alert-fatigue". Luna et al. (2007) described the need to "clean" the content of their commercially purchased knowledge base according to the clinical significance of drug-drug interactions. By creating a classification for DDIs in the system they customized the knowledge base for their organisation.

Spina et al. (2005) investigating the usefulness of different types of alerts in a CDS system in a group of primary care physicians stated that more tailored systems are needed, where DDI warnings on topical drugs should be avoided, when not relevant. Therefore drug formulations should be taken into account in a DDI knowledge base (Böttiger et al., 2009). Also interaction warnings should be suppressed, when drug monitoring is already in place.

Another option can be to suppress warnings on reorders for patients' medications as shown by Abookire et al. (2000). They found that overriding rates for drug allergy warnings increased from 48% to 83% for drugs being reordered for a single patient over a certain time period, suggesting that physicians tend to ignore warnings for the patients permanent medications, since they have handled and considered these alerts already once before. Consequently, tailoring systems focussing more on new ordered medications rather than on drug renewals would be another possibility to increase usefulness of CDS systems.

However, it will not be possible to develop knowledge bases and CDS systems fitting all needs. Personal adjustments seem to be necessary since physicians´ needs and their varying level of knowledge result in different perceptions of any CDS system.

6. Implementation of CDS systems

Healthcare agencies spend significant amounts of money on the development of clinical information systems, though often failing with successful implementation. Designing an effective approach for increasing end-user acceptance and subsequent use of IT- systems is a fundamental challenge. Successful implementation needs comprehensive approaches tailored to clinical settings and target groups taking individual, health care team, and organizational variations into account.

Wears & Berg (2005) described how implementation of any new technology into a clinical workplace triggers both changes in the workplace and in the use of technology, which itself triggers development of the technology (Figure 6). A workplace is described as a field where social behaviour meets technology and both influence each other.

It is also of great importance to consider the different interests in and views on a CDS system from users, administrators and vendors. Ash and colleagues (2003) described the complex interplay of physicians, administrators and IT- staff when implementing a computerized physician order entry (CPOE) system into a hospital setting. They looked at three important parts, which are always influenced by an implementation: the technical, organizational and personal part. Physicians thought the CPOE as technically cumbersome and time-consuming, forcing them to think like computers and click through various screens. They also felt that the CPOE was "forced" on them by hospital administration not taking into account the work situation which they believed was already overburdened. However, on a personal note they felt a need to master the system. The hospital administration thought the system technically to be cost-effective and delivering great statistics. People in the organisation felt pride in being at the forefront of technology. Personally they felt pride in having overcome the clinicians' resistance. The information technology staff perspective on the technical system was the urge and tendency to make the system even more useful, train the users and fulfil and develop the system according to the users wishes. Organisationally they tried to identify the right staff members for the implementation to reach everybody in the hospital. Personally they described enthusiasm for the benefits of the system, but at the same time they felt implementation as difficult and painful but useful in the long run. This study reflects the difficulties of a successful implementation taking into account the various expectations of different "interest groups".

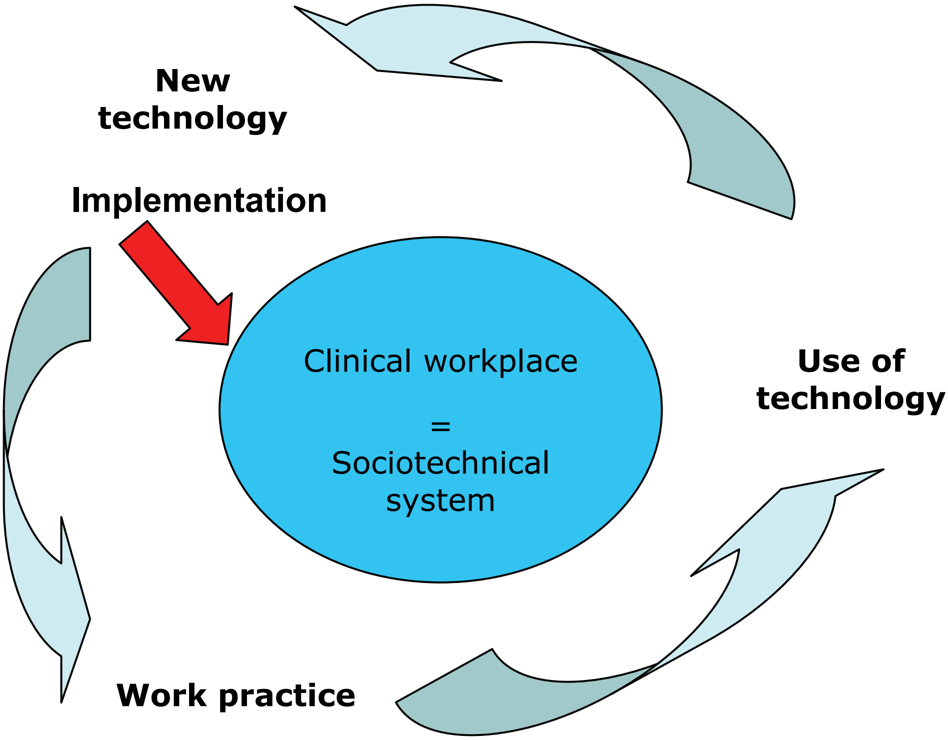


Fig. 6. Influences of technology in a clinical workplace environment. This figure shows that any new technology integrated into a clinical workplace will change work practice, which then will result in use of the technology different as planned from the beginning. This will trigger the development and change of the technical tool implemented (Wears & Berg, 2005)

Ash and colleagues (2003) derived four categories of principles for a successful implementation:

- computer technology
- personal principles
- organizational principles
- environmental issues.

These principles reflect the need to consider multiple issues during implementation and they highlight the relationship between technology, clinical information, people and organizational issues. Callen et al. (2008) described a Contextual Implementation Model (CIM), which is based on data from sites, where physicians use an existing CPOE system. The model acknowledges the complexity of the clinical environment and the requirements of the users. They concluded that implementation should start with a thorough analysis of the context where the CDS system will be implemented into. This analysis should include all three levels namely organisational, departmental and individual. Work practices have to be studied on an individual and department level. Computer literacy and keyboard skills have to be investigated among potential users and work requirements between departments have to be clarified to take the differences between organizations into account. Requirements of the CDS systems on individual and workplace level have to be investigated and differences can be included in the implementation plan so that during implementation one can accommodate the different needs. Targeted training programs can eliminate the problem with different keyboard skills and computer literacy. Analysis of organizational and team cultures will assist with modifying the cultures to increase receptiveness. They concluded that using the CIM model for implementation will facilitate the usage and benefit of any CDS system.

In a systematic review (Gruber et al., 2009), it was stated, that no single implementation strategy has proved to be completely effective. The authors defined a theoretical model for a computerised decision support system including five major steps in the life cycle of any CDS system (= Expanded Systems Life Cycle = ESLC):

- planning
- analysis
- design
- implementation
- maintenance

They identified risk zones for each phase and corresponding risk factors. Their analysis revealed that the highest number of failure and success were in the implementation zone focusing on preimplementation and "go-live" of the system. They also identified that training and education, attention to training, policy, process changes, and training to clinical content are key factors influencing the success or failure of a CDS system.

However, more research is needed to avoid costly errors in implementation. Studies focussing on barriers and incentives for changes should be performed focussing on various levels (namely the innovation itself, the professional, the patient, the social context, the organisational context, and the economic and political context) as suggested (Grol & Wensing, 2004).

7. Evaluation

Rigorously designed evaluations and research on the effectiveness of decision support systems are needed to assess their value in clinical practice and to identify areas for improvement in design and implementation. Kirkpatrick described four levels of evaluation in which the complexity of the behavioural change increases as evaluation strategies ascend to each higher level (Kirkpatrick, 1967). The four levels measure

- reaction to information
- learning
- behaviour
- results

Studies assessing effects of CDS systems on patient outcome are urgently needed. They are difficult to perform due to the length of time needed for the evaluation, the lack of reliable objective measures, and the number of potential confounding factors.

The selection of methodology to investigate an implementation of decision support systems is no different from choosing methods in any other type of research. A variety of study designs can be used to evaluate if decision support systems influence prescribing behaviour and patient outcomes. These studies include quasi-experimental designs (uncontrolled or controlled before-and-after studies and interrupted time series) and randomized controlled trials (RCTs) (Grimshaw et al., 2000). The RCT has the highest degree of evidence as non-randomized designs might introduce selection bias by including in the intervention group doctors or clinics that favour the particular intervention (Grimshaw et al., 2000, Stephenson & Imrie, 1998). The control group design considers other factors influencing the prescribing pattern such as seasonal variations in disease patterns, the introduction of new drugs and changes in treatment policies, the marketing activities of pharmaceutical companies and changes in regulatory policies (Grimshaw et al., 2000). However, due to ethical, practical and methodological reasons, they are seldom possible to apply when evaluating the impact of decision support systems. Therefore, well-designed quasi-experimental studies may be the method of choice.

Alternative research strategies include qualitative research methods to provide a deeper understanding of the subjective aspects of the interaction between healthcare professionals, patients and the electronic tools. The common feature of qualitative studies is that they do not primarily seek to provide quantified answers to research questions. The goal of qualitative research is the development of concepts which can help us to understand social phenomena in natural rather than experimental settings, giving due emphasis to the meanings, experiences, and views of all the participants (Pope & Mays, 1995). Examples of qualitative methods include in-depth interviews, focus group discussions, observations and various consensus methods.

Development and evaluation of a complex system, such as a CDS system and implementing it into the health care organisation require a multiple research approach i.e. method triangulation. The evaluation of the pilot study of the Janus decision support system had primarily a qualitative approach with focus on user satisfaction. Semi-structured qualitative interviews were performed with all users before, during and after the pilot study. By concentrating the evaluation on user satisfaction we gained data both on the technical failures as well as the physicians' attitudes to medical content and usefulness of the system and acceptance in clinical work. The evaluation and implementation were carried out by a multidisciplinary team within a small scale user clinic in order to be able to easily detect technical and practical obstacles (i.e. integration bugs) and even more serious potential quality problems of the pharmacological sources (i.e. pregnancy and breast-feeding alerts in the Swedish PDR) (Eliasson et al., 2006). Data and support were handled in a rapid way to be able to give direct feed-back to the user. Our experiences confirm that evaluations of small-scale pilot studies for proof of concept are important tools in the design of an optimal intervention that improves health care quality so that resources are used in an optimal way as stated by Harvey & Wensing (2003).

The results of the pilot study even helped us to identify factors, which have major impact on usefulness of the CDS system and user satisfaction and led to a two-part theoretical model for implementation and evaluation (Eliasson et al, 2006). This model considers both system-dependent and system-independent factors (Figure 7). The first part includes system-dependent factors, such as medical content, user friendliness and user support. The second system-independent part includes personal attitudes of the prescribers´ towards computer use as well as the attitude of the organisation towards implementing a CDS system.

Stockholm County Council conducts regular evaluations after pilot studies which we see as a cornerstone for development of successful electronic tools. The effectiveness of Janus toolbar and the frequency of its use, and users' characteristics are measured by questionnaires. Simultaneously, interviews are carried out to explore doctors' and other prescribers' experiences and perceptions of Janus toolbar. Those evaluations were used to

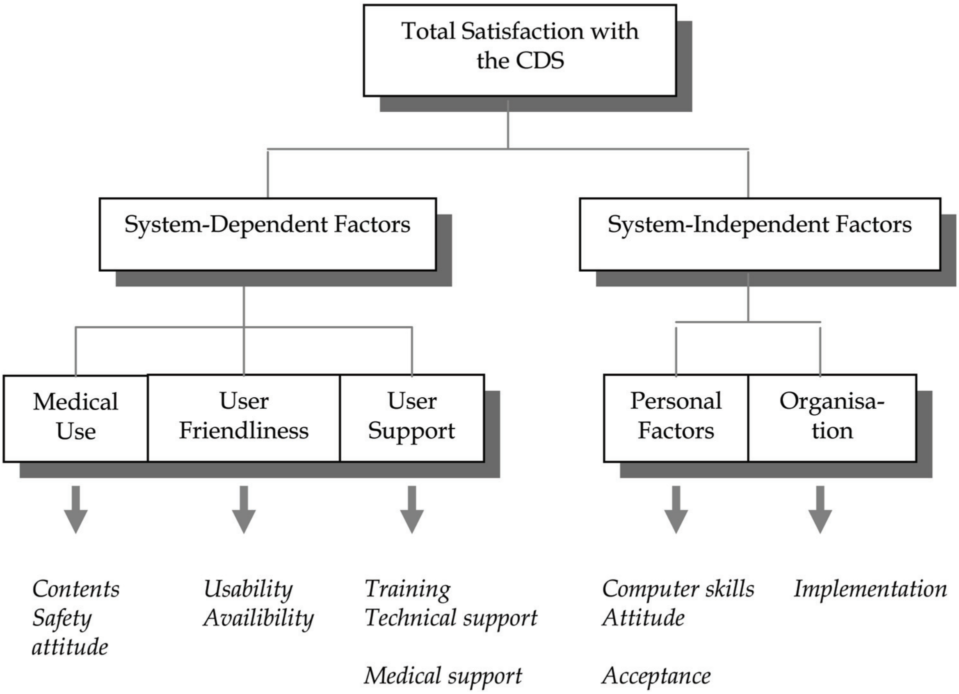


Fig. 7. A two-part theoretical model for evaluation of the CDS system taking system dependent (e.g. medical content, usability, support) and system independent factors (e.g. personal attitudes, organizational aspects) into account.

decide about the development of a new knowledge base for drug-drug interactions Sfinx, which is described above. The regular follow-ups over years showed results similar to the actual literature especially in terms of satisfaction, acceptance and intention to use (Krash, 2004, Ahearn & Kerr, 2003; Magnus et al., 2002). Physicians generally overrode the interaction warnings and expressed irritation on the irrelevant alerts, which often led to ignore them. Furthermore, physicians were dissatisfied with the usability, information and training of how to use the tool, and complained about technical barriers. Although physicians did not seem to use the tool in every day practice they underlined the clinical value and needs of it, i.e. being reminded of unknown /known drug-drug interactions and getting recommendations about how to avoid them. Prescribers were aware of the fact that the decision support system contributes to safer and more effective treatment of the patients. They were clear about their needs for the system and had good intentions to use it. However, even after thorough analysis of physicians' needs, we could observe that the system was not fully used after its implementation.

Some contributing factors are changes in expectations and intentions of the users from the initial discussion, to later implementation and the actual use at the work place when the database is integrated into daily work flow. Another influencing factor is that CDS systems integrated into daily practice suddenly offer more complete knowledge about patients' medications for the physicians, demanding new decisions and work tasks that GPs were not aware of. Physicians have different views on their responsibilities for diagnosis, drug treatment and follow-up of a patient resulting in different actions and variations of handling the information provided. Recently we have highlighted that there is a need for common and understandable rules on prescribing physicians´ responsibility in handling the total patients′ drug lists. These lists are made available to all prescribers through a newly implemented IT-tool (Rahmner et al., 2009). We can conclude that work flow, working environment and processes influence physicians′ behaviour to a greater extent than expected. Consequently, we still do not know how to design optimal CDS systems which affect and influence physicians′ behaviour in drug prescribing. The challenge for the future development and implementation of a CDS system into health care is to find a method to achieve and maintain expected changes in prescribing behaviour.

8. Summary

Knowledge bases provide the contents for any clinical decision support system. In this review we characterize the life cycle of a knowledge database to be used in drug prescribing. The various phases and the important issues in each phase are summarized in table 1. Knowledge bases need to fulfil and be tailored to the needs of the users. The focus of the content should be on practical use in a clinical environment, rather than covering the whole scientific area of a medical speciality. Standards are needed to be able to use knowledge bases across different electronic health care systems and countries, since clinical expertise is often the bottle neck for any development.

Integration of knowledge bases into CDS systems implemented into electronic health record system optimises their effectiveness by delivering patient specific reminders and alerts. The linkage between knowledge bases and CDS systems needs to be quality assured. Knowledge bases and CDS systems need to be surveyed through a management and administration system handling incidents and errors due to system or its content. Though many studies have shown the positive influence of CDS systems on physicians′ performance, there is still lack of understanding, when CDS systems improve performance. Outcome studies on patient care are lacking. Implementation of CDS systems has to be accompanied by staff education and training to assure acceptance and effectiveness even throughout the maintenance phase. More studies are needed with focus on actual improvement of patient safety and care instead of investigating physicians change in prescribing drugs.

With that in mind knowledge bases and CDS systems will prove to be helpful tools in the daily decision making process of any busy clinician when instituting and evaluating the drug therapy of a patient.

9. References

Abookire SA, Teich JM, Sandige H et al. (2000) Improving allergy alerting in a computerized physician order entry system. *Proc AMIA Symp.*, 2000., 2-6.

Ahearn MD, Kerr SJ. (2003). General practitioners′ perceptions of the pharmaceutical decision-support tools in their prescribing software. *MJA.*, 179.,34-37.

Ammenwerth E, Schnell-Inderst P, Machan C et al. (2008). The effect of electronic prescribing on medication errors and adverse drug events: a systematic review, *JAM Med Inform Assoc.*, 15., 585-600.

Ash JS, Gorma PN, Lavelle M et al. (2000) Multiple perspectives on physician order entry *Proc Amia Symp*. 27-31.

Life cycle phase of a knowledge base	Important issues for each life cycle phase
Development	Thorough analysis of physicians needs Standardisation of data structure, implementation of classification system, optimal linkage to drug registries
Quality assurance	Control of quality in key fields for linkage Introduction of semi-automated and manual processes for data auditing
Medical management & maintenance	Well documented and standardized procedures for knowledge base maintenance Root Cause Analysis for analysis of mistakes and follow ups of the planned actions
Providing knowledge bases at point of care	Integration into electronic health records for patient specific alerts Tailored systems with fast data access to avoid overalerting and increase acceptance
Implementation	Consider interests of users, organizations and vendors within the implementation plan Education, personal training, attention to process changes are key factors
Evaluation	Evaluation of small scale projects as important tools in the design of optimal interventions Regular evaluations necessary to secure optimal use of knowledge base or CDS system

Table 1. Summary of important messages for each step in the life cycle of a knowledge base.

Ash JS, Fournier L, Stavri PZ et al. (2003) Principles for a successful computerized physician order entry system implementation. *Proc Amia Symp.*, 36-40.

Ash SA, Berg M, Coiera E. (2004). Some unintended consequences of information technology in health care: The nature of patient care information system-related Errors, *J AM Med Inform Assoc.*, 11., 104-112.

Aspden P, Wolcott J, Bootman JL et al. (2006). Committee on identifying and preventing medication erros (2006). Preventing medication errors. Washington DC; *National Academies Press.*

Baorto D, Li L & Cimino JJ. (2009) Practical experience with the maintenance and auditing of a large medical ontology, *J Biomed Inform.*, 42., 494-503.

Bastholm Rahmner P. (2009) Doctors and drugs – How Swedish emergency and family physicians understand drug prescribing (Thesis). Medical Management Centre, Department of Learning, Informatics, Management and Ethics. Karolinska Institutet, Stockholm Sweden

Bastholm Rahmner P, Andersen-Karlsson E, Arnhjort T et al. (2004). Physicians´perceptions of possibilities and obstacles prior to implementing a computerised drug prescribing support system, *Int J Health Care Qual Assur Inc Leadersh Health Serv.*, 17., 4-5., 173-179.

Bates DW, Gawanda AA (2003). Improving safety with information technology, *N Engl J Med.*, 348., 25., 2526-2534.

Bates DW, Cohen MS, Leape LL et al.(2001). Reducing the frequency of errors in medicine using information technology, *J AM Med Inform Assoc.*, 8., 299-308.

Bergk V, Haefeli WE, Gasse C et al. (2005). Information deficits in the summary of product characteristics preclude an optimal management of drug interactions: a comparison with evidence from the literature, *Eur J Clin Pharmacol.*, 61., 5-6., 327-335.

Blumenthal D. (2009). Stimulating the adoption of health information technology, *N Engl J Med,* 360., 15., 1477-1479.

Böttiger Y, Laine K, Andersson ML et al. (2009), SFINX – a drug-drug interaction database designed for clinical decision support systems. *Eur J Clin Pharmacol.*, 65., 6., 627-633.

Callen JL, Braithwaite J, Westbrook JI. (2008). Contextual implementation model: a framework for assissting clinical information system implementations. *J Am Med Infrom Assoc.*, 15., 2., 255-262.

Dawes M, Sampson U. (2003). Knowledge management in clinical practice: a systematic review of information seeking behaviour in physicians, *Int J Med Inform.*, 71., 9-15.

Eliasson M, Bastholm, Forsberg P et al. (2006). Janus computerised prescribing system provides pharmacological knowledge at point of care – design, development and proof of concept. *Eur J Clin Pharmacol.*, 62., 251-258.

Eslami S, de Keizer NF, Abu-Hanna A. (2008), The impact of computerized physician medication order entry in hospital patients – A systematic review, *Int J Med Inform.*, 77., 365-376.

EU directive; 2007/47/EC (2007), http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:247:0021:0055:EN:PDF; accessed 2009-07-30.

Garg AX, Adhikari NKJ, McDonald H. (2006). Effects of computerised clinical decision support systems on practitioner performance and patient outcomes, *JAMA*, 293., 10., 1223-1238.

Gartner (2009), eHealth for a Healthier Europe! www.regeringen.se/content/1/c6/12/98/02/5b63bacb.pdf accessed 2009-07-27.

Godman B, Wettermark B, Hoffman M et al. (2009). Multifaceted national and regional drug reforms and initiatives in ambulatory care in Sweden: global relevance. *Expert Rev. Pharmacoeconomics Outcomes Res.*, 9., 1., 65-83.

Grimshaw J, Campbell M, Eccles M, Steen N. (2000). Experimental and quasi-experimental designs for evaluating guideline implementation strategies. *Family Practice*;17:S11-S18.

Grol R, Grimshaw J. (2003). From best evidence to best practice: effective implementation of change in patients´care. *Lancet*, 362., 9391., 1225-1230.

Grol R, Wensing M. (2004) What drives changes? Barriers to and incentives for achieving evidence-based practice. Med J Aust, 180., S57-60.

Gruber D, Cummings GG, Leblanc L. (2009) Factors influencing outcomes of clinical information systems implementation: A systematic review. *Comput Inform Nurs*, 27., 3., 151-163. quiz 164-165.

Gustafsson LL, Widäng K, Hoffmann M et al. (2003) Computerized decision support in drug prescribing II. A national database to provide up-to-date and unbiased information. In Swedish. *Lakartidningen*, 100., 15., 1338-1340.

Han YY, Carcillo JA, Venkataraman T et al. (2005) Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics* 116., 6., 1506-1512.

Harvey G, Wensing M. (2003). Methods for evaluation of small-scale quality improvement projects. *Qual Saf Health Care*. 12.,210-214.

Iedema RA, Jorm C, Long D et al. (2005) Turning the medical gaze in upon itself: Root cause analysis and the investigation of clinical errors. *Soc Sci Med*., 62., 7., 1605-1615.

Isaac T, Weissman JS, Davis RB. (2009) Overrides of medication alerts in ambulatory care. *Arch Intern Med*., 169.,3.,305-311.

Janusinfo; http://www.janusinfo.se/imcms/servlet/StartDoc; accessed 2009-07-30

Khajouei R, Jaspers WM. (2008). CPOE system and design aspects and their qualitative effect on usability. *Stud Health Technol Inform*. 136., 309-314.

Kirkpatrick DI. (1967) Evaluation of training. In Craig L., Bittel I., eds. Training and development handbook. New York: McGraw-Hill.

Ko Y, Abarca J, Malone DC. (2007). Practitioners´ view on computerized drug-drug interaction alerts in the VA system. *J AM Med Inform Assoc*., 14., 56-64.

Krahenbuhl-Melcher A, Schlienger R, Lampert M et al. (2007). Drug-related problems in hospitals. *Drug Safety*, 30., 5., 379-407.

Krall MA, Sittig DF. (2001). Subjective assessment of usefulness and appropriate presentation mode of alerts and reminders in the outpatient setting. *Proc Amia Symp*., 334-338.

Krash Bt (2004). Beyond usability: designing effective technology implementation systems to promote patient safety. *Qual Saf Health Care,* 13.,388-394.

Kuperman GJ, Reichley RM & Bailey TC. (2006) Using commercial knowledge bases for clinical decision support: opportunities, hurdles, and recommendations *JAMIA*., 13., 4., 369-371.

Lucas H. (2008), Information and communication technology for future health systems in developing countries. *Soc Sci Med*., 66., 10., 2122-2132.

Luna D, Otero V, Canosa D et al. (2007) Analysis and redesign of a knowledge database for a drug-drug interactions alert system. *Stud Health Technol Inform*, 129., 2., 885-889.

Magnus D, Rodgers S et al. (2002). GPs' views on computerized drug interaction alerts: questionnaire survey. *J Clin Pharm Ther*., 27.,377-382.

Norby U, Eiermann B, Tornqvist E, et al. (2006), Drugs and birth defects – a Swedish information source on the Internet. Paed Per Drug Ther., 7., 2., 89-112.

Palen TE, Raebel M, Lyons E. (2006). Evaluation of laboratory monitoring alerts within a computerized physician order entry system for medication orders. *Am J Manag Care*, 12., 7.,389-395.

Paterno MD, Maviglia SM, Gorman PN et al. (2009) Tiering drug-drug interaction alerts by severity increases compliance rates. *J Am Med Inform Assoc*., 16.,40-46.

Pope C, Mays N. (1995). Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. BMJ ;311:42-5

Rahmner P, Gustafsson LL, Holmstrom I. (2009). Who´s job is it anyway-Swedish general practitioners´perception of their responsibility for the patient´s drug list. *Ann Fam Med*; in press.

Rahmner PB, Gustafsson LL, Larsson J et al. (2009) Variations in understanding the drug-prescribing process: a qualitative study among Swedish GPs. *Fam Pract.*, 26., 2., 121-127.

Revere D, Turner AM, Madhavn A et al. (2007). Understanding the information needs of public health practitioners: A literature review to inform design of an interactive digital knowledge management system. *J Biomed Inform.*, 40.,410-421.

Schiff GD, Rucker D (1998). Computerized prescribing. Building the electronic infrastructure for better medication usage. *JAMA*, 279., 1024-1029.

Seidling HM, Al Barmawi A, Kaltschmidt J et al. (2007). Detection and prevention of prescriptions with excessive doses in electronic prescribing systems. *Eur J Clin Pharmacol.*, 63., 12., 1185-1192.

Shah NR, Seger AC, Seger DL et al. (2006) Improving acceptance of computerized prescribing alerts in ambulatory care. *J Am Med Inform Assoc.*, 13., 1., 5-11.

Shemeikka T, Gustafsson LL, Korkamz S. (2008) Following a Lex Maria case: safe computer support systems for drug prescribing required. In Swedish. *Läkartidningen* 105., 3., 3177-3178.

Sjöborg B, Bäckström T, Arvidsson LB et al. (2007). Design and implementation of a point-of-care computerized system for drug therapy in Stockholm metropolitan health region-Bridging the gap between knowledge and practice, *Int J Med Info*, 76., 7., 497-506.

Spina JR, Glassman PA, Belperio P et al. (2005) Clinical relevance of automated drug alerts from the perspective of medical providers. *Am J Med Qual*, 20., 1., 7-14.

Stephenson J., Imrie J. (1998). Why do we need randomised controlled trials to assess behavioural interventions? *BMJ.*, 316., 611-613.

Sweidan M, Reeve JF, Brian JA (2009). Quality of drug interaction alerts in prescribing and dispensing software. *Med J Aust* 190., 5., 251-254.

Tamblyn R, Huang A, Taylor L et al. (2008) A randomized trial of the effectiveness of on-demand versus computer-triggered drug decision support in primary care. *J Am Med Inform Assoc.*, 15., 430-438.

Van Rosse F, Maat B, Rademaker CMA et al. (2009). The effect of computerized physician order entry on medication prescription errors and clinical outcome in pediatric and intensive care: A systematic review, *Pediatrics*, 123., 1184-1190.

Weingart SN, Massagli M, Cyrulik A et al. (2009). Assessing the value of electronic prescribing in ambulatory care: A focus group study. *Int J Med Inform.* 78., 9., 571-578.

Wears RL, Berg M. (2005). Computer technology and clinical work: still waiting for Godot. *Jama* 293., 10., 1261-1263.

Wolfstadt JI, Gurwitz JH, Field TS. (2008). The effect of computerized physician order entry with clinical decision support on the rates of adverse drug events: A systematic review. *J Gen Intern Med.*, 23., 4., 451-458. *Jama.*, 299., 6., 685-687.

Wu AW, Lipshutz AKM, Pronovost PJ. (2008). Effectiveness and efficiency of root cause analysis in medicine. *Jama* 299., 6., 685-687.

Develop a Spatial Decision Support System based on Service-Oriented Architecture

Chuanrong Zhang
Department of Geography and Center for Environmental Sciences and Engineering, University of Connecticut, USA

1. Introduction

A Spatial Decision Support System refers to a computer system that assists decision-makers to generate and evaluate alternative solutions to semi-structured spatial problems through integrating analytical models, spatial data and traditional geoprocessing software so that individuals or groups can make feasible decisions (Armstrong 1993; Densham 1991; Malczewski 1996). It can allow easier decision-making by providing an easy access to geospatial data and analytical models. Many Spatial Decision Support Systems (SDSSs) have been developed for environmental and natural resources decision-making in recent years (e.g., Carver 1999; Jankowski et al, 1997, 2001, 2006; Van Der Perk et al., 2001). However, an important limitation of the SDSS applications is that they are not interoperable. The geospatial data and geoprocessing resources distributed by them cannot be shared and interoperated. While there is an increase in the number of SDSS applications over the past two decades, most of them did not take advantage of the Internet's distributed nature by sharing spatial data and geoprocessing software (Ostländer 2004; Rinner 2003). Several issues prevent the further development of SDSS applications.

One issue is that most SDSSs were developed independently of one another and they are typically standalone systems incapable of sharing and reusing existing data and processing functions. They have their own proprietary system designs, database storage structures, and process models. Thus, it is difficult to communicate and exchange spatial information among these systems, and decision-makers usually cannot integrate data and geoprocessing resources from these systems. Instead of direct integration, they have to spend a lot of money and time on taking a complex procedure to convert the heterogeneous information together. The integration of data and modelling software from disparate sources was beyond the technological capabilities of many potential users (Sengupta & Bennett 2003).

The second issue is the duplication problem caused by current SDSSs' incapability of sharing and reusing existing data and geoprocessing. Because of the lack of interoperability, accessibility and availability of data and information, redundant efforts are commonplace in the development of SDSS applications. While there is a massive increase in the number of SDSS applications over the past two decades, it is often the case that these applications were built with little knowledge of other applications with which they could share information. As a result, many agencies and companies are trying to maintain the databases and processing functions that coexist but are not integrated. There are many duplicate data and geoprocessing that occur in separate departments and applications. For example, to make a trip planning decision, a county transit agency may need replicate the street network that is managed by the state highway department and the utility data that is maintained by a local natural resource government agency. The costs attributed to the redundant and duplicated efforts are huge. There is a need to flexibly manage systems and processes and eradicate duplication efforts inside and between different departments and organizations. Further, duplication may cause inconsistency of replicated data.

Thirdly, it is time-consuming to obtain geospatial data when implementing SDSSs. Geospatial data capable of supporting desired analyses often did not exist or was too expensive to acquire (Sengupta & Bennett 2003). Although many geographical databases have been developed, geospatial data sharing and acquisition is still a big problem because of the heterogeneity of existing systems in terms of data modeling concepts, data encoding techniques, storage structures, and other related reasons (Devogele et al. 1998). In order to exchange information and share computational resources among heterogenous systems, conversion tools have to be developed to transfer data from one format into another. However, data conversion is costly and time consuming and may lead to inefficiencies and ineffectiveness in many time-critical decision-making applications, such as real time traffic management, which needs real-time access to diverse data to make quick decisions and take instantaneous actions (Zhang & Li 2005). For the development of a SDSS, the issue of how to aquire data rapidly from different sources becomes important. Decision supports demand that the right information be available at the right time for making the right decision.

Fourthly, while it is recognized that maintaining the most up-to-date geospatial information is important to aid in right decision making, it is difficult to keep consistent updated data in SDSS applications. To facilitate data sharing in SDSS applications, when data are updated at one source the change should be automatically updated in other closely associated applications. However, in current SDSS applications, automatically updating databases from disparate sources produced by different agencies cannot be realized with conventional database management.

Finally, it is costly and time consuming to develop a new SDSS from scratch. Although many small companies and government agencies want to develop SDSSs to make decision-making easier, they cannot afford it because of limited or declining resources. It is often more cost effective to reuse existing data and software via interoperable SDSSs than to develop new databases and custom software. A solution that builds on existing data and geoprocessing rather than starting a new one from scratch is needed. There is an increasing demand for the development of interoperable SDSSs to reuse geographical data and geoprocessing.

The emergence of OGC web services provides a way to overcome the heterogeneity problem of spatial databases and geoprocessing (Peng & Zhang 2004; Zhang et al. 2003a; Zhang & Li 2005). Users can "wrap" existing heterogeneous data into a web service and enable many potential clients to use the service. OGC’s web services represent an evolutionary, standards-based framework that may enable seamless integration of a variety of online geospatial data and geoprocessing (OGC Interoperability Program White Paper, 2001). Power (2003) suggested that the next generation of decision support systems should be primarily service-based. Rinner (2003) and Sugumaran and Sugumaran (2005) proposed web services and the Service-Oriented Architecture (SOA) to be the next generation infrastructure supporting decision-making. Realizing the great potential of web services, researchers began to move towards developing SDSS applications using OGC web services (Bernard et al. 2003; Keßler et al. 2005). In spite of this growing interest, little has been published about how to design and implement a workable interoperable SDSS using OGC web services.

The main objective of this chapter is to propose a framework of web services-based SDSSs for decision-making. The framework enables decision-makers to reuse and integrate geospatial data and geoprocessing from heterogeneous sources across the Internet. Based on the proposed framework, a prototype has been implemented to demonstrate how OGC web services and the SOA overcome the aforementioned issues and contribute to the development of interoperable SDSSs. The implemented prototype addressed how to find and integrate existing heterogeneous data from diverse sources for decision-making.

2. Proposed framework

2.1 Framework structure

A framework for web services-based decision-making system is proposed as shown in Figure 1. The main objectives of the framework are: (a) to enable geospatial data and geoprocessing sharing over the web; (b) to maximize productivity and efficiency with geospatial data and geoprocessing sharing; (c) to overcome data duplication and data maintenance problems; and (d) to make it easy to integrate with other SDSS applications. The framework is based on independent OGC web services and the SOA. It is essentially a collection of OGC web services, which communicate with each other by simple data passing or coordinating some activities.

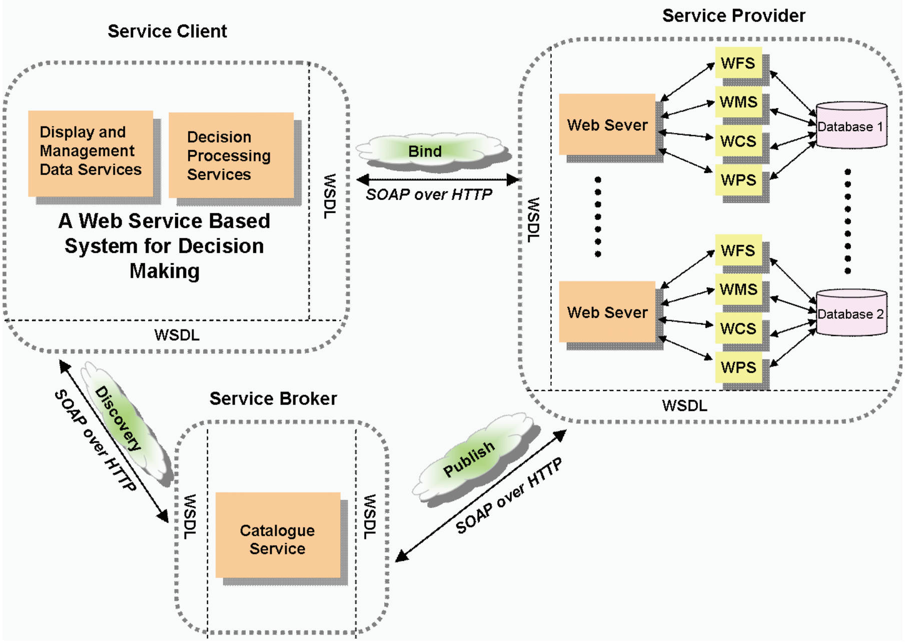


Fig. 1. A framework proposed for web services-based decision-making system.

The framework is composed of three elements: service provider, service broker and service client. Service provider supplies heterogeneous geoprocessing and geospatial data from disparate sources via OGC Web Processing Services (WPS) and data services such as Web Feature Services (WFS), Web Map Services (WMS) and Web Coverage Services (WCS). The OGC data services provide a basis to share spatial data with different data models and from different sources without data conversion. The decision geoprocessing services can be chained to build specific spatial decision support services. Service client helps decision-makers display and manage data services and access decision processing services to generate and evaluate alternative solutions to semi-structured spatial problems. Service client offers an easy interface to allow users to search and integrate geospatial data and geoprocessing from various service providers. Service broker provides a registry for available services. Service broker uses OGC Catalogue Services (CS) to register and manage the data and processing services and allow users to search for these services. Service clients search contents of catalogue services to find the datasets and services of interest, and they also can combine the data or processing services found through catalogue web services. The web services are connected via Web Service Description Language (WSDL) among the service provider, the service broker and the service client. The Simple Object Access Protocol (SOAP) binding over HTTP is employed for communication between web services via the Internet. SOAP essentially provides the envelope for sending web services messages. In general, the proposed framework ensures interoperability through open web services, which offer basic conditions for interoperability by using a standard exchange mechanism between diverse spatial data sources connected over the web.

The main benefit of the proposed framework is its capability of sharing and reusing existing data and geoprocessing from heterogeneous sources across the web. In the framework, different software vendors can be the sources of the data or geoprocessing environments. Decision-makers need not have prior knowledge of the original formats of the data and the original programming language of the geoprocessing. They can transparently exchange and integrate the heterogeneous data and geoprocessing over the web. Thus, the proposed framework may enable many existing proprietary legacy databases or geoprocessing to be reused and shared across application, enterprise, and community boundaries. The SOA employed by the proposed framework allows the system decomposed into several services that enable receiving a system of relatively independent distributed applications. With this loosely coupled nature, the proposed framework permits smart and fast modification of the system's obsolete data and geoprocessing, and quick integration of new data and geoprocessing into the application system, thus enhancing the flexibility in the reuse of geospatial data and geoprocessing. Therefore, it reduces investments in different SDSS applications by avoiding overlapping or repeatedly creating the same data, and leverages an organization's existing investment in data and applications by taking advantage of current resources. Since the solution is based on open standards, it has the potential to be a way of getting to the interoperability. The following sections introduce major concepts in the proposed framework - OGC web services and SOA.

2.2 OGC web services

Web services are described as reusable software components that interact in a loosely coupled environment, and they are designed to interoperate in a loosely-coupled manner. A web service can be discovered and used by other web services, applications, clients, or agents. Web services may be combined or chained to create new services. And they may be recombined, swapped or substituted, or replaced at runtime. Due to the fact that web services are based on XML standards, they are currently being used by enterprises for interoperability. Web services provide the interoperable capability of cross-platform and cross-language in distributed net environments (Anderson & Moreno 2003). As a result, companies may convert their applications to web services to make disparate applications interact.

OGC web services deal with geographic information on the Internet. OGC web services can be grouped into three different categories: data services, processing services and registry or catalog services (Figure 2). Data services are tightly coupled with specific spatial data sets and offer access to customized portions of the spatial data. Examples of data services include WFS (Web Feature Services), WMS (Web Map Services) and WCS (Web Coverage Services). WFS (OGC document 04-094 2005) allow a client to retrieve, query, and manipulate feature-level geospatial data encoded in GML (Geography Markup Language) from multiple sources. They are written in XML (Extensible Markup Language) and use an open-source standard GML (OGC document 02-023r4 2003) to represent features. GML data are stored in a universal format-- text format. Due to the universal format GML data can be easily integrate into other data across a variety of platforms and devices (Zhang et al. 2003a). As a standard data exchange format GML reduces the costly conversion processes among different format databases and can deliver vector data over the Internet (Zhang et al. 2003a). In proprietary systems, such as ESRI's ArcGIS, support of GML in WFS is through a DataStore. The DataStore can transform a proprietary data format such as Shapefile into the GML feature representation. There are two types of WFS - basic and transaction. "Basic" WFS only implement operations for describing and retrieving features over the web, and "transaction" WFS also implement operations for locking and modifying (creating, updating, and deleting) features across the web. One important property of WFS is that they can serve multiple feature types. Different features from different data stores can be integrated with WFS and clients do not realize that the features are retrieved from several sources (Zhang & Li 2005). WMS are capable of creating and displaying maps that come simultaneously from multiple heterogeneous sources in a standard image format such as SVG, PNG, GIF or JPEG (OGC document 04-024 2004). WMS provide three operation protocols: GetCapabilities, GetMap, and GetFeatureInfo. GetCapabilities allows a client to instruct a server to expose its mapping content and processing capabilities and return service level metadata. GetMap enables a client to instruct multiple servers to independently craft "map layers" that have identical spatial reference systems, sizes, scales, and pixel geometries. GetFeatureInfo enables a user to click on a pixel to inquire about the schema and metadata values of the feature(s) represented there. Unlike WFS which enable users to access specific feature DataStores in GML, WMS permit users to display spatial data and produce images of the data rather than to access specific data holdings. WCS provide access to potentially detailed and rich sets of geospatial information in forms that are useful for client-side rendering, multi-valued coverage, and input into scientific models and other clients (OGC document 03-065r6 2003). Rather than static maps (server-rendered pictures), WCS deliver coverage data (e.g. multi-spectral imagery or elevation data) in response to queries from HTTP clients. A WCS client can issue a GetCoverage request to obtain these numeric values for further processing or rendering on behalf of the user. WCS offer one or more layers, just like WMS, but do not render them for the user and therefore do not offer styles.

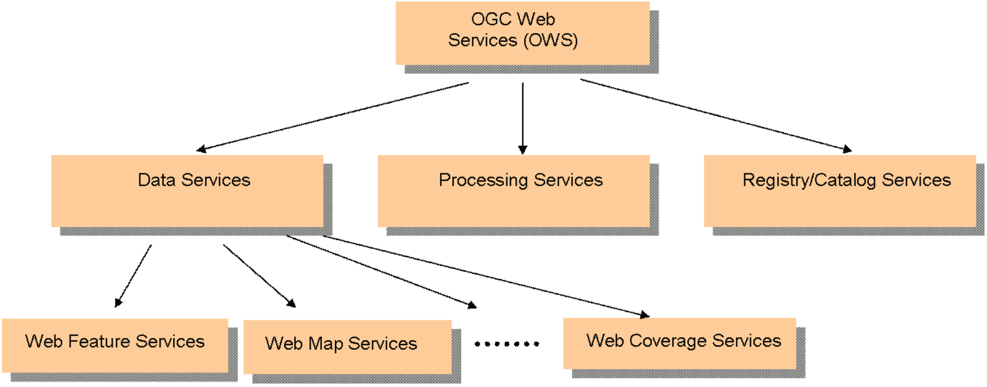


Fig. 2. OGC web services.

OGC WPS provide decision-makers access across the web to decision models that operate on spatially referenced data and help decision-makers achieve a higher effectiveness of decision-making while solving a complex and ill-defined spatial decision problem. WPS have mechanisms to identify the geospatial data required by the decision models, execute the calculation of the models, and manage the output from the calculation so that they can be accessed by the decision-makers. Both vector and raster data can be processed by WPS. The data can include image data formats such as GeoTIFF, or data exchange standards such as GML or Geolinked Data Access Service (GDAS). Three operations are required for a WPS interface (OGC document 05-007r4 2005): GetCapabilities operation allows decision-makers to request and receive back service metadata (or Capabilities) documents that describe the abilities of the implemented decision models; DescribeProcess operation grants decision-makers to request and receive back detailed information about one or more decision model(s) that can be executed, including the input parameters and formats, and the outputs; Execute operation allows decision-makers to run a specified decision model implemented by the processing services, using provided input parameter values and returning the outputs produced. To achieve interoperability, service providers must specify the specific implemented decision models in a separate document called an Application Profile.

OGC Catalog Services allow decision-makers to classify, maintain, register, describe, search and access information about web services. The Catalog Services provide catalogues for the above introduced OGC data services and processing services and support the ability to publish and search collections of descriptive information (metadata) for data, services, and related information objects (OGC document 04-021r2 2004). The essential purpose of a catalogue service is to enable decision-makers to locate, access, and make use of resources in an open, distributed system by providing facilities for retrieving, storing, and managing many kinds of resource descriptions. A catalogue service has two main functions -- discovery and publication. Discovery means that decision-makers seek to find resources of interest through simple browsing or by sophisticated query-driven discovery that specifies simple or advanced search criteria. The catalogue performs the search and returns a result set which contains all registry objects that satisfy the search criteria. OGC Catalog Services support distributed search. When decision-makers perform a distributed search, the request message is forwarded to one or more affiliated catalogues to enlarge the total search space.

In general, OGC web services conceal complexity of heterogeneous geospatial data and geoprocessing. They enable SDSS developers to integrate data and geoprocessing models into their applications without having to host the spatial data locally or build the models completely by themselves, and they also make possible to utilize many valuable existing legacy databases and geoprocessing to develop SDSSs.

2.3 Service oriented architecture

The concept of web service is based on the Service-Oriented Architecture (SOA). A pure architectural definition of SOA might be "an application architecture within which all functions are defined as independent services with well-defined invokable interfaces which can be called in defined sequences to form business processes" (Kishore et al. 2003). In another word, the SOA can be defined as a system in which resources are made available to other participants in the network as independent services that are accessed in a standardized way. The SOA provides for more flexible loose coupling of resources than traditional system architectures. With SOA, applications can access web services through the web without concern how each service is implemented. The interaction among various services in the SOA relies heavily on the standardized interface described by the WSDL and the web communication message protocol SOAP. WSDL defines all of the information necessary to invoke a web service such as what a web service can do, where it resides, and how to invoke it (W3C 2001). The WSDL provides a way for service providers to describe the basic format of web service requests over different protocols or encodings and thus helps improve interoperability between applications (OGC 04-060r1 2004). SOAP messages are encoded using XML and it is a simple XML based protocol for accessing a web service (W3C 2003) and is used for communication between applications running on different operating systems, with different technologies and programming languages.

In the SOA, three components - service provider, service client and service broker work together. A Service Provider publishes services to a Service Broker. A Service Client finds required services using a Service Broker and bind to them. The binding from the Service Client to the Service Provider should loosely couple the service. This means that the service requester has no knowledge of the technical details of the provider's implementation, such as the programming language, deployment platform, and so forth. The SOA aims to improve the ability of organizations to quickly create and reconfigure a SDSS to support new and rapidly changing situations. The key idea is to move away from monolithic systems, towards systems which are designed as a number of interoperable components, for enhanced flexibility and reuse. With SOA, each SDSS application can be built without a priori dependencies on other applications. The services in an application can be added, modified, or replaced without impacting other applications. This results in very agile systems, which can be flexibly adapted to changing requirements and technologies.

3. Prototype implementation

3.1 Prototype

The Lunan Stone Forest, or Shilin, is the World's premier pinnacle karst landscape. Located among the plateau karstlands of Yunnan Province, in southwest China, it is widely recognized as the type example of pinnacle karst, demonstrating greater evolutionary complexity and containing a wider array of karren morphologies than any other example

(Zhang et al. 2003b). The area is designated as a national park covering a protected Shilin area of 350 km2, and is organized into three zones with different protection levels. But no much evaluation work was done when the protected-area boundaries were delimited in 1984. The designation of these boundaries are mainly based on the scenery beautiful values of the Stone Forest Landscape, and it has no relationship with the karst landscape itself or its natural values. Further the boundaries are drawn on a small scale (1:1,000,000) geological map. They almost have no relationship with the topography characteristics such as road, river, topography line, or geological character. Thus, to a great extent it is even difficult for the administrative officials to know the direction of the boundaries and to find out their exact location, not to say for the public and local residents. This brings difficulty to carry out the accordingly conservation regulations.

An web-based SDSS for Lunan Stone Forest Conservation has been developed to provide a way to establish rational protective boundaries based on a variety of environmental and social criteria and render the location of the boundaries clear to the public (Zhang et al. 2005). However, the developed web-based SDSS was based on traditional Client-Server architecture and was implemented using traditional computer technologies such as Visual Basic 6.0, ESRI Mapobjects 2.1 and ESRI Mapobjects IMS 2.0 and ASP (Active Server Pages) (Zhang et al. 2005). Thus it is not an interoperable SDSS and has limitations for share and reuse of geographical data and geoprocessing although it indeed increased the public access to information and involvement in the decision-making processes for protective boundary designation decision-making processes. The objective of this case study is to develop an interoperable SDSS prototype to assist in protective boundary delimitation for Lunan Stone Forest Conservation based on the proposed framework shown in Figure 1. The interoperable SDSS prototype should render the location of the boundaries clear to the public. The prototype also should facilitate share and reuse of heterogeneous geographical data and geoprocessing over the web. The prototype covers several components in the proposed framework, such as using OGC WFS and WMS services to access the heterogeneous spatial data connected to heterogeneous legacy GISystems, using OGC WPS to access the multiple criteria decision model for delimitation of the protected-area boundaries, using OGC CS to register and discover the published WFS, WMS and WPS services, using WSDL as service interface to connect service providers, service brokers and service clients, using SOAP over HTTP for communication between web services over the web. Figure 3 illustrates the architecture of the prototype. The architecture consists of:

1. Data Service providers:
    - ESRI ArcGIS and PostGIS, which provide different format spatial data;
    - Geoserver (http://geoserver.sourceforge.net/html/index.php), an open-source software which enables full implementation of the OGC WFS and WMS specifications and serves ShapeFile and PostGIS data using WFSs and WMSs;
    - Java 2 Platform, Enterprise Edition (J2EE), the supporting environment for GeoServer;
    - Apache HTTP server, which serves as a web server for WFS and WMS;
    - Tomcat, a java servlet container, which provides web developers with a simple consistent mechanism for extending the functionality of a web server and for accessing web application GeoServer.
2. Web process service providers:
    - A Multiple Criteria Decision Model to incorporate the interacting biophysical and social-economic criteria such as geology, geomorphology, soil, vegetation,

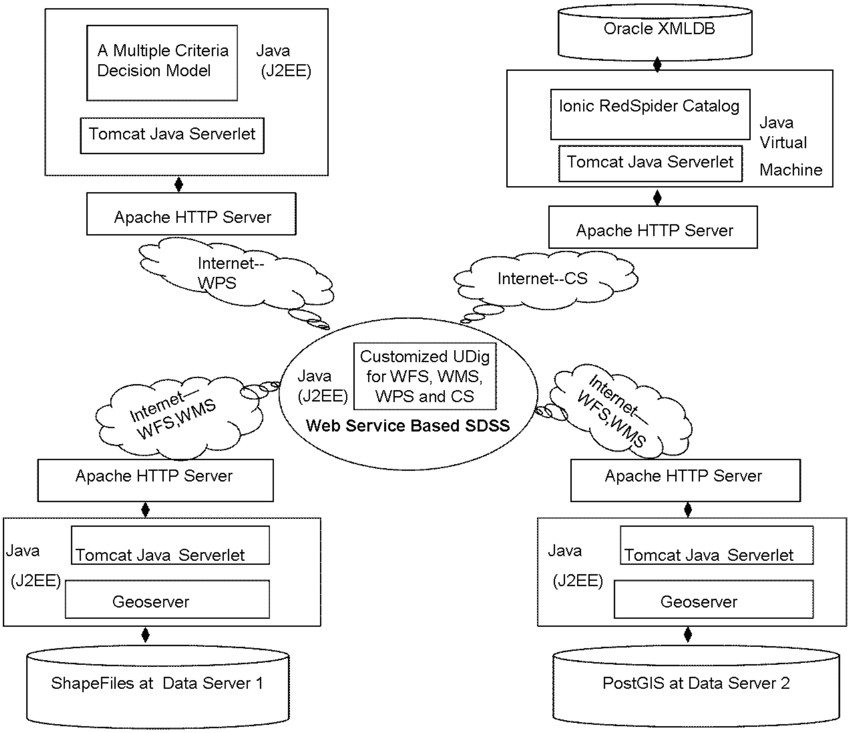


Fig. 3. Architecture of the implemented prototype.

population etc., into the delimitation of the protected-area boundaries, which is implemented as WPS using Java computer programming language;

- Java 2 Platform, Enterprise Edition (J2EE), the underlying developing environment for the Multiple Criteria Decision Model;
- Apache HTTP server, which serves as a web server for WPS;
- Tomcat, a java servlet container, which provides web developers with a simple consistent mechanism for extending the functionality of a web server and for accessing WPS.

3. Service brokers:
   - Ionic RedSpider Catalog software, which allows full implementation of the OGC CS specifications;
   - Oracle XMLDB, which provides capabilities for the storage and management of XML data for Ionic RedSpider Catalog;
   - Apache HTTP Server, which serves as a web server for catalogue services;
   - Tomcat Java Serverlet, which allows for accessing Ionic RedSpider Catalog;
   - Java Virtual Machine, the supporting environment for Ionic RedSpider Catalog.
4. Service clients:
   - Customized UDig software, which provides a user-friendly interface for decision-makers to query and access to web services such as WFS, WMS, WCS and WPS.

The same multiple-criteria decision model applied in the previous web-based SDSS (Zhang et al. 2005) was employed in this prototype but was recoded using Java computer programming language as web processing services. The multi-criteria evaluation approach was widely used in GIS literature (e.g. Carver 1991; Eastman et al. 1993). Among the many ways to integrate decision criteria, the weighted linear combination method is a popular one (Berry 1993; Hopkins 1977; Malczewski 2000) and was used to delimit different protected-area boundaries in this study. To rank the different protection level alternatives, the following formula was used:

$$S = \sum_{i=1}^{n} W_i C_i \tag{1}$$

where S is the suitability score with respect to the protection objective, W_i is the weight of the criterion i, C_i is the criterion score of i, and n is the number of criteria. The model has its own algorithm to make sure that $\Sigma W_i = 1$. By using formula (1), overall protective suitability scores were determined and the whole area was divided into several different level protection zones.

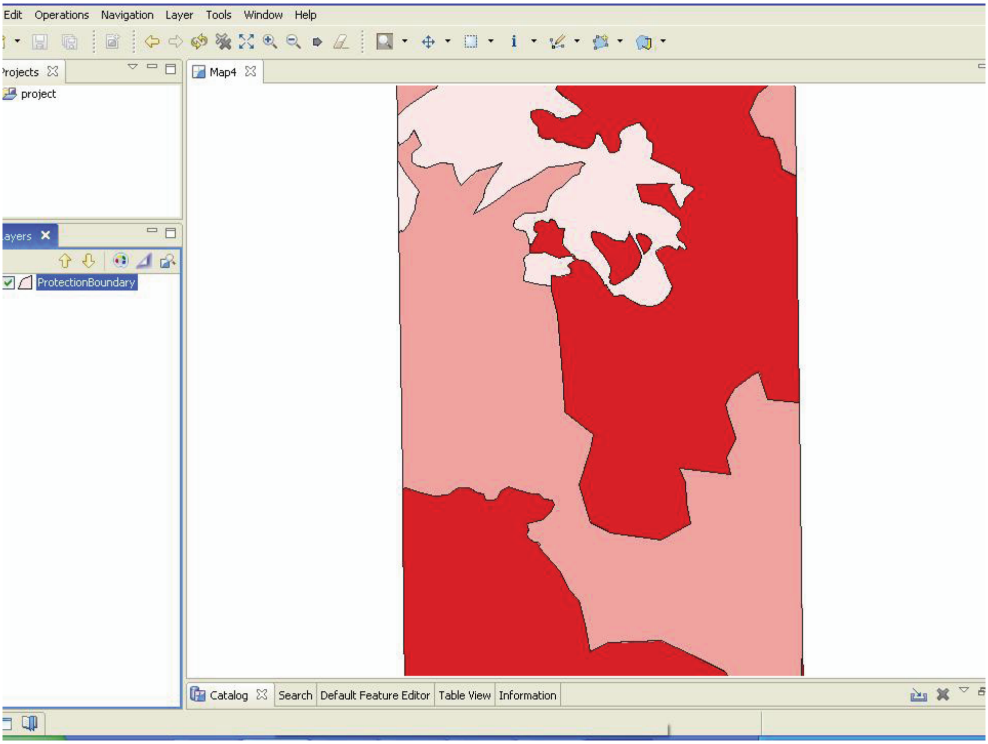


Fig. 4. One scenario of three different level protection areas delimited by using the prototype system.

3.2 Some experimental results

Through the prototype, decision-makers can delimit protective boundaries using the multiple criteria decision model based on a variety of biophysical and social economic criteria by employing OGC WFS, WMS, WPS and CS. Figure 4 shows one scenario of three different level protection areas that was delimited by using the prototype system. Note: in the prototype, the criteria data are stored in two different format databases (Shapfile and PostGIS) on two remote servers (http://140.146.179.29 and http://172.16.1.34). The multiple criteria decision model and the Ionic RedSpider Catalog software are held in the remote server (http://172.16,1.34).

Using the implemented prototype, decision-makers also can render the location of the boundaries clear to the public by aligning them with conspicuous landscape features such as water bodies, roads or buildings via employing WFS, WMS and CS (Figure 5). Note: in Figure 5 lake data (original format is Shapefile) come from the WMS from a remote server (http://172.16.1.34) and road data (original format is ArcSDE) come from the WFS from another remote server (http://140.146.179). The protective boundary data (original format is GML) come from WPS located at the same remote server with lake data (http://172.16.1.34).

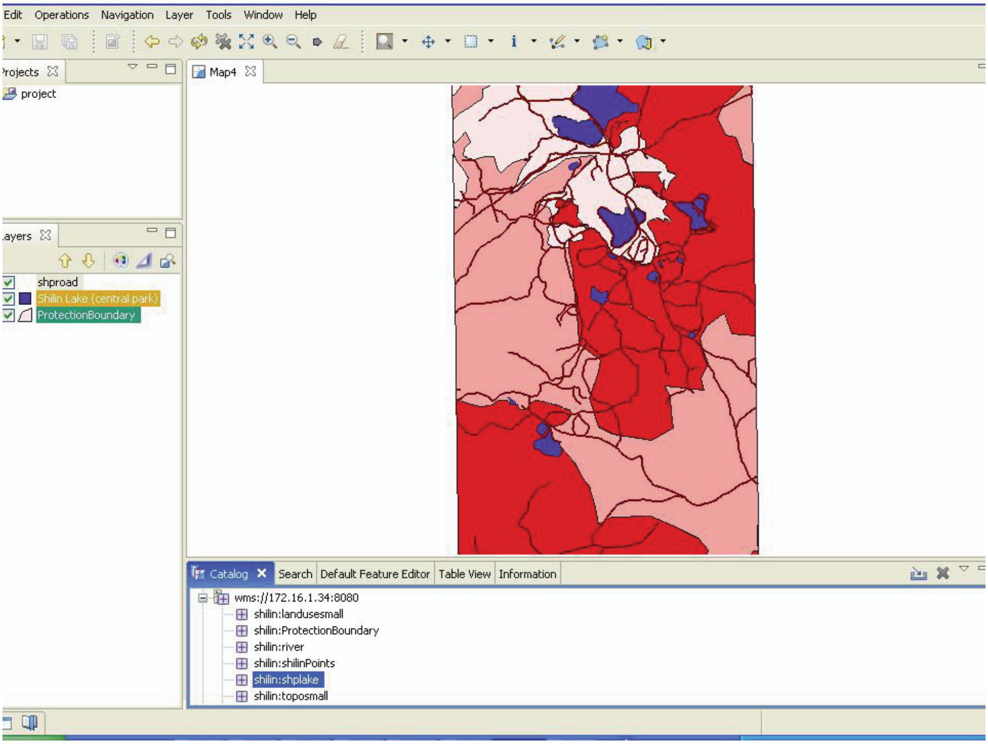


Fig. 5. Roads and lakes overlain on protective areas.

The following experimental results demonstrate some advantages of the web services-based prototype SDSS:

(1) The prototype system provides decision-makers with the ability to access and analyze heterogeneous criteria data in order to make better decisions for protected-area delimitation.

It allows the decision-makers access the heterogeneous criteria data from a variety of sources on the Internet. The criteria data, such as soil, vegetation, hydrology, geomorphology, land use, social and economic data, are stored in different databases with different formats. However, decision-makers can directly access these heterogeneous data sources without having to know specifically who might provide the data they want and the format of the data. They need not contact data providers by email or mail to get the files and convert them into a format they need to start the decision-making task. Figure 6 illustrates seamless and dynamic integration of geology data (original in Shapefile format) located at the data server http://140.146.179.29 and topography data (original in PostGIS databases) located at the data server http://172.16.1.34 by invoking the WFS and WMS services with little or no knowledge about the heterogeneous environments of the data providers. By seamless data integration the web services-based system not only promotes remote access and potential collaborative decision support applications. It also can reduce developing and maintenance costs and minimize data conflicts by avoiding redundant data.

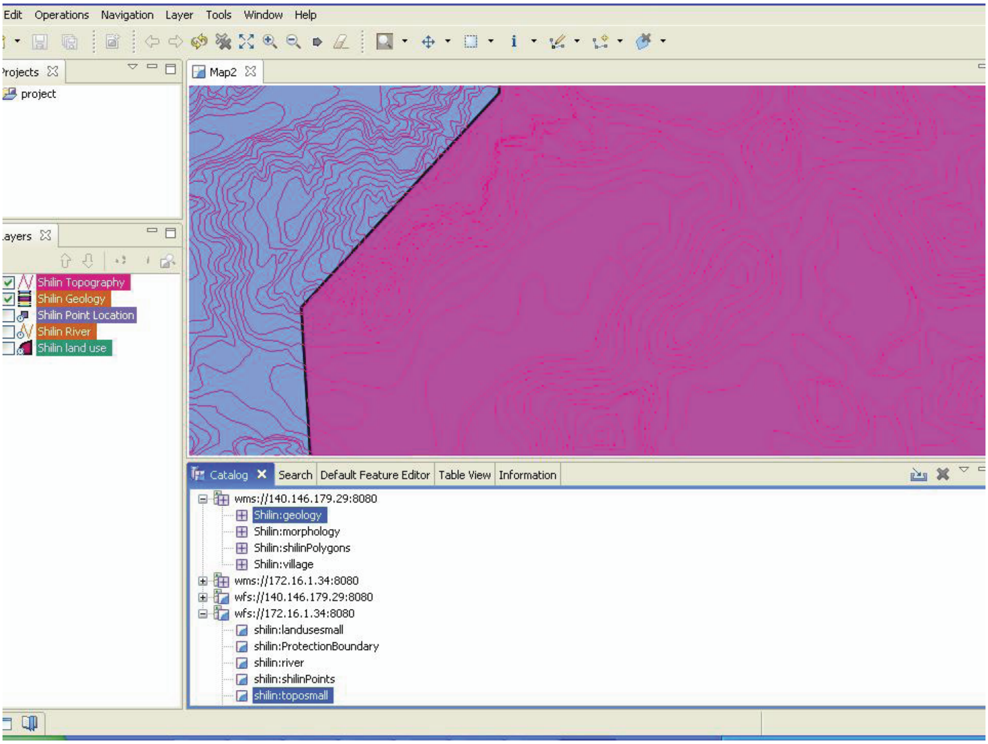
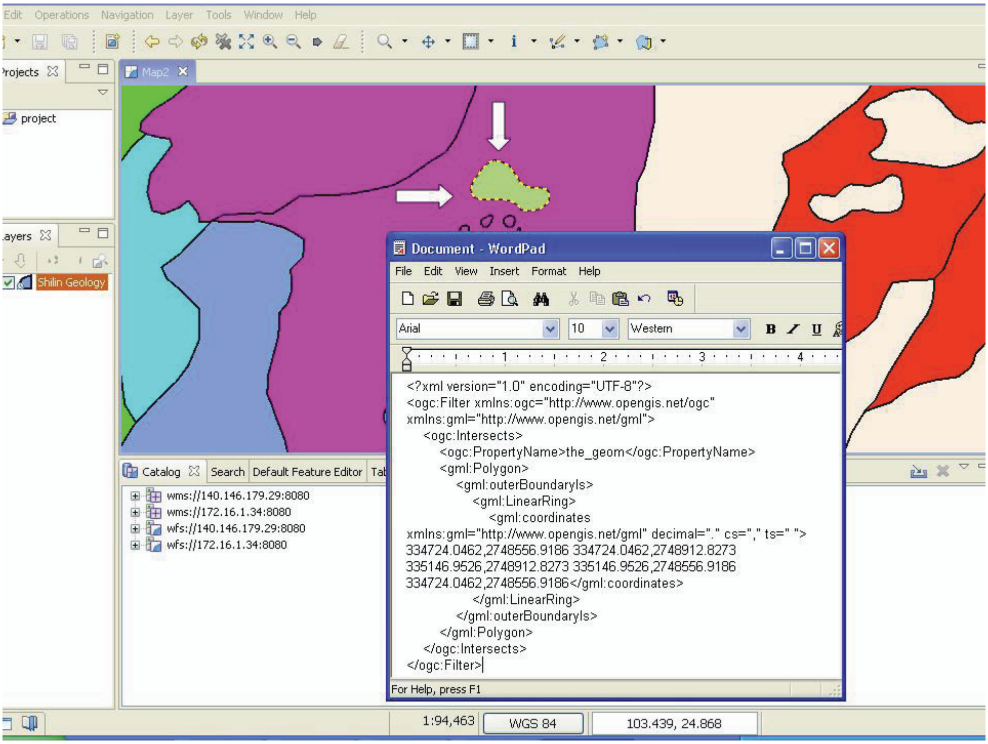


Fig. 6. Integrate different format data from disparate servers by invoking WFS and WMS services.

(2) The prototype system provides decision-makers with the ability of feature-level data search, access, and exchange in real time over the web. In some applications the actual

information needed by the decision-makers may be a subset of the information available. Decision-makers may need only several features of a data file or have interests only in a small area of a data file. Downloading entire datasets or data files will increase the time of data acquisition and analysis and affect the speed of decision-making. Because WFS deliver GML representations of simple geospatial features, decision-makers can access geographic feature data through WFS by submitting a request for just those features that are needed for a decision-making application. Figure 7 shows copy one geology polygon feature (the small polygon referred by two arrows) and paste it in GML format in a WordPad file over the web by WFS. Note: the original geology data format is a shapefile located in a remote data server http://140.146.179.29. The downloaded GML features may serve as input to decision models for small area decision-making processes.



wms://140.146.179.29:8080
wms://172.16.1.34:8080
wfs://140.146.179.29:8080
wfs://172.16.1.34:8080

```
<?xml version="1.0" encoding="UTF-8"?>
<ogc:Filter xmlns:ogc="http://www.opengis.net/ogc"
xmlns:gml="http://www.opengis.net/gml">
   <ogc:Intersects>
      <ogc:PropertyName>the_geom</ogc:PropertyName>
      <gml:Polygon>
         <gml:outerBoundaryIs>
            <gml:LinearRing>
               <gml:coordinates
xmlns:gml="http://www.opengis.net/gml" decimal="." cs="," ts=" ">
334724.0462,2748556.9186 334724.0462,2748912.8273
335146.9526,2748912.8273 335146.9526,2748556.9186
334724.0462,2748556.9186</gml:coordinates>
            </gml:LinearRing>
         </gml:outerBoundaryIs>
      </gml:Polygon>
   </ogc:Intersects>
</ogc:Filter>
```

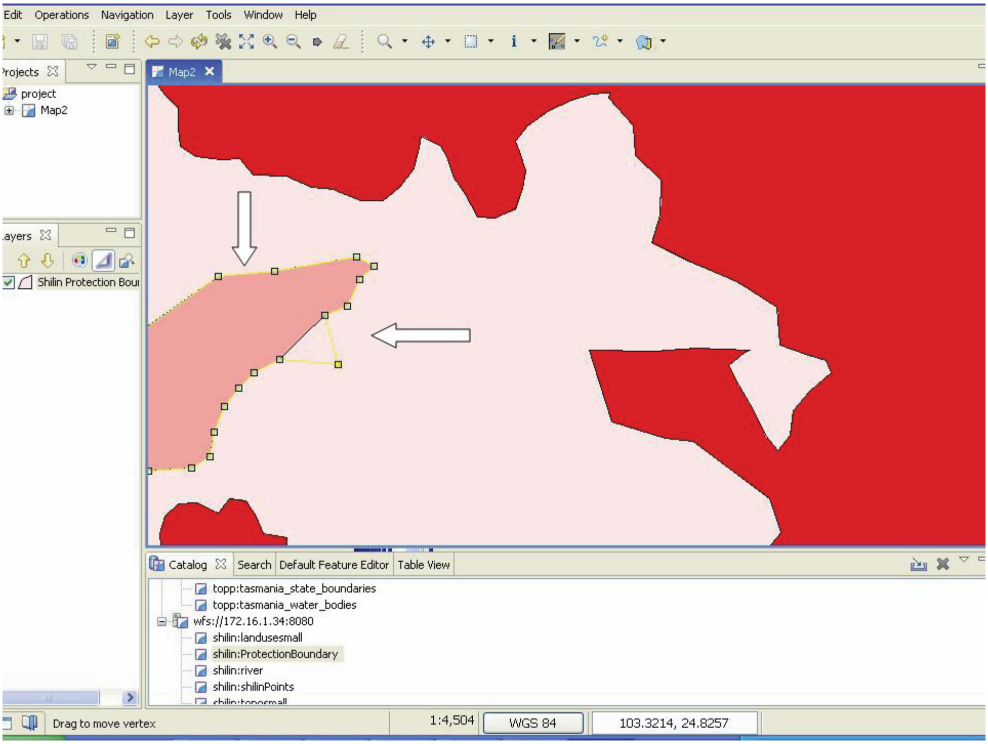
1:94,463 | WGS 84 | 103.439, 24.868

Fig. 7. Copy one geology polygon feature (the small polygon referred by two arrows) and paste it in GML format in a WordPad over the web by WFS.

(3) The prototype system allows decision-makers access the multiple-criteria decision model across the web via WPS. The WPS dynamically conduct spatial data analysis, compute the evaluation value and pass the evaluation results as input to WFS. The input and output data for the implemented WPS is in GML format, which are connected with WFS. Figure 4 displays one scenario of the different level protected-area boundaries calculated by WPS. In

this scenario, the first protection level (dark) covers almost all the limestone pinnacles and the lakes, which are considered by the karst scientist to be of great importance to the landscape; the second protection level (grey) includes nearby forests which have important influences on the local ecosystem; and the third protection level (white) contains less important protection targets including villages, farmlands, and tourism facilities, such as hotels, commercial stores, roads and parking lots. Since the multiple-criteria decision model is employed as web services, it provides the interoperable capability of cross-platform and cross-language and can be accessed and reused by other applications and organizations.

(4) The web services-based prototype system facilitates decision-makers access to the most up-to-date criteria data. With the WFS and WMS data maintenance of the prototype system becomes easy. Because the criteria data reside in the original databases, they are always updated. Unlike traditional SDSSs the data updated from one source have to be delivered or downloaded manually to its applications to maintain the changed data, the web services based prototype system automatically propagates the change or update of data. In the web services-based prototype system developers or decision-makers also can change or update criteria data or alternative solution maps remotely in disparate sources cross the web. They

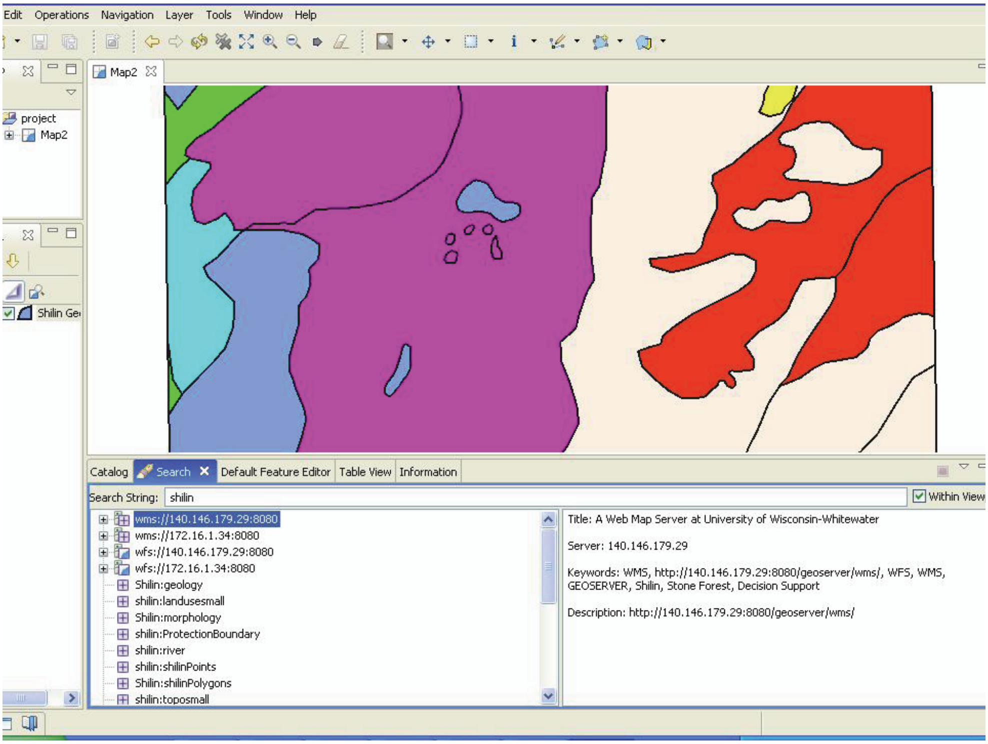


Drag to move vertex

Fig. 8. A decision-maker changes one scenario of protective boundaries created by another decision-maker in a remote server (http://172.16.1.34) over the web.

can create, delete and update geographic features in a remote database over the web using WFS. Figure 8 shows an example that a decision-maker changes one scenario of protective boundaries created by another decision-maker in a remote server (http://172.16.1.34) over the web. Changes to the protective boundaries are instantaneously relayed to other decision-makers and applications. This instant access to the most up-to-date information enables decision-makers avoid the tedious process of transferring data and facilitates the decision-making process. In this way inconsistencies generated by updates are minimized and enterprises collaboration for a specific joint project is supported.

(5) The catalogue services in the implemented prototype system enable decision-makers to dynamically discover and communicate WFS, WMS and WPS with a suitable resource provider. Decision-makers can search needed criteria data for the multiple criteria model from various resource providers in the registry by using keywords. Figure 9 illustrates the query results using keyword "shilin". All the service providers having "shilin" in their metadata, data or services are listed.



Search String: shilin

wms://140.146.179.29:8080
wms://172.16.1.34:8080
wfs://140.146.179.29:8080
wfs://172.16.1.34:8080
Shilin:geology
shilin:landusesmall
Shilin:morphology
shilin:ProtectionBoundary
shilin:river
shilin:shilinPoints
Shilin:shilinPolygons
shilin:toposmall

Title: A Web Map Server at University of Wisconsin-Whitewater

Server: 140.146.179.29

Keywords: WMS, http://140.146.179.29:8080/geoserver/wms/, WFS, WMS, GEOSERVER, Shilin, Stone Forest, Decision Support

Description: http://140.146.179.29:8080/geoserver/wms/

Fig. 9. Query web services using keyword "shilin".

Besides the aforementioned advantages, the web-services based prototype system has basic GIS functions enabling data investigation. For example, decision-makers can display and overlay different data layers, and can zoom in, zoom out, pan or query the attribute table of these data layers. Also, decision-makers can make different style maps for the WMS and WFS by changing styles inside the system or importing SLD (Style Language Descriptor) files from outside. Figure 10 illustrates different views of the same WFS data by importing different SLD files.

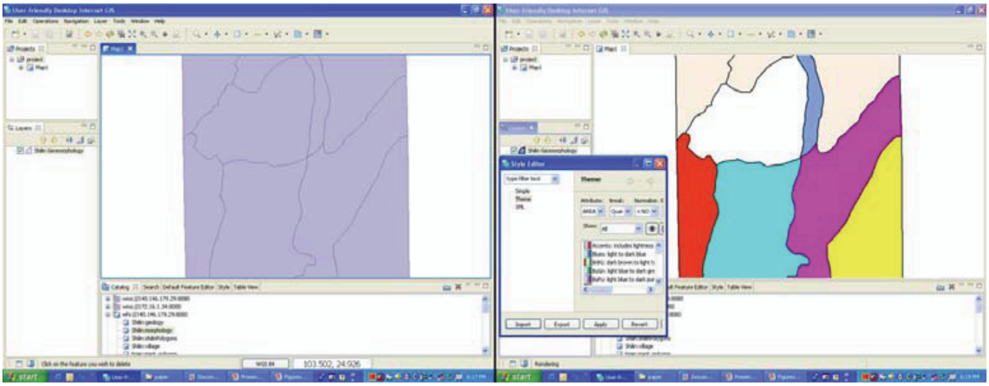


Fig. 10. Different views of the same WFS data by importing different SLD files.

4. Discussions and conclusions

This paper proposed a framework for web services-based spatial decision-making systems. A prototype has been implemented to demonstrate how to build an interoperable SDSS using OGC WFS, WMS, WPS and Catalogue Services based on the SOA. OGC WFS and WMS were used to access the heterogeneous spatial data and WPS were used to access the multiple criteria decision model. OGC Catalog Services were employed to locate geospatial data and processing services wherever they are located, and provide information on the services they find for the decision-makers. Results from the implemented prototype showed that the proposed framework provided an environment for interoperability technically via web services and standard interfaces. Information from any source may serve as input to the decision-making process in such systems. Decision-makers can access necessary geospatial information no matter where it resides, what format it takes and how quickly it changes. By reusing existing heterogeneous data and geoprocessing plus update and maintenance of data remotely across the web, the web services-based system provides a potential way to alleviate duplication problem and reduce related costs.

The proposed framework is particularly useful for organizations with scare resources such as limited time, expertise and finances to implement a SDSS. It is cost effective because it makes easier to distribute geospatial data and applications across platforms, operating systems, and computer language, and SDSS developers can find, access and use the information needed over the web. They no longer have to address the technical side of the SDSS to exploit its value because they do not need to develop and maintain whole databases and geoprocessing by themselves and can integrate existing geospatial data and functionality into their custom applications online.

Although the proposed framework offers the aforementioned advantages, it still has several issues which need further investigation. One issue is semantic interoperability. The proposed framework only resolves technical interoperability via web services and standard interfaces and it cannot resolve semantic heterogeneity problem in composition of web services. None of the XML-based standards such as WSDL and SOAP used by web services provide a means to describe a web service in terms of explicit semantics. Thus web services alone will not be sufficient to develop a real interoperable SDSS. Integration of web services and ontologies may offer a potential solution to the semantic heterogeneity problem. The second issue is performance. The framework uses WFS to deliver spatial vector data in GML format. However, the size of the GML files tends to be large especially when there are a large number of features included. The network and processing overhead associated with GML makes it inefficient for processing and storage performances. More research would greatly benefit from file compression algorithms and highly efficient parsing methods. The third issue is security. Using an identification and authentication that requires users to employ a login authentication may provide the first level of information security access control. One also can use the standard Secure Socket Layer and firewall based rules for security control at the transport level and digital signatures and/or encryption to protect specific parts of an XML/SOAP message at the application level. In addition, the Web Services Security Specification, which provides a complete encryption system, may be employed to add more security features to web services by using methods such as credential exchange, message integrity, and message confidentiality. Except for the three major issues discussed above, other issues such as the privacy issue, copyright issue and the data quality issue also need further study.

5. References

Anderson, G. & Moreno, S.R. (2003). Building web-based spatial information solutions around open specifications and open source software, *Transactions in GIS*, Vol. 7, pp. 447-66

Armstrong, M.P. (1993). Perspectives on the development of group decision support systems for locational problem solving, *Geographical Systems*, Vol.1, pp. 69-81

Bernard, L.; Ostländer, N. & Rinner, C. (2003). Impact Assessment for the Barents Sea Region – A Geodata Infrastructure Approach, *Proceedings of AGILE – 6th Conference on Geographic Information Science*. Lyon, France, pp. 653-661

Berry, J.K. (1993). Cartographic modeling: The analytical capabilities of GIS, In: *Environmental Modeling with GIS*, Goodchild M F, Parks B O and Steyaert L T (ed). New York, Oxford University Press, pp. 58-74

Carver, S.J. (1991). Integrating multi-criteria evaluation with geographic information systems, *International Journal of Geographical Information Systems,* Vol. 5, pp. 321-339

Carver, S.J. (1999). Developing Web-based GIS/MCE: Improving access to data and spatial decision support tools, In: *Spatial Multicriteria decision-making and analysis:*

A geographic information sciences approach, Thill JC(ed) New York, Ashgate, pp. 49-76

Devogele, T.; Parent, C. & Spaccapietra, S. (1998). On spatial database integration, *International Journal of Geographical Information Science*, Vol.12, pp. 335-352

Densham, P.J. (1991). Spatial decision support systems, In: *Geographical Information Systems: Principles and Applications,* Maguire D J, Goodchild M F and Rhind D W (ed) New York, John Wiley and Sons, pp. 403-412

Eastman, J.R. (1993). *IDRISI: A Grid Based Geographic Analysis System Version 4.1*, Worcester, MA: Clark University Graduate School of Geography

Hopkins, L.D.(1977). Methods for generating land suitability maps: A comparative evaluation, *Journal of the American Institute of Planning*, Vol. 43, pp. 386-400

Jankowski, P.; Nyerges, T. L.; Smith, A., Moore, T.J. & Horvath, E. (1997). Spatial group choice: a SDSS tool for collaborative spatial decision making, *International Journal of Geographic Information Science*, Vol.11, pp. 577-602

Jankowski, P.; Andrienko, G.L. & Andrienko, N.V. (2001). Map-Centered Exploratory Approach to Multiple Criteria Spatial Decision Making, *International Journal of Geographical Information Science*, Vol.15, pp. 101-127

Jankowski, P.; Robischon, S.; Tuthill, D.; Nyerges, T. & Ramsey, K. (2006). Design considerations and evaluation of a collaborative, spatio-temporal decision support system, *Transaction in GIS*, Vol.10, pp. 335-354

Keßler, C.; Rinner, C. & Raubal, M. (2005). An Argumentation Map Prototype to Support Decision-Making in Spatial Planning, *Proceedings of AGILE 2005 - 8th Conference on Geographic Information Science*, Toppen F and Painho M (ed). Estoril, Portugal, pp.135-142

Kishore, C.; Kerrie, H. & Edward, T.M. (2003). Migrating to a service-oriented architecture. www document, http://www-128.ibm.com/developerworks/library/ws-migratesoa/

Malczewski. J. (1996). A GIS-based approach to multiple criteria group decision-making, *International Journal of Geographical Information Systems,* Vol. 10, pp. 955-971

Malczewski, J. (2000). On the use of weighted linear combination method in GIS: common and best practice approaches, *Transactions in GIS,* Vol. 4, pp. 5-22

OGC document 02-023r4 (2003). OpenGIS Geography Markup Language (GML) implementation specification, version 3.00, www document, http://www.opengeospatial.org/ specs/?page=specs

OGC document 03-065r6 (2003). Web Coverage Service (WCS), Version 1.0.0. www document, http://www.opengeospatial.org/specs/?page=specs

OGC document 04-021r2 (2004). OpenGIS catalogue service specification, www document, http://www.opengeospatial.org/ specs/?page=specs

OGC document 04-024 (2004). Web Map Service, Version 1.3., www document, http://www.opengeospatial.org/specs/ ?page=specs

OGC document 04-060r1 (2004). OWS 2 Common Architecture: WSDL SOAP UDDI. www document, http://www.opengeospatial.org/ specs/?page=specs

OGC document 04-094 (2005) Web feature service implementation specification, version 1.1.0., www document, http://www.opengeospatial.org/specs/?page=specs

OGC document 05-007r4 (2005). OpenGIS Web Processing Service. www document, http://www.opengeospatial.org/specs/?page=discussion

OGC Interoperability Program White Paper (2001). Introduction to OGC Web Services. www document, http://ip.opengis.org/ows/010526_OWSWhitepaper.doc

Ostländer, N. (2004). Interoperable Services for Web-Based Spatial Decision Support. www document, http://www.agile-secretariat.org/Conference/greece2004/papers/P-13_Ostlander.pdf

Peng, Z.-R. & Zhang, C. (2004). The Roles of Geography Markup Language, Scalable Vector Graphics, and Web Feature Service Specifications in the Development of Internet Geographic Information Systems, *Journal of Geographical Systems*, Vol. 6, pp. 95-116

Power, D.J. (2003). Defining Decision Support Constructs, *Proceedings of the Seventh International Conference on Decision Support Systems (ISDSS'03),* Ustron, Poland

Rinner, C. (2003). Web-based Spatial Decision Support: Status and Research Directions, *Journal of Geographic Information and Decision Analysis* Vol.7, pp. 14-31

Sengupta, R.R. & Bennett, D.A. (2003). Agent-based modeling environment for spatial decision support, *International Journal of Geographical Information Science* Vol.17, pp. 157-180

Sugumaran, V. & Sugumaran, R. (2005). Web-based Spatial Decision Support System (WebSDSS): Evolution, Architecture, and Challenges. www document, http://mis.temple.edu/sigdss/icis05/fullProgram.htm

Van Der Perk, M.; Burema, J.R.; Burrough, P.A.; Gillett, A.G. & Van Der Meer, M.B. (2001). A GIS-based environmental decision support system to assess the transfer of long-lived radiocaesium through food chains in areas contaminated by the Chernobyl accident, *International Journal of Geographical Information Science*, Vol. 15, pp. 43-64

W3C. (2001). Web Services Description Language (WSDL), 1.1, www document, http://www.w3.org/TR/wsdl

W3C. (2003). SOAP Version 1.2 Part1: Messaging Framework. www document, http://www.w3.org/TR/soap12-part1/

Zhang, C.; Li, W.; Peng, Z.-R. & Day, M. (2003a). GML-based Interoperable Geographical Database, *Cartography* Vol. 32, pp. 1-16

Zhang, C, Day, M. & Li, W. (2003b). Land use and Land Cover Change in the Lunan Stone Forest, China, *Acta Carsologica,* Vol.32, pp. 161-174

Zhang, C. & Li, W. (2005). The Roles of Web Feature Service and Web Map Service in Real Time Geospatial Data Sharing for Time-Critical Applications, *Cartography and Geographic Information Science*, Vol.32, pp. 269-283

Zhang, C.; Li, W. & Day, M. (2005). Towards Establishing Effective Protective Boundaries for the Lunan Stone Forest Using an Online Spatial Decision Support System, *Acta Carsologica*, Vol.34, pp. 178-193

Zhang, C.; Li, W. & Day, M. (2006). Towards Rationalizing Protected-Area Designation in China Using a Web-Based Spatial Decision Support System, *Journal of Spatial Science*, Vol.51, pp.33-46

Spatial Decision Support System for Bank-Industry Based on GIS and Expert Systems Integration

Ana Maria Carnasciali and Luciene Delazari
Parana Federal University
Brazil

1. Introduction

This chapter performs a case analysis of the integrated use of Geographic Information Systems (GIS) and Expert Systems (ES) to assist decision-making process specifically on the question: where is the best place for a new retailer at the bank-industry? Choosing the best location for a new commerce unit is an important and complex decision because both spatial and symbolic variables are involved. The choice of locations requires strategic decisions based not only on common sense but also on the experience of experts, and this involves solid research. The integration of GIS with other special-purpose software can improve the potential of performed analysis. Particularly GIS integrated with ES can assist those tasks by making them less subjective. GIS can subsidize spatial marketing analysis by combining company socio-economic variables with the concurrence ones, while ES can store the expert's logical sequence of reasoning and even the importance order of symbolic variables. When the expert queries the ES, he or she gets data collected by other professionals otherwise dependant on them.

In bank industry those decisions are critical, because opening a new unit must take in account the spatial variables, as concurrence proximity, and symbolic variables concerned about the building where it will be located.

This chapter was based on a dissertation for Master Degree (Carnasciali, A. M. S., 2007) in which a study case was developed, on an analysis using integrated ES and GIS to open a new retailer of HSBC Bank Brazil S/A, in the city of Curitiba, Paraná – Brazil. A comparison was made between the methods used by the HSBC experts against the proposed system. Conclusions pointed out that the proposed system has contributed to an improvement in the decision-making process for the banking industry location problems, enlarging possibilities for spatial analysis and assisting the experts with subjective tasks.

2. Expert systems and geographic information systems

The use of expert systems (ES) aims to solve major problems which, otherwise, would demand too much effort from non-expert users. According to Waterman (1986) the key to the success of an ES is to identify the type of problem and the circumstances for which the ES is suitable:

- The problem can not be too large or complicated, for example, a specialist should take a few hours to solve it, not many days;
- Procedures should be established to solve the problem, and this solution should be consensus among the experts;
- Sources must exist to solve the problem (in the form of a system or roll of procedures) and should be accessible;
- The solution to the problem should not be based only on "common sense". In addition, Waterman (1986) also mentions that an ES is applicable when there is need to replicate the solution of the problem and the resources are not sufficient (there are few professionals, or professional work load is excessive). In this context, according to Rodriguez-Bachiller & Glasson (2004), the ES can be used to free the specialists to perform more complex tasks. That would allow the system to work as a learning support tool for non-specialists. This approach would result the ES to be an instrument of technology transfer, which according to the authors, makes them more attractive.

Expert Systems were first developed from Artificial Intelligence, aiming the design of computer systems which could be “trained” to perform specific operations. Examples are neural networks, software that can be trained to recognize, for instance, specific patterns detected by sensors, so they could identify the same patterns in other situations. At the same time, studies started aiming to understand how the brain performs certain operations to trying to capture and use this knowledge to solve problems as humans do.

This emphasis on the acquisition of knowledge promoted the interest in methods of knowledge representation to codify the knowledge applicable in particular situations. We investigated two methods of representation of knowledge: declarative knowledge, which describes a situation in a context, its elements and relationships. The semantic networks are the key to this approach, and were originally developed by Quillian in 1968, to represent the meaning of words, describing the objects in the classes to which they belong, their components and their characteristics. The second method, known as procedural knowledge, focuses on describing how to use the knowledge we have to solve the problem. Examples of this method are the production rules to represent logical solution, with "IF-then" rules that express how to infer values of certain variables (conclusions) from the knowledge of values of other variables (conditions) (Rodriguez-Bachiller & Glasson , 2004). The latter method was used in developing the experiment presented in this article.

The components of an ES are:

- Knowledge necessary to solve the problem, represented as If-then rules that are stored in so-called base of knowledge;
- The rules used in the chain of inference, called the inference engine;
- The interface that allows the user to provide any information necessary for the initiation of proceedings by the chain of inference.

A limitation presented by the ES is related to the difficulty of manipulating the spatial information with the traditional features of such systems. Furthermore, another class of programs has the special function of handling spatial data and their relationships: the Geographic Information Systems (GIS). One of the best definitions that may express a GIS was given by Maguire (1991): GIS can be seen simply as a spatially referenced database. According to Burrough & McDonnell (1998), GIS store spatial data, which describe the geography of the area (shape and position) and descriptive data, i.e., they provide both qualitative and quantitative information about spatial data. GIS allows the user to associate descriptive information to spatial entities.

GIS is the most appropriate system for the spatial analysis of geographic data. It has the ability to establish spatial relationships between elements. This ability is known as topology, i.e. the mathematical method used to define the spatial relationships. The structure of data in GIS, in addition to describing the location and geometry of the entities, defines relations of connectivity, adjacency, proximity, relevance, continence and intersection (Aronoff, 1989). Spatial data in a GIS are presented in the form of maps, which, within this system, may be object of a large number of transformation and manipulation processes. Examples are the operations of overlay, in which the maps are superimposed to produce new maps. The overlay operations are popularly known as the map algebra: the information on different maps receives weights and are then combined by arithmetic or logic operations. Example of this operation is the use of multi-criteria analysis to evaluate possible locations for a commercial activity. Other operations include "clipping", which means cutting part of a map at the limits of polygons, to obtain the descriptive statistics for the objects in the map, perform analysis and multivariate correlation and regression of values of different attributes on the map to define areas of influence of objects (buffers) such as, for example, areas in risk of flooding.

The ability of GIS can also be expanded with the creation of scripts by the user or with its integration with other systems. In this case, the use of integrated GIS and ES presents some advantages. The GIS gathers the data needed for spatial analysis in a unique digital base, stores the rank of importance of the spatial variables and still has all the operational advantages system, especially the ease of viewing combinations of different data. The ES stores the sequence of reasoning of the experts, as well as the rank of importance of symbolic variables, which is defined and accepted by all specialists. A new specialist, by consulting the ES, will get information that might take much more time to be collected with other specialists working in the institution.

2.1 Examples of integrated GIS and ES

The development of an integrated GIS to an ES is shown by Eldrandaly et al. (2003). The authors propose a system of spatial decision support that includes an ES and a GIS to select the best location of an industry. The main components of the system are: GIS, ES, user interface and multicriteria analysis using the AHP technique (analytical hierarchical process), used to define the weights of different variables. Once the knowledge base that supports the expert system is fairly limited, the authors consider the solution presented as a prototype and which can be extended to similar applications.

Zhu et al. (1996) developed a decision support system to assist in space planning and use of land in rural Scotland. The main functions are consultations, formulation and evaluation of models of land use through the integration of database and rules-based system. The system allows the user to specify what is their interest as well as the factors to be considered in evaluating the potential land use. The system can then formulate a model of land use that meets the user's preferences.

MacDonald (1996) developed a system of spatial decision support to help minimize the squandering of solid waste, integrating GIS and ES. To the author, the research is significant because it integrates various tools and provides easy to understand results. These tools include multi-criteria analysis, models of planning, sensitivity analysis and the presentation of results in a georeferenced way. As benefits of the research the author highlights the possibility of making scientific techniques more accessible to the sector of solid waste by allowing users to make the analysis more quickly and easily. The author also stresses that the system allows a better understanding of the sector of solid waste, since it presents several conflicts of interpretation, such as different environmental and economic approaches.

According to the papers presented, one realizes that the integration of GIS and ES may assist decision making in different applications, such as in selecting the best location of an industry, in the land use and planning in rural areas, in minimization of solid waste. Probably one of the factors that contribute to facilitate the decision process is the ability to gather more information, being the new information generated from the integration of GIS and ES, helpful in various analyses and planning tasks.

The literature on the subject is about the isolated use of either GIS or ES in problems involving location of bank branches. No integrated system of GIS and ES to assist location decision-making of a bank branch was found. Therefore, this article proposes the use of both systems jointly.

3. Integrating GIS and ES for defining the location of a new branch

3.1 Background

This study was conducted in the city of Curitiba, located in the State of Paraná, southern Brazil, as shown in Figure 1.

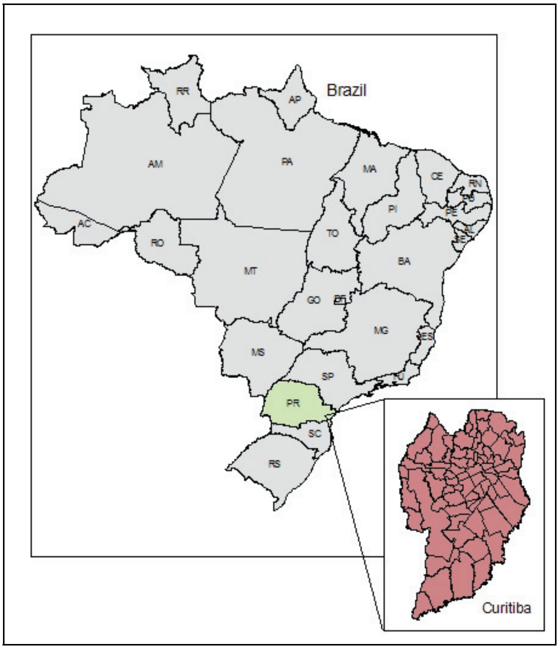


Fig. 1. Study Area

The HSBC Bank Brazil S/A - Multiple Bank is present in Brazil in about 550 municipalities, and has more than 1700 branches and service sites. In the city of Curitiba-PR, HSBC has a total of 29 branches, 49 in company dedicated offices (DO) and 41 sites of Electronic Services (SES) (in March 2006). The analysis of expansion, reduction or reshuffling the network of products or services is performed by three departments: Branch Network, DO and Payroll and Real Estate Administration. The Department of Branch Network performs feasibility studies of each branch. DO and Payroll Department performs feasibility studies of small and limited banking service sites. Real Estate Administration is responsible for selecting and contracting realties for the branches and DOs. Through outsourced service of real estate companies, buildings are identified as possible locations of a new branch, considering the existing branches, competitor bank branches, traffic generators and main streets. Subsequently, it is considered the best option based on rental values, physical structure and radius of action, ending the process with the concurrence of the Executive Board. The Department of Real Estate Administration does not make use of GIS or ES. The department receives maps (usually paper maps purchased in stationeries or newspaper stands or images obtained in internet) from the hired real estate companies providing base information on the surveyed properties and surroundings. In the case of paper maps, interest points are marked manually.

For the banking sector, selecting the best location for a new branch is an important and complex task, because it involves a number of variables, not only spatial but also symbolic and often includes personal opinions.

It is important to emphasize the use of ES as an aid tool in location decision-making. To identify available properties for a new branch, either for purchase or rental, consults to real estate companies and newspaper advertisements area necessary, and even local visits to the region, checking for signs or personal information. The use of ES becomes necessary because keeping an updated GIS database of available properties including surrounding characteristics such as transport and advertising possibilities would be too costly. Therefore, the task of deciding on the most appropriate site for a new branch can be improved through the integration of GIS and ES. The integration of systems improves the location decision in banking sector, because it extends possibilities for spatial analysis and assists specialists in tasks that involve subjective and often personal opinions. To make the choice of a new branch with the use of GIS and SE in an integrated way the activities are presented in the following topics.

3.2 Structuring geographic information system

The main analysis to be made concerns to the location decision, i.e., choosing the best location for the deployment of a new branch. The best region should be first considered and then the best spot in the chosen region. The analysis of the region was held jointly with experts from the HSBC Bank Brazil S/A of Curitiba, in the areas of Contracts, Law and Equity, on the digital base map of the city of Curitiba.

When selecting the region, the existing network of branches, in company dedicated offices (DO) and sites of Electronic Services (SES), both HSBC and competitors', should be considered. The street network, neighborhood limits and bus terminals were defined as the cartographic basis. The base map was obtained from IPPUC (Curitiba Institute for Research and Urban Planning). Moreover, information on traffic generator sites, such as colleges and universities, schools, supermarkets, hospitals, libraries, shopping malls are also relevant to the analysis.

The data that comprise the database of bank branches and other bank service sites, have been identified from the analysis to be performed in the GIS. The addresses were obtained from the Brazilian Central Bank (BC). The HSBC Court Cases Department provided the addresses of HSBC branches, DO and SES in Microsoft Excel (XLS) file for March 2006, which were compared with the list provided by BC and updated. The socioeconomic data were obtained from the Brazilian Institute of Geography and Statistics (IBGE) - Census of the year 2000.

3.2.1 Determination of spatial variables

Based on meetings with experts from the Department of Real Estate Administration of HSBC Curitiba, the relevant spatial variables for the region choice were defined: income, population density, population, traffic generators and growth rate of the neighborhoods. An important decision regarding the classification of numerical data is the number of classes and their limits. For the income variable, the ranges were defined based on purchasing power limits according to the HSBC categories Premier, Gold Class, Super Class and investor segment. For the population density, population, traffic generators and growth rate of the districts variables, several simulations were performed to define the limits of classes that best represent all the information in accordance with the understanding of experts. As there are different spatial variables that contribute to the choice of most appropriate site for a new branch, it was necessary to establish a hierarchy, i.e. an order of importance in order to identify the relative contribution of each one, as shown in Table 1.

<table>
<tr><th colspan="5">Hierarchy
+ ←→ -</th></tr>
<tr><th>Income
(BR R$)</th><th>Population density
(Inhab/Km^2)</th><th>Population
(Inhab)</th><th>Traffic generators
(#)</th><th>Growth rate of neighborhoods
(% per year)</th></tr>
<tr><td>196,78 a 750,00</td><td>0,01 a 25,00</td><td>1 a 500</td><td>5 a 10</td><td>-3,66 a 0,00</td></tr>
<tr><td>750,01 a 1500,00</td><td>25,01 a 50,00</td><td>501 a 750</td><td>11 a 20</td><td>0,01 a 1,00</td></tr>
<tr><td>1500,01 a 2500,00</td><td>50,01 a 100,00</td><td>751 a 1000</td><td>21 a 30</td><td>1,01 a 3,00</td></tr>
<tr><td>2500,01 a 5000,00</td><td>100,01 a 200,00</td><td>1001 a 1500</td><td>31 a 40</td><td>3,01 a 5,00</td></tr>
<tr><td>5000,00 a 11242,37</td><td>200,01 a 3216,15</td><td>1501 a 4227</td><td>41 a 64</td><td>5,01 a 16,88</td></tr>
</table>

Table 1. Hierarchy of Spatial Variables
Source: Carnasciali, A. M. S., 2007

3.2.2 Definition of the region

In the development of the GIS, the software ESRI ArcGIS 9.0 was used. The first step was to perform the geoprocessing (geographic positioning) of branches, DOs and SES of both HSBC and competition, totaling approximately 900 points, for later viewing in combination with socioeconomic data, traffic generators and others. The positioning process resulted in points on the center line of the streets, but as these centerlines could be the borders of neighborhoods, the points were moved to the left or right side of the street based on odd or even numbers of the property, respectively.

For the analysis of the regions, the limits defined by census tracts were considered, what allowed better detail compared with the limits of neighborhoods. According to the IBGE - Brazilian Institute of Geography and Statistics, the census tract is the territorial unit for data collection, formed by continuous area, with homogeneous occupancy and not cut but any relevant obstacle. The formation of a census tract for population survey is based on the number of private households. Each census tract has an average of 250 households.

The spatial variables received weights from the experts and then were combined, producing a grade for the region. An interval of five classes for the regions was established; that allowed the identification of the best scored ones. The selection of the region was conducted by experts from HSBC on the digital cartographic base, being observed, beyond the census tracts, the unities of the HSBC and competition, schools, main streets, as shown in Figure 2.

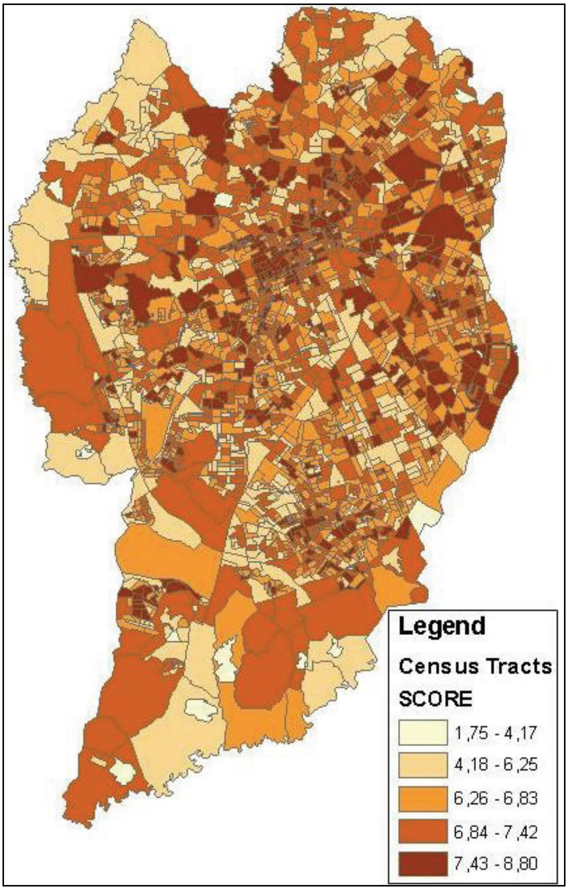


Fig. 2. Census tracts, HSBC branches, branches of other banks, schools and street network
Source: Carnasciali, A. M. S., 2007

The first choice of a neighborhood was Capão da Imbúia, an area with influence of the demand of neighbor city of Pinhais-PR, and an important attractive site, Unibrasil College. The Department of Real Estate Administration of HSBC contacted real state agents but as no suitable property was found in the area, another region was selected. This second option was for Sítio Cercado district, with emphasis on Izaac Ferreira da Cruz Street (Figure 3). According to the IBGE 2000 Census, Sítio Cercado has presented significant growth. The district has a population of 102,410 inhabitants, population density of 92.07 inhabitants/km² and a number of schools, supermarkets, and other services but only one branch of competitor bank: Banco Itaú S/A.

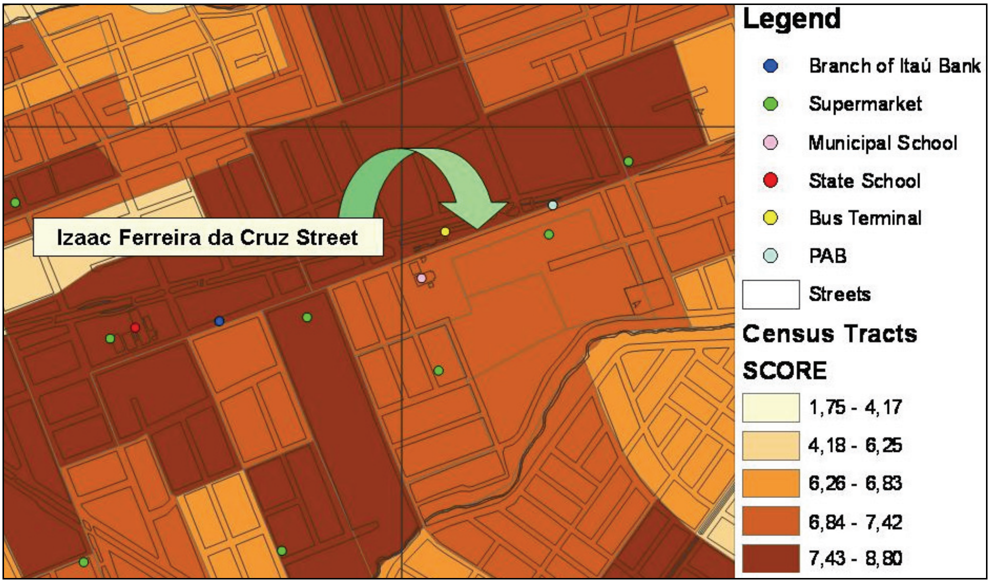


Fig. 3. Sítio Cercado District, stressing Izaac Ferreira da Cruz Street
Source: Carnasciali, A. M. S., 2007

3.3 Expert system development

Alongside the development of GIS, acquisition of knowledge was held, which is a decisive stage for the development of ES. It consisted of studies of books on marketing, and Geomarketing, regional economy as well as articles related to making decisions involving the best location of a point for location of banks and industries. Later, meetings were held with experts from the Department of Real Estate Administration of HSBC in order to gather information on how the decision on the best location to open a new branch is taken and what procedures are adopted, in an approach that aimed to make clear their reasoning. From these meetings with experts, some symbolic variables, relevant to the analysis of sites, were identified: region, purpose, constructed area, number of floors, price, location of the property/block, visibility, location of the property/street, street network, walking facilities, environment, transportation, access to street and building, access for the physically disabled, lighting and ventilation, toilets, parking, access to parking, use of parking, 2nd fire exit and advertising.

As there are different symbolic variables that contribute to choosing the most appropriate location for a new branch, experts have established a hierarchy, i.e. an order of importance between them in order to identify the relative contribution of each item (Table 2).

3.3.1 Software

For the development of ES, Shell Expert SINTA 1.1 was chosen. Shell Expert SINTA is a computational tool by the Applied Intelligent Systems of the Artificial Intelligence Laboratory of the Federal University of Ceará, that uses artificial intelligence techniques for automatic generation of expert systems. Some of its features include: the graphical interface, confidence factors, tools for debugging, possibility to include online help for each database, and others.

Hierarchy	Group	Symbolic Variables
	1	Region
		Purpose
+		Constructed area
		Number of floors
		Price
		Accesso to street and building
		Access for the physically disabled
		Lighting and ventilation
		Toilets
		Access to parking
		Advertising
		2nd fire exit
	2	Parking
		Surroundings
		Walking facilities
	3	Uso Estacionamento
		Location of the property/block
		Sistema Viário
		Visibility
-		Location of the property/street
		Transportation

Table 2. Hierarchy of symbolic variables
Source: Carnasciali, A. M. S., 2007

To manage the knowledge base, it was necessary to feed the system the following data: the variables (problems, factors that must be observed), the rules, the questions (interaction with the user's system specialist) and goals (the result of a query). A simplified architecture of Expert SINTO (Figure 4) is composed of:

- Knowledge: information (facts and rules) that the experts use;
- Base Editor: used for the implementation of the desired base;
- Inference Engine: sector of the ES responsible for deductions founded on the knowledge base;
- Global Database: evidences given by the user of ES during a consultation.

Once Expert SINTA already has an inference engine the concern is focused only on the representation of knowledge; the interprets this knowledge and runs it. The knowledge base generated in this work contains a base of information on how to determine the best point for the location of a new branch of HSBC Bank Brazil S/A. Menus were also prepared to help with explanations on the question that is being made. HelpScribble 7.6.1 was used to develop the Help menus.

The production rules used by Expert SINTA are a set of IF THEN rules, with the possibility of including logic connections to link the attributes in the scope of the database. An example is shown in Figure 5.

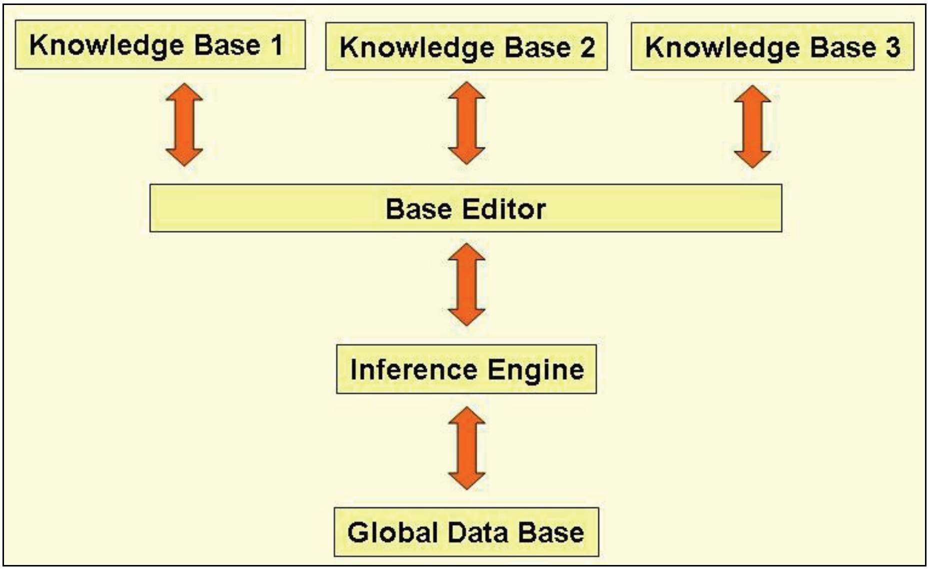


Fig. 4. Simplified architecture of Expert SINTA
Source: Adpated from Expert SINTA 1.1 Manual

IF purposei = Branch shared with department

THEN npurposei = 10 CNF 100%

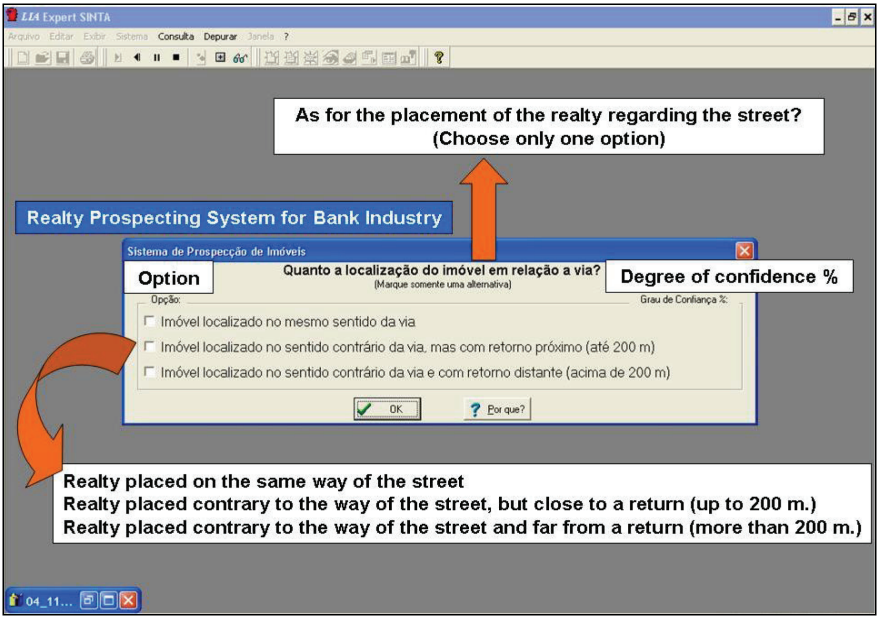
Fig. 5. Example of production rule in the proposed system
Source: Carnasciali, A. M. S., 2007

For the development of Expert System, named Realty Prospecting System (RPS), 178 rules were drawn up, totaling 22 questions conducted by the expert system (Figure 6) and the help menu (Figure 7). The system was developed entirely in Portuguese.

3.3.2 Definition of spot

The definition of the exact point of the location demands a prospection of realties in the area of interest. Four properties were selected in Sítio Cercado district, meeting the necessary features for a standard branch. The property prospectuses were: Building 1, located at 2710, Izaac Ferreira da Cruz Street; Building 2, located at 3330, Izaac Ferreira da Cruz Street; Building 3, also located at 3330, Izaac Ferreira da Cruz Street; and Building 4, at 1381, Pioneiros Street (Figure 8). The properties were visited for collecting the necessary information for assessing the feasibility of use.

By request of the Department of Real Estate Administration of HSBC Curitiba, real estate companies fill a form called Realty BID. The form contains information of the owner or dealer (name, phone/fax, email); of the property (address, area of land, built area, registration), identification / diagnosis of the property (tax statement, urban zoning), geographic and socioeconomic information of the region, purchase conditions, description of realty (number of floors, parking spaces, and others). Based on the reported information,



As for the placement of the realty regarding the street?
(Choose only one option)

Realty Prospecting System for Bank Industry

Sistema de Prospecção de Imóveis

Option

Quanto a localização do imóvel em relação a via?
(Marque somente uma alternativa)

Degree of confidence %

Opção:

Grau de Confiança %:

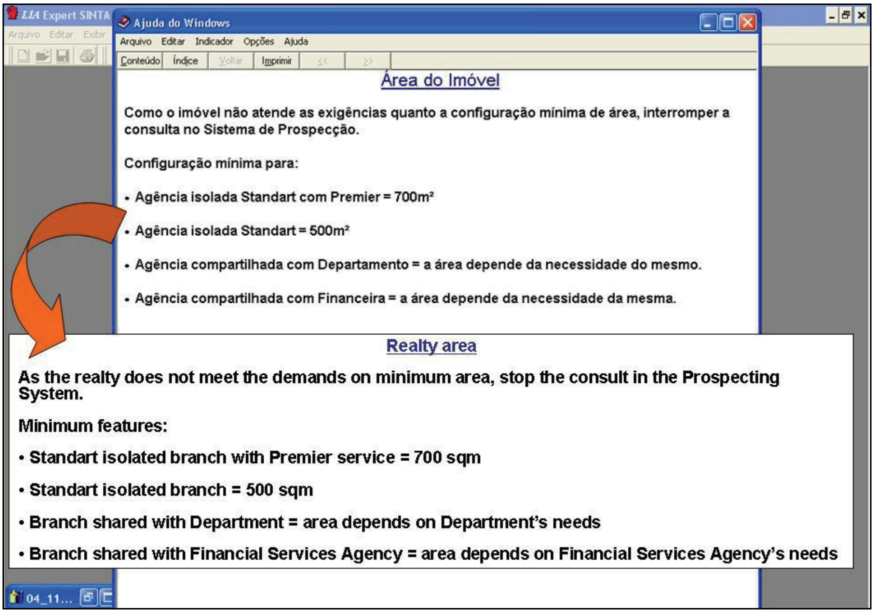
□ Imóvel localizado no mesmo sentido da via

□ Imóvel localizado no sentido contrário da via, mas com retorno próximo (até 200 m)

□ Imóvel localizado no sentido contrário da via e com retorno distante (acima de 200 m)

Realty placed on the same way of the street
Realty placed contrary to the way of the street, but close to a return (up to 200 m.)
Realty placed contrary to the way of the street and far from a return (more than 200 m.)

Fig. 6. Screen on question to the expert
Source: Carnasciali, A. M. S., 2007


LIA Expert SINTA
Arquivo Editar Exibir
Ajuda do Windows
Arquivo Editar Indicador Opções Ajuda
Conteúdo Índice Voltar Imprimir << >>
04_11...

Área do Imóvel

Como o imóvel não atende as exigências quanto a configuração mínima de área, interromper a consulta no Sistema de Prospecção.

Configuração mínima para:

- Agência isolada Standart com Premier = 700m²
- Agência isolada Standart = 500m²
- Agência compartilhada com Departamento = a área depende da necessidade do mesmo.
- Agência compartilhada com Financeira = a área depende da necessidade da mesma.

Realty area

As the realty does not meet the demands on minimum area, stop the consult in the Prospecting System.

Minimum features:

- Standart isolated branch with Premier service = 700 sqm
- Standart isolated branch = 500 sqm
- Branch shared with Department = area depends on Department's needs
- Branch shared with Financial Services Agency = area depends on Financial Services Agency's needs

Fig. 7. Help Menus
Source: Carnasciali, A. M. S., 2007

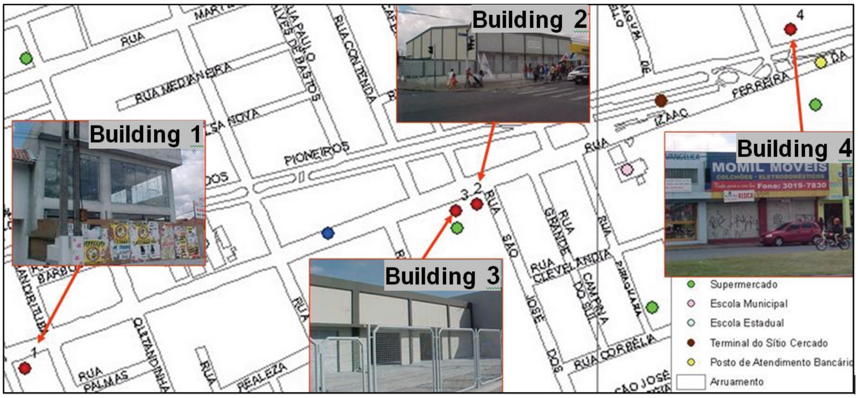


Fig. 8. Location of prospected realties
Source: Carnasciali, A. M. S., 2007

a pre-analysis, conducted by experts in Legal Matters, Maintenance, Fire & Safety, Trading, etc, is produced. In order to compare the used and proposed methods the author filled the forms of the selected properties. During this process the experts of the bank were asked to issue their opinions and justified them, listing the factors that influenced or assisted in their trial. The experts were also demanded to establish a ranking of the properties. Subsequently, the assessment was made using the Realty Prospecting System developed for the study.

3.4 Systems Integration

The Realty Prospecting System for the Banking Sector (RPS) integrates GIS and the specific ES (Figure 9). The user, in this case the expert of HSBC, is questioned whether there is a previously defined region, if not he or she is instructed to select it in the GIS. Once the region is defined, the prospecting process starts.

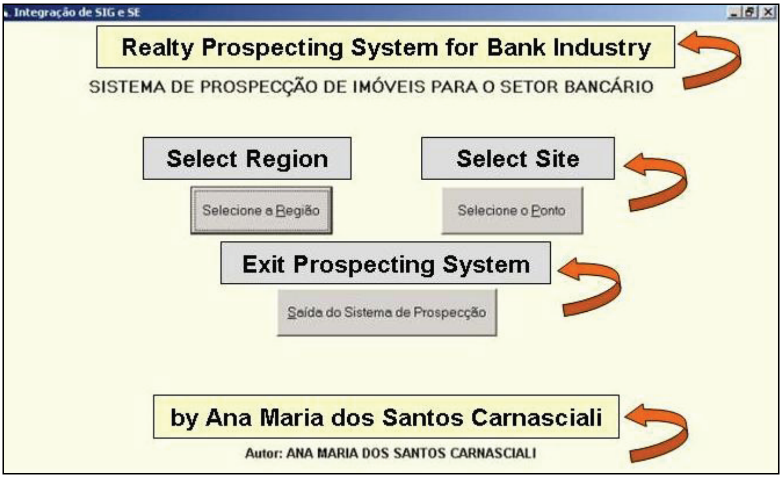
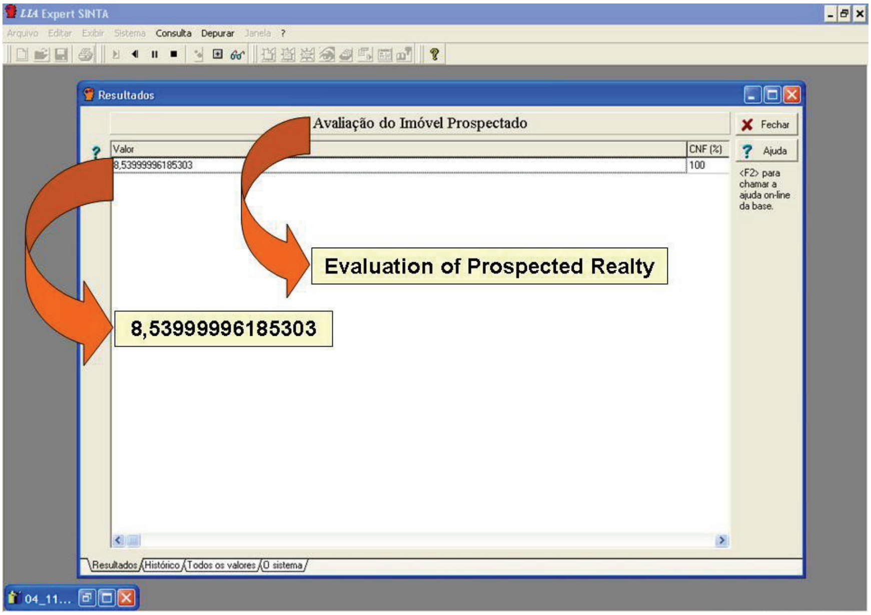

Integração de SIG e SE
Realty Prospecting System for Bank Industry
SISTEMA DE PROSPECÇÃO DE IMÓVEIS PARA O SETOR BANCÁRIO
Select Region
Selecione a Região
Select Site
Selecione o Ponto
Exit Prospecting System
Saída do Sistema de Prospecção
by Ana Maria dos Santos Carnasciali
Autor: ANA MARIA DOS SANTOS CARNASCIALI

Fig. 9. Realty Prospecting System for Bank Industry
Source: Carnasciali, A. M. S., 2007

Later, the expert returns to RPS for selecting the best spot. At the end of the consultation of each property the system displays a score and its position in a raking of scores (Figure 10). Thus the system assists the expert in selecting the best spot. Microsoft Visual Basic 6.0 was chosen as tool for the development of the integration between the systems.


LIA Expert SINTA
Arquivo Editar Exibir Sistema Consulta Depurar Janela ?
Resultados
Fechar
Ajuda
Resultados Histórico Todos os valores O sistema
04_11...

Avaliação do Imóvel Prospectado

Valor	CNF (%)
8,53999996185303	100

<F2> para chamar a ajuda on-line da base.

Evaluation of Prospected Realty

8,53999996185303

Fig. 10. Realty Prospecting System for Bank Industry
Source: Carnasciali, A. M. S., 2007

4. Results

The forms filled by the experts and the correlate ones produced by RPS were compared, resulting some considerations.

No expert considered all the variables that comprise the proposed system. It is important to emphasize that these variables they were defined by the experts as essential to the analysis, during the development. Just as an illustration of the importance of all variables in the analysis, we can mention that the variable access street/ building may indicate the need of construction work, which certainly will lead to an increase in costs and time. Only three variables, constructed area, access for people with disabilities mobility and price were considered by all experts, for at least one property. This shows that experts do not consider all the variables in the symbolic analysis either because some variables do not reflect their focuses or just by involuntary omission. All variables were considered qualitatively by three experts, a fact evidenced by the use of subjective terms and adjectives such as “desirable”, “above the desirable”, “well above the desirable level”, “below the desirable”, “excellent”, “good”, and others.

Tabulation was made of a hierarchy of prospected properties based on the results presented by the Realty Prospecting System the analysis of the experts, as shown in Table 3.

Hierarchy of prospected properties				
Building	**Realty Property System**	**Expert 1**	**Expert 2**	**Expert 3**
1	2nd	2nd	1st	3st
2	3rd	3rd	2nd	2nd
3	1nd	1st	2nd	1st
4	Disregard property	Discarded	2nd	Not feasible

Table 3. Hierarchy of prospected properties
Source: Carnasciali, A. M. S., 2007

Analyzing the forms filled by the experts the following considerations can be made:
a) The specialist 1 initially ranked properties 1 and 3 "tied". As a criterion for tiebreak he argued "... considering the small difference in rental value and that the need for adjustments is less costly for property 3 (located on ground floor), I believe that property 3 is our 1st choice." Even the expert, initially, reaching a tie, his final choice matched those given by the Realty Prospecting System. As this expert has been directly responsible for surveys of properties for HSBC for ten years, his analysis serves as a good parameter to evaluate the performance of the proposed system;
b) Opinions of expert 2 do not coincide with those given by the Realty Prospecting System and either with the ones mentioned by the other two experts. Interestingly, this expert was the one who considered most variables in the analysis. He pointed that the area of the building 4 is below the necessary, but does not disclose that this is a desirable item but not a cutline. As a specialist in property law with 5 years experience, in the role, he showed a trend to analyze failures of project as eliminatory letdowns, under considering the possibility of repair.
c) Expert 3 considered property 3 as his 1st option coinciding with the Realty Prospecting System. As for the 2nd and 3rd choice, the expert argues the property 1 "Property with desirable spot, at a corner, but with two floors..." and for building 2 “property with area above the desirable ... check the possibility to lease half of the building...". Therefore, 2nd choice was made based on assumptions, on the possibility of division. Comparing his choices and those made by the system, the result is different. However, it is important to emphasize that the system performs the analysis based on evidence and not on conjecture.

5. Conclusion

Integration of GIS and ES aims to help location decision-making for the banking sector. Through the GIS it was possible gathering the needed data for spatial analysis, storing the order of importance of spatial variables, and viewing a combination of the different data. The GIS has helped to expand the possibilities of spatial analysis for the selection of the region. The ES gives the definition of order of importance of symbolic variables, which were defined and accepted by all experts, and storages the sequence of reasoning, so that no variable was ignored or even forgotten during the process of selection of the spot. Some symbolic variables were, by the same specialist, now considered in the assessment of a property, sometimes not considered, showing lack of a sequence of reasoning. The task of selecting the best property was prepared by experts in a subjective manner. Realty Prospecting System shows, at the end of the consultation for each property a note that identifies which one obtained the highest score.

This work enables experts of Department of Real Estate Administration of HSBC Bank Brazil S/A - a comparison between the methods usually used and the integrated GIS and ES to help in location decision. According to the experts the use of the proposed system will benefit the professional involved with this task, providing technical bases for the analysis, and considering a large number of information necessary for the complete decision process, reducing potentially mistaken judgments, and the consequent risk of hiring buildings not suitable for the objectives outlined by the company.

Suggestions for the continuation of this work are:

a. Definition of areas of influence of the branches to identify possible blanks or overlaps in services. This would therefore be another spatial variable to be considered assisting in the selection of the region. However, there may be difficulties in dealing with this variable once its configuration may require reserved data such as the addresses of customers of each branch;
b. Implementation of options for reports of analyzed data for each property, as well as for generating graphs for scores achieved by each property;
c. Implementation of a system of Technical, Legal and Location Conformity Assessment of bank branches by integrating Geographic Information System and Expert System.

An emphasized recommendation is for the knowledge base of the ES to be increased, receiving information concerning the documentation of the property, including: approved project, property taxes, real estate registration, among others. Information concerning the documentation of the property, once considered simultaneously with the market value, constructed area, number of floors and others, could serve to prevent the hiring of a non conform property.

6. Acknowledgment

To HSBC Bank Brazil S/A – Department of Real Estate Administration for collaboration in developing the work.

To CAPES - Coordination for Improvement of Higher Education Personal - for granting the scholarship.

To CNPq - National Counsel of Technological and Scientific Development - for granting the Research scholarship - Process 3088892/2008-9.

To the Post Graduation Course in Geodesic Sciences of UFPR – The Federal University of Paraná – for financially supporting the publication.

7. References

Aronoff, S. (1989). *Geographic information systems: a management perspective*. WDL, ISBN: 0921804911, Ottawa.

Burrough, P.A. & Mcdonell, R. (1998). *Principles of Geographical Information System for Land Resource Assessment*. Claderon Press, Oxford. ISBN : 978-019823366, Oxford.

Carnasciali, A. M. S. (2007). *Integração de sistemas de informações geográficas e sistema especialista visando auxiliar a tomada de decisão locacional do setor bancário.* Dissertação (Mestrado em Ciências Geodésicas) – Universidade Federal do Paraná, Curitiba.

Eldrandaly, K., Eldin, N. & Sui, D. Z. (2003). A COM-based Spatial Decision Support System for Industrial Site Selection. *Journal of Geographic Information and Decision Analysis*. Vol. 7, No. 2, p. 72-92, ISSN: 1480-8943

Macdonald, M. L. (1996). A multi-attribute spatial decision support system for solid waste planning. *Computers, Environment and Urban Systems*. Vol. 20, No. 1, January 1996, p. 1-17, ISSN: 0198-971

Maguire, D.J. (1991*). Geographical information systems : principles and applications.* Longman Sci & Techn, ISBN: 0582056616, Harlow.

Rodriguez-Bachiller, A. & Glasson, J. (2004). *Expert Systems and Geographical Information Systems for Impact Assessment.* Taylor & Francis, ISBN: 0–415–30725–2, London.

Waterman, D. A. (1986). *A Guide to Expert Systems*. Addison-Wesley, ISBN: 0201083132, Canadá.

Zhu, X., & Aspinall, R. J. & Healey, R. G. (1996). ILUDSS: A knowledge-based system for strategic land-use planning. *Computers and Electronics in Agriculture*. .Vol. 15, No 4, October 1996, p. 279-301. ISSN: 0168-1699

12

A Web-Based Data Management and Analysis System for CO_2 Capture Process

Yuxiang Wu and Christine W. Chan
Energy Informatics Laboratory, Faculty of Engineering, University of Regina
Regina, Saskatchewan, S4S 0A2,
Canada

1. Introduction

Fossil fuels constitute a major energy resource for Canada. In 2002 alone, the production of oil, gas and coal contributed over $30 billion to the Canadian economy.

Fossil fuel is presently the world's most abundant, economical and reliable fuel for energy production. However, the industry now faces a major challenge because the production of fossil fuels including coal, crude oil and gas, and the processes currently used for energy production from such fuels, can have adverse environmental consequences. Hence, along with the positive economic advantages of energy production using fossil fuels come the responsibility of mitigating the consequent adverse environmental and climate-change impacts (Harrison et al., 2007).

Carbon capture and storage (CCS) is an approach for reducing carbon dioxide (CO_2) emissions to the environment by capturing and storing the CO_2 gas instead of releasing it into the air. The application of CCS to a modern conventional power plant could reduce CO_2 emissions to the atmosphere by approximately 80-90% compared to a plant without CCS (IPCC, Metz, & Intergovernmental Panel on Climate Change Working Group III, 2005). CO_2 capture technologies mainly include: chemical absorption, physical absorption, membrane separation and cryogenic fractionation. Among these technologies, chemical absorption of CO_2 is one of the most mature technologies because of its efficiency and low cost.

The highly complex CO_2 absorption process generates a vast amount of data, which need to be monitored. However, industry process control systems do not typically incorporate operators' heuristics in their intelligent control or data analysis functionalities. Our objective is to construct an intelligent data management and analysis system that incorporates such human experts' heuristics. The Data Analysis Decision Support System (DADSS) for CO_2 capture process reported in (Wu & Chan, 2009) is a step towards filling this gap in automated control systems. However, the DADSS is a standalone PC-based system with limited flexibility and connectivity. In this paper we present a web-based CO_2 data management and analysis system (CO_2DMA), which overcomes these limitations.

The system presented in this paper was built based on data acquired from the Pilot Plant CO_2 capture process of the International Test Centre for CO_2 capture (ITC), located at the University of Regina in Saskatchewan, Canada. The CO_2 capture process at the ITC is monitored and controlled by the DeltaV system (Trademark of Emerson Process

Management, U.S.A), which adopts the technology of Object-Linking and Embedding (OLE) for Process Control (OPC). OPC standards are widely used in industry process control and manufacturing automation applications ("OLE for process control," n.d.). More detailed information about OPC will be provided later.

The paper is organized as follows: Section 2 presents some background literature on decision support systems used for problem solving in engineering. Section 3 gives some background on knowledge acquisition and knowledge representation in the process of developing the web-based carbon dioxide data management and analysis system called CO_2DMA. Section 4 discusses software engineering techniques used in system development. Section 5 presents some sample test runs of CO_2DMA. Section 6 concludes the paper and presents some directions for future work.

2. Background

2.1 Amine-based CO_2 capture process

The purpose of CO_2 capture is to purify industrial gas streams before they are released into the environment and produce a concentrated stream of CO_2. Current post-combustion CO_2 capture technologies mainly include: chemical absorption, physical absorption, membrane separation, and cryogenic fractionation (Riemer, 1996). The selection of a technology for a given CO_2 capture application depends on a variety of factors, such as capital and operating costs of the process, partial pressure of CO_2 in the gas stream, extent of CO_2 to be removed, purity of desired CO_2 product, and sensitivity of solutions to impurities (i.e. acid gases and particulates, etc.) (White et al., 2003).

In recent years, chemical absorption has become the most commonly used technology for low concentration CO_2 capture, and amine solvents are the most widely used for chemical absorption. In the amine-CO_2 reaction, CO_2 in the gas phase dissolves in an aqueous amine solvent; the amines react with CO_2 in solution to form protonated amine (AMH+), bicarbonate ion (HCO3-), carbamate ($AMCO_2$-), and carbonate ion (CO32-) (Park et al., 2003). The system described in this paper is constructed based on ITC's CO_2 capture process, which primarily implements this chemical absorption as an industrial process for pilot run and research purposes.

Fig. 1 shows a process flow diagram of the CO_2 capture plant. Before the CO_2 is removed, the flue gas is pre-treated in the inlet gas scrubber, where the flue gas is cooled down, and particulates and other impurities such as sulfur oxide (SOx) and nitrogen oxide (NOx) are removed as much as possible. The pre-treated flue gas is passed into the absorption column by an inlet-gas feed blower, which provides the necessary pressure for the flue gas to overcome the pressure drop in the absorber. In the absorber, the flue gas and lean amine solution contact each other counter-currently. With the high temperature steam provided by the boiler, the amine selectively absorbs CO_2 from the flue gas. The amine solution carrying CO_2, which is called CO_2-rich amine, is pumped to the lean/rich heat exchanger, where the rich amine is heated to about 105 °C by means of the lean amine solution. The heated CO_2-rich amine enters the upper portion of the stripper. Then the CO_2 is extracted from the amine solution, which is now the lean amine solution. Most of the lean amine solution returns to the lean amine storage tank and then recycles through the process for CO_2 absorption. A small portion of it is fed to a reclaimer, where the degradation by-products and heat stable salts (HSS) are removed from the amine solution. The non-regenerable sludge is left behind in the reclaimer and can be collected and disposed. The CO_2 product

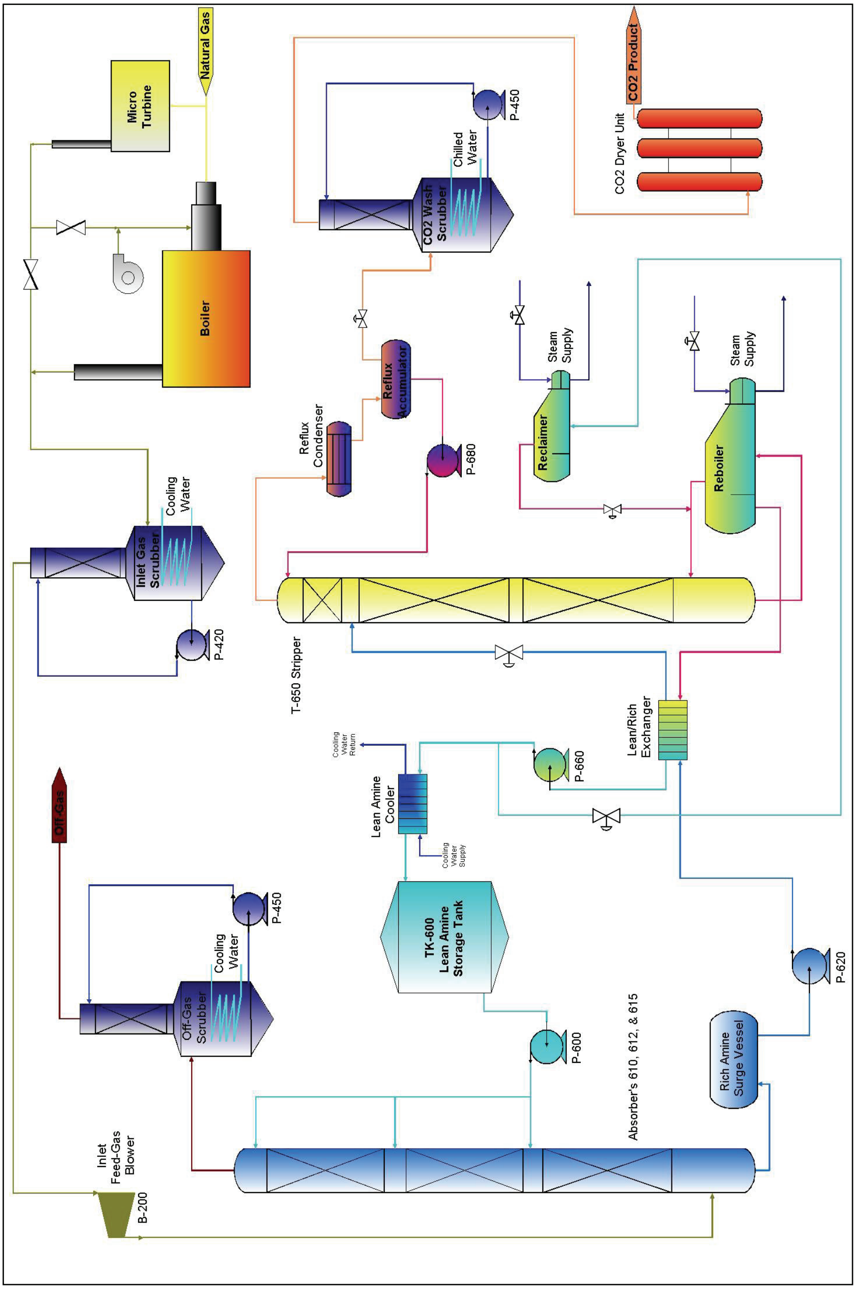


Fig. 1. Process flow diagram of the CO_2 capture process

and water vapour from the top of the stripper is passed through a reflux condenser. Most water is condensed inside the condenser, and the residual amine solvent is passed back to the stripper column reflux section so as to desorb CO_2 again. The CO_2 product enters a CO_2 wash scrubber, where the CO_2 gas is cooled to the desired temperature of 4 °C. From there, the CO_2 can be vented into the atmosphere or passed through a dryer and purification unit to produce food grade quality CO_2.

2.2 Decision support system

The decision support system (DSS) is a computerized system built for aiding the process of decision making (Turban, 1993). Decision support systems can supplement human cognitive deficiencies by integrating different sources of information and providing intelligent access to relevant knowledge. They can also help people select among well-defined alternatives based on formal or theoretical criteria and methods from engineering economics, operations research, statistics and decision theory. For problems that are intractable by formal techniques, artificial intelligence methods can also be employed (Druzdzel & Flynn, 1999).

Decision support systems (DSS) have been widely used in diverse domains, including business, engineering, the military, medicine and industrial areas. Some applications of decision support systems in process industries are described as follows. Geng et al. (2001) presented a knowledge-based decision support system that aids users select petroleum contaminant remediation techniques based on user-specified information. Szladow and Mills (1996) described a DSS that presents the application work-flows used for training new operators in five heavy industries including iron and steel, cement, mining and metallurgy, oil and gas, and pulp and paper. Flores et al. (2000) described an intelligent system that links multiple anaerobic systems for wastewater treatment to a common remote central supervisor via wide area networks. The local control systems have a hybrid structure, which is comprised of algorithmic routines for data acquisition, signal preprocessing, and calculation of plant operation parameters. Kritpiphat et al. (1996) developed an expert DSS for supervisory monitoring and control of a water pipeline network in a prairie city in Canada.

2.3 Web-based system

In the past decade, the World Wide Web has successfully demonstrated how the internet technologies can support information sharing and knowledge exchange. Technically, a web-based system can be defined as an application or service which resides in a server remotely or locally. The application or service can be accessed using a web browser from anywhere via the internet.

In the system requirement analysis and design stage of this project, a stand-alone application was considered but not adopted because of a number of reasons: (1) The stand-alone system relies on a particular data file as its data source. This is a limitation because it is not a flexible format that can be useful in future data analysis. (2) Knowledge and data sharing through a specific file will be difficult. (3) System and data access is limited to the station in which the system has been installed.

Due to these limitations, a web-based system was considered and adopted due to the following benefits (Liu & Xu, 2001):

- A web-based system has a reduced product development cycle time because of the increased collaboration among all areas of an organization and its supply chain, and the easy access to system information;

- A web-based system can make use of a full library of utilities, which help the developers avoid tedious coding and enable sharing of code common to various modules;
- A web-based system facilitates the uses in accessing the information of data source. When properly implemented, the web-based system can simplify many day-to-day user operations by managing and automating routine tasks, such as searching and completing data reports;
- Management of a project for constructing a web-based system is easier because a web-based system allows the system developers or maintainers to track the status of the system more effectively and facilitates validating the development work.

Therefore, a web-based CO_2 data management and analysis system was designed and developed; the knowledge engineering process for the system is described as follows.

3. Knowledge engineering of CO_2DMA

The knowledge engineering process (Jack Durkin & John Durkin, 1998) for building CO_2DMA involves the two primary processes of knowledge acquisition and knowledge representation.

3.1 Knowledge acquisition

The knowledge useful for developing a knowledge-based system refers to the problem solving activities of a human expert. Knowledge acquisition (KA) is the process of elucidating, analyzing, transforming, classifying, organizing and integrating knowledge and representing that knowledge in a form that can be used in a computer system (Druzdzel & Flynn, 1999) (Geng et al., 2001).

The objective of knowledge acquisition in this project is to obtain specific domain knowledge on how to filter and analyze CO_2 data. The knowledge was acquired during interviews with the expert operator, who is the chief engineer of the ITC pilot plant. Relevant knowledge on the existing process of filtering and analyzing CO_2 data includes:

- Data points or tags used: data obtained for 145 tags need to be analyzed for monitoring the CO_2 capture process.
- Steps or methods for filtering data: The original CO_2 data captured by the DeltaV system suffer from the deficiencies of being: (1) incomplete (lacking attribute values or certain attributes of interest, or containing only aggregate data), (2) noisy (containing errors, or outlier values that deviate from the expected), and (3) inconsistent (containing discrepancies in the tag names used to label attributes). In other words, there are reported errors, unusual values, redundant values and inconsistencies in the data recorded for some transactions. The data was pre-filtered by filling in missing values, eliminating noisy data, identifying or removing outliers, and resolving inconsistencies. The pre-filtering procedure involves the following four steps:

```
Step 1:  IF          Gas flow rate into absorber <= 4.4 OR
                     Gas flow rate into absorber >= 4.6,
         THEN        Delete this row of data

Step 2:  IF          Heat Duty <= 20000 – (20000 * 0.1) OR
                     Hear Duty >= 100000 + (100000 * 0.1),
         THEN        Delete this row of data

Step 3: IF          Input $CO_2$ fluid gas <= 9.0 OR
                     Input $CO_2$ fluid gas >= 12,
         THEN        Delete this row of data

Step 4: IF          Rebuilder steam flow rate >= 70000 – (70000 * 0.05) AND
                     Rebuilder steam flow rate <= 70000 + (70000 * 0.05) AND
                     Heat Duty >= 70000 – (70000 * 0.2) AND
                     Heat Duty <= 70000 + (70000 * 0.2),
         THEN        Keep this row of data
         ELSE IF     $CO_2$ production rate >= 0.7 – (0.7 * 0.1) AND
                     $CO_2$ production rate <= 0.72 + (0.72 * 0.1),
         THEN        Keep this row of data
         ELSE        Delete this row of data
```

Where	*Gas flow rate into absorber*:	Absorber inlet gas flow rate (1000 m^3/day);
	Heat Duty:	Energy used (BTU/lb-mole CO_2);
	Input CO_2 fluid gas:	Absorber input fluid CO_2 gas (CO_2%);
	Rebuilder steam flow rate:	Steam from rebuilder flow rate (kg/h);
	CO_2 production rate:	CO_2 production rate (tones/day).

3.2 Knowledge representation

The knowledge obtained from the KA process was represented in a number of classes, which are referred to as knowledge components. The components are organized into the class hierarchy shown in Fig. 2. Most of the classes were implemented as tables in a database.

The class of Project represents the concept of a CO_2 data analysis project. A project contains a Profile, Data, and Subsets. When a new analysis of the data is initiated, a new project is created.

The class of Profile stores the tags and data filtering steps the user uses in one analysis case. The class of tag represents a component from which parameter values are obtained. A tag has a name, the area it is in, the path, a description, the units, minimum/maximum values, and a flag for whether or not the tag is being used. The class of Step represents a filtering step. Each step has a name, a description, and a flag for whether or not it is used.

The class of Data is represented as a row of values which is identified by the date. The values are augmented with an inUse flag and a comment field to record if a value has been filtered and the reason for performing the filtering.

The class of Subset is a portion of the data. A subset has a name, a list of tags selected by the user, and the actual data. The class of SubsetData is very similar to the class of Data, except that it consists of only subsets of data. SubsetData is a row of data in the subset and only contains the date and parameter values.

4. System development of CO_2DMA

This section presents the structure of the CO_2DMA, and several software engineering technologies that were used during system development.

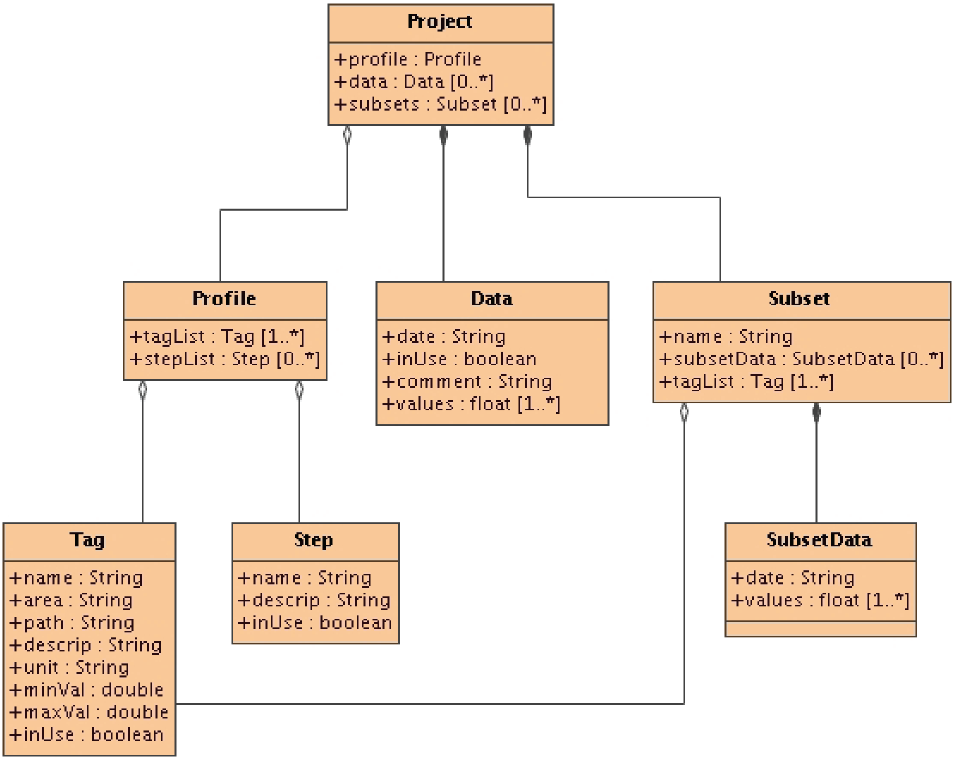


Fig. 2. Class Hierarchy of Knowledge Components

4.1 System structure

The CO_2DMA system consists of four main modules (see Fig. 3): (1) OPC Historical Data Access (HAD) Server module, (2) OPC Data Transporter module, (3) Database Server module, and (4) Web Server module. The OPC HAD Server usually resides in the same computer as the process control system, which is implemented in the DeltaV system at ITC. It is the repository where process data are stored, and which can be accessed only by programs built according to the HDA standards. The OPC Data transporter is a C# (Microsoft® software) program that runs along with the OPC HDA Server in the background. It continually reads data from the OPC HDA Server and converts the data into the appropriate types in order to transfer them into the Database Server. The Web Server component of the system is responsible for communicating with clients through the internet. The clients send request to and retrieve data from the Web Server. Both communication and data transfer are based on the HyperText Markup Language (HTML).

4.2 OPC and OPC transporter

OPC, which stands for Object-Linking and Embedding (OLE) for Process Control, is basically a series of standard specifications ("The OPC Foundation - Dedicated to Interoperability in Automation," n.d.). The OPC standard specifications support

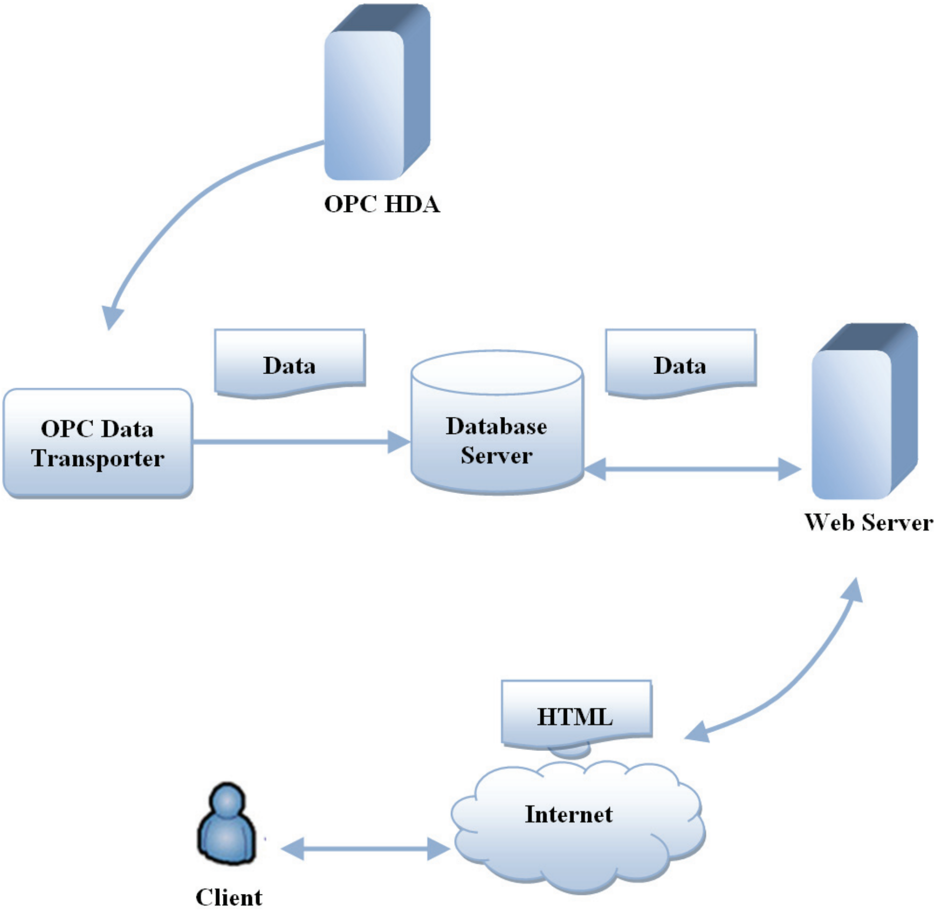


Fig. 3. Structure of CO_2DMA

communication of real-time plant data between control devices from different manufacturers ("OLE for process control," n.d.). The OPC Foundation maintains the standards. Since the foundation was created, more standards have been added.

The purpose of using OPC was to bridge Windows (Microsoft® software) based applications with process control hardware and software applications because the open standards support a consistent method of accessing field data from plant floor devices. The OPC servers define a common interface which can support different software packages to access data derived from the control devices.

Despite its advantages, the OPC technology suffers from two main weaknesses, which the ITC operators found can impede smooth operations of the CO_2 capture process:

The OPC technology, which includes the OPC servers and client applications, are developed based on the windows platform. This presents a problem when data and knowledge need to be shared with an application that is developed on a non-windows platform. In other words, interoperability among different platforms is not supported.

Only applications that support OPC protocols can access the data in the OPC Historical Data Access (HDA) Server, which is where the DeltaV data reside. Hence, data manipulation and analysis had to be handled by OPC client applications, and data cannot be reused by other computational tools that do not share OPC interfaces.

A generic database can address these limitations because it enables retrieving data from the real-time control system and storing them. Since we believe a generic database can render our system more flexible and the data reusable, the component called OPC Data Transporter was constructed for accessing, converting and sending data from the OPC HDA Server to the generic database. The OPC Data Transporter is an OPC client application written in C# (Microsoft® software) using the Historical Data Access (HDA) common library. Currently the transporter runs as a background program within the same machine as the DeltaV control system. It can also reside in a remote machine which physically connects to the control system. In either case, the data will be periodically captured from the OPC HDA Server and converted to the correct data type, then stored in the generic database. This approach supports isolating and protecting the process control system from outside interference, while enabling sharing of data and other useful information from the control system through the data repository implemented as the database.

4.3 Web server development

The Web Server plays a key role in the enhanced version of the DADSS because it acts as an intermediary between the database component and the user on the internet. The server was constructed using the LAMP software bundle, which includes:

- Linux, a Unix-like computer operating system.
- Apache, an open source HTTP Server.
- MySQL (Trademark of MySQL AB), multi-user SQL database management system (DBMS).
- PHP (Hypertext Preprocessor), a computer scripting language originally designed for producing dynamic web pages.

This LAMP bundle has become widely popular since its inception by Michael Kunze in 1998 because this group of free software could provide a viable alternative to commercial packages (“LAMP (software bundle) - Wikipedia, the free encyclopedia,” n.d.). Therefore, the LAMP bundle has been adopted for developing the Web server.

Usually the most time consuming part of building a web server is to program the entire site including design of the user interface as well as construction of the background logical layer. This process was often conducted in an ad hoc manner, based neither on a systematic approach, nor quality control and assurance procedures. Recently, different types of web application frameworks supporting different languages have been built. A web application framework is a software framework that is designed for supporting the development of dynamic websites, web applications and services; the framework is intended to simplify the overhead associated with common activity procedures in web development. The general framework usually provides libraries for database access, template frameworks, session management and code reuse.

In development of the web server, CakePHP (trademark of Cake Software Foundation) was adopted as the basic framework because of its detailed documentation and ease of use. Based on CakePHP, the system structure of the web server was designed and developed as shown in Fig. 4.

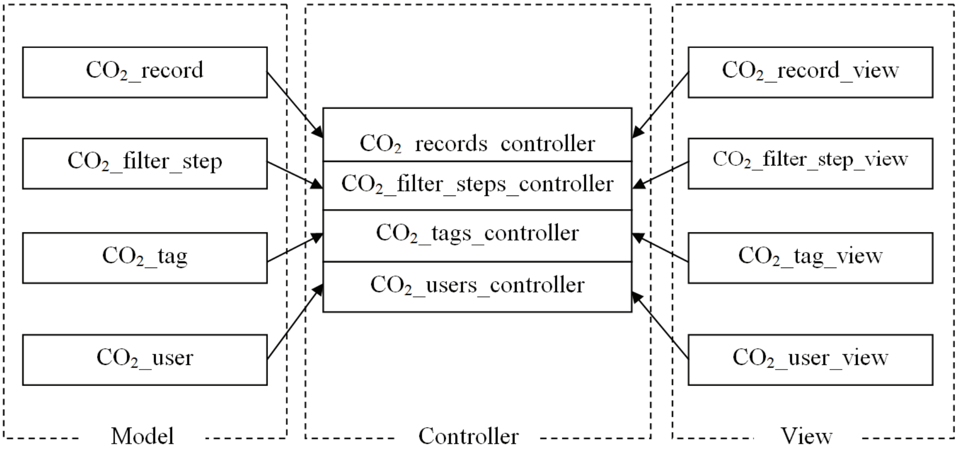


Fig. 4. Web Server Structure

As shown in Fig.4, the structure of the system follows the Model-View-Controller (MVC) architectural pattern (Gamma et al., 1995), which is one of the most commonly adopted application structural models in software development. Recently the model has become widely used in web application development. In the web server system, a model represents a particular database table, and its relationships to other tables and records. The Model also consists of data validation rules, which are applied when the model data are inserted or updated. The View represents view files, which are regular HTML files embedded with PHP code. This provides users with the web page display. The controller handles requests from the server. It takes user input which includes the URL and POST data, applies business logic, uses Models to read and write data to and from databases and other sources, and lastly, sends output data to the appropriate view file ("Basic Principles of CakePHP," n.d.). This system structure has the advantages of (1) modularizing the code and making it more reusable, maintainable, and generally better; and (2) encapsulating the knowledge captured from the expert operator, which was translated into procedures and methods using an object-oriented representation.

4.4 System security

With the proliferation of web-based applications, security is now one of the most crucial considerations in system development. This is also true for the CO$_2$DMA because it needs to be protected from outside interference. A number of steps were taken to ensure security:

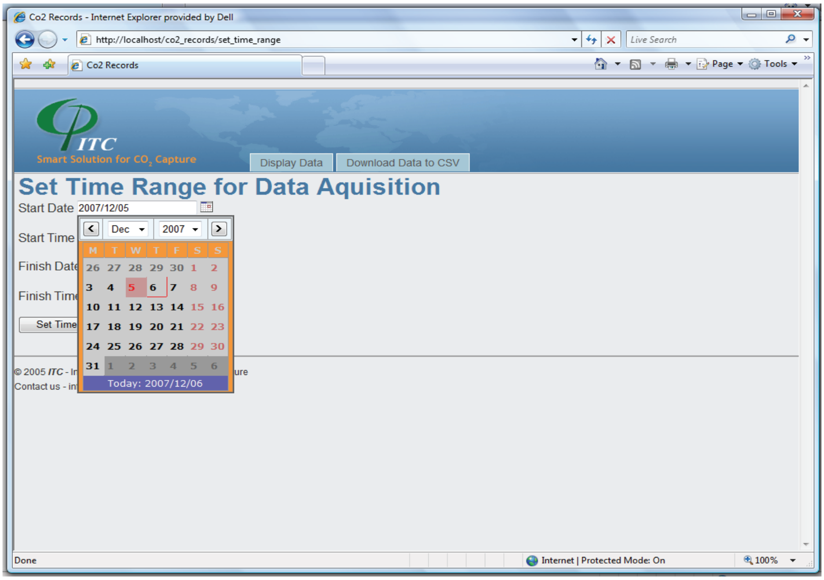
- The website can only be accessed by particular users with the correct user name and password. Accesses are filtered by IP address.
- Data transferred between the user's device and the web server will be encrypted by Secure Sockets Layer (SSL).
- The web server only has privileges to view data from the database server. The connection from the Web Server to the Database Server is read-only; therefore the Web Server cannot do any modifications to the Database.
- The OPC Data transporter module is responsible for transferring data from the DeltaV Server to the Database Server, but it can only access the HDA server and not the DeltaV process control system.

- Hardware Firewalls are configured between:
  a. the CO_2 capture process control system and the Database module
  b. the Database module and the Web Server module

5. Sample session

A sample session of running the system demonstrates how the CO_2DMA assist the operator in accessing and filtering data generated from the CO_2 capture process. When the user enters the website of the system, he/she can select either to display or download the data in order to obtain information about the CO_2 capture process. Hitting the button for display or download would trigger a display of the calendar, as shown in Fig. 5. The user can select the date and time range for which data are required. In response to the user's request, the system displays the entire set of tags, which stand for all the equipment in the process. The tag selection table, as shown in Fig. 6, also includes detailed descriptions such as area, path and unit of each tag. After selecting the tags, the user chooses the pre-filtering steps that are applicable to the data, as shown in Fig. 7. Finally the system displays the filtered data, which are presented in either the browser's viewable format (Fig. 8) or in CSV format and can be downloaded to the user's local machine.

The difference between the unfiltered data and filtered data can be revealed by examining the two sets of data on the sample variables of (1) CO_2 production rate, represented by 'Wet CO_2 out from v-680' and (2) 'Heat Duty', which were selected from the 145 variables monitored by the system. Two trend lines that approximate the data are drawn as shown in Fig. 9 and Fig. 10. The points in the plot of the unfiltered data in Fig. 9 are more scattered because of the high volume of noisy data. After filtering by CO_2DMA, more than 60 rows of noisy data were filtered out from the 590 rows, and the data points are closely clustered as shown in Fig. 10.


Co2 Records - Internet Explorer provided by Dell
http://localhost/co2_records/set_time_range
Live Search
Co2 Records
Page
Tools
ITC
Smart Solution for CO_2 Capture
Display Data
Download Data to CSV
Set Time
© 2005 ITC - In
ure
Contact us - in
Done
Internet | Protected Mode: On
100%

Set Time Range for Data Aquisition

Start Date 2007/12/05

Start Time

Finish Date

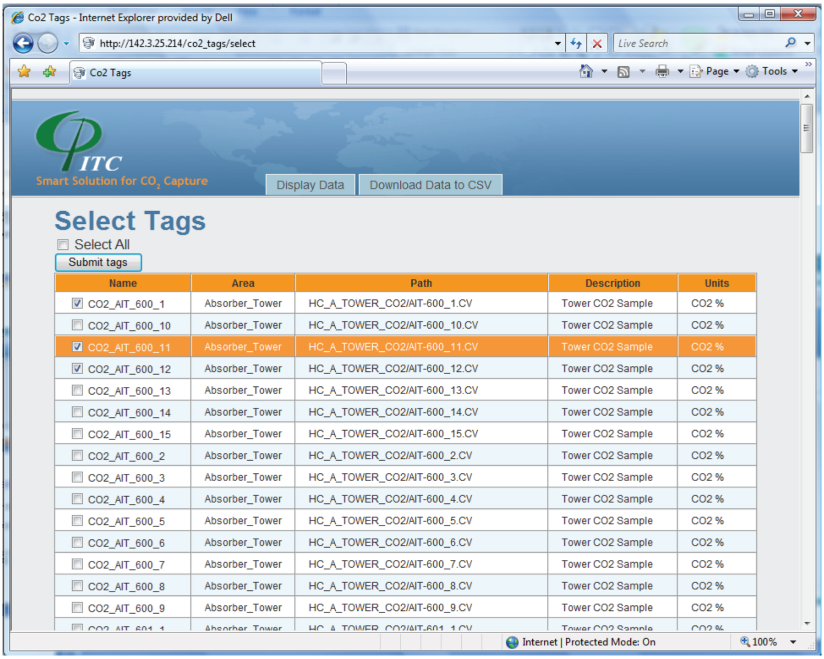
Finish Time

Dec 2007

M	T	W	T	F	S	S
26	27	28	29	30	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31	1	2	3	4	5	6

Today: 2007/12/06

Fig. 5. Date range selection


Co2 Tags - Internet Explorer provided by Dell
http://142.3.25.214/co2_tags/select
Live Search
Co2 Tags
Page
Tools
ITC
Smart Solution for CO_2 Capture
Display Data
Download Data to CSV
Internet | Protected Mode: On
100%

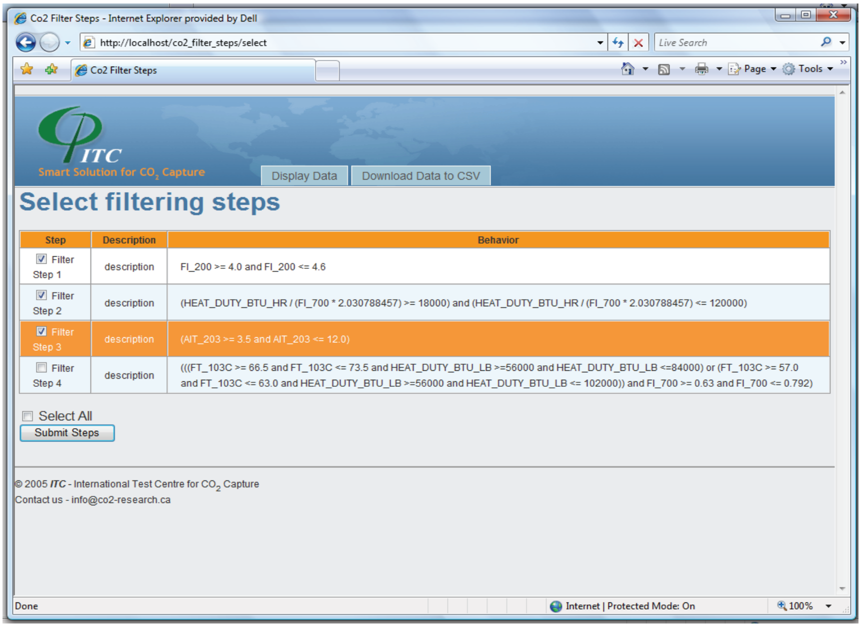
Select Tags

□ Select All

Submit tags

Name	Area	Path	Description	Units
☑ CO2_AIT_600_1	Absorber_Tower	HC_A_TOWER_CO2/AIT-600_1.CV	Tower CO2 Sample	CO2 %
□ CO2_AIT_600_10	Absorber_Tower	HC_A_TOWER_CO2/AIT-600_10.CV	Tower CO2 Sample	CO2 %
☑ CO2_AIT_600_11	Absorber_Tower	HC_A_TOWER_CO2/AIT-600_11.CV	Tower CO2 Sample	CO2 %
☑ CO2_AIT_600_12	Absorber_Tower	HC_A_TOWER_CO2/AIT-600_12.CV	Tower CO2 Sample	CO2 %
□ CO2_AIT_600_13	Absorber_Tower	HC_A_TOWER_CO2/AIT-600_13.CV	Tower CO2 Sample	CO2 %
□ CO2_AIT_600_14	Absorber_Tower	HC_A_TOWER_CO2/AIT-600_14.CV	Tower CO2 Sample	CO2 %
□ CO2_AIT_600_15	Absorber_Tower	HC_A_TOWER_CO2/AIT-600_15.CV	Tower CO2 Sample	CO2 %
□ CO2_AIT_600_2	Absorber_Tower	HC_A_TOWER_CO2/AIT-600_2.CV	Tower CO2 Sample	CO2 %
□ CO2_AIT_600_3	Absorber_Tower	HC_A_TOWER_CO2/AIT-600_3.CV	Tower CO2 Sample	CO2 %
□ CO2_AIT_600_4	Absorber_Tower	HC_A_TOWER_CO2/AIT-600_4.CV	Tower CO2 Sample	CO2 %
□ CO2_AIT_600_5	Absorber_Tower	HC_A_TOWER_CO2/AIT-600_5.CV	Tower CO2 Sample	CO2 %
□ CO2_AIT_600_6	Absorber_Tower	HC_A_TOWER_CO2/AIT-600_6.CV	Tower CO2 Sample	CO2 %
□ CO2_AIT_600_7	Absorber_Tower	HC_A_TOWER_CO2/AIT-600_7.CV	Tower CO2 Sample	CO2 %
□ CO2_AIT_600_8	Absorber_Tower	HC_A_TOWER_CO2/AIT-600_8.CV	Tower CO2 Sample	CO2 %
□ CO2_AIT_600_9	Absorber_Tower	HC_A_TOWER_CO2/AIT-600_9.CV	Tower CO2 Sample	CO2 %
□ CO2_AIT_601_1	Absorber_Tower	HC_A_TOWER_CO2/AIT-601_1.CV	Tower CO2 Sample	CO2 %

Fig. 6. Tag Selection


Co2 Filter Steps - Internet Explorer provided by Dell
http://localhost/co2_filter_steps/select
Live Search
Co2 Filter Steps
Page
Tools
ITC
Smart Solution for CO_2 Capture
Display Data
Download Data to CSV
Done
Internet | Protected Mode: On
100%

Select filtering steps

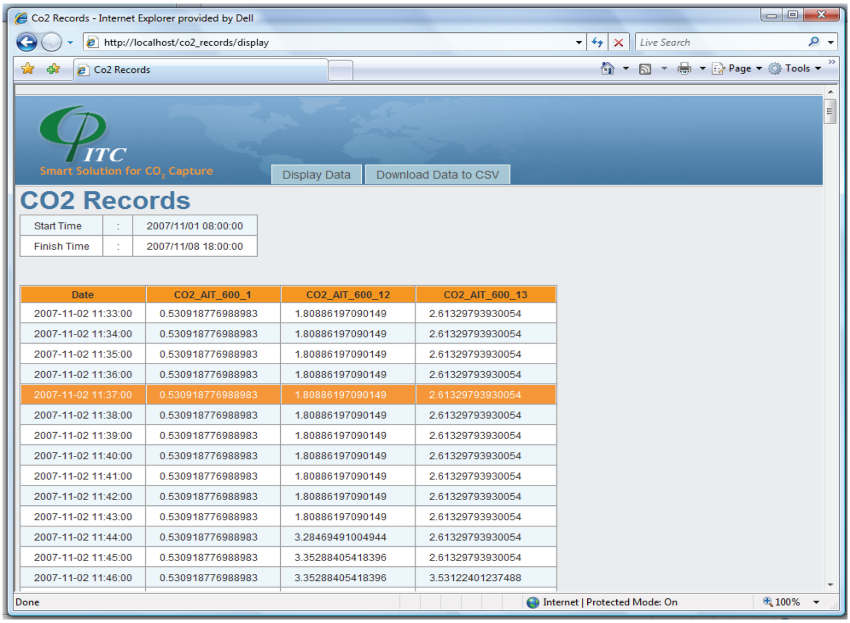
Step	Description	Behavior
☑ Filter Step 1	description	FI_200 >= 4.0 and FI_200 <= 4.6
☑ Filter Step 2	description	(HEAT_DUTY_BTU_HR / (FI_700 * 2.030788457) >= 18000) and (HEAT_DUTY_BTU_HR / (FI_700 * 2.030788457) <= 120000)
☑ Filter Step 3	description	(AIT_203 >= 3.5 and AIT_203 <= 12.0)
□ Filter Step 4	description	(((FT_103C >= 66.5 and FT_103C <= 73.5 and HEAT_DUTY_BTU_LB >=56000 and HEAT_DUTY_BTU_LB <=84000) or (FT_103C >= 57.0 and FT_103C <= 63.0 and HEAT_DUTY_BTU_LB >=56000 and HEAT_DUTY_BTU_LB <= 102000)) and FI_700 >= 0.63 and FI_700 <= 0.792)

□ Select All

Submit Steps

© 2005 *ITC* - International Test Centre for CO_2 Capture
Contact us - info@co2-research.ca

Fig. 7. Selection of filtering steps



Co2 Records - Internet Explorer provided by Dell
http://localhost/co2_records/display
Live Search
Co2 Records
Page
Tools
ITC
Smart Solution for CO₂ Capture
Display Data
Download Data to CSV
Done
Internet | Protected Mode: On
100%

CO2 Records

Start Time	:	2007/11/01 08:00:00
Finish Time	:	2007/11/08 18:00:00

Date	CO2_AIT_600_1	CO2_AIT_600_12	CO2_AIT_600_13
2007-11-02 11:33:00	0.530918776988983	1.80886197090149	2.61329793930054
2007-11-02 11:34:00	0.530918776988983	1.80886197090149	2.61329793930054
2007-11-02 11:35:00	0.530918776988983	1.80886197090149	2.61329793930054
2007-11-02 11:36:00	0.530918776988983	1.80886197090149	2.61329793930054
2007-11-02 11:37:00	0.530918776988983	1.80886197090149	2.61329793930054
2007-11-02 11:38:00	0.530918776988983	1.80886197090149	2.61329793930054
2007-11-02 11:39:00	0.530918776988983	1.80886197090149	2.61329793930054
2007-11-02 11:40:00	0.530918776988983	1.80886197090149	2.61329793930054
2007-11-02 11:41:00	0.530918776988983	1.80886197090149	2.61329793930054
2007-11-02 11:42:00	0.530918776988983	1.80886197090149	2.61329793930054
2007-11-02 11:43:00	0.530918776988983	1.80886197090149	2.61329793930054
2007-11-02 11:44:00	0.530918776988983	3.28469491004944	2.61329793930054
2007-11-02 11:45:00	0.530918776988983	3.35288405418396	2.61329793930054
2007-11-02 11:46:00	0.530918776988983	3.35288405418396	3.53122401237488

Fig. 8. Sample of filtered data

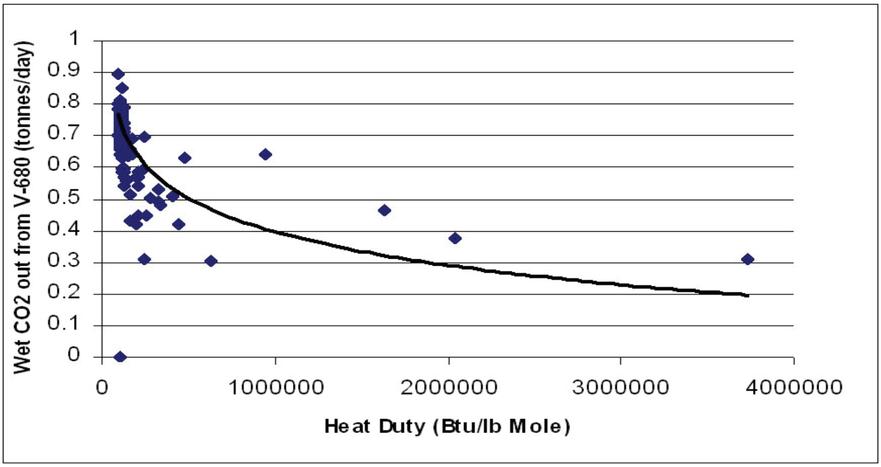


Fig. 9. Plot of data before filtering

6. Conclusion and future works

A web-based data management and analysis system for the CO_2 capture process called CO_2DMA has been developed. The system has a user friendly interface and therefore does not require a steep learning curve for the user. Since the system is built as a web service application, there is no need to install any software in the user's computer. By automatically

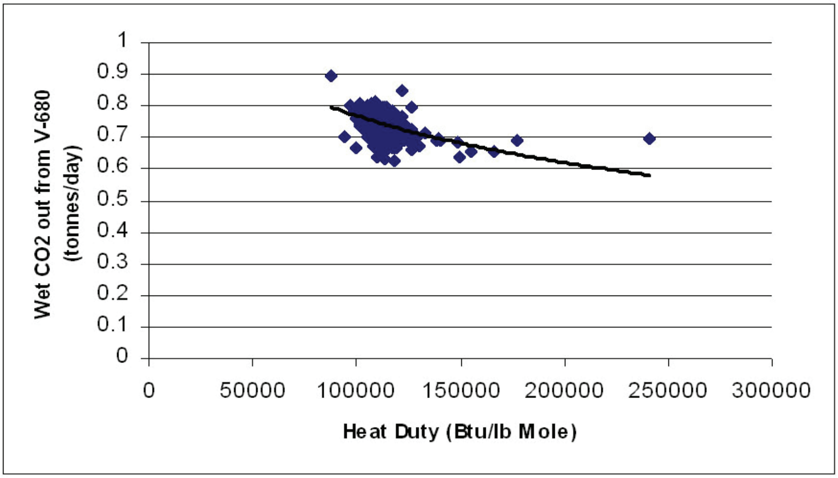


Fig. 10. Plot of data after filtering

filtering and processing hundreds of fields of raw data, the CO₂DMA frees users from having to perform data filtering manually; hence, it improves efficiency of the data filtering process.

Future work for enhancing system efficiency involves saving the user's preferred filtering procedures in a historical configuration file. With this enhancement, the user can simply retrieve their preferred configurations from the configuration file, thereby avoiding the step of selecting the filtering criteria. We also plan to add curve fitting and graphing functions to the system so that the filtered data can be processed for visual displays inside the system instead of being exported to Microsoft Excel (Trademark of Microsoft Office) for further charting. Automation of the data filtering step is only the first step in our research agenda. Future objectives include building system modules for analyzing the data for prediction, planning and control of the CO_2 capture process using artificial intelligence techniques.

7. Acknowledgement

The authors are grateful for the generous support of grants from the Canada Research Chair Program and Natural Science and Engineering Research Council of Canada. We would like to acknowledge the help from members of the Energy Informatics Lab, Robert Harrison and Chuansan Luo. We also would like to thank Dr. Paitoon Tontiwachwuthikul, Don Gelowitz and Dr. Raphael Idem for their valuable suggestions, technical guidance and insights.

8. References

Basic Principles of CakePHP. (n.d.). . Retrieved March 29, 2009, from http://book.cakephp.org/view/13/Basic-Principles-of-CakePHP.

Druzdzel, M. J., & Flynn, R. R. (1999). *Decision Support Systems. Encyclopedia of Library and Information Science. A. Kent*. Marcel Dekker, Inc.

Durkin, J., & Durkin, J. (1998). *Expert Systems: Design and Development* (p. 600). Prentice Hall PTR. Retrieved March 27, 2009, from http://portal.acm.org/citation.cfm?id=551328.

Flores, J., Arcay, B., & Arias, J. (2000). An intelligent system for distributed control of an anaerobic wastewater treatment process. *Engineering Applications of Artificial Intelligence*, *13*(4), 485-494. doi: 10.1016/S0952-1976(00)00015-4.

Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1995). *Design patterns. Elements of reusable object-oriented software*. Retrieved March 27, 2009, from http://adsabs.harvard.edu/abs/1995dper.book.....G.

Geng, L., Chen, Z., Chan, C. W., & Huang, G. H. (2001). An intelligent decision support system for management of petroleum-contaminated sites. *Expert Systems with Applications*, *20*(3), 251-260. doi: 10.1016/S0957-4174(00)00063-4.

Harrison, R., Yuxiang Wu, Nguyen, H., Xiongmin Li, Gelowitz, D., Chan, C., et al. (2007). A Decision Support System for Filtering and Analysis of Carbon Dioxide Capture Data. In *Electrical and Computer Engineering, 2007. CCECE 2007. Canadian Conference on* (pp. 1380-1383). doi: 10.1109/CCECE.2007.347.

IPCC, Metz, B., & Intergovernmental Panel on Climate Change Working Group III. (2005). *IPCC special report on carbon dioxide capture and storage: Special Report of the Intergovernmental Panel on Climate Change* (p. 431).

Kritpiphat, W., Tontiwachwuthikul, P., An, A., Chan, C. W., & Cercone, N. (1996). SUPERVISORY AND DECISION-SUPPORT SYSTEM FOR INTELLIGENT MONITORING AND CONTROL OF A PIPELINE NETWORK. In *First International Conference on Intelligent Systems in Process Engineering: Proceedings of the Conference Held at Snowmass, Colorado, July 9-14, 1995* (p. 355). American Institute of Chemical Engineers.

LAMP (software bundle) - Wikipedia, the free encyclopedia. (n.d.). . Retrieved March 29, 2009, from http://en.wikipedia.org/wiki/LAMP_(software_bundle).

OLE for process control. (n.d.). .

Park, J., Yoon, S. J., & Lee, H. (2003). Effect of Steric Hindrance on Carbon Dioxide Absorption into New Amine Solutions: Thermodynamic and Spectroscopic Verification through Solubility and NMR Analysis. *Environmental Science & Technology*, *37*(8), 1670-1675. doi: 10.1021/es0260519.

Riemer, P. (1996). Greenhouse gas mitigation technologies, an overview of the CO_2 capture, storage and future activities of the IEA Greenhouse Gas R&D programme. *Energy Conversion and Management*, *37*(6-8), 665-670. doi: 10.1016/0196-8904(95)00237-5.

Szladow, A. J., Ltd, L. T., & Mills, D. (1996). INTELLIGENT SYSTEMS IN HEAVY INDUSTRY. In *First International Conference on Intelligent Systems in Process Engineering: Proceedings of the Conference Held at Snowmass, Colorado, July 9-14, 1995* (p. 237). American Institute of Chemical Engineers.

The OPC Foundation - Dedicated to Interoperability in Automation. (n.d.). . Retrieved March 27, 2009, from http://www.opcfoundation.org/.

Tony Liu, D., & William Xu, X. (2001). A review of web-based product data management systems. *Computers in Industry*, *44*(3), 251-262. doi: 10.1016/S0166-3615(01)00072-0.

Turban, E. (1993). *Decision Support and Expert Systems: Management Support Systems* (p. 960). Prentice Hall PTR. Retrieved March 27, 2009, from http://portal.acm.org/citation.cfm?id=541815.

White, C. M., Strazisar, B. R., Granite, E. J., Hoffman, J. S., & Pennline, H. W. (2003). Separation and capture of CO_2 from large stationary sources and sequestration in geological formations--coalbeds and deep saline aquifers. *Journal of the Air & Waste Management Association (1995)*, *53*(6), 645-715.

Wu, Y., & Chan, C. W. (2009). A data analysis decision support system for the carbon dioxide capture process. *Expert Systems with Applications*, 36(6), 9949-9960. doi: 10.1016/j.eswa.2009.01.064.

13

Case Studies of Canadian Environmental Decision Support Systems

William Booty and Isaac Wong
Environment Canada
Canada

1. Introduction

This chapter will discuss two different decision support systems that we have developed for Canadian environmental applications. We will first discuss how these systems utilize data and models to solve domain-specific problems and focus on effectiveness rather than efficiency in the decision making processes. In particular we will discuss how they are useful in better understanding the complex interaction between land and water and how they also provide a method to make informed resource management decisions and that they require the integration of scientific data, information, models and knowledge across multi-media (air, land and water), multi-disciplines and diverse landscapes. We will discuss how modelling is an important asset of any environmental decision support system (EDSS), particularly considering the high cost of full scale field work. Modelling presents a cost effective approach to assess the impact on the environment.

We will discuss how a typical EDSS needs to be developed to address the issues of linking multi-media models at different geospatial scales, how it provides interfaces that can accept, select, link and recalibrate discipline-specific component models, and how it can seek optimal solutions for a given domain problem.

We will also discuss that very often the EDSS is built around the concept of a management user interface to assist policy makers in their decision making. The technical users employ other tools to build model inputs, execute, and calibrate and validate the models while the management or policy makers view the inputs and outputs of the system that the technical users have built. This will allow management to investigate the analytical results based on robust science built by the researchers. Key functionality includes mapping and visualization of the results, scenario gaming and key statistical analyses of the results.

The first example we will discuss is the Environmental Effects Modelling Statistical Assessment Tools Decision Support System. We will discuss how it provides a user-friendly data analysis, display and decision support tool for Canada's federal environmental effects monitoring program for pulp and paper and mining industries. We will describe how the tool allows the assessment of the effects of effluent from industrial or other sources on fish and benthic populations. We will explain that in many of our EDSS systems, it is coupled with artificial intelligence such as expert systems to guide the users in the right direction. We will explain how the results are used in assessing the adequacy of existing regulations for protecting aquatic environments. We will explain how the design of such an EDSS has benefited from significant input from scientists, researchers, other end-users, system developers, modellers and Geographic Information System (GIS) specialists. The integration includes data, maps and models with user-friendly tools, including data input/output views, map input/output views, and modelling result views for interpretation, further analysis, conclusion and recommendation with the aid of the expert system approach.

The second example of an EDSS is one that has been developed by Environment Canada to provide policy makers with a tool to help in examining management options for dealing with the impacts of land use on water for agricultural issues in Canada. The system deals with both temporal and spatial consistency among component models, where the output from one model is used as input to another in a sequence of linked calculations. In this example, the dynamic landscape model generates land use maps for various land use scenarios that can be used either in a single storm event non-point source pollutant model such as the Agricultural Non-Point Source Pollutant Model (AGNPS) (Young et al., 1987), or in a continuous time non-point source pollutant model such as the Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998). It also provides the ability for scenario gaming, and testing, pollutant source tracing and the determination of optimal solutions. Examples will be provided of its application to a watershed in Ontario Canada.

Finally we will summarize the effectiveness of these systems and some insights as to how they might be improved and future directions.

Understanding complex environmental problems and making informed resource management decisions requires the integration of scientific data and knowledge across multiple disciplines and diverse landscapes. Ever increasing demands for timely, accurate and spatially explicit information require environmental modellers to deploy the latest information technology to provide decision support for various departmental priorities, such as global climate change, point source and non-point source pollution, lake eutrophication, biodiversity and ecosystem sustainability.

Decision Support Systems (DSS) (Alter, 1980) are computer-based interactive human-computer decision-making systems that assist policy makers in decision making processes. These systems utilize data and models to solve domain-specific problems and focus on effectiveness rather than efficiency in decision making processes. They also make informed resource management decisions and require the integration of scientific data, information, models and knowledge across multi-media (air, land and water), multi-disciplines and diverse landscapes in better understanding complex environmental issues.

In order for any decision support system to be a success, a proper design process is critical. The design of any environmental decision support system (EDSS) should come from a diversified functional group. They are scientists, environmental modellers, decision support system developers, computer programmers and component specialists such as Geographic Information Systems. Each of them contributes certain aspects of the system and how all the pieces fit into the system seamlessly. The system's blueprint should come from scientists who understand what is most required. Figure 1 depicts a schematic of an EDSS concept. At a glance, one can see that this kind of system offers a generic framework to integrate data, text, maps, objects, images, videos, environmental models and knowledge with user-friendly tools, including database management systems, mapping systems, visualization, advanced statistics, analytical functions and expert systems/artificial intelligence tools to produce outputs for interpretation, integration, post analysis and recommendation.

An EDSS should be able to handle data rich and poor situations. It should be functionality rich so the EDSS users not only perform the existing analytical routines but also expand to incorporate new ideas. Thus, a good EDSS should consist of the following functions and features:

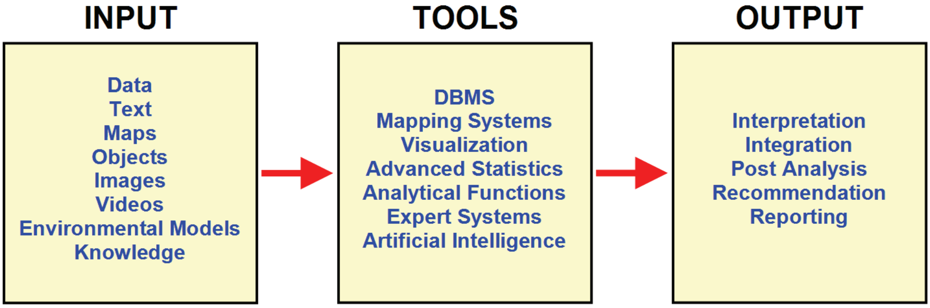


Fig. 1. Conceptual diagram of an environmental decision support system

1. User-friendly interface
2. Integrate and apply multiple tools such as environmental models, database management system, geographical information system, advanced statistics, artificial intelligence and, visualization in the analytical process
3. Assist users in integrating and inspecting complex and multiple input data and output results files that are at different spatial and temporal scales
4. Simplify the existing analytical process and allow room for any expansion
5. Provide an avenue to explore "what-if" option to compare scenarios

2. Case 1: environmental effects monitoring decision support system

2.1 Background

In Canada, pulp and paper and mining industries are currently required to conduct EEM under the Pulp and Paper Effluent Regulations (PPER) and Metal Mining Effluent Regulations (MMER). The Canadian EEM program is the first regulated, mandatory program of this magnitude in the world. The EEM involves iterative phases of monitoring and reporting. Each monitoring cycle of EEM requires submission of data and reports by pulp and paper mills and mines across Canada (Walker et al. 2002). This paper uses pulp and paper mill data as an example but the expert system also applies to the metal mining industry.

The key components of the EEM program include a fish survey to assess effects on fish and a benthic invertebrate community survey to assess effects on fish habitat. For the purposes of EEM, an effect is generally defined as a statistically significant difference in fish or benthic invertebrate community endpoints measured between an area exposed to effluent and a reference area or a statistically significant gradient in these endpoints from the exposure to reference areas. For the EEM, certain specified data (the effect endpoints, see Table 1 for fish endpoints) that are generated from the fish survey, benthic invertebrate community survey and fish usability studies were designated to assess the presence and level of effects.

The current objective of the EEM program is to evaluate the effects of pulp and paper or mining effluents on the aquatic environment. The program uses both field monitoring and laboratory approaches to directly or indirectly assess the health of fish, habitat impairment and concerns relating to human consumption of fish. It is also being used to discriminate between pulp and paper mill or mine-related effects and other natural or anthropogenic stressors. The program is designed to generate all relevant data and knowledge which may

Type of response	Effect endpoint	Statistical procedure
Survival	Age	ANOVA
Energy use	Size-at-age (body weight against age) Relative gonad size (gonad weight against body weight)	ANCOVA ANCOVA
Energy storage	Condition (body weight against length) Relative liver size (liver weight against body weight)	ANCOVA ANCOVA

Table 1. Endpoints to be used for determining exposure-associated effects on fish as designated by statistically significant differences between exposure and reference areas.

be used for analysis of both spatial and temporal trends in a way that can be reliably interpreted.

It should be pointed out that not all effects identified in the EEM will represent damage to fish, fish habitat, or the usability of the fisheries resources, but the effects can represent differences or gradients that may reflect changes to the ecosystem associated with the effluent. Detailed information on the effects, including the magnitude, geographic extent, and possible cause are used in the management of the aquatic resources.

Early in the EEM program it became clear to the EEM office that the assessment results coming back from the various industries were not always being carried out correctly or consistently. This was due to several factors. The first factor is that the collected data needs to be processed into a form that the statistics packages can accept, as well as carrying out log transformations if necessary and looking for outliers. Another factor is that the outputs of the results are often scattered because of multiple endpoints that need to be examined. It is also difficult to prepare the data for each statistical end point test, to make a decision as to whether there is a significant effect, and then to be able to move on to another end-point test procedure, if it is required. Finally, the results of all of the statistical tests need to be retrieved from various files and reviewed. The EEM Statistical Analysis Tool Decision Support System (EEM-SAT DSS) was conceptualized and developed to solve these problems with the aid of an expert system module. This chapter focuses on the development of the EEM-SAT DSS as well as examples of outputs from the system rather than the development of the EEM approach itself. The reader is directed to the following publications for details on the EEM approaches and designs (Lowell et al., 2002; Walker et al., 2002; Environment Canada, 2000a, 2002).

2.2 Key design considerations of the EEM-SAT decision support system

The flow schematic of the EEM-SAT procedures is shown in Figure 2. The EEM-SAT Expert System was developed to solve the problems of inconsistent data entry, data sampling collection, human induced errors, and the misinterpretation of the EEM statistical functions that were detected by the EEM office. There are many statistical treatments and routines required for testing the endpoint effects. In addition, reference sites must be established with domain experts to ensure the appropriateness and quality of these sites. The EEM procedures are tedious, time consuming and error prone if they are not automated. In the wake of these shortcomings, an EEM-SAT DSS was developed to fully integrate with the EEM statistical database. Since the test data is collected on site, it is more appropriate to allow the consultant at each company to use the automated procedure in the DSS to analyze the most recent test data. Therefore, the development of the DSS takes great care in terms of defining the correct reference data for each company, the data validation and outlier screening, the implementation of rigorous statistical routines for the endpoint effect analysis and the site assessment report which indicates the overall condition of the fish and benthic invertebrate communities. When an extreme effect is detected, a company can take the proper course of action for remediation.

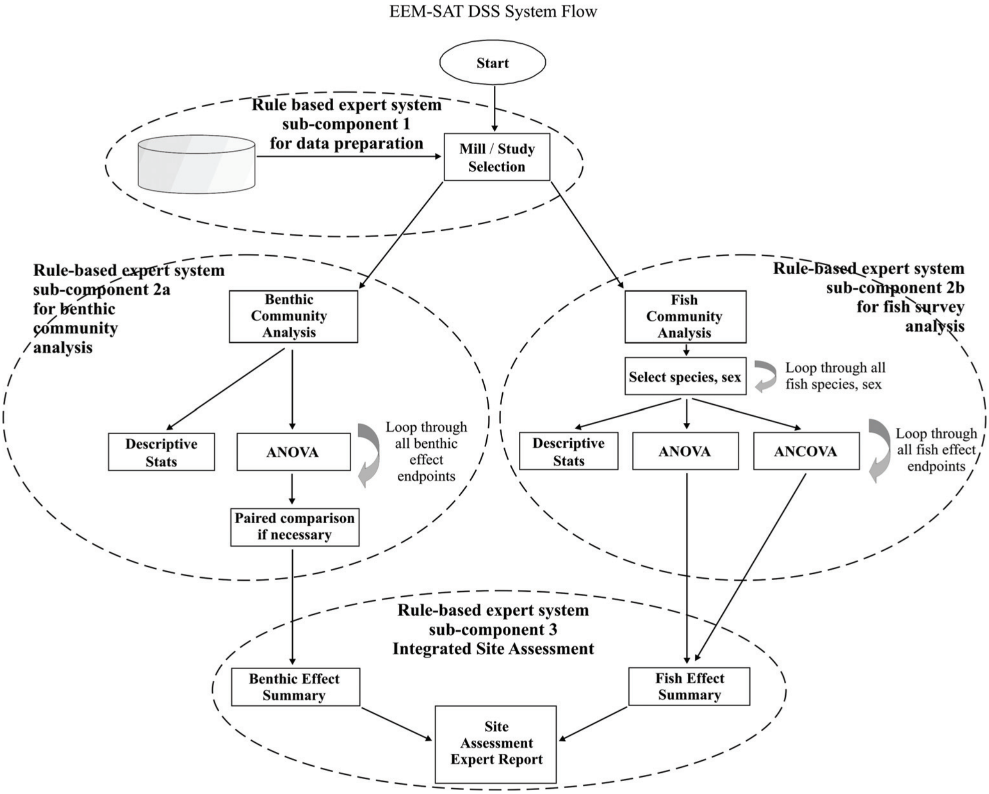
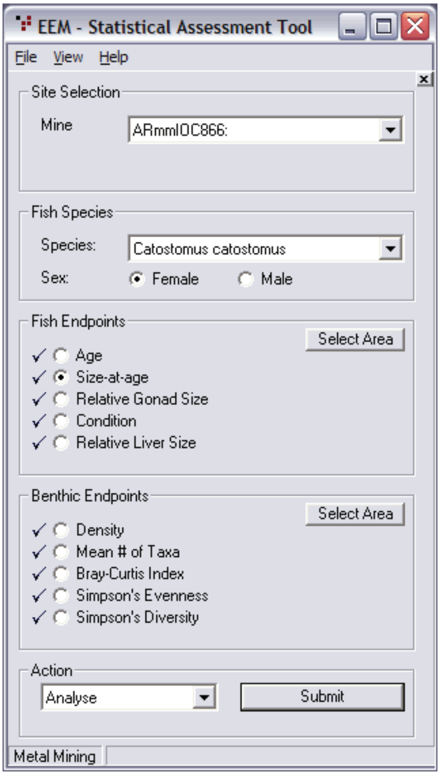

EEM-SAT DSS System Flow
Start
Rule based expert system sub-component 1 for data preparation
Mill / Study Selection
Rule-based expert system sub-component 2a for benthic community analysis
Benthic Community Analysis
Descriptive Stats
ANOVA
Loop through all benthic effect endpoints
Paired comparison if necessary
Rule-based expert system sub-component 2b for fish survey analysis
Fish Community Analysis
Select species, sex
Loop through all fish species, sex
Descriptive Stats
ANOVA
ANCOVA
Loop through all fish effect endpoints
Rule-based expert system sub-component 3 Integrated Site Assessment
Benthic Effect Summary
Fish Effect Summary
Site Assessment Expert Report

Fig. 2. EEM- SAT schematic diagram

The main driver of the EEM-SAT DSS is its rule-based expert system. The expert system provides control in the automation of the EEM-SAT procedures, linking appropriate test datasets with the reference datasets, utilizing the appropriate statistical procedures for each of the endpoint effects and determining the magnitude of the effect if it is found to be present. It has been developed based upon the RAISON Decision Support System framework (Booty et al., 2001; Lam et al., 2004), with interfaces constructed using Visual Basic. Expert systems, a form of artificial intelligence, are human computer systems that perform problem-solving tasks in a specific domain (Ignizio, 1991; Buchanan and Shortliffe, 1985). The systems are useful tools in numerous application areas including environmental domains. They apply heuristic rules to encode domain knowledge, together with inference engines, in order to deduce conclusions from information that the users provide. Decision support systems are systems that employ various techniques that include artificial intelligence. In particular, expert systems can be integrated with more classical techniques of functionality such as statistics, mapping and/or data retrieval to form systems that provide more effective decision support in a study domain. The expert system module of the EEM-SAT system acts as a wizard. It has several expert system components including rule bases about the limits of its applicability, what kind of input data it requires, how to remove outliers, how to estimate its parameters from available information, how to extract data from the database, how to execute the statistical routines and correctly interpret the statistical results. In general, the expert system component assists users with less expertise in EEM to better perform their job. The EEM-SAT DSS is an example of a typical environmental decision support system. Figure 3 illustrates the user interface of the EEM-SAT DSS. The EEM-SAT expert system components contain only small knowledge bases, but they dramatically improve the functionality of the EEM-SAT DSS. The three major components of the expert system are described below.



EEM - Statistical Assessment Tool

Site Selection
Mine: ARmmlOC866:

Fish Species
Species: Catostomus catostomus
Sex: ● Female ○ Male

Fish Endpoints
✓ ○ Age
✓ ● Size-at-age
✓ ○ Relative Gonad Size
✓ ○ Condition
✓ ○ Relative Liver Size

Benthic Endpoints
✓ ○ Density
✓ ○ Mean # of Taxa
✓ ○ Bray-Curtis Index
✓ ○ Simpson's Evenness
✓ ○ Simpson's Diversity

Action: Analyse

Metal Mining

Fig. 3. EEM-SAT DSS user interface

2.3 EEM-SAT DSS expert system components

2.3.1 Intelligent data preparation and screening process

The expert system module guides the users to provide the appropriate data and information. The users can either submit the information interactively or via the database.

The system builds rules based on expert knowledge to return the pulp and paper mill and mill IDs that have data for certain types of fish species and invertebrate analysis. It provides an effective means to identify outliers based on robust statistical techniques. Finally, it also examines the nature of the data and determines if it is necessary to perform data transformations before submitting the data for the statistical routines. The rules used for fish are as follows:

1. If a pulp and paper mill is selected, then access the database to return available fish species and determine if there are enough benthic data for the EEM statistical analyses.

2a. If enough data is available, then the system displays the data and asks the user to select species and appropriate statistical tests.

2b. If not enough data is available, then it advises the user of the data gap, which data areas are lacking and signals to the DSS that statistical routines cannot proceed.

3. If fish species and benthic data are selected, then it advises the user of any potential outliers.

4. If fish species and benthic data are selected and the outlier detection procedure is complete, then examine the data to determine if a logarithmic transformation of data is required.

5a. If data transformation is required, then perform the necessary data transformation and submit the processed data to the fish and benthic statistical routines.

5b. If data transformation is not necessary, then submit the data to the fish and benthic statistical routines.

The rules for the benthic organisms are as follows:

1. Isolate data: data is selected on the basis of study ID, species and gender/sex. If there is not enough data for the study, the system will reject the request and inform the user to supply the missing information for the analysis.
2. Log_{10} transform dependent and independent variables (if necessary): logarithmic transformations are preferred in the endpoint effect analysis simply because biological measures are often considered to be logarithmic or exponential scale. One of the criteria of the ANOVA and ANCOVA procedures is that the data should be normal. This includes body weight, total length, fork length, standard length, age, gonad weight and liver weight.
3. Checking for outliers: scatter plots of fresh weight vs. age, gonad weight vs. fresh weight, fresh weight vs. length, and liver weight vs. fresh weight (all variables are in logarithmic scale) are presented. When a scatterplot illustrates outliers, the user should be given an opportunity to identify, modify, and/or delete data. Although there is no formal guidance on screening data on the basis of studentized residuals, a rule of thumb is that when the studentized residual exceeds 4, this indicates that the observation may be unusual and the observation should be removed and the analysis should be rerun.

In addition, the user should pay attention to the high ''leverage'' values. Those observations tend to potentially skew the observed relationship in one direction or another. There is professional judgment used to determine whether the data with high leverage should be excluded. A common approach is that if they grossly skew the expected relationship, then exclusion should be considered.

2.3.2 Integrated advanced statistics for fish community end-point analyses

The EEM DSS uses both field monitoring and laboratory approaches to directly or indirectly assess the health of fish, habitat impairment, and concerns relating to human consumption of fish. The objective of assessing impairment is met through assessment of the endpoint effects in both fish and benthic communities. The overall fish analysis pathway and its endpoint effects are shown in Figure 4. Fish monitoring for the EEM DSS involves monitoring both sexes of two sentinel species at reference and exposure areas to assess if there are differences in the growth, reproduction, survival or condition of fish populations. Sex differences are common due to differences in overall energetic requirements between male and female fish. Effect endpoints include weight at age, size-at-age, relative gonad size, liver weight, and condition factor for fish and taxon richness, Simpson's diversity and evenness indices, and Bray–Curtis index for benthos, as shown in Table 1. Simpson's diversity index is a measure of the character of a community that takes into account both the abundance patterns and the taxonomic richness of the benthic invertebrate community. It is

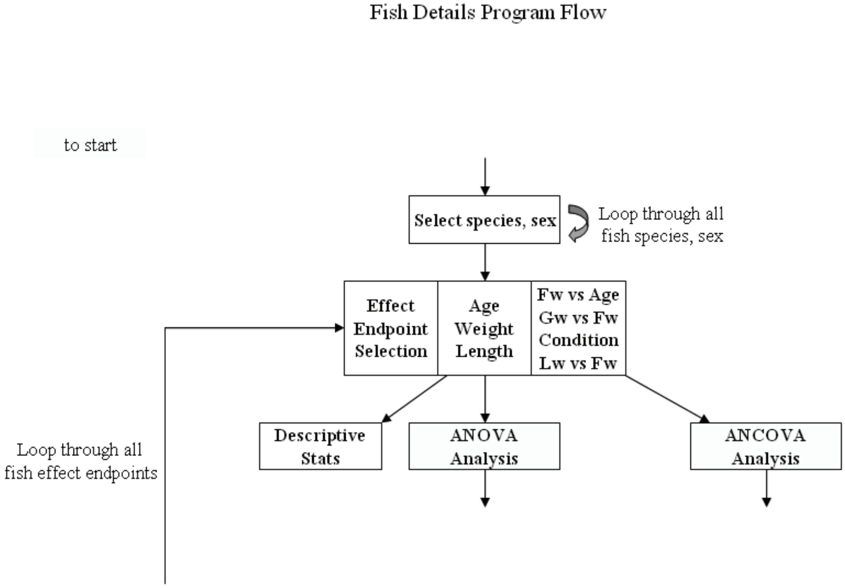


Fig. 4. Overall fish analysis pathway and its endpoint effects schematic.

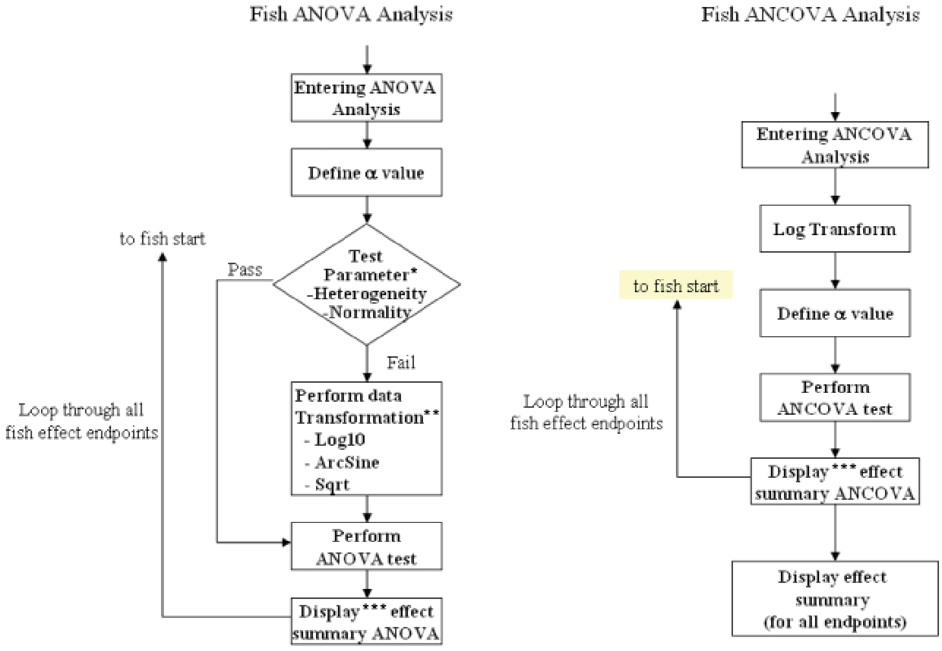
Type of response	Effect endpoint	Statistical procedure
Survival	Age	ANOVA
Energy use	Size-at-age (body weight against age) Relative gonad size (gonad weight against body weight)	ANCOVA ANCOVA
Energy storage	Condition (body weight against length) Relative liver size (liver weight against body weight)	ANCOVA ANCOVA

Table 1. Endpoints to be used for determining exposure-associated effects on fish as designated by statistically significant differences between exposure and reference areas.

calculated by determining, for each taxonomic group, the proportion of individuals that it contributes to the total sample. For the Simpson's evenness index, evenness measures how similar the frequencies of the species are. When all the frequencies are equal, evenness is one. Frequency imbalance lowers the Simpson's evenness index. The Bray–Curtis index is an index that measures the degree of difference in community structure (especially community composition) between sites. This measure helps to evaluate the amount of dissimilarity between benthic invertebrate communities at different sites.

A summary of the EEM-SAT fish analysis expert system rules is as follows:

1. Analyses of variance (ANOVA) test (Figure 5): carry out ANOVA to test for differences between areas, and calculate means and standard deviations for each key variable for both areas (reference and exposure). Once the groups are identified to be significantly different, the user needs to determine which pairs differ. This is done using Tukey's HSD post hoc test (SYSTAT 11 Statistics II, 2005). Assumptions for ANOVA are that the data for reference and exposure populations are normally distributed; the variances are


Fish ANOVA Analysis
Entering ANOVA Analysis
Define α value
Test Parameter* -Heterogeneity -Normality
Pass
Fail
to fish start
Loop through all fish effect endpoints
Perform data Transformation** - Log10 - ArcSine - Sqrt
Perform ANOVA test
Display*** effect summary ANOVA
Fish ANCOVA Analysis
Entering ANCOVA Analysis
Log Transform
Define α value
to fish start
Perform ANCOVA test
Loop through all fish effect endpoints
Display*** effect summary ANCOVA
Display effect summary (for all endpoints)

*Note:
1) View residual plots for reference, exposed
2) Check for patterns to determine if transformation is required. If so, which transformation is appropriate.
3) Screen or indicate outlier, if required

**Note: The type of transformation will be based on the recommendation of the science committee

***Note: All results convert back to non-transformed state for display

Fig. 5. Fish ANOVA and ANCOVA Analysis schematic diagrams

equal between the reference and exposure populations and the error terms are independently distributed.

2. Analyses of covariance (ANCOVA) and regression (Figure 5): this is done using the General Linear Model (GLM) (Environment Canada, 2000a). In particular, a contrast statement is used to test relationships among reference vs. exposure. The test is composed of two parts. It is carried out first to determine whether the slopes are approximately parallel. If the slopes are parallel, it then requires determining if the elevations of the regressions are significantly different. ANCOVA combines the features of ANOVA and regression, and can be used to compare regressions among treatments (i.e. reference vs. effect areas). Assumptions of ANCOVA are that the residuals are normally and independently distributed with zero mean and a common variance; the independent variable (covariate) is fixed and measured without error; the relationship has the form specified (linear regression) and the slopes of regression lines among areas are equal.

2.3.3 Integrated advanced statistics for benthic community end-point analyses

For the purposes of the EEM program an ''effect on the benthic invertebrate community'' means a statistical difference between benthic invertebrate community measurements taken in an exposure area and a reference area (e.g., control/impact design) or a statistical difference between measurements taken at sampling areas in the exposure area that indicate gradually decreasing effluent concentrations (e.g., a gradient design). The EEM-SAT DSS program only pertains to control/impact analyses at this stage of development. This design uses ANOVA to detect differences between reference and exposure areas. The six basic study designs and their associated statistical procedures are shown in Table 2.

Study Design	Statistical procedure
Control/impact	ANOVA/ANCOVA
Multiple Control/Impact	ANOVA/ANCOVA
Simple Gradient	Regression/ANOVA/ANCOVA
Radial Gradient	Regression/ANOVA/ANCOVA
Multiple Gradient	ANCOVA
Reference Condition Approach	Multivariate/ANOVA/ANCOVA

Table 2. Statistical procedure used to evaluate exposure-associated effects on benthic invertebrates for each of the six basic study designs.

The benthic analysis procedures are represented in Fig. 6. The expert system rules for the benthic invertebrate are the same as the fish analyses with different effect endpoints. The effect endpoints for benthic invertebrate analyses are abundance (density), mean number of taxa, Bray–Curtis index, evenness and Simpson's diversity index. These descriptors are largely summary metrics selected to encompass the range of effects, which may be a result of mine or pulp and paper mill effluents.

The α value is set to 0.10 for both fish and benthic invertebrate in determining effects. This is because the fish and benthic invertebrate community survey should minimally have sufficient statistical power analysis to detect an effect size of two standard deviations. Table 3 indicates that α and ß should be able to achieve both Type I and II errors at 0.10 since sample size is usually about 5 for both fish survey and benthic community studies.

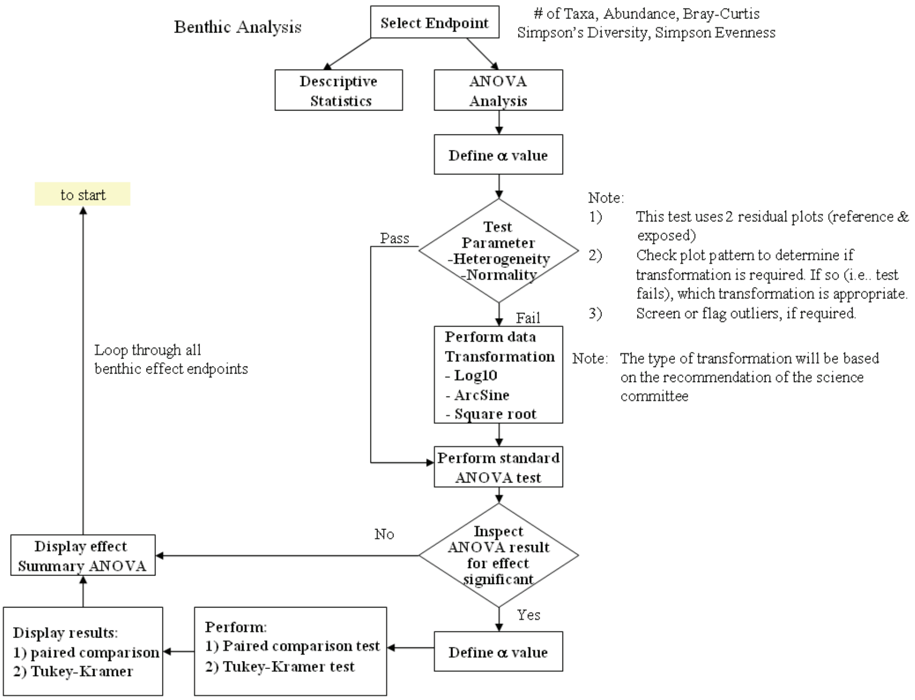

Benthic Analysis
Select Endpoint
# of Taxa, Abundance, Bray-Curtis
Simpson's Diversity, Simpson Evenness
Descriptive Statistics
ANOVA Analysis
Define α value
to start
Pass
Test Parameter -Heterogeneity -Normality
Note:
1) This test uses 2 residual plots (reference & exposed)
2) Check plot pattern to determine if transformation is required. If so (i.e.. test fails), which transformation is appropriate.
3) Screen or flag outliers, if required.
Fail
Loop through all benthic effect endpoints
Perform data Transformation - Log10 - ArcSine - Square root
Note: The type of transformation will be based on the recommendation of the science committee
Perform standard ANOVA test
No
Display effect Summary ANOVA
Inspect ANOVA result for effect significant
Yes
Display results: 1) paired comparison 2) Tukey-Kramer
Perform: 1) Paired comparison test 2) Tukey-Kramer test
Define α value

Fig. 6. Benthic Analysis schematic diagram.

A	1-β			
	0.99	0.95	0.90	0.80
0.01	14	11	10	8
0.05	11	8	7	5
0.10	9	7	9	4

Table 3. Sample size required to detect an effect size of ± 2 standard deviations

2.3.4 Integrated site assessment expert report to interpreting the analysis results

The expert system module ingests all the relevant information and the outputs for the current analysis. Specifically, a detailed summary of the current state of the analyses is provided. It will highlight and provide detailed information of potential troubled areas of the statistical analyses. For example, what tests do the fish species fail and by how much? This report allows the users to deal with the problem properly. The rules used in this subcomponent of the system include:

1. If the statistical routines are successfully completed, then extract metadata information of the pulp and paper mill, the name of the statistical routines and the summary results of each test.
2. If some of the results are outside of the normal bounds, then extract the detailed information of those tests and include in the summary results. The detailed information

includes the name of the test and all statistical data such as the degree of freedom, the test value, the critical value and the confidence level. The system will proceed to produce a complete and integrated report to the user to pinpoint the critical results of the full analyses.

2.4 An EEM-SAT DSS example

2.4.1 Fish size-at-age (effect endpoint)

In this example of the decision support process, the rates of growth are described by the relationship of size (as weight or length) to age. Over the entire life span of the fish, this relationship is curvilinear, with the rate of increase declining as fish approach the limit of their life span. Size-at-age may be estimated by calculating the regression relationship between body size (weight or length) and age for each sampling area (reference and exposure). Calculation of mean age is meant as a gross reflection of the age distribution of adult fish collected from each area. The EEM-SAT DSS automatically $\log_{10}$ transforms all ANCOVA based analyses. It also uses weight as the covariate. Body weight is corrected by subtracting the gonad weight and liver weight from the body weight prior to analysis. The EEM-SAT DSS follows the two-step (slopes, then intercept) analyses. If slopes are not significantly different at the alpha value specified by the user, then the test for intercepts (least squared means) proceeds automatically. If slopes are significantly different, the software will not test for differences in intercepts. Endpoints analyzed by ANCOVA (size-at-age, relative gonad size, condition, and relative liver size) have only one component in the EEM-SAT: 1) Effect Analyses. Any descriptive measures associated with the analyses are included in the ANCOVA result table. An example is shown in Figure 7. In this example, the effect endpoint "size at age" is examined. Using ANCOVA testing the slope difference between the "Reference" and "Effects" data shows that test p-value is 0.854 and is larger than the α, (0.10), therefore, there is no significance difference. However, testing difference of the means indicates that the test p is almost zero and is statistically significant against α, (0.10).

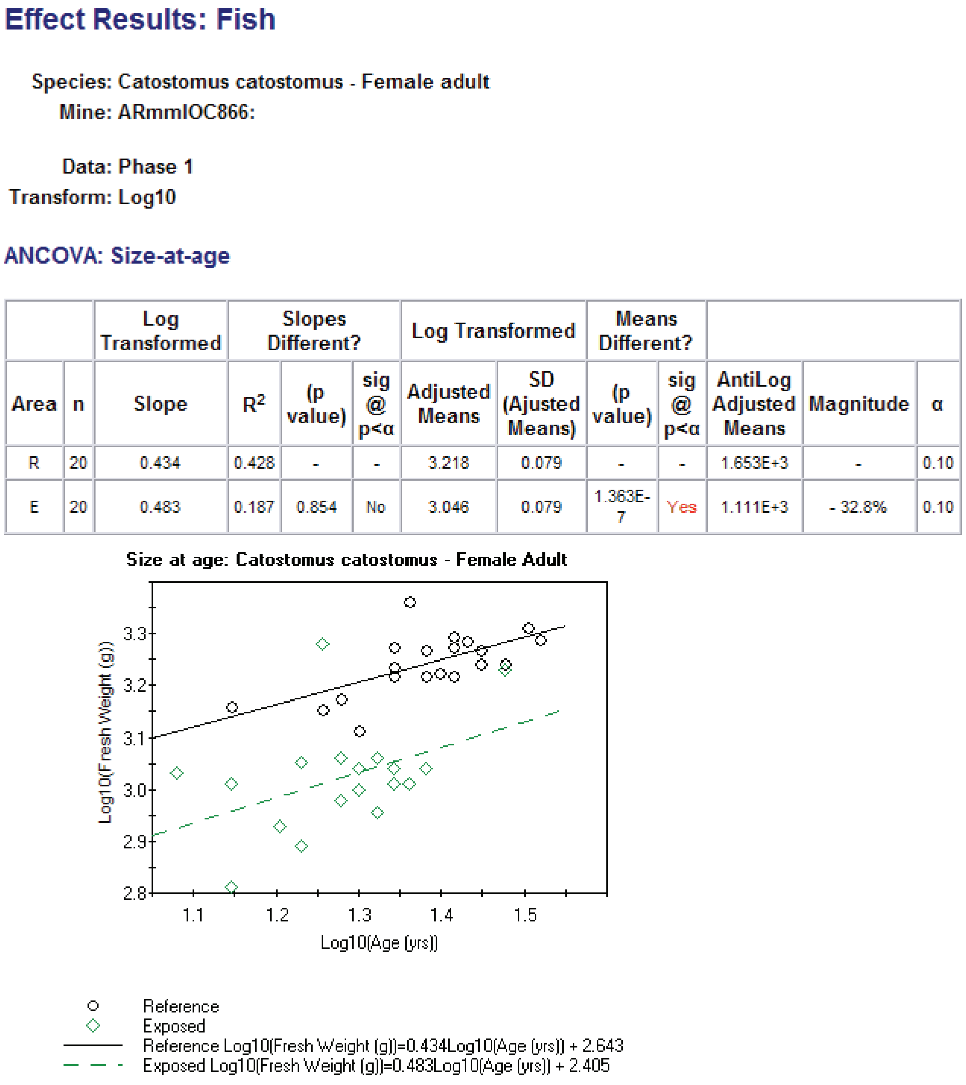
2.4.2 Benthic endpoints

All benthic endpoints are analyzed by ANOVA and have two components in the EEM-SAT: 1) Descriptive Statistical Analyses and 2) Effect Analyses. Examples of an effects analysis of a site (name deleted for privacy) are presented below. In Figure 8 for mean number of taxa, the test p-value is less than α (0.10) indicating that there is a significant difference (magnitude of -45%) between the reference and the near field sites. The Bray-Curtis Index results are shown in Figure 9 where it can be seen that the test p-value is less than α (0.10), with a magnitude of +83.8%, again indicating a significant effect.

2.4.3 Integrated site assessment expert report

The system summarizes the results of the analyses for each site as shown in Figure 10. It can be seen that for fish species Catostomus catostomus, all of the endpoints show an effect. For benthos, all but the Simpsons Evenness endpoint indicates a significant effect.

The EEM-SAT DSS file open and save options allow users to generate scenarios for comparison with different levels of significance. In this program, a key point is to maintain high probabilities of correctly identifying areas that are actually impaired (high statistical power), while still maintaining low probabilities of falsely concluding that impairment has occurred in nonimpacted areas (low α values).



Effect Results: Fish

Species: Catostomus catostomus - Female adult
Mine: ARmmlOC866:

Data: Phase 1
Transform: Log10

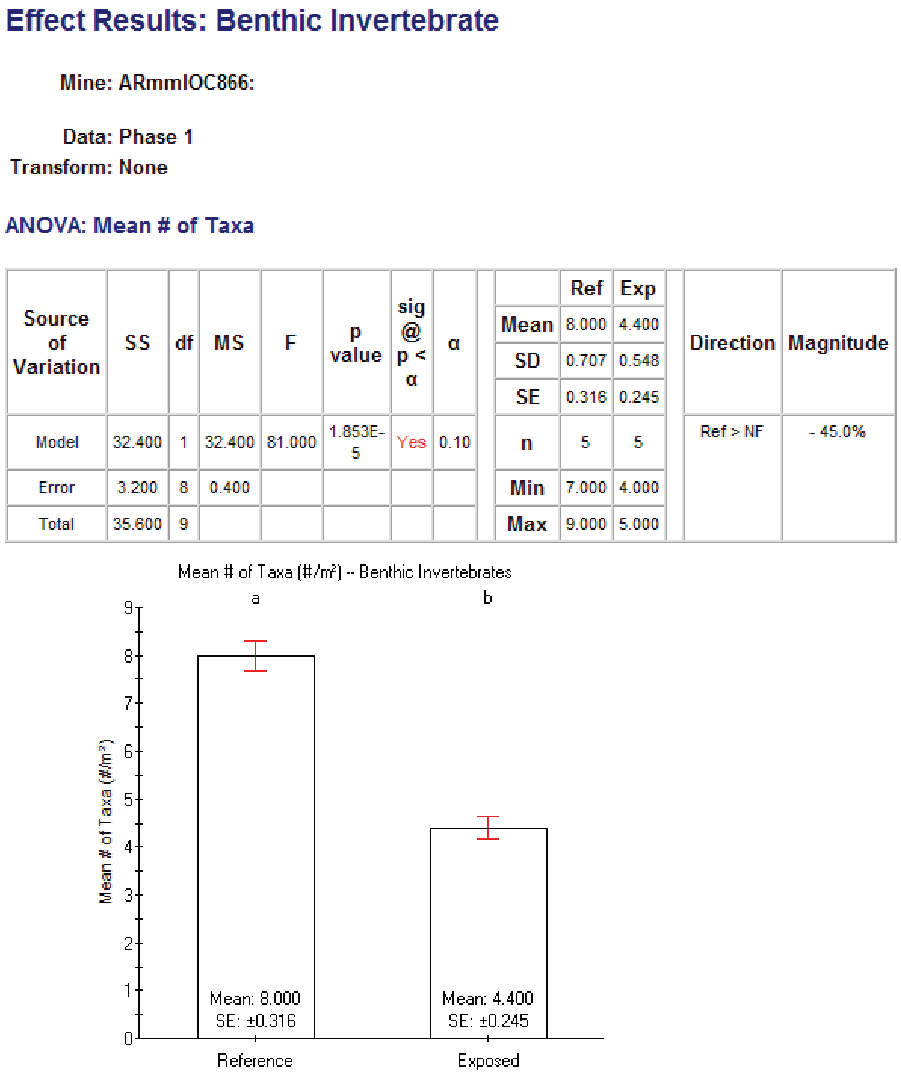
ANCOVA: Size-at-age

		Log Transformed	Slopes Different?			Log Transformed		Means Different?				
Area	n	Slope	R^2	(p value)	sig @ p<α	Adjusted Means	SD (Ajusted Means)	(p value)	sig @ p<α	AntiLog Adjusted Means	Magnitude	α
R	20	0.434	0.428	-	-	3.218	0.079	-	-	1.653E+3	-	0.10
E	20	0.483	0.187	0.854	No	3.046	0.079	1.363E-7	Yes	1.111E+3	- 32.8%	0.10

Size at age: Catostomus catostomus - Female Adult

○ Reference
◇ Exposed
—— Reference Log10(Fresh Weight (g))=0.434Log10(Age (yrs)) + 2.643
– – Exposed Log10(Fresh Weight (g))=0.483Log10(Age (yrs)) + 2.405

Fig. 7. ANCOVA fish size-at-age analysis example results



Effect Results: Benthic Invertebrate

Mine: ARmmIOC866:

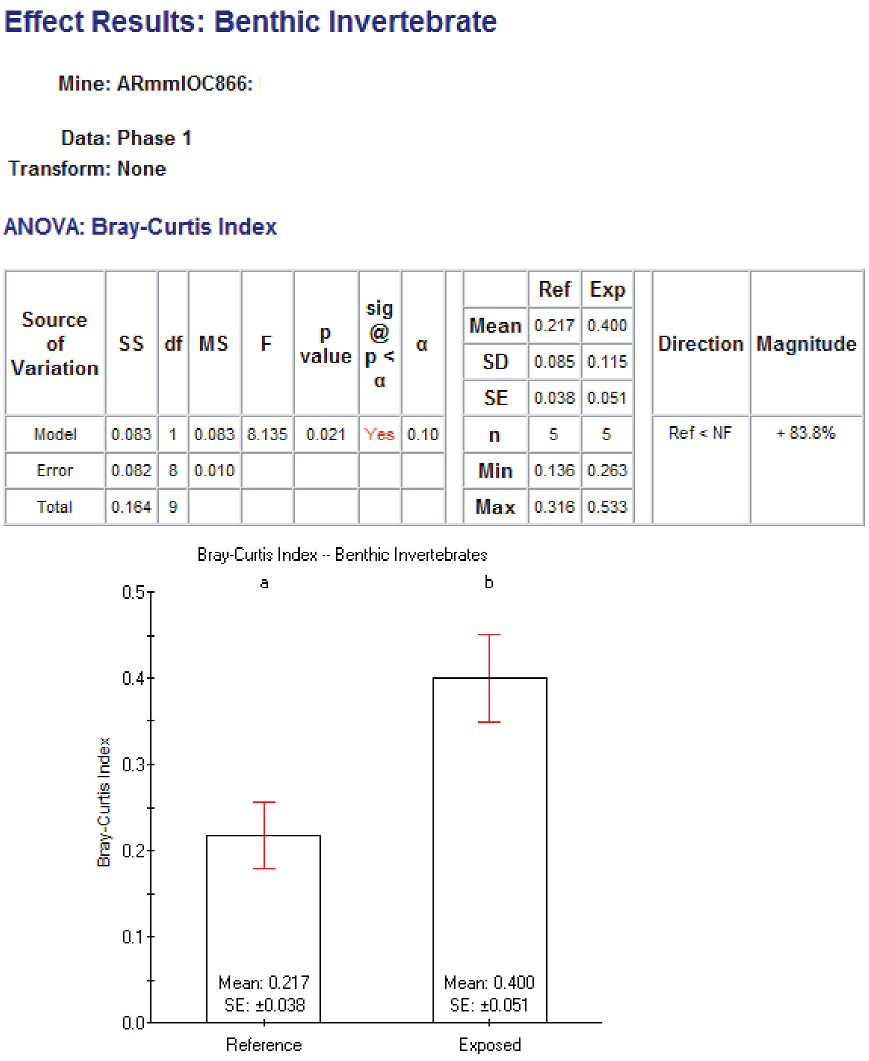
Data: Phase 1

Transform: None

ANOVA: Mean # of Taxa

Source of Variation	SS	df	MS	F	p value	sig @ p < α	α		Ref	Exp	Direction	Magnitude
								Mean	8.000	4.400		
								SD	0.707	0.548		
								SE	0.316	0.245		
Model	32.400	1	32.400	81.000	1.853E-5	Yes	0.10	n	5	5	Ref > NF	- 45.0%
Error	3.200	8	0.400					Min	7.000	4.000		
Total	35.600	9						Max	9.000	5.000		

Fig. 8. Benthic Mean number of taxa example results.



Effect Results: Benthic Invertebrate

Mine: ARmmIOC866:

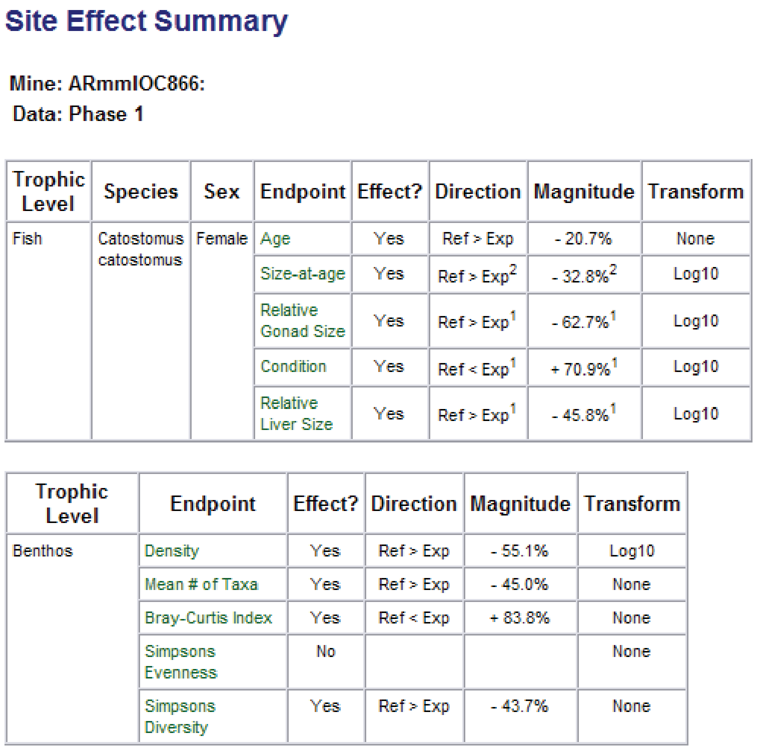
Data: Phase 1

Transform: None

ANOVA: Bray-Curtis Index

Source of Variation	SS	df	MS	F	p value	sig @ p < α	α		Ref	Exp	Direction	Magnitude
								Mean	0.217	0.400		
								SD	0.085	0.115		
								SE	0.038	0.051		
Model	0.083	1	0.083	8.135	0.021	Yes	0.10	n	5	5	Ref < NF	+ 83.8%
Error	0.082	8	0.010					Min	0.136	0.263		
Total	0.164	9						Max	0.316	0.533		

Fig. 9. Benthic Invertebrate example Bray-Curtis Index results.



Site Effect Summary

Mine: ARmmIOC866:
Data: Phase 1

Trophic Level	Species	Sex	Endpoint	Effect?	Direction	Magnitude	Transform
Fish	Catostomus catostomus	Female	Age	Yes	Ref > Exp	- 20.7%	None
			Size-at-age	Yes	Ref > Exp[2]	- 32.8%[2]	Log10
			Relative Gonad Size	Yes	Ref > Exp[1]	- 62.7%[1]	Log10
			Condition	Yes	Ref < Exp[1]	+ 70.9%[1]	Log10
			Relative Liver Size	Yes	Ref > Exp[1]	- 45.8%[1]	Log10

Trophic Level	Endpoint	Effect?	Direction	Magnitude	Transform
Benthos	Density	Yes	Ref > Exp	- 55.1%	Log10
	Mean # of Taxa	Yes	Ref > Exp	- 45.0%	None
	Bray-Curtis Index	Yes	Ref < Exp	+ 83.8%	None
	Simpsons Evenness	No			None
	Simpsons Diversity	Yes	Ref > Exp	- 43.7%	None

Fig. 10. Effect summary results for fish and benthic invertebrates

3. Case 2: land and water integration decision support system

3.1 Background

Agricultural activities such as animal farming, grazing, plowing, pesticide spraying, irrigation and fertilizer applications can cause non-point source or diffuse pollution. Nutrients and sediment are two of the main agricultural pollutants affecting water quality that result from these activities. Nutrients such as phosphorus and nitrogen are minerals that can be applied to enhance plant growth and crop production. When they are applied in excess of crop needs, the excess nutrients are often attached to soil particles that can be carried by overland water runoff from land into the aquatic ecosystems. The nutrients can cause excessive algae and aquatic plant growth in rivers and streams; cloud the water; reduce the amount of sunlight reaching aquatic plants; cover fish spawning areas and food supplies; greatly increase the costs of water treatment; reduce swimming and water recreation activities; create a bad smell; kill fish; and accelerate aging of rivers and lakes. Besides fisheries and recreation effects, these pollutants also have harmful effects on drinking water supplies and wildlife. Thus, there exists an important linkage between the land and the water. Environmental performance standards for floral and faunal communities in terrestrial ecosystems are based on assessments and forecasts of land cover and land use in agricultural regions. Performance standards for aquatic community structure in streams are based on assessments and forecasts of flow regime, sediment levels and nutrient concentrations. However, the physico-chemical condition of a stream is strongly affected by catchment characteristics, including land cover and land use, but also by basin shape, surficial geology and soil structure. Thus, land cover and land use patterns will have profound impacts on both water quantity and quality, and aquatic biodiversity.

3.2 Key design considerations of the LWIDSS

Modelling is an important asset of any environmental decision support system. With the high cost of full scale field work, modelling presents a cost effective approach to assess the impact on the environment. In the LWIDSS, the emphasis is on the terrestrial and aquatic models that are commonly used to assess agricultural impacts.

Dynamic landscape modelling helps decision makers assess the consequences of alternative management scenarios at the landscape scale. It simulates land use scenarios that are characterized by different assumptions about management practices. The results are in the form of GIS spatial layers that can be evaluated at the landscape level or can be fed into other component models such as non-point source pollutant models to evaluate the impact of land to water quality. This approach differs from other modelling efforts in that it does not confine itself to just one model or a given set of models. Rather, it provides an open architecture framework that accepts any component model within the system that can be linked to other component models in the causal chain, be it a dynamic landscape model for land use scenario creation or a non-point source pollutant model for sediment and nutrients assessment.

Land use scenarios are integrated with watershed hydrology models to develop flow, sediment and nutrient performance standards in streams to protect aquatic biodiversity. In addition to the scenario representing the present day (current), others are developed to explore different land use cases (e.g., agricultural intensification). Validated and calibrated hydrologic models use these scenarios to estimate water quantity and quality parameters. These parameters are then used to forecast aquatic biodiversity according to empirically-derived relationships between stream flow, sediment and nutrient regimes and biotic condition. Benthic algal and invertebrate communities, as well as fish communities, function as the biotic endpoints of streams and rivers to gauge ecosystem integrity.

Non-point source pollutant modelling in general can be a large and complex process requiring great quantities of input and generating vast amounts of output because of its nature of trying to simulate real-world processes. Dealing with such sums of data, both inputs and outputs, can be daunting to those who are trying to understand and extract knowledge and information from them. Not all modelling programs contain tools for visualizing the results, comparing multiple sets of results, performing post-analysis or managing/organizing the data from different model runs. Typically, after executing the models, different software are employed to look at the output and to perform further statistical analysis and these can be time consuming procedures by themselves. It is apparent that it would be very useful to have an integrated set of tools in a single software system that performs these tasks (Lam et al., 1998). This would make those doing modelling more productive by allowing them to examine the results in a more efficient manner. More time can be spent on modelling and less time on manipulating data to move it into software programs.

The LWIDSS addresses the issues of linking multi-media models at different geospatial scales. It provides interfaces that can accept, select, link and recalibrate discipline-specific component models. It can seek optimal solutions for best management practices such as buffer strip widths for sediment and nutrients reduction based on feedback of individual component models. The design of the LWIDSS has benefited from significant input from scientists, researchers, other end-users, system developers, modellers and GIS specialists. Figure 11 illustrates the schematic diagram of the LWIDSS. At a glance, the LWIDSS offers a medium to integrate the data and information of the land and water by providing a number of necessary functionalities. The integration includes data, maps and models with user-friendly tools, including data input/output views, map input/output views, and modelling result views for interpretation, further analysis, conclusion and recommendation. The LWIDSS is designed as a framework and can be easily adapted to any watershed as portability is an important aspect of design consideration.

The LWIDSS is built around the concept of a management user interface to assist policy makers in their decision making. The technical users employ other tools to build model inputs, execute, and calibrate and validate the models; the technical aspect of the modelling process is beyond the scope of this paper. The friendly interface provides a platform for the management or policy makers to view the inputs and outputs of the system that the technical users have built. This will allow management to investigate the analytical results based on robust science built by the researchers. Key functionality includes mapping of the results, scenario gaming and key statistical analyses of the study area.

The LWIDSS design also calls for both temporal and spatial consistency among component models as the output from one model is used as input to another in a sequence of linked calculations. For example, the dynamic landscape model generates land use maps for various land use scenarios that can be used either in a single storm event non-point source pollutant model such as Agricultural Non-Point Source Pollutant Model (AGNPS), which is grid based or in a continuous time non-point source pollutant model such as Soil and Water Assessment Tool (SWAT), which is vector based.

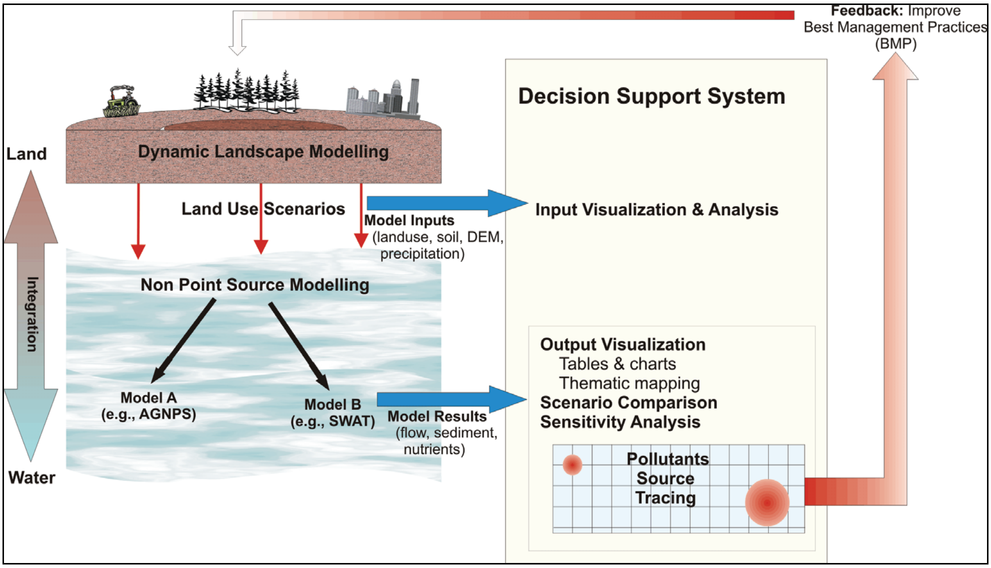


Fig. 11. Schematic diagram of the LWIDSS.

A major requirement for policy makers is to use the LWIDSS to predict results, including evaluation of different scenarios on land and water and optimization of BMP. These should be obtained in a relatively short computational timeframe with a friendly user interface. Using the modelling approach to understand the non-point source pollutant problem is important for providing the assessment of the impacts of land-water integration. In addition, implementing a scenario gaming approach would allow decision makers an opportunity to understand the problem based on different possible scenarios and to make viable decisions to manage the problem more effectively and to minimize the impacts. The feedback among component models is critical to the whole integration process. Different models can complement each other with their strengths. For example, the AGNPS model excels in identifying pollutant "hot spots" and the SWAT model can take the information on the "hot spots" to further evaluate the optimal solution of BMP.

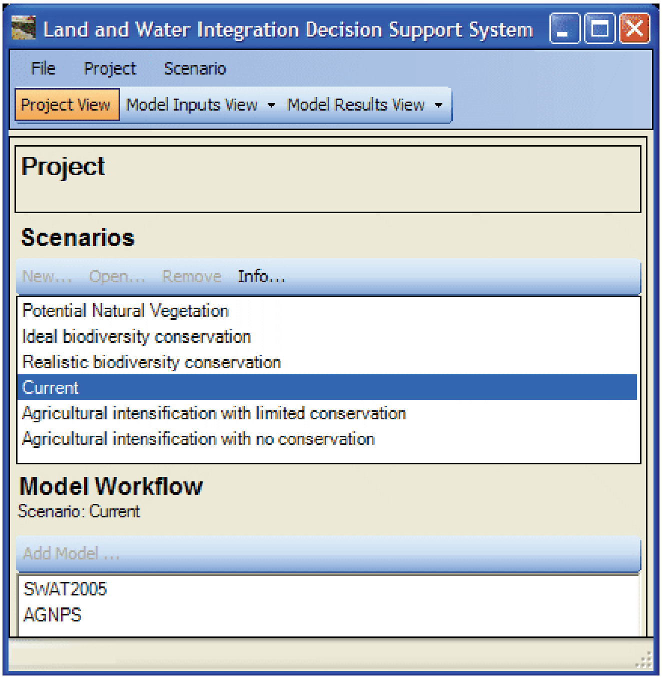
3.3 Integrated analysis approach

The integrated analysis approach will be described through an example which decision makers setting environmental policies could potentially face. The problem is to investigate a study area in Ontario, Canada to assess its current state in terms of aquatic ecosystem health and to forecast future conditions if agriculture practices need to be intensified because of increasing demands for crop production. We identify the Raisin River Watershed, an agricultural watershed in Ontario, Canada as a pilot example for the development of the LWIDSS. It is selected to be the primary focus because agricultural activities can have an impact on the environment, in particular on the aquatic ecosystem. Current watershed conditions can be evaluated by gathering empirical data such as that done through water quality sampling. This usually involves sending people out into the field across the watershed at regular time intervals, collecting the water samples and analyzing the samples back at the laboratory. This method can be quite costly both in terms of time and money due to amount of labour and transportation and equipment requirements. In fact, it may not be feasible at all because of budget constraints or watershed size. Also, sites located in difficult to reach areas and poor weather can complicate the process. An alternate and more cost effective method is to make use of computer models to try to predict the water quality by simulating real-world physical processes occurring in nature (Leon et al., 2004). More specifically, some type of non-point source pollutant model is required to predict the level of pollutants (e.g., sediments) in the water system. Non-point source pollutant models such as AGNPS and SWAT, by themselves, are useful tools in aiding decision makers with setting up best management practices to improve water quality. But, when these models are coupled to a decision support system such as the LWIDSS, their effectiveness is enhanced. The LWIDSS provides value-added decision support functions to the models which otherwise may be lacking or deficient. In addition, it can make the overall decision support process easier and more productive for decision makers. Figure 12 illustrates the user interface of the LWIDSS. We will explain the development of these decision support processes in detail next.

3.4 Land water integrated decision support system components

3.4.1 Review of inputs

One of the first steps in any modelling work is to identify all of the input that is required by the model. In the case of watershed-based non-point source models, the physical characteristics of the watershed need to be described to the model for it to simulate the physical processes such as the hydrology. GIS map layers are used to define the soil texture, the land use and the surface topography (Digital Elevation Model) of the Raisin River



Project

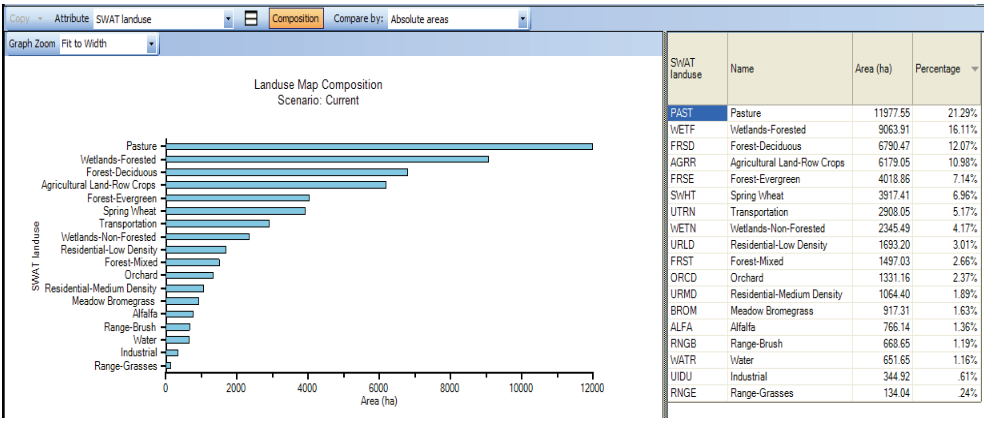
Scenarios

Potential Natural Vegetation
Ideal biodiversity conservation
Realistic biodiversity conservation
Current
Agricultural intensification with limited conservation
Agricultural intensification with no conservation

Model Workflow
Scenario: Current

SWAT2005
AGNPS

Fig. 12. User Interface of the LWIDSS



SWAT landuse	Name	Area (ha)	Percentage
PAST	Pasture	11977.55	21.29%
WETF	Wetlands-Forested	9063.91	16.11%
FRSD	Forest-Deciduous	6790.47	12.07%
AGRR	Agricultural Land-Row Crops	6179.05	10.98%
FRSE	Forest-Evergreen	4018.86	7.14%
SWHT	Spring Wheat	3917.41	6.96%
UTRN	Transportation	2908.05	5.17%
WETN	Wetlands-Non-Forested	2345.49	4.17%
URLD	Residential-Low Density	1693.20	3.01%
FRST	Forest-Mixed	1497.03	2.66%
ORCD	Orchard	1331.16	2.37%
URMD	Residential-Medium Density	1064.40	1.89%
BROM	Meadow Bromegrass	917.31	1.63%
ALFA	Alfalfa	766.14	1.36%
RNGB	Range-Brush	668.65	1.19%
WATR	Water	651.65	1.16%
UIDU	Industrial	344.92	.61%
RNGE	Range-Grasses	134.04	.24%

Fig. 13. Reviewing model inputs in the LWIDSS. A chart and table of land use composition in the watershed is presented to identify the primary uses of the land.

Watershed. In addition, non-spatial data such as climate data including precipitation and temperature may also be necessary to simulate rainfall and evapotranspiration (transport of water from the surface to the air). Reviewing the input data of the models within the LWIDSS using multiple formats including maps, graphs and tables (Figure 13) is an important aspect in the overall decision support process. It allows policy makers an opportunity to check and possibly confirm whether or not the results produced from the model seem reasonable. Going back and examining the input data can also help to get some insight into why one is getting certain results when they may be expecting something different. The ability of the LWIDSS to overlay data from several GIS layers allows policy makers the ability to identify areas where data quality may be poor and thus require more attention. The ability to integrate data allows more valuable information to be generated.

3.4.2 Model calibration and validation

Non-point source models are run for the first time using current watershed conditions to assess the present situation (i.e., the current water quality). The accuracy of the model prediction is evaluated by comparing the results to known observation data (water quantity and quality). If the predictions are poor, then model variables are adjusted and the model is re-run until the predictions are satisfactory (unique watershed characteristics not reflected by the input data and data accuracy can cause the model to behave differently). This iterative process is referred to as model calibration and validation. The LWIDSS assists in this procedure by providing a platform for examining and visualizing the model outputs in a variety of different ways such as through graphs and tables and for comparing the outputs to the observation data. Model output from non-point source models are typically expressed as amounts of pollutants in the water (i.e., concentration). Results can be quickly accessed because they are organized by location within the watershed and by predicted parameter (e.g., stream flow and sediment). Results can also be filtered by time period and can be summarized to a broader timeframe. Predicted model outputs are compared to observations using plots of observations over predicted results (Figure 14) or using statistics such as regression coefficient *R* or Nash Sutcliffe coefficient (Nash & Sutcliffe, 1970) for quantitatively assessing the prediction accuracy.

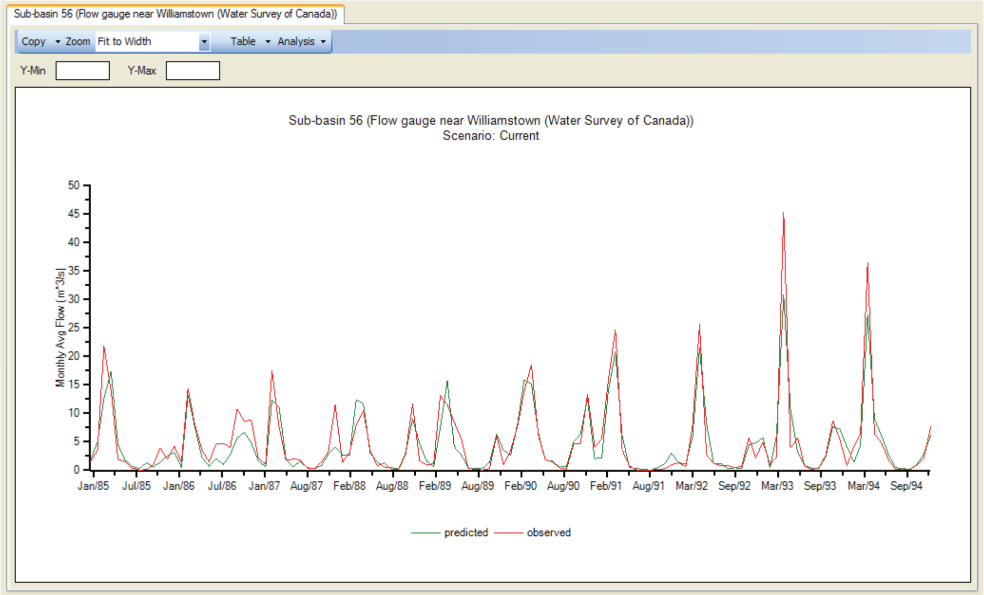


Fig. 14. Reviewing model results in the LWIDSS for the purpose of model calibration. A plot of predicted flow (model) and observed flow (from a monitoring station) is examined to gauge prediction accuracy.

3.4.3 Scenario gaming and comparison

After a model is calibrated, it is now ready to be used for prediction in other situations, real or hypothetical. These are commonly referred to as scenarios. Suppose, in the future, there will be a need to increase the amount of agriculture to produce higher crop yields. The question that needs to be addressed is: how much of an impact will this agricultural intensification scenario have on the environment? The only feasible and cost effective means of arriving at an answer is through modelling. This scenario requires that a new hypothetical land use map be developed by experts using dynamic landscape models. This land use map will undoubtedly involve increasing the amount of land available for farming. The LWIDSS should provide the ability for policy makers to review and compare the land use maps from multiple scenarios (Figure 16), thus allowing them to provide comments and feedback to the landscape modellers. The non-point source model is run using the new agricultural land use map (and with the other input data staying constant). Policy makers then use the LWIDSS to analyze the results for the new agricultural intensification scenario and to also perform a scenario comparison (Figure 15) against the current scenario (present day land use map) through scenario gaming (Wong et al., 2007). This allows them to assess how much the pollution is predicted to increase when compared to the current environment and for them to plan the next course of action.

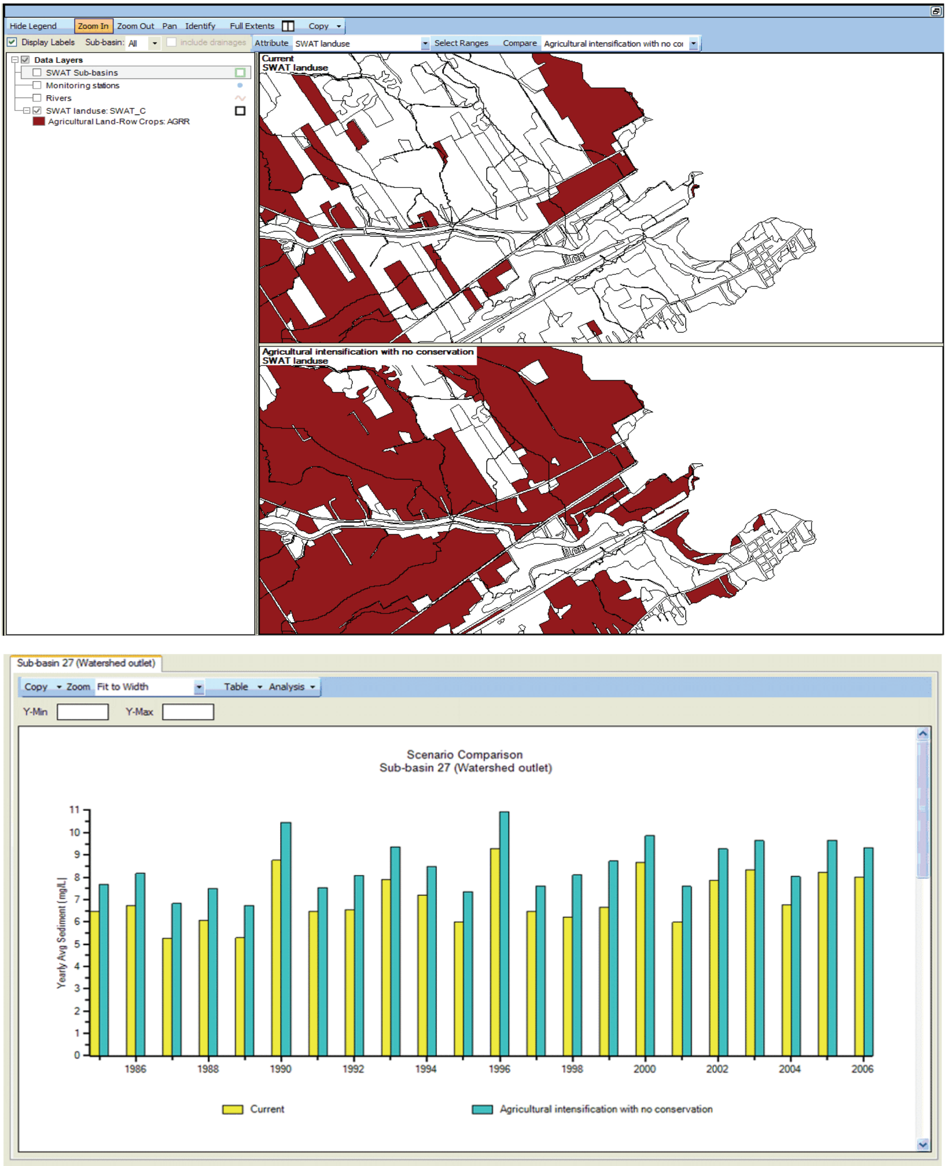
3.5 Integrated modelling assessment

If the policy makers, after examining the model results, determine (based on current standards or guidelines) that the agricultural intensification scenario produces too much pollution, then what can be done to reduce the environmental impact (while still maintaining similar levels of agriculture needs)? An LWIDSS which has the capability of locating sources of pollutants through a source tracing analysis would be valuable to a policy maker in finding areas (“hot spots”) whose pollution contribution exceeds a set threshold. These hot spots can then be targeted for change. They can be mapped and this information feeds back to the dynamic landscape model so that it uses this knowledge to update the existing agricultural land use map by applying best management practices to the hot spot areas. For example, filter strips could be added near water bodies to prevent agricultural pollutants from entering water resources. The non-point source model is then re-run and the results are compared to the previous set in the LWIDSS (Figure 16). If the amount of pollution reduction is unacceptable, then the set of target areas are updated and the whole feed back mechanism between the non-point source model and the landscape model is repeated until an “optimal” solution for best management practices is found.

4. Conclusions

Environmental decision support systems have been implemented in Environmental Canada. They are proved to be practical. Not only the modellers increase their productivity with the assist of the EDSSs, government officials and policy makers also gain valuable insights by using these EDSSs to set policy guidelines and objectives.

The EEM-SAT DSS has been tested by the EEM National office as well as by external users and has proven to substantially improve the ability of the user to generate accurate and consistent analyses that are required under the Canadian EEM program. The responses from external users have been positive. A significant proportion of the time and effort spent on



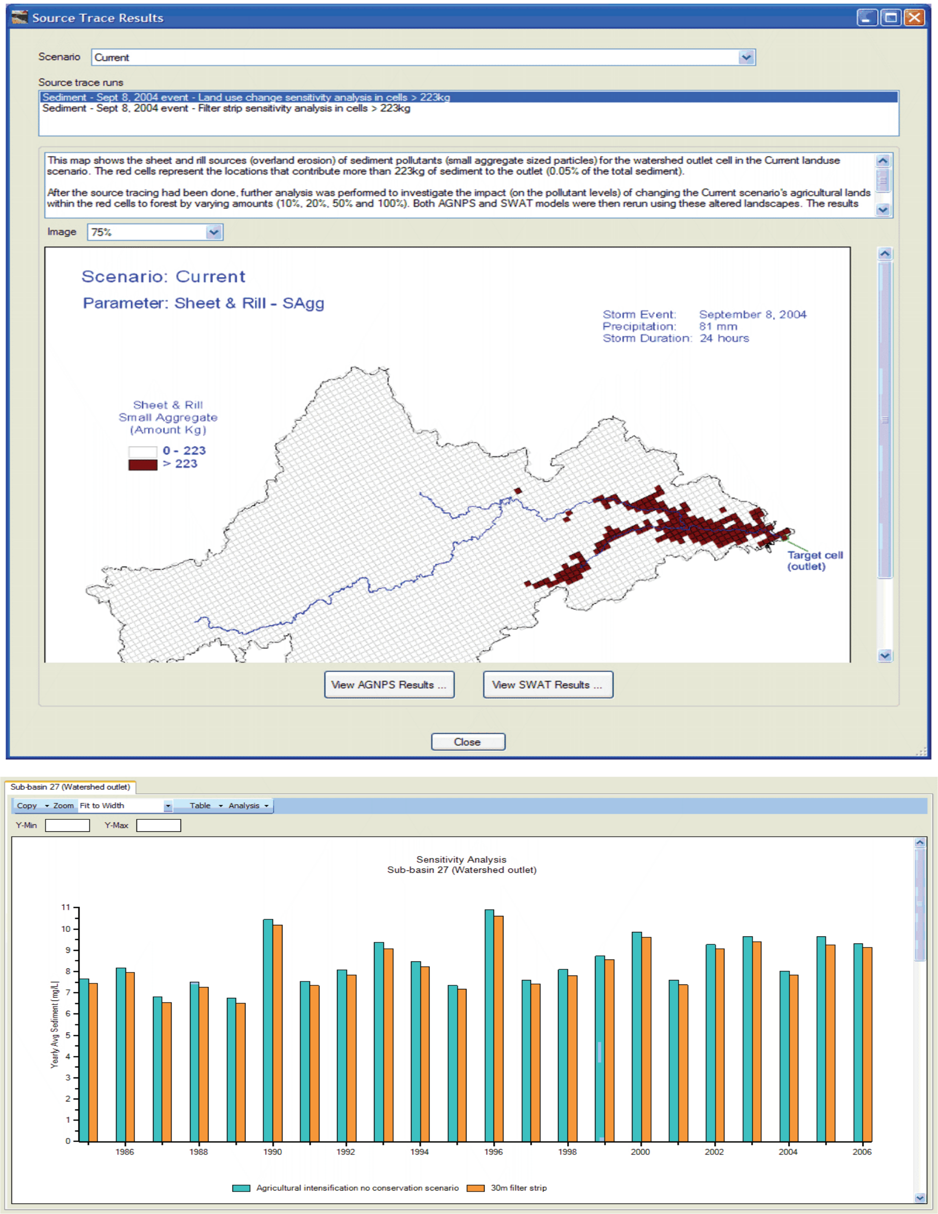
☑ Display Labels Sub-basin: All ☐ include drainages Attribute SWAT landuse Compare Agricultural intensification with no co

Sub-basin 27 (Watershed outlet)

Copy Zoom Fit to Width

Y-Min Y-Max

Fig. 15. Reviewing model input (top) and results (bottom) in the LWIDSS for the purposes of scenario comparison. Top: Maps comparing land use for the two scenarios (top map: current; bottom map: agricultural intensification). Shaded areas are agricultural lands. Bottom: a chart comparing sediment concentrations predicted at the watershed outlet between two land use scenarios is examined to investigate the impact of land use to water quality.



Source Trace Results

Scenario: Current

Source trace runs

Sediment - Sept 8, 2004 event - Land use change sensitivity analysis in cells > 223kg

Sediment - Sept 8, 2004 event - Filter strip sensitivity analysis in cells > 223kg

This map shows the sheet and rill sources (overland erosion) of sediment pollutants (small aggregate sized particles) for the watershed outlet cell in the Current landuse scenario. The red cells represent the locations that contribute more than 223kg of sediment to the outlet (0.05% of the total sediment).

After the source tracing had been done, further analysis was performed to investigate the impact (on the pollutant levels) of changing the Current scenario's agricultural lands within the red cells to forest by varying amounts (10%, 20%, 50% and 100%). Both AGNPS and SWAT models were then rerun using these altered landscapes. The results

Image: 75%

Scenario: Current

Parameter: Sheet & Rill - SAgg

Storm Event: September 8, 2004
Precipitation: 81 mm
Storm Duration: 24 hours

Sub-basin 27 (Watershed outlet)

Fig. 16. Reviewing results from source tracing analysis. Top: map displaying locations in the watershed that contribute higher amounts of sediment to the watershed outlet is analyzed to find potential target areas for best management practices. Bottom: a chart comparing sediment concentrations before and after best management practices (application of filter strips).

developing the EEM-SAT DSS involved the data entry design, expert system rulebase for data preprocessing and QA/QC, expert system rulebase for logic control of the endpoint analyses for both fish and benthic invertebrates, as well as the rulebase for the site effects analyses and current status reporting.

The release of EEM-SAT DSS makes the statistics requirements of the Canadian EEM robust and transparent. It reduces the likelihood of incorrect data retrieval, improper data manipulation, misuse of statistics procedures and misinterpretation of the statistics results. With repeated beta testing and user usability analysis, the system has become a seamless and friendly user interface for its users based on sound science to ensure accuracy and effectiveness for users of all levels. The system addresses three key improvement areas that otherwise requires a considerable amount of time and effort by the user. These areas include proper data manipulation, the statistics analyses and results interpretation. Efforts were also made to ensure that the design would be generic enough to allow the system to work for other sector data analyses such as metal mining.

Designed for policy makers, the LWIDSS system is an effective analytical, planning and management tool to interpret and report modelling and scenario gaming results of the watershed-based modelling. Its framework can be used to determine the impact assessment of various land use scenarios on sediment and nutrients. The integrated decision support processes are captured in the LWIDSS and can be used for other similar watersheds if appropriate data is available. The results of the modelling and scenario gaming can be linked with other systems using a common data exchange interface.

The design of the LWIDSS facilitates integration of diverse information, ranging from various land use scenario map layers, to soil texture map layer, to Digital Elevation Model data, and to water quality data. It is also designed to provide an opportunity to query a variety of databases, to visualize spatial and/or temporal patterns, and to analyze model input data and output results using the DSS customized tools. Inherent to the LWIDSS architecture are relational databases with common design structures for data integration to be used in the modelling and scenario gaming framework. The use of modelling and scenario gaming will allow decision-makers to explore potential responses of land and water integration to hypothetical situations, i.e. to answer the "what-if" question. The feedback loop among the models is important for the policy makers to explore the best possible management options and the course of action for pollution control.

We are continuing to strive to improve the development of future EDSSs. The use of automation and expert system rules to help guide modellers in the decision process will greatly reduce the uncertainty by using the appropriate tools and approaches. In addition, bringing the EDSSs to the web will allow timely update of models, data and information thus increases the usability of the EDSSs by removing data and model inconsistency and duplication.

5. References

Alter, S.L. 1980. *Decision support systems: current practice and continuing challenges,* Addison-Wesley, Reading, MA.

Arnold J.G., Srinivasan R., Muttiah R.S., Williams J.R. 1998. Large area hydrologic modeling and assessment. Part I, model development. *Journal of the American Water Resource Association* 34(1):73–89.

Booty, W.G., Lam, D.C.L., Wong, I.W.S., and Siconolfi, P. 2001. Design and implementation of an environmental decision support system. *Environmental Modelling and Software* 16, pp. 453-458.

Buchanan, B.G. and Shortliffe, E.H. 1985. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, Addison-Wesley, Reading, MA.

Environment Canada. 2000a. Data Interpretation Guidance for Environmental Effects Monitoring. *National EEM Office*, Environment Canada, Ottawa, Ontario.

Environment Canada 2002. Metal Mining EEM Guidance Document for Aquatic Environmental Effects Monitoring, *National EEM Office*, Science Policy and Environmental Quality Branch, Ottawa Canada.

Ignizio, J.P. 1991. *Introduction to Expert Systems: The Development and Implementation of Rule-Based Expert Systems*, McGraw-Hill, 402 pp.

Lam, D.C.L., Puckett, K.J., Wong, I., Moran, M.D., Fenech, G., Jeffries, D.S., Olson, M.P., Whelpdale, D.M., McNichol, D., Mariam, Y.K. and Minss, C.K. 1998. Anintegrated acid rain assessment model for Canada: from source emission to ecological impact, *Water Quality Res. J. Canada*, 33, 1-17.

Lam, D.C.L., Leon, L.F., Hamilton, S., Crookshank, N., Bonin, D. and Swayne, D.A. (2004). Multi-Model Integration In A Decision Support System: A Technical User Interface Approach For Watershed And Lake Management Scenarios. *Environmental Modelling and Software* 19, pp. 317-324.

Leon, L.F., Booty, W.G., Bowen, G.S., and Lam, D.C.L. 2004. Validation of an agricultural non-point source model in a watershed in southern Ontario, *Agricultural Water Management*, 65, 59-75.

Lowell, R.B., Hedley, K, and Porter, E. 2002. Data Interpretation Issues for Canada's Environmental Effects Monitoring Program. *Water Qual. Research J. Canada* 37(1), pp.101-117.

Nash, J.E., and Sutcliffe, J.V. 1970. River flow forecasting through conceptual models: Part I – A discussion of principles, *Journal of Hydrology*, 10, 282-290.

SYSTAT 11 Statistics II, 2005. "*General Linear Models*", Pages 139-205, Richmond, CA

Walker, S.L., Hedley, K, and Porter, E. 2002. Pulp and Paper Environmental Effects Monitoring in Canada: An Overview. *Water Qual. Research J. Canada* 37(1), pp. 7-19.

Wong, I., Bloom, R., NcNicol, D.K., Fong, P., Russell, R. and Chen X. 2007. Species at risk: Data and knowledge management within the WILDSPACE decision support system, *Environmental Modelling and Software*, 22, pp. 423-430.

Young, R., Onstad, C., Bosch, D. and Anderson, W. 1987. *Agricultural non-point source pollution model: A watershed analysis tool*, Conservation Research Report 35, Agricultural Research Service, U.S. Department of Agriculture, Washington, D.C.

Expanding Decision Support Systems Outside Company Gates

Petr Bečvář[1], Jiří Hodík[2], Michal Pěchouček[2], Josef Psutka[3],
Luboš Šmídl[4] and Jiří Vokřínek[2]
[1]CertiCon a.s.
[2]Czech Technical University in Prague, Faculty of Electrical Engineering, Department of Cybernetics;
[3]University of West Bohemia, Pilsen;
[4]Center of Applied Cybernetics
Czech Republic

1. Introduction

This chapter presents a set of different decision support systems to demonstrate different ways in which the decision support process can be extended outside a single company.

First, the ExtraPlanT system is introduced. It was originally built as a multi-agent production planning system (ExPlanTech), but later it was enhanced with a wide range of other features. One of them is a set of interfaces for remote access including a thin and thick client, an interface for portable devices and a voice interface based on an automatic speech synthesis. This system demonstrates a possibility of how the decision-making process can be extended – towards the user. A manager on a business trip as well as a tele-worker or a nomad-worker has full access to the decision support system. ExtraPlanT system also contains an enterprise-to-enterprise planning agent and a material-resource-agent. Both agents support cooperation between trusted partners by sharing information in the decision support process. This technology allows an outsourcing company or a material supplier, equipped with a compatible system, to share available capacities and obtain basic planning information.

The second system introduced is called e-Cat. It is designed as a competency register for a Virtual Organization, a consortium or a professional community. This web catalogue system combines centralized and de-centralized elements in order to preserve member's autonomy while ensuring data consistency and safety of the community. Emphasis is put on the intelligent match of searched and available competencies. If the exact match cannot be reached, a generalization and a specialization can be used to find a suitable partner. The main use of the e-Cat system is in formation and evolution of Virtual Organization.

The last system is called DSS (Decision Support System) and it is directly focused on the operational management of Virtual Organizations. It supports the Virtual Organization manager by simulating possible future performances of the organization. The embedded scheduler provides the user with a wide range of additional features: impending deviations alerting, suggestions of local adaptation, alternative of Virtual Organization configuration analyzer, what-if analysis, and other.

2. ExtraPlanT multi-agent system for production and supply chain planning

Besides collecting and presenting data to the user, the modern decision support systems also provide extensive analysis of these data. Such analysis can include simulation, knowledge-based analysis, data mining, and other approaches based on artificial intelligence. These modern approaches require a great deal of computer resources and direct access to company data sources, which can be a drawback in modern, de-centralized and nomadic environment. A straightforward solution to this problem is an application server that runs all the analysis within company gates and user interface clients that make the results accessible worldwide (Shen at al., 2005). These clients can be *thick* clients which must be installed by the user or *thin* clients which run in the user's web browser.

For users within the enterprise, *thick client* technology is usually used: the unformatted results are transferred and presented by a client program running on a personal computer. *Thin client* technology is more suitable for users outside the enterprise, so that the results are formatted by the application server into a presentable format (e.g. HTML, ECMA Script[1] or ActionScript[2]) and presented to the user by a standardized device.

The potential of thin and thick client approach is demonstrated in this section using an example of ExtraPlanT decision support system and their Extra-enterprise and Enterprise-to-enterprise features (Bečvář at al., 2007).

While original ExtraPlanT interfaces adhered to the concept of thin client, contemporary interfaces usually use Rich Internet Application technique to enrich thin clients with the features originally common to thick clients.

2.1 Internal system architecture

The ExtraPlanT decision support system has been developed by Gerstner Laboratory at the Czech Technical University in Prague, the CertiCon Company[3].

ExtraPlanT is an extension of ExPlanTech (Pěchouček at al., 2002b) intra-enterprise production scheduling in *Small and Medium-sized Enterprises* (SMEs) with project-oriented production. ExPlanTech implements a ProPlanT (Mařík at al. 2000) reference multi-agent architecture. The system is composed of non-trivial agents providing different system functionality. The number of agents changes only to reflect changes in the factory configuration or to add system features.

The basic set of agents that have to be always present in ExPlanTech system is called *Core Agents* and it implements planning and resource management. The purpose of this core functionality is to aid a human user in determining the resources needed for a particular project, creating production plans, and balancing the utilization of internal resources, and also to warn the user if there are insufficient resources for the existing project.

The Core Agents can be extended by a number of optional *Additional Agents* implementing other related features. In ExPlanTech project, these agents were responsible for interaction with legacy factory systems, and they provided thick client user interfaces.

The Set of Additional Agents was largely extended within ExtraPlanT project – three types of agents were added:

- Agents for performance measurement and supervision

[1] http://www.ecma-international.org/publications/standards/Ecma-262.htm

[2] http://www.adobe.com/devnet/actionscript/

[3] http://www.certicon.cz/

- Agents providing various kinds of user interfaces
- Agents for extra enterprise cooperation.

An example of the configuration is depicted in Figure 1.

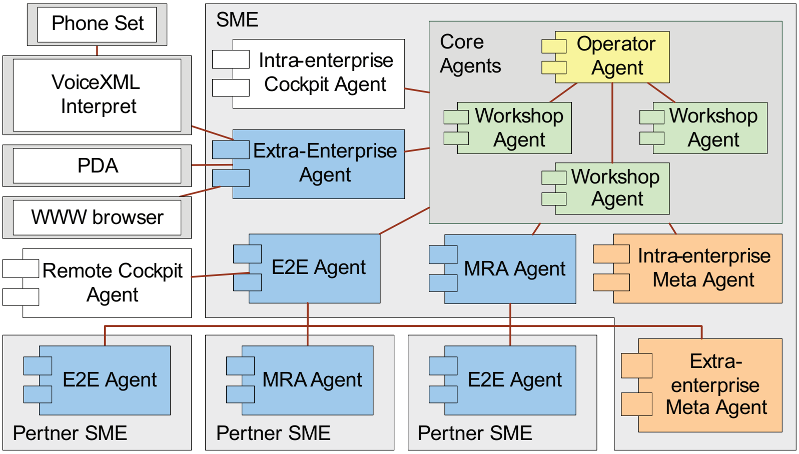


Fig. 1. Simplified architecture of the ExtraPlanT system.

The first group is represented by the *Intra-Enterprise Meta-Agent*, dedicated to the visualization and analysis of medium and long-term manufacturing processes (Hodík at al., 2005) and *Extra-enterprise Meta-Agent* for the analysis of external cooperation with suppliers and partners.

The second group of Additional Agents provides a variety of possibilities for the user to interact with the decision support system. It implements several versions of thin and thick clients mentioned in Section 2 and further described in Section 2.2.

The third group contains agents for extra enterprise cooperation. In ExtraPlanT system *Material Resource Management Agent* and *Enterprise-to-Enterprise Agent* were developed to help optimizing material resource manipulation and supply chain relationships (see Section 2.3) (Pěchouček at al., 2005).

2.2 Extra-enterprise user interfaces

A significant part of ExtraPlanT project focused on the availability of decision support outside the company network. At the dawn of networked or virtual organizations, nomadic working and globalized economy, it was clear that managers of even small enterprises need to make decisions anywhere and with proper support. The project attempted to exploit all the technology available at that time to provide decision support in such scenarios. The set of use cases also included an option to offer a limited version of this access to a trusted partner or customer to be able to track the work of their order.

The thin client technology consists of the *Extra-Enterprise Agent* (EEA) (Hodík at al., 2005) that acquires data from Core Agents and relays them to a WEB-tier of Java EE[4] multi-tiered application running an application server. The EEA is a member of the agent community

[4] http://java.sun.com/javaee/index.jsp

and it uses an asynchronous inter-agent messaging to communicate with other agents. *FIPA Request Protocol* is used to query other agents periodically for new data and *FIPA Subscribe Protocol*[5] is used when the other agent is able to inform EEA autonomously about data changes.

The WEB-tier converts data into HTML pages etc., thus making them viewable using standard devices. Four versions of thin client were implemented: standard HTML output to be viewed by WWW browser, simplified HTML layout for PDA devices (XHTML was not an adopted standard yet), WML cards to be viewed by WAP enabled cell phones (an example of technology that faded away before becoming widely used), and VoiceXML to be processed by VoiceXML Interpret and accessed by common telephone set (see section 2.4).

The thick client technology (also named *Remote Cockpit Agent*) assumes an installation of dedicated software on a user's computer, but it later provides the user with a more convenient way to handle the decision support system. Strong emphasis is put on the security of extra-enterprise communication: public key cryptography in common with the secure JADE platform and HTTPS protocol is employed to provide the maximal security level.

2.3 Agents for Enterprise-to-Enterprise cooperation

Agents for enterprise-to-enterprise (E2E) cooperation represent an early attempt to involve external partners directly into the decision support process. Unlike the concept of virtual organizations introduced below, the E2E solution used in ExtraPlanT is mainly viewed from the point of view of a single company. Also, the whole solution requires the installation of dedicated software on the application server of each company involved in the cooperation.

Enterprise-to-enterprise Agent

This agent connects ExtraPlanT system to compatible external software systems. The main mission of the E2E Agent is to augment the decision support process by proactively exchanging data with extra enterprise partners. This technique is focused on finding possibilities of tasks outsourcing in the case of exhausted local resources and on advertising free local capacities to the cooperators.

On the enterprise-to-enterprise level, each partner is represented by one agent and these agents connect together on a pear-to-pear basis. The background of E2E Agents is hidden for the others so that the cooperation does not depend on a particular infrastructure used by each partner.

Knowledge-based social model helps E2E Agents to propose which partner could best cover the needs of the company. The agent can take into account various aspects of cooperation and make decision upon many criteria (such as reputation, history of cooperation, recommendation, due dates, price, expected quality, etc.). This model is continuously updated with every new contract or information.

Material Resource Agent

Material Resource Agent (MRA) was designed to support the decision process in another major task of supply chain management. A material resource handling differs from free capacities sharing (material resources are storable, often requiring precise specification…) and, therefore, a dedicated agent with different control algorithm and data access is required.

[5] http://www.fipa.org/

The material resource handling is implemented on both intra-enterprise and enterprise-to-enterprise level. On the intra-enterprise level, the MRA is equipped with an adequate user interface or a database connection to be able to obtain data from case specific data sources. On the enterprise-to-enterprise level, the communication abilities have been enhanced to allow exchange of relevant data in a community of MRA and E2E Agents.

2.4 Voice interface

The ExtraPlanT telephony interface was implemented using a VoiceXML technology. VoiceXML is a standard[6] that allows dividing telephony application into two independent parts by defining the interface between them – the VoiceXML language. The parts are (i) low-level speech recognition and synthesis engine (*VoiceXML Interpreter*) and (ii) high-level domain-dependent application. The two modules and their interaction are depicted in Figure 2.

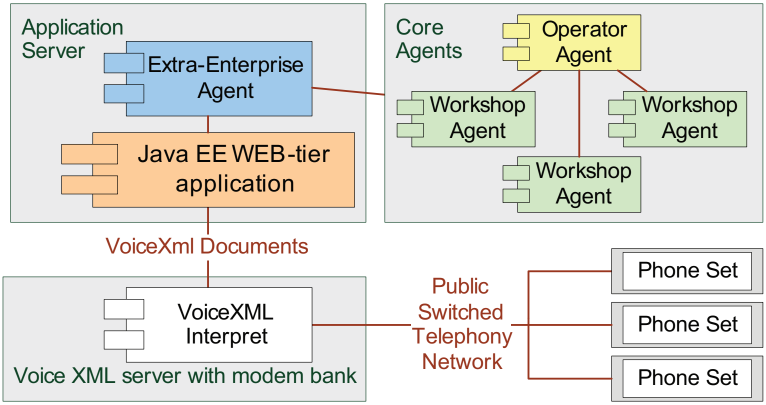
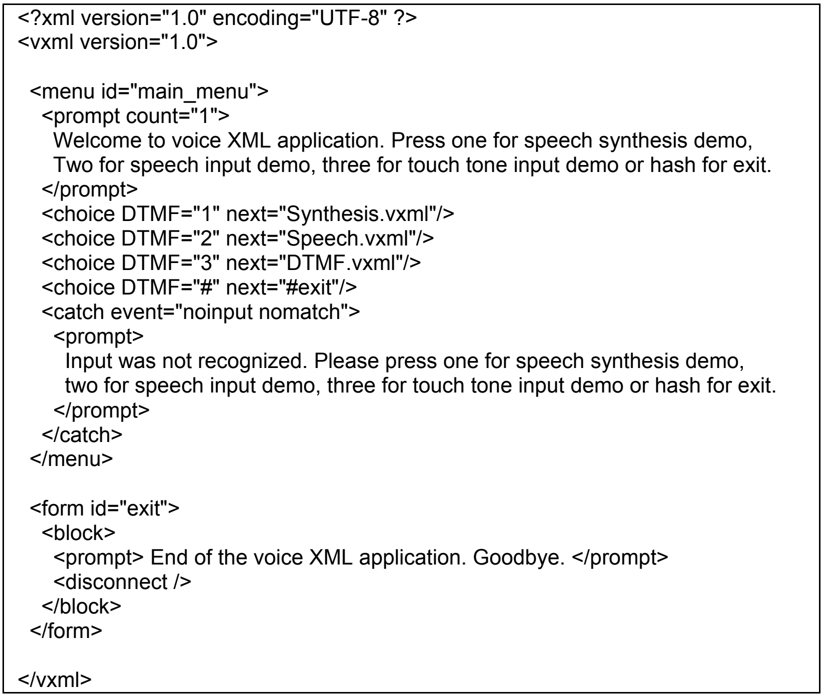


Fig. 2. Structure of the telephony interface and its connection to ExtraPlanT Core Agents.

The ExtraPlanT telephony interface was implemented using a VoiceXML Interpreter (Šmídl at al., 2002), which is available in the Center of Applied Cybernetics at the University of West Bohemia in Pilsen[7]. This interpreter uses an engine for speech recognition and synthesis (Müller at al., 2000) which was developed in the Department of Cybernetics at the same university in cooperation with the SpeechTech[8] company. The engine contains a text-to-speech module (Matoušek at al., 2004) that produces intelligible and natural speech.

The high-level application does not need to be concerned with speech processing and instead focuses on the structure of dialogs and transferred data. The interaction between the user and the system is viewed as a sequence of dialogs described by the VoiceXML documents. Dialogs can feature synthesized speech, digitized audio, recognition of spoken speech and DTMF (Dual-tone multifrequency, also known as Touch Tone) key input. An example of a VoiceXML Document is depicted in Figue 3.

[6] http://www.w3.org/TR/voicexml/
[7] http://ui.kky.zcu.cz/en
[8] http://speechtech.cz/index-en.php



```
<?xml version="1.0" encoding="UTF-8" ?>
<vxml version="1.0">

  <menu id="main_menu">
    <prompt count="1">
      Welcome to voice XML application. Press one for speech synthesis demo,
      Two for speech input demo, three for touch tone input demo or hash for exit.
    </prompt>
    <choice DTMF="1" next="Synthesis.vxml"/>
    <choice DTMF="2" next="Speech.vxml"/>
    <choice DTMF="3" next="DTMF.vxml"/>
    <choice DTMF="#" next="#exit"/>
    <catch event="noinput nomatch">
      <prompt>
        Input was not recognized. Please press one for speech synthesis demo,
        two for speech input demo, three for touch tone input demo or hash for exit.
      </prompt>
    </catch>
  </menu>

  <form id="exit">
    <block>
      <prompt> End of the voice XML application. Goodbye. </prompt>
      <disconnect />
    </block>
  </form>

</vxml>
```

Fig. 3. Example of a VoiceXML document as used in the telephony interface.

A telephony interface, like any other type of interface, has certain advantages and drawbacks. The main advantage is its accessibility in situations when a computer cannot be used. Interaction with the system is natural and does not require any special training.

However, speech has certain disadvantages when used as a computer system's output. The speech is sequential and time-consuming, meaning that the user hears pieces of information one at a time, requiring him or her to remember all the previous messages. If a large amount of information is necessary for a particular decision, it is, therefore, difficult to ensure that the user remembers everything at the appropriate moment (Balentine & Morgan, 1999).

Two techniques have been developed to help mitigate this disadvantage: *Dynamic Prompt Wording* and the *Two-level Communication Model*.

Dynamic Prompt Wording

Every user who has some frequent experience with common telephony applications (e.g. phone banking) becomes annoyed with lengthy interaction. Explanations and polite phrases that are useful during the first interactions soon become useless and obstructing. A telephony interface of decision support systems is used daily by a small group of trained users. To optimize the descriptiveness of the machine output according to a user's experience, the *Dynamic Prompt Wording* technology has been developed.

The wording of each utterance is selected dynamically from among several possible versions. The longest, most detailed versions are used for novice users. For more experienced users, the interface chooses a shorter version, without explanations or polite phrases. All versions of all utterances used in one dialog are defined in a XML file and they are selected dynamically according to the user's experience with the particular dialog (Bečvář at al., 2004).

Two-level Communication Model

Spoken presentation, unlike graphical interface, is not able to display information in parallel. While a graphical interface can simply display a graph and let the user select from it any desired information, the user of speech interface has to receive pieces of information one by one (Balentine & Morgan, 1999).

Two possible solutions are available: (i) to transmit all data and let the user record and analyze them or (ii) to analyze the data on the server side and transmit only the desired conclusion.

The *Two-level Communication Model* presumes the use of both solutions. If approach (i) is used, we refer to it as the *Fact-finding Mode*, while approach (ii) is called the *Analytical Mode* of the interface.

The Analytical Mode requires a new component, called the *Analytical Module,* which uses the knowledge based approach to estimate which information is important for the user, to obtain relevant data and to present the information early in the conversation in the form of short summaries (Bečvář at al., 2007).

If the problem is detected, further information is usually necessary in order to solve it. The telephony interface then switches to the Fact-finding Mode, which presumes that the user is prepared to record the received information for later analysis. In the Fact-finding Mode the interface simply transmits all the facts that are potentially relevant.

3. Virtual organizations

SMEs usually cannot satisfy complex customer needs on their own, so collaboration is needed to cover all the business aspects (Boughzala & Zacklad, 1999) (Říha at al., 2002). To increase their competitiveness and to gain more business opportunities, SMEs form alliances.

The set of individual partners (SMEs) that share information about their resources, all agreeing to form possible coalitions, is called *alliance*. The alliance is regarded as a long-term cooperation agreement among the partners. A *coalition* is defined as a set of partners who agreed to fulfill a single, well-specified goal. A coalition, unlike an alliance, is thus usually regarded as a short-term agreement between collaborative partners (Pěchouček et al., 2002a). In contemporary business terms, the coalition of SMEs is commonly called Virtual Organization (VO).

The VO is a kind of networked organization, having been defined many times. Common key-terms related to the VO are: (i) specific form of a network organization, (ii) formed by autonomous and mutually independent partners, (iii) single body towards the customer, (iv) dissolved after its mission fulfillment and (v) allowing individual partners to concentrate on their core competencies. Some of definitions also require the following features for the networked organization to be accepted as a VO: (vi) sharing risk with partners, and (vii) use of information technology for the organization coordination.

The *VO lifecycle* (Camarinha-Matos & Afsarmanesh, 1998) consists of four main phases: *creation*, *operation*, *evolution* (or adaptation) and *dissolution,* as shown in Figure 4.

The decision making process in the environment of a VO differs from the same process in a closed, centralized organization in several important ways. In a closed organization the manager has full access to internal information of the managed units as well as full control

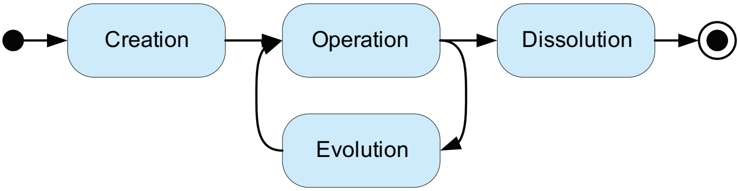


Fig. 4. VO life-cycle

over them. The VO manager is limited by the concluded contract and the willingness of managed VO members to cooperate, provide information and update the contracts.

Information and communication systems are crucial for effective decision making in VO (Adesta, 2005). Although standard tools (web sites, emails, databases, etc.) work well, they are separate and do not offer a possibility of integrating knowledge from various sources and of ensuring common and consistent view for all partners. Section 4 describes a profile catalog tool mainly intended for the creation and evolution phase of VO, while Section 5 deals with a simulation tool usable in the creation, operation and evolution phase.

4. e-Cat – system for competency management in virtual organization

Unlike ExtraPlanT, the e-Cat system was intended for use in networked environment of alliance of SMEs from the beginning. As another example of early systems, it focused on a single aspect of virtual organization and on a single stage of virtual organization life cycle.

e-Cat is a prototype of a tool designed to support partner search during the creation or evolution phase of VO lifecycle by proper and consistent profile and competency management. If SMEs are able to quickly find a partner with appropriate abilities, competencies and services, they can together collaboratively cover new business opportunities.

4.1 Competency management

Since the terms "competency", "competency class", "competency instance" and "profile" are used in several slightly different meanings, it is important to summarize the competency management terminology (Biesalski, 2003).

For the purpose of e-Cat the definitions presented in (Hodík at al., 2009) are used:

- **Competency** is an ability to perform business processes which are supported by necessary available resources, practices and activities, allowing the organization to offer products/services.
- **Competency class** defines the existence of the competency in the world; it distinguishes it from other existing competencies and defines relations to them. Competency class may be extended by the definition of means used to measure the level and robustness of the competency. If a class is not specific enough, specializing classes may exist. Thus, every competency class can have its generalizing and specializing class(es).
- **Competency instance** always refers exactly to one competency class and to one subject. If the competency class defines Competency Evidences, the competency instance can optionally assign values to them.
- **Profile of subject** contains two main elements: (i) general information about the partner, and (ii) a set of competency instances derived from the competency classes.

4.2 The e-Cat system

The e-Cat is an agent-based prototype of a tool for competency management in alliances of SMEs. The technology used is based on a distributed set of agents, representing individual members supported by centralized elements. Such a hybrid peer-to-peer network architecture (Friese at al., 2002) enables effective cooperation in a heterogeneous distributed environment where agents ensure maximal independence between alliance members and private knowledge preservation.

Data are stored and exchanged in the XML format. The schema describing competency classes and profiles is based on HR-XML[9] schema designed for competency description in human resources management. Some elements that are meaningful only for human resource management have been removed, and the schema was extended by means of defining taxonomy of competencies.

For each competency class, a set of generalizing, specializing and related competencies can be defined. Cycles are not allowed. If the competency class has no generalizing competency, it is one of the *roots* of taxonomy structure. If the competency class has multiple generalizing competencies, it will be displayed on multiple positions in taxonomy trees in the user interface.

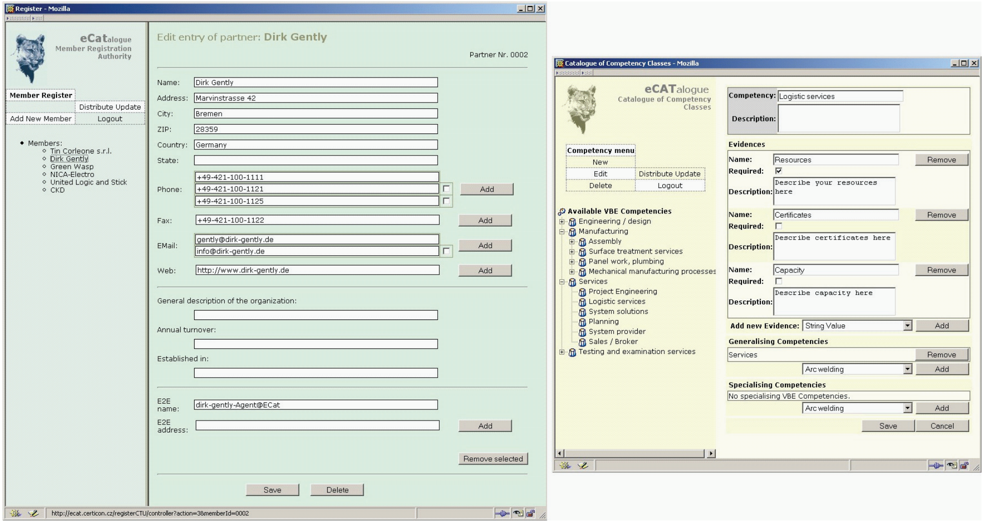
Alliance members provide other members with its profile containing their competencies. Each competency in the profile is inherited from a competency class defined in the Catalogue of Competency Classes. The user interface to the system is provided by thin clients through ordinary web browsers. Some screenshots of the interfaces can be found in Figure 5.

The e-Cat consists of following subsystems:

- **Distributed Profile Catalogue** keeps, manages and distributes members' profiles. Each member is responsible for keeping his profile up-to-date and consistent with set of competency classes, shared within the alliance. e-Cat system distributes the profile to other members and provides read-only access to profiles of other members. The catalogue is equipped with an intelligent search tool able to use generalizing and specializing competencies when an exact match cannot be found.
- **Catalogue of Competency Classes** provides partners with a list of competency classes that can be advertised through e-Cat. Each class is defined by their exact description, attributes and location within competency taxonomy. The Catalogue of Competency Classes is a centralized element, maintained by a competency expert. It ensures coherence in the common schema of competencies.
- **Members Registration Authority** is another centralized element that maintains a list of members of the alliance. For each member, all the data necessary for identification in the real and virtual environment are kept. The centralization allows the alliance management to control the members entering and leaving the community and prevents a member from pretending to act as another company.

These systems are implemented by three types of agents: *Members Registration Authority Agent*, *Catalogue of Competency Classes Agent* and *SME Agent*. The first two agents with appropriate user interfaces are intended to be deployed on alliance management servers and maintained by responsible experts of alliance support institutions. Each alliance member is represented by single SME Agent and each agent may be deployed on member's servers (see Figure 6). To increase the utilization of computer resources, several members can share a

[9] http://ns.hr-xml.org/23/HR-XML-23/CPO/Competencies.html



- Members:
  - Tin Corleone s.r.l.
  - Dirk Gently
  - Green Wasp
  - NICA-Electro
  - United Logic and Stick
  - CKD

Edit entry of partner: **Dirk Gently**

Partner Nr. 0002

Name: Dirk Gently
Address: Marvinstrasse 42
City: Bremen
ZIP: 28359
Country: Germany
State:
Phone: +49-421-100-1111
+49-421-100-1121 □
+49-421-100-1125 □
Fax: +49-421-100-1122
EMail: gently@dirk-gently.de
info@dirk-gently.de □
Web: http://www.dirk-gently.de

General description of the organization:

Annual turnover:

Established in:

E2E name: dirk-gently-Agent@ECat
E2E address:

http://ecat.certicon.cz/registerCTU/controller?action=38memberId=0002

Available VBE Competencies
- Engineering / design
- Manufacturing
  - Assembly
  - Surface treatment services
  - Panel work, plumbing
  - Mechanical manufacturing processes
- Services
  - Project Engineering
  - Logistic services
  - System solutions
  - Planning
  - System provider
  - Sales / Broker
- Testing and examination services

Competency: Logistic services
Description:

Evidences

Name: Resources
Required: ☑
Description: Describe your resources here

Name: Certificates
Required: □
Description: Describe certificates here

Name: Capacity
Required: □
Description: Describe capacity here

Add new Evidence: String Value

Generalising Competencies

Services
Arc welding

Specialising Competencies

No specialising VBE Competencies.
Arc welding

Fig. 5. Screenshots of the e-Cat user interfaces

single server and deploy their agents on a third-party computer. SME Agents can also be deployed on alliance management servers. In an extreme case, the distributed part of the catalogue can be omitted and the whole system can run on a single central server.

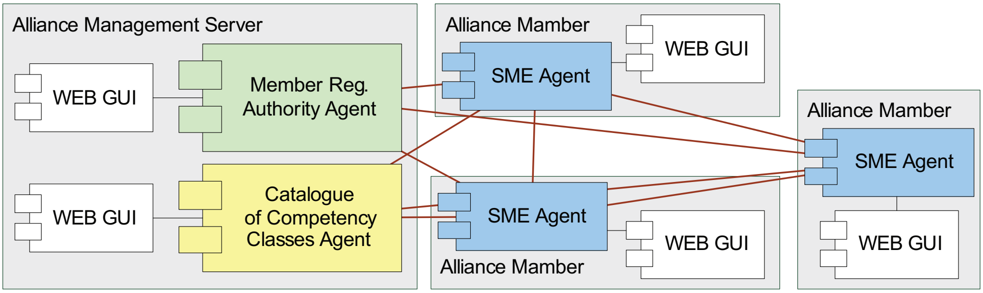


Fig. 6. Deployment diagram of centralized and decentralized elements of e-Cat

e-Cat system is designed to be robust against failure of any of its agent or communication link. Each agent maintains local data cache, which is used during the inaccessibility of another agent, and which is replaced by actual data as soon as the connection is restored. Centralized elements provide alliance members with: (i) ensuring the common understanding of competencies within the whole alliance, and (ii) maintaining identification information of alliance and thus maintaining the security measures against intruders or alliance members with malicious behavior.

4.3 Usage scenario

Several user scenarios were considered to be performed with e-Cat:

- **Joining the alliance and creating new profile.** A new member of the alliance installs required software on the company server or another available server. The newly installed system requires addresses of Members Registration Authority Agent and Catalogue of Competency Classes Agent to operate. When the new installation of SME Agent is finished, the company can be included in Members Registration Authority and thus be added to the alliance. From now on, agents of other alliance members can obtain the address of a new agent from the Members Registration Authority so that they can exchange the profile information.

- **Announcing a competency.** When an alliance member decides to offer some services to other members, the competency class for such services is instantiated in its profile. The appropriate competency class is found in Catalogue of Competency Classes. If the proper class does not exist in the catalogue, it can be added in cooperation with the manager of Catalogue of Competency Classes, or a generalizing competency can be used.
- **Creating a new competency.** When a request for adding a new competency class to the catalogue appears, the competency expert reviews the request and decides whether or not to accept it and adapt the catalogue. SME Agents are automatically notified once the catalogue is updated.
- **Looking for a provider of a competency.** The search engine of the e-Cat system offers various attributes for finding a potential partner in the alliance. If the search result is unsatisfactory, the user can decide to use the taxonomy structure to find a partner providing generalizing or specializing competency.

4.4 Exploitation

The prototype of the e-Cat system was used by CertiCon Corporation to develop an IRIS system – a competency catalogue portal application. The IRIS system was developed for a CertiCon customer to be a central point of an alliance of SMEs carrying business in IC design. In contrast to e-Cat, the IRIS system is not based on de-centralized multi-agent architecture but utilizes classical database and application server architecture.

IRIS system uses a concept of competency classes and competency instances that proved successful in e-Cat system. It is extended by a set of other features like more advanced competency search and extended profile of alliance member. IRIS system is integrated with Typo3[10] a content management system that allows each member to enrich their profile with any desired content.

5. DSS – decision support system for simulation of virtual organization performance

The Decision Support System (DSS)[11] concentrates on supporting operational management of VO. The user of the DSS is the VO manager/coordinator. The DSS may be used simultaneously with the e-Cat system or independent of it – these two systems are complementary. The dominant lifecycle phase for using the DSS is the evolution phase, but it may significantly facilitate the decision making process during the creation as well as operation phases. The place for use of the DSS in the VO lifecycle is highlighted in Figure 7.

[10] http://typo3.org/

[11] The internal name of this activity was e-Dog (electronic Decision Optimization Guide) but it was not published as the official one. Therefore, there is no special name for this decision support system.

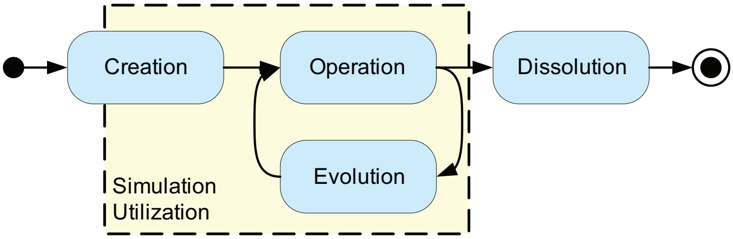


Fig. 7. Scope of the simulation in the VO life-cycle

DSS allows the VO manager to proactively detect and evaluate changes in the VO operations, and it provides suggestions for the adaptations. The simulation analysis supports this process as well as suggests answers to “what-if questions” before the change is applied or the explored situation has occurred (Miller at al., 2002). Such simulation also supports the VO manager in discovering potential bottlenecks and the possibilities of their removing.

The core of DSS is formed by a VO operations simulator based on agent-based rescheduling concept, originally developed for intra-enterprise scheduling in the ExPlanTech project. The original concept is modified so that each VO member is represented by a set of ExPlanTech Core “Workshop” Agents; each of the agents represents one production unit, also known as the competency cell (Neubert at al. 2001). The agent-based core of the system is supplemented by a group of tools for management of VO configurations, plans, and schedules, interfaces for online gathering of VO operational data from dedicated automated tools, tools for configurations of VO operation simulation as well as its (re)scheduling, and analytical tools for evaluation of the alterative VO configurations and schedules. The DSS architecture and prototype implementation are described in more detail in (Hodík & Stach, 2007).

5.1 Virtual organization management toolkit

DSS is a component of the Organization Management toolkit (VOM), which is a complex system for management of distributed workflows of the VOs. The VOM toolkit consists of several interconnected subsystems:

- **VOMod** (VO-Model) is a core maintaining the VO configuration, including the definition of its task and consortium.
- **SID** (Supporting Indicator Definition) is a tool for managing a configuration of performance indicators necessary to be monitored during the VO lifecycle.
- **DI3** (Distributed Indicator Information Integrator) is a technology for collection of VO operational data (according to the defined performance indicators) distributed among the VO members.
- **MAF** (Monitor and Finance) is a tool for analyzing the measurements of the current state of the VO as it is.
- **DSS** (Decision Support System) is a tool for the simulation of the possible VO performance according to the actual state of the VO configuration and the configuration given by the VO manager.

Figure 8 presents VOM components and their high-level dependencies (Hodík at al., 2007) (Negretto at al., 2008).

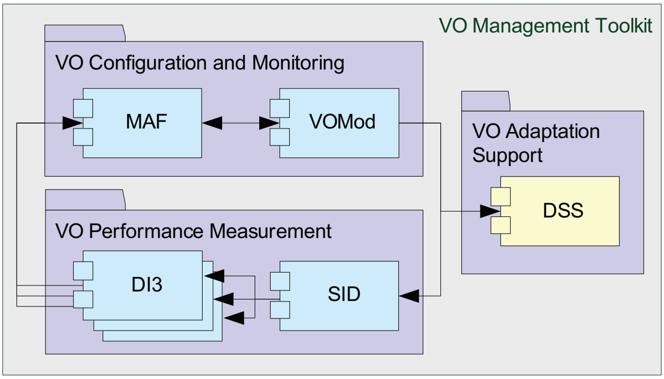


Fig. 8. Architecture of the Virtual Organization Management Toolkit

DSS may be run as a stand-alone application. If connected to the other VOM toolkit components, it utilizes operational data as they are collected and stored by it. Otherwise, such (as well as virtual) data has to by provided manually.

5.2 Simulation based what-if-analysis

The what-if-analysis provided by DSS is primarily based on current state of the VO, and provide an outlook on possible future progress. For the decision making process, the what-if-analysis is an important tool to prove the hypothesis and intended actions before putting them into operation. The simulation runs according the VO state and allows several configurations of the constraints to the simulated future. The three main steps of the what-if-analysis are:

1. Simulation Configuration
2. Simulated VO Performance (the simulation itself)
3. Simulation Evaluation

During the simulated VO performance, the simulated runs are influenced by defined constraints, and the rescheduling invoked by them is executed as many times as configured. Finally, the results are collected and evaluated. For the activity flow of the simulation based what-if-analysis see Figure 9. The final decision about modifications of the VO schedule and configuration adaptations are up to the VO manager and her negotiation with the other VO members.

The simulation configuration may be divided into two partially independent steps: (i) configuration of the VO state for the beginning of the simulation, and (ii) configuration of the “ifs”, i.e. the simulated future. The state of the VO is given by the online operational data collecting tools, or it is defined manually by the VO manager. The simulation future configuration includes definitions and probabilities of the performances deviations (positives as well as negatives), which are represented by (global) *events* and *behavioral models*.

The events do not relate to any particular VO member; they represent situations influencing defined task that occurs out of control of the responsible member. The behavioral models describe the expected will and ability (described by a set of Beta distributions) of the VO members to keep the scheduled dates if participating on any task.

The events and behavioral models may postpone the start of the processing and/or lengthen or shorten it, or they may make the task impossible to be accomplished by the currently

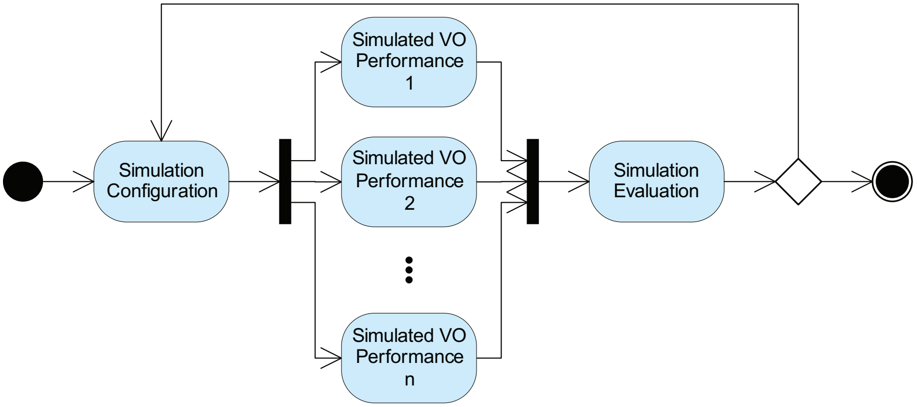

Simulated VO Performance 1
Simulation Configuration
Simulated VO Performance 2
Simulation Evaluation
Simulated VO Performance n

Fig. 9. Activity diagram of the what-if-analysis

responsible member. In such a case, the processing of the task must be negotiated with an alternative member. The events as well as behavioral models are applied to the schedule according to the simulated run of the VO. When the completion time of the task is influenced, the rest of the schedule may be (according to the configuration of the simulation) rescheduled by the right-shift rescheduling or by total rescheduling of the tasks with processing not started yet.

After finishing all simulation runs, the results are collected and mutually compared and evaluated according to the original VO schedule and configuration. The simulation evaluation contains three components: (i) graphic overview of the original schedule together with simulated performances, (ii) various schedule details (e.g. critical path, makespan, and working load) evaluation, and (iii) set of suggestions provided by the simple built-in rule-based system. Examples of the GUI presenting the simulation analysis are in Figure 10.

5.3 DSS verification

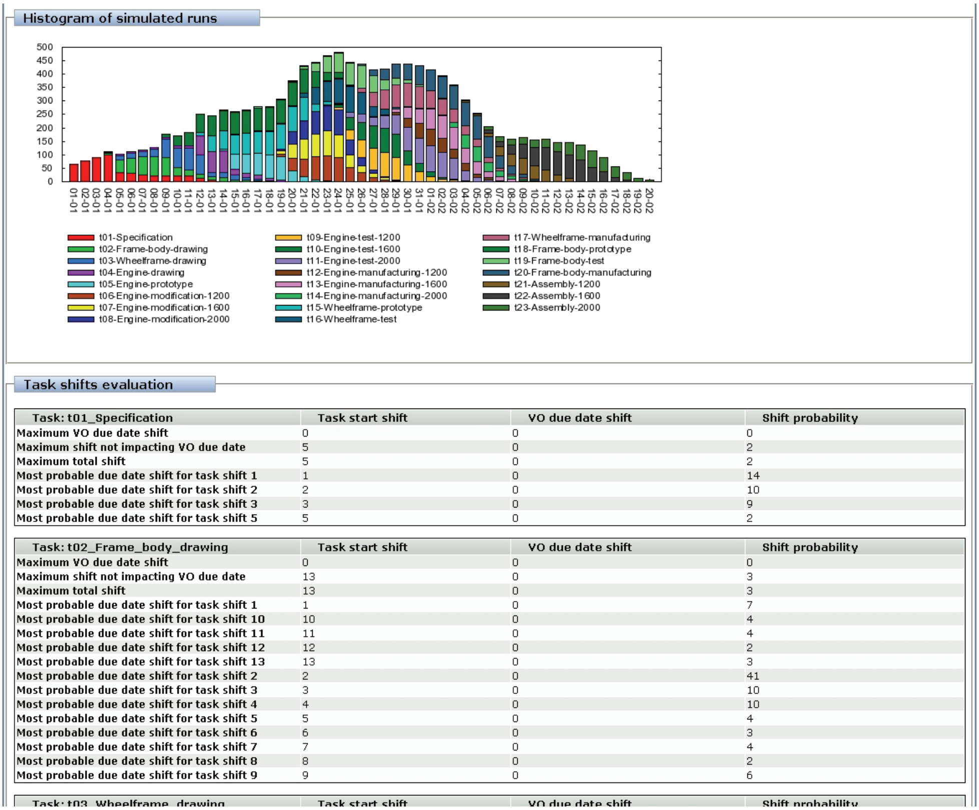
The DSS was designed, implemented and verified as a component of the VOM toolkit (all components in form of research prototype) within the ECOLEAD[12] project. The pilot prototype was successfully used in real business environment by Virtuelle Fabrik[13].

The key features and innovations of DSS are:

- Autonomous agents based modeling (new quality in VO simulation)
  - Each VO member is modeled separately to consider the individuality of the VO members, their behaviors and capabilities.
- Modeling with uncertainty
  - Member’s behavioral models based on past performances and experience.
  - Generation of random performance variations allows proving the VO configuration robustness. Statistical simulation based on the Monte-Carlo method
- Integrated prototype (component of the VOM toolkit)
  - Integration allows providing simulation based on the latest VO operational data.

[12] http://ecolead.vtt.fi/

[13] http://www.virtuelle-fabrik.com



Task: t01_Specification	Task start shift	VO due date shift	Shift probability
Maximum VO due date shift	0	0	0
Maximum shift not impacting VO due date	5	0	2
Maximum total shift	5	0	2
Most probable due date shift for task shift 1	1	0	14
Most probable due date shift for task shift 2	2	0	10
Most probable due date shift for task shift 3	3	0	9
Most probable due date shift for task shift 5	5	0	2

Task: t02_Frame_body_drawing	Task start shift	VO due date shift	Shift probability
Maximum VO due date shift	0	0	0
Maximum shift not impacting VO due date	13	0	3
Maximum total shift	13	0	3
Most probable due date shift for task shift 1	1	0	7
Most probable due date shift for task shift 10	10	0	4
Most probable due date shift for task shift 11	11	0	4
Most probable due date shift for task shift 12	12	0	2
Most probable due date shift for task shift 13	13	0	3
Most probable due date shift for task shift 2	2	0	41
Most probable due date shift for task shift 3	3	0	10
Most probable due date shift for task shift 4	4	0	10
Most probable due date shift for task shift 5	5	0	4
Most probable due date shift for task shift 6	6	0	3
Most probable due date shift for task shift 7	7	0	4
Most probable due date shift for task shift 8	8	0	2
Most probable due date shift for task shift 9	9	0	6

Task: t03_Wheelframe_drawing	Task start shift	VO due date shift	Shift probability

Fig. 10. of user interface for VO simulation evaluation

- Software-as-service implementation (on-demand service)

Users are provided with on-line access to the tools that they could not afford or maintain individually.

6. Conclusion

The three systems presented were created in different times, demonstrating the development of new trends in the decision support systems domain. In the beginning, the systems were focused on aid intra-enterprise decision support and relayed on internal data only.

Later, the cooperation with trusted partners was incorporated. This cooperation required the installation of dedicated software, establishing secure connection between partner's IT systems and, usually, modifications in their business process. These difficulties resulted in a very limited number of systems that were successfully deployed among SMEs.

The contemporary situation is much more favorable for the successful deployment of similar systems. First of all, the systems themselves have evolved – they are more robust, more elaborate and they exploit the research effort of the past years to provide decision support in multiple useful scenarios and to extend their interoperability.

Moreover, the business environment has also changed. SMEs are more aware of the idea of VO and they understand the advantages of cooperation in networked environment. Users are also more accustomed to electronic cooperation and distributed tools. Ideas like software as a service or tools like Google Docs have become part of common business life and largely support the trust in electronic tools.

Future decision support systems will be designed to naturally involve the partners of the company in the decision process. If these partners are part of an alliance, a networked organization or a virtual organization, their data or intentions can be partially obtained using electronic tools and used as an input for decision making. The behavior and intentions of partners, those who are not collaborative enough, can be simulated and used in the decision making as well.

Contemporary technology, which combines centralized and distributed elements, can implement such a system with satisfactory reliability while preserving the partner's independence.

7. Acknowledgements

The ExtraPlanT system has been developed with the support of the EUTIST-AMI IST-2000-28221 project.

Development of voice interface to ExtraPlanT system has been supported by the Ministry of Education of the Czech Republic under Project 1M0567.

Systems e-Cat and VOM (including DSS systems) has been developed with the support of the ECOLEAD project FP6 IP 506958.

The research described is part-funded by the EC FP6 project I*PROMS Network of Excellence and by the Ministry of Education, Youth and Sports of the Czech Republic grant No. MSM 6840770038. The opinions expressed herein are those of the named authors only and should not be taken as necessarily representative of the opinion of the European Commission.

8. References

Adesta, E. Y. T. (2005). A Strategic Planning Procedure to Support Progress Towards Extended Enterprise. *Proc. of the Eleventh Int. Conf. on Information and Computation Economies ICE-2005* (pp. 371-377). Centre for Concurrent Enterprising.

Balentine, B.; & Morgan, D. P. (1999). *How to Build a Speech Recognition Application*, Enterprise Integration Group, San Ramon, California, USA

Bečvář, P.; Pěchouček, M. & Šmídl, L. (2004). Telephony interface of the ExtraPlanT multi-agent production planning system. *Proc. Berliner XML Tage 2004 - Dialogsysteme mit XML-Technologien*, Berlin, Germany

Bečvář, P.; Šmídl, L.; Psutka, J. & Pěchouček, M. (2007). An Intelligent Telephony Interface of Multiagent Decision Support Systems. *IEEE Trans. Syst., Man, Cybern. C,* Vol. 37, No. 4, (pp. 553–560)

Biesalski, E. (2003). Knowledge management and e-human resource management. *FGWM 2003*, Karlsruhe.

Boughzala, I. & Zacklad, M. (1999). Cooperation engineering for the extended enterprise, *Proceedings of the Human Centered Processes Conference* (pp. 119–127), Brest, France.

Camarinha-Matos, L. M. & Afsarmanesh, H. (1998). Virtual Enterprises: Life cycle supporting tools and technologies, *Handbook of Life Cycle Engineering: Concepts, Tools and Techniques* (pp. 535-571). Chapman and Hall.

Friese, T.; Freisleben, B.; Rusitschka, S. & Southall, A. (2002). A Framework for Resource Management in Peer-to-Peer Networks, *Proceedings of NetObjectDays 2002* (pp. 4-21), Springer.

Hodík, J.; Bečvář, P.; Pěchouček, M.; Vokřínek, J. & Pospíšil, J. (2005) ExPlanTech and ExtraPlanT: multi-agent technology for production planning, simulation and extra-enterprise collaboration. *International Journal of Computer Systems Science & Engineering,* Vol. 20, No. 5 (pp. 357– 367).

Hodík, J. & Stach J. (2007). Virtual organization simulation for operational management. *Innovative Production Machines and Systems: Third I*PROMS Virtual International Conference*, 2-13 July, 2007. Whittles Publishing, 2007.

Hodík, J.; Mulder, W.; Pondrelli, L.; Westphal, I. & Hofman, R. (2007). ICT services supporting virtual organization management. *Innovative Production Machines and Systems: Third I*PROMS Virtual International Conference*, 2-13 July, 2007. Whittles Publishing, 2007.

Hodík, J.; Vokřínek, J. & Bečvář, P. (2009). Support for Virtual Organisation Creation - Partners' Profiles and Competency Management. *International Journal of Agent-Oriented Software Engineering*. Vol. 3, (pp. 230-251). ISSN 1746-1375.

Mařík, V.; Pěchouček, M.; Štěpánková, O. & Lažanský, J. (2000). ProPlanT: Multi-Agent System for Production Planning. *Applied Artificial Intelligence,* Vol. 14, No. 7

Matoušek, J.; Romportl, J.; Tihelka D. & Tychtl Z. (2004). Recent Improvements on ARTIC: Czech Text-to-Speech System, *Proceedings of INTERSPEECH 2004 - ICSLP, 8th International Conference on Spoken Language,* Vol. III, pp. 1933–1936, Jeju, Korea

Miller, J. A.; Cardoso, J. & Silver, G. (2002). Using simulation to facilitate effective workflow adaptation. *SS '02: Proceedings of the 35th Annual Simulation Symposium,* Washington, DC, USA. IEEE Computer Society.

Müller, J.; Psutka, J. & Šmídl L. (2000). Design of Speech Recognition Engine, In: *Lecture Notes In Computer Science; Vol. 1902; TDS '00: Proceedings of the Third International Workshop on Text, Speech and Dialogue*, (pp. 259–264), Springer-Verlag, London, UK

Negretto, U.; Hodík, J.; Kral, L.; Mulder, W.; Ollus, M.; Podrelli, L. & Westphal, I. (2008). VO Management Solutions. *Methods and Tools for Collaborative Networked Organizations*. Heidelberg: Springer.

Neubert, R.; Langer, O.; Görlitz, O. & Benn, W. (2001). Virtual enterprises-challenges from a database perspective. *ITVE '01: Proceedings of the IEEE workshop on Information technology for Virtual Enterprises,* (pp. 98-106), Washington, DC, USA. IEEE Computer Society.

Pěchouček, M.; Mařík, V. & Bárta, J. (2002a). A knowledge-based approach to coalition formation. *IEEE Intelligent Systems,* , Vol. 17, No. 3, (pp. 17–25).

Pěchouček, M.; Říha, A; Vokřínek, J.; Mařík, V. & Pražma, Ř. (2002b). ExPlanTech: applying multi-agent systems in production planning. *International Journal of Production Research,* Vol. 40, No. 15, (pp. 3681–3692)

Pěchouček, M.; Vokřínek, J. & Bečvář, P. (2005). ExPlanTech: Multiagent Support for Manufacturing Decision Making. *IEEE Intell. Syst.* Vol. 20, No. 1, (pp. 67–74)

Říha, A.; Pěchouček, M.; Vokřínek, J. & Mařík, V. (2002). From intra-enterprise towards extra-enterprise production planning, *Knowledge and Technology Integration in Production and Services* (pp. 349–356). New York, Kluwer Academic / Plenum Publishers.

Shen, W.; Lang, S. & Wang, L. (2005). iShopFloor: an Internet-enabled agent-based intelligent shop floor. *IEEE Trans. Syst., Man, Cybern. C,* Vol. 35, No. 3, (pp. 371–381)

Šmídl, L.; Müller, L. & Psutka, J (2002). VoiceXML Based Telephone Dialog System Providing Access to Entrance Examination Results Stored in the University Database, *Proc. SCI 2002*, Orlando, July 2002

15

Design and Implementation of a Decision Support System for Analysing Ranking Auction Markets for Internet Search Services

Juan Aparicio, Erika Sanchez, Joaquin Sanchez-Soriano and Julia Sancho
Center of Operations Research (CIO). University Miguel Hernandez of Elche
Spain

1. Introduction

Nowadays, Internet is the usual platform for people around the world to search for firms offering specific services. However, many Internet search engines provide useless lists due to the fact that they are extremely long or not very well organized. This has been the starting point for some Internet search service providers to create new systems for ranking firms according to different searching engines.

One outstanding example of these providers is the giant Google. Google has developed an auction mechanism (see, for example, (Krishna, 2002) or (Klemperer, 2004) for details on auction mechanisms) for firms to advertise their services on the Internet, known as Google Adwords system. Under this mechanism, when a consumer searches for firms offering specific services, the results for a particular keyword (or group of keywords) are ranked in descending order according to what previously the firms have bid. Then, when a consumer clicks on the name of the firm listed on the search site, this firm has to pay the provider an amount equal to the bid price regardless of whether the consumer finally purchases or not. This way of ranking firms has several benefits over other possibilities. On one hand, the provider offers pay-for performance service since firms pay only when a consumer clicks on their corresponding hyperlink. On the other hand, each firm is encouraged submitting a new bid anytime to change the order at which it appears on the list.

This issue has been studied previously in the literature. (Lim & Tang, 2006) introduced a one-stage game for two firms that captures the advertising mechanism of a search service provider. So, game theory allowed them to analyze the firm's optimal bidding strategy and assess the impact of several parameters on the provider's revenue. Nevertheless, it just was the first attempt to analyse bidding behaviour arising from this type of situations, since their model presents several limitations. First, Lim and Tang's model is limited to only two firms, each one with just three feasible bids. Secondly, Lim and Tang's model does not take into account the dynamic interactions among firms (e.g., fluctuating coalition structures) since they assume just one stage. For these reasons, it seems suitable to extend the analysis of (Lim & Tang, 2006) to other more complex situations.

In this chapter we describe software which could be used as a decision support system tool or framework for analysing, at least from an academic point of view, ranking auction markets for Internet search service providers. The software tool is based on the behaviour of the firms in a realistic market, thus many different parameters are considered. Taking into account that the problem is really complex from a mathematical point of view, the results are obtained by simulation. This kind of approach using computational tools to analyse a problem is very often in engineering problems because of their mathematical complexity and it has been also used to analyse economic problems as markets based on auction mechanisms. For instance, (Sancho et al., 2008) provide a simulation framework to analyse competitive electricity markets, (Atkins et al., 2007) provide an agent based computational framework to study large commodity markets or (Mehlenbacher, 2009) studies signal averaging in English auctions using a multi-agent system. Additionally, some computational experience is reported to illustrate what kind of results we could obtain.

Finally, we would like to point out that other related papers are (Feng et al., 2007), who focused their analysis on how to improve the seller's expected revenue by enforcing a reserve price in ranked items auctions, and (Sancho et al., 2009), who deal with auction situations arising from Internet search service providers but considering a cooperative approach.

The rest of the chapter is organised as follows. In Section 2 we provide a description of the Internet ranking auction situation and introduce the main parameters involved in the problem. In Section 3 we introduce software tool, describe its main elements and how it works. Furthermore we include some computational experience. Finally, Section 4 concludes.

2. Brief description of the Internet ranking auction

In this section we formally introduce the Internet ranking auction situation and the parameters we use in the developed software tool for analysing such situations. Our approach involves analysing the problem from a competitive point of view, i.e., we are considering that the firms will compete to obtain a better position on the list because that is profitable for them. The position of a firm on the list will depend on the money each firm agrees to pay per click and, hence, the strategies of the firms would be their possible bids and their goals being focused on maximizing their expected profit.

In particular, we consider a multi-stage situation in which an arbitrary number, n, of firms, each owns a homepage, are planning to list their names (links) under the same group of keywords in order to obtain as many visits as possible. Indeed, each visitor is a potential client to buy their products or to contract their professional services. To this end, they resort to an Internet search service provider, as could be Google or Yahoo. As the firms are interested in being on the top of the list they should pay some amount of money to the Internet search service provider in order to avoid the usual ordering provided by the searching engine used by the Internet search service provider. Furthermore the Internet search service provider could vary the order of the firms on the list from one period to another according to the paid money by them. In this sense, we are considering different periods and thereby the problem is dynamic.

We denote by T the total number of periods, t=0,…,T-1. And, we denote by N_t the total number of customers who use the aforementioned group of keywords to conduct a search in period t (day, hour, minute, etc.). Additionally, we denote by N_{dt} the total number of disloyal customers in period t, i.e., those who do not have a clear preference among all the firms on the list and therefore they can click any homepage link. Consequently, $N_t - N_{dt}$

will be the number of loyal customers in period t. We assume that loyal customers always visit only the site of their preferred firm. In this context, l_i will represent the market share of firm i over the set of all loyal clients interested on that particular group of keywords. In other words, l_i is the proportion of visits that firm i receives from all loyal customers (it is obvious that $\sum_{i=1}^{n} l_i = 1$).

On the other hand, p_j denotes the proportion of clicks from the disloyal customers that a firm in position j on the list will receive. In this way, if firm i is ranked in position j in period t then it will receive a total number of clicks in that period equals to the sum of the clicks received from its loyal customers and the clicks received from the disloyal customers, in formula

$$c_{it} = l_i\left(N_t - N_{dt}\right) + p_j N_{dt} . \tag{1}$$

The unitary reward per customer of firm i in period t, when a customer clicks on the link to enter in its homepage i, is denoted here by θ_{it}. In order to obtain their position in period t, the firms have to make a bid. These bids are the amount of money that firms agree to pay for each click received. Finally, they achieve the position in the ranking corresponding to their bids taking into account that all submitted bids are arranged in decreasing order. Therefore, firm i must only make a single bid and their final profit will be given by

$$\left(\theta_{it} - b_{it}\right)\left(l_i\left(N_t - N_{dt}\right) + p_j N_{dt}\right) . \tag{2}$$

Therefore, we assume that firms pay for all clicks from both loyal and disloyal customers.
Whereas the revenue for the provider is given by

$$\sum_{i=1}^{n} b_{it}\left(l_i\left(N_t - N_{dt}\right) + p_j N_{dt}\right) . \tag{3}$$

Other input which the Decision Support System uses is the average expense that any customer spends for each click to the firms' homepages. It is denoted here as e. However a customer not always spends that money when she enters in the homepage of a firm, therefore there is uncertainty about the expense happens or not and so we will denote by π the probability of such expense happens.

Since firm i does not have information on the proportion of clicks that a firm receives from the disloyal customers, we assume that each firm i has a private forecast, f_{ijt}, about p_j in period t. In practice, all these estimations can be obtained by each firm using whatever market information and statistical tool at its disposal. Additionally, we assume that the private forecast of each firm can be updated over time with the new information obtained from the previous periods. For this reason, we use the subscript t to highlight that f_{ijt} is the firm i estimation of p_j obtained with the information available for firm i until period t.

Finally, after receiving all the bids in period t, the Internet service provider announces the ranking and the bid of each firm for this period and so on. Since the bids are revealed in period t, each firm has incentives to submit a new bid for period t+1 in order to try to change or keep its position in the ranking, i.e., the order at which the firm appears on the list.

3. Fundamentals and development of the Decision Support System for Internet ranking auctions

Based on the above description of the Internet ranking auction situation, we have implemented a software tool, using C++ Builder 6, which we have tried to reflect the reality of the firms' bids and carry out the ranking of these bids, considering also the dynamic component of the situation.

The uncertainty about the number of clicks for each position has been modelled through the private forecasts f_{ijt}. And both the forecasts and other parameters will be updated over time by means of specific algorithms that we will explain below.

On the other hand, we have resorted to simulation for analyzing the firms' behaviour. The reason is not hard to see. Considering more than two firms and three strategies leads to a mathematically intractable scenario and therefore the simulation approach seems suitable and reasonable to deal with.

3.1 General outline of the application

The developed and implemented Decision Support System works taking into account the risk profiles defined for each firm which participes in the ranking auction, the total number of customers (loyal and disloyal), the average expense per click, the proportion of clicks per position, the average reward for each firm, the private forecasts and the number of periods to be simulated. Once all necessary inputs to start the simulation have been introduced, which constitute the inicial working conditions, the implementation cycle is the following.

(a) *Storing inputs*. The system stores information about firms; customers; expenses; number of simulations; number of periods to be simulated; and bids.

(b) *Computation of the variables of interest*. When the parameters which are necessary to start the simulation have been introduced into the system, the number of total customers is obtained through simulation. Then, for each customer, the system determines whether she is loyal or disloyal. If she is loyal, the system determines to which firm. In any case, by simulation, the system obtains the values of all variables necessary to run on the simulation. If we are in the first period (t=0), then the system simulates all the bids with the initial information introduced into the system and ranks the firms. All this information is then saved. If we are in a general intermediate period t, before simulating the firms' bids, the system updates θ_{it} and f_{ijt} for all firms using the information obtained from previous periods and then it simulates the firms' bids and ranks them.

(c) *Presentation of results*. Once the system has simulated the target period, all obtained data are reported in a practical and friendly format such as spreadsheets, which allow us to store on the hard disk the results of all simulations carried out. The results are sorted in different spreadsheets showing with tables and figures the following information: b_{it}, θ_{it} , e, f_{ijt} and c_{it} for each period t, t=0,…,T-1, which we consider relevant to analyse a particular Internet ranking auction situation. This way in which the simulation results are presented eases to analyse them using the different mathematical and statistical utilities that the most of spreadsheets usually have.

Figure 1 shows the flow chart which represents and summarises the operating model of the application we are describing. We can observe that the general structure of the application is very simple consisting basically of two consecutive cycles, one for the customers' behaviour and another for the firms' bids.

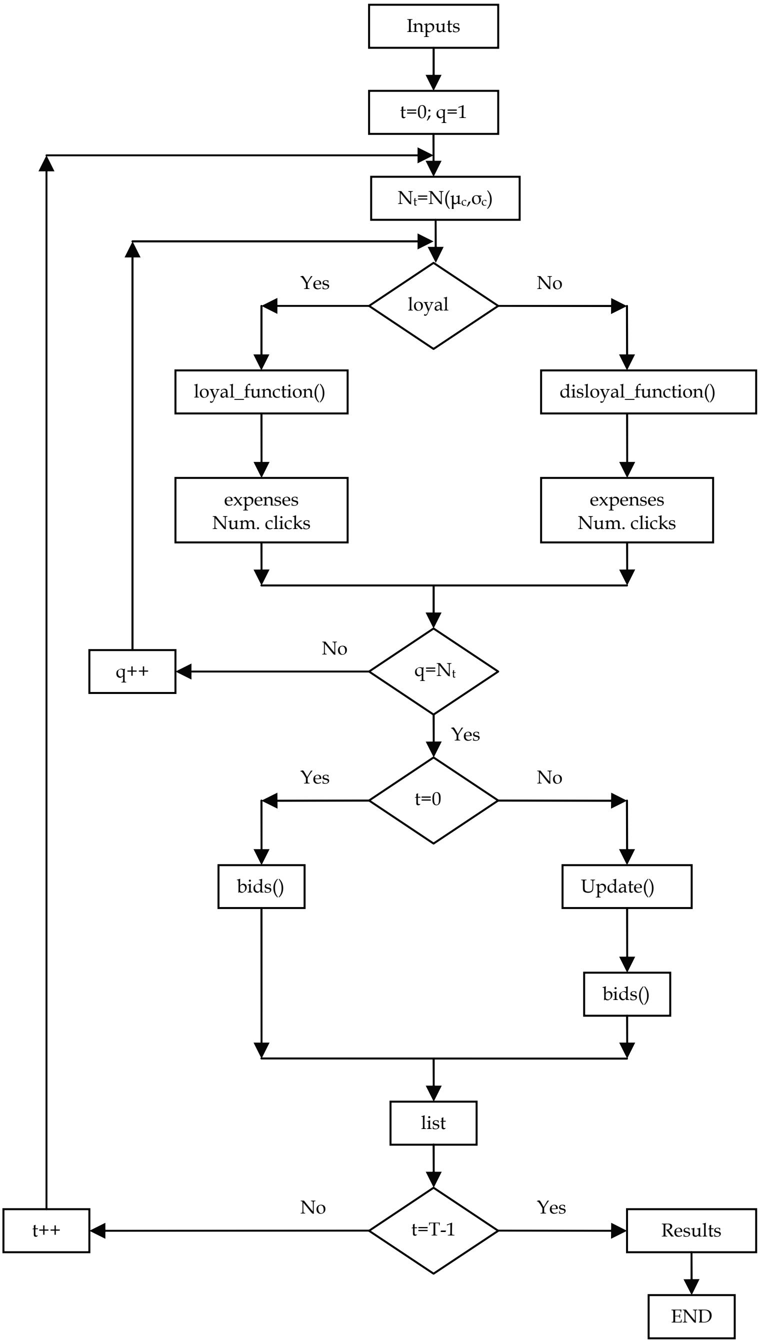
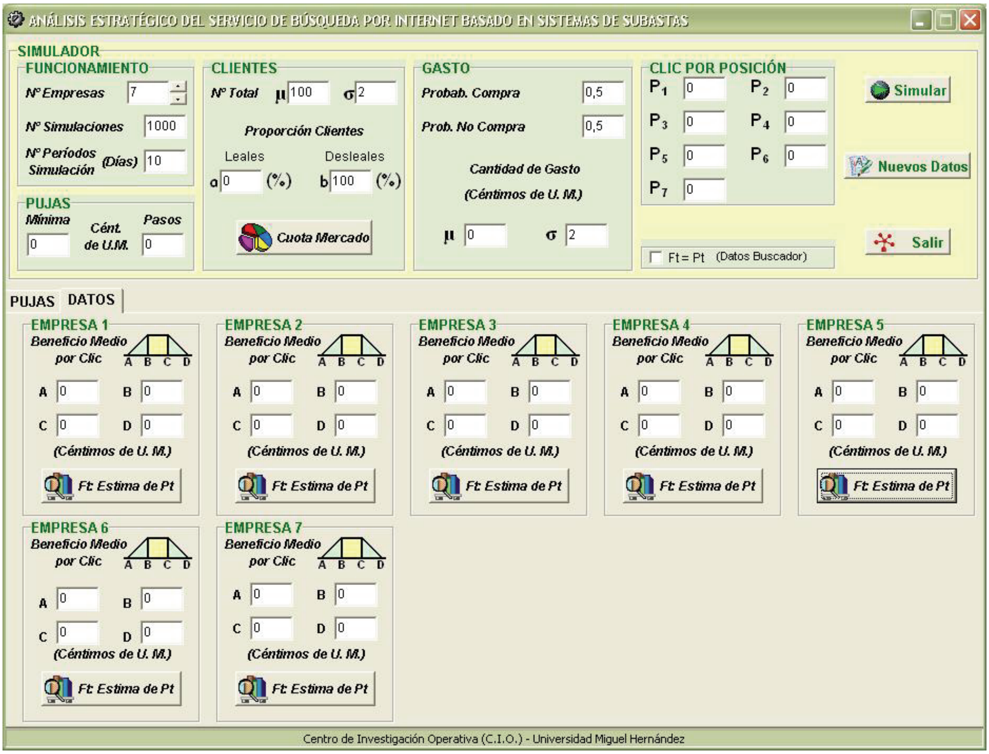


Fig. 1. The flow chart which represents the operating model of the application

We should also describe the main aspects of the software developed. One advantage of this system is the clarity which the different parameters are dealt with, and also its easy use. The software consists of a graphic and intuitive interface through which the user can introduce all the parameters necessary to carry out the simulation of an Internet ranking auction. This interface consists of only one window (see Figure 2) in which the main parts of the application with very different characteristics are shown. In the following subsections we will show and explain in detail each of these parts. Finally, another advantage of this software tool is the presentation of the results in spreadsheets, because, as we said before, it eases the posterior analysis of the obtained results from the simulation.



ANÁLISIS ESTRATÉGICO DEL SERVICIO DE BÚSQUEDA POR INTERNET BASADO EN SISTEMAS DE SUBASTAS

SIMULADOR

FUNCIONAMIENTO
Nº Empresas 7
Nº Simulaciones 1000
Nº Períodos Simulación (Días) 10

PUJAS
Mínima 0 Cént. de U.M. Pasos 0

CLIENTES
Nº Total μ 100 σ 2
Proporción Clientes
Leales a 0 (%) Desleales b 100 (%)

GASTO
Probab. Compra 0,5
Prob. No Compra 0,5
Cantidad de Gasto (Céntimos de U. M.)
μ 0 σ 2

CLIC POR POSICIÓN
P_1 0 P_2 0
P_3 0 P_4 0
P_5 0 P_6 0
P_7 0

☐ Ft = Pt (Datos Buscador)

Beneficio Medio por Clic (Céntimos de U. M.)

	A	B	C	D
EMPRESA 1	0	0	0	0
EMPRESA 2	0	0	0	0
EMPRESA 3	0	0	0	0
EMPRESA 4	0	0	0	0
EMPRESA 5	0	0	0	0
EMPRESA 6	0	0	0	0
EMPRESA 7	0	0	0	0

Centro de Investigación Operativa (C.I.O.) - Universidad Miguel Hernández

Fig. 2. The main application window. [Analisis estrategico del servicio de busqueda por Internet basado en sistemas de subastas (Strategic analysis of the Internet search service based on auction systems), funcionamiento (starting conditions), empresas (firms), pujas (bids), leales (loyals), desleales (disloyals), gasto (expense), probab. compra (purchase probability), cuota mercado (market share), nuevos datos (new data), salir (exit), datos buscador (search service provider data), beneficio medio por clic (average profit per click), estima de (estimation of)]

3.2 Parts of the Decision Support System

3.2.1 System variables

This part allows us to configure the internal operations of the Decision Support System. First, we have to enter as an input the number of firms (the limit depends on the features of the computer in which we are running the software). Other necessary inputs are the number of simulations and the number of periods to simulate. The first input is related to the times in which the system repeats the calculations to determine suitable values for different parameters. Specifically, the system tries to estimate the probability of being in any position of the ranking conditioned to a particular submitted bid. Nevertheless, this process will be shown in detail later on.

3.2.2 Bids

In this part we need to introduce two parameters. The first one is related to the reserve price, which the provider can impose in the ranking auction. In many instances, providers (sellers, in general) reserve the right to not provide the service if the price determined in the auction is lower than some threshold. This threshold amount is called "the reserve price". The system takes into account this possibility. Obviously, if such reserve price does not exist, then we must only introduce a value zero for this parameter. The second parameter is the amount that must be added to the reserve price to build the set of feasible bids. Let us denote this amount as Δ. Therefore the minimum feasible bid will be the reserve price, r, the maximum feasible bid will be the unitary reward, θ_{it}, and the possible bids in between will be calculated as r+kΔ, k=1,2,…,K where K is an integer number such that $K \le (\theta_{it} - r)/\Delta$.

3.2.3 Customers

We must also enter as an input the total number of customers who use the Internet service each period t. To this end, we introduce into the system this parameter modeled by a Gaussian distribution of mean μ_c and standard deviation σ_c, therefore we are considering the possibility that the number of customers is not constant along the number of periods under consideration. For each period t, the system simulates an execution of a Gaussian distribution $N(\mu_c, \sigma_c)$ and, afterwards, rounds off this value to obtain an integer number of customers.

Additionally, it is necessary to introduce into the system the percentage of loyal and disloyal customers in the (simulated) market. We assume that this information is shared for all the participating firms in the auction.

Finally, we have a button to add information about the firms' market shares, l_i. All these inputs will be used for the system, by simulation, to determine the total number of customers, the disloyal clients and the number of loyal clients for each firm.

3.2.4 Expenses

The main aim of this part is to provide estimations of the customers' expenses, once they have clicked on a particular homepage. In order to obtain this information, we have to enter as an input the probability of a customer purchasing the product or contracting the service from any firm on the list. If, finally, a customer purchases the product, then it is necessary to know how much she spends. For this reason, we introduce into the system the amount of expense per client modeled again by a Gaussian distribution $N(\mu_e, \sigma_e)$. All these parameters will be used later in the simulation stage.

3.2.5 Clicks per position on the list

It is necessary to introduce into the system information on the proportion of clicks that a firm receives from the disloyal clients when a firm is ranked in position j, p_j, for all j=1,…,n.

In this framework, we assume that these parameters depend solely on the ranking. In addition, we point out that p_j, for all j=1,…,n, is private information of the Internet search service provider. Consequently, the firms, through whatever market information gathering techniques at its disposal, need to have a private forecast, f_{ijt}, about the parameter p_j. These estimations help firms to make a decision about what bid to submit in each period t. Our system updates the private forecasts over time from the number of clicks obtained in the positions at which firms appear in each of the simulated periods.

The system also allows us to simulate that the Internet search service provider discloses information on p_j, for all j=1,…,n, to the firms so as to check whether this strategy encourages them to bid more aggressively or not. The question, in this case, is to verify whether reducing the uncertainty on the number of clicks per position on the list implies that firms bid more aggressively.

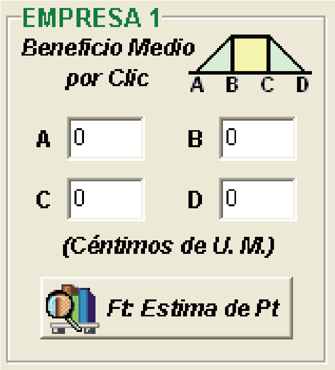
3.2.6 Information about firms

3.2.6.1 Data

As for the part of the framework devoted to the firms, we need to enter as an input the unitary reward per customer and the private forecast of p_j for each firm in period t=0, i.e., θ_{i0} and f_{ij0}, respectively (see Figure 3).

In our system θ_{i0} has been characterized as a trapezoidal fuzzy number (A,B,C,D). Therefore, it is necessary to introduce these four parameters. We use a trapezoidal fuzzy number to represent the (imperfect) knowledge of the firm about the unitary reward per customer.

In the first period, t=0, the system simulates a value for θ_{i0} from the fuzzy number. Later, the system will update the reward per customer over time, θ_{it}, t=1,…,T-1, using information about the customer's expenses and other variables involved in the problem. We should also point out that θ_{i0} could be defined as a crisp number. To this end, it would be enough to consider A=B=C=D.



A 0
B 0
C 0
D 0

Fig. 3. Part of the application devoted to the firm 1. [Empresa (firm), beneficio medio por clic (average profit per click), centimos de u.m. (cents of monetary unit), Estima de (estimation of)]

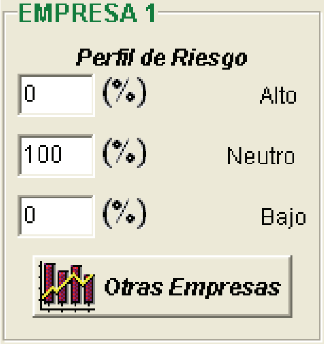
On the other hand, the system needs information about the estimation of the percentage of clicks per position for each firm. In other words, we have to introduce into the system the perceived percentage of disloyal clients who will visit the firm's homepage depending on the position at which it appears on the list, i.e., f_{ij0}. We note that this parameter is defined only for the first period to simulate, t=0, because after that initial period, the system automatically updates it over time for each firm obtaining the value of f_{ijt}, for all t=1,…,T-1. The strategies of the firms can vary from one period to another as a consequence of the updated perceived estimations of f_{ijt} since the firms will bid more aggressively to obtain positions with higher f_{ijt}, i.e., to obtain more visits and hence a higher expected profit.

3.2.6.2 Risk aversion and behaviour

Other data we have to enter as an input is the risk profile of firm i, i=1,…,n, and at the same time the risk profile of firm k (k=1,…,i-1,i+1,…,n) following the perception of firm i about it. Obviously firms' risk profile is directly related to the shape of the utility function of each firm. We denote here the utility function of firm i by UF_i. Following von Neumann-Morgenstern tradition (see (von Neumann & Morgenstern, 1944)), a firm i is risk averse if UF_i is a convex function, risk neutral if UF_i is a linear function and risk loving if UF_i is a concave function. In particular, in this software we assume that UF_i is a square root function for the first case, the identity function for the second case and, finally, a square function for the third case. Particularly, in our system we have modeled a more general situation. We allow firms to behave in a different way over time. In other words, in a period t firm i could behave as a risk averse, risk neutral or risk loving player depending on a probability distribution. We have to enter into the system as an input this distribution for each firm i, i=1,…,n (see Figure 4). In a similar manner, for running, the system needs the perception of firm i about the risk profile of each firm k, k=1,…,i-1,i+1,…,n (see Figure 5). So, the system in each period will simulate a value from the risk profile to determine the type of the utility function of each firm. Therefore we are considering that a firm to make a decision not only takes into account its risk profile but also its perception about the risk profiles of its competitors.

3.3 Main algorithms implemented

In this section we show the main algorithms that have been implemented with the task of calculating the number of loyal and disloyal customers and their expenses, the firms' bids, the ranking in each period, the updates values for the private forecasts of the clicks per position, the parameters of the fuzzy numbers, etc.



EMPRESA 1

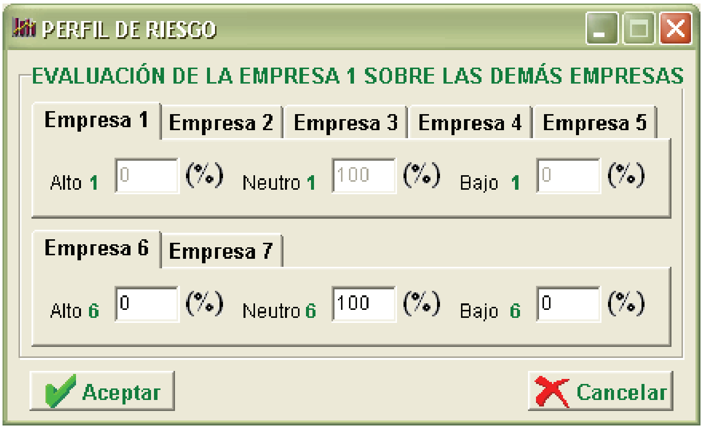
Perfil de Riesgo

0 (%) Alto

100 (%) Neutro

0 (%) Bajo

Fig. 4. Risk profile of the firm. [Empresa (firm), perfil de riesgo (risk profile), alto (high), neutron (neutral), bajo (low), otras empresas (other firms)]



PERFIL DE RIESGO

EVALUACIÓN DE LA EMPRESA 1 SOBRE LAS DEMÁS EMPRESAS

Alto 1 0 (%) Neutro 1 100 (%) Bajo 1 0 (%)

Alto 6 0 (%) Neutro 6 100 (%) Bajo 6 0 (%)

Fig. 5. Perception of firm 1 about the risk profile of the rest of competitors. [Perfil de riesgo (risk profile), empresa (firm), evaluacion de la empresa 1 sobre las demas empresas (perception of firm 1 about the risk profile of the rest of firms), alto (high), neutro (neutral), bajo (low), aceptar (OK), cancelar (cancel)]

3.3.1 Loyal and disloyal customers

Here we are going only to show the algorithm used for a disloyal client because the case of a loyal client requires an easier algorithm.

For any period t, once the system has simulated a value for the total number of customers who are going to use the Internet search service in that period, we need to know whether each customer is loyal or disloyal. To this end, we make use of the information previously introduced into the system (see Section 3.2.3) about the percentage of loyal and disloyal clients in the market. So, we simulate a random variable, which simply follows a Bernoulli distribution B(p) where p is the probability to be loyal, that determines whether a customer q is or not loyal. If q is, finally, a loyal customer then, using the firms' market shares, l_i, we can determine, by simulating an execution of the multinomial distribution $M(l_1, l_2, \ldots, l_n)$, which is her preferred firm among all the participants on the list.

Now, let us assume that a particular customer q is disloyal. Then, the system follows the algorithm described in Figure 6. In particular, for each position on the list j, j=1,..,n, the system simulates an execution of a uniform random variable $U_j[0,1]$, we call this execution by u_j. After that, if u_j is lower than p_j (the proportion of clicks associated to the position j) then the system understands that the disloyal customer q clicks on the homepage which appears in position j, otherwise the system considers that customer q does not click it. On the other hand, we assume that a disloyal customer could click on all the homepages on the list if she is willing to, unlike loyal customers which only click on their preferred firm's homepage. Therefore, the sum of all p_j, j=1,2,…,n, could greater than 1.

When u_j is lower than p_j for a position j, then the system simulates the amount of money that customer q spends in the firm which appears at position j. We note that both the probability of clicking, purchasing and how much to spend in the site do not depend on which firm is but the first probability depends on the position while the others two are always the same for all customers, firms and positions. Therefore, in this sense, the position on the list plays a crucial role in our approach because it makes the difference in the expected revenues of the firms. However, once a customer enters in a homepage her behaviour is not affected by the position, the firm or anything else. Consequently, the important question in this setting is whether a customer clicks or not the link to enter in a site.

Regarding the algorithm to simulate the expenses of customer q, let us suppose that $u_j<p_j$, i.e., customer q clicks on the homepage placed in position j. Then, the system simulates an execution of a uniform random variable over the interval [0,1]. If the obtained value is lower than the probability of purchasing (introduced previously into the system as an input), it means that the customer q will spend his money on the products of the firm which appears in position j. Afterwards, the system simulates the expenses by means of an execution of a Gaussian distribution $N(\mu_e,\sigma_e)$ (see Section 3.2.3).

Finally, the system saves for each customer q the positions on the list she visited and the expenses she spent in each visited position.

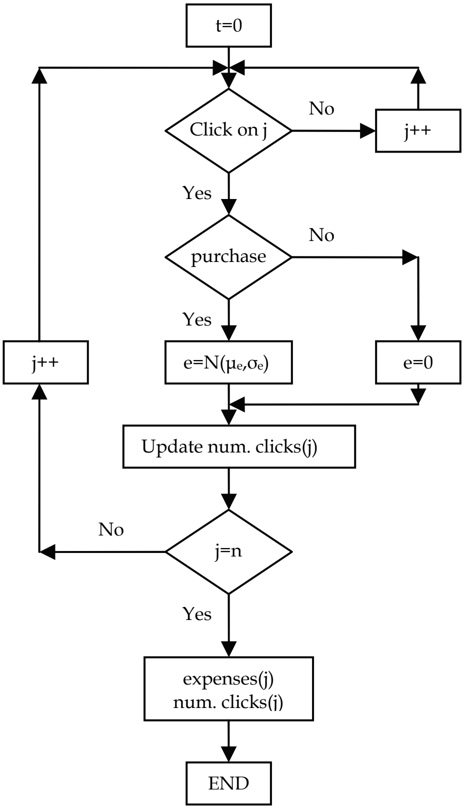


Fig. 6. The flow chart of the function for the disloyal customers

The algorithm for loyal customers only consists of the expenses part of the algorithm for disloyal customers for this reason it is omitted. Furthermore, we note loyal customers only spend money in their preferred firms' homepages while disloyal customers could spend money in several or all firms' homepages. On the other hand, it is nevertheless true that we could have considered that the probability of purchasing and/or spending change when a previous purchases has been done but the present approach is enough for our purposes and that extension or modification is left for further versions of this software tool. In fact, this modification would stress more the role of the position in this kind of situations.

3.3.2 Updating process

This process involves modifying the value of the variables that change over time. The firms' strategies can vary depending on the value of the simulated parameters in previous periods, since it is important for them to improve their estimations on the parameters used for them to make a bid. Therefore firms update their information available incorporating the data obtained from the previous periods in order to improve their knowledge about some system parameters relevant for them. In this section, we briefly show which parameters are updated by the firms and, additionally, how this process is carried out.

The parameters that we consider relevant for the firms from a strategic point of view, and hence they will be modified period by period, are the following: f_{ijt} , the estimation of the percentage of clicks per position, and θ_{it} , the unitary reward per customer who clicks on firm i's homepage.

Regarding estimations f_{ijt}, the Internet search service provider knows the real value of the percentage of clicks per position on the list, p_j, j=1,…,n. However, each firm at the end of period t only knows the number of received clicks on its homepage with absolute certainty. Therefore, firm i in position j at the end of period t can just update the estimation f_{ijt+1}. This new estimation can be calculated as the percentage of received clicks from the disloyal customers, i.e, we compute the ratio of the number of disloyal clients who click on firm i's homepage to the total number of disloyal clients. In order to know the number of clicks due to disloyal customers, we calculate the total number of clicks (from loyal and disloyal clients) minus the number of clicks from loyal customers. Firm i knows the total number of clicks received after playing period t, denoted by c_{it}, because each firm has a counter on its homepage. And the number of clicks from loyal customers to firm i is calculated by multiplying the total number of customers who visited the Internet search service in period t, N_t, times the proportion of loyal customers in the market and, finally, times the market share of firm i. It is worth to note that we consider N_t is common knowledge to all firms, because we assume that the search service provider publishes this information when the auction for period t is over. This assumption is not restrictive since it is important for the search service provider to advertise the number of customers using its search services in order to attract more firms over all when N_t is large enough.

As a consequence of the previous calculation, firm i has a first new estimation f_{ijt+1} on p_j, it will be denoted as $f_{ijt+1}(1)$. Nevertheless, we assume that each firm can improve the accuracy of the estimation $f_{ijt+1}(1)$. To do that, we will use the previous estimation f_{ijt} on p_j. First, we define a discrepancy index (DI) for measuring the difference between $f_{ijt+1}(1)$ and, the previous estimation, f_{ijt}:

$$DI= | f_{ijt} - f_{ijt+1}(1) | / f_{ijt+1}(1) \tag{4}$$

Depending on the value of DI, the system will weight each one of these estimations on p_j to build a new compound forecast. We only distinguish three cases:

Case 1: DI<=0.25

In this case, both f_{ijt} and $f_{ijt+1}(1)$ have relevant information about the real percentage of clicks for position j on the list. Therefore we consider that both estimations on p_j are equally credible. So, we update the estimation of f_{ijt+1} by the following expression.

$$f_{ijt+1}=(f_{ijt} + f_{ijt+1}(1))/2 \tag{5}$$

Case 2: 0.25<DI<=0.5
In this case, f_{ijt} and $f_{ijt+1}(1)$ are a little different. In this case we consider more credible estimation $f_{ijt+1}(1)$ than estimation f_{ijt}. Therefore, we use a weight of 2/3 for $f_{ijt+1}(1)$ and a weight of 1/3 for f_{ijt} to capture this feeling on the estimations. In this way, we have to calculate:

$$f_{ijt+1}=(2f_{ijt} + f_{ijt+1}(1))/3 \tag{6}$$

Case 3: DI>0.5
This is the more extreme case. Here f_{ijt} and $f_{ijt+1}(1)$ are clearly different. In this case, f_{ijt} is very far to the estimation obtained with the data of period t, i.e., $f_{ijt+1}(1)$. Therefore we consider that estimation $f_{ijt+1}(1)$ is much more credible than estimation f_{ijt}. Hence, we update firm i's estimation on p_j exclusively with the value of $f_{ijt+1}(1)$.

$$f_{ijt+1}=f_{ijt+1}(1) \tag{7}$$

Overall, the updating process is carried out to update the firms' private forecasts of the percentage of clicks per position. Given the available information that firm i has at the end of period t, we exclusively update the estimation of f_{ijt+1}, where j is the position at which the firm i appears on the list during that period. It means that we will use the previous forecasts f_{ikt}, where k=1,…,j-1,j+1,…n, to estimate f_{ikt+1} because firm i does not have new information about those positions. Therefore, each firm is able to update the information about only one position at the end of each period. On the other hand, it is not difficult to modify the software tool in order to consider another procedure or additional cases to update estimations f_{ijt}.

It is worth to note that the system allows us to simulate a situation where the search service provider publishes the real parameters, i.e., p_j, j=1,…,n. Due to that option, we can study the behaviour of the firms (regarding their bids) when they have more and better information about the relevance of the different positions for their interests.

On the other hand, as we said above, other parameters that will be modified over time are θ_{it}'s, i.e., the unitary reward per customer who clicks on firms' homepages. Remember that θ_{it} is modeled in this system as a trapezoidal fuzzy number (A_{it}, B_{it}, C_{it}, D_{it}). Regarding this, in order to update θ_{it} period by period, the system will revise separately each one of the four parameters A_{it}, B_{it}, C_{it} and D_{it}. Obviously, this updating process depends on the money spent by the customers in the firms' homepages. We know that the gross revenue obtained by the firm i at the end of period t will coincide with the loyal and disloyal customers' expenses. Then, in order to obtain the reward per click, we calculate the ratio of the gross revenue to the total number of clicks. We will denote this value by $\hat{\theta}_{it}$. The system uses $\hat{\theta}_{it}$ as a tool to update the four parameters defining the trapezoidal fuzzy number. Specifically, the system uses the following expressions:

$$A_{it+1}=0.9A_{it}+0.1\,\hat{\theta}_{it} \tag{8}$$

$$B_{it+1}=0.9B_{it}+0.1\,\hat{\theta}_{it} \tag{9}$$

$$C_{it+1}=0.9C_{it}+0.1\,\hat{\theta}_{it} \tag{10}$$

$$D_{it+1}=0.9D_{it}+0.1\,\hat{\theta}_{it} \tag{11}$$

Therefore, the reward per click corresponding to firm i at the end of period t will be modeled as a trapezoidal fuzzy number (A_{it+1}, B_{it+1}, C_{it+1}, D_{it+1}). So, the system will simulate a value from (A_{it+1}, B_{it+1}, C_{it+1}, D_{it+1}) to determine an estimation of the new reward per click θ_{it+1}. We note that for the rewards per click we consider a position more conservative than for the percentage of clicks per position on the list, i.e., the firms update more slowly their estimations on the reward per click. However, as in the previous updating algorithm, it is not difficult to modify the software tool to consider another procedure of updating the rewards per click, for example, following the same idea as in the case of the percentage of clicks per position on the list. This kind of possibility provides certain flexibility to the software tool implemented with respect to consider other additional situations not included in the version presented in this chapter.

3.3.3 Bids

In this section we show how the system assesses a finite set of feasible bids for each firm in order to determine their optimal bid. The optimal bid will be the feasible bid which maximizes expected value of the utility function of the evaluated firm. In order to determine the optimal bid of firm i, i=1,…,n, the system carries out several steps.

First of all, we need to define the risk profile for firm i. Also, we have to specify the number of bids that firm i will consider as feasible to submit to the Internet search service provider. The system will select one of these bids as the optimal one for firm i in period t according to its utility function UF_i.

Secondly, we have to determine the probability of appearing in each position j on the list, for all j=1,…,n, once the firm has submitted a particular bid. This process is carried out for each one of the feasible bids for firm i, for all i=1,…,n.

Third, we calculate the expected value of the utility function for each one of the feasible bids. In this set, a bid is chosen so as to maximize the expected utility, following our assumptions. Roughly speaking, this is the procedure to determine the optimal strategy for firm i.

Obviously, this process is carried out for each period t, t=0,…,T-1, and for each firm i, i=1,…,n, obtaining the simulated bids and the ranking list for each period.

Overall, the determination of the optimal bid depends on the risk profile, the average reward per click, the private forecast about the number of clicks per position, and the perception that each firm has about the rest of competitors, as we show next.

Step 0. The system executes a realization of the trapezoidal fuzzy number (A_{it}, B_{it}, C_{it}, D_{it}) to obtain an average reward per unit θ_{it} to be used in the following steps.

Step 1. Let us assume that we are working with firm i in period t. A similar process is performed for all periods and all participating firms in the ranking auction.

The system simulates the risk profile for the firm (see Section 3.2.6.2). So, we obtain an integer number, 0, 1 or 2, corresponding to a risk loving, a risk neutral or a risk averse player, respectively. Once the system has simulated the risk profile, we know how firm i will behave. Nevertheless, we need to know the set of feasible bids as well.

Step 2. Following the inputs introduced into the system and the values obtained in the previous steps, the number of feasible bids will be K+1, where K is an integer number such that $K \le (\theta_{it} - r)/\Delta$ (see Section 3.2.2). In particular, the set of feasible bids is obtained by means of the simple expression $r+k\Delta$, where $k=0,1,...,K$. The number of bids to be evaluated can be as greater as one likes because the tool allows us to choose the minimum amount Δ to be considered (see Section 3.2.2).

Step 3. Once we know the set of feasible bids and the risk profile, we have to determine the probability of appearing in each position on the list since a firm is not able to know its position on the list before submitting a particular bid. Therefore, in particular, we are interested in determining the corresponding probability distribution associated to each feasible bid b_{it}. We will use this probability distribution to calculate the expected value of the utility function given bid b_{it}.

Obviously, firm i's profit not only depends on firm i's bid but also on the bids of the rest of competitors. Therefore, it is also necessary that the firm assumes a certain kind of behaviour for each of its competitors.

In order to make clear the above point, we first show the expression of the expected value of the utility function given bid b_{it} for firm i:

$$E[UF_i]=\sum_{j=1}^{n}UF_i\left((\theta_{it}-b_{it})\left(l_i(N_t-N_{dt})+f_{ijt}N_{dt}\right)\right)P(j/b_{it}), \quad (12)$$

where $P(j/b_{it})$ denotes the probability of appearing in position j after submitting the feasible bid b_{it}.

Taking into account that computing mathematically the above probabilities is almost intractable, we will approximate them by simulation. To this end, we will use the ratio of the number of times that firm i has appeared in position j, after submitting b_{it}, to the total number of simulations. In other words, the system simulates an auction as many times as the user introduced into the system. In each auction, firm i will always submit the same bid b_{it}, and will keep fix the same risk profile. Regarding the competitors, the process is more sophisticated. First, the system simulates a value from the fuzzy number of firm i θ_{it} (k), for all k=1,…,i-1,i+1,…,n, in order to obtain the average reward per click which each competitor use. Secondly, the system multiplies that realization of θ_{it} (k) times a factor which depends on the perception of firm i about the risk profile of each firm k. This factor has been modeled in the system by a Beta(a,b) distribution. The parameters a and b have been defined in a different way depending on the kind of risk profile of the firm. If the firm is risk averse then we consider a=2 and b=3. If the firm is risk neutral then we consider a=3 and b=3. And if the firm is risk loving then a=3 and b=2 (other values for parameters a and b could be easily considered just modifying the corresponding part of the code). In this way, the probability density function is asymmetric with a high left tail, symmetric, and asymmetric with a high right tail, respectively, following the natural bidding behaviour of the firms. Finally, the system builds the bid of competitor k, k=1,…,i-1,i+1,…,n, by means of the following expressions

$$b_{kt}=\theta_{it}\,(k)\cdot Exc(Beta(a_k,b_k)), \quad (13)$$

where $Exc(Beta(a_k,b_k))$ is an execution of the distribution $Beta(a_k,b_k)$.

We note that we consider that each firm knows neither the average reward per unit of the others nor a particular estimation on them, therefore they use their own knowledge about the average reward per unit to evaluate the possible averages reward per unit which can be used by their competitors. In some sense, each firm considers that its knowledge on the average reward per unit is good enough and the other firms have the same (or very similar) information about that.

Step 4. The system calculates the expected value of the utility function, (12), for the bid b_{it}. This procedure is repeated for each feasible bid for firm i. In this way, the system is able to select the optimal bid, i.e., the feasible bid with the highest expected value of the utility function.

To end this section, it is worth noting that all firms that participate in the auction submit their optimal bids. So, the system ranks the firms in descending order according to all these bids. Consequently, the system is able to build a list for each period t, t=0,…,T-1.

3.4 Some computational experience

In order to show how the system works we use two stylized and simple examples one with only two firms and another with five firms. In the first case, the example has the following characteristics:

Inputs	**Value**
Num. of firms	2
Num. Simulations	1000
Num. of periods	10
Num. Customers	N(1000,5)
% disloyals	100%
Prob. of purchasing	0.2
Expenses	1 €
Reserve price	5 cent/€
Δ	0.1
p_1	0.9
p_2	0.2

Table 1. Example with two firms

In this first example, we work with only two risk neutral firms, during a period of 10 days, assuming that all the customers which visit the Internet search engine are disloyal, the probability of purchasing is 0.2 and when a customer makes the decision of purchasing her expenses is constant and equals to 1€. On the other hand, the proportion of clicks received if the firm is ranked first clearly higher than if the firm is ranked second. Also, we assume that both firms are symmetric. In other words, both present the same features. In particular, the starting average reward per click is modelled by means of the trapezoidal fuzzy number (17, 17.5, 19.5, 20.0). Regarding the private forecasts about the number of clicks per position on the list, we consider that the firms' estimations deviate significantly from the actual value of the parameters, p_1 and p_2. In particular, we consider that $f_{110}=f_{120}=f_{210}=f_{220}=0.5$, therefore, the firms evaluate that the position is not relevant to obtain more clicks and hence a higher expected revenue. Therefore, one could expect that the firms bid for the first day the reserve price to appear on the list in whatever position. Finally, we assume that each firm believes that its rival is risk neutral as well.

In Figure 7 the bidding strategy for each firm over time is shown. As it can be seen, on the first day, both firms submit the reserve price as optimal bids as one could expect without any analysis. As it was noted before, it is due to the fact that the number of clicks the firms will receive if they are ranked first or second is little sensitive. In other words, the firms think that they will receive the same number of visits independently on the position at which they appear. Therefore, they have little incentives to bid aggressively. However, since the private forecasts about these parameters change over time (see Figure 8, for Firm 2 we obtained a similar figure) showing increasingly the importance to be ranked first on the list, the firms bid more aggressively. In some way, they are having additional information about the real number of visits per position and it allows them to improve the forecast accuracy of p_j, j=1,2. Since p_1>>p_2 the firms have incentives to bid higher.

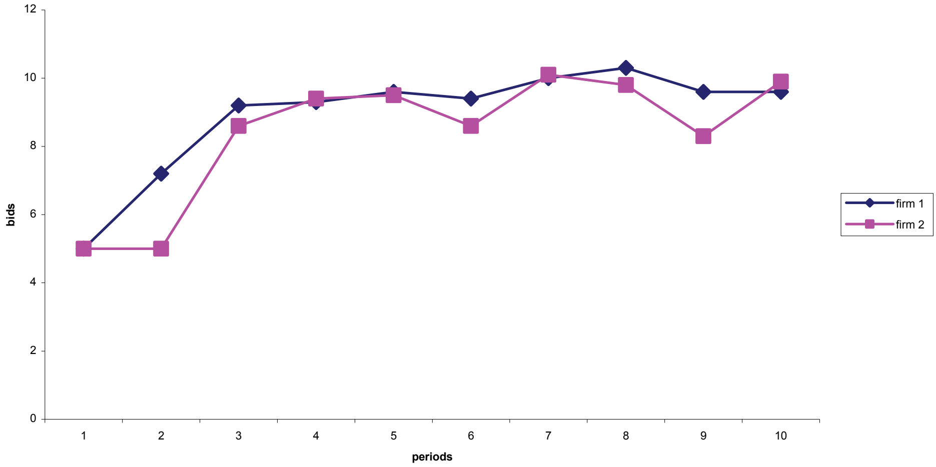


Fig. 7. Bids of the two firms for the studied period

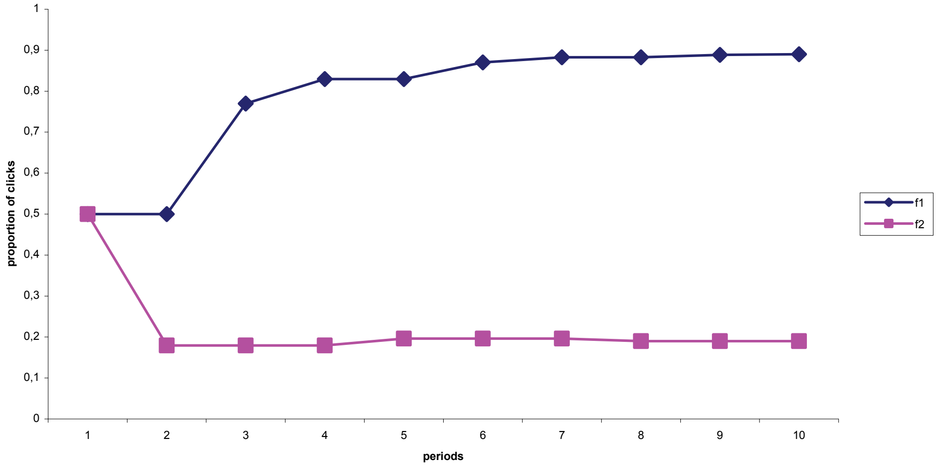


Fig. 8. Evolution of Firm 1's private forecasts about p_j

Finally, from Figure 7 and Figure 8 we can observe that the optimal bid for both firms after the ten periods is approximately 10 cents and in few periods the firms have a reasonable good estimation on the parameters p_1 and p_2.

In order to show how the system works under more competition, we consider a second example with five firms. The particular characteristics of this numerical example are shown in the following table:

Inputs	Value
Num. of firms	5
Num. Simulations	1000
Num. of periods	10
Num. Customers	N(1000,5)
% disloyals	100%
Prob. of purchasing	0.2
Expenses	1 €
Reserve price	5 cent/€
Δ	0.1
p_1	0.9
p_2	0.7
p_3	0.3
p_4	0.2
p_5	0.1

Table 2. Example with five firms

Apart from the characteristics given in Table 2, we also assume that the five firms are risk neutral and symmetric again. In particular, we modelled θ_{kt} by the trapezoidal fuzzy number (17, 17.5, 19.5, 20). On the other hand, the private forecasts about p_j, j=1,…,5, will be 0.8, 0.7, 0.6, 0.3 and 0.1, respectively. It is worth noting that in this second example the starting estimations are more realistic than in the previous example. It should imply that firms use a more aggressive bidding strategy from the first period, t=0.

Next in Figure 9 we show the information about the bids submitted to the system for each firm period by period. We observe that the optimal bid for firms after the ten periods analysed is close to 12.

In this case, the value of the parameters encourages aggressive bidding even in the first period. Unlike the previous example, the firms submitted a bid strictly greater than the reserve price in period t=0. Note also that in the final periods the average bid with five firms is greater than the average bid with two firms. Obviously, it is consequence of the intrinsic competition of both examples.

Regarding the estimation of p_j, each firm learnt over time about the real proportion of clicks per position. In this way, at the last period, the private forecasts about these parameters are very close to the actual values (see Figure 10).

It is worth noting that the system generates more information about the auction over time than the above presented. For example, number of clicks received, expenses related to the customers, the evolution of the reward per click, etc. Once the simulation has been completed, all the data obtained are presented in a practical format such as spreadsheets, which allows us to store on the hard disk the results of all the simulations carried out. Nevertheless, we only wanted to show briefly some of the results that the developed software is able to yield.

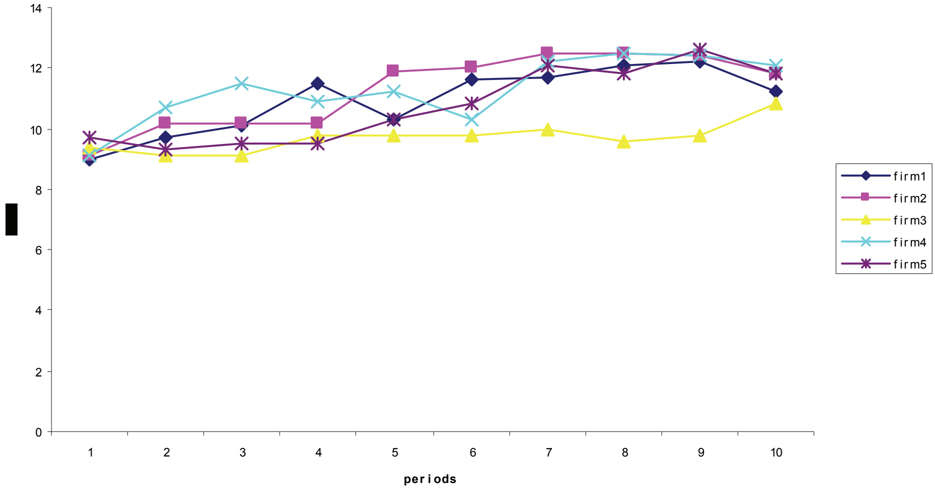


Fig. 9. Bids of the five firms for the studied period

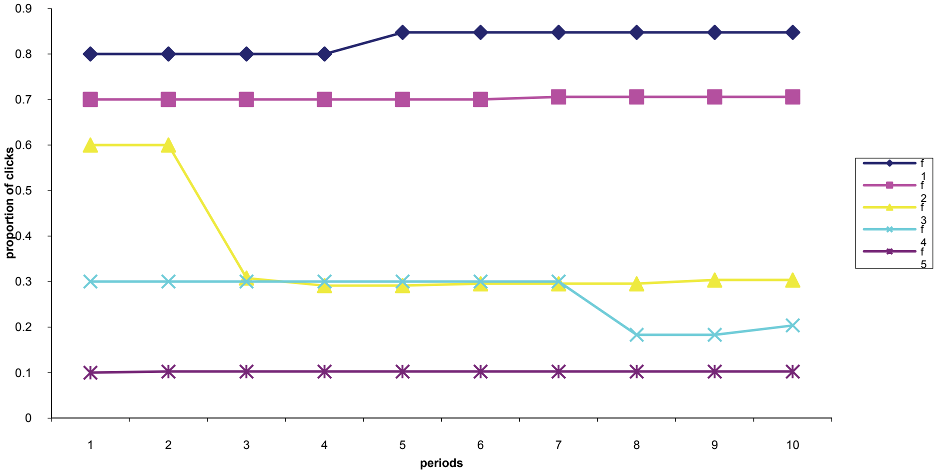


Fig. 10. Evolution of Firm 1's private forecasts about p_j

4. Conclusions and further research

The main objective of this work has been to develop a simple software tool in the form of a Decision Support System or computational framework for analysing ranking auction markets for Internet search service providers which could be useful for economic scholars or practitioners. The particular features of this tool make possible to simulate in a clear and simple way the bidding behaviour of a set of firms when facing ranking auctions situation on the Internet.

This tool could be interesting for analysing different aspects of Internet search engines. In particular, the tool provides information about how the Internet search engine could induce firms to bid more aggressively, or whether it is beneficial for the provider to disclose more information regarding the number of clicks per position to the firms, the effect of collusion or coordination, etc. On the other hand, the software tool has been implemented to give the possibility to modify easily some parts (in particular some algorithms) in order to consider other situations not included in the present version.

We would like to finish mentioning some additional topics for further research on the Decision Support System considered in this chapter. First, we could use the software to check whether the results obtained by (Lim & Tang, 2006) for only two firms are correct for a greater number of firms. Secondly, we could analyse how a multiple period auction affects firms' bidding strategies. Third, we could also study how collusion over time can distort the final results of the auction. Overall, we view our approach as a building block or framework for developing further analysis.

5. Acknowledgements

The research of Juan Aparicio is supported by Ministerio de Ciencia e Innovacion of Spain through the project MTM2009-10479. On the other hand, Joaquin Sanchez-Soriano acknowledges the financial support from the Ministerio de Educacion y Ciencia of Spain through the project MTM2008-06778-C02-01.

6. References

Atkins, K.; Marathe, A. & Barrett, C. (2007) A computational approach to modelling commodity markets. *Computational Economics*, Vol. 30, No. 2, September 2007, 125-142, ISSN 0927-7099

Feng, J.; Shen, Z.M. & Zhan, R.L. (2007). Ranked items auctions and online advertisement. *Production and Operations Management,* Vol. 16, No. 4, July-August 2007, 510-522, ISSN 1059-1478

Klemperer, P. (2004). *Auctions: Theory and Practice*, Princeton University Press, ISBN 978-0-691-11925-0, USA

Krishna, V. (2002). *Auction Theory,* Academic Press, ISBN 978-0-12-426-297-3, USA

Lim, W.S. & Tang, C.S. (2006). An Auction Model Arising from an Internet Search Service Provider. *European Journal of Operational Research,* Vol. 172, 956-970, ISSN 0377-2217

Mehlenbacher, A. (2009) Multiagent System Simulations of Signal Averaging in English Auctions with Two-Dimensional Value Signals. *Computational Economics*, Vol. 34, No. 2, September 2009, 119-143, ISSN 0927-7099

Sancho, J.; Sanchez-Soriano, J.; Chazarra, J.A. & Aparicio, J. (2008). Design and implementation of a decision support system for competitive electricity markets. *Decision Support Systems,* Vol. 44, 765-784, ISSN 0167-9236

Sancho, J.; Sanchez-Soriano, J.; Pulido, M.; Llorca, N. & Aparicio, J. (2009). Ranking auctions: a cooperative approach. *Working Paper, Center of Operations Research,* CIO-2009-5, 1-23, ISSN 1576-7264

von Neumann, J. & Morgenstern, O. (1944). *Theory of Games and Economic Behavior,* Princeton University Press, ISBN 978-0-691-13061-3, USA

16

A Fuzzy – Based Methodology for Aggregative Waste Minimization in the Wine Industry

Ndeke Musee, Leon Lorenzen and Chris Aldrich
Department of Process Engineering, University of Stellenbosch
South Africa

1. Introduction

The wine industry generates large quantities of waste annually, including organic solid wastes (solids, skins, pips, marc, etc.), inorganic solid wastes (diatomaceous earth, bentonite clay, perlite), liquid waste (cleaning wastewater, spent cleaning solvents, cooling water), and gaseous pollutants (carbon dioxide, volatile organic compounds, ammonia, sulphur dioxide, etc.) (Chapman et al., 2001; Musee, 2004a; Musee et al., 2007). Several factors give rise to these diverse waste streams (Musee, 2004a; Musee et al., 2007), however, only the most salient ones are highlighted here. Firstly, wine production evolved from a cottage industry to a global industry. Because of their antiquated origin, the design and development of many wineries made no provision for in-plant modern waste minimization (WM) approaches. Secondly, because the wine industry is dependent on an agricultural feedstock (grapes), the resultant waste streams tend to have a high concentration of organic material. This is because the grape feedstock cannot be altered, replaced, or eliminated before the vinification process begins – if the finished wine quality is to remain consistent. And finally, although auxiliary process feedstock, such as filter aids and diatomaceous earth are essential for clarifying the wine, they cannot be incorporated into the final product. Consequently, the clarification agents constitute part of the waste streams generated from the wine industry. In view of these unique constraints facing the wine industry, among others, necessitates the development of appropriate WM strategies to address the waste management challenges facing the wine industry (Musee et al., 2007).

In recent years, there has been continuous pressure on the operating profits of wine makers, mainly owing to increasing competitiveness in the global wine market. This can be attributed to increased variety of wine brands, rise in operational and input material costs, as well as the emergence of an onerous environmental regulatory framework in many wine producing countries (Bisson et al., 2002). Notably, the impact of stringent environmental legislation on the cost of production is expected to continue to be a key determinant in the international competitiveness of wine products (Katsiri & Dalou 1994; Massette, 1994; Müller, 1999). This, and a combination of other powerful intrinsic and external drivers should motivate the wine industry to consider the possibility of incorporating WM strategies as an integral part of wine making processes. As such, the identification and implementation of appropriate WM strategies should be part of the drive to reduce the cost of wine production – particularly in the context of ensuring its future sustainability.

In practice, vinification processes are characterized by complex interactions amongst different production processes. As a result, any effective attempt to enhance winery waste management is likely to require a solution comprising of several WM strategies, and implemented concurrently. However, such an undertaking is dependent on the identification of suitable strategies, and secondly, a careful assessment of each strategy to determine its likely influence in addressing the overall WM problem in the wine industry. Moreover, the assessment of each strategy would inevitably entail the use of multiple screening criteria, such as technical feasibility, economic and social imperatives as well as environmental integrity. Unfortunately, the application of different criteria for the ranking of WM strategies is complicated by the lack of quantitative operating data presently available in the wine industry (Musee et al., 2006a).

Nonetheless, to be effective, decision support tools designed to facilitate waste management in the wine industries should ideally be able to exploit the qualitative data available, as these data constitute a vital component of industry knowledge. Fuzzy logic (Zadeh, 1965; Bonissone, 1997; Yen & Lugari, 1998; Ross, 2004) provides such a platform. Previously fuzzy logic has been applied in developing rational solutions for complex real world problems (Bonissone, 1997), and offering interpretable results (Setnes et al., 1998). For example, successful applications of fuzzy logic have been demonstrated in domains such as process design (Huang and Fan, 1995), water quality assessment (Ocampo-Duque et al., 2006), manufacturing (Büyüközkan & Feyzioğlu, 2004), safety (Gentile et al., 2003), sustainability (Phillis & Andriantiatsaholiniaina, 2001; Gagliardi, et al., 2007; Musee & Lorenzen, 2007; Prato, 2007), and hazardous waste classification (Musee et al. 2006b, Musee et al, 2008a; Musee et al, 2008b).

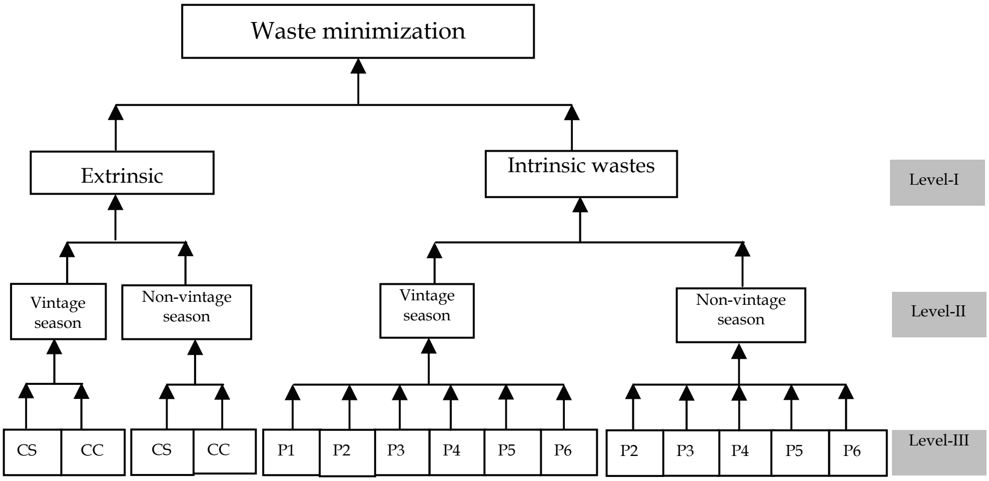
Musee and co-workers (Musee et al., 2003; Musee et al., 2006a) studied a fuzzy logic approach to support decision making in the wine industry that entailed the ranking of WM strategies based on experts' opinions. Because experts hold widely different opinions, this approach yielded decisions associated with a high degree of uncertainty. In the current work, this drawback is addressed using a fuzzy logic framework by combining the ranking of expert opinions with operational data to improve the analysis and selection of WM strategies in the context of wine production. The merits of the proposed approach will be illustrated with two case studies.

This chapter is organized as follows. Section 2 provides an overview of the waste management in the wine industry, and the tools applied to model such a highly unstructured problem. The tools used in modelling the wine waste management problem comprised of; the screening and ranking indices, qualitative reasoning in developing various probable scenarios, and the fuzzy logic. In Section 3, a case study on WM in the wine industry is introduced, where a conceptual model – the intelligent decision support system together with mathematical equations – and how the knowledge stored in different knowledge rule bases were linked to effectively evaluate WM in the wine industry context. Section 4 presents results derived from the model, and a discussion on their application to real-world winery operations with respect to WM. The main findings of the chapter are presented in Section 5.

2. Basics of waste management in the wine industry

2.1 Hierarchical evaluation of wineries

The hierarchical analysis of process systems has its origin in the hierarchical decision approach developed by Douglas (1988). In this chapter, hierarchical analysis was applied to decompose the waste management problem in the wine industry. The vinification process was decomposed into several subtasks, followed by the identification of the most influential variables concerning: (i) the degree of recovery of products and by-products during the production processes; (ii) the quantity and quality of effluent generated during cleaning and sanitization processes; and (iii) the quantity of chemicals consumed during cleaning and sanitization processes.



CS: Cleaning and sanitation; CC: Chemical consumption; P1: Crushing and destemming processes; P2: Transfer processes and operations; P3: Filtration, P4: Pressing; P5: Fermentation; P6: Bottling and packaging.

Fig. 1. Analysis of the vinification processes using a hierarchical approach for the identification of WM strategies.

Fig. 1 depicts a hierarchical model of vinification processes. In this study, the operational variables were decomposed into three levels based on literature survey and interviews with experts knowledgeable on waste management practices and norms in the wine industry. In Level-I, different waste types generated from the vinification processes were classified as intrinsic (process) or extrinsic (utility). Detailed description of waste classification in the wine industry has been presented elsewhere (Musee et al., 2007), and will not be repeated here. In this study, the breadth of the adopted waste classification approach ensured that no waste stream was left unaccounted for. In addition, the model gave rise to consistent and robust results. These aspects will be elucidated in details in Sections 3 and 4.

Owing to the seasonality of the vinification processes and high value-added nature of the product (wine) – wine production is an ideal candidate for both batch and semi-batch manufacturing techniques. This causes a wide variation in the composition of the waste streams - characterized by strong seasonal dependence. In Level-II, the vinification process is characterized by two seasons, viz. the vintage and non-vintage season. It should be noted that a definitive distinction between vintage and non-vintage season is more or less dependent on the processes that take place during a given period on the vinification calendar (Chapman et al., 2001).

The vinification process was further decomposed based on the seasonality of the wine production which led to the identification of the most predominant processes under each season as described in Level-III (see Fig. 1). The next task was to identify how each process, for instance; crushing and destemming, fermentation, or filtration, etc., contributed to the final waste matrix in a given vintage season. A close scrutiny of the waste streams indicated that the final effluent quantity and composition, product and by-products losses, as well as the quantity of chemicals consumed had a strong dependence on the vinification season, as well as the processes operated in a given season. And the final task entailed the development of a systematic approach for identifying WM strategies under each process. The adopted methodology comprised of a three-step sequential approach, namely; waste source identification, qualitative waste causative analysis, and the formulation of feasible alternatives for WM in the wine industry (Musee, 2004; Musee et al., 2007).

2.2 Screening and ranking index

In an earlier study (Musee et al., 2006a), experts were asked to rank the WM strategies based on their waste management experience in the wine industry. However, the approach was found to be cumbersome, owing to the large number of strategies to rank (see details in Musee et al., 2007). This resulted in inconsistencies in the final ranking of WM strategies. This drawback is addressed in this chapter by using a more rigorous ranking approach that includes the use of a WM index (WMI). The WMI was used to assign dimensionless scores to each strategy in terms of its overall potential degree of influence on a specific targeted system output (e.g. chemical usage, effluent quality, etc). The WMI was more effective than the less formal approach previously reported by Musee et al. (2006a), because the influence of the experts' subjective perceptions or personality aspects were eliminated.

Generically, the evaluation criteria for WM alternatives often use economic functions which often lead to unintended consequences of identifying suboptimal solutions. Another demerit of relying solely on economic criteria lies in their inherent bias for identifying inferior alternatives purely based on cost. These criteria tend to favour options geared towards waste treatment above those that promote waste reduction, elimination, reuse, or recycling. In this investigation, a multifaceted criterion was used, accounting for the unique and operational constraints experienced by managers and operators in the wine industry. Ranking and screening of waste streams and pollution prevention systems have been extensively discussed by several researchers (Hanlon & Fromm, 1990; Balik & Koraido, 1991; Crittenden & Kolaczkowski, 1995; Smith & Khan, 1995; Allen & Rosselot, 1997). These indices were found to be strongly dependent on the nature of the industry, or problem under consideration, as well as the databases accessible in the domain under study. In this chapter, the Smith and Khan's Index (Smith & Khan, 1995) commonly referred as the pollution prevention index was modified to suit the limitations of the accessible data in the wine industry.

For the purpose of producing a broader and more acceptable prioritization of the derived WM strategies, each strategy was evaluated based on a set of multiple criteria, including the position of a given strategy in the WM hierarchy. Consequently, source reduction was assigned a higher priority in comparison to reuse and recycling. Different criteria were assigned different weights as shown in Table 1. The reason being, each criterion had a different impact on the reduction of the overall quantity of waste generated and the capital costs required for its successful implementation. It should be noted that the Smith and Khan Index (Smith & Khan, 1995) was comprised of source reduction, reuse, and waste treatment in accordance to the EPA pollution prevention hierarchy (USEPA, 1988). However, in this study, the waste treatment, payback period and depth of solution criteria were excluded from the index, owing to waste management challenges unique to the wine industry. The rest of the criteria used in describing WM solutions were recycling, degree of waste reduction expressed in percentage, ease of implementing a given strategy, and the capital cost of a given solution. The weights assigned in each of the criteria were in descending order in accordance with the foregoing description, such that capital cost was assigned the lowest weight of 1, whereas source reduction had the highest weight of 10^5. Note that the percentage of waste reduction was expressed in qualitative terms – and generically was referred as the waste reduction possibility. Thus, the WMI for a given strategy was computed using the relation:

$$\mathrm{WMI} = \mathrm{SR}\times 10^5 + \mathrm{Re}\times 10^4 + \mathrm{R}\ \times 10^3 + \mathrm{WRP}\times 10^2 + \mathrm{IP}\times 10^1 + \mathrm{CC}\times 10^0 \qquad (1)$$

where SR: source reduction, Re: Reuse or recovery, R: reclaim or recycle, WRP: waste reduction possibility, IP: implementation potential, and CC: capital cost.

Criteria	Weight	Activity		Index value
Source Reduction (SR)	10^5	Elimination		1.00
		Minimize	High	0.75
			Medium	0.50
			Low	0.25
Reuse/Recovery (Re)	10^4	Full		1.00
		Partial		0.67
		Low		0.33
Reclaim/Recycling (RR)	10^3	Full		1.00
		Partial		0.67
		Low		0.33
Waste Reduction Possibility (WRP)	10^2	No reduction (nr)		0
		Low reduction (lr)		1
		Moderate reduction (mr)		2
		High reduction (hr)		3
Implementation Potential (IP)	10^1	Procedure change (pc)		5
		Material substitution (ms)		4
		Preventive maintenance		3
		Retrofit equipment (re)		2
		New equipment		1
Capital Cost (CC)	10^0	No cost (nc)		5
		Low cot (lc)		4
		Moderate cost (mc)		3
		High cost (hc)		2
		Very high cost (vhc)		1

Table 1. WM index for the wine industry (adapted from the Smith and Khan Pollution Index (Smith & Khan, 1995).

2.3 Qualitative reasoning

The concept of qualitative reasoning (Bobrow, 1984; Kleer & Brown, 1984) was applied to aid in representing and making available general and physical knowledge commonly used by engineers, scientists and managers to address the environmental problems experienced in the wine industry – without invoking mathematics of continuously varying quantities and differential equations. This is because qualitative reasoning provides the most suitable platform to represent numerous qualitative abstractions specific to the wine industry through creation of quantitative models, without the necessity for rigorous mathematical computations. The use of qualitative symbolic representations and discrete quantities aided in modeling the complex behavior of different vinification processes and unit operations. Therefore, the qualitative approach aided in predicting the behavior of processes and unit operations satisfactorily because only a small number of qualitative variables were required to describe the system. As such, the qualitative reasoning was used to describe the qualitative 'states' attainable with or without implementation of a given strategy in order to mitigate against waste generation, or in improving the management of inevitable waste streams.

Consequently, a qualitative model was developed as point of departure without the necessity for detailed information on the implementation of various WM strategies in the winemaking process. For example, consider the WM strategy where a counter current method is applied to reduce the effluent generated during cleaning. Using qualitative reasoning, it is feasible to predict at least three possible 'states' – which satisfies the condition of using a small number of qualitative variables – after the strategy is applied. The states were derived from casual observations, or based on experience from previous measurements where the primary goal was to model the level of the strategy's actual 'effectiveness' after implementation.

The effectiveness of applying a counter current WM strategy in a specific winery was described by three 'states', namely; effective, partially effective, not effective. Therefore, there were three possibilities regarding the final effluent quantity that can be predicted using a qualitative reasoning approach, viz.; high potable water usage (if the strategy is poorly or not implemented), moderate potable water usage (if the strategy is partially implemented), and low potable water usage (if strategy implemented adequately).

Practically, the implementation of a given WM strategy can yield a continuum of possibilities ranging from the best case scenario (adequate implementation) to the worst case scenario (poor or no implementation). To attain such possibility may necessitate an increase on the number of predictable states by the model from three to nine. This has the advantage of broadening and increasing the sensitivity of the solution space of the decision support system. This was achieved by combining the degree of belief, or level of confidence (CF) the user expresses on a particular response regarding the implementation of a given strategy. Three levels of CF values were specified in this work, and were combined through simple algebraic multiplication to the dimensionless scores representing the qualitative values of a given strategy to expand the predictable states from three to nine. If the qualitative values of a given strategy are assigned dimensionless scores, x_{i1}, x_{i2}, x_{i3}, and CF values y_1, y_2, and y_3: then the possible predictable states for a single strategy or action can be modelled by the relation:

$$(x_{i1}, x_{i2}, x_{i1})\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = (x_{i1}y_1, x_{i1}y_2, x_{i1}y_3, x_{i2}y_1, x_{i2}y_2, x_{i2}y_3, x_{i3}y_1, x_{i3}y_2, x_{i3}y_3) \qquad (2)$$

where the CF values were fixed at 1.00, 0.75 and 0.50 for y_1, y_2, and y_3, respectively, and i denotes the strategy under consideration. Note that $x_{i1}y_1$, $x_{i2}y_1$, $x_{i3}y_1$ are equal to the original three predictable states represented by the values x_{i1}, x_{i2}, x_{i3}, correspondingly.

The next task was to determine a suitable methodology of aggregating the quantitative outputs derived using Eq. 2 that influences a given variable, which in turn exerts a direct impact on the magnitude of a specific targeted system output (e.g. effluent quantity). In this study, the aggregation process was based on the following premise. No single strategy could adequately address the WM problem in the wine industry in a given process or unit operation. Hence, several strategies had to be implemented concurrently in an integrated manner in order to address the waste management challenges in the wine industry adequately. Therefore, the variables used as linguistic input values into the fuzzy model were functions of sums of individual strategies influencing it (Section 3). Equally important, to determine the crisp numerical value of a given variable, all scores from strategies influencing it were summed and normalized using a mathematical expression of the form:

$$Var_k = \frac{\sum_{i=1}^{n} F(x_{ni}, y_j)}{Max\left(\sum_{i=1}^{n} F(x_{ni}, y_j)\right)} \times S_m \tag{3}$$

where Var_k is the k$^{\mathrm{th}}$ variable, x_{ni} is the dimensionless score assigned to a strategy's qualitative value, y_j is the user's level of confidence to a given response, n is the total number of strategies influencing the k$^{\mathrm{th}}$ variable, j= 1, 2, 3 with fixed numerical values of 1.00, 0.75 and 0.50, correspondingly; and S_m is the m$^{\mathrm{th}}$ standardization coefficient where its values were 10 for m=1, or 100 for m=2. The aggregation principle applied in this study is schematically represented in Fig. 2.

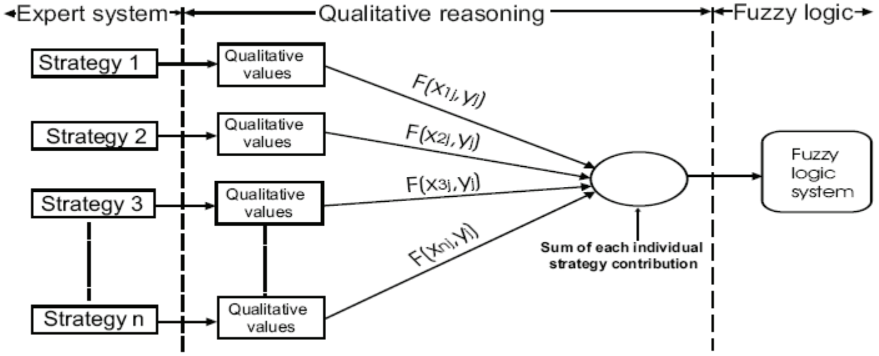


Fig. 2. A conceptual framework based on qualitative reasoning to transform qualitative values into fuzzy numbers.

2.4 Fuzzy logic

The fundamentals of fuzzy set theory (Zadeh, 1965; Yen & Lugari, 1998) are well known (Klier & Yaun, 1988; Zimmermann, 1991; Yager & Zadeh, 1992; Yager & Filev, 1994; Yen & Lugari, 1998; Mamdani & Assilian, 1999). Thus, only the most salient features of fuzzy systems essential for designing and developing an intelligent decision support system are summarized. Fuzzy logic generalizes ordinary or classical sets in an attempt to model and simulate human linguistic reasoning particularly in domains characterized by incomplete, imprecise, vague and uncertain data and knowledge. As such, fuzzy logic being a soft computing tool has the "ability to compute with words", and therefore, provides rational and well reasoned out solutions for complex real world problems (Bonissone, 1997), such as WM in the wine industry (Musee et al., 2004a; Musee, 2004b; Musee et al., Musee et al., 2005; Musee et al., 2006a).

2.4.1 Fuzzy membership functions

In a fuzzy system, the variables are regarded as linguistic variables to aid 'computation with words'. A linguistic value is defined as a variable whose value is a fuzzy number, or is a variable defined in linguistic terms (Lee, 1990). Each linguistic value, LV, is represented by a membership function $\mu_{LV}(x)$. The membership function associates each crisp input, say X_A, with a number, $\mu_{LV}(x_A)$ in the range [0,1]. Essentially $\mu_{LV}(x_A)$ represents the grade of the membership of x_A in LV, or equivalently, the truth value of proposition 'crisp value A is LV'. The overlapping of the membership functions allows an element to belong to more than one set simultaneously, and the degree of membership into each set indicates to what extent the element belongs to that particular fuzzy set (Fig. 3).

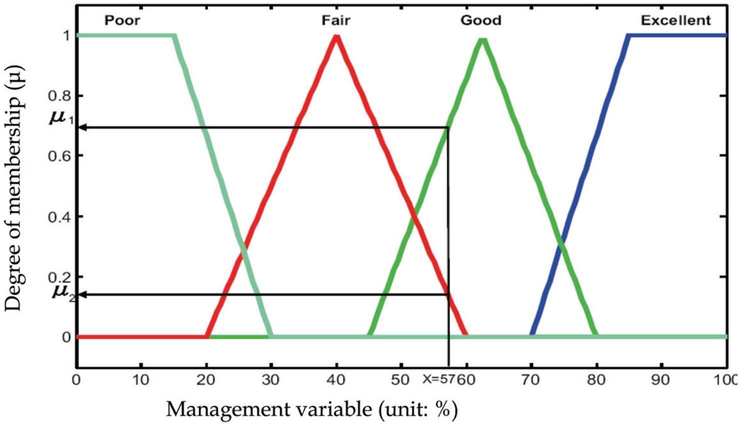


Fig. 3. Triangular and trapezoidal membership functions for the effluent quality management variable with fuzzy linguistic terms: poor, fair, good, and excellent.

To illustrate the functionality of the membership functions for the purpose of determining the linguistic value of a given variable, a simple example is provided. Let the crisp input value for the effluent quality management linguistic variable be x = 57 in a universe of discourse of 0 to 100. Then, according to Fig. 3, an input of 57 generates two membership functions μ_{AC} (x_i), viz. $\mu_1 = 0.686$ in the fuzzy set labelled Good and $\mu_2 = 0.150$ in the set labelled Fair. Note that the rest of linguistic values Poor and Excellent each had a membership function of zero. By applying the max-min fuzzy inferencing algorithm (Lee, 1990) where the membership function values are $\mu_1 = 0.686$ and $\mu_2 = 0.150$, then the linguistic value was determined as Fair (Min [(0.686, 0.150)] =0.150).

2.4.2 Knowledge representation

In a fuzzy rule-based modelling system, knowledge is represented by use of linguistic IF-THEN rules. Ultimately, this renders the knowledge library (base) the core of the system and, the breadth and quality of the knowledge determine the capacity of the system to render useful intelligent decision support. Generically, the premise and conclusion parts of the fuzzy rules are of the form:

$$R^{(l)}\text{: IF } x_1 \text{ is } F_1^1 \text{ AND}\ldots\ x_n \text{ is } F_n^l,$$

$$\text{THEN } y \text{ is } G^l \tag{4}$$

where F^l_i and G^l are fuzzy sets, $x = (x_1, \ldots , x_n)^T \in U$ and $y \in V$ are linguistic input and output variables, respectively, with l = 1, 2, …, M. Practically, the fuzzy IF-THEN rules provide a convenient framework for incorporating human experts' knowledge in fuzzy expert systems. Each fuzzy IF–THEN rule (Eq. 4) defines fuzzy set $F^l_1 \times \ldots \times F^l_n \Rightarrow G^l$ in the product space U × V.

The knowledge necessary for decision making regarding WM in the wine industry was encapsulated in several rule bases with a total of 152 rules. Additionally, the rules were systematically encoded into different hierarchically interlinked rule bases to ensure their easy accessibility at different levels of the decision support execution, and on the other hand, to minimize the overall number of rules in the rule bases. The design of hierarchically interlinked rule bases is analogous to consulting several experts on a certain problem, to derive a final conclusion that takes into account each individual opinion. The model is flexible, robust, and allows the user to choose initial values or adjust the rules in any knowledge base on the basis of operational realities related to the vinification process or processes under scrutiny. Notably, the use of few IF-THEN rules has the merit of aiding in validating the functionality and the contribution of each rule in a given rule base.

2.4.3 Fuzzy inferencing

The core of decision making in a fuzzy logic system is the inference engine. Fuzzy inferencing is used to derive an aggregated output from a particular knowledge base using the rules coded in specific rule bases. In practice, many fuzzy inferencing methods have been developed, with the so-called max-min and max-dot or max-prod (Lee, 1990; Mendel, 1995; Yen & Lugari, 1998) being the most popular. In this case the max-min fuzzy inferencing algorithm proposed by Mamdani & Assilian (1999) was applied. According to the Mamdani-Assilian inferencing algorithm, the truth values of the fuzzy output variables are clipped, such that the area under the clip line determines the outcome of the rule. Finally, a defuzzifier converts the fuzzy aggregate membership grades generated from the inference engine into non-fuzzy output values. Again, there are various approaches to defuzzification (Mazumoto, 1995; Mendel, 1995). The most common is Yager's centroidal method (Yager, 1980), and was applied in this study because of its sensitivity in comparison to other techniques (Yager & Zadeh, 1992).

To illustrate how the fuzzy inference system aided in diagnosing the WM in the industry, a brief description of its salient features and functionalities are presented. Notably, each of the four knowledge sub-modules had a set of features in the form of data, information, and knowledge stored in various interlinked data bases and rule bases as described in Section 3. The hierarchical reasoning structure of each knowledge sub-module can be generically summarized as follows:

1. The linguistic set of strategies/actions were transformed through qualitative reasoning, as well as ranking and screening processes into dimensionless scores at the first hierarchical level (Level-III) of each knowledge sub-module (Fig. 1) using Eqs. 1 and 2.

2. The qualitative or linguistic strategies/actions were broadly grouped into two or three fuzzy linguistic input variables. For example, strategies affecting the effluent quantity were aggregated into three linguistic input variables, viz. organic matter removal, equipment efficiency, and effluent quantity (volume) management. These linguistic variables were used in evaluating the targeted system output (e.g. effluent quantity) at various hierarchical levels of a given knowledge sub-module (see Figs. 2 and 6).
3. The inference for computing the crisp numerical input variable values was performed through solving a series of algebraic summation equations in a specific knowledge sub-module. For instance, to evaluate the degree of product and by-products recovery Eqs. 9 and 10 (in Section 3) were used to compute the fuzzy crisp inputs for the organic material recovery and the general management variables, respectively.
4. The crisp inputs derived in step 3 for a given targeted system output (e.g. chemical usage or effluent quality) were fed automatically into the fuzzy model to derive a final aggregated and ranked output depending on a given set of user's specified inputs. Note that the final fuzzy model crisp output signified a measure of given winery's performance with respect to the targeted system output such as chemical usage, effluent quality, etc.

To illustrate the system's functionality, let's consider the data and knowledge stored for evaluating product and by-products recovery before wet cleaning and sanitization processes. Assume that from steps 1 to 3, the product and by-products recovery (PBR) and generic management crisps values were computed as 0.48 and 55% in their respective domains of discourse. After coding the crisp input values into the fuzzy product and by-products recovery rule base module – resulted in firing four IF-THEN rules in the rule base as shown in Fig. 4. To infer the final system output, first, each of the four activated rules had to be evaluated individually. The linguistic rules and the evaluation findings are as follows:

Rule #10: IF PBR is Moderate AND Generic management is Good
THEN Effective PBR_{eff} is Moderate
EVALUATION: Min (0.30, 0.57) = 0.30

Rule #11: IF PBR is Moderate AND Generic management is Fair
THEN Effective PBR_{eff} is Low
EVALUATION: Min (0.30, 0.25) = 0.25

Rule #14: IF PBR is Low AND Generic management is Good
THEN Effective PBR_{eff} is Low
EVALUATION: Min (0.96, 0.57) = 0.57

Rule #15: IF PMR is Low AND Generic management is Fair
THEN Effective PBR_{eff} is Low
EVALUATION: Min (0.96, 0.25) = 0.25

The fuzzy model through fuzzification process derived the evaluation results as follows. A PBR input of 0.48 produced 0.96 and 0.30 degrees of membership in the fuzzy sets Low and Moderate, respectively. Similarly, a crisp input of 55% for the generic management variable yielded membership degrees of 0.25 and 0.57 in the fuzzy sets Fair and Good, respectively. The clipped membership functions derived from the four activated rules in the rule base

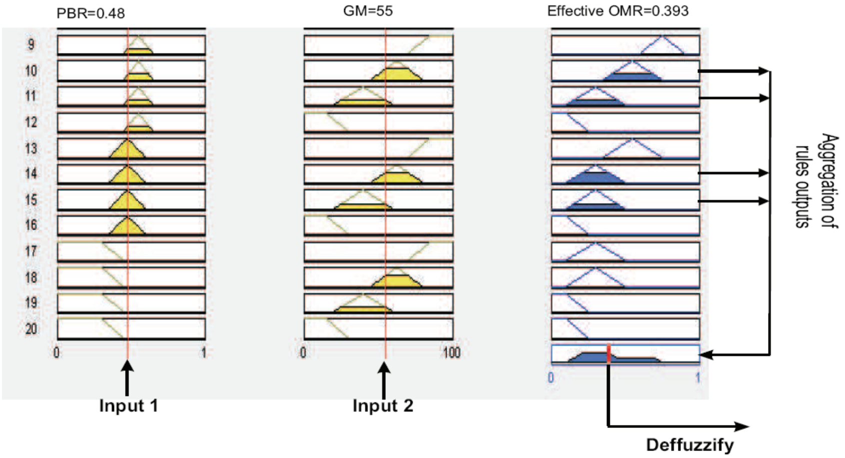


Fig. 4. Fuzzy inferencing mechanism using Mamdani-Assilian model to evaluate the product and by-products recovery (PBR) with two input-variables, and four activated rules.

were aggregated through the defuzzification process into a numerical score (Fig.4). The aggregated system numerical output of the four fuzzy sets signified an overall estimation of the recovered product and by-products, and was determined using the disjunction (max) operator based on the Yager's centroidal defuzzification method given by the expression:

$$Z^* = \frac{\sum_{j=1}^{n} \mu(\omega_j)\dot{\omega}_j}{\sum_{j=1}^{n} \mu(\omega_j)} \qquad (5)$$

Applying Eq. 5 on the four fuzzy set outputs yielded a crisp output of 0.393 which was linguistically ranked as Low recovery of product and by-products.

3. Intelligent decision support system development

A systematic methodology for data acquisition as well as knowledge inferencing and manipulation was developed for the purpose of representing diverse findings concerning different aspects of WM in the wine industry. The unique protocol for data handling was essential to ensure that the final findings were transparently computed, hence, could easily be interpreted by the targeted end-users. Generically, the development of an intelligent decision support system comprised of three steps, namely knowledge acquisition, knowledge representation, and inference mechanism. In the following sections, salient aspects on each of the above steps in the context of WM in the wine industry are summarized.

3.1 Knowledge acquisition

Knowledge acquisition entailed the sourcing of data, information and knowledge concerning WM in the wine industry. The knowledge was manually collected through conducting interviews with experts and extensive literature reviews. The knowledge was broadly classified as generic knowledge (GK) or specific knowledge (SK) (Fig. 5). The GK comprised of WM techniques and practices that were universally applicable to a wide range of processes and unit operations owing to their repetitive (routine) character. Conversely, SK focused on WM techniques and strategies for specific process or unit operation, and particularly targeting the intrinsic waste streams (Musee et al., 2007).

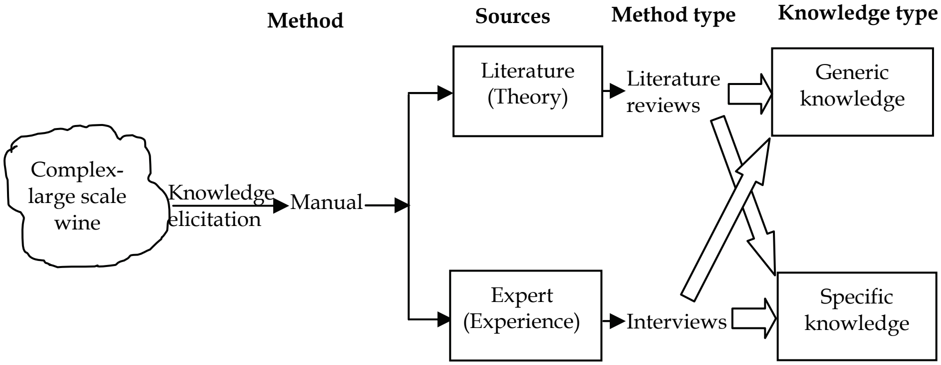


Fig. 5. Data and knowledge sources, knowledge type accessible, and knowledge acquisition for WM in the wine industry.

Equally important, the strategies for either knowledge type (SK or GK) were dependent on the target output under consideration. For instance, the knowledge-type strategies for evaluating the products and by-products recovery were different from those of effluent quality both in the case of intrinsic or extrinsic waste streams. The combining of generic and specific knowledge types and exploitation of the synergies between them lead to the development of rule bases that captured a good degree of WM strategies in the context of the wine industry. Only the knowledge in product and by-products sub-module, as well as effluent quantity sub-module are briefly described in the following sections.

3.1.1 Products and by-products losses/recovery sub-module

Solid wastes from the vinification processes contain commercially valuable products (wine) and by-products. However, depending on the way the solid waste streams are handled, either the products/by-products recovery can be optimized, or huge losses are incurred. The degree of product and by-product recovery is dependent on the vinification processes and unit operations under consideration, or the vinification season (vintage or non-vintage). As a result, the loss of the product and by-products has a direct impact on the final effluent quality and quantity as they resulted into the liquid waste streams – mainly the wastewater.

In order to compute the degree of product and by-products recovery in a given vinification process, strategies related to intrinsic and extrinsic factors, the vintage season, and experts' estimations of relative contribution of the potential losses for each process under scrutiny – were taken into account. Notably, only strategies that had a direct link to recovery, or loss of the products or by-products were taken into account in this sub-module. Experts were asked to provide an estimation of product and by-products overall impact on the quality and quantity of the effluent on a scale of zero to one. An example of heuristics from two experts is presented in Table 2. The values were estimates in relative terms between different processes and unit operations on the basis of a given expert's opinion. During the computation of the final product and by-products the average of values of the six experts who provided inputs were used, for the vintage and non-vintage seasons.

Process/unit operations	Effluent quantity				Effluent quality			
	Vintage		Non-vintage		Vintage		Non-vintage	
	EP1	EP2	EP1	EP2	EP1	EP2	EP1	EP2
Crushing/destemmi	0.10	0.25	0.00	0.00	0.15	0.30	0.00	0.00
Wine transfers	0.25	0.15	0.35	0.25	0.30	0. 10	0.40	0.35
Filtration	0.15	0.10	0.20	0.25	0.10	0.10	0.10	0.15
Pressing	0.20	0.20	0.15	0.15	0.15	0.30	0.10	0.20
Fermentation	0.25	0.25	0.20	0.20	0.25	0.20	0.25	0.15
Bottling/packing	0.05	0.05	0.10	0.15	0.05	0.10	0.15	0.15
Sum of weights	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**

Table 2. Two experts' approximation concerning potential losses of product and by-products under different processes and vinification seasons. Note: Values are based on South Africa vinification processes, EP: expert opinion.

An estimation of the quantities of products and by-products recovered during a given season was computed as follows. First, qualitatively the degree to which the generic and specific strategies were implemented in a given facility was evaluated using the qualitative reasoning approach. Secondly, the computed values owing to generic and specific implementation of the strategies were added. The additive value was taken as an indication of the degree to which product and by-products were recovered from surfaces and equipment before wet cleaning and sanitization processes commenced. Thirdly, to ensure uniformity and interpretability of values in a given process or unit operation, and consequently in the entire vinification process, the computed values were normalized. For example, assume for process i, the values obtained after the implementation of the generic and specific strategies are A_i and B_i, respectively. Then, the normalized value N_i for process i is given by the expression:

$$N_i = \frac{A_i + B_i}{A_{it} + B_{it}} \tag{6}$$

where A_{it} and B_{it} are the maximum values if all the generic and specific strategies were adequately implemented in process i.

Thus, the total recovery of products and by-products (PBR) for the entire vinification process is approximated by a linearly weighted expression:

$$PBR = \sum_{i=1}^{6} (N_{is} \beta_i) \tag{7}$$

where PBR is the total recovered product and by-products from all processes and unit operations in season s defined in the range 0 to 1. Zero means no recovery, hence maximum losses whereas one implies maximum recovery of the product and by-products. $ß_i$ is the weight specified by the waste management experts in the wine industry for processes i (i=1, 2, …, 6) during season s (s = 1 (vintage), 2 (non-vintage)) satisfying the condition:

$$\beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 = 1 \tag{8}$$

Note that processes i (i = 1, 2, …, 6) represent crushing and destemming, transfer systems, filtration, pressing, fermentation, and bottling and packing processes, respectively. N_{is} is an index in the range ($0 \leq N_{is} \leq 1$) representing the recovered organic material in a specific process i before wet cleaning starts and is computed using the relation:

$$N_{is} = \frac{\sum_{i=1}^{n}(W_{nA} \times CF_k) + \sum_{i=1}^{m}(W_{mB} \times CF_k)}{Max\left(\sum_{i=1}^{n}(W_{nA} \times CF_k) + \sum_{i=1}^{m}(W_{mB} \times CF_k)\right)} \tag{9}$$

where W is the dimensionless score assigned to the qualitative values of each strategy; A symbolizes specific strategies; B symbolizes generic strategies; n is the number of specific strategies considered in process i; and m is the number of generic strategies considered to improve product and by-products recovery in all processes and unit operations under consideration; CF_k is a measure of the degree of belief the user has on a given response regarding a particular practice or strategy; k = 1; 2; 3 whose values were fixed at 1.00, 0.75, and 0.50, respectively.

A schematic representation of computations for ranking the product and by-products recovery in a given season is shown in Fig. 6. The generic and specific knowledge derived for the product and by-products recovery sub-module are presented in Tables 3 and 4, respectively. The second variable that exerted influence on the product and by-products recovery is the generic management (GM_{PBR}). This variable was determined by qualitatively evaluating the last four strategies in Table 3. GM_{PBR} was evaluated and normalized using the expression:

$$GM_{PBR} = \frac{\sum_{i=1}^{q}(W_{qC} \times CF_k)}{Max\left(\sum_{i=1}^{q}(W_{qC} \times CF_k)\right)} \times 100 \tag{10}$$

GM_{PBR} was defined in the discourse of 0 to 100. The value 0 represents the worst management scenario, while 100 imply the best management attainable in a specific facility.

In real plant practices, overall effective product and by-products recovery is a function of PBR (operating and technological solutions) and GM_{PBR} (management-related aspects) variables. Since both variables contain uncertainty (vagueness) and are linguistically quantified, the overall effective product and by-products recovery was evaluated using the fuzzy mathematical formalism of the form:

$$PBR_{eff} = f\left(PBR, GM_{PBR}\right) \tag{11}$$

where f is a fuzzy logic function, PBR_{eff} is defined in the range 0 to 1, such that 0 represents a scenario where no recovery of product and by-products takes place before wet cleaning occurs while 1 signifies the best case scenario representing optimal recovery of product and by-products.

Influencing factors	Qualitative levels of action	WMI[a]	Rank[b]
Prevention/and or reduction of products	Effective		
and byproducts dispersion.	Fair	100,451	2
	Low		
Institutionalization of process procedures	High		
for waste dispersion elimination/	Moderate	100,355	3
and dispersion.	Low		
Percentage of waste streams segregation.	High (65-100%)		
	Moderate (40-65%)	75,292	4
	Low (below 40%)		
Frequency of spillages, leakages,	High (70-100%)		
incidentals and accidents during	Medium(30-70%)	50,313	6
the operations.	Low (under 30%)		
Levels of progressive prioritization of education on waste management in a winery at the following personnel:			
Senior management.	High		
	Medium		
	Low		
Skilled workers.	High		
	Medium		1[c]
	Low		
Unskilled workers.	High		
	Medium		
	Low		
Levels maintenance of facilities and	Routinely done		
equipment.	Often done	75,241	5
	irregulary/not done		
Time laspe between end of a process	Immediately		
or operation and commencement of	After sometime	25,155	7
products and byproducts recovery.	After long time		

Table 3. The rankings of the generic strategies influencing the recovery and handling of product and by-products for intrinsic wastes[1].

3.1.2 Effluent quantity (volume) sub-module

Fig. 7 depicts a hierarchical model for evaluating the effluent quantity generated during the cleaning and sanitization processes in a winery. The aggregated effluent quantity is a function of three linguistic variables, namely: the organic matter removal (OMR_s), equipment efficiency (EE_v), and effluent quantity (volume) management (M_v). The strategies influencing effluent management are summarised in Table 5. The effluent quantity management variable (M_v) is a function of generic effluent quantity management (M_{gv}) and the generic management of product and by-products (GM_{PBR}) variable at Level-II as depicted in Fig. 7.

[1] [a]WMI: waste minimization index discussed in section 2.2; [b]Ranking was used to facilitate the process of assigning dimensionless scores to aid in the evaluation of the degree to which product and by-products were recovered from a given process or unit operation. [c]The extent to which various strategies were effectively implemented in a winery for minimizing or eliminating waste as function of training and awareness of personnel at all levels. The training and education factor was ranked as the most significant in this category based on expertise knowledge.

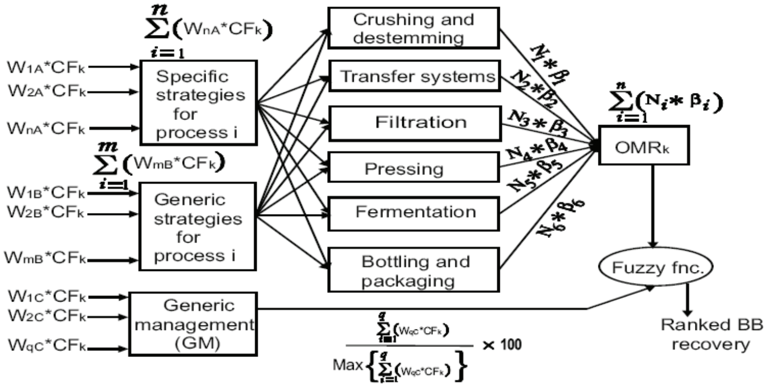


Fig. 6. Hierarchical structure for evaluating overall product and by-products recovery.

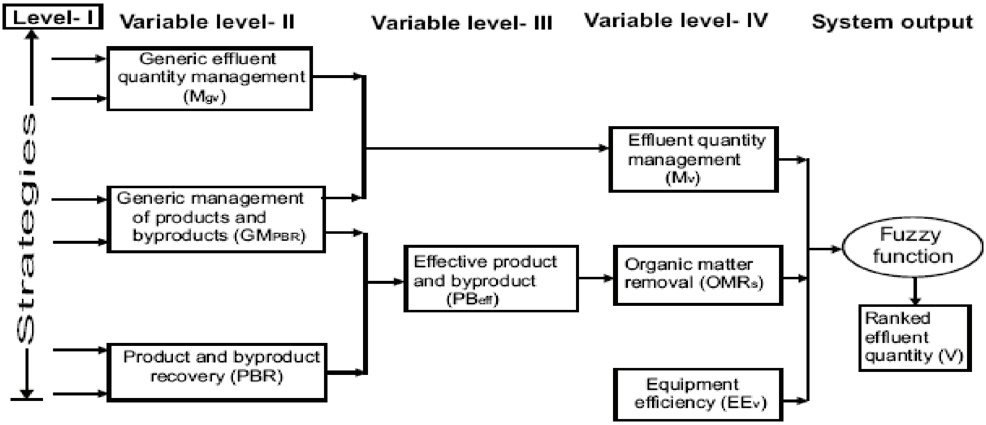
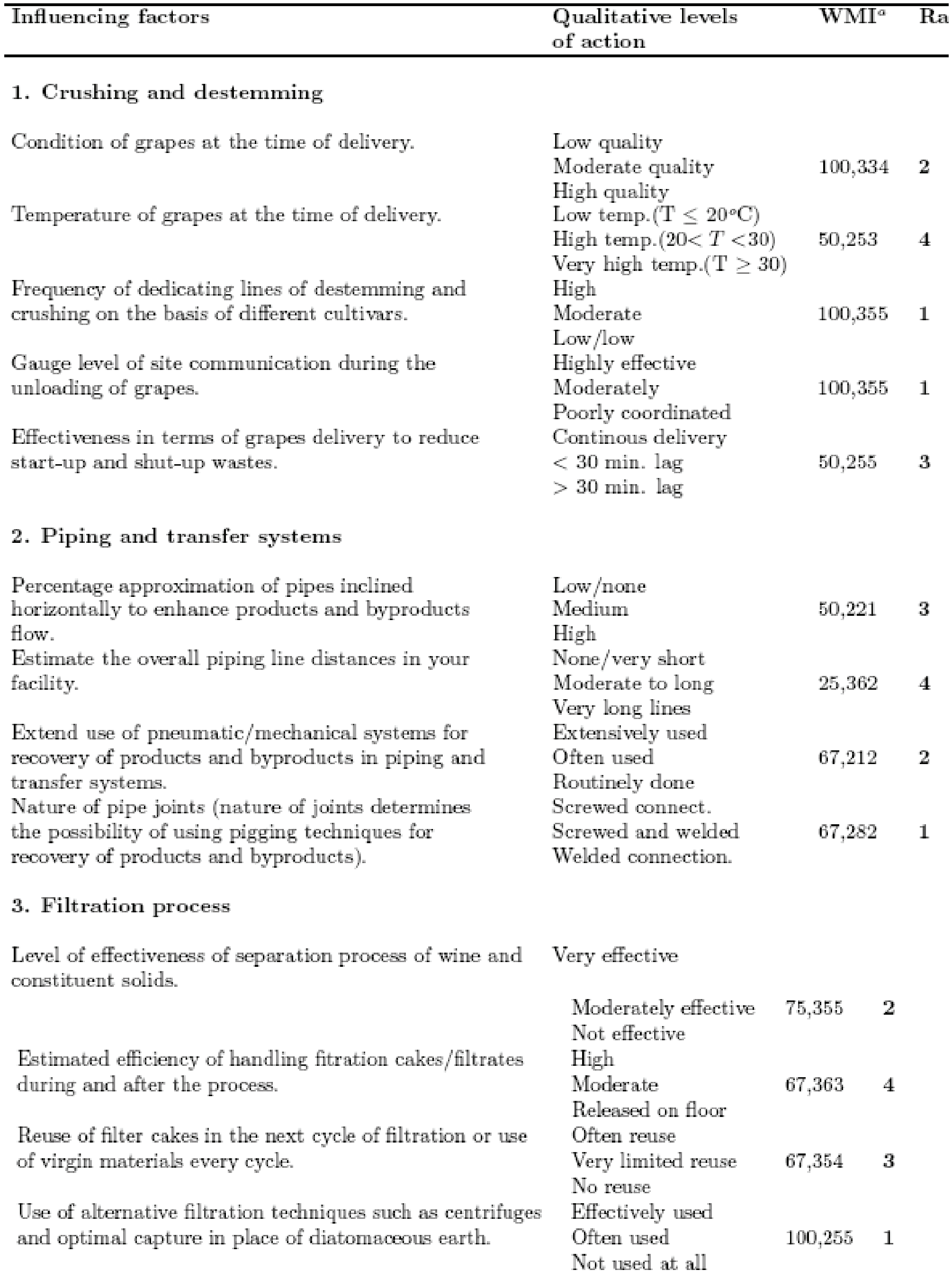

Level- I
Variable level- II
Variable level- III
Variable level- IV
System output
Strategies
Generic effluent quantity management (Mgv)
Generic management of products and byproducts (GMPBR)
Product and byproduct recovery (PBR)
Effective product and byproduct (PBeff)
Effluent quantity management (Mv)
Organic matter removal (OMRs)
Equipment efficiency (EEv)
Fuzzy function
Ranked effluent quantity (V)

Fig. 7. Hierarchical model structure to evaluate the effluent quantity during the cleaning and sanitization processes.

Note that the crisp numerical value for the generic effluent quantity management, (M_{gv}), is defined as:

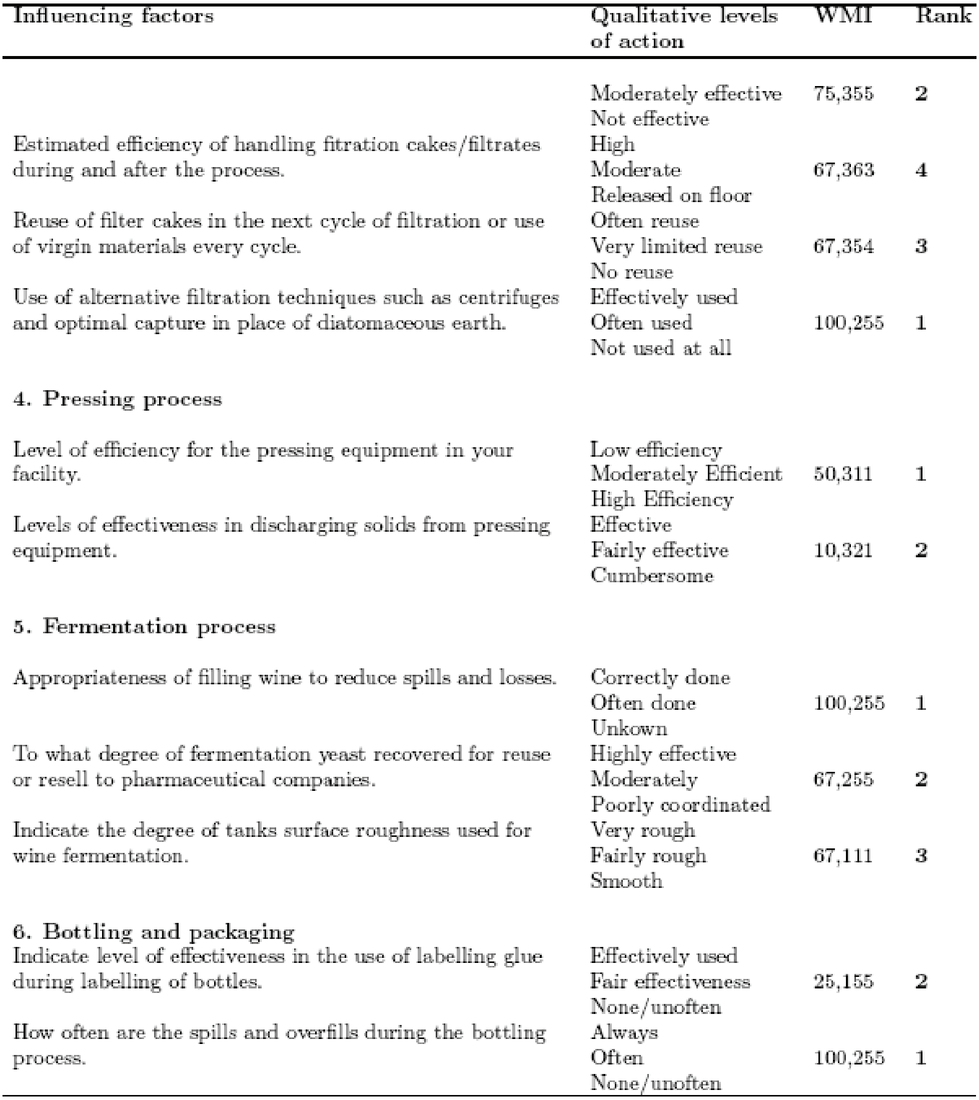
$$M_{gv} = \frac{\sum_{i=1}^{n}(W_{ns}^{'} \times CF_k)}{Max\left(\sum_{i=1}^{ns}(W_{ns}^{'} \times CF_k)\right)} \times 100 \tag{12}$$

M_{gv} is defined in the range 0 to 100. 0 implies the worst management scenario whereas 100 signifies the best effluent quantity management achievable in a given winery; W'_{ns} is a dimensionless score for the ith strategy, i= 1; 2; 3; … n in season s.
The generic management of product and by-products (GM_{PBR}) is computed using Eq. 10. Therefore, the effective effluent quantity management linguistic variable, M_v, is defined by the relation;



Influencing factors	Qualitative levels of action	WMI^a	Ra
1. Crushing and destemming			
Condition of grapes at the time of delivery.	Low quality Moderate quality High quality	100,334	2
Temperature of grapes at the time of delivery.	Low temp.(T $\leq$ 20°C) High temp.(20< T <30) Very high temp.(T $\geq$ 30)	50,253	4
Frequency of dedicating lines of destemming and crushing on the basis of different cultivars.	High Moderate Low/low	100,355	1
Gauge level of site communication during the unloading of grapes.	Highly effective Moderately Poorly coordinated	100,355	1
Effectiveness in terms of grapes delivery to reduce start-up and shut-up wastes.	Continous delivery < 30 min. lag > 30 min. lag	50,255	3
2. Piping and transfer systems			
Percentage approximation of pipes inclined horizontally to enhance products and byproducts flow.	Low/none Medium High	50,221	3
Estimate the overall piping line distances in your facility.	None/very short Moderate to long Very long lines	25,362	4
Extend use of pneumatic/mechanical systems for recovery of products and byproducts in piping and transfer systems.	Extensively used Often used Routinely done	67,212	2
Nature of pipe joints (nature of joints determines the possibility of using pigging techniques for recovery of products and byproducts).	Screwed connect. Screwed and welded Welded connection.	67,282	1
3. Filtration process			
Level of effectiveness of separation process of wine and constituent solids.	Very effective Moderately effective Not effective	75,355	2
Estimated efficiency of handling fitration cakes/filtrates during and after the process.	High Moderate Released on floor	67,363	4
Reuse of filter cakes in the next cycle of filtration or use of virgin materials every cycle.	Often reuse Very limited reuse No reuse	67,354	3
Use of alternative filtration techniques such as centrifuges and optimal capture in place of diatomaceous earth.	Effectively used Often used Not used at all	100,255	1

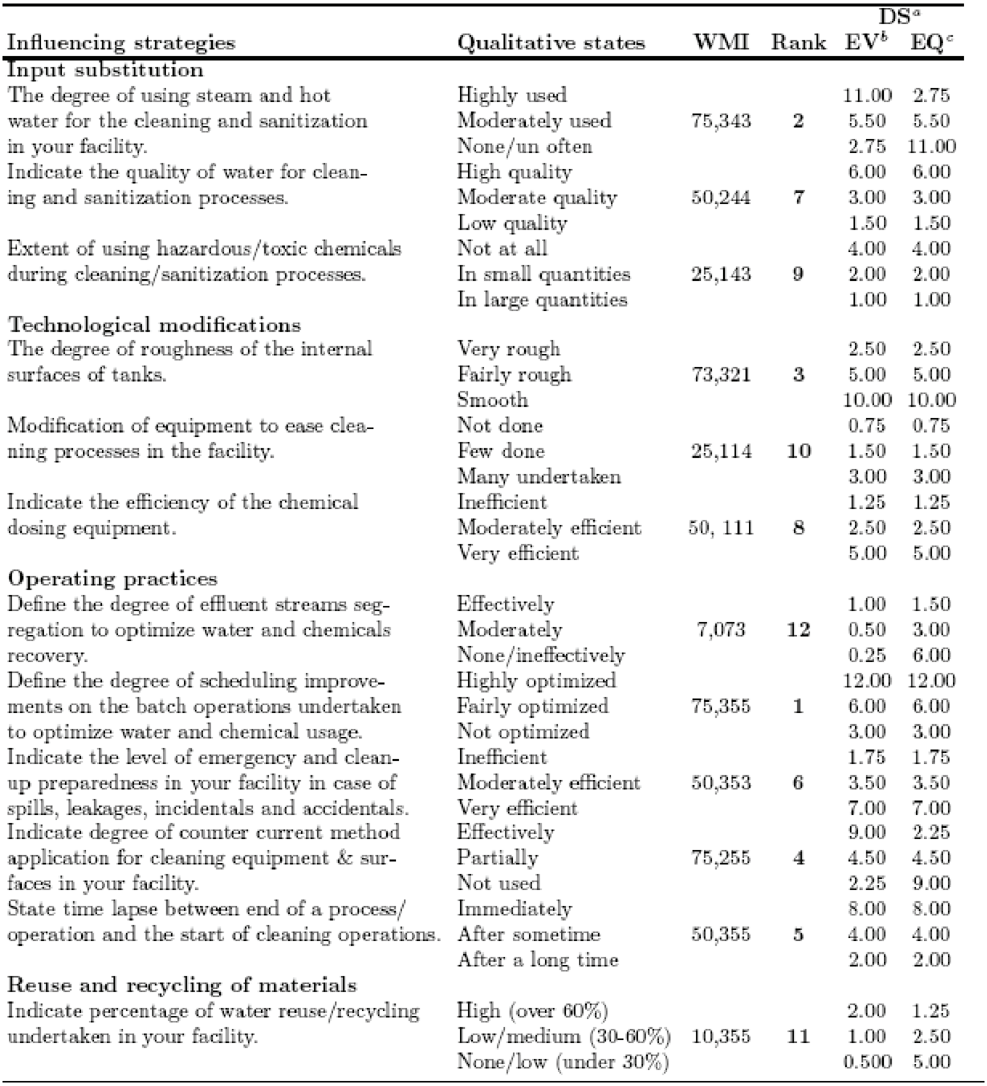
Table 4. continued …



Influencing factors	Qualitative levels of action	WMI	Rank
	Moderately effective Not effective	75,355	2
Estimated efficiency of handling fitration cakes/filtrates during and after the process.	High Moderate Released on floor	67,363	4
Reuse of filter cakes in the next cycle of filtration or use of virgin materials every cycle.	Often reuse Very limited reuse No reuse	67,354	3
Use of alternative filtration techniques such as centrifuges and optimal capture in place of diatomaceous earth.	Effectively used Often used Not used at all	100,255	1
4. Pressing process			
Level of efficiency for the pressing equipment in your facility.	Low efficiency Moderately Efficient High Efficiency	50,311	1
Levels of effectiveness in discharging solids from pressing equipment.	Effective Fairly effective Cumbersome	10,321	2
5. Fermentation process			
Appropriateness of filling wine to reduce spills and losses.	Correctly done Often done Unkown	100,255	1
To what degree of fermentation yeast recovered for reuse or resell to pharmaceutical companies.	Highly effective Moderately Poorly coordinated	67,255	2
Indicate the degree of tanks surface roughness used for wine fermentation.	Very rough Fairly rough Smooth	67,111	3
6. Bottling and packaging			
Indicate level of effectiveness in the use of labelling glue during labelling of bottles.	Effectively used Fair effectiveness None/unoften	25,155	2
How often are the spills and overfills during the bottling process.	Always Often None/unoften	100,255	1

Table 4. Rankings of the specific strategies influencing the recovery and handling of product and by-products of intrinsic wastes, under each process or unit operation[2].

[2] Symbols and ranking criteria is the same as in Table 3.



				DS[a]	
Influencing strategies	Qualitative states	WMI	Rank	EV[b]	EQ[c]
Input substitution					
The degree of using steam and hot water for the cleaning and sanitization in your facility.	Highly used			11.00	2.75
	Moderately used	75,343	2	5.50	5.50
	None/un often			2.75	11.00
Indicate the quality of water for cleaning and sanitization processes.	High quality			6.00	6.00
	Moderate quality	50,244	7	3.00	3.00
	Low quality			1.50	1.50
Extent of using hazardous/toxic chemicals during cleaning/sanitization processes.	Not at all			4.00	4.00
	In small quantities	25,143	9	2.00	2.00
	In large quantities			1.00	1.00
Technological modifications					
The degree of roughness of the internal surfaces of tanks.	Very rough			2.50	2.50
	Fairly rough	73,321	3	5.00	5.00
	Smooth			10.00	10.00
Modification of equipment to ease cleaning processes in the facility.	Not done			0.75	0.75
	Few done	25,114	10	1.50	1.50
	Many undertaken			3.00	3.00
Indicate the efficiency of the chemical dosing equipment.	Inefficient			1.25	1.25
	Moderately efficient	50, 111	8	2.50	2.50
	Very efficient			5.00	5.00
Operating practices					
Define the degree of effluent streams segregation to optimize water and chemicals recovery.	Effectively			1.00	1.50
	Moderately	7,073	12	0.50	3.00
	None/ineffectively			0.25	6.00
Define the degree of scheduling improvements on the batch operations undertaken to optimize water and chemical usage.	Highly optimized			12.00	12.00
	Fairly optimized	75,355	1	6.00	6.00
	Not optimized			3.00	3.00
Indicate the level of emergency and clean-up preparedness in your facility in case of spills, leakages, incidentals and accidentals.	Inefficient			1.75	1.75
	Moderately efficient	50,353	6	3.50	3.50
	Very efficient			7.00	7.00
Indicate degree of counter current method application for cleaning equipment & surfaces in your facility.	Effectively			9.00	2.25
	Partially	75,255	4	4.50	4.50
	Not used			2.25	9.00
State time lapse between end of a process/ operation and the start of cleaning operations.	Immediately			8.00	8.00
	After sometime	50,355	5	4.00	4.00
	After a long time			2.00	2.00
Reuse and recycling of materials					
Indicate percentage of water reuse/recycling undertaken in your facility.	High (over 60%)			2.00	1.25
	Low/medium (30-60%)	10,355	11	1.00	2.50
	None/low (under 30%)			0.500	5.00

Table 5. The rankings and assignment of dimensionless scores for the generic strategies that influences the effluent quality and quantity during vinification[3].

[3] [a]DS: Dimensionless scores; [b]EV: Effluent quantity; [c]EQ: Effluent quality

$$M_v = \left(\frac{M_{gv} + GM_{PBR}}{2} \right) \tag{13}$$

where M_v is defined in the range 0 to 100, where 0 implies the worst effluent quantity management in a winery, while 100 signifies excellent effluent quantity management.

In this study, what is regarded as organic matter removal (OMR_s) in winery operations is generically referred herein as product and by-products recovery, and has been discussed in section 3.1.1. Therefore, taking OMR_s in variable Level-IV (see Fig. 7) to be equivalent to PBR_{eff} in variable Level-III, then the crisp numerical input value for this variable was computed by multiplying Eq. 11 by 10 to express the OMR_S values in the range of 0 to 10. The expression for evaluating OMR_S is:

$$\mathrm{OMR_s} = \mathrm{PBR_{eff}} \times 10 \tag{14}$$

At Variable Level-IV, the last important variable also influencing the final effluent quantify is the efficiency of the equipment used for cleaning and sanitization purposes. Therefore, the equipment efficiency was used as a measure of the water, or steam quantity delivered from the equipment per unit time. The cleaning equipment with high efficiency (e.g. high pressure cleaners) had the merit of reducing the effluent quantity generated per unit time.

However, if the organic matter were present on the surfaces and equipment being cleaned, then the high efficiency of the cleaning equipment would have had a negative impact on the resultant effluent quality. This is because high concentrations of organic matter and chemicals in the effluent degrades its quality. On the other hand, if the equipment efficiency were low, there was likelihood of effluent quality being high, owing to dilution. These heuristics were used to model the equipment efficiency in relation to its potential impacts on the effluent quantity, as well as effluent quality. The crisp numerical input of the cleaning efficiency (EE_v) variable in Variable Level-IV was estimated using the heuristics mentioned above, and the fact that optimal efficiency of cleaning equipment ranges between 60 to 70%. Thus, EE_v was computed using the expression:

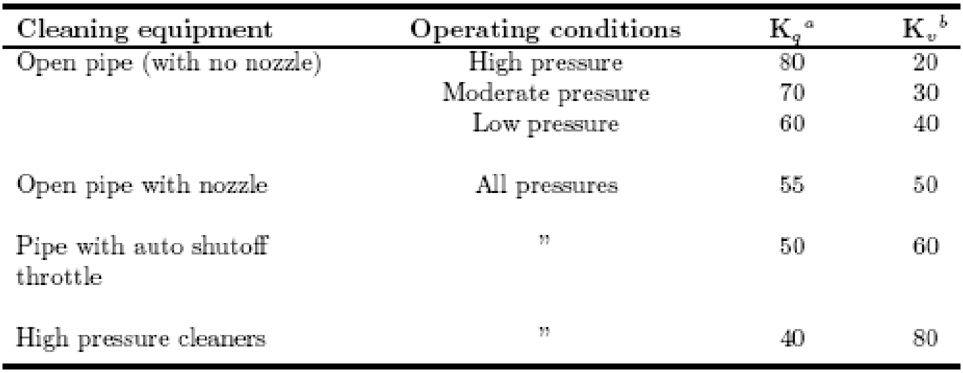
$$\mathrm{EE_v} = \mathrm{K_v} \ \mathrm{X} \ \mathrm{CF_R} \tag{15}$$

where K_v is a constant and a function of the cleaning equipment used with respect to its influence on the effluent quality. A summary of the K_v values is presented in Table 6. CF_R represents the degree of belief the user has on the cleaning equipment efficiency under use; R = 1, 2, … 6 with values fixed at 1.0; 0.9; 0.8; 0.7; 0.6; and 0.5, correspondingly. Note that EE_v ranged from 0 to 100, where 0 signified the lowest equipment efficiency 100 implied the highest efficiency.

Thus, the resultant effluent quantity generated (V) was evaluated using the results obtained from Eqs. 12-15 given by the expression:

$$\mathrm{V} = \mathrm{f}\left(\mathrm{M_v}, \mathrm{OMR_S}, \mathrm{EE_v}\right) \tag{16}$$

V was defined in the range 0 to 1, where 0 represented the best case scenario signifying prudent management of potable water through an integrated management of the influencing factors, whilst 1 implied the worst case scenario representing large effluent quantities of water generated during the cleaning and sanitizing processes. Similar equations were developed for the chemical usage and quality of effluent generated during the vinification process (Musee, 2004a).



Cleaning equipment	Operating conditions	K_q[a]	K_v[b]
Open pipe (with no nozzle)	High pressure	80	20
	Moderate pressure	70	30
	Low pressure	60	40
Open pipe with nozzle	All pressures	55	50
Pipe with auto shutoff throttle	"	50	60
High pressure cleaners	"	40	80

Table 6. K values for modelling the cleaning equipment efficiency effect on final effluent quality and effluent quantity[4].

3.1.3 System architecture

The architecture of the proposed knowledge-based decision support systems (KBDSS) is depicted in Fig. 8. The design of the architecture was based on two factors. Firstly, to ensure that it can handle the diverse, qualitative, and incomplete knowledge essential for decision making with respect to WM in the wine industry. And secondly, to guarantee system flexibility in terms of the ability to accommodate future expansions and incorporation of additional features, new tasks, new knowledge and information without restructuring and developing the entire system code from scratch. The former challenge was addressed by using a hybrid of expert systems and fuzzy logic, while the latter was achieved by designing the system using a modular approach. The system structure comprised of four components, namely; the knowledge base (rule base and data base), graphical user interface (GUI), fuzzy inference engine, and the knowledge acquisition and maintenance module. Fig. 8 illustrates the structure and information flow in the fuzzy logic expert system from a top level modular approach.

The knowledge base consisted of the rule base and the data base. The data base management module entailed managing and combining different kinds of knowledge stored in the system in form of generic knowledge, specific knowledge, mathematical models, or in the if-then rule format. The mathematical models module was used to compute the crisp numerical values which served as decision input variables into the fuzzy inference mechanism module. Easy access and interactions between the data base, rule base and the inference engine facilitated the identification of a solution dependent on user inputs, as well as preloaded knowledge stored in various rule bases. In this case, the fuzzy rule base was designed to evaluate different system outputs classified as functional groups, viz. effluent quality, effluent quantity, chemical consumption, and effectiveness of handling intrinsic oriented wastes. The if-then rules under each module were associated with effluent quality (48), effluent quantity (48), chemical consumption (36), and product and by-products recovery (20).

[4] [a]K_q: The constant used in computing the efficiency of the cleaning equipment when considering the effluent quality; [b]K_v: The constant used in computing the efficiency of the cleaning equipment when considering the effluent volume.

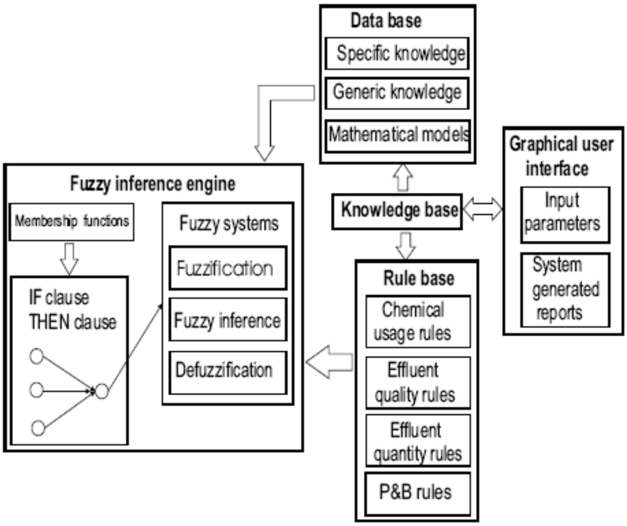


Fig. 8. The fuzzy logic expert system functional architecture for evaluating WM in the wine industry.

The GUI system provided seamless interaction of various components between the user and the data, information as well as knowledge stored in different data bases and rule bases. The GUI provided a convenient environment for data entry, specific module evaluation, overall system evaluation, and display of system results. This aided free interaction of the user and system through "buttons" that made it easy to understand and transparently illustrate how the system results were evaluated at various stages, before the final findings were displayed on the screen.

4. Results and discussions

In this section, the functionality of the proposed decision support system is illustrated by use of two case studies, viz.; the product and by-products recovery and the effluent quantity system outputs. Under each system output, the decision support system's ability to diagnose WM in the wine industry will be demonstrated by examining the final specific outputs under different user defined operational conditions. The derivation of the system results was executed in accordance to the defined procedure described in section 2.4.3.

4.1 Evaluation of product and by-products recovery

Generally there is no unique solution in terms of product and by-products recovery during vinification, owing to the complexity of interactions among different processes and human actions, as well as spatial and temporal factors. Nonetheless, the actions of operators and WM strategies adopted by the management were ranked as the most influential factors on the overall product and by-products recovery. The evaluation of product and by-products recovery required 29 inputs distributed as follows: Four generic, 20 specific and five management-related strategies (Tables 3 and 4). The system inputs for four different runs based on the user's responses, as well as degrees of confidence on each response are summarized in Table 7. Also, the screening and ranking index was applied in each set of specific and generic strategies to ensure that the dimensionless scores assigned reflected the relative importance of each strategy towards the recovery of product and by-products in a given process or unit operation (Tables 3 and 4).

To demonstrate how the proposed algorithm was applied in diagnosing WM in terms of product and by-products recovery, consider the inputs presented in Table 7. Under each run, the contribution of each process or unit operation for the product and by-products recovery was computed and normalized on a scale of 0 to 1 using Eq. 9. For instance, during the pressing process the inputs for the generic and specific strategies yielded scores of 10.500 and 7.125, respectively, under Run 1. The computed values were then normalized to determine the effective contribution of the pressing process to the overall recovery of product and by products, and yielded a value of 0.5508 (using Eq. 9).

The procedure was repeated in the rest of the processes, and the results are summarized in Table 8. To determine the overall weighted recovery of product and by-products in the vintage season (PBR^v), the computed values from each process (derived using Eq. 9) were multiplied with the experts' weightings using Eq. 7. For Run 1, the computation yielded a value of 0.5603. On the other hand, the generic waste management linguistic variable (GM) for handling product and by-products recovery was evaluated using Eq. 10. The GM value based on Run 1 user inputs yielded an aggregated value of 45.35%. Consequently, under Run 1, the crisp input values to the fuzzy model (Fig. 6) for the PBR^v and GM were 0.5603 and 45.35%, respectively. These crisp inputs were evaluated in a fuzzy model (using Eq. 11) in a rule base of 20 if-then rules and yielded a fuzzy effective product and by-products recovery (PBR^v_{eff}) of 0.345, which was linguistically ranked as Low.

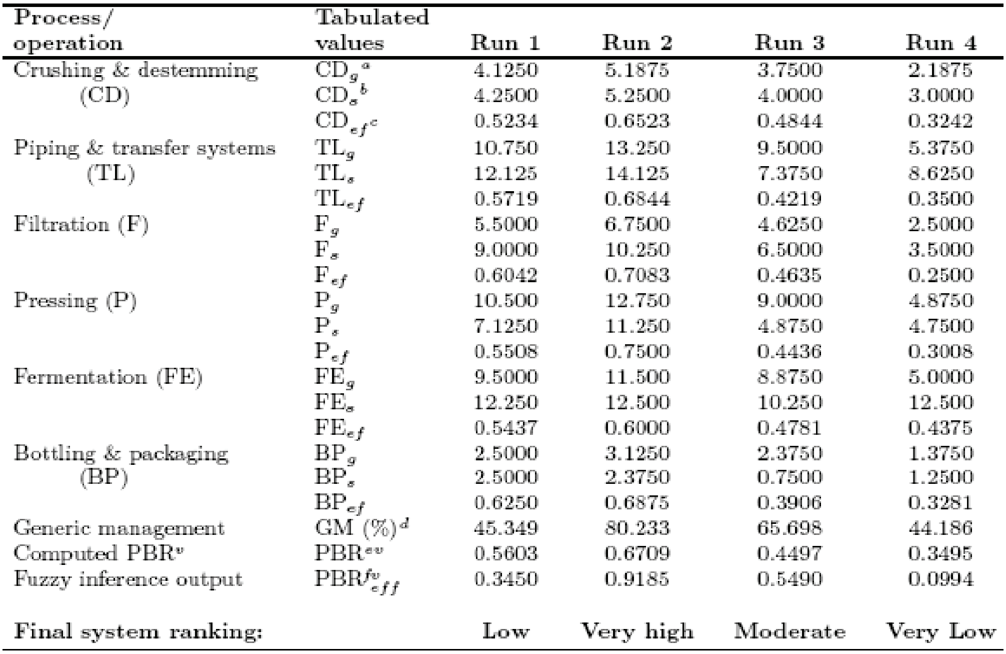
The product and by-products recovery ranked as Low under Run 1 implies large losses of wine and other potential feedstock materials for recovering by-products – eventually ending up into the wastewater streams. Consequently, the effluent quality and effluent quantity are likely to be impacted negatively. For instance, the wine and its associated solid waste streams – reduce the quality of the effluent – as evidenced by high chemical oxygen demand and suspended solids in various wastewater streams. Also, it may lead to high use of cleaning water to remove wine residues and other solids from equipment surfaces. A similar methodology was applied in Runs 2 to 4 and the results are presented in Table 8 based on the specified users' inputs presented in Table 7. Under Run 2, the crisp inputs computed for the GM and PBR^V linguistic variables were 0.67 (on a scale of 0–1) and 80% (on a scale of 0–100), respectively. These crisp input values were fed into the fuzzy model, resulting in the GM and PBR^V variables being linguistically labelled as Good and Very High, respectively, according to the rule base shown in Fig. 4. The defuzzified modular numerical output for the effective PBR^V_{eff} was 0.910, and linguistically labelled as Very High. This signifies the merit of adopting an integrated approach in implementing WM strategies during wine production.

By comparing the simulation results in Run 1 and Run 3, one would expect a higher recovery for the product and by-products in Run 1. However, this is not the case, because experience has shown that wineries where the management is rated as Good (as in Run 3) there is a higher possibility of improving product and by-products recovery than in situations where it is ranked Poor (as in Run 1). Thus, the fuzzy rules were designed to reflect this practical reality, which may otherwise be impossible to account for using data driven deterministic approaches. In addition, this also illustrates the necessity for training of personnel and ensuring good operating practices in order to enhance effective WM in the

	User input factor choices:[a]			
Qualitative factors	**Run 1**	**Run 2**	**Run 3**	**Run 4**
Generic Equip. factors				
Dispersion reduction	M(1.00)	H(0.75)	L(1.00)	L(0.50)
Enforcement of procedures	H(0.75)	H(1.00)	M(0.75)	L(1.00)
Streams segregation	M(1.00)	L(0.50)	H(1.00)	M(0.75)
Spills, leaks, etc frequency	L(1.00)	M(1.00)	M(1.00)	M(1.00)
Specific factors				
Crushers and destemmers				
Grapes condition	M(0.75)	M(1.00)	L(1.00)	M(1.00)
Grapes temperature	H(1.00)	M(0.5)	M(1.00)	H(0.50)
Lines dedication	M(1.00)	H(1.00)	M(0.75)	M(0.75)
Site communications	M(0.75)	H(0.75)	H(1.00)	L(1.00)
Grapes deliveries	H(0.75)	M(1.00)	M(0.50)	L(1.00)
Pipe and transfer systems				
% of inclined pipes	H(1.00)	M(1.00)	L(0.50)	M(0.75)
Overall piping lengths	M(0.75)	M(0.75)	H(0.50)	M(0.50)
Mechanical recoveries	H(0.75)	H(1.00)	M(1.00)	L(1.00)
Type of pipe joints	L(1.00)	H(0.75)	M(0.75)	H(0.75)
Filtration process				
Effectiveness of separation	M(0.75)	M(1.00)	M(0.50)	L(0.50)
Handling eff. of filtration cakes	H(0.50)	H(0.75)	L(1.00)	L(1.00)
Reuse of filter cakes	L(1.00)	M(0.75)	H(1.00)	M(1.00)
Alt. filtration methods	H(1.00)	H(1.00)	H(0.50)	M(0.50)
Pressing Process				
Pressing equip. efficiency	H(0.75)	H(1.00)	M(0.75)	M(1.00)
Eff. of solids discharge	M(0.75)	H(0.75)	M(0.75)	L(1.00)
Fermentation process				
Wine filling in fermentors	M(0.50)	H(0.75)	M(1.00)	H(1.00)
Yeast recovery from fermentors	H(1.00)	M(0.50)	H(0.50)	M(0.75)
Roughness of tank surfaces	M(0.75)	H(0.75)	M(0.75)	L(1.00)
Bottling and packaging				
Effi. of using glue	H(0.50)	M(0.75)	M(0.50)	L(1.00)
Reduction of spills	H(1.00)	H(1.00)	L(1.00)	H(0.50)
Generic Man. factors				
Training & awareness				
• SM[b]	M(0.50)	H(1.00)	H(0.75)	H(0.50)
• SW[c]	H(0.75)	H(0.75)	H(1.00)	M(1.00)
• UW[d]	M(0.75)	M(1.00)	L(1.00)	M(0.75)
Equipment maintenance	H(0.50)	H(0.75)	M(1.00)	L(0.50)
Time lag period	M(1.00)	H(0.75)	M(0.75)	H(0.50)

Table 7. User's inputs for evaluating product and by-products recovery during the vintage season[5].

[5] [a]All the user inputs are ranked as High (H), Medium (M), and Low (L) for simplicity purposes; [b]Senior management; [c]Skilled workers; [d]Unskilled workers.



Process/ operation	Tabulated values	Run 1	Run 2	Run 3	Run 4
Crushing & destemming (CD)	CD_g[a]	4.1250	5.1875	3.7500	2.1875
	CD_s[b]	4.2500	5.2500	4.0000	3.0000
	CD_{ef}[c]	0.5234	0.6523	0.4844	0.3242
Piping & transfer systems (TL)	TL_g	10.750	13.250	9.5000	5.3750
	TL_s	12.125	14.125	7.3750	8.6250
	TL_{ef}	0.5719	0.6844	0.4219	0.3500
Filtration (F)	F_g	5.5000	6.7500	4.6250	2.5000
	F_s	9.0000	10.250	6.5000	3.5000
	F_{ef}	0.6042	0.7083	0.4635	0.2500
Pressing (P)	P_g	10.500	12.750	9.0000	4.8750
	P_s	7.1250	11.250	4.8750	4.7500
	P_{ef}	0.5508	0.7500	0.4436	0.3008
Fermentation (FE)	FE_g	9.5000	11.500	8.8750	5.0000
	FE_s	12.250	12.500	10.250	12.500
	FE_{ef}	0.5437	0.6000	0.4781	0.4375
Bottling & packaging (BP)	BP_g	2.5000	3.1250	2.3750	1.3750
	BP_s	2.5000	2.3750	0.7500	1.2500
	BP_{ef}	0.6250	0.6875	0.3906	0.3281
Generic management	GM (%)[d]	45.349	80.233	65.698	44.186
Computed PBR^v	PBR^{v}[e]	0.5603	0.6709	0.4497	0.3495
Fuzzy inference output	PBR^{v}_{eff}[f]	0.3450	0.9185	0.5490	0.0994
Final system ranking:		Low	Very high	Moderate	Very Low

Table 8. System outputs for product and by-products recovery based on inputs in Table 7[6].

wine industry. Under Run 4, the system ranking indicates a Very Low recovery of product and by-products owing to Very Low (0.35) aggregated value for PBR^V and Fair (44%) rated generic management. To improve the performance of such a winery, it would be necessary to adopt an integrated approach in which diverse WM strategies – are implemented simultaneously and optimally.

4.2 Evaluation of effluent quantity

The second case study illustrates the suitability of the qualitative-quantitative model proposed in this paper to address WM challenges in the wine industry focusing on the effluent quantity generated during vinification processes. The user inputs and the system aggregated outputs are presented in Table 9. The linguistic inputs of the fuzzy model were the effluent quantity management (M_v), organic matter removal (OMR_s), and equipment efficiency (EE_v) – computed using Eqs. 13, 14, and 15, respectively. Nine runs were executed to illustrate the model suitability to assess WM challenges regarding effluent generation. Runs 1, 4, and 8 show ineffective implementation of WM as evidenced by the ranking of

[6] [a]g: denotes the generic factors' contribution.; [b]s: denotes the specific factors' contribution; [c]ef: denotes ; the total effective contribution of a given process or unit operation; [d]GM: generic management variable.
[e]PBR^v: organic matter handling during vintage season; [f]$\mathrm{PBR}^v_{\mathrm{eff}}$: effective organic matter handling during the vintage season ranking computed using the fuzzy if-then rules.

	User input factor choices[a]								
Qualitative factors	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8	Run 9
Input substitution (IS)									
Hot water/steam usage	M(1.00)	H(0.75)	M(0.75)	M(0.50)	M(0.75)	M(1.00)	M(0.50)	L(0.50)	L(0.50)
Quality of water used	H(0.75)	H(1.00)	H(0.75)	L(0.50)	H(0.75)	H(0.75)	L(0.50)	M(0.75)	M(0.75)
Hazardous and toxic chems.	M(0.50)	H(0.75)	H(1.00)	L(1.00)	H(1.00)	M(0.50)	L(1.00)	H(1.00)	H(1.00)
Technological mod. (TM)									
Vessels surface roughness	M(0.50)	H(1.00)	H(0.75)	H(0.50)	H(0.75)	M(0.50)	H(0.50)	M(0.75)	M(0.75)
equipment modifications	L(1.00)	H(0.75)	M(0.75)	L(1.00)	M(0.75)	L(1.00)	L(1.00)	H(1.00)	H(1.00)
Eff. of chem. dosing equips.	M(1.00)	H(1.00)	M(0.75)	L(1.00)	M(0.75)	M(1.00)	L(1.00)	L(1.00)	L(1.00)
Operating practices (OP)									
Degree of effluent segregation	H(0.75)	H(1.00)	H(0.75)	M(0.50)	H(0.75)	H(0.75)	M(0.50)	M(0.75)	M(0.75)
Scheduling of batch processes	L(0.75)	H(0.75)	H(0.75)	H(0.75)	H(0.75)	L(0.75)	H(0.75)	M(1.00)	M(1.00)
Emergency preparedness	M(0.75)	H(0.75)	M(1.00)	L(1.00)	M(1.00)	M(0.75)	L(1.00)	M(0.75)	M(0.75)
Counter current method appl.	H(0.50)	H(1.00)	M(0.50)	M(0.50)	M(0.50)	H(0.50)	M(0.50)	L(1.00)	L(1.00)
Time lag period	M(1.00)	H(0.75)	M(0.75)	H(0.50)	M(0.75)	M(1.00)	H(0.50)	M(0.75)	M(0.75)
Reuse and recycling (RR)									
% of water reuse/recycling	M(0.50)	H(1.00)	H(1.00)	H(0.50)	H(1.00)	M(0.50)	H(0.50)	L(1.00)	L(1.00)
IS_{ef}	0.1410	0.2212	0.1619	0.0577	0.1619	0.1410	0.0577	0.0978	0.0978
TM_{ef}	0.0737	0.2212	0.1346	0.0897	0.1346	0.0737	0.0897	0.1026	0.1026
OP_{ef}	0.1811	0.3878	0.2372	0.2212	0.2372	0.1811	0.2212	0.1811	0.1181
RR_{ef}	0.0064	0.0256	0.0256	0.0128	0.0256	0.0064	0.0128	0.0064	0.0064
$\mathrm{M}_{gv}(0-1range)$	0.4022	0.8558	0.5593	0.3814	0.5593	0.4022	0.3814	0.3894	0.3894
GM (0-1 range)	0.4535	0.8023	0.6570	0.4420	0.8023	0.6570	0.8023	0.4535	0.8023
M_{v}	42.787	82.905	60.815	41.164	68.080	52.963	59.185	42.145	59.587
OMR_{ve}	3.4500	9.1850	5.4900	0.9940	9.1850	5.4900	9.1850	3.4500	9.1850
EE_{v}	40.000	72.000	42.000	21.000	72.000	72.000	64.000	35.000	72.000
EV	0.7979	0.0937	0.4330	0.9170	0.0991	0.2370	0.2004	0.8161	0.1030
Effluent quantity ranking	**H[b] vol.**	**VL[c] vol.**	**M[d] vol.**	**VH[e] vol.**	**VL vol.**	**L[f] vol.**	**L vol.**	**H vol.**	**VL vol.**

Table 9. User inputs and WM analysis results for effluent quantity during the vinification under the vintage season.

the effluent quantity as High or Very High. Although the inputs in each run reflected different operational conditions, the aggregated values for the final effluent quantity management ($\mathrm{M_v}$) ranged between 41% and 43%, and linguistically ranked as Poor.

In addition, the cleaning equipment efficiency ($\mathrm{EE_v}$) was ranked linguistically Very Low as the values for Runs 1, 4, and 8 were 40%, 21%, and 35%, respectively. This mostly likely indicates that open pipes (without nozzles) were used for cleaning purposes (see Table 6 for low $\mathrm{K_v}$ values). Besides, the organic matter removal from equipment and surfaces before cleaning and sanitization started was ranked Poor by the fuzzy model as evidenced by the final computed values listed in Table 9. Therefore, one way of minimizing the quantity of effluent generated under such operational scenarios represented by Runs 1, 4, and 8 is to target low cost strategies, particularly focussing on the removal of organic matter and fixing nozzles to the cleaning pipes in order to improve the cleaning equipment efficiency. Secondly, the suggested options are easy to implement and operate without need for costly training of personnel.

To illustrate the viability of the proposed alternatives using low cost alternatives, consider the simulation findings of Runs 1 and 6. The $\mathrm{M_{gv}}$ was left unchanged at 0.4022 as technological modifications and finding of suitable alternative input substitutes were viewed as costly, not easy to implement, and requiring long payback periods. However, the GM variable, which is easy to implement in a winery was varied from 0.454 in Run 1 to 0.657 in Run 6, and consequently changed the $\mathrm{M_v}$ from 43% to 53%. Concurrently, the $\mathrm{OMR_{ve}}$ and $\mathrm{EE_v}$ fuzzy model inputs were varied from 3.45 to 5.49, and 40% to 72% (through replacing open pipes with high pressure cleaner), respectively. The fuzzy model in Run 6 (using Eq. 16) yielded an effluent quantity linguistically labelled as Low.

Therefore, easily implemented, low cost alternatives have been demonstrated to reduce the quantity of effluent during the vinification processes. Similarly, the same procedure was applied to illustrate how high quantities of effluent can be reduced in cases such as Runs 3 and 5; Runs 4 and 7, as well as Runs 8 and 9. The breadth of the solutions provided by the proposed framework indicates its versatility and robustness in addressing WM related to effluent quantity. Note that Run 2 was presented as a base case illustrating an ideal winery, where various WM strategies were adequately implemented, and therefore no further action was necessary. Therefore, besides the ranking of the effluent quantity or other system outputs, such as chemical usage and effluent quality, the proposed framework also offers a suitable diagnostic tool for the wine industry to identify areas where they can improve their overall waste management performance.

5. Conclusions

In summary, the proposed knowledge-based decision support system provides a systematic approach for evaluating and diagnosing the unstructured WM problem in the wine industry by way of processing the user inputs (both qualitative and quantitative) to compute a given winery performance in terms of effluent quantity, product and by-products recovery, chemical usage, or effluent quantity. The system results are in a format that can be easily read, understood, or altered by the user. This is because the final system outputs are expressed in the form of performance indexes (range [0, 1]), and therefore, the proposed decision model offers a transparent and robust tool for assessing the performance of a given winery – with respect to WM. Secondly, the system incorporated data, information and knowledge sourced from experts that can aid in facilitating efficient decision-making regarding WM in the wine industry. Thirdly, as the wine industry has dearth of statistical data regarding WM unlike the chemical industry, fuzzy logic and qualitative reasoning soft computing approaches were applied to aid in evaluating WM in this industry.

Presently, there is no evaluation-framework tool that can assess WM performance of a given winery through integration of both qualitative and quantitative data. Thus, the integrated methodology proposed in this paper serves a suitable tool to achieve this objective as well as to aid in automating WM analysis in the wine industry. Consequently, based on the integrated framework presented, the WM can be evaluated and ranked at different levels of aggregation – clearly identifying areas that may need improvement to optimise resources utilization and reduce operational costs. And finally, the system has the merit of reducing the time, effort, and resources required in undertaking extensive WM in the wine industry. The suitability of the approach has been demonstrated through two worked case studies, each with several different functional scenarios.

6. Acknowledgements

The financial support of the NRF, University of Stellenbosch, Wintech (doctoral grants), and THRIP are acknowledged.

7. References

Allen, D.T. & Rosselot, K.S. (1997). *Pollution prevention for chemical processes*. Wiley & Sons, New York, ISBN 978-0-4711-1587-8.

Balik, J.M. & Koraido, S.M. (1991). Identifying pollution prevention options for a petroleum refinery. *Pollution and Prevention Review*, 1, pp. 273-293.

Bisson, L.F.; Waterhouse, A.L.; Ebeler, S. E.; Walker, M.A. & Lapsley, J.T. (2002). The present and future of the international wine industry. *Nature*, 418, pp. 696–699.

Bonissone, P.P. (1997). Soft computing: the convergence of emerging reasoning technologies. *Soft Computing*, 1, pp. 6–18.

Bowbrow, D.G. (1984). Qualitative reasoning about physical systems: An introduction. *Artificial Intelligence*, 24, pp. 1–5.

Büyüközkan, G. & Feyzioğlu, O. (2004). A new approach based on soft computing to accelerate the selection of new product ideas. *Computer Industry*, 54(20), pp. 151–167.

Chapman, J.; Baker, P. & Wills, S. (2001). *Winery wastewater handbook: Production, impacts and management*, Adelaide, South Australia: Winetitles, ISBN 978-1-8751-3035-1.

Crittenden, B.D. & Kolaczkowski, S.T. (1995). *Waste minimization: a practical guide* Institution of Chemical Engineers, Rugby, U.K., ISBN 978-0-8529-5342-6.

Douglas, J. M. (1988). *Conceptual design of chemical processes*, McGraw-Hill New York, ISBN 978-0-0710-0195-3.

Gagliardi, F.; Roscia, M. & Lazaroiu, G. (2007). Evaluation of sustainability of a city through fuzzy logic. *Energy*, 32, pp. 795–802.

Gentile, M.; Rogers, W.J. & Mannan, M. S. (2003). Development of an inherent safety index based on fuzzy logic. *American Institution of Chemical Engineers Journal*, 49(4), pp. 959–968.

Halim, I. & Srinivasan, R. (2002). Systematic waste minimization in chemical processes 2: Intelligent decision support system. *Industrial Engineering and Chemistry Research*, 41, pp. 208–219.

Hanlon, D. & Fromm, C. (1990). *Waste minimization assessments*, H.M. Freeman (ed). Hazardous Waste Minimization, McGraw-Hill, Singapore, pp. 71–126, ISBN 978-0-0702-2043-0.

Hilson, G. (2003). Defining "cleaner production" and "pollution prevention" in the mining context. *Mineral Engineering*, 16, pp. 305–321.

Huang, Y.L. & Fan, L. T. (1995). *Intelligent process design and control for in-plant waste minimization*, A.L. Rossiter (ed.). Waste minimization through process design, McGraw-Hill, New York, pp. 165–180, ISBN 978-0-0705-3957-0

Katsiri, A. & Dalou, F. (1994). Wine and Distillery effluents in Greece: Main results of the SPRINT AQUANET program, In: Proceedings of international specialized Conference on Winery Wastewaters, 20-22June, Narbonne, France, pp.25–30.

Kleer, J.D. & Brown, J.S. (1984). A qualitative physics based on confluences. *Artificial Intelligence*, 24, pp. 7–83.

Lee, C.C. (1990). Fuzzy logic in control systems: fuzzy logic controller-Part I. *IEEE Transactions on Systems, Man, and Cybernetics*, 20, pp. 218–404.

Mamdani, E.H. & Assilian, S. (1999). An experiment in linguistic synthesis with a fuzzy logic controller, *International Journal of Humans-Computer Studies*, 51:135–147.

Massette, M. (1994). Wineries facing regulation, In: Proceedings of international specialized Conference on Winery Wastewaters, 20-22 June, Narbonne, France, pp.13-18.

Mazumoto, M. (1995). Improvement of fuzzy control, H. Li, M.M. Gupta (eds.). *Fuzzy Logic and Intelligent Systems*, Kluwer Academic Publishers, Norwell, Massachusetts, pp. 1–16, ISBN 978-0-7923-9575-1.

Mendel, J. (1995). Fuzzy logic systems for engineering: a tutorial. *Proceedings of IEEE*, 83(3), pp. 345–377.

Müller, A.M. (1999). Government of South Africa Gazette No. 20526.8, October, 1999. Government Notice, Department of Water Affairs and Forestry, Section 21(e) 1999.

Musee, N.; Lorenzen, L. & Aldrich, C. (2003). Integrated intelligent decision support system for waste minimization in wine making processes, South African Institute of Chemical Engineers (SAIChE) Congress, 3–5 September, Sun City: South Africa [full paper in CD-Rom], ISBN 0-958-46096-7, 10 pp.

Musee, N. (2004a). An integrated approach to waste and energy minimization in the wine industry: a knowledge-based decision methodology, PhD Dissertation (unpublished): Department of Process Engineering, University of Stellenbosch, South Africa

Musee, N.; Lorenzen, L. & Aldrich, C. (2004b). Integrated intelligent decision support system: is it a viable tool for WM in the wine industry? *Chemical Technology*, 2:5–8.

Musee, N.; Lorenzen, L. & Aldrich, C. (2005). A fuzzy-based heuristic methodology for waste minimization analysis: A case study in the wine industry, Proceedings in the 7th World Congress of Chemical Engineering, Glasgow, Scotland, ISBN 0-852-9549-48

Musee, N.; Lorenzen, L. & Aldrich, C. (2006a). Decision support for WM in wine-making processes. *Environmental Progress*, 25(1), pp. 56–63.

Musee, N.; Lorenzen, L. & Aldrich, C. (2006b). An aggregate fuzzy hazardous index for composite wastes. *Journal Hazardous Materials*, 137 (2), pp. 723–733.

Musee N, Lorenzen L, Aldrich C. Cellar WM in the wine industry: a systems approach. *Journal of Cleaner Production*, 15, pp. 417–431.

Musee, N. & Lorenzen, L. (2007). Assessing sustainability of gold mining operations using fuzzy logic methodology, J. Avraamides, G. Deschênes, D. Tucker (eds.). By- and Co-products and the environment, Australasian Institute of Mining and Metallurgy, Carlton Victoria, Australia, pp 261 – 263. ISBN 978-1-9208-0674-3.

Musee, N.; Lorenzen, L. & Aldrich, C. (2008a). New methodology for hazardous waste classification using fuzzy set theory: Part I. Knowledge acquisition. *Journal of Hazardous Materials*, 154 (1-3), pp. 1040-1051.

Musee, N; Aldrich, C. & Lorenzen L. (2008b). New methodology for hazardous waste classification using fuzzy set theory: Part II. Intelligent decision support system. *Journal of Hazardous Materials*, 157, pp. 94-105.

Ocampo-Duque, W.; Ferré-Huguet, N.; Domingo, J.L. & Schuhmacher, M. (2006). Assessing water quality in rivers with fuzzy inference systems: a case study. *Environmental International*, 32(6), pp. 733–742.

Phillis, Y.A. & Andriantiatsaholiniaina, L.A. (2001). Sustainability: an ill-defined concept and its assessment using fuzzy logic. *Ecological Economics*, 37(3), pp. 435-456.

Prato, T. (2007). Assessing ecosystem sustainability and management using fuzzy logic. *Ecological Economics*, 61, pp. 171–177.

Ross, T. J. (2004). Fuzzy logic with engineering applications. John Wiley & Sons, New York, ISBN 978-0-4708-6075-5

Setnes M, Babuska R, Vergruggen HB. Rule-based modeling: precision and transparency. IEEE Tran Syst Man Cyber Part C- Appl Rev 1998;28(1):165–169.

Smith, R.L. & Khan, J.A. (1995). *Unit operations database for transferring WM solutions*. A.P. Rossiter (ed.). Waste Minimization through Process Design. McGraw-Hill, New York, pp. 133-148, ISBN 978-0-0705-3957-0

USEPA, (1988). *Waste Minimization Opportunity Assessment Manual, Office of Research and Development.* United States Environmental Prevention Agency 1988, Cincinnati, OH. Publication No. EPA/625/7-88/003.

Yager, R.R. (1980). A general class of fuzzy consecutives. *Fuzzy Sets and Systems*, 4, pp. 103–111.

Yager, R.R. & Zadeh, L. A. (1992). An introduction to fuzzy logic applications in intelligent systems. Kluwer Academic Publishers, Boston, ISBN 978-0-7923-9191-3

Yager, R.R. & Filev, D.P. (1994). Essentials of fuzzy modelling and control. Wiley, New York, ISBN 978-0-4710-1761-5

Yen, J. & Lugari, R. (1998). *Fuzzy logic: Intelligence, control and information*. Prentice Hall, Upper Sandle River, New Jersey, ISBN 978-0-1352-5817-0

Zadeh, A.L. (1965). Fuzzy sets. *Information Control*, 8, pp. 338–353.

Zimmermann, H.-J. (1991). *Fuzzy Set Theory and its Applications* (second ed.). Kluwer Academic Publishers, Norwell, Massachusetts, ISBN 978-0-7923-9075-6

17

Prospects of Automation Agents in Agribusiness (Hop Industry) Decision Support Systems Related to Production, Marketing and Education

Martin Pavlovic[1] and Fotis Koumboulis[2]
[1]International Hop Growers' Convention,
[2]Halkis Institute of Technology,
[1]Slovenia
[2]Greece

1. Introduction

Difficulty for decision making in modern agribusiness has increased significantly, since it involves a large number of strongly interrelated factors that affect the satisfaction of the performance criteria describing product quality, production timing and cost. In addition, decision making has to take into account regulations and restrictions concerning the safety of the personnel, the environmental protection and the energy saving. Moreover, agriculture is becoming more commercialized, as farmers are competing with other farmers all over the world. To face these challenges modern agricultural practices must be adopted; however, they require appropriately educated and informed farmers (Abdon & Raab, 2004). So, the question is how to provide the required knowledge and information to farmers, even to those without a high education level.

A powerful tool to circumvent these difficulties is the technological area of agroinformatics and concerns the use of Information Management and Decision Support Systems (IMDSS). They aim to monitoring all functions of an agricultural process and facilitating decision making by proposing scenarios towards satisfying specific performance criteria and restrictions. IMDSS may perform several operations: monitoring the agricultural process, action planning and proposal of scenarios, processing of measurement data to extract information regarding the production cost and the product quality, fault diagnosis and alarm management.

Decision Support Systems (DSS) have been extensively used in industrial applications to support the human-supervisor decisions regarding assurance of efficient and safe processes operation (Lambert et al., 1999; Sanchez et al., 1996). The degree of automation in decision making is the major factor of differentiation between DSS. The DSS characterized by the lower degree of automation simply facilitate decision making by offering information to the operator; in an upper stage DSS incorporate decision-making units that simply propose actions without the jurisdiction of activation. DSS classified in the highest degree of automation completely replace - in certain activities - the human operator. The development of Automation Units for industrial decision making, being implemented as software agents, which may be incorporated in an abundance of commercial Supervisory Control and Data Acquisition (SCADA) products, has been proposed (Koumboulis & Tzamtzi, 2005). The proposed Automation Agents for Decision Support Systems (AADSS) may cover a wide range of industrial applications, providing decision support of the highest degree of automation.

IMDSS specifically oriented for agribusiness applications are met both in the international practice and literature (McCown, 2002; Parrott et al., 2003). Most of them are designed to serve specific sectors of agribusiness, like cotton management (McCown, 2002), vegetable processing industry (Berlo van, 1993), soybean management (Welch et al., 2002), wheat cultivar selection and fertilization (McCown, 2002), irrigation (Mira da Silva et al., 2001), etc. Moreover, in most cases the proposed DSS are designed to support only a limited range of the decisions to be taken by the farmer. For example, a DSS called PCYield is designed to answer the following two questions aiming to support decisions concerning soybean cultivar selection: What yield range might be expected? And, what happens if irrigation is withheld for a time and the weather is dry? (Welch et al., 2002). An interesting case is DSS supporting farm planning (Recio et al., 2003), which involves a significant range of farming decisions, like scheduling of field tasks, investment analysis, machinery selection and cost/benefit analysis.

Increasing the functionality of DSS with additional characteristics have been proposed, for the example the use of operational research and management's science tools in order to integrate weather forecasts in decision making (Recio et al., 2003). Of special interest is the development of DSS that incorporate tools from agribusiness logistics (Berlo van, 1993; Folinas et al., 2003; Biere, 2001).

Several approaches have been used for developing DSS in agribusiness, based either on operational research/management science, on heuristic search or other artificial intelligence techniques (Recio et al., 2003). Three of the most popular are linear programming approaches, dynamic programming approaches, as well as model-based simulation approaches; a fuzzy logic based approach has also been proposed (Thangavadivelu & Colvin, 1997) for scheduling tillage operations. Despite the research efforts for the development of DSS for agribusiness, farmers seem to be reluctant to involve DSS in their work; thus, the range of application of DSS in agriculture remains significantly smaller than in industry. According to McCown (2002), the two variables long recognized as key to user acceptance is perceived usefulness and ease of use.

The present work proposes stepwise development of Automation Agents for Decision Support Systems (AADSS) for agribusiness – applied in hop industry. Following Koumboulis and Tzamtzi (2005) for industrial applications, the automation agents will be implemented as software units, which may be incorporated in several commercial SCADA products. The proposed agents aim to support decision making in a significant range of the agribusiness operation, extending from cultivation techniques to farm planning and commerce of products, by undertaking to execute actions, like monitoring the agricultural process, providing e-learning functionalities, fault diagnosis, e-commerce support, planning based on logistics and processing of weather information. Thus the farmer is strongly supported with regard to all decision making that concerns the actions to be performed for all stages of agribusiness, from production to commerce. Moreover, the AADSS will provide a very friendly and easy to use graphical user interface, exploiting the graphical user interface capabilities of commercial SCADA products.

2. Automation agents to support agribusiness management

The proposed Automation Agents will be based on modern techniques of sustainable agriculture, so as to face the restrictions and difficulties of agricultural environment, such as high complexity, presence of uncertain and time changing factors, like weather, and restriction of natural sources, like water.

The proposed Automation Agents are the Operator Agent and the Supervisory Agents. The Supervisory Agents are the E-learning Agent, the Monitoring Agent, the Fault Diagnosis Agent, the Weather Information Agent, the Ecommerce Agent and the Logistics Agent. Their functionality aims to supervise all aspects of hop industry, from production to commerce, and moreover to provide e-learning services to farmers' group. It is important to note that the knowledge data base used for the E-learning Agent is dynamically adapted with any new information derived from processing the data gathered from the process. The Operator Agent exploits information gathered from all the aforementioned Automation Agents in order to support the farmer's decision making by proposing possible scenarios to the farmer. The interconnection between the Operator and the Supervisory Agents is illustrated in Figure 1.

Operator Agent

The Operator Agent will be implemented according to the DAI-DEPUR architecture, an integrated and distributed artificial intelligence supervisory architecture, proposed by Sanchez et al. (1996) for a waste-water treatment plant. The DAI-DEPUR architecture has also been used for the implementation of an Operator Agent for industrial applications (Koumboulis & Tzamtzi, 2005). Below we present in short the DAI-DEPUR architecture as described by Sanchez et al. (1996).

The DAI-DEPUR architecture comprises the data level, the distributed knowledge level, the reasoning level and the supervisory level.

The data level comprises the data collection system, that receives data from the process sensors, and the data base management system. The distributed knowledge level comprises distributed agents each of which processes validates and monitors the available information for a specific subsystem of the process, in order to describe the subsystem's behavior. Their conclusions are sent to the supervisory level in order to diagnose the whole plant state. The distributed knowledge level comprises simulation modules, as well as knowledge acquisition modules. The role of the knowledge acquisition modules is two-fold; first a conceptual clustering of data is performed, that leads to a representation of the process domain of operation in terms of classes, and then conjunctive classification rules are determined.

The reasoning level manages a Case Library that contains information about previously experienced situations, as well as solutions that have been followed in the past for each of these situations. Besides, the reasoning level may evaluate proposed solutions using simulation. The Case Library is dynamically enriched with information regarding the newly experienced situations.

The supervisory level gathers information from the distributed knowledge and the reasoning level, to determine the current status of process operation and coordinate the rest of Automation Agents. In our case the supervisory level is responsible for all decisions undertaken by the Operator Agent, that is:

1. Alarm management
2. Exchange of information between the Automation Agents
3. Configuration of the supervisory agents
4. Derivation and proposal of optimal, or at least suboptimal scenarios of actions to be undertaken by the farmer, based on specific performance requirements.

Due to the complexity of the agricultural processes, in conjunction with the large number of factors to be taken into account, it is a usual case in agribusiness to deal with competitive goals that is criteria that cannot be simultaneously achieved. For example, the improvement of the product quality usually increases the production cost; then scenarios proposed to the farmer should be based on a compromise between competitive design goals, performed to achieve the optimal result, regarding the economic issues of the process, like production cost and product quality, as well as energy or natural resources saving, while satisfying constraints imposed by environmental protection rules. This compromise may be formulated as an optimization under constraints problem (Koumboulis & Tzamtzi, 2005).

Supervisory Agents

Six Supervisory Agents are proposed based on the following approaches, selected to serve the needs of agribusiness:

1. E-Learning Agent
2. Monitoring Agent
3. Fault Diagnosis Agent
4. Weather Information Agent
5. E-commerce Agent
6. Logistics Agent

The **E-Learning Agent** aims to educate farmers with regard to required information and knowledge concerning the production techniques and the commerce of the specific product. The E-Learning Agent comprises a historic module implemented by a database, and a knowledge processing unit. The historic module contains required information background in order to make decisions concerning the agricultural process. The historic module stores measurements regarding the agricultural process and the weather, technical and scientific information concerning for example cultivation methods, plant diseases, special characteristics of each plant variety, effect of the weather on the plant, etc.; market information, like the available stock for each plant variety, current prices, current request from a national or a global market, etc., as well as legislation rules and restrictions. The knowledge processing unit is used in order to dynamically adapt the historic module based on the newly gathered information. The E-Learning Agent exploits also the information stored in the Case Library of the Operator Agent's reasoning level concerning previously experienced situations.

The **Monitoring Agent** aims to provide the farmer, as well as the other Agents, all the necessary information regarding the current status of the agricultural process. For this purpose, the Monitoring Agent exploits the data level of the Operator Agent that collects information from the agricultural process sensors, like measurements of humidity and temperature, height of the plant, etc. These data are processed in order to derive information about the current status of the agricultural process, as for example the plant growth. The information provided by the Monitoring Agent is also used by the distributed knowledge level of the Operator Agent.

The **Fault Diagnosis Agent** aims to a timely diagnosis of any fault that may occur in the agricultural process, like plant diseases, harm caused by insects, malfunctioning of the irrigation system, problems due to weather conditions, etc. Fault detection is achieved using the information provided by the Monitoring Agent and it is based on specific fault detection rules provided by the historic module of the E-Learning Agent. Whenever a specific situation deviates from the rules provided by the historic module of the E-Learning Agent, data are either processed by the knowledge process unit of the E-Learning Agent or they are sent to an expert, in order to derive new rules that will enrich the fault diagnosis capabilities of the Agent. The Fault Diagnosis Agent may also use the functionality of the distributed knowledge level of the Operator Agent.

The **Weather Information Agent** aims to support the decision making process regarding the issues that concern weather conditions. This Agent gathers information from meteorological services, combining it with measurements from the cultivation site. Based on this information it provides short term prediction of weather conditions. Moreover, the Weather Information Agent cooperates with the historic module of the E-Learning Agent in a two fold manner: first it exploits information from the historic module to support the weather prediction process; secondly, the information gathered by the Weather Information Agent is used to enrich the historic module.

The **E-Commerce Agent** aims to support the sales through internet of agricultural products and it can also support the farmer in buying row material through internet. The E-Commerce Agents comprises an intelligent unit to use the market information stored in the historic module, in order to decide whether each buying or selling action is profitable for the farmer.

The **Logistics Agent** aims to support the farmer concerning decisions on planning and controlling an efficient flow and storage of raw materials, intermediate and final products from point of origin to point of consumption, so as to achieve specific performance requirements, while simultaneously satisfying constraints (Berlo van, 1993; Folinas et al., 2003; Biere, 2001). The performance requirements concern the product cost and quality, the satisfaction of the market demand, the optimal exploitation of the processing equipment and cultivation area, etc. The restrictions concern the storage capacity, the maximum allowed time of storage, the available cultivation capacity and the processing capacity, etc. The planning provided by the Logistics Agent should take into account time changing and uncertain factors, like the weather conditions and the demand of the market. The Logistics Agent may consider the whole logistical chain incorporating agriculture, processing industry and market, as was proposed in (Berlo van, 1993); in this case, the DSS should communicate with corresponding information systems of the processing industry and market sectors, in order to derive the data required.

3. Embedding automation agents in SCADA systems

The implementation of Automation Agents should satisfy certain functional specifications, like interactive communication, real time processing and precision of calculations. To achieve these goals, the software development of the Automation Agents is based, following Koumboulis and Tzamtzi (2005), on international standards for open architecture, which will assure the embedment of the software in several commercial SCADA products. These systems are modern Information Systems with build-in capabilities of network communication that have widely been used in modern industry. The subsystems of a SCADA system are:

1. Bilateral communication with the process: data acquisition from sensors and command transmission though appropriate actuators,
2. (Graphical User Interface (GUI), through which the operator monitors and commands the process and
3. Automation Agents for Decision Support System.

The AADSS functions, provided by most of the commercial SCADA products, are: storage of info to historical modules, statistical data processing, formulation of reports, and alarm management. The Automation Agents presented in the previous sections will utilize the network communication, collection, registration, depiction and data process capabilities provided by modern SCADA. For example, the data level of the DAI-DEPUR architecture will be implemented by the data processing unit of the SCADA system, while the alarm management operation performed by the supervisory level of the DAI-DEPUR architecture will be supported by the alarm handling unit of the SCADA system. Besides communication of Automation Agents with the human-operator will take place via the graphical user interface of SCADA system. The embedment of the Automation Agents in SCADA systems

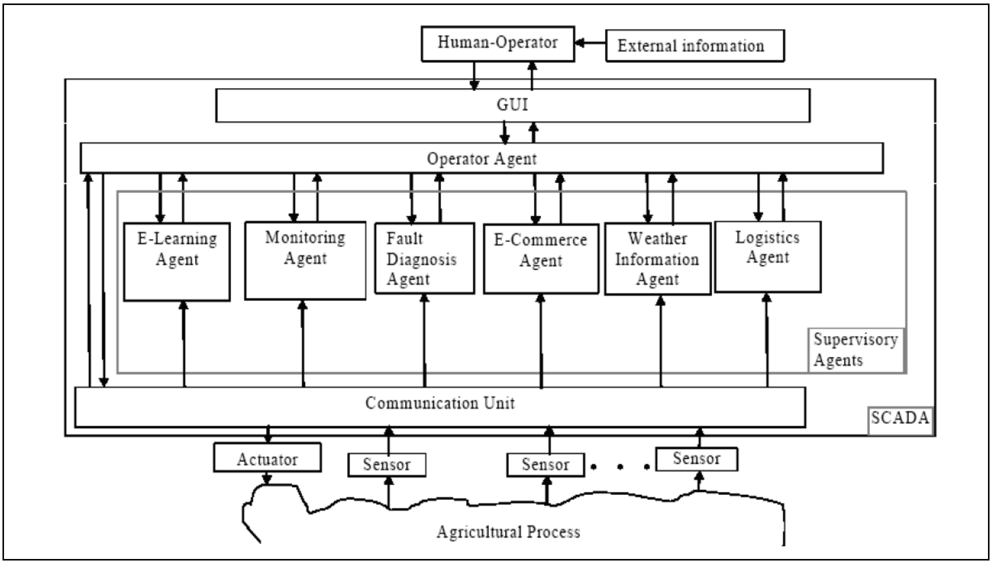


Fig. 1. Embedment of Supervisory and Automation Agents for Agribusiness in SCADA systems

is achieved using the latest technology of modern software tools, supported by most of modern commercial SCADA programs, like the OPC (OLE for Process Control) that allows bilateral real time data transfer between different types of equipments, as well as established standards of object-oriented communication between heterogeneous applications of Windows operating systems, like COM and ActiveX. The embedment of Automation Agents into SCADA systems is presented (Figure 1).

4. Model application in a hop industry

An implementation of the proposed scheme is planed to take place within a framework of a research project for a case study in a hop industry that is related predominantly to a brewing industry. Beer brewing is an intricate process encompassing mixing and further elaboration of four essential raw materials, including barley malt, brewing water, hops and yeast. Particularly hops determine to a great extent typical beer quality such as bitter taste, hoppy flavour, and foam stability. Adding different hop varieties to the wort kettle produces the typical beer aromas.

The benefits and opportunities for a hop industry to establish its informative and useful information management system is mainly due to the following reasons:

1. Hops are a classic internationally agricultural traded commodity, one of the few internationally traded goods bought and sold on world markets without any major economic restrictions, that is on the real basis of supply and demand. Thus, to remain competitive in global hop industry hop producers must respond to the changing needs of the brewing community by providing quality raw material from appropriate hop varieties.
2. The demands on production techniques, varieties, quality and preparation of export producing hops is changing constantly (Pavlovic et al., 2003; Pavlovic & Koumboulis, 2004). For the sake of international character of hops as well as stakeholders' initiatives,

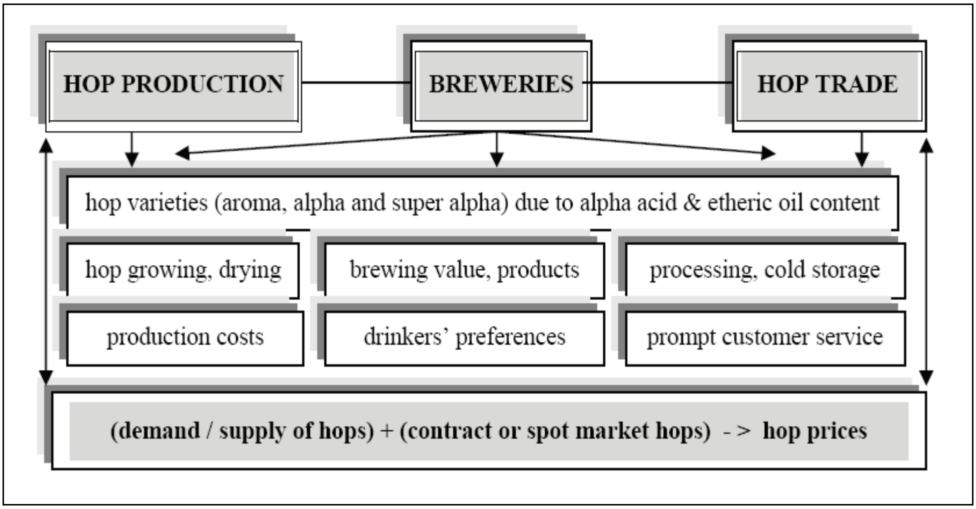


Fig. 2. Quality and business circle in a hop industry

the International Hop Growers' Convention (IHGC) with 33 companies and organisations from 20 countries is supporting activities to improve the information management for a benefit of its members (Figure 2).

Furthermore, also benefits related to raw material quality management for the brewing industry are expected. The four largest breweries such as Anheuser-Busch InBev (Belgium/USA) with its share of world beer production of 22,2%; SABMiller (UK) - 11.0%; Heineken (NL) - 9.0%; Carlsberg (DK) - 6.1%; produced in 2007 around 50% of the world beer quantity. However, there are also many small scale breweries that may benefit from the AADSS model information results.

Regardless the size of a brewing company, the product principles as well as the objectives regarding technology and purchasing are more or less the same: to use raw materials as economically as possible in the brewing process (Pavlovic & Koumboulis, 2004).

5. Conclusions

The proposed Automation Agents are planed to be applied in a decision support system for a hop industry. However, they represent generalised tools applicable to some extent also to other agribusiness sectors. The proposed AADSS is particularly suited for large agricultural production units, since for small units the decision making process is significantly easier. Of special interest are farms with several plots of different crops, where the global farm resources have to be appropriately distributed among the plots, as well as farms simultaneously coordinated, as is the case of agricultural cooperatives.

The proposed Agents may contribute to farmers' education, to early correct fault diagnosis and incorporation of modern agricultural techniques, to process historical data taking into account several factors in order to support decision making, and to support the farmer in e-commerce activities. The incorporation of Automation Agents in DSS will contribute considerably in the decongestion of the responsibilities of the farmer. Thus farmers are assisted to focus on higher level operations like production planning and commerce. On the other hand, production costs are decreased, as the performance criteria can be achieved in a more precise manner and with less human effort. Finally, the environmental protection may benefit significantly since the proposed agents will contribute to a better exploitation of natural sources, as well as energy saving. The implementation of the Automation Agents is based on open architectures, so as to give the capability of integration in a crowd of commercial SCADA systems, fact that simplifies significantly the development and implementation procedure of the Automation Agents. What is more, the graphical user interface of SCADA systems provides a user friendly environment that will increase the farmers' acceptance.

6. References

Abdon, B. R. & Raab, R. T. (2004). Knowledge sharing and distance learning for sustainable agriculture in the Asia Pacific Region: The role of Internet. New directions in a diverse planet. Proc. of the 4th Int. Crop Science Congress, Brisbane, Australia, Sept. 26. – 10. Oct. 2004. http://www.cropscience.org.au/icsc2004/

Berlo van, J.M. (1993). A decision support tool for the vegetable processing industry; an integrative approach of market, industry and agriculture'. Agricultural Systems 23: 91-109.

Biere, A.W. (2001). Agribusiness logistics: An emerging field in agribusiness education'. *In*: IAMA World Food and Agribusiness Symposium. Sydney, Australia, 27. June, 2001. http://www.ifama.org/conferences/2001Conference/Papers/Area%20I/Biere_Arlo.PDF

Folinas, D.; Vlachopoulou, M.; Manthou, V. & Manos, B. (2003). A web-based integration of data and processes in the agribusiness supply chain. EFITA 2003 Conference. Debrecen, Hungary: 143-149.

Koumboulis, F. N. & Tzamtzi, M. P. (2005). Automation Agents Embedded in Industrial Decision Support Systems. *In*: International Conference on Intelligent Agents, Web Technologies and Internet Commerce – IAWTIC'2005. Vienna, Austria. pp: 51-57.

Lambert, M.; B. Riera & Martel, G. (1999). Application of functional analysis techniques to supervisory systems. Reliability Eng. & System Safety 64: 209-224.

McCown, R. L. (2002). Changing systems for supporting farmers decisions: Problems, paradigms, and prospects. Agric. Systems 74: 179-220.

Mira da Silva, L.; Park, J. R.; Keatinge, J. D. H. & Pinto, P. A. (2001). A decision support system to improve planning and management in large irrigation schemes. Agric. Water Manag. 51: 187-201.

Parrott, L.; Lacroix, R. & Wade, K. M. (2003). Design considerations for the implementation of multi-agent systems in the dairy industry. Computers and Electronics in Agric. 38: 79-98.

Pavlovic, M.; Koumboulis, F. N.; Tzamtzi, M. P. & Karras, D. A. (2003). Model of information management and data exchange on a global hop supply. Proc. of the 2nd Int. Symp. on Intelligent Information Technology in Agriculture. Beijing, China, pp. 20-23.

Pavlovic, M. & Koumboulis, F. M. (2004). Methodology of an IHGC Market Supply Data Collation. (www.ihgc.org). Hop Bull. 11: 17-24.

Pavlovic, M.; Luo, X.; Kosir, I. J.; Virant, M. & Gu, F. (2006). Expansion of the Chinese hop and brewing industry. Hop Bull. 13: 61-69.

Pavlovic, M.; Koumboulis, F. N.; Tzamtzi, M. P. & Rozman, C. (2008). Role of automation agents in agribusiness decision support systems. Agrociencia 42: 913-923.

Recio, B.; Rubio, F. & Criado, J. A. (2003). A decision support system for farm planning using AgriSupport II. Decision Support Systems 36: 189-203.

Sanchez, M.; Cortes, U.; Lafuente, J.; Roda, I. R. & Poch, M. (1996). DAIDEPUR: An integrated and distributed architecture for waste-water treatments plants. Artificial Intelligence in Eng. 1: 275-285.

Thangavadivelu, S. & Colvin, T. S. (1997). Fuzzy-Logic-Based decision support system for scheduling tillage operations. Eng. Applications of Artificial Intelligence 10: 463-472.

Welch, S. M.; Jones, J. W.; Brennan, M. W.; Reeder G. & Jacobson, B. M. (2002). PCYield: Model-based decision support for soybean production. Agric. Systems 74: 79-98.

18

Automatically Building Diagnostic Bayesian Networks from On-line Data Sources and the SMILE Web-based Interface

Anucha Tungkasthan, Nipat Jongsawat, Pittaya Poompuang,
Sarayut Intarasema and Wichian Premchaiswadi
Graduate School of Information Technology in Business, Siam University
Thailand

1. Introduction

One of the most difficult obstacles in the practical application of probabilistic methods is the effort that is required for model building and, in particular, for quantifying graphical models with numerical probabilities. The construction of Bayesian Networks (BNs) with the help of human experts is a difficult and time consuming task, which is prone to errors and omissions especially when the problems are very complicated or there are numerous variables involved. Learning the structure of a BN model and causal relations from a dataset or database is important for extensive BNs analysis.

In general, the causal structure and the numerical parameters of a BN can be obtained using two distinct approaches. First, they can be obtained from an expert. Second, they can also be learned from a data set. The main drawback of the first approach is that sometimes there is not enough causal knowledge to establish the structure of the network model with certainty and estimation of probabilities required for a typical application is a time-consuming task because of the number of parameters required (typically hundreds or even thousands of values). Thus, the second approach can initially help human experts build a BN model and they can make it applicable at a later time. In practice, some combination of these two approaches is typically used. This paper essentially focuses on using the second approach.

This paper presents a practical framework for automating the building of diagnostic BN models from data sources obtained from the WWW and demonstrates the use of a SMILE web-based interface to represent them. This work proposes the following components: 1) an RSS agent that automatically gathers RSS feeds from diverse data sources in the WWW environment, 2) a transformation/conversion tool that transforms and converts the collected data for both continuous and discrete valued data sets 3) a reasoning engine that has the ability to learn and build the causal structure for BN models from data and provide functionality to perform a diagnosis, 4) the visualization of BN models on a website, and 5) a diagnosis of the BN model and the resulting reports. This article is organized as follows: Section 2 presents a little more detail about the basic concepts of Bayesian networks and tools. Section 3 addresses related work. Section 4 describes the design and implementation of a practical framework for automating the building of diagnostic BN models from online data sources in more detail. Section 5 presents a conclusion and discusses some perspectives and ideas for future work. An acknowledgement is described in section 6.

2. Fundamentals

This section is intended to describe the fundamentals of Bayesian networks and the core reasoning engines of SMILE web-based interface development. They are described in the following sections.

2.1 Bayesian network and Bayesian updating

Bayesian networks (also called belief networks, Bayesian belief networks, causal probabilistic networks, or causal networks) (Pearl, 1988) are acyclic directed graphs in which nodes represent random variables and arcs represent direct probabilistic dependencies among them. The structure of a Bayesian network is a graphical, qualitative illustration of the interactions among the set of variables that it models. The structure of the directed graph can mimic the causal structure of the modeled domain, although this is not necessary. When the structure is causal, it provides a useful, modular insight into the interactions among the variables and allows for a prediction of the effects of external manipulation. Nodes of a Bayesian network are usually drawn as circles or ovals. A Bayesian network also represents the quantitative relationships among the modeled variables. Numerically, it represents the joint probability distribution among them. This distribution is described efficiently by exploring the probabilistic independence among the modeled variables. Each node is described by a probability distribution conditional on its direct predecessors. Nodes with no predecessors are described by prior probability distributions. Both the structure and the numerical parameters of a Bayesian network can be elicited from an expert. They can also be derived from data, as the structure of a Bayesian network is simply a representation of interdependencies in the data and the numbers are a representation of the joint probability distributions that can be inferred from the data. Finally, both the structure and the numerical probabilities can be a mixture of expert knowledge, measurements and objective frequency data.

Bayesian updating, also referred to as belief updating, or somewhat less precisely as probabilistic inference is based on the numerical parameters captured in the model (Cooper, 1990). The structure of the model which is an explicit statement of the interdependencies in the domain helps in making the algorithms for Bayesian updating more efficient (Dagum & Luby, 1997). All algorithms for Bayesian updating are based on a theorem proposed by Rev. Thomas Bayes (1702-1761) and are known as Bayes Theorem. Belief updating in Bayesian networks is computationally complex. In the worst case, belief updating algorithms are NP-hard (Cooper 1990). There exist several efficient algorithms, however, that make belief updating in graphs consisting of tens or hundreds of variables tractable. Pearl developed a message-passing scheme that updates the probability distributions for each node in a Bayesian network in response to observations of one or more variables (Pearl, 1986). Lauritzen and Spiegelhalter, Jensen et al, and Dawid proposed an efficient algorithm that first transforms a Bayesian network into a tree where each node in the tree corresponds to a subset of variables in the original graph (Lauritzen & Spiegelhalter, 1988: Jensen et al., 1990: Dawid, 1992). The algorithm then exploits several mathematical properties of this tree to perform probabilistic inference. Several approximate algorithms based on stochastic sampling have been developed. Of these, best known are probabilistic logic sampling (Henrion, 1988), likelihood sampling (Shachter & Peot, 1989: Fung & Chang, 1989), and backward sampling (Fung & del Favero, 1994), Adaptive Importance Sampling (AISBN) (Cheng & Druzdzel, 2000), and Approximate Posterior Importance Sampling (APIS-BN) (Yuan & Druzdzel, 2003). Approximate belief updating in Bayesian networks has also been shown to be worst case NP-hard (Dagum & Luby, 1993).

2.2 The core reasoning engines

The core reasoning engines of the web-based interface development capability consist of SMILE (Structural Modeling, Inference, and Learning Engine), SMILEarn, and JSMILE. SMILE is a reasoning engine that is used for graphical probabilistic models and provides functionality to perform diagnosis. SMILEarn is used for obtaining data from a data source, preprocessing the data, and learning the causal structure of BN models. JSMILE is used for accessing the SMILE library from the web-based interface. This section provides some more detailed information about SMILE, SMILEarn and JSMILE wrapper.

SMILE (Structural Modeling, Inference, and Learning Engine) is a fully platform independent library of functions implementing graphical probabilistic and decision-theoretic models, such as Bayesian networks, influence diagrams (IDs), and structural equation models (Druzdzel, 1999). Its individual functions, defined in the SMILE Application Programmer Interface (API), allow creating, editing, saving, and loading graphical models, and using them for probabilistic reasoning and decision making under uncertainty. SMILE can be embedded in programs that use graphical probabilistic models as their reasoning engines. Models developed in SMILE can be equipped with a user interface that best suits the user of the resulting application. SMILE is written in C++ in a platform-independent manner and is fully portable. Model building and the reasoning process are under full control of the application program as the SMILE library serves merely as a set of tools and structures that facilitates them.

SMILEarn extends the functionality provided by SMILE. It provides a set of specialized classes that implement learning algorithms and other useful tools for automatically building graphical models from data. It is a C++ library that contains a set of data structures, classes, and functions that implement learning algorithms for graphical models and includes other functionality (such as data access, storage and preprocessing) that can be used in a model in conjunction with SMILE. Although SMILEarn is a module of SMILE, which means that it requires SMILE to be used, but one can use SMILE without the need to install and use SMILEarn.

JSMILE is a library of java classes for reasoning about graphical probabilistic models, such as Bayesian networks and influence diagrams. It can be embedded in programs that use graphical probabilistic models as a reasoning engine. It is a wrapper library that enables access to the SMILE and SMILEXML C++ libraries from java applications. JSMILE is not limited to stand-alone applications. It can also be used on the back-end side of a multi-tiered application.

3. Related works

There are various kinds of software applications that can be used to create decision theoretic models, learn the causal structure, and perform diagnosis based on BNs and IDs. There are both commercial and non-commercial software applications available. The commercial software applications are widely used in a business environment. Many of them are integrated into business analysis software and used particularly for solving difficult business problems. The non-commercial software applications are extensively used for the educational purposes. This article reviews only the most relevant subset of non-commercial software applications based on BNs and IDs.

B-Course is an analysis tool that was developed in the fields of Bayesian and causal modelling (Mylltmaki et al., 2002). It is a free web-based online data analysis tool, which allows users to analyze data for multivariate probabilistic dependencies. It also offers facilities for inferring certain type of causal dependencies from the data. B-Course is used via a web-browser, and requires the user's data to be a text file with data presented in a tabular format typical for any statistical package (e.g., SPSS, Excel text format). It offers a simple three step procedure (data upload, model search, and analysis of the model) for building a BN dependency model. After searching the model, B-Course provides the best model to the user via a report. Users can continue to search for the next best model but they must make the decision for selecting the best model that fits their needs. Selecting the best model is sometimes very difficult for inexperienced users. In B-Course, there are no structural learning algorithms provided for the user to aid in selection. The analysis method, modelling assumptions, restrictions, model search algorithms, and parameter settings are totally transparent to the user.

Elvira is a tool for building and evaluating graphical probabilistic models (Lacave et al., 2007). It is a non web-based application. It is implemented in Java, so that it can run on different platforms. It contains a graphical interface for editing networks, with specific options for canonical models (e.g., OR, AND, MAX, etc.), exact and approximate algorithms for discrete and continuous variables, explanation facilities, learning methods for building networks from databases, algorithms for fusing networks, etc. Elvira is structured as four main modules: (1) data representation- containing the definition of the data structures that are needed for managing BNs and IDs in Java, (2) data acquisition- including the classes that are necessary for saving and loading a network from either a file or a database, (3) processing - implementing the algorithms for processing and evaluating models, and (4) visualization - defining the Elvira graphical user interface (GUI) which obviously makes use of the classes that are included in the previous modules.

GeNIe (Graphical Network Interface) is a versatile and user friendly development environment for building graphical decision models (Druzdzel, 1999). The original interface was designed for SMILE which is described in a previous section. GeNIe may be seen as an outer shell to SMILE. GeNIe is implemented in Visual C++ and draws heavily on the Microsoft foundation classes. GeNIe provides numerous tools for users such as an interface to build Bayesian network models or influence diagrams, to learn the causal relationships of a model using various algorithms, and to perform model diagnosis. In order to use GeNIe efficiently, the GeNIe software must be installed and the user should have some background knowledge about probabilistic graphical models and become familiar with the tools provided in GeNIe.

Poompuang, et al presents a development environment for building graphical decision-theoretic models based on Bayesian networks and influence diagrams working on the website by utilizing an original engine called "SMILE" (Poompuang, et al., 2007). They propose the idea of building and developing graphical decision-theoretic models on a web page in order to overcome such the limitation of Bayesian belief network software developed on a windows-based platform, which makes the models not easily portable and is limited in its graphical representation across multiple system platforms. They present a prototype of Bayesian network models and influence diagrams in a World Wide Web environment, which can be displayed by a standard web browser.

Tungkasthan, et al presents a visualization of BN and influence Diagram models on a website (Tungkasthan et al., 2008). They develop an application based on the Macromedia Flash and Flash Remoting technologies. The application model on the client side is constructed by using the Macromedia Flash and the connection between a client and web server is developed by using the Flash Remoting technology. They use the capability of Marcomedia Flash and Flash Remoting technology to build richer, more interactive, more efficient, and more intuitive user interfaces for their applications than are possible with other web technologies such as JSP and Java applets. Their applications also provide a powerful, intuitive drag-and-drop graphical authoring tool that is comfortable for the users and have quick-loading and dynamic interfaces.

Jongsawat, et al presents a technique to dynamically feed data into a diagnostic Bayesian network model and a web-based user interface for the models (Jongsawat et al., 2008). In their work, the BN model (the students' attitude towards several factors in a college enrolment decision) is fixed and the data obtained from an online questionnaire are saved into a database and transferred to the model. The user can observe the changes in the probability values and the impact the changes have on each node in real-time after clicking on a belief update button. Users can also perform Bayesian inference in the model and they can compute the impact by observing values of a subset of the model variables on the probability distribution over the remaining variables based on real-time data.

Jongsawat, et al presents a SMILE web-based interface that permits users to build a Bayesian network causal structure from a dataset or database and perform Bayesian network diagnosis through the web (Jongsawat & Premchaiswadi, 2009). There are several learning algorithms such as Greedy Thick Thinning, PC, Essential Graph Search, and Naive Bayes provided for the user. The user can just select the desired learning algorithm and adjust its parameter settings to learn the model structure. After building the BN structure, the user is able to quantify uncertain interactions among random variables by setting observations (evidence) and use this quantification to determine the impact of the observations. The SMILE web-based interface was developed based on SMILE, SMILEarn, and SMILE.NET. It uses a novel, user-friendly interface which interweaves the steps in the BN analysis with brief support instructions on the web page.

4. Design and implementation

The following steps in this section describe how a practical framework and SMILE web-based interface are designed and implemented for automating the building of diagnostic BN models from online data sources. The structure of the proposed framework is presented in Fig. 1.

In an article by the U.S. News & World Report's "World's Best Universities rankings based on the Times Higher Education-QS World University Rankings in 2009" was selected to be the case study and the source of information for the BN model construction. The top 400 world's best universities were reported. There are six categories in each rank to be scored and reported on the web site. They consist of the following items: Academic Peer Review, Employer Review, Student to Faculty, International Faculty, International Students, and Citations per Faculty. In the data preparation process, we built RSS feeds from these online sources. The sample of RSS feeds is described below.

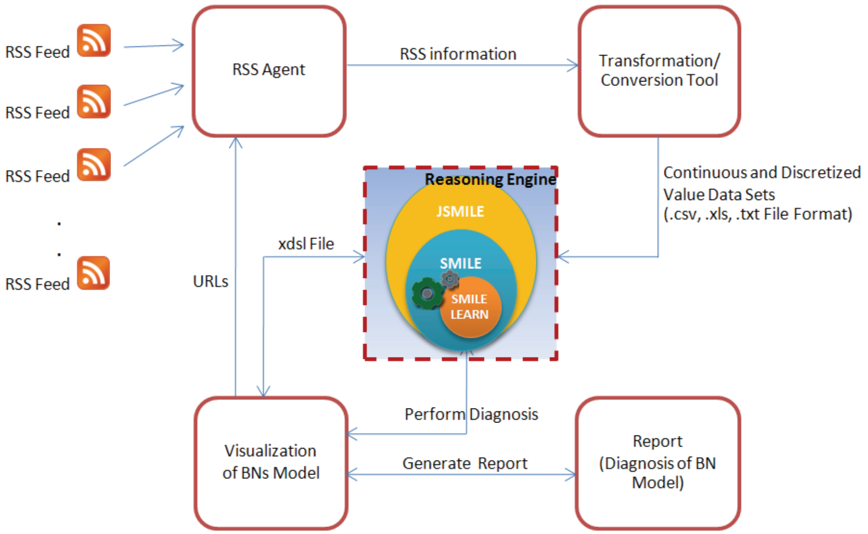
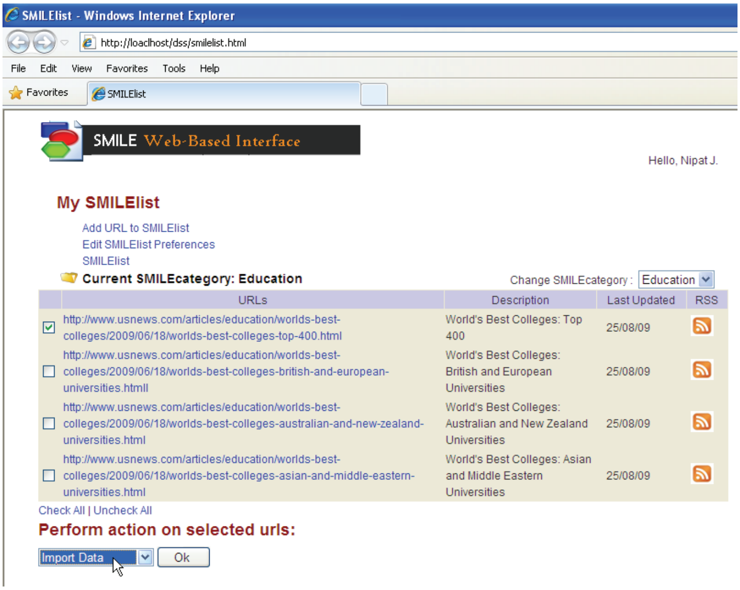

RSS Feed
RSS Feed
RSS Feed
RSS Feed
RSS Agent
RSS information
Transformation/ Conversion Tool
Continuous and Discretized Value Data Sets (.csv, .xls, .txt File Format)
Reasoning Engine
JSMILE
SMILE
SMILE LEARN
xdsl File
URLs
Perform Diagnosis
Visualization of BNs Model
Generate Report
Report (Diagnosis of BN Model)

Fig. 1. A practical framework for automating the building of diagnostic BN models from online data sources

```
<?xml version="1.0" encoding="utf-8"?>
<rss version="2.0">
<channel>
        <title>Top 400 world's best universities</title>
        <link>http://www.usnews.com/articles/education/worlds-best-colleges/2009</link>
        <description> Top 400 world's best universities were reported</description>
        <lastBuildDate> Mon, 10 Aug 2009  18:37:00 GMT</lastBuildDate>
        <language>en-us</language>
<item>
        <uname>Harvard University</uname>
        <academicPeerReviewScore>100</academicPeerReviewScore>
        <employerReviewScore>100</employerReviewScore>
        <studentToFacultyScore>96</studentToFacultyScore>
        <internationalFacultyScore>87</internationalFacultyScore>
        <internationalStudentsScore>81</internationalStudentsScore>
        <citationsPerFaculty>100</citationsPerFaculty>
</item>
<item>
        <uname>Yale University</uname>
        <academicPeerReviewScore>100</academicPeerReviewScore>
        <employerReviewScore>100</employerReviewScore>
        <studentToFacultyScore>89</studentToFacultyScore>
        <internationalFacultyScore>71</internationalFacultyScore>
        <internationalStudentsScore>98</internationalStudentsScore>
        <citationsPerFaculty>100</citationsPerFaculty>
</item>
.
.
.
</channel>
</rss>
```

Beyond the creation of RSS feeds of the top 400 world's best universities, we also created RSS feeds for other topics such as World's Best Colleges in Asian and Middle Eastern Universities, Australian and New Zealand Universities, British and European Universities, and Canadian Universities.

First, the RSS agent is used to collect and read the provided RSS feeds according to the agent's predefined URLs. Fig. 2 shows the screenshot of the SMILE web-based interface. Users click on "Add URL to SMILElist" option to add URLs that contains RSS feeds to the list. Next, they click on the "SMILElist" option to display all added URLs in the table and then click on the checkbox to select the URLs (RSS feeds) that are required for the Bayesian belief network construction, then choose "Import Data" from the data list, and finally click on the OK button to pass the URLs to the agent. The agent gathers the RSS feeds according to the specified URLs and passes them to the transformation/conversion tool.


SMILElist - Windows Internet Explorer
http://loaclhost/dss/smilelist.html
File Edit View Favorites Tools Help
Favorites
SMILElist
SMILE Web-Based Interface
Hello, Nipat J.

My SMILElist

Add URL to SMILElist
Edit SMILElist Preferences
SMILElist

Current SMILEcategory: Education

Change SMILEcategory : Education

	URLs	Description	Last Updated	RSS
☑	http://www.usnews.com/articles/education/worlds-best-colleges/2009/06/18/worlds-best-colleges-top-400.html	World's Best Colleges: Top 400	25/08/09	▨
□	http://www.usnews.com/articles/education/worlds-best-colleges/2009/06/18/worlds-best-colleges-british-and-european-universities.htmll	World's Best Colleges: British and European Universities	25/08/09	▨
□	http://www.usnews.com/articles/education/worlds-best-colleges/2009/06/18/worlds-best-colleges-australian-and-new-zealand-universities.html	World's Best Colleges: Australian and New Zealand Universities	25/08/09	▨
□	http://www.usnews.com/articles/education/worlds-best-colleges/2009/06/18/worlds-best-colleges-asian-and-middle-eastern-universities.html	World's Best Colleges: Asian and Middle Eastern Universities	25/08/09	▨

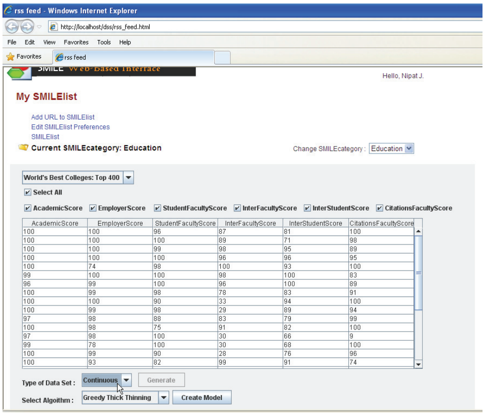
Check All | Uncheck All

Perform action on selected urls:

Import Data | Ok

Fig. 2. Screenshot of SMILE web-based interface

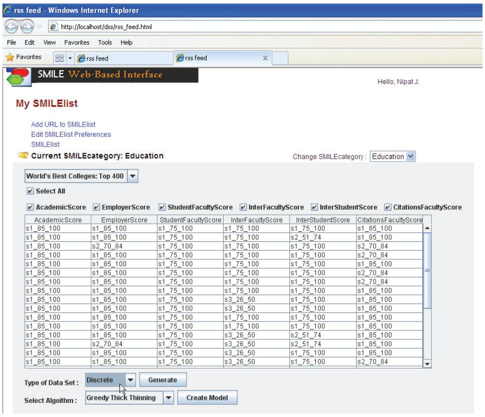
Second, the transformation/conversion tool provides a method to cope with pre-processing the data obtained from the RSS feeds and handles missing values in the dataset, checks the compatibility and integration of collected data, and converts them into two separate data sets (continuous and discrete valued data sets). Fig. 3(a) and Fig. 3(b) show continuous and discrete valued data sets displayed using a datagrid view on the web. The users choose one of the issues from the list in a combo box above the data grid table. The checkbox for each variable is automatically displayed and the variables and their data are loaded into the datagrid table. Users can choose some of the variables by clicking on the checkbox or choose the "Select All" option to select all variables to be included for learning and building the causal structure of the BN model. If they select the type of data set (continuous or discrete) from the list in a combo box below the datagrid table, the selected data set is loaded and displayed in the datagrid view. To select a "Discrete" value data set, they must specify the properties of each variable such as state name, lower bound value, and upper bound value for each state of the variables. The screenshot of defining and editing the properties of each variable are shown in Fig. 4(a) and Fig. 4(b). Next, they select a learning algorithm (Thick Thinning, PC, Essential Graph Search, and Naive Bayes) from the list in a combo box below and then click on the "Create Model" button to pass these parameter settings to the core engine.


rss feed - Windows Internet Explorer
http://localhost/dss/rss_feed.html
File Edit View Favorites Tools Help
Favorites
rss feed
SMILE Web-Based Interface
Hello, Nipat J.
My SMILElist
Add URL to SMILElist
Edit SMILElist Preferences
SMILElist
Current SMILEcategory: Education
Change SMILEcategory : Education
World's Best Colleges: Top 400
☑ Select All
☑ AcademicScore ☑ EmployerScore ☑ StudentFacultyScore ☑ InterFacultyScore ☑ InterStudentScore ☑ CitationsFacultyScore

AcademicScore	EmployerScore	StudentFacultyScore	InterFacultyScore	InterStudentScore	CitationsFacultyScore
100	100	96	87	81	100
100	100	100	89	71	98
100	100	99	98	95	89
100	100	100	96	96	95
100	74	98	100	93	100
99	100	100	98	100	83
96	99	100	96	100	89
100	99	98	78	83	91
100	100	90	33	94	100
100	99	98	29	89	94
97	98	88	83	79	99
100	98	75	91	82	100
97	98	100	30	66	9
99	78	100	30	68	100
100	99	90	28	76	96
100	93	82	99	91	74

Type of Data Set : Continuous
Select Algoithm : Greedy Thick Thinning

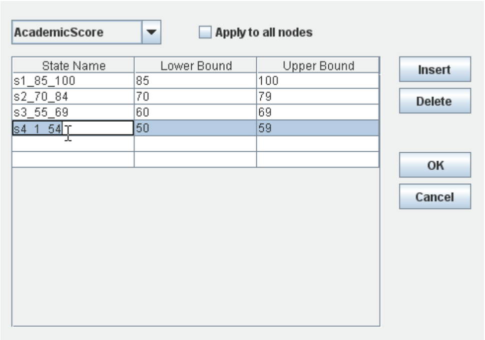
Fig. 3(a). A continuous valued data set


rss feed - Windows Internet Explorer
http://localhost/dss/rss_feed.html
File Edit View Favorites Tools Help
Favorites
rss feed
rss feed
SMILE Web-Based Interface
Hello, Nipat J.
My SMILElist
Add URL to SMILElist
Edit SMILElist Preferences
SMILElist
Current SMILEcategory: Education
Change SMILEcategory : Education
World's Best Colleges: Top 400
☑ Select All
☑ AcademicScore ☑ EmployerScore ☑ StudentFacultyScore ☑ InterFacultyScore ☑ InterStudentScore ☑ CitationsFacultyScore

AcademicScore	EmployerScore	StudentFacultyScore	InterFacultyScore	InterStudentScore	CitationsFacultyScore
s1_85_100	s1_85_100	s1_75_100	s1_75_100	s1_75_100	s1_85_100
s1_85_100	s1_85_100	s1_75_100	s1_75_100	s2_51_74	s1_85_100
s1_85_100	s2_70_84	s1_75_100	s1_75_100	s1_75_100	s2_70_84
s1_85_100	s1_85_100	s1_75_100	s1_75_100	s1_75_100	s2_70_84
s1_85_100	s1_85_100	s1_75_100	s1_75_100	s1_75_100	s1_85_100
s1_85_100	s1_85_100	s1_75_100	s1_75_100	s1_75_100	s2_70_84
s1_85_100	s1_85_100	s1_75_100	s1_75_100	s1_75_100	s2_70_84
s1_85_100	s1_85_100	s1_75_100	s1_75_100	s1_75_100	s1_85_100
s1_85_100	s1_85_100	s1_75_100	s3_26_50	s1_75_100	s1_85_100
s1_85_100	s1_85_100	s1_75_100	s3_26_50	s1_75_100	s1_85_100
s1_85_100	s1_85_100	s1_75_100	s1_75_100	s1_75_100	s1_85_100
s1_85_100	s1_85_100	s1_75_100	s1_75_100	s1_75_100	s1_85_100
s1_85_100	s1_85_100	s1_75_100	s3_26_50	s2_51_74	s1_85_100
s1_85_100	s2_70_84	s1_75_100	s3_26_50	s2_51_74	s1_85_100
s1_85_100	s1_85_100	s1_75_100	s3_26_50	s1_75_100	s1_85_100
s1_85_100	s1_85_100	s1_75_100	s3_26_50	s1_75_100	s2_70_84

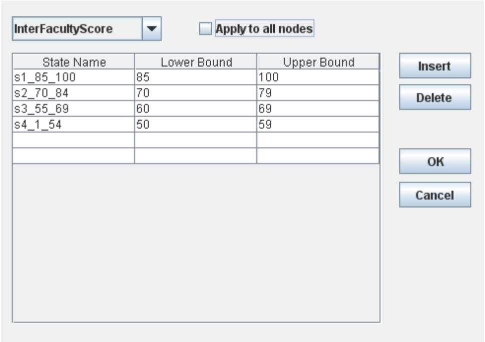
Type of Data Set : Discrete
Select Algoithm : Greedy Thick Thinning

Fig. 3(b). A discrete valued data set


AcademicScore

☐ Apply to all nodes

State Name	Lower Bound	Upper Bound
s1_85_100	85	100
s2_70_84	70	79
s3_55_69	60	69
s4_1_54	50	59


InterFacultyScore

☐ Apply to all nodes

State Name	Lower Bound	Upper Bound
s1_85_100	85	100
s2_70_84	70	79
s3_55_69	60	69
s4_1_54	50	59

Fig. 4(a). Define the properties for the variable Fig. 4(b). Edit the details of each variable

Third, the core reasoning engine receives continuous or discrete valued data sets, learns and builds the causal structure of the BN model, and performs diagnosis analysis on the BN model. This is one of the most important components of the framework. It consists of JSMILE, SMILEarn, and SMILE. JSMILE is the outermost shell of the core reasoning engine. It is used as a wrapper for accessing the SMILE and SMILEarn library from the SMILE web-based interface. It calls the functions, passes the parameter values to SMILE and SMILEarn and receives the return values. SMILEarn is actually a module of SMILE. It is used to cope with incoming data from previous steps, pre-processes the data (removing or filling missing values in the dataset, discretization of continuous values, etc.), and learns the causal structure of the BN model. It provides a set of specialized classes that implement learning algorithms and other useful tools for automatically building graphical models from data.

The sample of source code shown below is the class that implements the greedy thick thinning procedure for learning the structure and parameters of a Bayesian network.

```
class DSL_network;
class DSL_dataset;
class DSL_greedyThickThinning
    {
    public:
            DSL_greedyThickThinning()
            {
                        maxParents = 5;
                        priors = K2;
                        netWeight = 1.0;
            }

            int Learn(const DSL_dataset &data, DSL_network &net);
            enum PriorsType { K2, BDeu };
            PriorsType priors;
            int maxParents;
            double netWeight;
            typedef std::vector<std::pair<int, int> > IntPairVector;
            IntPairVector forcedArcs;
            IntPairVector forbiddenArcs;
            IntPairVector tiers;
    };
```

There are several fields of this class that are used for defining some details of the learning algorithm. For example, the “maxParents” field defines the maximal number of parents a node can have. The “priors” field defines the type of priors (K2 method). The learning method, Learn (const DSL_dataset &data, DSL_network &net), performs the actual learning procedure. The first argument is the input dataset. The result from the learning procedure is stored in the DSL_network which is the second argument. The method returns DSL_OKAY if learning was successful and an error code otherwise.

SMILE is mainly used for graphical BN models and provides the functionality to perform a diagnosis. With SMILE diagnosis, a user can determine the state of the network by performing tests or observations. A user is able to select a test and perform it by setting the evidence for the test.

Fourth, the visualization part of the BN model is automatically built by using a java applet and placed on the SMILE web-based interface. In the BN model, the states of each variable or node are automatically altered in real-time when the RSS feeds are updated. The agent checks for the updated RSS feed and loads metadata about its content into the BN model. The update belief function of the core engine is called to update the probability values of the states of each variable and displays them on the model. Fig. 5 depicts a simplified representation of the communication between a web browser and a web server. In the first and second steps, the client makes requests for a web page from the web server. The web server returns an applet back to client side. In the third and last step, the applet will handle and communicate with a servlet application that runs on the web server. Fig. 6 shows the BN model on the SMILE web-based interface. Users select “Picking” from the list in a combo box and clicks on each node to drag and drop it independently on the applet area. They can zoom in or out of the model by clicking on “zoom in/zoom out” button. They click on the

"Update Belief" button to update the probability values of all variables in BN model. The tooltip text for the updated probability results appears when they move the mouse cursor over any node. The updated probability results on each node are shown in Fig. 7.

Lastly, for BN diagnosis, the user is allowed to perform a model diagnosis by entering observations (evidence) for some of the context and evidence variables. Fig. 8 shows the screenshot of the BN model diagnosis. The user begins the BN model diagnosis by performing a right click on a node and selects the state for setting the evidence for the test. After setting the evidence, they click on the "Update Belief" button to update the model.

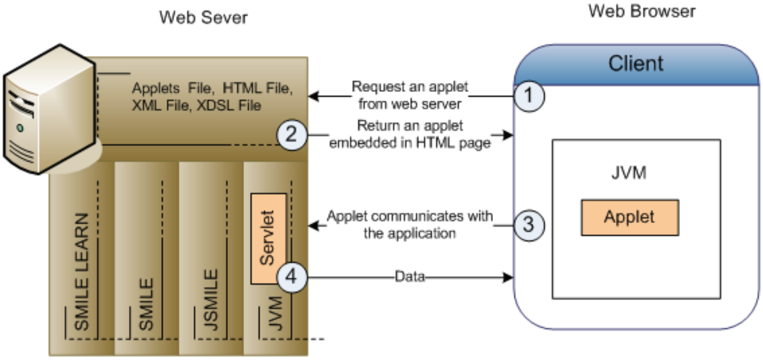
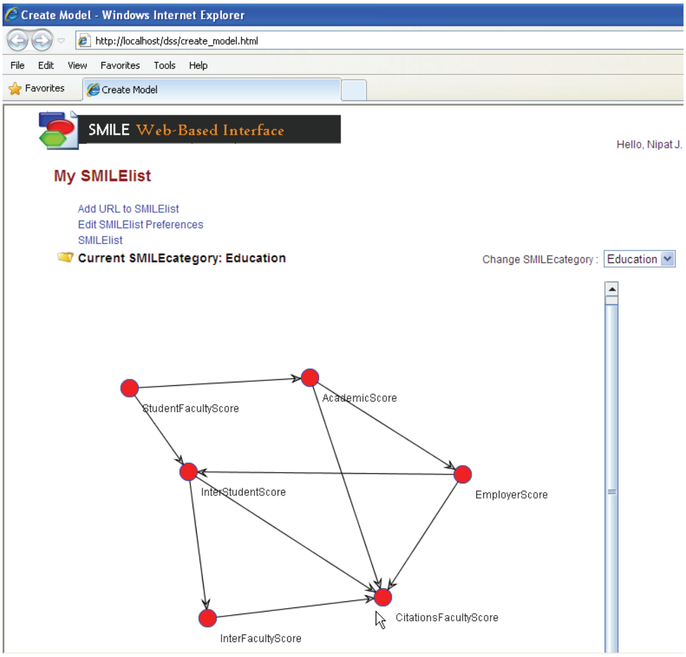

Web Sever
Web Browser
Applets File, HTML File, XML File, XDSL File
SMILE LEARN
SMILE
JSMILE
JVM
Servlet
Client
Request an applet from web server
Return an applet embedded in HTML page
Applet communicates with the application
Data
1
2
3
4
JVM
Applet

Fig. 5. SMILE web-based interface architecture


Create Model - Windows Internet Explorer
http://localhost/dss/create_model.html
File
Edit
View
Favorites
Tools
Help
Favorites
Create Model
SMILE Web-Based Interface

Hello, Nipat J.

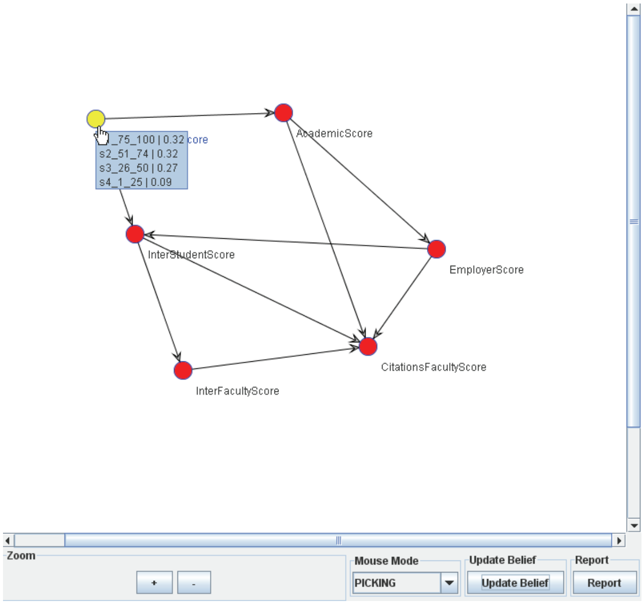
My SMILElist

Add URL to SMILElist
Edit SMILElist Preferences
SMILElist

Current SMILEcategory: Education

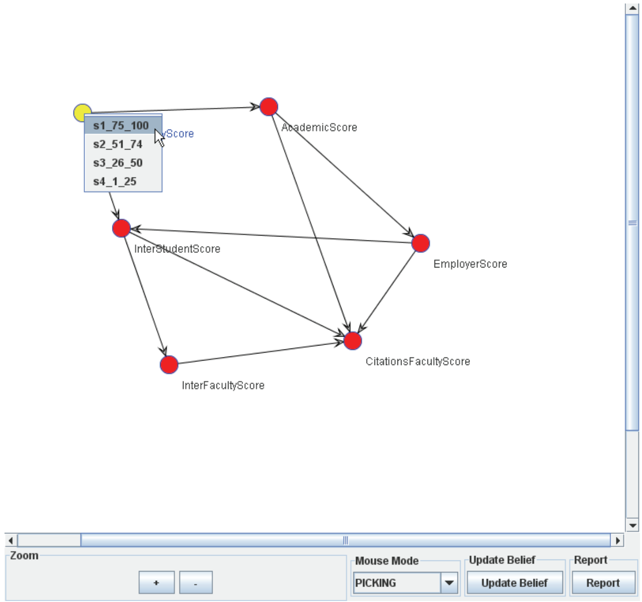
Change SMILEcategory : Education

Fig. 6. Screenshot of the BN model on SMILE web-based interface



Mouse Mode PICKING

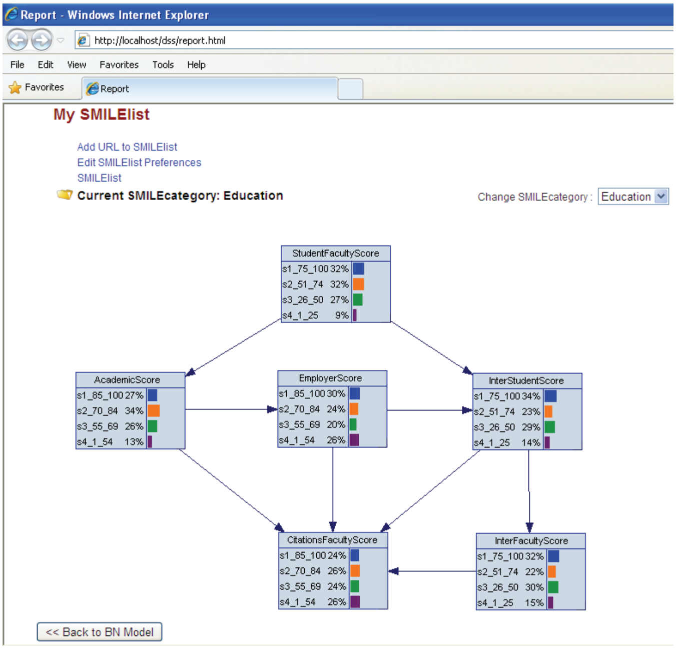
Fig. 7. Screenshot of the probability values on the node after updating belief



Mouse Mode PICKING

Fig. 8. Screenshot of a BN model diagnosis

Next, they click on the "Report" button to display a graphical representation of the BN model. The report graphically displays a BN model and its probability values with a horizontal bar graph. It is shown in Fig. 9 and Fig. 10.


Report - Windows Internet Explorer
http://localhost/dss/report.html
File
Edit
View
Favorites
Tools
Help
Favorites
Report
<< Back to BN Model

My SMILElist

Add URL to SMILElist
Edit SMILElist Preferences
SMILElist

Current SMILEcategory: Education

Change SMILEcategory : Education

StudentFacultyScore	
s1_75_100	32%
s2_51_74	32%
s3_26_50	27%
s4_1_25	9%

AcademicScore	
s1_85_100	27%
s2_70_84	34%
s3_55_69	26%
s4_1_54	13%

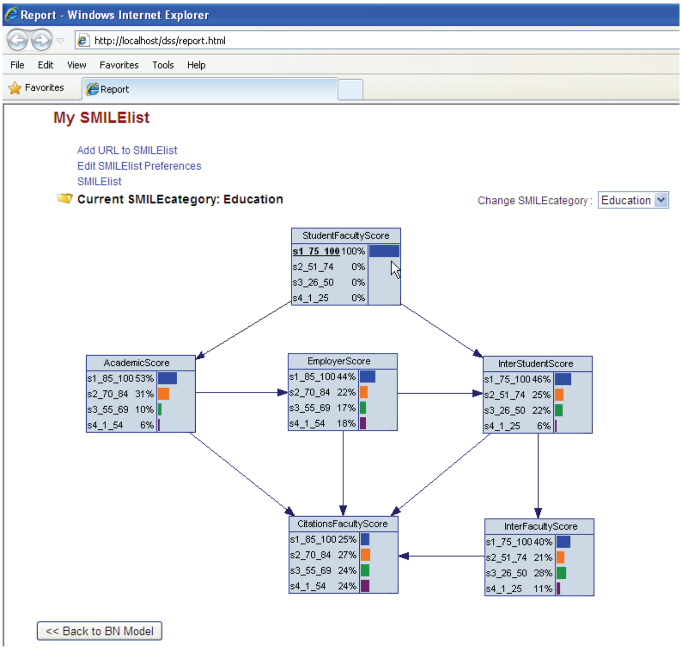
EmployerScore	
s1_85_100	30%
s2_70_84	24%
s3_55_69	20%
s4_1_54	26%

InterStudentScore	
s1_75_100	34%
s2_51_74	23%
s3_26_50	29%
s4_1_25	14%

CitationsFacultyScore	
s1_85_100	24%
s2_70_84	26%
s3_55_69	24%
s4_1_54	26%

InterFacultyScore	
s1_75_100	32%
s2_51_74	22%
s3_26_50	30%
s4_1_25	15%

Fig. 9. Screenshot of a graphical report on a BN model


Report - Windows Internet Explorer
http://localhost/dss/report.html
File
Edit
View
Favorites
Tools
Help
Favorites
Report
<< Back to BN Model

My SMILElist

Add URL to SMILElist
Edit SMILElist Preferences
SMILElist

Current SMILEcategory: Education

Change SMILEcategory : Education

StudentFacultyScore	
s1_75_100	100%
s2_51_74	0%
s3_26_50	0%
s4_1_25	0%

AcademicScore	
s1_85_100	53%
s2_70_84	31%
s3_55_69	10%
s4_1_54	6%

EmployerScore	
s1_85_100	44%
s2_70_84	22%
s3_55_69	17%
s4_1_54	18%

InterStudentScore	
s1_75_100	46%
s2_51_74	25%
s3_26_50	22%
s4_1_25	6%

CitationsFacultyScore	
s1_85_100	25%
s2_70_84	27%
s3_55_69	24%
s4_1_54	24%

InterFacultyScore	
s1_75_100	40%
s2_51_74	21%
s3_26_50	28%
s4_1_25	11%

Fig. 10. Screenshot of a graphical report on a BN model (Set Evidence)

5. Conclusion

This paper presented a practical framework for automating the building of diagnostic BN models from data sources obtained from the WWW and demonstrates the use of a SMILE web-based interface to represent them. The framework consists of the following components: RSS agent, transformation/conversion tool, core reasoning engine, and the SMILE web-based interface. The RSS agent automatically collects and reads the provided RSS feeds according to the agent's predefined URLs. A transformation/conversion tool provides a method to handle the pre-processing of the data obtained from the RSS feeds and handles missing values, checks the compatibility and integration of collected data, and converts them into two separate continuous and discrete valued data sets. The core reasoning engine learns and builds the causal structure for the BN model, and performs probabilistic inference for the Bayesian belief network. A SMILE web-based interface permits users to perform Bayesian network diagnosis through the web. They can quantify uncertain interactions among random variables by setting observations (evidence) and use this quantification to determine the impact of observations. A graphical structure for representing the BN model is shown as a report and displayed to the user.

The two main principles utilized in the proposed framework and SMILE web-based interface were transparency and ease of use. Our future work will focus on improving a decision-oriented diagnosis approach. The SMILE web-based interface has been extended to cope with influence or relevance diagrams. The next version of the application will allow users to quantify a decision maker's decision options and preferences and use these to determine an optimal decision policy.

6. Acknowledgement

The authors would like to thank the Decision Systems Laboratory, University of Pittsburgh for supporting documents, and source file of the engines: Structural Modeling, Inference, and Learning Engine (SMILE), SMILEarn, and JSMILE wrapper. All necessary files and documentations have been obtained from the Decision Systems Laboratory's web site. It is available at http://genie.sis.pitt.edu.

7. References

Lacave, C. ; Luque, M. and Díez, F.J. (2007). Explanation of Bayesian Networks and Influence Diagrams in Elvira. *IEEE Transactions on Systems, Man, AND Cybernetics-Part B: Cybernetics*, Vol. 37, No. 4, p. 952-965.

Yuan, C. & Druzdzel, M.J. (2003). An Importance Sampling Algorithm Based on Evidence Pre-propagation. *Nineteenth International Conference on Uncertainty in Artificial Intelligence*, Acapulco, Mexico, p. 624-631.

Dagum, P., & Luby, M. (1993). Approximate probabilistic reasoning in Bayesian belief network is NP-Hard. *Artificial Intelligence*, Vol. 60, p. 141-153.

Dagum, P. & Luby, M. (1997). An Optimal Approximation Algorithm for Bayesian Inference, *Artificial Intelligence*, Vol.93, p.1-27.

Jensen, F.V.; Olesen, K. G. and Andersen, S.K. (1990). An Algebra of Bayesian Belief Universes for Knowledge-Based Systems. *Networks: Special Issue on Influence Diagrams*, Vol.20, No. 5, August 1990, p.637-659.

Fung, R., & Chang, K.C. (1989). Weighing and Integrating Evidence for Stochastic Simulation in Bayesian Networks. *In Proceedings of the Fifth Conference on Uncertainty in Articial Intelligence (UAI-89)*, p.209-219, New York, N.Y., Elsevier Science Publishing Company, Inc.

Fung, R., & DelFavero, B. (1994). Backward Simulation in Bayesian Networks. *In Proceedings of the Tenth Annual Conference on Uncertainty in Artificial Intelligence (UAI–94)*, p.227-234, San Francisco, C.A., Morgan Kaufmann Publishers.

Cooper, G.F. (1990). The Computational Complexity of Probabilistic Inference using Bayesian Belief Networks, *Artificial Intelligent*, Vol. 42, No. 2-3, p. 393-405.

Cheng, J. & Druzdzel, M.J. (2000). AIS-BN: An Adaptive Importance Sampling Algorithm for Evidential Reasoning in Large Bayesian Networks. *Journal of Artificial Intelligence Research (JAIR)*, Vol. 13, p. 155-188.

Jongsawat, N.; Poompuang, P. & Premchaiswadi, W. (2008). Dynamic Data Feed to Bayesian Network Model and SMILE Web Application. *In Proceedings of IEEE on Software Engineering: Artificial Intelligence, Networking, and Parallel/Distributed Computing*, p. 931-936, Phuket, Bangkok, Thailand.

Jongsawat, N. & Premchaiswadi, W. (2009). A SMILE Web-Based Interface for Learning the Causal Structure and Performing a Diagnosis of a Bayesian Network. *Proceedings of IEEE on Systems, Man, and Cybernetic, Systems Science & Engineering (SMC2009)*, San Antonio, Texas, USA.

Druzdzel, M.J. (1999). SMILE: Structural Modeling, Inference, and Learning Engine and GeNIe: A Development Environment for Graphical Decision-Theoretic Models. *In Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI–99)*, p. 902-903, Orlando, FL.

Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems. *Networks of Plausible Inference,* San Mateo, CA, Morgan Kaufmann Publishers.

Pearl, J. (1986). Fusion, Propagation, and Structuring in Belief Networks. *Artificial Intelligence*, Vol. 29, No. 3, p. 241-288.

Mylltmaki, P.; Silander, T.; Tirri, H. & Uronen, P. (2002). B-Course a Web-based Tool for Bayesian and Causal Data Analysis. *International Journal on Artificial Intelligence Tools*, Vol. 11, No. 3, p. 369-387.

Poompuang, P.; Kungtasthan, A. ; Jongsawat, N. & Sutheebanjard, P. (2007). Graphical Decision-Theoretic Models on the Web. *Proceedings of Knowledge Management*, p. 163-170, Bangkok, Thailand.

Shachter, R.D. & Peot, M.A. (1989). Simulation Approaches to General Probabilistic Inference on Belief Networks. *In Uncertainty in Artificial Intelligence*, p. 221-231, New York, N.Y., Elsevier Science Publishing Company, Inc.

Lauritzen, S.L. & Spiegelhalter, D.J. (1988). Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems (With Discussion). *Journal of the Royal Statistical Society Series B*, Vol. 50, No 2, p.157-224.

Tungkasthan, A.; Poompuang, P. & Premchaiswadi, W. (2008). SMILE Visualization with Flash Technologies, *Proceedings of IEEE on Software Engineering: Artificial Intelligence, Networking, and Parallel/Distributed Computing*, p. 551-556, Phuket, Bangkok, Thailand.

Decision Support System Based on Effective Knowledge Management Framework To Process Customer Order Enquiry

Dr. Chike. F. Oduoza,
Reader in Process Engineering and Management,
School of Engineering and Built Environment,
University of Wolverhampton
WV1 1LY
United Kingdom

1. Introduction

The customer enquiry stage in a business management environment is very challenging as it strongly influences future workload of production management activity. At this stage, customers generally make enquiries requesting product delivery in terms of quantity, delivery date and sales price. Firms usually need to respond to these enquiries before customers can confirm the corresponding quotes and finally the enquiries may be translated into customer orders. A firm's potential profitability depends crucially on selecting a proper subset of enquiries to fulfill, delay or turn away. Such a decision is the responsibility of the sales and marketing department. However, it is the customer who makes the final decision as to whether to order or not and how many to order, based largely on the satisfaction derived from the enquiry. At present, there are very few, efficient, and effective, simple and easy to implement methodologies, to help businesses manage existing in house knowledge in order to respond to ordering enquiries at the customer enquiry stage.
Literature review (Harris, 2009) shows that:

- Existing management systems do not support all key activities within the enterprise product development process.
- Only a few systems provide the necessary decision support throughout product development and in most cases are more geared to specific tasks with stand alone functionality.
- Few decision support management systems provide the means for the exploitation of manufacturing constraints and product knowledge
- Available knowledge management systems do not provide the means to identify, capture, formalise, present and utilise tacit knowledge
- The lack of exploitation of information and knowledge has a major impact on overall product development process
- Knowledge and experiences gained from existing projects are sometimes poorly documented and therefore are not available for reuse in other related projects

- Transfer of knowledge from previous projects will no doubt affect the quality, efficiency, cost and time to market of developing new products and processes.

The aim of this chapter is to propose the application of a comprehensive decision support system for knowledge and information management during customer order enquiry with a view to minimise cost, achieve quality assurance and enhance time to market especially in new product development. This report will be illustrated by case studies demonstrating how effective and robust knowledge management would support decision making especially at the order enquiry stage during product development. It will also present a DSS that highlights the influence of negotiation on customer due dates in order to achieve forward or backward planning with a resultant profit maximization outcome if the strategy is carefully implemented. A mathematical model will be developed that links profit maximization with screening customer / order enquiries and thereby decide whether or not to proceed with an inquiry by balancing capacity against demands placed on it. In the long term it is expected that the decision support system will be capable of assessing future customer orders / enquiries based on previous experience.

2. The significance of knowledge management in decision making (processing order enquiries)

Experimental data is classed as raw or discerned elements and when these elements are patterned in a certain way, data becomes transformed into information. When rules or heuristics are applied to information, knowledge is then created as actionable information for producing some value added benefit. The knowledge that is created and shared amongst organisational members can be categorised into two typical forms of knowledge - Tacit and Explicit (Polanyi 1996).

Tacit knowledge is highly personal, context-specific, and therefore hard to formalise and communicate, this type of knowledge is stored in the human brain, such as in personal belief, expertise, perspective and values formed as a result of experience. On the other hand, **explicit knowledge** is defined as public knowledge and covers those aspects of knowledge that can be articulated in formal language and can be easily transmitted among individuals using information technology. There are two basic strategies for managing knowledge (Hansen 1999; Swan *et al.*, 1999) as follows;

- ***Codification strategy*** is based on codifying the knowledge and storing it in artefacts and databases where it can be accessed.
- ***Personalisation strategy*** is where the knowledge is tied up to the persons who develop the knowledge and therefore the sharing of that knowledge is achieved only by personal interactions.

Vast amounts of work have been carried out on how knowledge is utilized in each organizational activity; especially in marketing, costing, design, manufacture etc, however there is limited research available on how knowledge from all the activities within product development can be pulled together and utilized to provide decision support throughout the product life cycle. Knowledge in product development environment is considered to consist of four different activities.(Harris, 2009)

1. Identification; the identification of knowledge required to develop new products, including product specifications, process, tooling, and material capabilities
2. Capture; how the knowledge is captured stored and retrieved.

3. Formalize and Present; how knowledge can be formalized and presented to ensure its use in existing and future projects.
4. Utilization; how the knowledge identified, captured and formalized can be integrated into products and decisions, and applied in other projects.

Western and Japanese cultures both view knowledge differently; while the Japanese view knowledge as being primarily tacit (not easily seen or expressible), the West focuses on explicit knowledge expressed in words and numbers and therefore more easily communicated than tacit knowledge. Perceptions of knowledge seem to be rooted in culture. Nonaka, and Takeuchi (1995), adopt a traditional definition of knowledge as" justified personal belief" and closely tied to an individual's or group's values and beliefs. On the other hand, Miller and Morris (1999) suggest that knowledge is gained when theory, information, and experience are integrated. Enterprise knowledge is thought to be a dynamic mix of individual, group, organisational and inter organisational experiences, values, information and expert insights. This concept originates in the minds of the individual knowledge workers, and then emerges as knowledge workers interact with each other and the environment, and finally knowledge is leveraged for efficient customer management and competitive advantage.

Knowledge management is thought to be a discipline whose major objective is to develop methods and tools for detecting, leveraging, distributing and improving the knowledge assets of organisations (Cortes et al (2001). Knowledge management background is thought to comprise organisational theory, information systems, knowledge representation, and human and machine learning. It is thought to be the systematic, goal oriented application of measures to steer and control the tangible and intangible knowledge assets of organisations, with the aim of using existing knowledge inside and outside of these organisations to enable the creation of new knowledge and generate value, innovation and improvement out of it. (Jaime et al 2006).

Global business enterprises are aware that access to essential operations information will enable them maintain competitive advantage and thereby stay one step ahead of other businesses. They therefore need to develop an effective knowledge management strategy both for the benefit of their employees and customers in order to support decision making process and thereby remain sustainable. Knowledge management concept has increasingly become fashionable however, many organisations are still unable to develop and leverage knowledge to enhance business performance. In most cases organisational knowledge are fragmented, sometimes difficult to locate and therefore to leverage, share and reuse. Tacit knowledge exists in the minds of employees and therefore may not be available to process customer queries and enquiries. There is a need therefore to develop robust decision support systems to capture, store, share and leverage data, information and knowledge. Decision support systems will enable the transformation of tacit to implicit knowledge to be shared and leveraged for improved decision making. They will also enable the conversion of explicit to implicit knowledge a process of internalisation.

There are various perspectives of knowledge management; strategic knowledge management – deals with pinpointing opportunities to find, distribute and transfer knowledge related to long term goals of an organisation; tactical knowledge management finds, distributes and transfers knowledge for the medium term organisational goals; operational knowledge management is associated with daily or short term operations (Young, et al, 2007). They developed a knowledge management framework demonstrating how design information and knowledge, manufacturing information and knowledge, operations information and knowledge and disposal information and knowledge all add up to shape the total product information for the product life cycle. Mustafa and Robert (2003), have also developed a knowledge-based decision support system (KBDSS) suitable for short term scheduling in flexible manufacturing systems and strongly influenced by the tool management concept to provide a significant operational control tool for a wide range of machining cells, where a high level of flexibility is demanded. The benefits are more efficient cell utilisation, greater tool flow control and a dependable way of rapidly adjusting short term production requirements. **Development of a robust knowledge-based system to support the decision making process is made necessary by the inability of decision makers to promptly address all the questions posed by potential customers at the enquiry stage and also to diagnose efficiently many of the malfunctions that arise at machine, cell, and entire system levels during manufacturing.**

Knowledge management models and frameworks are categorised based on their purpose; They could be predictive (predicting what a system's behaviour will be), explanatory / descriptive (enables past observations to be understood as part of an overall process) or prescriptive - provides a picture of the real world if certain rules are applied (Small & Sage, 2006)

Decision support systems would play a major role in information requirements determination in the system life cycle ranging from design through to manufacture, operations and final disposal. Unfortunately existing support methodologies especially in product design and manufacture focus on how to specify requirements once they are determined but however don't enable optimal determination of those requirements. There is therefore a need for a decision support system equipped with a structured knowledge base to help information analysts in the critical decision task of determining requirements especially in system design / manufacture. A typical application of the proposed decision system is in the determination of requirements for product design / manufacture to promptly respond to customer order enquiries in a business environment. Lui and Young (2007) have discussed key information models and their relationships in manufacturing decision support in three different scenarios. They confirmed that global manufacturing businesses are benefiting from the information and knowledge support provided by modern IT tools such as Product Life Cycle Management (PLM), Enterprise Resource Planning (ERP), and Customer Relations Management (CRM). For instance two types of decision making in a global manufacturing scenario can be distinguished; decisions associated with product configuration and decisions associated with project coordination

Figure 1 shows the concept of knowledge derived from theory, information and experience and extended to include wisdom – which is tacit in nature and could be described as successfully applied knowledge. For instance when an enterprise faces an order enquiry from customer, it should be able to respond to the enquiry based on available information, theory and experience. When there is a repeat enquiry in the future the enterprise should now be able to respond promptly based on what it has previously learned and also from a previous successfully applied knowledge.

Figure 2 demonstrates information / knowledge management framework for a product life cycle showing all the sources and phases at which information about the product can be derived. Each of the phases has data and knowledge that describe the characteristics of that stage in the product life cycle. Such information are useful to both the design and manufacture engineers and also to the customer who would need a full understanding of product attributes to enable optimal design, manufacture and guaranteed product

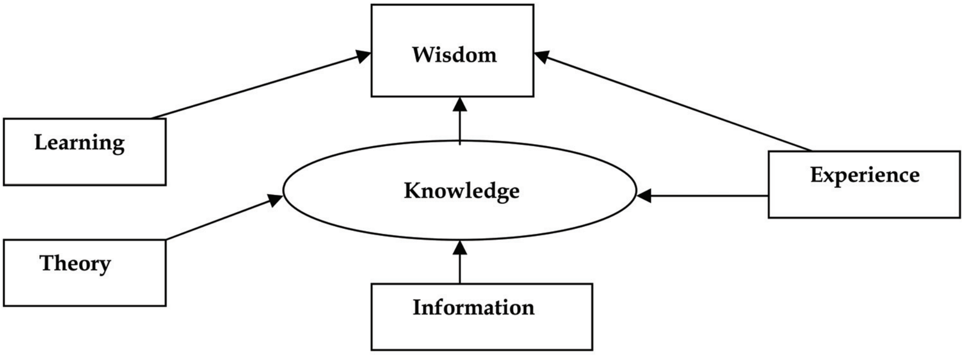


Fig. 1. Knowledge derived from Theory, Information and Experience

performance. At the order enquiry stage a potential customer would be made aware of all relevant information including data on product functionality, durability, efficiency, energy requirement, etc, while the sales representative will negotiate on product specification and requirements, mode of operation, value added, cost price, delivery due date, maintenance requirements, etc. The sales department working in collaboration with the design / production department will also establish that they can deliver what has been promised to the customer within the due date. This will involve material requirement planning, supplier management, production scheduling and planning, outsourcing requirement, quality assurance etc. An order is confirmed only when there is an agreement / contract established between

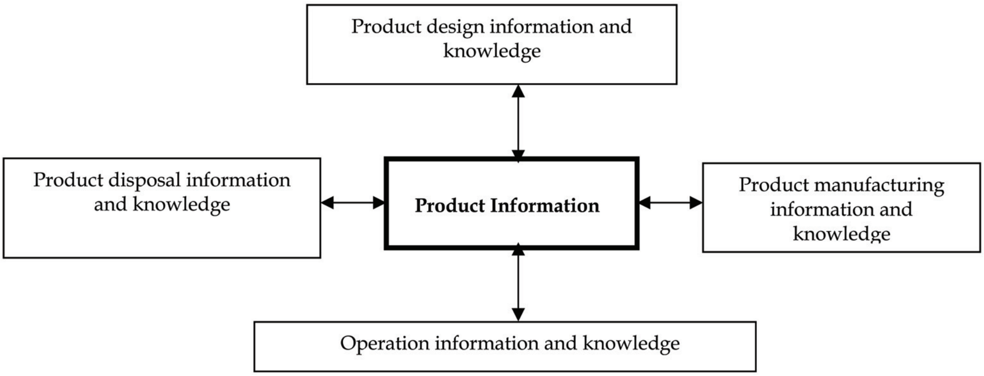

Product design information and knowledge
Product disposal information and knowledge
Product Information
Product manufacturing information and knowledge
Operation information and knowledge

Fig. 2. Information and Knowledge Framework Derived Based on Product Life Cycle

3. Methodologies to respond to customer ordering enquiries – a literature review

Customer Relationship Management (CRM) has attracted the attention of both academics and practitioners. CRM focuses on managing the relationship between a company and its current and prospective customers and is key to success for many organisations (Gebert et al 2003). In CRM, one of the important concerns is how to offer improved levels of customer service and support by means of a variety of ideas, approaches, and tools (Sheikh, 2003). It provides a tool that will enable prompt decision making about an enquiry in a make to order environment to the mutual satisfaction of both the customer and the enterprise. Parente, Pegels and Nallan (2002) surveyed production and sales managers and their findings indicate that the internal relationship between sales and production is important to the customer, especially in Engineer-To-Order production situation. In today's highly customer-centric competitive market, improving customer service level would be crucial for firms to increase their competitiveness. While it is a challenging task it is worth the effort according to Xu et al (2002). Currently, research into customer order enquiry is scanty yet initial customer enquiries constitute the gateway for building a sustainable customer relationship with an organization. The focus of a previous study by Kingsman *et al.* (1996) was on the management of manufacturing lead time and job release at the cutomer enquiry stage. They argued that dealing properly with enquiries is a major problem that make-to-order companies face, and a lack of co-ordination between sales and production at the customer enquiry stage often leads to confirmed orders being delivered later than promised and a possible production at a loss. Hendry and Kingsman (1993), indicated that a hierarchical production planning system specifically designed for the make-to-order sector of industry is necessary for customer enquiry management. The aim of the hierarchical system was to control the delivery and manufacturing lead times of all orders processed by a firm. However, details of how to respond to customer enquiries, especially for a large set of enquiries, were not elaborated on in their study. They only highlighted the problem and tried to work out a feasible solution.

Kingsman et al (1993) researched on the integration of marketing and production planning in make to order companies and concluded that a major problem confronting firms is the gap between sales/marketing and production functions. This lack of coordination often led to confirmed orders being delivered later than the due date by the sales team. Sometimes orders are produced at a loss, or could lead to a delay of other orders with a consequence of additional costs.. In another study, Hendry (1992) developed a methodology in which two decision levels, the customer enquiry and job release stages were addressed and linked. A decision support system (DSS) was finally developed to assist in planning the capacity at the customer enquiry stage in make-to-order companies. In order to make the response realistic, satisfying and competitive, the literature emphasizes that it is imperative to integrate marketing and production planning. Ulusoy and Yazgac (1995) argued that cooperation between production and marketing departments appear to have a large impact on the well being of a firm. They have therefore, developed a multi-period, multi-product model with the objective of profit maximization reflecting the characteristics of both departments. The advertising efficiency and price of the products are determinable within their model. Similarly, Kingsman *et al.* (1993) discussed possible approaches that depended on estimating routinely the probability of winning an enquiry order, dependent on many factors including price and lead-time etc. Olumolade and Norrie (1996) have developed a decision support system for scheduling in a customer-oriented manufacturing environment. The aim of their research was to assess schedulability prior to assigning parts for scheduling. Their system comprised four basic modules – the demand, material management, tool management and system status modules. Halsall and Price (1999) have also presented a DSS approach to support production planning and control in smaller companies and argued that SME companies would benefit from a manufacturing DSS in which the links between customer orders and manufacturing operations were maintained throughout the duration of the production planning process.

Xiong *et al.*(2003) recently proposed a DSS framework suitable for the management of customer enquiries for SMEs. Their studies indicate that the DSS approach plays a very important role in assisting SMEs to respond to enquiries at the customer enquiry stage. Enterprise Resource Planning System (ERP) can help firms automate their order entry, process customer order and keep track of order status. They are also used to plan capacity and create daily production schedule for manufacturing plants (Yen et al 2002). However, such software may be expensive and inappropriate for SMEs to use for only processing enquiries at the customer enquiry stage. Implementation of such software is complex, and elaborate, and requires huge initial investment and continuing maintenance expenditure (Halsall et al 1999 and Xiong et al 2006). Most recently, Oduoza and Xiong (2007) developed a decision support system framework to process customer order enquiries in SMEs.

3.1 Order processing at the customer enquiry stage, a challenge for businesses

The customer enquiry stage has a strong impact on the production workload and well being of businesses. At this stage, customer enquiries need to be transferred to customer orders and planned for in the next production run. If a firm fails to achieve enough customer orders, its production capacity would be underutilized and waste occurs. A key objective for businesses and SMEs in particular is to maximise profits and minimize waste while processing customer requirements. Generally, such a decision is based on the acceptance or rejection of an enquiry or could even involve negotiation with customers in order to protect the interest of both parties. This process needs to be carefully handled to maintain a firm's market credibility. Customer enquiry management therefore plays a very major role in the business operations of enterprises and for SMEs it is often difficult to properly manage this essential part of their business.

When orders become confirmed, firms generally schedule them against receipts of materials/components and the standard manufacturing lead-time. This forms the basis for Material Requirement Planning (MRP) or Manufacturing Resource Planning (MRPII). Stadtler and Kilger (2000), however, argued that such a method often led to unrealistic production plans because it assumed infinite supply of materials and capacity beyond the standard lead-time and creating supply recommendations based on order backlog. Additionally, such a method did not aim at processing enquiries at the customer enquiry stage, but at planning confirmed orders for the next production procedure. More importantly, customer enquiry is only a prelude to ordering and cannot be simply considered as an actual order. A lot of enquiries might sometimes only request a delivery date, a delivery quantity or sales price for a product without any commitment to ordering. A customer often makes a similar enquiry to several companies at the same time in today's e-business market environment and to a large extent, the decision as to whether an enquiry can be transformed to an order depends primarily on the satisfaction of the customer to the preliminary enquiries / responses from companies. The more responsive firms are in terms of speed, and quality of delivery, the more feasible it may be able to secure the current order and subsequently future orders. However, it seems almost impossible for businesses to provide a proper response without the help of practical and useful tools and techniques. Although there are commercially available software that can assist firms in dealing with orders including automating order entry, processing such orders and keeping track of order status, these systems often require information and function integration, and may therefore have a complicated system structure. Their implementation therefore, is complex, and requires huge initial investment and continuing maintenance expenditure (Halsall and

Price, 1999).. However, an SME company generally only has limited budget, and may not afford to implement such a system for dealing only with customer enquiries. In addition, the lack of professional expertise and technical support in SMEs more especially makes it difficult to make decisions if and when they finally prepare to implement such a system. Consequently, a relatively simple and practical tool would be very useful to guide SMEs in the rapid processing of customer enquiries.

As far as the decision on how to respond to customer sales department of a firm is concerned, the lack of effective coordination between different functions such as production and marketing departments could affect the reliability of the response. For most business enterprises, the production department is often confronted with unrealistic delivery dates for incoming orders. This usually arises when the marketing department often quotes a price and delivery date to maximise their chance of winning the order, however, the production department would need to reconcile impending demand with available resources, capacity utilization, production routing, etc. Because of a deficiency in an integrated information management system, the coordination of different functions may become difficult to achieve.

3.2 Responsive customer enquiry management

Speed of delivery and quality of responses to enquiries seem to be the two major factors affecting customer enquiry management (Xiong et al, 2003 and Xiong et al, 2006). The efficiency of such responses determines how fast the enquiry can be followed through. At the crucial stage, the time spent in responding to an enquiry comprises the time between the receipt of the enquiry and the completion of the response. By and large, a firm would endeavour to decrease this time to enhance the responsiveness to process enquiries as customers would not appreciate a long waiting time for a response. In today's highly competitive market, a customer would quickly resort to another supplier if an early response from a previous vendor is not achieved. The quality of response to customer enquiry is a measure of effectiveness of the management system. A reliable and feasible response is determined by the probability of keeping the delivery promise after making a response. In many cases, achieving response efficiency in terms of speed and as well as delivering error free orders could be conflicting and may sometimes prove difficult. Therefore, to optimise response efficiency and order delivery to specification simultaneously should be major objectives for effective customer enquiry management. By adopting suitable appropriate techniques, method and tools, it becomes possible to achieve customers' main priorities, such as precise delivery time, exact quantity required and also affordable sales price. For example, a successful customer enquiry management might allow customers to enquire and order via the Internet using a build-in-order model and ensuring they get what they want while also enabling the company to cover its cost.

Typically, an enquiry comprises information about requested quantity, delivery date (DD) and price for a product. This gives rise to a three-dimensional response surface constructed with quantity, delivery date and sales price as the axes, thus providing a possible solution, or guideline, for quoting orders for customers and negotiating with customers. There are generally two levels for the dimensions of DD and sales price, fixed level and flexible level. If the customer has specified a DD and a price in its enquiry, the time frame and price is generally considered as fixed. The firm also should be able to check whether it can benefit from accepting the order and whether sufficient capacity is available to produce the new order in addition to existing jobs that have already been confirmed for that time period. Under this circumstance, the firm cannot change the requested DD and price without renegotiating with customers in advance. On the other hand, if an enquiry does not specify a fixed DD and price, a feasible DD and a proper price are defined and included in the response in terms of other objectives such as prioritization, sales expectation, accounting, and capacity, etc. Such information can be used for quoting, approving and negotiation. Typical examples for such an enquiry are those with an unusually huge quantity requested for a particular product or orders with a very short delivery time. Normally, the response to such an enquiry must be considered and approved by production management, e.g. the tender vetting committee, in the firm.

4. Development of a decision support system to manage order enquiry processing

The procedure for responding to an enquiry is basically a multi-stage decision process (Kingsman et al 1996). The initial decision is whether or not to prepare a bid, and if so, how much effort to put into the specification and estimation process. The Make To Order company has the choice to put in a lot of effort to prepare a competitive bid or produce a quick estimate with a high safety margin to allow for errors and unforeseen problems and for further negotiation. Consideration has to be given to the accuracy of the cost estimates produced, the feasibility of being able to produce the order within the current work load at specified delivery times and finally possible overrun costs. If the decision is to go ahead to produce an order, the next step is to provide a response based on the three basic elements of the enquiry – Due Date, Quantity required and Sales Price. In the light of this challenge, a suitable approach is to develop a DSS (Decision Support System) environment to assess different options provided by the system and then make a final well-informed choice. For example, the detailed impact a DD would have on the workload can be estimated graphically, and also the implications of sticking to the DD while adjusting manufacturing capacities could be assessed for decision-making. Also, the decision to further negotiate with customers on a bid could be made if enquiry is considered very important. This involves combining robust business rules with powerful computing capability for effective decision-making. While it does not present an optimal solution, it provides the flexibility for users to consider alternative courses of action before making a decision. By using a DSS approach, potential feasible solutions arising from customer enquiries are evaluated against predefined objectives on the basis of which decisions are made. The major objectives of such a DSS approach are to,

- bridge the gap between sales and marketing functions on the one hand, and manufacturing, product development, finance, human relations, etc on the other hand.
- provide a guideline for negotiation between company and customers
- enable decision making in cases of uncertainty.
- optimize production capacity and material availability

4.1 Prerequisites of proposed decision support system

The architecture of the proposed DSS approach is shown in Fig. 3. The entire system is based on databases which provide all necessary data of customer, production capacity and materials availability, accounting, product and customer, etc. The whole process dealing with enquiries is streamlined and controlled within the environment and a set of interfaces are provided to integrate tools that equip the system with the flexibility to access necessary real time information, assess different options, and undertake a sensitivity analysis. The DSS is broken down into the following modules to enable the right business decision-making.

- **Web Based Enquiries:** Verify incoming enquiries for suitability of delivery (based on available production capacity, process capability, time constraint, potential profit) by the firm. A request could be rejected at this initial stage of the pre-screening process.
- **Due Date Check:** This module checks that the requested due date is feasible based on available production capacity and materials availability. If it is not feasible, some capacity adjustments may have to be made such as overtime usage, operator reallocation or even production rescheduling. The following three scenarios may then become possible;
  - **Negotiate Due Date, while aiming at a fast order**
    DD may be already set by company although this is the most critical success factor to the potential customer
  - **Negotiate Due Date and keep as a slow order**
    DD may be already set by company but this is not a critical success factor to the potential customer
  - **Due Date is fixed**
    The due date is fixed by the potential customer and not negotiable.
- **Evaluation of Enquiries:** Given limited available production capacity, enquiries need to be evaluated in order to select a subset of enquiries, which will be fulfilled. Such evaluation can be based on objectives such as reducing the inventory cost and/or increasing the potential profitability of orders. The DSS should provide several objectives to facilitate the needs of a variety of users. Hence, mathematical models may be combined with judgmental rules to improve the accuracy of the evaluation process.

Typical examples of judgmental rules include,

- Profitability of fulfilling the order
- Importance of the customer
- Value of the order and effect on future business
- Possibility for a repeat order
- Balance of workload for work centers
- Entry into new market
- Process capability to handle order

- **Financial Outcome:** The price for ordering a product for delivery on a given due date is assessed in terms of overall profit maximization and available capacity.
- **Enquiry Audit and Approval:**. The objective of this module is to audit the enquiry process prior to approval by management for important and/or unusual enquiries. These enquiries would strongly influence production planning and scheduling, hence a high level approval is required to finalise an order before confirmation is sent to customers.
- **Capacity Planning:** It is essential to allocate available capacity to priority orders which suggests a policy for order screening and portfolio management. The purpose of this module is to provide management with the capability to assess alternative courses of action necessary to fulfill customer orders. Typical outputs for capacity adjustment are; choice of alternative materials, critical paths/items, recommended purchase orders for critical materials, and request to expedite/de-expedite purchased orders, as well as the plan to assign overtime and reallocate operators between different work centers.

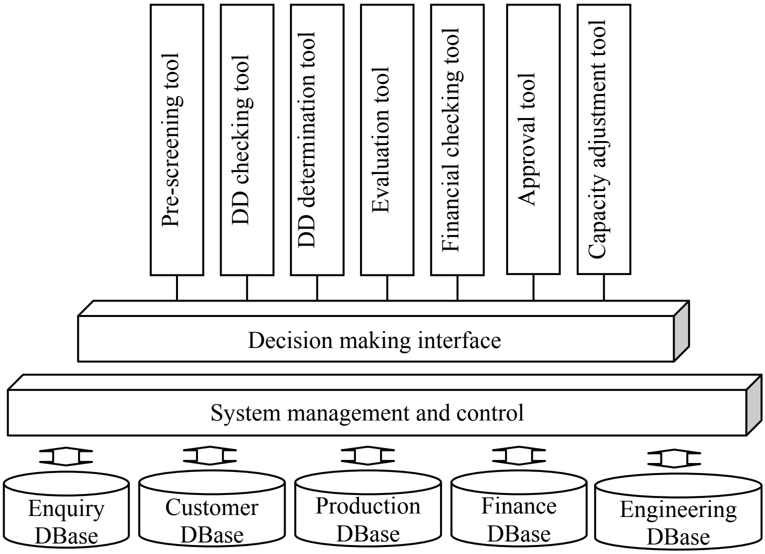


Fig. 3. Generic DSS architecture for customer enquiry management suitable for SMEs (Oduoza & Xiong, 2007)

The nature of the link between the proposed DSS approach and other production planning functions will depend on other systems that a company has in place. The proposed DSS can be envisaged as a front-end system to deal with enquiries in the first instance before they are translated into customer orders and then enter the production planning process. It also could be used at tender vetting committee meetings or in the preparation for such meetings. For enquiries that are considered not important enough to convene a committee meeting, the person responsible for providing a response to such enquiries would need to use it.

4.2 Framework constructs of the decision support system

Successful construction of the framework for the decision support system will depend on essential parameters such as available to promise (ATP) which is a function of material availability and available capacity necessary to manufacture the desired product. These parameters will now be discussed in greater detail.

4.2.1 Available To Promise (ATP) and its determination

The first concern for the proposed DSS is the definition of a criterion to measure the capability to meet customer requirements. Here, we propose to use a concept termed the available-to-promise (ATP) which is a bucketized quantity typically used on weekly basis. It is a standard quantity capable of being produced during a time period based on material availability of all components that assemble or manufacture the requested product. Therefore, product structure, described as bill-of-material (BOM), is essential in ATP computation. Typically the ATP computation complexity increases as the product BOM becomes more complex

Figure 4 shows a flow chart for decision support process based on customer flexibility in terms of due date for delivery. When due date is fixed b y the customer then it is defined by a backward planning to accommodate a new order subject to materials availability, potential profit and available capacity. However, a flexible due date can enable a forward planning. Overall, the final acceptability of an enquiry will be based on due date, available capacity, process capability, potential profit and materials availability.

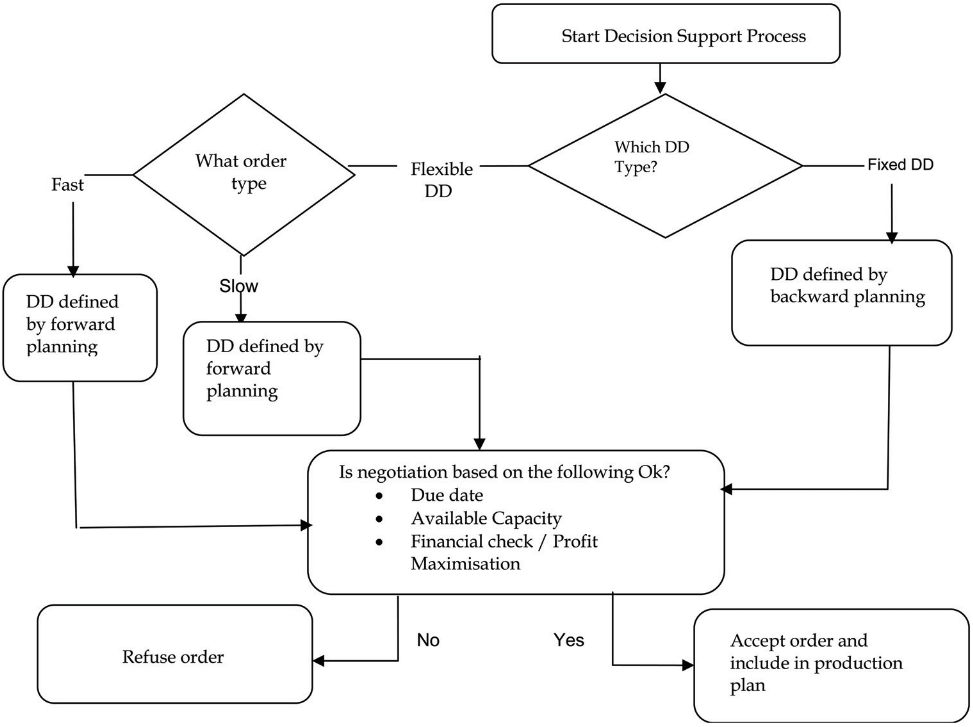

Start Decision Support Process
Which DD Type?
Flexible DD
Fixed DD
What order type
Fast
Slow
DD defined by forward planning
DD defined by forward planning
DD defined by backward planning
Is negotiation based on the following Ok?
Due date
Available Capacity
Financial check / Profit Maximisation
No
Yes
Refuse order
Accept order and include in production plan

Fig. 4. Flow chart for decision support process based on negotiation with customer

Xiong *et al*. (2003), Oduoza and Xiong (2007) presented a dynamic BOM approach for handling the complex ATP computation for products with multi-level BOM. A dynamic BOM is a two-level BOM, which is generated dynamically in terms of the materials availability of different components during BOM explosion. Through an iterative process to generate a set of dynamic BOMs, the ATP can be accumulated through exploding BOM from top downwards by the associated computation approach. The process in which a set of dynamic BOMs is generated is shown in Fig 5. There are three dynamic BOMs generated sequentially in correspondence with the ordinary product BOM shown at the left side of Fig. 5. For the product BOM, Dynamic BOM 1 is first created due to materials shortage of component C_2 which is described as *Critical Item*. Corresponding to Dynamic BOM 1, ATP is initially determined by quantity per component and lead time of all components, C_1, C_2, …, C_I, required for the process.

4.2.2 Production capacity

The production capacity required for this manufacturing process is then checked against the available production capacity. If the available capacity is sufficient, material availability

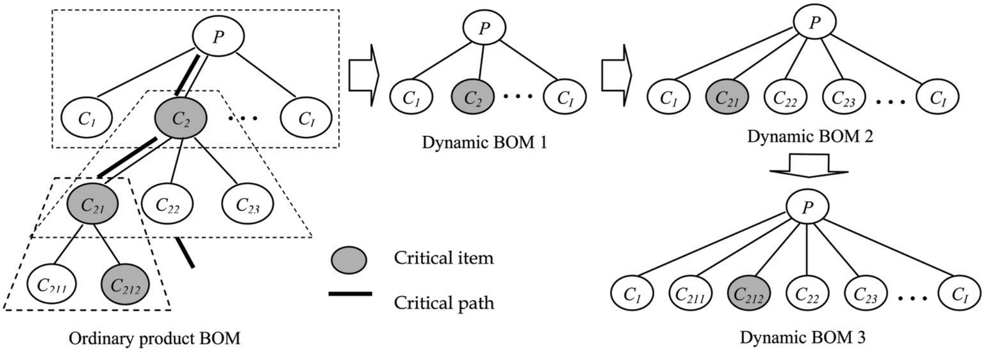


Fig. 5. The generation process of a set of dynamic BOMs during BOM explosion

restricts this production process; otherwise, the production capacity will be under constraint. If C_2 is identified as the Critical Item in Dynamic BOM 1, a new dynamic BOM (Dynamic BOM 2), is formed by replacing it with its direct child components, C_{21}, C_{22} and C_{23}, and similar ATP computation is carried out for Dynamic BOM 2. This process continues until a Critical Item is one bottom-level component in a dynamic BOM, such as C_{212} in Dynamic BOM 3. Details of the dynamic BOM based ATP computations have been previously described by Xiong *et al* (2003). In a recent short communication, Framinan and Leisten (2007) have recommended a reformulation of the model to redefine the inventory holding costs arguing that the model in its current form can lead to certain problems regarding the constraints and the objective function.

In the dynamic BOM based ATP computation presented above, both materials availability and production capacity are effectively reconciled and accounted for. ATP computed on this basis is thus more realistic and reliable too. Because it is easy to put into consideration the materials availability and related production capacity constraints in two-level BOM structure, such dynamic BOM based ATP computation is a suitable approach to determine the real time fulfillment capability with respect to customer enquiries.

4.2.3 Application of linear programming to enhance decision support

It is essential to extend traditional mathematical programming models incorporating intrinsic uncertainty to business decision making but without the assumption of model coefficients. Oliveira and Antunes (2007) have demonstrated the application of multiple objective linear programming models with interval coefficients for decision-making in uncertainty while Polacek et al (2007) applied variable neighbourhood search algorithm to schedule periodic customer visits in order to minimize travel time for salespersons. In a separate study Klashner and Sabet (2007) presented a new DSS design model for complex critical decision-making and partial empirical evaluation derived from a field study at a power utility control centre while most recently Power and Sharda (2007) reviewed model driven Decision Support Systems (DSS) built on the basis of decision analysis, optimization and simulation technologies. The authors presented issues that users need to be aware of in the operation of the decision support system and also emphasized on the user interface, as well as behavioural issues in decision support systems.

Zorzini, Corti and Pozzetti (2008) have also proposed an interpretative framework to identify the contextual factors impacting company choices during decision making at the customer enquiry stage. They presented a model that formalizes the decision process for setting due dates categorized as (a) negotiable due date, fast order (DD set by company but delivery time performance is the most relevant critical success factor), (b) negotiable due date, slow order (DD can be set by company and delivery time performance is not the main critical success factor) and (c) fixed DD (DD is fixed by customer).

Under limited Available To Promise quantities, it is imperative to evaluate a set of enquiries in order to select a subset to fill so that certain business objectives can be achieved. The objectives of this process vary from company to company, and are usually predefined based on company's policy and business philosophy. However, one of the very important business principles is to maximise company's revenue from processing customer enquiries. Assuming that all enquiries are requested for one product only and the DD of every enquiry is fixed, the model to evaluate customer enquiries is derived as follows.

Indices:

i - index of customer enquiries, $i \in I$, where I is the number of enquiries

t - index of time buckets, $t \in T$, where T is length of planning horizon

Parameters:

t_i - requested time bucket for enquiry i, $i \in I$

$E_i(t_i)$ - quantity required by enquiry i in time bucket t_i, $t_i \in T$ and $i \in I$

$p(t)$ - sales price per unit of product in time bucket t

$ATP(t)$ - ATP quantity used for filling customer demands in time bucket t

c_h - unit inventory holding cost per time bucket

c_p – unit manufacturing cost

c_l – unit lost sales cost

Decision variables:

α_i - binary variable stating whether to accept enquiry i

β_{ti} - fraction of $ATP(t)$ allocated to enquiry i

Objective function:

$$\underset{i\in I, t\in T}{Max}\ profit = f_{revenue} - f_{cost} \tag{1}$$

where, $f_{revenue}$ is the revenue from accepting customer enquiries, and f_{cost} is the cost incurred from accepting these enquiries.

$$f_{revenue} = \sum_{i=1}^{I} [\alpha_i * E_i(t_i) * p(t_i)] \tag{2}$$

$$f_{cost} = f_m + f_l + f_h \tag{3}$$

where, f_m, f_l, and f_h represent manufacturing cost, lost sales cost and inventory holding cost respectively, and they are defined as follows.

$$f_m = \sum_{i=1}^{I} [\alpha_i * E_i(t_i) * c_p] \quad \textbf{(manufacturing cost)} \tag{4}$$

$$f_l = \sum_{i=1}^{I} \sum_{t=1}^{T} [(1-\alpha_i) * E_i(t_i) * c_l] \quad \textbf{(lost sales cost)} \tag{5}$$

$$f_h = \sum_{t=1}^{T}\{c_h * [ATP(t) * (T - t)]\} \ \text{-} \ \sum_{i=1}^{I}[c_h * \alpha_i * E_i(t_i) * (T - t_i)] \quad \textbf{(holding cost)} \tag{6}$$

Constraints:

1. Customer requested quantity

$$\sum_{t=1}^{T} \beta_{ti} = \alpha_i \quad \forall i \in I \tag{7}$$

2. ATP quantity

$$\sum_{i=1}^{I} \beta_{ti} * E_i(t_i) \le ATP(t) \quad \forall t \in T \ , \ t_i \ge t \tag{8}$$

3. Fraction of ATP in time bucket *t* allocated to enquiry *i*

$$0 \le \beta_{ti} \le 1 \quad \forall i \in I \ \ \forall t \in T \tag{9}$$

4. Variable constraints

$$\alpha_i = 0 \text{ or } 1 \quad \forall i \in I \tag{10}$$

As described above, ATP is based on available working time of associated work centers as well as the materials availability of all related components. The objective of the model, in Equation (1), is to maximize the profit from accepting a subset of customer enquiries. Equation (2) defines the revenue objective while Equation (3) represents the cost incurred from accepting certain customer enquiries which includes manufacturing cost (Equation 4), lost sales cost (Equation 5) and inventory holding cost (Equation 6). Inventory holding cost f_h consists of two items; the first item represents the total inventory holding cost without accepting any customer enquiry; the second item is the decreasing inventory holding cost from accepting certain enquiries. Equation (7) defines the allocation fraction β_{ti}, generating a quantitative relationship between the sum of allocation fractions from ATP to a specific enquiry and decision variable α_i. Equation (8) ensures that the allocated ATP quantity for all orders within each time bucket must not exceed the ATP quantity in that time bucket.

The model proposed above is a mixed 0-1 linear programming model, and its global optimum can be obtained by using commercially available optimization solver such as LINGO. This mathematical model provides an adaptive combination of every cost as well as profit associated with acceptance of customer enquiries. It can thus help firms analyze real life management decisions.

4.2.4 Sensitivity analysis

In order to facilitate the assessment of every option to respond to customer enquiries, sensitivity analysis is imperative in other to run through customer enquiries effectively. For instance, in some cases, it may be necessary to take special actions by adjusting materials and production capacity in order to reduce the lead time. This is especially true in cases of large orders from customers or demand for a very short DD. Since the definition of capacity referred to in the proposed DSS includes materials availability and production capacity, the following information may be used in the DSS for adjusting materials availability.

- Alternative components
- Critical component and implications for related material shortage, and
- Delivery lead time and lot size for critical component

In the proposed DSS, the variables affecting material availability includes sourcing for alternative materials, issue of critical paths / items, purchase orders for critical materials, and expedite /de-expedite purchase orders.

At the planning stage, the production capacity is usually planned on a weekly basis by means of the forecast values of the total workload on the shop floor (Hendry, 1992). However, when a specific enquiry (for example an order for a large quantity) is received, the DSS should be able to adjust the production capacity to allow for the special needs for this particular order. The two most common methods for adjusting the production capacity – assigning overtime and reallocating operators between different work centers – can be easily incorporated into this proposed DSS. The overtime is usually assigned to some bottleneck work centers to increase their available working time. This is necessary to expedite such orders that may be delayed if no action were taken. The method of reallocating operators between different work centers will be appropriate if there is an imbalance of workload across the shop floor.

5. Knowledge management to support decision making during new product development – a case study

This section describes the case application of knowledge management to support decision making during new product development. Harris (2009) has developed a knowledge based framework to support new product development and describes logistical details of how knowledge acquired at the product / process design stage feeds into manufacture. The captured knowledge and lessons learnt from this initial cycle enhance/improve future product design in terms of tooling, manufacture and product attributes and requirements.

Figure 6 shows a knowledge based framework to support decision making during product development (Oduoza, Harris and Al-Ashaab, 2010). The system describes two major databases; product data archive, and a manufacturing knowledge base both linking to manufacture and product / process design activities. System end users, manufacturing staff and project engineers all have access to the two databases from their work stations and can extract relevant information when necessary. It clearly highlights how enterprise business strategy is linked to both product and process design and manufacture. The framework shows that details of the product and process design can be accessed from a product data archive comprising; product, manufacture, and tooling requirements and also the associated production history. The product archive feeds into the manufacture knowledge base which contains all the relevant information on machine specifications and capabilities, material requirements and planning, and quality assurance issues. The framework also describes details of how the product design department links and collaborates with manufacturing / production section in terms of procedure, technology and project management.

Product Data Archive

The product data archive comprises information, manufacture and tooling modules. It also highlights the product manufacturing history.

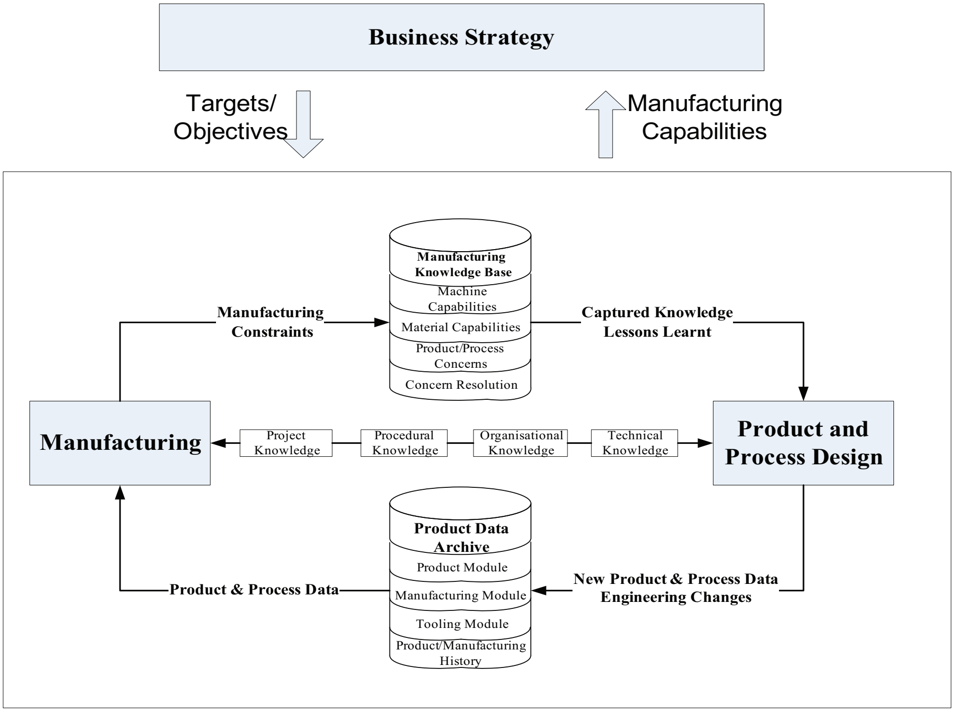


Fig. 6. Knowledge Based Framework to Support Product Development (Oduoza, Harris and Al-Ashaab, 2010)

Product module.

The product module details all the product specifications especially the drawing of the product, material specifications, quality assurance requirements, packaging, and delivery specifications. Each element of the product module highlights further details showing previous revisions, and issue dates.

Manufacture module.

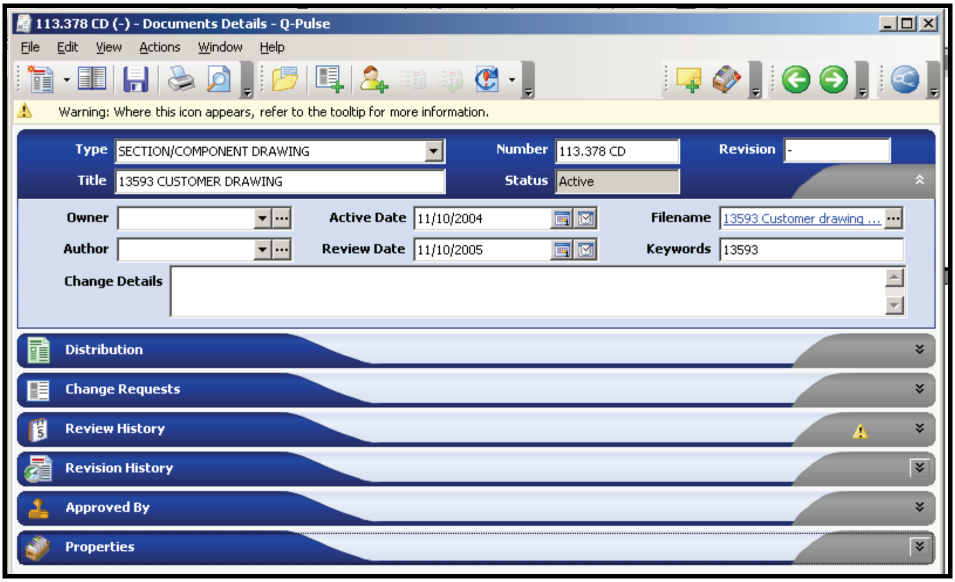
Decision making during the product development process is quite often difficult as not all stakeholders have access to all the necessary information and knowledge related to the manufacturing process. This is made easier by assembling all the necessary information relating to the manufacturing process such as process capabilities, capacity, machine size and speed, standard operating procedures etc necessary to create a manufacturing module. The manufacturing module in the product data archive can be accessed in the form of charts, spreadsheets and system specifications. In summary, the Manufacturing Module contains the relevant knowledge necessary to identify the manufacturing methods and processes for new product development.

Tooling module

The tooling module consists of a variety of tool designs used in roll forming, cut off, piercing, setting sheets, programs for the production lines, and all the necessary information to produce and set tooling on the machines.

Product manufacturing history.

Product manufacturing history provides details of each product from the first enquiry, through to new product introduction and manufacture, and details all engineering changes relating to the product, manufacture and tooling modules.



113.378 CD (-) - Documents Details - Q-Pulse

Warning: Where this icon appears, refer to the tooltip for more information.

Type: SECTION/COMPONENT DRAWING
Number: 113.378 CD
Revision: -
Title: 13593 CUSTOMER DRAWING
Status: Active
Owner:
Active Date: 11/10/2004
Filename: 13593 Customer drawing ...
Author:
Review Date: 11/10/2005
Keywords: 13593
Change Details:

Fig. 7. Document details screen for part number 113.378 (Oduoza, Harris and Al-Ashaab, 2010)

Figure 7 shows a screen shot of the document details screen for a given part number in the product data archive

The data archive system is split into two major areas;

a. Document list screen, where documents can be searched by:

- Document Type.
- Part No.
- Document No.
- Revision level.
- Document owner.
- Customer.
- Product Market sector.
- Type of product.

b. Document details screen which provides information about the specific document and includes:

- Document type.
- Document number.
- Document status
- Document revision
- Document author
- Document owner
- Document active date; date of issue

- File name and hyperlink
- Distribution list
- Change requests
- Document review history
- Document revision history
- Document approval
- Document associated properties; product customer, product type, product market sector.

The document control side of the product data archive provides a central storage area for all documents related to the development and manufacture of products. It prevents the use of obsolete documents and only presents the most current revision to the user. Documents are securely managed to prevent uncontrolled modifications, drafts or copies. Records pertaining to document change, history, approval and distribution are securely held for each revision of a document and are readily available for review. Document registers are also automatically updated with approved changes, in other to eliminate the risk of human error while allowing for effective engineering change control.

Enquiry/engineering control.

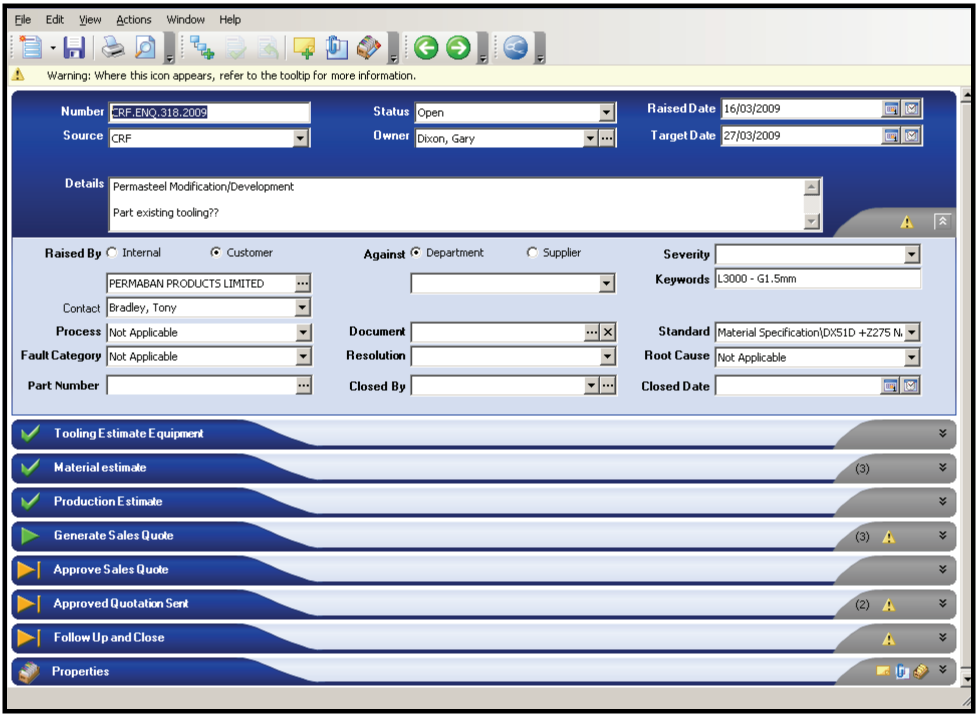
A fundamental part of the product development system is the enquiry/engineering change request phase It is paramount that all information received from the customer and information / knowledge created internally are correctly documented and made available to all to avoid controversy.

Approximately 500 new enquiries and engineering change requests are received by the case company each year and these enquiries comprise insufficient details to quote and product tooling being quoted which is already available. The procedure starts with the enquiry being received from the customer/potential customer followed by data entry onto the system of the following; drawings, concepts and or specifications for the product etc. Enquiry will be declined or discontinued if customer fails to provide all the relevant information that will enable response. When all the relevant information become available, it is given a unique enquiry number, with drawings, product specifications and additional information such as customer and contact details are attached to the enquiry (embedded into the database). All departmental and sectional heads are alerted of a new enquiry on the system with hyperlink to open up the enquiry and make necessary input.

Figure 8 shows a screen shot of the enquiry page displaying general information relating to the enquiry; the enquiry number, date raised, status, ownership of enquiry, customer and customer contact, brief description of the enquiry/engineering change request, and the material specification of the product (provided by the customer). All drawings, specifications and communications relating to the enquiry sit as file attachments in the properties field shown at the bottom of the screen.

As soon as the enquiry has been entered into the system the sequence of steps to be taken is indicated in the drop down tabs in the bottom half of the screen:

- Tooling estimate; this field records tooling required, tooling costs and whether or not the part can be manufactured.
- Materials cost estimate; this field records estimated costs of material to produce the product and cannot be completed until the tooling stage is complete; the tooling stage confirms the material content of the component.
- Production estimate; this field records anticipated production speed, output/hour, utilization and manning levels.



Warning: Where this icon appears, refer to the tooltip for more information.

Number: CRF.ENQ.318.2009
Source: CRF
Status: Open
Owner: Dixon, Gary
Raised Date: 16/03/2009
Target Date: 27/03/2009
Details: Permasteel Modification/Development

Part existing tooling??

Raised By ○ Internal ● Customer
PERMABAN PRODUCTS LIMITED
Contact: Bradley, Tony
Process: Not Applicable
Fault Category: Not Applicable
Part Number:
Against ● Department ○ Supplier
Document:
Resolution:
Closed By:
Severity:
Keywords: L3000 - G1.5mm
Standard: Material Specification\DX51D +Z275 N
Root Cause: Not Applicable
Closed Date:

Fig. 8. Screen shot showing an open enquiry (Oduoza, Harris and Al-Ashaab, 2010)

- Sales quote; the system generates a sales quote only when it obtains detailed information on tooling, materials, and production estimates. The sales quote is then vetted and approved by the manufacturing director prior to release to the customer.
- Dispatch of approved sales quote; this field details when the sales quote is sent and prompts for a copy to be added in the properties field.
- Follow up and close order; If an order is received, details of the order are entered into this field, where the customer has declined the quotations, reasons are added and the enquiry is closed.

The enquiry process is sequential and the system will not allow further details to be entered until the previous stage is complete. All persons involved in the enquiry process are e-mailed when each stage is completed.

In the case organization, the enquiry system has been running for approximately 18 months, and has reduced the time taken to quote to an average of 5 days, and due to the robustness of the system it prevents quotations being raised without the necessary and appropriate information.

Manufacturing knowledge base.

The manufacturing knowledge base is a database with the information and knowledge required to support the product and process design activities. It aims to capture decisions taken and their context in order to support effective decision making in future product development activities. The database contains: machine capabilities, material capabilities, product/process concerns and problem resolution.

Machine capabilities

The manufacturing capabilities element of the manufacturing knowledge base is utilized to select suitable manufacturing equipment and production lines for the manufacture of a specific product. The selection of suitable manufacturing equipment is carried out in conjunction with the product module information.

Material capabilities

The material capabilities element of the manufacturing knowledge base is probably the most important aspect to be considered when developing a new product. Selection of the wrong material could have consequence on tooling specification / estimate.

6. Construction industry enquiry on material fabrication using cold roll forming process - a case study.

Cold roll forming is a manufacturing process to transform flat metal strips into various profile shapes achieved by means of forming rolls. This study describes a customer order enquiry received from a client in the construction industry, for the fabrication of lightweight lattice beams, joists and trusses to be utilized in a wide variety of buildings such as schools, hotels, sports halls, superstores and industrial buildings. The product specification was a very high strength to weight ratio, with the lattice beams required to be manufactured in depths ranging from 220mm - 3000mm, and able to span up to 38m depending on building type and application.

The customer proposed a product profile, which was analyzed in order to appreciate manufacturability in terms of physical size, shape, and suitability of the material to cold roll form. The product profiles were assessed by the relevant departments and quotation was put forward, however, tooling costs, were deemed too expensive for the project to be commercially viable. The tooling cost for the three profiles was assessed in excess of £90,000, however the customer requested that tooling costs should be no greater than £60,000. A review meeting held with the client to determine the critical profile requirements and what features if any could be modified came up with the following material recommendations:

- Yield strength of 420 N/mm^2 minimum
- Minimum structural properties and section modulus were specified.
- The overall width of the component was reviewed.
- Design to cost of tooling should not be higher than £60,000 achieved by designing the same set of tooling for identical profiles (spacers to lengthen or shorten the straight parts of the profile referred to as modular/platform tooling)

However, to fulfil this customer order requirement poses a conflict as follows;

- The tighter the specified radius of the product the more material that would be required to form the profile, thus increasing manufacturing product cost.
- There will be an increase/decrease in the sectional properties of the profile. (Dependent on shape and load direction)

Consequently, further analysis of profile structural characteristics was necessary to confirm that increasing the radius of the profile did not alter its structural properties beyond the specified limits. To standardize tooling profiles fully also required dimensional modifications, with two options considered, to either standardize the bottom forming or the top forming rolls. Standardizing the top forming rolls would result in additional set up time whereby all the rolls would have to be stripped off the rolling mill in order to change from one profile to the next.

Information highlighted in this case study can be stored as a part of a knowledge management system to serve as a point of reference for future order enquiries in other to minimise on cost, time and material resources. It will also enable the design of precise and accurate product dimensions.

7. Conclusion

It is essential to manage initial customer enquiries channeled to an organisation satisfactorily as this could facilitate order winning and could even be more rewarding if it became a lucrative project. Enquiry management could certainly be made easier if the sales / marketing agent is fully aware of the nature of the enquiry and whether or not this could be fulfilled. In other to provide a suitable response, the sales agent of a manufacturing enterprise, for instance, should be able to discuss the skills of the work force, and the capability of their business process to manage the fulfillment of that order. Sometimes the respondent to the enquiry is not fully aware of the implications of the enquiry and therefore cannot provide a satisfactory answer to the potential customer and consequently the order could be lost if not properly managed. It is therefore essential to put in place a decision support system to help the respondent to answer crucial questions at this critical stage of developing a potential lasting relationship with a customer.

This chapter has presented a decision support system driven by a robust knowledge management framework to aid the respondent (sales / marketing agent of the firm) at the enquiry stage to provide accurate and useful information which could confirm or transform an enquiry into an order. In a typical scenario (cited in this study) relevant to new product development, the framework comprises product data archive (product database consisting of product, manufacture and, tooling modules and the product manufacturing history) and the manufacturing knowledge base (describes machine and material capabilities, product / process concerns and any concerns resolution). When these two databases (from the knowledge management framework) are properly coordinated, they provide a vast amount of information about the firm which is beneficial for operations management and also useful to potential clients who would want to do business with the organization.

DSS approach proposed here will assist firms and other businesses involved in customer enquiry / order management to make proper and well-informed decisions for customer enquiries. By providing the desired flexibility to assess different responses and experiment with alternative courses of action, the speed and efficiency of making an informed response to customer enquiry can be significantly improved.

The major contributions from this study are:

1. the problems confronting businesses in managing order enquiries at the customer enquiry stage have been highlighted.
2. proposal of a DSS framework to enable effective and efficient management of order enquiries at the customer enquiry stage was initiated. This is significant especially for SMEs who lack the essential resource to respond to such enquiries as well as manage capacity scheduling, planning and materials availability.
3. mathematical model that links profit maximization with screening customer / order enquiries and thereby decide whether or not to proceed with an enquiry by balancing capacity against demands placed on it.

4. review of due date as an important DSS parameter and if negotiable could affect an enquiry outcome
5. construction of a flexible DSS structure suitable for web based customer enquiries, and a recommendation for its implementation.
6. decision support system driven by a robust knowledge management framework to aid the respondent (sales / marketing agent of the firm) at the enquiry stage in providing accurate and useful information which could help to confirm or transform an enquiry into an order

8. Acknowledgement

The author wishes to thank Alan Harris for providing relevant data on the knowledge management framework and also to Mohua. H. Xiong for a previous collaboration on decision support systems framework.

9. References

Cortes, U, Sanchez-Marre, M, Sanguesa, R, Comas, J, Roda, I.R, Poch, M and Riano, D (2001), *Knowledge Management in Environmental Decision Support Systems,* AI Communications, 14, 3-12.

Gebert, H., Geib, M., Kolbe, L., Brenner, W (2003)., *Knowledge-enabled customer relationship management: integrating customer relationship management and knowledge management concepts*, Journal of Knowledge Management, 7(5), 107-123.

Halsall, D.N., and Price, D.H.R(1999)., *A DSS approach to developing systems to support production planning control in small companies*. International Journal of Production Research; 37(7),1645-1660

Hansen, M.T.,(1999). *The search-transfer problem: the role of weak ties in sharing knowledge across organization subunits*. Administrative Science Quarterly 44, 82–111

Harris, A (2009), *A knowledge based framework to support lean product development,* MPhil Thesis, University of Wolverhampton, UK.

Hendry, L.C., and Kingsman, B.G (1993)., *Customer enquiry management: part of a hierarchical system to control lead times in make-to-order companies*. Journal of Operational Research; 44(1); 61-70.

Hendry, L.C (1992)., *COPP: a decision support system for managing customer enquiries*. International Journal of Operations and Production Management, 12(11); 53-64.

Jaime, A, Gordoni, M, Mosca, J and Vinck, D, (2006), *From Quality Management to Knowledge Management in Research Organisations,* International Journal of Innovation Management, Vol. 10, 197 – 215.

Kingsman, B.G., Hendry, L., Mercer, A. and De Souza A (1996)., *Responding to customer enquiries in make-to-order companies – problems and solutions*. International Journal of Production Economics. 46/47, 219-231

Kingsman, B.G., Lee, W., Hendry, L.C., Alan, M., Elaine, W (1993)., *Integrating marketing and production planning in make-to-order companie*s. International Journal of Production Economics, 30/31 ;53-66.

Klashner, R and Sabet, S (2007), *A DSS design model for complex systems; lessons from mission critical infrastructure,* Decision Support Systems; Vol. 43; 990 - 1013

Liu, S and Young, R.I.M (2007), *An exploration of key information models and their relationships in global manufacturing decision support,* Proc. IMechE, Vol. 21, Journal of Engineering Manufacture, 711-724

Miller, W. L and Morris, L (1999) *Fourth Generation R&D: Managing Knowledge, Technology and Innovation*, Wiley, N.J

Mustafa, O and Robert, B (2003), *A Knowledge-Based Decision Support System for The Management of Parts and Tools in Flexible Manufacturing Systems*, Decision Support Systems, Vol. 35, 487-516),

Nonaka, I and Takeuchi, H (1995), *The Knowledge Creating Company: How Japanese Companies Create the Dynamics of Innovation*, Oxford University Press, New York.

Oduoza, C.F. and Xiong, M.H (2007), *A decision support system framework to process customer order enquiries in SMEs,* International Journal of Advanced Manufacturing Technology, Vol. 42, 398-407

Oduoza, C.F, Harris, A, and Al-Ashaab, A (2010), *Knowledge based framework used in decision support,* Manuscript in preparation for International Journal of Manufacturing Technology Management

Oliviera, C and Antunes, C.H (2007), *Multiple objective linear programming models with interval coefficients – an illustrated overview*, European Journal of Operational Research, Vol 181;1434-1463

Olumolade, M and Norrie, D (1996), *A decision – support system for scheduling in a customer – oriented manufacturing environment*, Integrated Manufacturing Systems; 7/3; 38-46

Parente, D.H., Pegels, C.C., Nallan S (2002)., *An exploratory study of the sales-production relationship and customer satisfaction*, International Journal of Operations & Production Management; 22(9); 997-1013

Polacek, M, Doerner, K.F, Harti, R.F., Kieche, G, and Reimann, M (2007), *Scheduling periodic customer visits for a traveling salesperson*, European Journal of Operational Research; Vol 179; 823-837

Polanyi, M (1966), *The tacit dimension*, London: Routledge and Kegan Paul

Power, D.J, Sharda, R, *Model –driven decision support systems: concepts and research directions 2007;* Decision Support Systems, Vol 43, 1044-1061

Sheikh, K., *Manufacturing Resource Planning (MRPII) with an introduction to ERP, SCM, and CRM*. McGraw- Hill; 2003.

Small, C.T, Sage, A.P, (2006), *Knowledge Management and Knowledge Sharing: A Review, Information* Knowledge Systems Management, 5, 153 – 169.

Swan J, Newell S, Scarbrough H, Hislop D (1999). *Knowledge management and innovation: networks and networking* Journal of Knowledge Management; 3(4).

Stadtler, H., Kilger, C.(2000), *Supply Chain Management and Advanced Planning – Concepts, Models, Software and Case Studies*, Springer Press.

Xiong, M.H., Tor, S.B., Khoo, L.P., Bhatnagar, R., and Venkat, S.(2003a) *Framework to managing customer enquirers for SMEs*; Singapore-MIT Alliance Symposium, Singapore.

Xiong, M.H., Tor, S.B., Bhatnagar, R., Khoo, L.P. and Venkat, S (2006)., *A DSS approach to managing customer enquiries for SMEs at the customer enquiry stage*, International Journal of Production Economics; Vol. 103; 332-3461

Yen, D.C., Chou, D.C., and Chang, J (2002)., *A synergic analysis for Web-based enterprise resource planning systems*. Computer Standards and Interfaces; 24(4); 337-346.

Young, R.I.M, Gunendran, A.G, Cutting-Decelle, and Gruninger, M (2007), *Manufcaturing Knowledge Sharing in Product Life Cycle Management a Progression Towards the Use of Heavy Weight Ontologies*, Internatnional Journal of Production Research, Vol. 45, 1505-1519

Zorzini, M, Corti, D, and Pozzetti, *A, Due date (DD) quotation and capacity planning in make-to-order companies: Results from an empirical analysis*, International Journal of Production Economics, 2008, Vol. 112, 919-933

20

Decision Support for Web-based Prequalification Tender Management System in Construction Projects

Noor Maizura Mohamad Noor and Rosmayati Mohemad
Universiti Malaysia Terengganu
Malaysia

1. Introduction

In the ninth Malaysian Plan, the government attempts to optimize the use of Information Communication and Technology (ICT) infrastructure in delivering information to the people (Unit, 2006). Web-based application is being used as a medium to distribute useful information to the people in the most effective way. Most of the public sectors in Malaysia are utilizing Web system in their daily practices, but in several complicated processes, they are still using conventional way in processing the related documents. One of the examples is tendering processes in construction industry.

Tendering processes in construction industry normally will consume longer time to process all tender documents and there are some standard procedures to be followed. Standard procedures for construction in Malaysia are underlined by Department of Work (JKR). As the government is moving towards electronic tendering for the construction industry at the national level, most of public sectors publish open tender advertisements in government Web services but application to handle electronic tendering processes in construction industry has not been completely developed yet. Thus, the construction industry remains to be the most complex and fragmented industry in Malaysia (Weng & Alsagoff, 2006).

Prequalification is an initial phase in tendering processes where all tender documents will be screened out to select the compliance contractor. Prequalification is important as it can minimize risk of construction projects. In this phase, the contractor has to fulfill the tender documents according to the client request. It also provides an analysis on the tender documents by filtering the qualified contractors that have completed all tender documents properly.

Since some of tendering processes are confidential and involves many parties such as clients, consultants and constructors, we proposed Web-based application for prequalification tender management system in construction project or known as PreQTender to facilitate these processes. The proposed model is expected to benefit in terms of security of tender documents, reducing tender administration in terms of workload and paperwork, thus increasing productivity and efficiency in daily responsibilities. Furthermore, handling documents electronically is also to ensure fair and transparency processes.

The PreQTender provides an automated decision making process that reduces the use of manpower and processing time of the tenders. Web-based Decision Support System (DSS) has reduced technology barriers and made it easier and less costly to serve decision relevant timely information to the client wherever and whenever they may need it (Delen et al., 2007). The PreQTender is being developed to support contractor selection process.

This paper is organized as follows. First section described the current practices of tendering process. Second section continues with literature review. Third section is explanations of problem statement. Fourth section illustrates the research framework of PreQTender. Fifth section is expected result and lastly ends with conclusions.

2. Literature review

2.1 Current practices of tendering processes - background

Tender is an invitation to offer for an item or work. According to Du et al., tendering process is an invitation to those relevant parties to make an offer to the principal, which must be capable of accepting the offer, thereby creating a legally binding contract (Du et al., 2004). Principal is any party inviting and receiving tenders whilst the client may include a contractor. Tenderer is any party whose submitting tenders, including contractor, subcontractor and supplier (Du et al., 2004).

Tenders can be divided into several types such as open, restricted and negotiation tender. Open tender is a tender that offered to any interested contractors. Negotiated tender is carried out under special circumstances whereby is done when the clients need some expertise that capable in doing such projects. Restricted tender is applied when purchase authority has procedure in place which will definitely lead to the award of a contract. The procedure is restricted because contractors are first invited to express an interest and those contractors that have qualified against certain criteria are invited to submit tender (Du et al., 2004). Tendering processes is a complex process. A typical one involves lots of business procedures such as tender specification preparation, tender advertisement, tender aggregation, tender evaluation, tender awarding, and contract monitoring (Ng et al., 2007).

2.2 Prequalification tendering process

Prequalification tendering process is to identify qualified constructors based on some criteria as requested by clients. Besides that, this stage also certifies all the prequalification documents that are required to be submitted by constructors. Prequalification stage is generally preferred by clients to minimize the risks and failures. It also will enhance the performance levels of selected contractors. (Palaneeswaran & Kumaraswamy, 2001).

The prequalification tendering practices are different between countries according to the rules, regulations and procedures to be followed. In the study done by Palaneeswaran and Kumaraswamy, they examined several prequalification practices in Hong Kong, Australia and USA (Palaneeswaran & Kumaraswamy, 2001).

Construction Industry Development Agency (CIDA) in Australia has recommended three categories of prequalification criteria namely mandatory, additional and reserved. Technical, financial, quality assurance, time performance, human resource management, skill, occupational health and safety are considered as mandatory whilst claims performance and research development are regarded as additional and reserved. However, construction industry in Hong Kong has identified several different prequalification criteria such as experience, corporate, workload, support functions, resources and facilities. Whilst in USA, different public clients use various prequalification ratings.

Other research has been done by Topcu where the study has determined four main prequalification criteria in construction industry in Turkey including ability to complete projects, expertise, experienced technical staff and resources (Topcu, 2004).

2.3 Scenarios of electronic tendering system

The transparent procurement procedure powered by the Electronic Tendering System has been regarded as one of the most important merits embedded in the digital government. Taiwan government demonstrates the framework of electronic tendering system which in turn benefit both government bodies and vendors in terms of time and cost barriers (Liao et al., 2002). This framework is benefit as it simplifies original tendering process where the tasks of obtaining, submitting and opening tender are done via Internet. Continuation to this, in December 2002, the Public Construction Commission statistical reports showed that over 100,000 public agencies have utilized the Electronic Tendering System to upload the procurement documents and over 200,000 firms have attempted to supply construction projects, goods, and services by downloading the documents via the Electronic Tendering System (Chu et al., 2004).

Besides that, Indian government is operating The Indian Government Tenders Information System as the central Source for tenders offered by the Central and State Governments and other public bodies across India (Goverment, 2008). Kajewski et al. have reported on electronic tendering for construction industry in Australia, United Kingdom, United States of America (USA) and Canada (Kajewski et al., 2001). The report shows that electronic tendering has provided different facilities between countries even though it has similar functions. To make the Web-based more effective, integration between Web-based and DSS technology is implemented.

2.4 Web-based DSS:A new transform technology

Web-based has been used widely as it plays an important role in distributing information. Web technology enables the user to access a wide array of information and do transaction processing services easily at any places. Web is a platform of choice for building DSS.

DSS can be categorized into five different types such as Knowledge Driven, Communication Driven, Data-Driven, Document Driven and Model-Driven (Noor et al., 2006, Power and Kaparthi, 2002). In this research, it concentrates on Communication Driven and Model-driven. The integration of DSS and Web technology become popular as its gives a lot of benefit and make it more convenient to the user to access the DSS technology through the Web environment (Bhargava et al., 2007). DSS is a complex system that helps analyze decisions or choose between different options. A model-driven DSS places emphasis on statistical analysis, financial optimization or simulation. These are used by managers and staff members of a business, or people who interact with the organization, for a number of purposes depending on how the model is set up such as scheduling, decision analyses and others. DSS can be deployed via software or hardware in stand-alone personal computers, client or server systems, or the Web.

Prequalification tendering with Web-based Decision Support System (DSS) is a new transform technology that attempts to apply from manual tendering processes to electronic tendering system in order to make the process more efficient and effectives. Due to the growing interest in the Web, there are many on-going efforts to develop and implement Web-based DSS in various areas, such as health care, private companies, government, and education (Bhargava et al., 2007).

2.5 Advantages of using web-based DSS

Recently, both electronic business and electronic government are increasing their demands for more online data analysis and decision support. A Web-based DSS is a complex software system. It may integrate multidisciplinary data sources and related tools to generate value-added information to support decision-making (Zhang & Goddard, 2007).

One of Web-based DSS benefit is it can reduce management cost. According to Power, using Web-based DSS can reduce cost of operations, administration support and maintenance as well as end user training cost (Power & Kaparthi, 2002). Another advantage of Web-based DSS is it provides a trust security services. Security services have been defined for electronic tendering system with consideration for its legal nature where only authenticated parties will be allowed to access the system (Betts et al., 2006, Du et al., 2004). Moreover, Web-based DSS does not require any specific support from additional software and it is also more accessible and provides an interactive and unique interface.

2.6 Decision making models in tendering processes

Tendering process has multi criteria to be considered where each element will be considered as an important element in selecting qualified constructors. These criteria can be divided into qualitative and quantitative attributes.

There are several published models for selection process such as Analytic Hierarchy Process (AHP), Artificial Neural Networks (ANN), Multi Attribute Analysis, Multi Attribute Utility, Case-based Reasoning (CBR), Fuzzy Set Prequalification, Knowledge Based System (KBS), Dimensional Weighting Aggregation (DWA) and PERT model (El-Sawalhi et al., 2007, Holt, 1998).

AHP has been widely adopted to support multi criteria decision. Al-Dughaither has demonstrated prequalification multi criteria decision making model using AHP (Al-dughaither, 2006).

AHP is the most powerful and flexible weighted scoring decision making process to help people set priorities and make the best selection when both qualitative and quantitative aspects of a decision need to be considered (Cziner et al., 2005). AHP allows group decision making where group members can use their experience, values and knowledge to break down the contractor prequalification problem into a hierarchy and solve it by the AHP steps (Banaitiene & Banaitis, 2006).

By using AHP, it allows the decision makers to break down a decision into smaller parts, proceedings from the goal to criteria to sub criteria down to the alternative courses of action. AHP has been used to overcome the difficulties of the prequalification process. It is a sophisticated structured mathematic procedure and it is easy to implement for different applications (Al-dughaither, 2006).

3. Problem statement

In Malaysia, many of the public sectors offer a tender advertisement on the Web, but all documents related during tendering processes are still handled manually. In the manual tendering processes, the client hires a consultancy team made up of experts, such as architects, designers, project managers, quantity surveyors and other construction expertise, while contracting the construction to the selected contractor (Palaneeswaran & Kumaraswamy, 2000).

Manual tendering processes can be long, in term of duration, often taking three month or longer, which is costly for both contractor and supplier organizations (Ng et al., 2007). Weng and Alsagoff have identified many problems and issues faced in Malaysia construction industry and several are listed here such as problems in authenticating contractor status, insufficient copies of documents available due to high demand, voluminous tender documents to be fulfilled by constructors, incomplete information/documents, delay due to corrections and amendments before issuance of documents, voluminous documents to vet through, possible for leaking of restricted information of tender documents, possible mix up of documents, problems in issuing and collecting addendums, lack of information for decision making, inconsistency of tender evaluation and uncertainties in validity of information used in tender evaluation (Weng & Alsagoff, 2006).

According to Ng et al., the large volume of papers needs a lot of manpower to arrange tender documents (Ng et al., 2007). This tendering process uses a lot of space to store the tender documents and it usually costly to both client and contractor. The public tendering processes imposed by the government, are aiming at reducing the possibility of waste and abuse of public money (Hameri & Nordberg, 1999). Preparing tendering documentation and conducting tender obtaining processes requires much labor which is costly for suppliers. The management of paper-based documents as product samples and confidential information presents an obstacle (Kajewski et al., 2001, Liao et al., 2002).

Constructors are required to face several crucial phases in tendering processes before awarding a contact. The phases are including tender specification preparation, tender advertisement, tenderer prequalification, tender aggregation, tender evaluation, tender awarding, contract monitoring and others. Conventional prequalification process takes longer time in processing the tender documents. According to Russell, prequalification is the process of screening contractors where the minimal capabilities below which any potential contractors would not be considered for the evaluation phase (Russell, 1992). It is a process of evaluating and determining the competency of companies that appear qualified to perform construction services that meet the client's expectations for such services. The prequalification procedure is often chosen to minimize risk (Topcu, 2004).

The selection of qualified constructors gives confidence to clients in terms of selecting eligible contractor who is believed to achieve the project goals (El-Sawalhi et al., 2007). Contractor selection is a critical activity that plays a vital role in the overall success of any construction project (Liao et al., 2002, Palaneeswaran & Kumaraswamy, 2001). In the construction project, the selection of an appropriate contractor is the most critical for project success (Banaitiene & Banaitis, 2006). To choose the right person for the right project requires a right decision making selection of main contractors for construction work. Contractor selection is one of the main decisions made by the clients. Objectives of the tender system in construction even in the global contexts remain unchanged, that is to devise a most efficient framework to select capable contractors who can complete the construction project within set parameters of time, money, and quality (Alsagoff & Weng, 2006).

Prequalification is also important in the construction process as it is ensuring that the project is built on time, within the budget and based on quality level the owner expects. All the tenders that submitted by the contractor will be checked for the completeness of the documents required, documents verification, registration validation, and mandatory requirement.

This research proposes a design for prequalification tendering process with the aim of using Web technology as a medium to distribute information to the contractors. The objective of this research is to analyze, design and develop an overall framework of the prequalification tendering processes, namely as PreQTender, and finally test and evaluate the reliability of Web application for prequalification tendering processes in order to make it works properly on real system. This research focuses on tender for construction projects in Malaysia and manages the prequalification tendering process such as obtaining and submitting tender process and also covers the process of selecting the potential contractors that eligible to be evaluating into the evaluation stage. An overall workflow has been created in order to see the prequalification tendering processes clearly.

4. Research framework

4.1 Workflow

Figure 1 shows the overall workflow of PreQTender tendering process. The process starts with invitation to tender (ITT) by client until the selection of potential contractors. Each interested contractor has to submit all tender documents as requested by client. Contractors are allowed to update application before completely submitting tenders in between time given. Addendum is a condition where contractor's tender has corrections or modification. PreQTender checks prequalification for each contractor who has successfully submitted the tender based on specific criteria as stated by client. Then, PreQTender will evaluate the compliant contractors and finally generate short-listed of qualified contractors before go into evaluation stage.

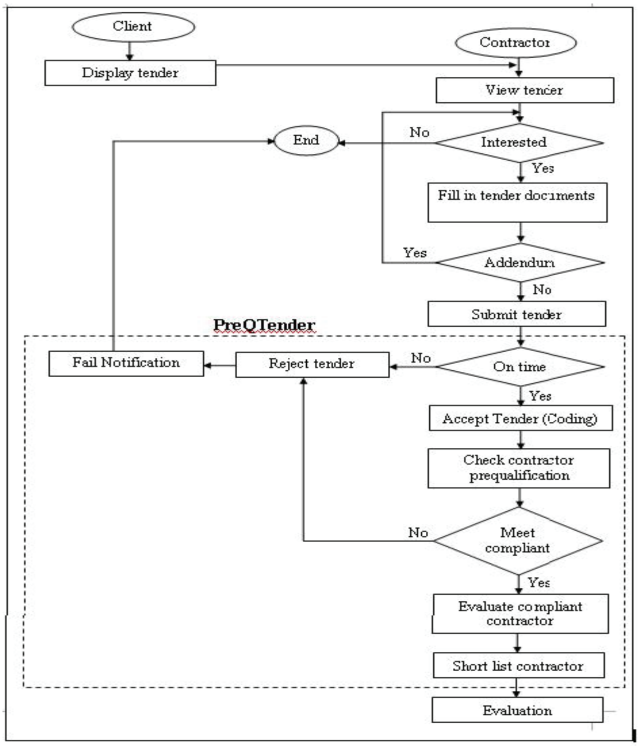


Fig. 1. Overall Workflow of PreQTender

Figure 2 shows the workflow of the contractor selection process. Before short listing the potential constructors, PreQTender will check and control every tender document received (form A-GA) to ensure the completeness of information and documents. The compulsory documents such as copy of bank statement and company account audited for last two years will be thoroughly checked by PreQTender as well as supporting documents. For the supporting documents the contractor has to enclose a copy of bank report of financial records, a letter certified completeness of work for each projects involved, Employees Provident Fund (EPF) for every technical staff of the company, academic qualification of each technical staff and report of current works performance for each project.

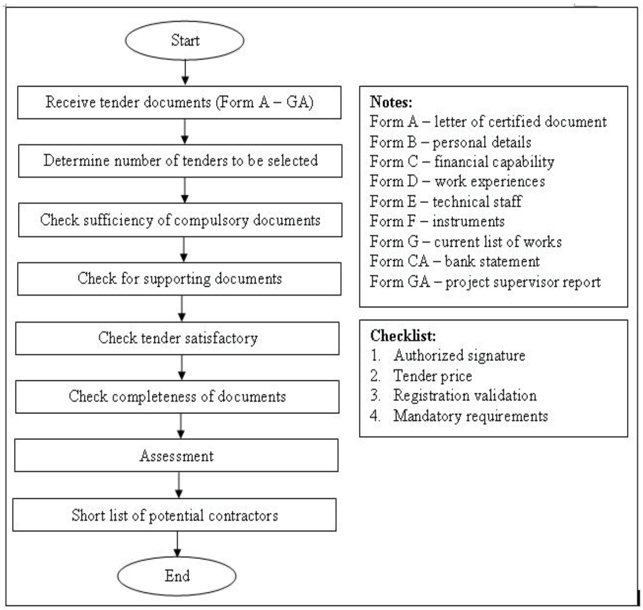

Start
Receive tender documents (Form A – GA)
Determine number of tenders to be selected
Check sufficiency of compulsory documents
Check for supporting documents
Check tender satisfactory
Check completeness of documents
Assessment
Short list of potential contractors
End
Notes:
Form A – letter of certified document
Form B – personal details
Form C – financial capability
Form D – work experiences
Form E – technical staff
Form F – instruments
Form G – current list of works
Form CA – bank statement
Form GA – project supervisor report
Checklist:
1. Authorized signature
2. Tender price
3. Registration validation
4. Mandatory requirements

Fig. 2. Work flow of Contractor Selection Process

4.2 Framework

Figure 3 depicts the framework of PreQTender where it consists of five main modules. The modules are preparation of tender documents, obtaining of tender documents, submission of tender documents, opening tender documents, and prequalification tendering analysis. Client involves in preparation and opening of tender documents while contractor involves in obtaining and submission of tender documents. Before the contractors enter to the obtaining module, they are required to make certain payment.

4.3 Software architecture

The World Wide Web and database technologies have been employed to facilitate the functions of PreQTender. Figure 4 shows the software architecture of PreQTender. World-Wide-Web browser is used to access the Web server via the HTTP protocol.

Web-based DSS support three quarter architecture which a Web browser sends HTML request using the hypertext transfer protocol (HTTP) to a Web server. The Web server

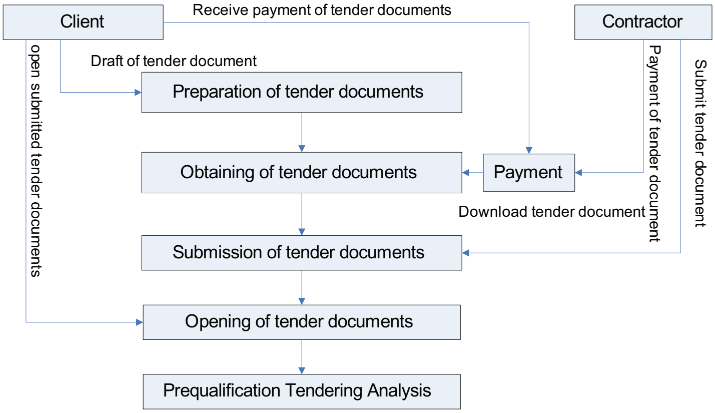

Client
Receive payment of tender documents
Contractor
Draft of tender document
open submitted tender documents
Preparation of tender documents
Payment of tender document
Submit tender document
Obtaining of tender documents
Payment
Download tender document
Submission of tender documents
Opening of tender documents
Prequalification Tendering Analysis

Fig. 3. Framework of PreQTender

processes these requests using a common Gateway Interface Script (CGI) script. The CGI handles model processing, SQL generation, post-SQL processing, HTML formatting. Application server sends requests to a modelling program or a database server. Tools like Java and JavaScript are improving the display of results and the interactive analysis of data and models. In order to require user interaction, scripting languages such as JavaScript and Hypertext PreProcessor (PHP) are used.

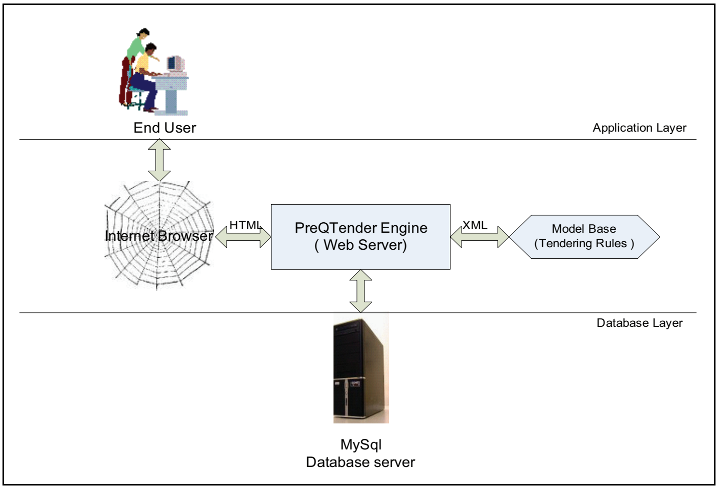


Fig. 4. Software Architecture of PreQTender

Prequalification tendering engine is a collection of software procedure written in PHP and hosted on Apache Web server. This Web server will retrieve the information from the model-based in doing the decision making for contractor selection. Model-based is the part where DSS element is embedded here. This model supports the decision making process where all the tender documents will be processed before storing them into the database. Then all the information will be stored in MySQL Database. The results of the selected contractors will be returned to the user's Web browser for display.

4.4 Phases in PreQTender

There are three phases involved in PreQTender as stated below.

Phase 1: Analyzing on Prequalification Tendering Process.

In this phase, analysis on prequalification criteria for tendering processes has been done in order to identify current problems and possible solutions. Analyzing is the process of understanding the problem and the requirements for a workable solution.

Phase 2: Designing and Developing PreQTender.

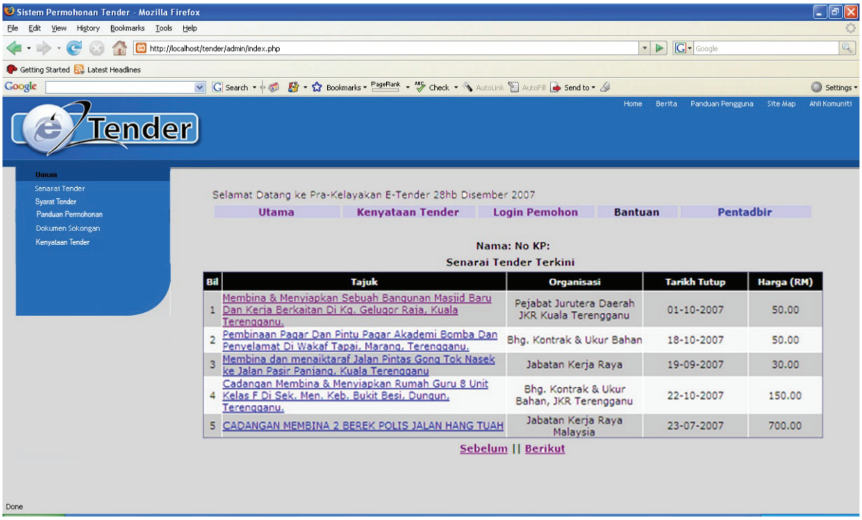
DSS is a concept that helps decision maker in processing large quantity set of tender documents that have been submitted by interested contractors. Web-based application is used to apply the conventional process to electronic format. Specific method and tools is used in developing PreQTender. In this phase, the application design is developed on the basis requirements, scopes and objectives that have been identified previously. It involves design framework and model for the system and validates it against requirement and present to the client for approval.

Phase 3: Testing and Evaluating PreQTender.

In this phase, the actual code based on the design is created and tested against requirements and test cases. The system will be tested in order to make sure the reliability of the system that has been developed.

5. Expected results

In this paper, we have presented a Web technology application with DSS to design the tendering process for construction projects. PreQTender will provide secure process and can be accessed only by the registered members. Currently, we are working on design and developing phase. Figure 5 depicts the list of available tenders while Figure 6 shows the interface that setting the priority of criteria. Client is responsible to determine this task.



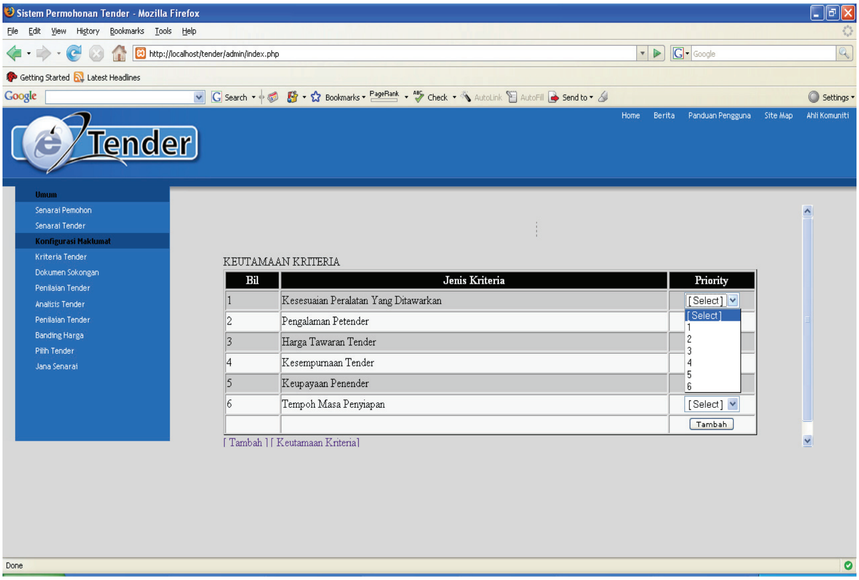
Selamat Datang ke Pra-Kelayakan E-Tender 28hb Disember 2007

Nama: No KP:

Senarai Tender Terkini

Bil	Tajuk	Organisasi	Tarikh Tutup	Harga (RM)
1	Membina & Menyiapkan Sebuah Bangunan Masjid Baru Dan Kerja Berkaitan Di Kg. Gelugor Raja, Kuala Terengganu.	Pejabat Jurutera Daerah JKR Kuala Terengganu	01-10-2007	50.00
2	Pembinaan Pagar Dan Pintu Pagar Akademi Bomba Dan Penyelamat Di Wakaf Tapai, Marang, Terengganu.	Bhg. Kontrak & Ukur Bahan	18-10-2007	50.00
3	Membina dan menaiktaraf Jalan Pintas Gong Tok Nasek ke Jalan Pasir Panjang, Kuala Terengganu	Jabatan Kerja Raya	19-09-2007	30.00
4	Cadangan Membina & Menyiapkan Rumah Guru 8 Unit Kelas F Di Sek. Men. Keb. Bukit Besi, Dungun, Terengganu.	Bhg. Kontrak & Ukur Bahan, JKR Terengganu	22-10-2007	150.00
5	CADANGAN MEMBINA 2 BEREK POLIS JALAN HANG TUAH	Jabatan Kerja Raya Malaysia	23-07-2007	700.00

Fig. 5. List of Tenders


Sistem Permohonan Tender - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://localhost/tender/admin/index.php
Google
Getting Started Latest Headlines
Google Search Bookmarks PageRank Check AutoLink AutoFill Send to Settings
Home Berita Panduan Pengguna Site Map Ahli Komuniti
eTender
Umum
Senarai Pemohon
Senarai Tender
Konfigurasi Maklumat
Kriteria Tender
Dokumen Sokongan
Penilaian Tender
Analisis Tender
Penilaian Tender
Banding Harga
Pilih Tender
Jana Senarai
Done

KEUTAMAAN KRITERIA

Bil	Jenis Kriteria	Priority
1	Kesesuaian Peralatan Yang Ditawarkan	[Select] [Select] 1 2 3 4 5 6
2	Pengalaman Petender	
3	Harga Tawaran Tender	
4	Kesempurnaan Tender	
5	Keupayaan Penender	
6	Tempoh Masa Penyiapan	[Select]
		Tambah

[Tambah] [Keutamaan Kriteria]

Fig. 6. Priority of Criteria

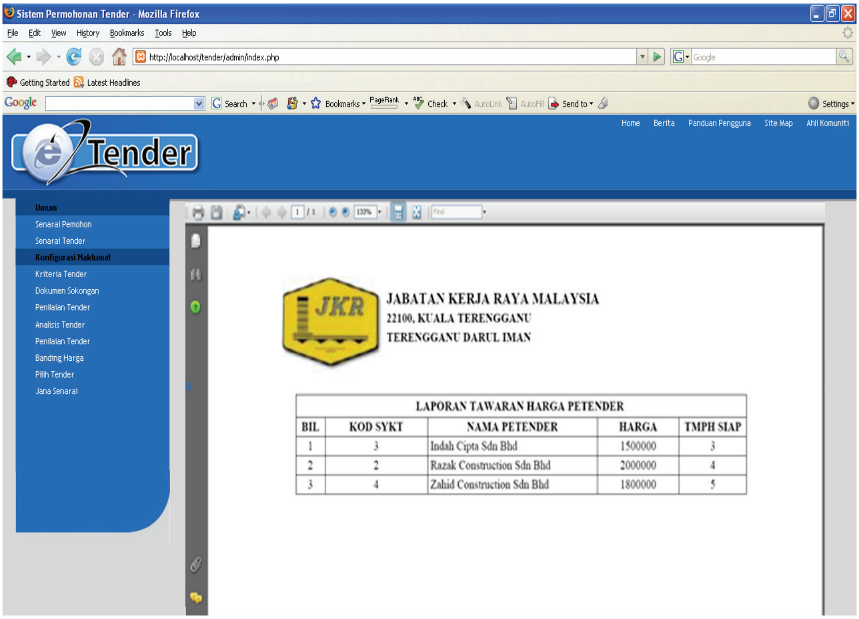
The PreQTender user interface has menu-data driven dialogues that offer a friendly environment for the user to perform the tasks that are currently available in the PreQTender. This user interface is also responsible for the overall control of the PreQTender, accessing and changing information with other menu within the PreQTender. The user interface has been implemented as a set of Web pages. The Web-based user interface enables the users to register, submit, retrieve process and manipulates data. Menus are created within each Web page. Registered users can perform operations on the project data by selecting the appropriate menu item from the Web page.

This new way of distribution information will provide better communications process between the contractor and client. Obtaining and submitting tender can be accessed easily via the Internet. The contractors also can manage their accounts and at any time. All of the information will be stored safely in database and it can prevent the data from lost or corrupted. Contactor selection process will be easier by using this prequalification process which takes shorter time and it can be conducted at any location. Figure 7 shows generated tender reports. This PreQTender will become more reliable, secured and in the end it will enhance the productivity on the client perspective.

The proper designed system can simplify the contractor selection and recent improvements in Internet technology and connectivity provide an opportunity to make the process of tendering for construction works more transparent and efficient.

6. Conclusion

This Web application for the tendering processes is hope to improve the manual tendering processes and yet make it convenient to the contractors and clients to manage the tenders. It is also will increase the integrity and transparency of the prequalification tendering processes.


Sistem Permohonan Tender - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://localhost/tender/admin/index.php
Getting Started Latest Headlines
Home Berita Panduan Pengguna Site Map Ahli Komuniti
eTender
Umum
Senarai Pemohon
Senarai Tender
Konfigurasi Maklumat
Kriteria Tender
Dokumen Sokongan
Penilaian Tender
Analisis Tender
Penilaian Tender
Banding Harga
Pilih Tender
Jana Senarai

JABATAN KERJA RAYA MALAYSIA
22100, KUALA TERENGGANU
TERENGGANU DARUL IMAN

LAPORAN TAWARAN HARGA PETENDER				
BIL	KOD SYKT	NAMA PETENDER	HARGA	TMPH SIAP
1	3	Indah Cipta Sdn Bhd	1500000	3
2	2	Razak Construction Sdn Bhd	2000000	4
3	4	Zahid Construction Sdn Bhd	1800000	5

Fig. 7. Tender Reports

In addition, this system integrates the structure of various Integrated Web-based DSS services, including tender invitation document preparation, uploading, tender obtaining, tender submitting, tender opening and tender analysis. Owing to this system, tendering efficiency is increased and the possibility for tender collusion is severely reduced. Contractor participation is thus encouraged and a nation's competitive ability is consequently increased.

7. References

Al-Dughaither, K. (2006) A Multicriteria Decision Making Model For Contractors Pre-Qualification. *Joint International Conference On Computing And Decision Making In Civil And Building Engineering*. Montreal,Canada.

Alsagoff, S. A. & Weng, E. L. C. (2006) Lines, Chains And Other (Restraint) Devices: Revisiting Porter's Value Chain With A Malaysian Case Study In The Implementation Of Electronic Tendering For The Construction Industry. *Proceedings Of The 10th Pacific Association Of Quantity Surveyors Congress Singapore.* Singapore.

Banaitiene, N. & Banaitis, A. (2006) Analysis Of Criteria For Contractor's Qualification Evaluation. *Technological And Economic Development Of Economy*, 12, 276-282.

Betts, M., Black, P., Christensen, S., Dawson, E., Du, R., Duncan, W., Foo, E. & Nieto, J. G. (2006) Towards Secure And Legal E-Tendering. *Journal Of Information Technology In Constructio*, 11, 89-102.

Bhargava, H. K., Power, D. J. & Sun, D. (2007) Progress In Web-Based Decision Support Technologies. 43, 1083-1095.

Chu, P.-Y., Hsiao, N., Lee, F.-W. & Chen, C.-W. (2004) Exploring Success Factors For Taiwan's Government Electronic Tendering System: Behavioral Perspectives From End Users. *Government Information Quarterly,* 21**,** 219-234.

Cziner, K., Tuomaala, M. & Hurme, M. (2005) Multicriteria Decision Making In Process Integration. *Journal Of Cleaner Production,* 13**,** 475-483.

Delen, D., Shardaa, R. & Kumara, P. (2007) Movie Forecast Guru: A Web-Based Dss For Hollywood Managers. *Decision Support System,* 43**,** 1151-1170.

Du, R., Foo, E., Boyd, C. & Fitzgerald, B. (2004) Defining Security Services For Electronic Tendering. *Proceedings Of The Second Workshop On Australasian Information Security, Data Mining And Web Intelligence, And Software Internationalisation.* Dunedin, New Zealand

El-Sawalhi, N., Eaton, D. & Rustom, R. (2007) Contractor Pre-Qualification Model: State-Of-The-Art. *International Journal On Project Management,* 25**,** 465-474.

Goverment, I. (2008) The Indian Goverment Tenders Information System.

Hameri, A.-P. & Nordberg, M. (1999) Tendering And Contracting Of New, Emerging Technologies. *Technovation,* 19**,** 457- 465.

Holt, G. D. (1998) Which Contractor Selection Methodology? *International Journal Of Project Managemen,* 16**,** 153-164.

Kajewski, S., Tilley, P., Crawford, J., Remmers, T., Chen, S.-E., Lenard, D., Brewer, G., Gameson, R., Kolomy, R., Martins, R., Sher, W., Weippert, A., Caldwell, G. & Haug, M. (2001) Electronic Tendering: An Industry Perspective. Queensland University Of Technology.

Liao, T. S., Wang, M. T. & Tserng, H. P. (2002) A Framework Of Electronic Tendering For Government Procurement: A Lesson Learned In Taiwan. *Automation In Construction,* 11**,** 731-742.

Ng, L. L. N., Chiu, D. K. W. & Hung, P. C. K. (2007) Tendering Process Model (Tpm) Implementation For B2b Integration In A Web Services Environment. *Proceedings Of The 40th Annual Hawaii International Conference On System Sciences* Ieee Computer Society.

Noor, N. M. M., Papamichail, K. N. & Warboys, B. (2006) An Integrated Web-Based Decision Support System For Tendering Processes. *Joint International Conference On Computing And Decision Making In Civil And Building Engineering.* Montreal, Canada.

Palaneeswaran, E. & Kumaraswamy, M. (2000) Contractor Selection For Design/Build Projects. *Journal Of Construction Engineering And Management,* 126**,** 331-339, .

Palaneeswaran, E. & Kumaraswamy, M. (2001) Recent Advances And Proposed Improvemnets In Contractor Prequalification Methodologies. *Building And Environment,* 36**,** 73-87.

Power, D. J. & Kaparthi, S. (2002) Building Web-Based Decision Support Systems. *Studies In Informatics And Control,* 11**,** 1-12.

Russell, J. S. (1992) Decision Models For Analysis And Evaluation Of Construction Contractors. *Construction Management And Economics,* 10**,** 185-202.

Topcu, Y. I. (2004) A Decision Model Proposal For Construction Selection In Turkey. *Building And Environment,* 39**,** 469-481.

Unit, E. P. (2006) Ninth Malaysia Plan 2006-2010. Putrajaya, Prime Minister's Department.

Weng, E. L. C. & Alsagoff, D. S. A. (2006) E-Readiness: How Ready Are We? The National E-Tendering Initiative - A Malaysian Experience. *Proceedings Of The 10th Pacific Association Of Quantity Surveyors Congress.* Singapore.

Zhang, S. & Goddard, S. (2007) A Software Architecture And Framework For Web-Based Distributed Decision Support Systems. *Decision Support Systems,* 43**,** 1133-1150.

21

Decision Support Systems used in Disaster Management

Marius CIOCA and Lucian-Ionel CIOCA
"Lucian Blaga" University of Sibiu
Romania

1. Introduction

The informational society is emerging as a new stage in the development of human society, by intense use of information in all fields of activity. The technological support of the new society is being built through the convergence of three major sectors: information technology, communication technology and digital content production. The development of new communication and information technology means is crucial to increasing competition, improving services and communication between institutions (Bizoi, 2007).

The initial concept of **Decision Support System (DSS)**, even though it was coined before the PC era, focused on the use of interactive calculation in semistructured decision-making (Alter, 2002).

The decision support systems are a distinct class of information systems. They integrate specific with general-use decision support information devices to form a constitutive part of the organizational global system (Filip, 2004).

In 1995, Clement identified four factors which determine the difficulty degree of the decision-making process (Hellstom & Kvist, 2003). The first, and altogether the most important factor is the *complexity of the problem*. The human factor has a limited capacity of perceiving and solving complex problems and, therefore, builds simplified mental models of real situations. Even if these models are applied in the best way possible, any simplification may lead to defective decisions. The second factor is given by the *uncertainty degree of the problem*, and the third is the fact that, in most cases, *several different objectives are set*. A certain decision may be right in the short run, but may prove wrong in the long run and vice versa. The last factor presented by Clement and which we should also consider refers to the *different conclusions that may be derived from different perspectives*, especially when several people are involved in the decision-making process.

In order to make good decisions, the decision maker must be well informed, must have access to high-quality models (from simple, implicit models to sophisticated mathematical models) and to "adequate" information. A decision support system may make all these conditions achievable (Hellstom & Kvist, 2003).

Considering the activities that the DSS supports, the elements of the decision-making model are (Demarest, 2005):

- a *decision maker* – an individual or a group responsible for making a particular decision;
- a *set of inputs of the decision-making process* – data, numerical or qualitative models for interpreting data, previous experiences with similar data sets or decisional situations

and diverse rules of a cultural or psychological nature, or constraints associated to the decision-making process;

- the *decision-making process* proper – a set of steps, which are more or less clearly defined, for transforming input data into output data as decisions;
- a *set of output data of the decision-making process*, including the decisions proper and (ideally) a set of evaluation criteria for the decisions which take into account the needs, problems or objectives at the root of the decision-making process.

1.1 A short history

According to (Keen & Scott, 1978), the concept of Decision Support System has emerged from two main areas of research: theoretical studies focused on organizational decision-making conducted by the researchers of the Carnegie Institute of Technology during the 1950s and 1960s and the technical work on interactive computer systems carried out at the Massachusetts Institute of Technology in the 1960s. The concept of Decision Support System became an area of research on its own in the mid - 1970s before gaining in intensity during the 1980s.

Executive Information Systems, Group Decision Support Systems and Organisational Decision Support Systems emerged in the mid - and late 1980s from single user and model - oriented Decision Support Systems.

According to (Aggarwal, 2001), the evolution of the DSS may be divided into four generations: the first DSS generation focused on data; the second DSS generation focused on improving the user interface; the third DSS generation focused on models and the fourth, the present-day generation, was obtained by introducing new analytical web-based applications.

As a short conclusion, the Decision Support Systems belong to a multidisciplinary environment, including database research, artificial intelligence, human-computer interaction, simulation methods, software engineering and telecommunications.

Thus, the concept of Decision Support Systems is an almost established concept, but which is still growing due to the integration (incorporation) of several individual and relatively newer technologies (object orientation, expert systems, advanced communications), from which it "extracts" new valences and strengths. Concurrently, the vitality of the concept is stimulated by the growing tendency of integrating processes and functions with all industrial systems, environment management systems, etc. (Filip, 2004).

1.2 Definitions

The definitions provided during the last 30 years for DSS show, according to (Keen, 1987), "both what DSS is and what it is not", with consequences on both the scientific basis, and the credibility of the decision support applications.

Essentially, a DSS is a computerized system which improves the activity of decision-makers situated on different levels in the chain of command (from supervision of different processes to leading positions in politics). At the same time, DSS stimulates the decision-maker to improve the decisional process and make the right decisions in order to obtain high and quickly visible performances (decision effectiveness) (Filip & Bărbat, 1999).

As early as 1980, (Sprague, 1980) observed that the initial definition of the Decision Support Systems – computerized interactive systems which support decision-makers in using data and models to solve unstructured problems was too restrictive, and thus, the definition was expanded to include any system involved in the decision-making process.

This expansion of the definition made the concept of Decision Support Systems an umbrella term for different types of systems, many of which having no connection with the initial idea of Decision Support Systems (Alter, 2002). If, initially, Decision Support Systems were instruments for large companies, today, they also address small companies too. These instruments have changed and will change considerably the way in which decisions are made. They enable the individual or organisational decision-maker to manage more effectively the volume and complexity of information and better co-ordinate activities.

1.3 DSS characteristics and functions

The characteristics specific to a DSS depend on the type of decision the systems have been designed for (Bellorini & Lombardi, 1998). However, numerous authors have suggested a series of "standard" characteristics any DSS should possess. Considering the results obtained by (Parker & Al-Utabi, 1986) after studying 350 sources on the same subject (Bellorini & Lombardi, 1998) and the essential characteristics emphasized by (Filip, 2004) we may synthesize a list of DSS characteristics:

- to provide support and improve, not replace, human reasoning; the user maintains control over the DSS at all times.
- to assist managers in the decision-making process connected with unstructured and semi-structured problems, which cannot be solved through simple reasoning and judgment, or through any other classes of information systems;
- to be flexible and adaptable in relation to the changes in the context of the decision and support as many (or even all) decision process stages as possible;
- to be focused on characteristics in order to make it more user-friendly to less proficient users (managers on all levels, a single decision-maker or a group) and not be limited to the computerisation of some methods of working used before the implementation of the system, but to facilitate and stimulate new approaches (to ensure support for a variety of decision processes and for different styles);
- to combine the use of analytical models and techniques with data access functions; the data and information in the system should be obtained from various sources;
- to improve the efficiency of the decision process, rather than its effectiveness, focusing on the increase in productivity and the quality, suitability and applicability of decisions, rather than on the time and cost of decision.

In conclusion, the main characteristics of a DSS are:

- it alleviates efforts, amplifies decision-makers' capacity and its purpose is not to replace them or transform them into mere agents who adopt mechanically solutions provided by the computer;
- its purpose is to approach semi-structured problems, in which sections of the analysis effort could be computerised, but the decision-makers use their own reasoning to control the decision process.

1.4 DSS classification

The systems that used to provide support in the decision process have been named by specialists Decision Support Systems or Decision Management Systems. Recently, terms such as artificial intelligence, data mining, on-line analytical processing, knowledge management have been used for systems whose objective was to inform and assist managers in the decision process (Muntean, 2003).

Because of the existence of a huge number of terms, which have caused many problems to DSS research, several criteria, have been proposed for a classification of Decision Support Systems (Suduc, 2007).

Undoubtedly, numerous DSS classifications have been developed in time, but we shall restrict below to those classifications which are enough relevant and encompassing to the subject in discussion.

Donovan and Madnick (1977), quoted by Turban (1998), divided DSS, according to the nature of the decisional problem, into two categories:

- institutional DSSs facilitate solving structured problems within an organisation;
- ad-hoc DSSs facilitate solving semi-structured problems, which are not usually anticipated.

Hackathorn and Keen (1981), quoted by Turban (1998), identified three categories of DSS:

- single-user DSSs;
- group DSSs;
- organisational DSSs.

Steven Alter, quoted by Muntean (2003) proposed in 1980 a classification of the Decision Support Systems according to "the degree to which the system's output can directly determine the decision", independently from problem type, functional area or decisional perspective. Thus, seven categories of Decision Support Systems were proposed, divided into two super-classes:

- Data-oriented DSSs
  - File Drawer Systems, whose purpose is to automate certain manual processes and provide access to data items. They address people who have operational responsibilities (operators, clerks, workshop supervisors). Currently, this category includes simple query and reporting instruments which access transactional systems;
  - Data Analysis Systems, which facilitate the analysis of current and historical data, in order to produce reports for managers. Data analysis is required for budget analysis, business opportunities analysis, investment effectiveness analysis, etc. Today, this category includes a large number of data warehouse applications;
  - Analysis Information Systems, which provide access to a multitude of support databases for the decisional process, as well as a series of simple models in order to supply information necessary for solving particular decisional situations. This category includes today the OLAP systems, frequently used in sales forecasting, competition analysis, production planning, etc.
- Model-oriented DSSs
  - Systems oriented on Accounting and Financial Models. The models employed are "what-if" and "goal-seeking" and they are frequently used in producing profitability estimates for new products, estimative balances, etc.
  - Systems oriented on Representational Models, which use simulation models to estimate consequences; they are used extensively in risk analysis, in production simulation etc.;
  - Systems oriented on Optimisation Models which help producing optimal solutions for different activities;
  - Systems oriented on Suggestion Models, which carry out the logical process that leads to a suggested decision for activities with a certain degree of structuring (such as determining the rate of updating insurance, models for the optimisation of bond supply, etc.).

1.5 Advantages and limitations

Filip (2007) identifies four advantages and six limitations, as shown below:

- advantages
  - direct (or intermediated) work with the decision support system may contribute to improving the individual's decisional capacity;
  - increase in work productivity by extending capacity of decision-makers to directly process information;
  - expanding decision-makers' individual capacities leads to improved decisions, as a result of a better analysis;
  - being an artificial object, the decision support system is objective and impartial;
- limitations
  - the system lacks human traits: creativity, intuition, imagination, responsibleness or the instinct of self-preservation;
  - because of hardware and software limitations, there could be consequences which lead to insufficient qualities (regarding correctness and completeness) of knowledge accumulated within the system and in the limited possibilities of communication between decision-maker and the DSS;
  - in order to be effective and efficient, the system must be designed with a specific purpose in mind, for a specific field of use and a specific type of relative decision problems;
  - the DSS is designed as a component part of the global computer system of the organisation, from which it derives the necessary data. Thus, there may be compatibility problems between computer systems;
  - terminological issues and problems related to the significance of certain aspects approached by DSS may arise because of the cultural differences between developers and users;
  - the system may be used only partially and terminological issues may arise if the system documentation is cumbersome or poorly structured.

1.6 Disaster prevention DSSs

DSSs are extensively applied in environmental protection. They are used in pollution control, in water resources management and rationing, in flood control and forecasting, in agriculture for pest control, in forestry, in the prevention of epidemic diseases, etc.

The following are examples of systems used in activities related to ensuring the balance of ecosystems and in environmental protection:

- the TELEFLEUR (TELEmatics-assisted handling of FLood Emergencies in URban areas) funded by the European Commission, for the development of an operational system for the prevention and management of floods, which combines telematic technologies with advanced meteorological and hydrological forecasting encapsulated in a decision support system. The DSS was tested in the following areas: Liguria, Italy and Greater Athens, Greece;
- in Italy, the "Dipartimento di Informatica" (the Department of Informatics), at the "La Sapienza" University in Rome, have developed a DSS for flood control and prevention, based on Web technologies. The computer system has a distributed architecture, collecting data from distant sources. The decisional system simulates scenarios using the collected data and makes quantitative and qualitative predictions. The system also

provides a decision risk analysis for flooded areas. The DSS integrates an expert system in its architecture which uses experience and data accumulated from previous similar situations to make decisions;

- the L-THIA (Long-Term Hydrologic Impact Assessment) developed by Purdue University, United States of America, is one of the best systems used for monitoring and controlling the hydrological impact of climate change. It is integrated with GIS, a database management Oracle system and special user interfaces designed for users who are not very familiar with decision support computer systems. The data is collected through Web technologies, with PERL codes. The system provides the user with hydrologic maps which may be used to analyse the current situation and simulate hydrologic flow control scenarios;
- another DSS used in environmental protection, air and soil pollution control has been developed by University of Ljubljana, Slovenia. The proDEX system is developed in Python. It is dedicated to complex environment pollution issues, integrates with relational distributed databases and uses GIS in its architecture;
- in Chinese universities, the departments of informatics teach courses on DSSs, focused on different areas of application. Thus, the Shanghai Jiao Tong University focuses on developing Decision Support Systems for durable development and environmental protection. These Decision Support Systems integrate with artificial intelligence technologies – intelligent agents.
- research in the field has been carried out in Romania as well, as a result of the global effects which influence the environment and which produced major damages between 2004 and 2006 in our country. The "Lucian Blaga" University of Sibiu has developed a system, funded by the state budget, based on cross-platform (UNIX, Windows, etc.) Open Source technologies such as the PHP application server, the MySQL database server, and the Apache Web server. The purpose of this project was to create a flood (disaster) warning system, which will be presented as a case study in subchapter 4.

2. Decision support technologies and tools

The complexity of durable development issues require rational decisions, and decision-making is becoming increasingly difficult especially in the field of environmental protection. Due to the advances in decision theory and the study of decision support systems, new decision support methods and instruments have been developed. However, designing and building instruments able to assist the decision-maker in making decisions for complex issues is a highly demanding task (Boboşatu, 2008).

Using certain techniques and methods which are generally accepted in DSSs may contribute to the improvement of risk identification and prevention processes. These involve making strategic decisions with the synergetic contribution of different committees and groups of experts. Their activity is frequently hindered by physical, temporal and cognitive barriers. In addition, these methods and techniques used by DSSs are simply general recipes for approaching specific decisional situations. In practice, they need reinterpretation, refinement, adaptations and additions.

The use of a DSS for risk identification and prevention will enable the decision-makers to turn to the best account the intellectual capital needed for its application. As most decision-makers mainly focus on content and less on procedure, this is not as easy and well documented as it appears to be. Moreover, the use of a DSS contributes to the reduction of co-ordination malfunctions in the case of collective decision processes and facilitates the integration of intermediary reports obtained through the application, co-ordination and aggregation of the methods and techniques employed.

An important role in the development of DSSs is played by the concept of OLAP (On-Line Transaction Processing) with the technologies it is based on: ROLAP (Relational Online Analytical Processing), MOLAP (Multidimensional Online Analytical Processing) and HOLAP (Hybrid OLAP).

In MOLAP, data is stored in multidimensional cubes. The data is not stored in a relational database, but in proprietary formats. This technology has the following advantages:

- excellent performance (the MOLAP cubes are built for quick data interrogation and are optimal for "Slice and Dice" operations);
- ability to perform complex calculations (all calculations are generated the moment the cube is created, and thus results are obtained very fast).

ROLAP is an alternative to MOLAP (Multidimensional OLAP). While both analytical technologies, ROLAP and MOLAP, are designed so as to allow data analysis through a multidimensional data model, ROLAP is significantly different from MOLAP as it requires additional storage space and calculations. ROLAP instruments access data in the relational database and generate SQL interrogations to calculate information adequately, when required by an end-user (Filip, 2004).

ROLAP enables the user to create additional tables in the database (aggregate data tables) which sum up data in any desired size combination. The advantages of this technology are:

- scalability in handling a large volume of data, especially models with millions of members;
- the data is stored in relational databases that can be accessed through any SQL reporting instrument;
- ROLAP instruments are more performant in handling non-aggregate fact tables (for example, text descriptions), while MOLAP instruments are less performant when interrogating those elements.

The HOLAP technology tries to combine the advantages of ROLAP and MOLAP to obtain faster performances. When detailed information is needed, HOLAP allows "drill through" operations which retrieve data directly from the relational database.

The concepts of OLAP and data warehouse are complementary. The data warehouse collects the information needed by decision-makers starting from the data source and its main objective is to centralise decisional information by ensuring the integration of extracted data, their coherence and the preservation of their evolution. That is why, the implementation of data warehouses is based on ROLAP. The data warehouse maintains data integrity and feeds data storehouses.

Data storehouses are the result of extracting a part of the information in the data warehouse required by the decision process and are useful to a class of decision-makers for their specific analysis needs, case in which they are oriented on analysis subjects. The data storehouses efficiently support the OLAP analysis processes, and they are implemented by using the MOLAP technology (Boboşatu, 2008).

Usually, Web-oriented DSSs use a "three-tier" or "four-tier" architecture (Power, 2002) and enable a decision-maker to send a request through a Web browser (Internet Explorer, Netscape, etc.) to the Web server through HTTP (Hipertext Transfer Protocol). The Web server processes the request using a program, or a script and displays the result in the decision-maker's Web browser who placed the request. Web applications are designed to enable any authorised user to interact with them through a Web browser and an Internet (Intranet) connection. Usually, the code of the application is located on the remote server, and the user interface is displayed on the user's browser. The instruments for developing Web-oriented DSSs are still new and rather complex. Many decision-makers have heard of HTML, but this is just a small part of the multitude of instruments used in developing a DSS. In general, decision-makers are bombarded with terms and acronyms such as Web Server API (Application Programming Interface), Java applets and servlets, Java Script in HTML pages, ActiveX and Plugins,.NET components, etc.

3. The environment and disaster management

Man lives in an environment which is permanently exposed to a diversity of more or less dangerous situations, generated by numerous factors. Extreme natural phenomena such as: storms, floods, drought, landslides, earthquakes and others, in addition to technological accidents (severe pollution, for example) and conflicting situations, may influence directly the life of every person and that of society as a whole. These phenomena, also termed as catastrophes or disasters (or hazards, to use a geographical term), must be precisely known so that they could be dealt with promptly. The reduction and mitigation of the effects of such disasters require a thorough interdisciplinary study of hazards, of vulnerability and risk, as well as proper dissemination of information. Informatics is called upon to play its own role in this field.

In this context, *hazard* is the probability that a potentially dangerous phenomenon might appear and affect both the environment and human beings. Thus, hazard is a natural or anthropic phenomenon, harmful to the human being, whose consequences appear because of the fact that safety measures have been exceeded. Natural hazards are a form of interaction between man and the environment, in which certain adaptation limits of society are exceeded. The presence of human society is mandatory so that these hazards may be possible. If an avalanche takes place in Antarctica, for example, it is nothing else but a natural phenomenon. If the same phenomenon occurs in Făgăraş Mountains, where a cabin or a road is affected, we are facing a natural hazard.

Vulnerability emphasises the degree in which people and their possessions are exposed to hazards, it indicates the level of damages which a certain phenomenon may produce and it is expressed on a scale from 0 to 1, 1 meaning the total destruction of property and loss of human lives in the affected area. The destruction of the environment triggers an increase in vulnerability. For example, deforestation produce greater erosion and trigger landslides, faster and more powerful freshets, and an increased vulnerability for settlements, access ways and communication networks.

Risk is defined as the probability that people and their property be exposed to a hazard. Risk is the probable level of deaths, injuries, damages produced by a certain natural phenomenon or group of phenomena, in a certain place and time. The elements of risk are: population, property, access ways and communication networks, economic activities, etc. exposed to risk in a certain area.

Floods are natural phenomena and a component of the Earth's natural hydrologic cycle. Floods are natural phenomena which have always influenced the development of human society; they are the most common natural disasters on Earth and they produce the greatest number of deaths and the greatest damages all over the world. In the same time, floods determined people to change their approach to such natural disasters, from regarding them as a *caprice* of nature, to man's attempt to *fight* floods, to *defend* himself from floods and then to *prevent* floods.

Under the present-day circumstances, when profound climate changes have occurred and their effects such as floods and other disasters, a disaster/flood warning system such as the one described below in subchapter 4 is extremely useful (and complementary the other systems).

4. Decision support systems for disaster management. Case study

This subchapter presents a case study, the results of a research contract, funded by he state budget between 2007-2008, and carried out by the authors and the research team at the "Lucian Blaga" University of Sibiu.

Flood management is made easier by the fact that floods occur in predictable locations and in most cases warning is possible. That is why the project is very useful as it uses mathematical simulation and modelling in case of disaster and because it proposes an effective flood warning method.

The lifecycle of a disaster has three stages:

- the prevention/warning stage;
- the disaster stage;
- the post-disaster stage.

The project described below is included in the first stage, given its main objectives: warning the population through a fast flood warning system, "collecting" data and sending it to the dispatcher to be disseminated to the institutions in charge with such situations and the population in the affected area.

4.1 Specific objectives

Starting from a critical analysis of all other similar systems in the world and in Romania, we may infer that the theme dealt with substantially improves human knowledge in this field of study and brings several original elements through the development of a "fast flood warning system", as current warning strategies do not include such a solution, as well as through the data-collection method it presents. The present-day warning systems are the telephone, the fax machine, television, radio, etc. and not the SMS alternative we have developed in this project and which we are describing in this subchapter. This approach might be termed as a "relatively new" and viable alternative, given current circumstances when natural phenomena are becoming increasingly frequent and their consequences increasingly serious.

The research results are applicable to various institutions such as prefectures, disaster prevention county committees, county councils, disaster prevention local councils, civil protection county inspectorates, environmental protection county agencies, etc.

4.2 General presentation

Under disaster circumstances (water inrush, floods, etc.), the first systems that are going to "fail" are, in this order: TV cable, the power supply network, and possibly the mobile phone network. As shown below in figure 1, the system designed to inform the citizen on the imminence of a flood "short-circuits" – in general – between points 3 and 4 (5) and respectively 2-5. Thus, we consider our solution feasible and realistic as it provides a practical alternative to existing warning systems. (Cioca, M. et. al., 2008), (Cioca, M. et. al., 2009).

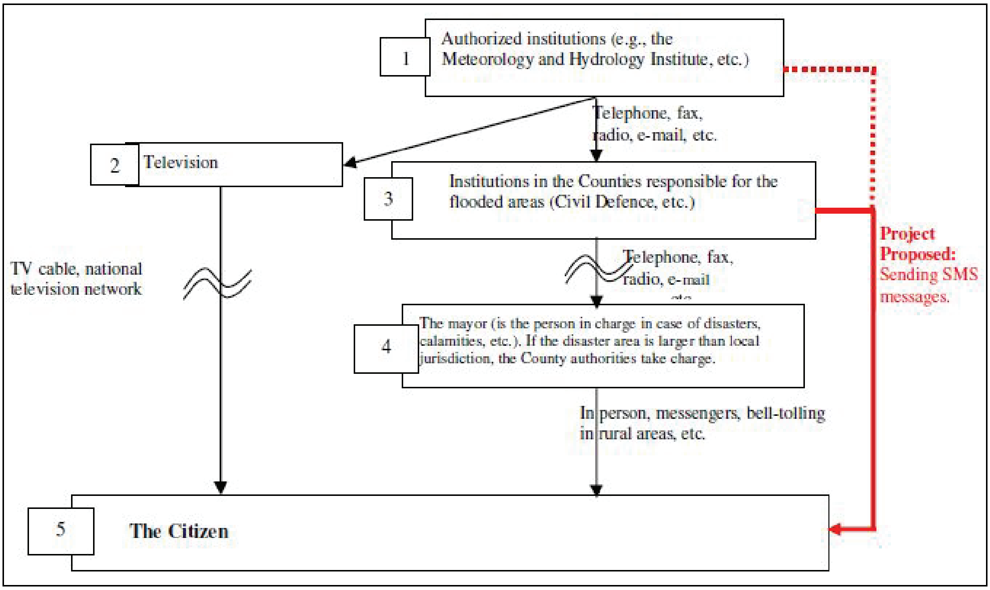

1
Authorized institutions (e.g., the Meteorology and Hydrology Institute, etc.)
Telephone, fax, radio, e-mail, etc.
2
Television
3
Institutions in the Counties responsible for the flooded areas (Civil Defence, etc.)
TV cable, national television network
Telephone, fax, radio, e-mail
4
The mayor (is the person in charge in case of disasters, calamities, etc.). If the disaster area is larger than local jurisdiction, the County authorities take charge.
In person, messengers, bell-tolling in rural areas, etc.
5
The Citizen
Project Proposed: Sending SMS messages.

Fig. 1. Disaster Warning System Scheme

The first part of the paper deals with the warning system – described briefly in section 4 – while its second part deals with the data collection and dispatch system (section 5).

The system constructed during the development of the project and presented in this present paper is viable also because mobile telephony has gained more subscribers than fixed telephony; some families own even 2 or 3 mobile phones so there is a huge chance that at least one member of the family receives the warning message.

In figure 1 (Cioca, M. et. al., 2007) (with a continuous line) the system is implemented at a regional (county) scale and may be easily expanded to a national one (with dotted line); if implemented at a national level, the system will radically reduce the number of situations when man is taken "by surprise" by raging waters.

In conclusion, in case of a water inrush, citizens may find themselves in one of the following situations:

- they are not informed on the matter;
- they are informed too late;
- they are informed in time (this has been rather uncommon in rural areas in between 2004 and 2008).

The warning system increases the chances that citizens might find themselves in the last of the three categories mentioned above.

The system:

- enables the users to send disaster-warning text messages;
- to the authorities (or directly to citizens);

- is as feasible and platform independent as possible;
- is easy to use and user-friendly to users with intermediate computer skills.

4.3 Technical specifications – dispatcher system

4.3.1 Fundamental architecture

Client-Server

The Client - Server architecture is one of the most commonly used in application development and it ensures the division of the application operation logical model into smaller functional units (Daconta et al., 2003), (Fensel et al., 2002), (Shadbolt et al., 2006), (Sheth et al., 2003).

Http (Https) Web Architecture

The application is based on a web platform. The advantages of using a web platform are:

- the fact that the communication protocol has already been implemented and tested on an international scale (HTTP/HTTPS);
- at the client level, no specialized software is required; a web browser is the only requirement; thus, the architecture can be implemented on a large number of systems with no particular configuration of these stations;
- system upgrading is a simple operation and requires just a single modification of the segments in the specified system.

Data transmission security is ensured by using the HTTPS, which provides a secure channel between the client station and the system residing on the server.

4.3.2 Technologies

Client

The client requires a minimum number of software applications, which,as we shall see below, are rather inexpensive:

- Web browser (Firefox, Internet explorer >5.5, Netscape) – to access the system residing on the server;
- the access to the host server (intranet, internet) – to ensure this interconnectivity for this station. Any communication media may be usually employed, from wire to wireless technologies;
- hardware requirements are those of the abovementioned software application.

Server

The server software requirements are the following:

- Operation System: Linux Based – better stability both in terms of security and performance;
- programming language: PHP 5.x;
- SGBD: MySQL 5.x;
- WEB Server: Apache 2.x.

The optimum system hardware will be specified only when the solution has been implemented on the host system. The higher the number of users, the more powerful should the hardware be to ensure optimum performance (Jeffery & Kacsuk, 2004), (Laszewski & Wagstrom, 2004).

Specialized Hardware

GSM communication is ensured by a GSM Modem and if the number of SMS messages is too high, the Bulk SMS Message service provided by a local mobile operator could be a good choice.

The Bulk SMS Message service is a service provided by certain mobile operators which p the interconnection between the local system and the mobile operator.

One of the drawbacks of such a system is the fact that the connection may be interrupted on account of implementation reasons (cable faults), but on the other hand this system is able to send a large number of SMS alerts.

4.4 System architecture

4.4.1 Basic architecture

The basic system architecture is presented in (Cioca et al., 2009).

The required software is presented in the (Cioca, 2008).

4.4.2 Detailed architecture

a. Risk Levels

Five levels of risk are defined within the system:

- Disaster;
- High risk;
- Medium risk;
- Low risk;
- Minor risk.

These levels are to be used on a regular basis in the entire warning process.

b. Dangerous Events

The following (main) dangerous events are defined within the system:

- Earthquake;
- Floods;
- Fire.

The above are just examples; other events may also be defined.

c. Regionalization

This process involves designating certain areas where dangerous events are likely to occur (villages, cities, etc.). This regionalization process divides a territory in order to make it manageable during the occurrence of external events that require supervision.

d. Scenarios

This process involves assigning one or more areas to a person in charge, depending on the type of dangerous event.

e. Users

There are several types of users:

a. General Administrator

This user is in charge with the entire system, its configuration, user management, etc.

The most important operations:

- User management;
- Operation Parameters Management (number of messages sent per time unit, etc.);
- Monitoring the warning log.

b. Local Operator

This user performs the most important task, i.e. to initiate alert process. This user selects the scenario, the area and this request is sent to the system which informs the persons in charge.

c. Personal Operator

This user updates the contact information (especially phone numbers, contact details, etc).

The basic scheme of the entire SMS warning application is presented in the figure 2.

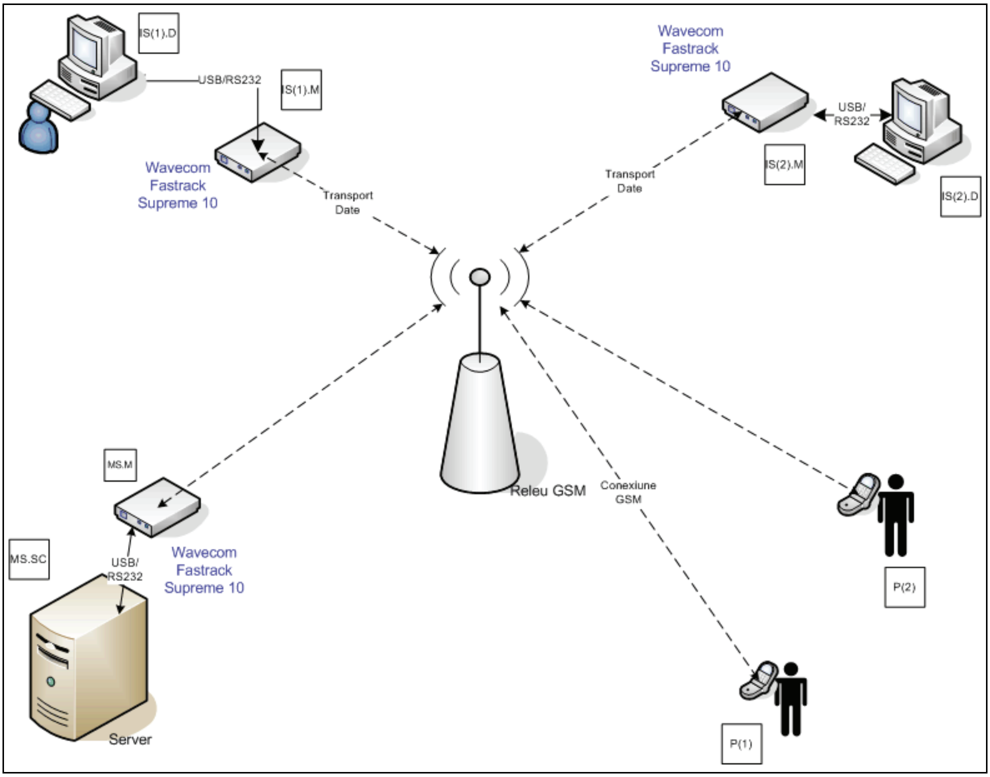


Fig. 2. Basic Scheme of the Disaster Warning Process (the dispatcher sends SMS to people in danger)

4.5 Communication between the system and the GSM environment

4.5.1 General details

In this implementation, we have used a Fastrack Modem M1306B:
General product specifications:

- the best GSM/GPRS connectivity characteristics;
- internationally tested solution;
- 900/1800 Mhz dual band;
- AT interface, provides connectivity to a wide range of equipment;
- hardware connection: RS232 cable.

4.5.2 General presentation

The characteristics are presented in detail below:
Standard access to environment:

- 900 Mhz
- E-GSM, ETSI GSM

GPRS compatible:

- Class 10
- PBCCH support
- Coding schemes CS1 to CS4

Interfaces:

- RS232(v.24/V2.8)
- Baud Rate: 300, 600,1200,2400,4800,9600,19200, 38400,57600,115200
- 3 v SIM Interface
- AT Command set V.25 and GSM 07.05 & 07.07
- Open AT interface form embedded application SMS:
- Text & PDU
- Point to Point (MT/MO)

Audio:

- Echo cancelation
- Noise Reduction
- Telephony
- Emergency Calls

4.5.3 Communication protocol implementation; interface instructions

In order to transmit a text message (SMS) by using a modem the following instructions should be used on the serial interface. At this moment, just this modem function is to be used. The modem is able to perform many other tasks, but this is the only one implemented so far.

AT	Initiate model connection
OK	Result
AT+CMGF=1	Setting mode – SMS Mode
OK	Result
AT+CMGW=”+0740******”	Setting Number
>Text Sample	Message ends with ^Z
+CMGW: 1	Message Index
OK	Result
AT+CMSS=1	Send message
+CMSS: 20	Send message index
OK	Result

4.6 Technical specification – data collection system

4.6.1 Communication between the system and the GSM environment

The system has been conceived as a minimal system, in which active components are preset to known GPS locations (or locations of some other type).

Operation

Basically, the system functions as follows: it collects data, it sends it to the central station where the danger is rated. If the danger rate is higher than a default level, the system initiates an alert request for the staff in charge.

The basic scheme of the entire data-collection application is presented in the figure 3.

Component parts

The system contains the following entities:

- MS – Main Server or main data collection, processing, and emergency alert server;
- ISS1… n – individual static (known coordinates) data collection stations, i.e. stations operated by specialized personnel who observed and manually enters data into the system;
- P1…n – Personnel in charge

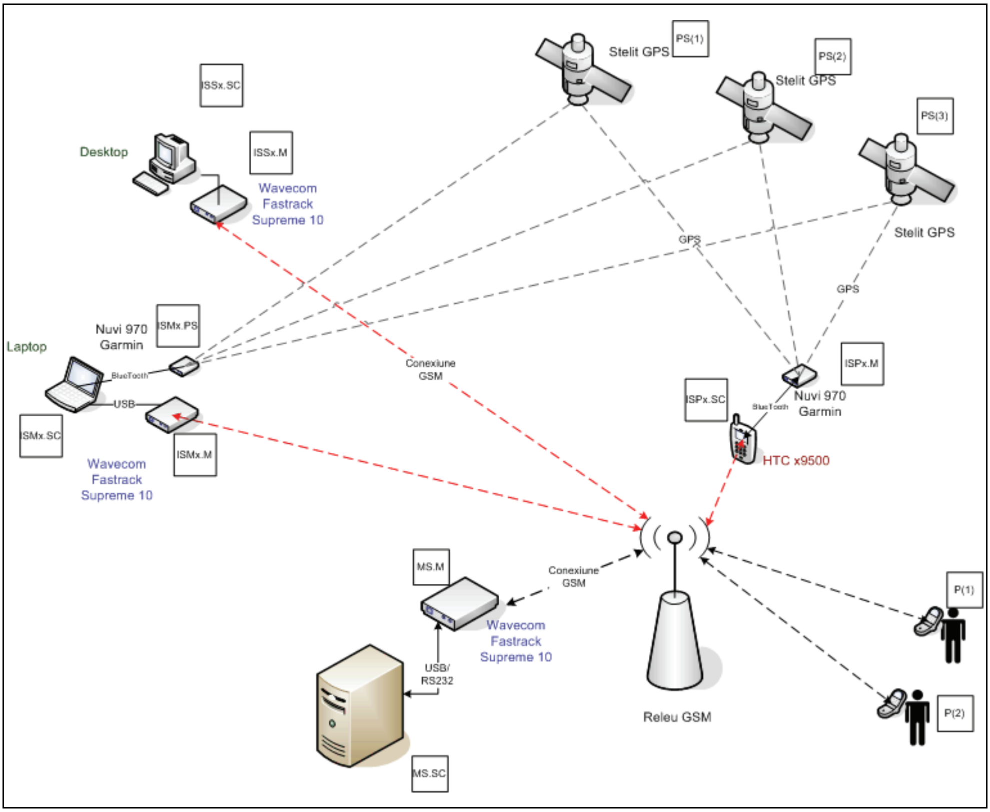


Fig. 3. Basic Scheme of the Data Collection and Dispatch Process

4.6.1.1 Main Server or MS

Consists of the following subcomponents:

- Computing system (MS.SC);
- GSM receiver/transmitter (MS.M).

(Technical) Operation

- the computer system (MS.SC) checks on a regular basis if there are any new messages (collected data). This operation is to be performed through OpenAT commands sent on the RS232 interface for the attached modem (MS.M);
- the collected data is processed and if one of the system parameters exceeds the specified limit, the warning process is initiated, the persons in charge are alerted (Px), through OpenAT commands sent through the RS232 interface for the GSM modem.

4.6.1.2 ISS

Individual static data collection system having (GPS coordinates known)
It is composed of:

- a computer system (desktop) (ISS(x).SC);
- GSM transmitter (ISS(x).M).

(Technical) Operation

- the user enters data into the system (temperature, water level, etc.)
- the computer system (ISSx.SC) archives data ad through OpenAT commands data is transmitted to the GSM modem (ISSx.M).

The modem sends the data through the GSM network to the Main Server MS.

4.6.2 The dynamic system

It is conceived as a dynamic system, in which all active components are mobile, and their location is determined by the GPS equipment.

Operation:

Basically, the system collects data, sends it to a central processing system, where the danger rate is analyzed. If the danger rate is higher than a specified limit, then the computer system initiates an alert request for the personnel in charge with crisis management.

Component parts:

The system contains the following entities:

- MS – Main Server or main data collection, processing, and emergency alert server;
- ISM1… n – individual data-collection mobile stations, operated by specialized personnel who observe and enter manually data into the system
- PS1...n – Positioning satellites
- P1...n – Personnel in charge

4.6.2.1 MS – Central Processing, Collection, and Alerting System

It includes the following subcomponents:

- computer system (MS.SC)
- GSM receiver/transmitter (MS.M)

(Technical) Operation

- the computer system (MS.C) checks new messages on a regular basis (collected data). This operation is performed through OpenAT commands send throughteh RS232 interface for the attached modem (MS.M).
- collected data is processed and if one of the parameters exceed the specific limit, the warning process is initiated, the persons in charge are alerted (Px), through OpenAT commands sent through the RS232 interface for the GSM modem.

4.6.2.2 ISS

Individual static data collection system (known GPS coordinates) It includes the following subcomponents:

- a computer system (desktop) (ISS(x).SC);
- GSM transmitter (ISS(x).M).

(Technical) Operation

- the user enters data into the system (temperature, water level, etc.)

- the computer system (ISSx.SC) archives data and through OpenAT commands data is transmitted tothe GSM modem (ISSx.M) through the RS232 interface.
- the modem sends the data through the GSM network to the Main Server MS.

4.6.2.3 ISM

Individual dynamic data collection system (unknown GPS coordinates, may be determined through ISM(x).PS)
It includes the following subcomponents:

- a computer system (laptop) (ISM(x).SC);
- GSM transmitter (ISM(x).M);
- Positioning system (ISM(x).PS).

(Technical) Operation

- the user enters data into the system (temperature, water level, etc.);
- ISMx.SC receives data from ISMx.PS through the Bluetooth interface;
- the computer system (ISMx.SC) archives data (GPS coordinates + Collected Data) and through OpenAT commands data is transmitted to the GSM modem (ISM(x).M);
- the modem sends the data through the GSM network to the Main Server MS.

4.6.2.4 ISP

Individual dynamic data collection system (unknown GPS coordinates, may be determined through ISP(x).SC).
It includes the following subcomponents:

- a computer system (PDA, Smartphone) (ISP(x).SC);
- Positioning system (ISP(x).PS).

(Technical) Operation

- The user enters data into the system (temperature, water level, etc.);
- ISPx.SC receives data from ISPx.PS through the Bluetooth interface;
- ISPx.SC sends the data through the GSM network to the Main Server MS.

4.7 Further developments

Future research will consider the following:

- Extending the mobile data collection platform; a possible scenario is presented in the (Cioca, 2008).
- the possibility that the person in charge might send back a code to the server through which the system would be able to make decisions (create new alerts, distribute the alert to other levels);
- adding sensors to the system that would help sending alerts automatically, or at least in an aided manner; the sensors might be installed in the field, in key locations, which would thus allow human operators to be assign a different task.

4.8 Level and impact area of the results. Other similar approaches

The critical analysis conducted through the method described in (Cioca et al., 2007) of the present-day disaster management and warning systems, has led to the conclusion that the SMS warning system is very useful and has not been implemented in Romania. Worldwide, there are several devices that helped us build the system and the software modules able to handle the device presented above. Moreover, besides this SMS warning system, the real-time data collection system is an equally important element.

The subject this project is a new one in our field of study; even though there are a few weak attempts of some international researchers to approach the matter, no Romanian researchers have tackled it so far (Cioca, 2008).

Nevertheless, after a long and painstaking struggle to find similar approaches, we have identified the following:

- in August 2004, the Dutch government funded a project of LogicaCMG involving the development of a natural-disaster and terrorist-attack warning system (http://edition.cnn.com/2005/TECH/11/09/dutch.disaster.warning/);
- JNW, the first company specialized in SMS warning systems in case of tsunamis in Sri Lanka(http://www.groundviews.org/2007/09/13/sms-news-alerts-uringemergencies-the-experience-of-jnw-and-the-tsunami-warning-of-13th-september-2007/).

5. Conclusions

In a constantly changing economic and social environment, organisations, managers, specialists in finance and accounting, people in charge with warning the population in case of disasters, etc. must make important decisions caused by the mobility of internal and external factors.

Decisions made in this context must balance advantages and disadvantages, forecast short-term, medium-term and long-term consequences on the activity of an organisation or community which may be affected by disasters, and be assessed before implementation. Decision support systems are meant to meet such requirements. The development of such systems is a time-consuming operation and it can only be carried out by specialised personnel. The modeling, formalisation and implementation efforts of knowledge in the field are substantial. Numerous modelling methods, support systems for creating representations and implementing solutions have been developed (Donciulescu et al., 1986), (Donciulescu et al., 1985), (Filip, 2008), but the mere operation of data collection is extremely difficult. Once the system has been created, it must pass a series of tests, all the defects must be corrected and only then it may be exploited under the direct supervision of those who implemented it.

The motivation to develop a complex system for the management of the environment and public dissemination of environment-related information is twofold:

- to provide the managers who make decisions in environmental issues with a complex environment management system, which enables them to make scientifically verified decisions, based on principles derived from ecology; such principles are: preservation of ecological balance, and biodiversity (genofund and ecofund), reducing water, air and soil pollution, reasonable exploitation of natural resources; the principles for approaching a complex environment management system and for the public dissemination of information related to the environment are as well presented;
- in addition, a system for the public dissemination of information related to the environment and a disaster warning system for the authorities are required to protect the population and their property from natural disasters (floods, drought, landslides, avalanches, severe pollution of air, water and soil and other disasters).

An important contribution to the development of the complex environment management system and public dissemination of information related to the environment is the *integration of subsystems (component modules) with the complex environment management system*. In this context, the architecture of the complex environment management system and public dissemination of information related to the environment emphasises the component modules and the interaction among them. DSSs dedicated to the environment require a system of mathematical models, of the simulation and control for the assessment of environmental risk (floods, landslides, drought, etc.) and their consequences, a disaster warning system, an Internet system for the management of environmental data, an environment management expert system, and finally, pilot systems for environment management and public dissemination of information related to the environment.

In other words, in order to develop performant and complex DSSs for disaster management, on the one hand, a multidisciplinary approach is required, an approach which should bring together specialists in various fields, such as: environmental sciences, GIS, geography, mathematics, informatics, organisations in charge with dealing with such situations; on the other hand, a global approach which should unite institutions and people from different countries, as such phenomena are not restricted to certain areas on Earth and they may occur anywhere; a transfer of know-how between partners/researchers worldwide brings mutual benefits to all and might prevent the loss of human lives and material damages.

6. References

Aggarwal, A.K., (2001). *A Taxonomy of Sequential Decision Support Systems*, University of Baltimore, USA, http://www.informingscience.org/proceedings/IS2001 Procee dings/pdf.aggarwalEBKAtaxa.pdf)

Alter, S. (2002). A work system view of dss in its fourth decade, *Eighth Americas Conference on Information Systems*, 2002, pg. 150-156

Bellorini, N.; & Lombardi, M. (1998). *Information and Decision Support Systems with GIS Technology*, http://pc-ambiente.como.polimi.it/dida/tesi/Svezia.pdf

Bizoi, M. (2007). Sisteme suport pentru decizii. Utilizare. Tehnologie.Construire, *PhD report 1*, Academia Română

Boboşatu, F. (2008). Sisteme avansate de asistare a deciziilor. *PhD Thesis*, Universitatea Politehnica, Bucureşți

Cioca, M. (2008). Sisteme suport pentru decizii utilizate in managementul dezastrelor, *Research Report,* Romanian Academy Research Grant, no. 215/18.04.2008

Cioca, M.; Cioca, L.I.; & Buraga, S.C. (2007). Spatial [Elements] decision support system used in disaster management, *First IEEE International Conference on Digital Ecosystems and Technologies*, Cairns AUSTRALIA, pp. 235-240

Cioca, M.; Cioca, L.I. & Buraga, S.C. (2008). SMS Disaster Alert System Programming, *2nd IEEE International Conference on Digital Ecosystems and Technologies*, Phitsanuloke THAILAND, pp. 489-493

Cioca, M.; Cioca, L.I. & Mihăescu, L. (2009). Infrastructure and System Programming for Digital EcoSystems used in Natural Disaster Management, *3rd IEEE International Conference on Digital Ecosystems and Technologies*, Istanbul, TURKEY, (to appear)

Daconta, M.C.; Obrst, L.J. & Smith, K.T. (2003). *The Semantic Web*, John Wiley & Sons

Demarest, M. (2005). Technology and Policy in Decision Support Systems, (http://www.dssresources.com)

Donciulescu D.; Filip F.G. & Popescu, Th. (1986). Computer Aided Decision Making in Water Retention and Allocation. In *H. P. Geering and M. Mansour (Eds.). Large Scale Systems Theory and Applications.* Pergamon, Press, Oxford, pp. 861-865

Donciulescu, D.; Filip F.G. & Guran, M. (1985). DSS in Water Resources. In *Proc. IFAC/IFORS Conf. On Cont. Sci. and Technol. For Development, CSTD' 85, Academic Publishers*, Beijing, pp. 1312-1322

Donovan, J.J. & Madnick, S.E. (1977). Institutional and Ad Hoc DSS and Their Effective Use, pp. 79-88, (http://web.mit.edu/smadnick/www/papers/J010.pdf)

Fensel, D.; Bussler, C. & Maedche, A. (2002). Semantic Web Enabled Web Services, Proceedings of *the International Semantic Web Conference, Lecture Notes in Computer Science 2342*, Springer-Verlag

Filip, F.G. (2004). *Sisteme suport pentru decizii*, ISBN 973-31-2232-7, Editura Tehnică, Bucureşti

Filip, F.G. (2007). *Sisteme suport pentru decizii*, ISBN 978-973-31-2308-8, Editura Tehnică, Bucureşti

Filip, F.G. (2008). Decision Support and Control for Large-Scale Complex Systems, A*nnual Reviews in Control (Elsevier) 32 (1)*, ISSN: 1367-5788

Filip, F.G. & Bărbat, B. (1999). *Informatică industrială: Noi paradigme şi aplicații*, Editura Tehnică, Bucuresti

Hackathorn, R.D., Keen, P.G.W (1981). Organizational strategies for personal computing in decision support systems", *MIS Quarterly*, *5*(3), pp.21-27

Hellstom, P. & Kvist, T. (2003). Evaluation of decision support modules and human interfaces using the TopSim simulator, *Future Train Traffic Control Project*, Report 4, Appendix 3, (http://www.it.uu.se/research/project/ftts/reports/C4B3.pdf)

http://edition.cnn.com/2005/TECH/11/09/dutch.disaster.warning/

http://www.groundviews.org/2007/09/13/sms-news-alerts-duringemergencies-the-experience-of-jnw-and-the-tsunami-warning-of-13th-september-2007/

Jeffery, K. & Kacsuk, P. (2004). *Grids: the Next Generation*, ERCIM News, 59

Keen, P. G. W. & M. S. Scott Morton (1978). Decision support systems: an organizational perspective, *Reading, Mass., Addison-Wesley Pub. Co*

Keen, P.G.W. (1987). Decision support systems: the next decade. *Decision Support Systems*, *3*, pp. 253-265

Laszewski, Von G. & Wagstrom, P. (2004). *Gestalt of the Grid*, Tools and Environments for Parallel and Distributed Computing, John Wiley & Sons

Muntean, M. (2003). Perfecționarea sistemelor suport de decizie în domeniul economic, *PhD Thesis*, Academia de Studii Economice, Facultatea de Cibernetică, Statistică şi Informatică Economică, Bucureşti

Parker, B. J. & Al-Utabi, G. A. (1986). Decision support systems: the reality that seems to be too hard to accept? OMEGA, *Int. J. Management Science 14(2)*

Power, D. J. (2002). *Decision support systems: Concepts and Resources for Managers*. Quorum Books, Westport, Connecticut

Shadbolt, N.; Hall, W. & Berners-Lee, T. (2006). The Semantic Web Revisited, *IEEE Intelligent Systems*, 21(3)

Sheth, A. et al. (2003). Relationships at the Heart of Semantic Web: Modeling, Discovering, and Exploiting Complex Semantic Relationships, *Enhancing the Power of the Internet*, Springer-Verlag

Sprague, Jr.,R.H. (1980). A framework for the development of decision support systems. *MIS Quarterly, 4(4)*, republicat în: (Sprague, Watson, 1993), pp. 3-28

Suduc, A.M. (2007). Sisteme support pentru decizii. Utilizare. Tehnologie.Construire, *PhD report 1*, Academia Română

Turban, E. (1998). *Decision Support Systems and Intelligent Systems*, Prentice Hall Inc. (Sixth Edition, 2001).

22

Security as a Game – Decisions from Incomplete Models

Stefan Rass, Peter Schartner and Raphael Wigoutschnigg
Alpen-Adria Universität Klagenfurt
Austria

1. Introduction

Securing a computer system is always a battle of wits: the adversary tries to locate holes to sneak in, whereas the protector tries to close them. Transmitting messages through a publicly accessible medium whilst having the content concealed from the adversary's eyes is traditionally accomplished using mathematical transformations. These are practically irreversible, unless some additional information – called the key – is available, making the secret accessible for the legitimate holder(s) of the key. Ever since the concept of perfect secrecy has rigorously been formalized by Shannon (1949), it has been known that unbreakable security is bought at the cost of keys that equal the message in terms of length. In addition, the key is required to be random and must be discarded immediately after usage. This pushed the concept of unconditional security out of reach for implementation in computer networks (though diplomatic and military applications existed), until 1984, where the idea of quantum cryptography was born by Bennett & Brassard (1984). The unique feature of this novel type communication is its usage of information carriers other than electrical pulses. By encoding bits in the polarization plane of single photons, the information becomes essentially not cloneable, as Wootters & Zurek (1982) have shown, and any attempt can be detected. This rendered the one-time pad practical in real-life electronic networks and unconditional security no longer needed to remain a dream.

Classical cryptography widely relies on unproven conjectures regarding the difficulty of solving computational problems. The field of public key cryptography draws its power from the infeasibility of reverting simple algebraic operations within large finite groups, but no proof has yet been discovered that rules out the existence of efficient algorithms to solve those problems. The sole indicator of security is thus the absence of any publication proving the assumptions wrong. But there is yet no other indication than pure hope for this to be true. Symmetric techniques, although conceptually different, come with no better arguments to support their security. Although these may lack much of the structure that public key systems enjoy and are thus harder to analyze, a rigorous proof of security or mathematical framework for proving security is also not available.

In this work, we attempt taking a step towards providing a rigorous and easy-to-use decision-theoretic framework for proving security. Results are formulated with applications to quantum networks, but we emphasize that the framework is in no way limited to these.

1.1 The problem of perfect end-to-end secrecy

Quantum cryptography claims to bring perfect secrecy to a given line, but speaking honestly, it is no more than this. Using a carrier that is sufficiently fragile to rule out copying it, naturally raises the question of how much distance can be bridged? In fact, nowadays available quantum cryptography allows for communication over a distance of up to 144 km, as demonstrated by Schmitt-Manderbach et al. (2007), but arbitrary distances can yet not be bridged. Although theoretical results due to Lo & Chau (1999) indicate that the noise problem can be overcome, making arbitrary ranges theoretically possible, building networks is inevitable for a global roll-out. Existing solutions mostly rely on trusted relay for that matter. However, why attack the quantum line, if attacking a relay node is fully sufficient?

Under the assumption of perfectly protected lines, recent results indicate that without pre-existing secrets that are exclusively known to the sender and the receiver, end-to-end-security is only achievable under hard constraints on the network topology. To be more precise, let G be a graph that models a network. Let $V(G)$, $E(G)$ be the sets of vertices and edges of G, and assume the sender s and receiver r to be parts of G, that is $\{s, r\} \subseteq V(G)$. The adversary can be modelled by a set $A \subseteq 2^{V(G)\setminus\{s, r\}}$ (the powerset of $V(G)\setminus\{s, r\}$), that is we assume that a selection of subsets of vertices can be compromised. If k such sets can become conquered simultaneously, then we face a k-active adversary. An infected vertex v is assumed fully under the adversary's control, so a message passing through v can be read, blocked or modified and v is free to create as many new messages as desired. There is no limitation on computational power or knowledge of the adversary.

If removing from G the vertices in any k sets in the adversary structure A cannot disconnect s and r in G, then we call the graph $A^{(k)}(s,r)$-subconnected. If, by doing so, the network cannot be disconnected at all, then the graph is said to be $A^{(k)}$-subconnected.

Referring to these notions, a network permits perfectly secure message delivery from s to r if and only if the graph G is $A^{(2)}(s,r)$-subconnected. The reader may consult Ashwin Kumar et al. (2002) for a proof. Different, yet no less stringent requirements are imposed by Wang & Desmedt (2008): among related results, the following necessary condition best highlights the difficulty of achieving unconditional security in a real-life network: if for $u \geq 1$, $3(k - u) + 1 \geq k + 1$ directed node-disjoint paths from s to r exist, then a necessary condition for perfectly secure message transmission from s to r against a k-active adversary is that there are u directed node disjoint paths (these u paths are also disjoint from the $3(k - u) + 1$ paths from s to r) from r to s.

The described adversary model applies to many situations, as for example machines running certain software may all suffer from the same security holes. Networks equipped with devices from different vendors may be considered vulnerable if one vendor's devices turn out to be insecure. A k-active adversary would correspond to k vendors cooperating, or equivalently arise, if k vendors obtained the same malicious module from a single fraudulent manufacturer, turning a heterogeneous set of products into a possible backdoor for an adversary.

1.2 Decision theory and system security

Many results either guarantee or rule out perfectly secret communication, but this might not be satisfactory. If perfectly secure communication is not possible, then how much is achievable with the given resources? A variety of security metrics has been proposed, but a measure of security is yet missing. This work summarizes a decision-theoretic approach to quantifying risk in terms that can be specified to best suit the application at hand.

Protecting business assets is the core goal that security engineers are in pursuit of, so measuring the quality of a protection mechanism in terms of values of the protected asset is certainly a more convincing argument than hoping that relaying nodes in quantum networks are trustworthy, or no efficient solution algorithm for some computational problem has yet been discovered.

The problem of measuring security has been tackled by a vast number of authors. Assessing security is commonly achieved by security metrics or scores, whereas the latter is considered for sole comparative purposes and does not have an interpretation on its own. The common vulnerability scoring scheme (see Houmb & Franqueira (2009) is one example for a scoring technique. Other taxonomies like proposal of Innerhofer-Oberperfler & Breu (2009) are as well subjective and may help decision-makers, but are not designed to support a further mathematical treatment.

Decision support systems like CAULDRON by Mas (2008) cook up reports generated by vulnerability scanners to boil down a vast amount of information to a manageable lot of recommendations. The models we describe can naturally benefit from these systems, and are thus considered an add-on for a standard topological vulnerability analysis (cf. Jajodia et al. (2005) for details on the latter). In particular, our results will generalize the assertions about the (im)possibility of perfectly secure communication as cited above. An approach that is closely related to our model has been given by Ying et al. (2006) and Mavronicolas et al. (2005). These approaches consider less general scenarios than we do, and suffer from the need for accurate adversary models. We demonstrate how this requirement can elegantly be dropped, while simultaneously simplifying a subsequent analysis.

2. Modelling

The modelling approach proposed in the following requires identification of security primitives of a given network. Its core ingredient is an enumeration of possibilities for transmission and parameter selection, and its output will be a game-theoretic model. For convenience of the reader, we review some necessary basics of game theory and multipath transmission, to illustrate the required input for a powerful model

2.1 Game-theoretic foundations

It is useful to collect some tools from game-theory that will help establishing the results presented here. A *(non-cooperative n-person) game* $\Gamma = (N, S, H)$ is a triple composed from a set $N = \{1,2, \dots , n\}$ of players being able to choose actions from their corresponding strategies within the set of sets $S = \{PS_1, PS_2, \dots, PS_n\}$, such that the *i*-th player, when taking the action $s_i \in PS_i$ from his set PS_i of possible pure strategies, receives the payoff $u_i(s_i, s_{-i})$, where $u_i \in H$ and s_{-i} denotes the vector of pure strategies chosen by *i*'s opponents. The set H thus comprises the set of payoff functions for each player. A probability distribution over the set PS_i is called a *(mixed) strategy*. We will exclusively refer to mixed strategies in the following, and denote the set of distributions over PS_i as S_i (note that the set of pure strategies is included in the set of distributions by considering a pure strategy as a Dirac-mass located at the pure strategy s_i). A *(Nash-)equilibrium* is a strategy-profile $s^* = (s_1^*, \dots, s_n^*)$ such that

$$u_i(s_i, s^*_{-i}) \leq u_i(s^*_i, s^*_{-i}) \quad \forall \mathrm{i} \in \mathrm{N}.$$

In other words, no player can benefit by solely deviating from the equilibrium strategy. The possibility of a gain when several players cooperate is not ruled out however. This is not topic of this work.

If all strategy sets are finite (assumed in the following), then the utility for a mixed strategy is the expected (average) utility over an infinite number of repetitions of the game. In other words, if (x, y) are the strategies (discrete probability distributions) of player 1 and 2, respectively, then the expected utilities are given by the bilinear forms

$$u_1(x, y) = x^T Ay \text{ and } u_2(x,y) = x^T By,$$

where $A \in \mathbb{R}^{n \times m}$, $B \in \mathbb{R}^{n \times m}$ (for $n = |PS_1|$, $m = |PS_2|$) are the *game-matrices*. The full two-player game is denoted as the triple $\Gamma = (\{1, 2\}, \{S_1, S_2\}, \{A, B\})$. If $A = -B$, then the game is called zero-sum, and its *value* $v(\Gamma)$ is given as the value of the function at the saddle-point, which is

$$v(\Gamma) = \max_x \min_y x^T Ay.$$

It can be determined upon linear optimization, as described by Schlee (2004).

2.2 Multipath transmission

We have already summarized two results characterizing the possibility of perfectly secure message transmission in the introduction. A popular for circumventing the person-in-the-middle attack is relying on several paths, over which messages are propagated independently. We shall not burden ourselves with the intricate details of error-correcting codes and how these relate to the concepts of secret sharing, and refer the reader to McElice & Sarwate (1981) for details. Recent protocols embodying the ideas of correctable shares for multipath transmission are found in the work of Fitzi et al. (2007) and Wang & Desmedt (2008), and we confine ourselves to remarking that perfectly secure message transmission is possible under a few assumptions:

1. An encoding is available that allows to divide a message into pieces such that any subset (of pieces) of limited cardinality does not provide any information about the secret itself. This is achieved by standard secret-sharing, as we will summarize later.
2. The network topology ensures the existence of several node-disjoint paths that connect any two nodes in the network. Results from graph theory (see Chartrand & Zhang (2005)) characterize suitable networks. Procedures for building such topologies from scratch have been developed in Rass (2005), and algorithms for determining optimal extensions of existing networks have been devised by Rass & Schartner (2009b).

Error-correction facilities that are inherently available within some secret-sharing schemes can be exploited to further increase security and reliability, however, are not a must for our upcoming considerations.

Let us review a simple form of secret sharing here that will become a theoretical asset for later conclusions. Given a secret $s \in [0, n - 1]$, choose t random numbers $r_i \in [0, n - 1]$, and set $r_{t+1} = s \oplus r_1 \oplus \ldots \oplus r_t$, where $\oplus$ denotes the bitwise exclusive or operation. It is evident that unless all values $r_1, \ldots, r_t, r_{t+1}$ are known, the secret remains one-time pad encrypted with the exclusive-or of the unknown components, and thus perfectly concealed. The values $r_1, \ldots, r_t, r_{t+1}$ are the *shares* that arose from s. The core idea of multipath transmission is to send each share over its own channel that does not intersect any other channel in the network. Unless an adversary has $(t+1)$ nodes conquered, no information about s can leak out. Practical multipath transmission protocols utilize a more sophisticated form of secret-sharing, where shares are created as points on a chosen (random) polynomial. Unless a sufficient number of such points are known, the polynomial, and therefore the secret it embeds in its constant term, remains protected from the adversary's eyes. The advantage over the previously sketched scheme is its resilience against loss of shares up to the extent of the threshold. This comes at the price of higher computational effort, as calculations have to be performed in large finite fields.

The methodology that is presented in the following naturally captures a much wider range of situations; however, we stick with a multipath scenario for illustrative purposes.

2.3 Setting up the model

Given a network at hand, mapping it into a model that permits decision-theoretic treatment proceeds in several steps. Each step is expanded below, starting with a definition capturing some terminology. Notice that in the sequel, we explicitly consider secret and reliable transmission, which is assumed available in various ways over the given network. Degrees of freedom that are available to the sender of a message comprise the following: transmission paths, encoding schemes (including encryption) and protocol parameters. We will assume a multipath transmission scenario (for otherwise perfect secrecy over multi-hop connections is ruled out under weak conditions as shown previously), and take the encoding to be fixed (as prescribed by the hardware devices). However, we can determine the path through the network.

Definition 1: A *pure strategy* is a set of node-disjoint paths that connect a sender Alice to a receiver Bob. The set of all pure strategies is denoted as *PS*. A *mixed strategy* is a probability distribution over *PS*. We denote the set of all such (discrete) distributions over *PS* as *S*, and refer to $x \in S$ simply as a *strategy*.

Speaking in game-theoretic terms, we refer to a pair of honest instances Alice and Bob as player 1, and call player 2 the adversary. Consequently, the sets of pure strategies are PS_1 and PS_2, with corresponding strategy sets S_1 and S_2. The methodology comprises five steps:

Step 1. **Identification of pure transmission strategies**: The expert shall enumerate all degrees of freedom that a sender enjoys when initiating a transmission. This in particular includes all sets of node-disjoint paths that can be used for multipath transmission. All meaningful choices are collected in the set PS_1 of pure strategies.

Step 2. **Identification of pure attacks strategies**: The expert shall enumerate all nodes that are vulnerable to an attack. This could be an assumption of the number of nodes that can simultaneously be compromised or a more complex adversary structure. In particular, this analysis should account for software security flaws that could be exploited. The finite set of options that are open to the adversary makes up the set PS_2 of pure strategies.

Step 3. **Setting up the utility taxonomy**: The expert shall specify a scoring scheme that applies to the outcome of a transmission. Examples include the binary set $I = \{0, 1\}$ with 0 meaning failure of a transmission, and 1 indicating a successful secret delivery. Finer discrete or even continuous scales can be based on a message priority ranking, or on the amount of Shannon-entropy that a message is tied to. We will frequently use the set $I = \{0, 1\}$ in the following.

Step 4. **Setting up the model matrix**: For every combination $(s_i, s_j) \in PS_1 \times PS_2$, identify the outcome of an attack and assign to the variable u_{ij} the score according to the ranking I. Notice that $u_{ij} = u_1(s_i^1, s_j^2)$ is the utility for the honest parties. For instance, with I = {0, 1}, if a successful transmission using the pure strategy $s_i \in PS_1$ would resist an attack via strategy $s_j \in PS_2$, then we set u_{ij} = 1. Otherwise we would write u_{ij} = 0 to indicate the adversary's success. The *model matrix* is denoted as A and has the entries u_{ij}.

Step 5. Analysis and Conclusions: The model matrix is the sole ingredient for any further analysis of the model. Conclusions are obtained from the results to follow.

Example 1: To illustrate the modelling process, consider a network topology as shown in Fig. 1 with two instances Alice and Bob who wish to communicate.

Modeling step 1: Assume that Alice and Bob have picked three pairs of shortest node-disjoint paths, disregarding other possibly longer paths. So player 1's set of pure strategies is denoted as PS_1 = {s_1^1, s_2^1, s_3^1}, and given by

- s_1^1: Use paths (s, 1), (1, 2), (2, t) and (s, 3), (3, 6), (6, t),
- s_2^1: Use paths (s, 1), (1, 2), (2, t) and (s, 4), (4, 5), (5, t)
- s_3^1: Use paths (s, 3), (3, 6), (6, t) and (s, 4), (4, 5), (5, t)

Modeling step 2: Eve strategies for attacking are given by PS_1 = {s_1^2, s_2^2, s_3^2}, where

- s_1^2: Compromise nodes 1 and 3,
- s_2^2: Compromise nodes 1 and 4,
- s_3^2: Compromise nodes 3 and 4,

modeling a situation in which two out of three vulnerable nodes {1, 3, 4} can be compromised simultaneously.

Modeling step 3: The utility taxonomy is chosen as I = {0, 1}, where 0 indicates failure of a secret transmission, and 1 means success. This scale considers loss of any secret content equally harmful.

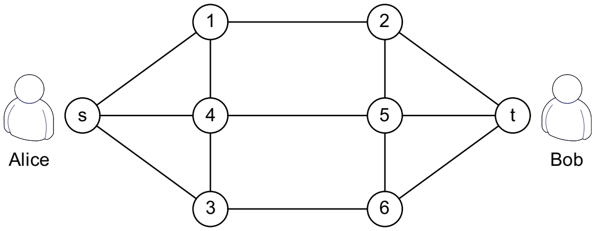


Fig. 1. Example Network Topology

Modeling step 4: Writing down every possible combination of pure strategies in a matrix, with entry 1 if the attack fails, we end up with the following table, directly representing the utility function u_1: $PS_1 \times PS_2 \to \{0, 1\}$, specified by a tableau (game-matrix, model-matrix):

u_1	s_1^2	s_2^2	s_3^2
s_1^1	0	1	1
s_2^1	1	0	1
s_3^1	1	1	0

The final step is the formal analysis of the model. We defer this until the formal results have been presented.

3. Decisions from incomplete models

An accurate game-theoretic model would call for the specification of the adversary's payoffs in order to optimally count his intrusion attempts. Unfortunately, we have no method of reliably eliciting the intentions and benefits that an attacker gains. Furthermore, we may be unable to observe our unknown opponent's payoffs at all, which rules out any chance of learning the adversary's payoff structure from experience. The game-theoretic model is thus *incomplete* in two respects:

1. We have no way of reliably determining the utility for the second player.
2. We have no mechanism of detecting our own success, nor can we observe the adversary's success. This may not apply for scenarios in which the adversary is active, so that an intervention can be detected, but a passively eavesdropping intruder will most likely remain undetected.

The remedy is switching to a zero-sum game, assuming the adversary's intentions and payoffs to be the precise opposite of our own ones. Though intuitively evident, the validity of this approach is formally founded (see Rass & Schartner (2009a) for a proof):

Lemma 1: Let $\Gamma = (\{1, 2\}, \{S_1, S_2\}, \{A, B\})$ be a bi-matrix game. Set $n = |PS_1|$, $m = |PS_2|$ and let A, B be the payoff matrices $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{n\times m}$ for players 1 and 2, respectively. Let $\Gamma_0 = (\{1, 2\}, \{S_1, S_2\}, \{A, -A\})$ be the zero-sum game from player 1's perspective, i.e. with the payoff of player 2 being the negative payoff of player 1 (disregarding the unknown matrix B), and let v denote the value of the game Γ_0. Then

$$v \leq (x^*)^T A y^*$$

for all existing Nash-Equilibria (x^*, y^*) of the game Γ.

This is the formal permission to use $(-A)$ as a substitute for the adversary's payoff, for getting a lower bound on the achievable utility. In other words, unless the adversary's purpose is truly opposite to our own one, we can only be off better than expected. Also, the bound cannot be improved, as examples by Rass (2009) demonstrate. In the following, we denote a random variable X with discrete distribution x by writing $X \sim x$.

Definition 2 (Loss): Let $i \in \{1, 2\}$ denote a player in a two-person game with pure strategy set PS_i, and S_i denoting the associated (mixed) strategy space. Assume that the utility function u: $PS_1 \times PS_2 \rightarrow I \subset \mathbb{R}^+$ to be a mapping into a compact set I. The *loss* is a random variable L measuring the difference between the actual and the possible outcome under the chosen pure strategies. It is defined as $L := (\max I) - u(X, Y)$, where $X \sim x \in S_1, Y \sim y \in S_2$.

Based on this, we can define *risk* as the expected loss. This is in alignment with the definition of risk as the product of probability and damage, as used by the German BSI (2008), as well as Hammer (1999).

Definition 3 (Risk): With the notation from Definition 2, player i's *risk* (for $i \in \{1, 2\}$) when choosing the strategies $x \in S_1$ and $y \in S_2$, is defined as the expected loss under this strategy choice, namely $r_i(x, y) = \mathrm{E}(L) = (\max I) - \mathrm{E}_{x,y}(u_i(X, Y))$, where for the random variables X and Y have the discrete distributions x and y, respectively, and the risk is dependent on the choices of player i's opponent.

It is straightforward to reformulate Definition 2 and Definition 3 for more than two entities. However, this general formulation is not required in the sequel, and thus omitted. The core concept upon which we can analyze security in a decision-theoretic sense is introduced through

Definition 4 (Vulnerability): Let $A \in I^{n \times m}$ be the model matrix set up by Alice and Bob, where these two have $n = |PS_1|$ pure strategies for communicating, facing an adversary with $m = |PS_2|$ pure strategies to choose from. Assuming that the set $I \subset \mathbb{R}^+$ to be compact, the *vulnerability* $\rho(A)$ is defined as

$$\rho(\mathrm{A}) := (\max I) - v(\Gamma_0),$$

where $v(\Gamma) = \max_x \min_y x^T A y$ is the value of the associated zero-sum game (see Lemma 1) $\Gamma_0 = (\{1, 2\}, \{S_1, S_2\}, \{A, -A\})$ and S_1, S_2 are the probability spaces over PS_1, PS_2, respectively.
The vulnerability is directly derived from the game-matrix A, which we shall refer to as the model matrix in the following. Summarizing the construction, this matrix displays the utility for the honest parties, for each possible (pure) transmission and (pure) attack strategy.
Example 2: The value of the game with the matrix given previously is found as $v = {}^1/_3$, so that the vulnerability comes to $\max I - v = 1 - {}^1/_3 = {}^2/_3$.

4. Results

An immediate consequence of the definition of vulnerability and Lemma 1 is the following result. We refrain from stating the proofs for the cited assertions and refer the interested reader to the work of Rass (2009).
Theorem 1 (Rass (2009)): If the message valuation scale is binary in the sense that every message scores 1 when being delivered successfully, and zero when getting deciphered by Eve, then $\rho(A)$ is the maximum probability of a concealed message becoming disclosed.
Capturing utility in terms of entropy permits quantification of the expected leak of information. In other words, the decision-theoretic approach yields a measure of secrecy-capacity of a network via a corollary to Theorem 1:
Theorem 2 (Rass (2009)): For a (random) message M with Shannon-entropy $\mathrm{H}(M)$, the amount h by which the adversary's uncertainty (Shannon-entropy) is decreased upon a (secret) communication between Alice and Bob satisfies

$$h \leq \rho(A) \cdot \mathrm{H}(M)$$

where $\rho(A)$ is the vulnerability, and the model matrix is set up with the binary scale $I = \{0, 1\}$ (i.e. a 1 in the matrix corresponds to one successful secret delivery, and 0 means failure).
Example 3: Knowing that the vulnerability of our example model is 0.667, this is the maximum probability of having a secret communication disclosed when communicating over the network. Theorem 2 states that no more than two thirds of any secret information will leak out in the long run average.
Whereas Theorem 1 and Theorem 2 capture long-term average secrecy of a channel, the decision on whether or not the next transmission should be started calls for a measure of risk for a single concrete transmission. If L is the random variable that measures the loss (i.e. the difference between the maximum utility and the actual utility) of the next transmission, then the next result upper-bounds the probability for loosing more than indicated by the vulnerability (in accordance with Theorem 1).
Theorem 3 (Rass (2009)): Let the secret communication between Alice and Bob be modelled by a bi-matrix game Γ and let Γ_0 be the associated zero-sum game as in Lemma 1. Let A be

the model-matrix. If Alice and Bob play an equilibrium strategy for Γ_0, then for $\varepsilon \geq 0$, the random loss $L \in I$ satisfies

$$\Pr(L > \rho(A) + \varepsilon) \leq \exp\left(-\frac{2\varepsilon^2}{C}\right)$$

where $\rho(A)$ is the vulnerability and the constant C is the solution of the optimization problem

$$C = \min_{t \in [0, \min(I) + \max(I)]} \sum_{i=1}^{n} \left(\max_j \left| a_{ij} - t \right| \right)^2$$

Determining the constant C is easy and requires only polynomial effort. The optimization problem is convex so the solution is unique. Asking for the loss that a sequence of messages can cause requires taking possible interdependencies of the messages into account. This rules out applying Theorem 3 repeatedly to upper-bound the probability of the joint loss. Instead, one can prove the following assertion to hold for several, possibly interdependent transmissions.

Theorem 4 (Rass (2009)): Let A denote the model matrix of the honest participants Alice and Bob. Assume $I \subset \mathbb{R}^+$ to be compact. If n (possibly interdependent) messages are transmitted over the network, then for any $\varepsilon \geq 0$,

$$\Pr(\min_{1 \leq i \leq n} L_i \geq \rho(A) + \varepsilon) \leq \exp\left(\frac{-n\varepsilon^2}{2(\max(I) - \min(I))^2}\right)$$

if L_i denotes the loss if the adversary mounts an attack on the i-th transmission.

5. Applications

The framework sketched here is general and applies to a wide range of scenarios. Despite its initial purpose being the security assessment of quantum networks, the results apply to any finite two-person game. Future research includes applications to classical networks, as well as considering more general communication scenarios like broadcasting.

5.1 Perfectly secure transmission

A conclusion that can be drawn from Theorem 4 is the possibility of perfectly secure communication over an arbitrarily insecure channel. Assume $I = \{0, 1\}$, so that utility 1 (cf. step 3 of the modelling process) corresponds to secure and secret delivery, and 0 corresponds to successful adversarial extraction of the secret. Consider the event that for n messages, $\min_i L_i \geq 1$. Since the i-th loss $L_i \in I = \{0, 1\}$, this implies $L_i = 1$ for all $i = 1, 2, \ldots, n$. In other words, the upper bound given by Theorem 4 refers to the even that the adversary extracted all messages from the sequence. Letting n become large, this probability will decay exponentially fast, which means that with overwhelming probability, at least one message will remain concealed and secure. If an (n,n)-secret sharing as described in previous paragraphs is employed, then the probability of extracting any secret content is negligible. Notice that none of the results presented relies on a hypothesis about the adversary's intentions or a mechanism of detecting the success of a transmission. Hence, we can draw strong conclusions from a game-theoretic model that we cannot even fully specify. The formal statement of the above intuition is

Theorem 5 (Rass (2009)): Let the pair (Alice, Bob) set up their model matrix with binary entries $u_{ij} \in \{0, 1\}$, where $u_{ij} = 1$ if and only if a message can securely be delivered by choosing the i-th pure strategy, and the adversary uses his j-th pure strategy for attacking. Then $\rho(\text{A}) \in [0, 1]$, and

- for any $\varepsilon > 0$, if $\rho(\text{A}) < 1$, then a protocol exists so that Alice and Bob can communicate with an eavesdropping probability of at most ε.
- if $\rho(\text{A}) = 1$, then the probability of the message becoming extracted by the adversary is equal to 1.

Rephrasing the implication of Theorem 5 reveals that the possibility of secure communication is completely characterized by the vulnerability. This is a significantly stronger result than the ones presented by Wang & Desmedt (2008), as its assertion is valid for any given network topology. In particular, it opens the possibility of optimizing a given topology, as we will show later. The vulnerability is compatible with the security concept as given by

Definition 5 (Wang & Desmedt (2008)): Let a message transmission protocol be given, and call *adv* the adversary's transcript of an eavesdropped execution. Assume the transmission protocol to take a message m and a random sequence r of coin-flips as input, and denote the adversary's information as $adv(m, r)$. Furthermore, let m_A denote Alice's input, and let m_B denote Bob's final output of the protocol.

- Let $\delta < ½$. A message transmission protocol is *δ-reliable* if, with probability at least $1 - \delta$, Bob terminates with $m_A = m_B$. The probability is taken over the possible choices of m_A and the coin-flips of all nodes. If $\delta = 0$, then the protocol is said to be *reliable*.
- A message transmission protocol is *ε-private* if, for every two messages m_0, m_1 and every r, $\sum_c \left|\Pr(adv(m_0, r) = c) - \Pr(adv(m_1, r) = c)\right| \le 2\varepsilon$ that is, if the distributions of the adversary's views for transmissions of m_0, m_1 differ at most by 2ε in the 1-norm. The probabilities are taken over the coin-flips of the honest parties, and the summation is over all possible values of the adversaries view. A 0-private protocol is called *perfectly private*.
- A message transmission protocol is (ε, δ)-secure, if it is δ-reliable and ε-private. A (0,0)-secure protocol is called *perfectly secure*.

This definition is perfectly compatible with our understanding of vulnerability, as indicated by the following

Theorem 6 (Rass (2009)): The vulnerability is a measure of privacy and reliability in the sense of Definition 5 because if Alice and Bob set up their model matrix with entries

$$u_{ij} = \begin{cases} 1, & \text{if the message is delivered successfully;} \\ 0, & \text{otherwise} \end{cases}$$

for every strategy combination of the honest parties and the adversary, then the protocol is ρ-reliable, where $\rho = \rho(A)$ is the vulnerability. If Alice and Bob set up their model matrix with entries u_{ij} such that

$$u_{ij} = \begin{cases} 1, & \text{if the adversary learns nothing about the secret content;} \\ 0, & \text{otherwise} \end{cases}$$

for every strategy combination of the honest parties and the adversary, then the protocol is (2ρ)-private, where $\rho = \rho(A)$ is the vulnerability.

5.2 Denial-of-service resilience

Denial of service scenarios are of particular interest in the field of quantum cryptography, because communication is aborted upon detecting the presence of an adversary. Since there is no evasive mechanism that could ensure the function of a link under eavesdropping, a denial of service is most easily mounted in a quantum network. Any small physical damage may raise the error rate sufficiently to logically cut the link. Other examples include classical (electrical) networks, or even distributed attacks on computer networks via bot-nets. All of these can easily be modelled and analyzed with our approach.

Modelling scenarios with random influences (such as intrusion detection systems that have only a high probability of preventing an attack, but offer no provable security assurance) is straightforward by switching from a deterministic utility function to a random one. Basically, this amounts to replacing a random outcome by its expectation. Examples are networks that employ intrusion detection and prevention mechanisms. These mechanisms evade an intruder with a certain probability, but not with certainty, so the possible outcomes $u_{ij} = 1$ (successful transmission), or $u_{ij} = 0$ (transmission failure) occur with probabilities p and $1 - p$, respectively. In that case, one would set up a matrix over the set $I = [0, 1]$ instead of $I = \{0, 1\}$, and replace each random utility U_{ij} with its expected value $\mathrm{E}(U_{ij}) = p \cdot u_{ij}$.

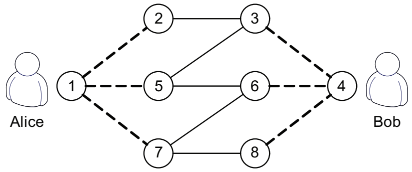


Fig. 2. Network in jeopardy of a DoS-Attack

Example 4: Take the network topology as shown in Fig. 2, and assume the adversary with threshold two, i.e. no more than two links can be attacked simultaneously. Moreover, assume that only the dashed links are assumed vulnerable, which corresponds to the assumption that the neuralgic point is the last mile connecting the peer's machine to the (quantum) backbone. The links inside the network are assumed protected from unauthorized access. This time, we are interested in whether or not Alice (node 1) and Bob (node 4) are in danger of suffering a denial-of-service attack, that is, the adversary's purpose is cutting the channel between the two by exploiting the eavesdropping detection facility of quantum cryptography. Setting up the game with a binary matrix with entries

$$u_{ij} = \begin{cases} 1, & \text{if removing the edges in } s_j^2 \text{ leaves the chosen path } s_i^1 \text{ intact;} \\ 0, & \text{otherwise (Eve has blocked the path)} \end{cases}$$

we end up with a 9×15 matrix. Solving the game yields the value $v = {}^1/_3$, so $\rho = 1 - v = {}^2/_3$. The assertion of Theorem 1 is not limited to pure communication, and we may directly conclude that the probability of a successful denial-of-service is at most $\rho = {}^2/_3$. Let Alice and Bob retransmit their message in case of failure. Then the probability of mounting a denial-of-service is for, say 100 messages, by Theorem 5 no more than $\exp(-100({}^1/_3)^2/2) \approx 0.00386$ (choose the maximal ε, which is $\varepsilon = {}^1/_3$).

5.3 Constructing networks with optimal security

Decision-makers that ought to assess several security enhancements for an existing network may be interested in an objective measure of security. The vulnerability as given in Definition 4 provides a natural scoring functional that can be optimized under given budget constraints.

Sticking with an eavesdropping scenario for simplicity, consider a network whose topology does not permit perfectly secure message transmission. This could be the case if a company owns fibre-optic lines and wants to enter the market as a backbone provider for quantum networks. Such secure delivery services are most interesting in R&D-scenarios, where several spatially separated departments work on highly valuated research projects with the need to exchange sensitive data regularly. Different, yet equally important, examples are secure backups, which should be located outside the company's premises (due to fire protection requirements).

In this section, we consider the first of the following two questions, where the second is straightforward to tackle.

- Given a set of environmental and monetary constraints, what is the best security we can achieve under the given conditions?
- Given a minimum security level, what is the cheapest extension to the network that achieves the desired security?

Since the vulnerability as defined up to now refers to only two players, one needs a more general tool: for a graph $G = (V, E)$, and a set of instances $U \subseteq V$, we will consider the maximum vulnerability over each pair of communicating nodes in the set U in the graph G. This quantity is

$$R(U,G) := \max_{s,t\in U, s\neq t} \rho(A(s,t))$$

where $\rho(A(s,t))$ is the vulnerability that is derived from the model matrix A, which now depends on the specific pair (s, t). For simplicity, we restrict ourselves to extensions on the link level of a network. That is, given a graph $G = (V, E)$, and a set E' of links that can technically be realized, we seek a minimum cost extension $\widetilde{E} \subseteq E' \subseteq V \times V$ such that the extended topology $G' = (V, E \cup \widetilde{E})$ has minimum vulnerability. Costs of various types (staff, maintenance, etc) can be captured through a vector-valued function $c: E' \to \mathbb{R}^d$, where $d \geq 1$. The components of c refer to different types of costs that cannot be merged into a single functional. Having specified some constraint $M \in \mathbb{R}^d$ we ought to solve the following nonlinear optimization problem:

$$\begin{aligned} &\min_{\widetilde{E}\subseteq E'} R(U, G(V, E \cup \widetilde{E})) \\ s.t. \quad & c(\widetilde{E}) \leq M \end{aligned} \tag{1}$$

By reducing this problem to the 0-1-Integer programming problem, as done by Rass (2009), we obtain the following

Theorem 7: The optimization problem (1) is at least NP-hard.

Unfortunately, this result provides some evidence that we can hardly do better than solving the optimization procedure as follows:

1. Enumerate all feasible extensions to the network. That is, find all extensions that obey the cost constraints.
2. Determine the vulnerability of the extended network for each of these cases, and select the extension with the least vulnerability.

This method can be applied to illustrate the optimization through the following

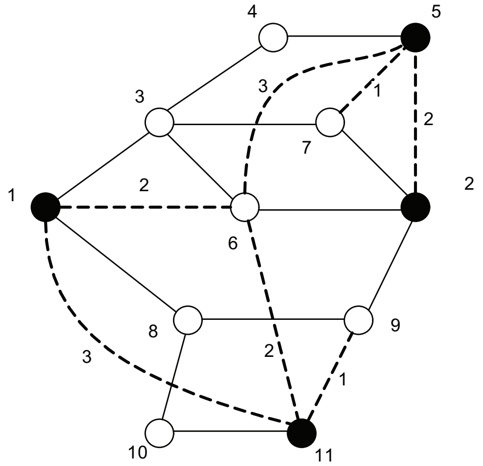


Fig. 3. Initial Network

Example 5 (Rass & Schartner (2009b)): Given a graph G as shown in Fig. 3, suppose that the set E' of candidate extensions comprises the dashed links, i.e. $E' = \{(1, 6), (1, 11), (2, 5), (5, 6), (5, 7), (6, 11), (9, 11)\}$. The weights of the links are the costs for building them. It is easy to verify that the network vulnerability excluding the dashed links is 1, which is due to node 5's inability to communicate with any other node, once an adversary with threshold at least 1 compromises node 4. In our example, assume that each pair within the set of communicating nodes $C = \{1, 2, 5, 11\}$ uses a (2, 2)-secret sharing scheme, and that the adversary has threshold 2, i.e. Eve can compromise any two nodes in the network, except for {1, 2, 5, 11}. Enumerating the paths that a fixed pair can use by $i = 1, 2, \ldots$, and Eve's possible attacks by $j = 1, 2, \ldots$, the game-matrix has an entry $u_{ij} = 0$ if and only if Eve's attack is a cut between the i-th pair (alternatively, Eve mounted a person-in-the-middle attack), and 1 otherwise. The cost functional is assumed additive and scalar-valued. Given a budget limit of $M = 13$, we ask for a selection of dashed links that gives us the minimum achievable network jeopardy. Solving the optimization problem, we can find 14 different solutions, each of which satisfies the budget limit and provides a network jeopardy of $^1/_3$ (in a real-life situation, the set of admissible solutions may further be restricted to limitations on cable length, or other additional constraints). One such solution (the cheapest among the candidates, with cost 9) is $\widetilde{E} = \{(1,11),(5,6),(5,7),(6,11)\}$, and the resulting, security enhanced network, is depicted in Fig. 4.

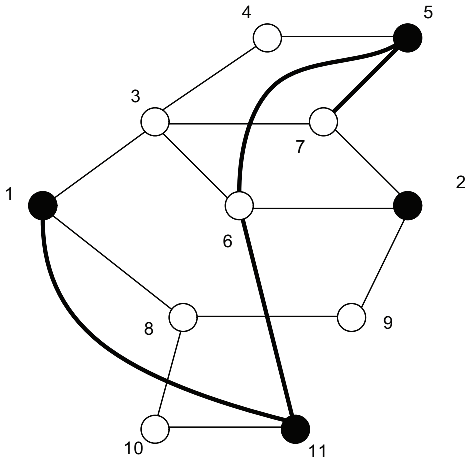


Fig. 4. Optimized Network

6. Conclusion

We devised a simple to use framework for analyzing complex scenarios with tools from game theory. Motivated by recent results regarding the impossibility of perfect end-to-end secrecy, we point out decision theory as a valuable tool for obtaining strong general results in the field of system security. Unlike many other approaches, our method is not limited to a specific scenario, and can equally well be applied to tackle confidentiality, authenticity, integrity and availability aspects of a system.

The framework has yet been formulated for the communication of two instances, and generalizations to broadcast scenarios are a direction for future research. The vulnerability measure that we obtained may also be used with time-series forecasting techniques to build an automatic alarming system that keeps track of ongoing evolution and predicts the future security of the given system.

A third avenue of future research is the analysis of the optimization problem. Some steps of turning the problem into a standard mixed integer optimization problem have been accomplished (see Rass (2009) for details), and commercial software packages can be used for tackling the arising problems. The results allow casting the problem of designing a secure network into a combinatorial optimization problem, using a widely automated procedure. Protocol parameters, protocols themselves, transmission paths, and most other parameters can be enumerated automatically. This way, we can automatically create the strategy sets for the honest instances. Network vulnerability scanners help identifying the attack strategies, so these can be set up in an automated manner as well. It is easy to set up the game-matrix, even if random influences are considered. Finally, the analysis, optimization and prediction of future values can also be handed over to software solutions, making the methodology flexible, efficient and a valuable add-on for security analysis in a broad range.

7. References

Ashwin Kumar, M.; Goundan, P. R.; Srinathan, K. & Pandu Rangan, C. (2002), On perfectly secure communication over arbitrary networks, *in 'PODC '02: Proceedings of the twenty-first annual symposium on Principles of distributed computing'*, ACM, New York, NY, USA, pp. 193–202.

Bennett, C. & Brassard, G. (1984), Public key distribution and coin tossing, in '*IEEE International Conference on Computers, Systems, and Signal Processing.*', IEEE Press, Los Alamitos.

BSI (2008), *IT-Grundschutz-Kataloge – 10. Ergänzungslieferung*, Bundesamt für Sicherheit in der Informationstechnik. http://www.bsi.bund.de/gshb/, English version (from 2005) available at http://www.bsi.de/gshb/intl/index.htm.

Chartrand, G. & Zhang, P. (2005), *Introduction to Graph Theory*, Higher education, McGraw-Hill, Boston.

Fitzi, M., Franklin, M. K., Garay, J. & Vardhan, S. H. (2007), Towards optimal and efficient perfectly secure message transmission, in S. Vadhan, ed., '*Theory of Cryptography, 4th Theory of Cryptography Conference, TCC 2007*', Lecture Notes in Computer Science LNCS 4392, Springer, pp. 311–322.

Hammer, V. (1999), *Die 2. Dimension der IT-Sicherheit: Verletzlichkeitsreduzierte Technikgestaltung am Beispiel von Public Key Infrastrukturen*, DuD-Fachbeiträge, Vieweg.

Houmb, S. H. & Franqueira, V. N. L. (2009), Estimating ToE risk level using CVSS, in '*Proceedings of the International Conference on Availability, Reliability and Security*', IEEE Computer Society Press, pp. 718–725.

Innerhofer-Oberperfler, F. & Breu, R. (2009), An empirically derived loss taxonomy based on publicly known security incidents, in '*Proceedings of the International Conference on Availability, Reliability and Security*', IEEE Computer Society Press, pp. 66–73.

Jajodia, S., Noel, S. & O'Berry, B. (2005), *Massive Computing*, Springer US, chapter Topological Analysis of Network Attack Vulnerability, pp. 247–266.

Lo, H.-K. & Chau, H. F. (1999), 'Unconditional security of quantum key distribution over arbitrarily long distances', Science 283, 2050–2056. arXiv:quant-ph/9803006.

Mas (2008), 'Combinatorial analysis utilizing logical dependencies residing on networks (CAULDRON)'. http://ait.gmu.edu/~csis/.

Mavronicolas, M., Papadopoulou, V., Philippou, A. & Spirakis, P. (2005), *Internet and Network Economics*, LNCS, Springer, chapter A Graph-Theoretic Network Security Game, pp. 969–978.

McElice, R. & Sarwate, D. (1981), 'On sharing secrets and Reed-Solomon codes', *Communications of the ACM* 24(9), 583–584.

Rass, S. (2005), *How to send messages over quantum networks in an unconditionally secure manner*, Technical Report TR-syssec-05-05, Universität Klagenfurt, Forschungsgruppe Systemsicherheit.

Rass, S. (2009), *On Information-Theoretic Security: Contemporary Problems and Solutions*, PhD thesis, Klagenfurt University, Institute of Applied Informatics (to appear).

Rass, S. & Schartner, P. (2009a), Game-theoretic security analysis of quantum networks, in '*Proceedings of the Third International Conference on Quantum, Nano and Micro Technologies*', IEEE Computer Society, pp. 20–25.

Rass, S. & Schartner, P. (2009b), Security in quantum networks as an optimization problem, in '*Proceedings of the International Conference on Availability, Reliability and Security*', pp. 493–498.

Schlee, W. (2004), *Einführung in die Spieltheorie*, Vieweg.

Schmitt-Manderbach, T., Weier, H., Fürst, M., Ursin, R., Tiefenbacher, F., Scheidl, T., Perdigues, J., Sodnik, Z., Kurtsiefer, C., Rarity, J. G., Zeilinger, A. & Weinfurter, H. (2007), 'Experimental demonstration of free-space decoy-state quantum key distribution over 144 km', *Physical Review Letters* 98(1), 010504. http://link.aps.org/abstract/PRL/v98/e010504

Shannon, C. (1949), 'Communication theory of secrecy systems', *Bell System Technical Journal* 28, 656–715.

Wang, Y. & Desmedt, Y. (2008), 'Perfectly secure message transmission revisited', *IEEE Transactions on Information Theory* 54(6), 2582–2595.

Wootters, W. K. & Zurek, W. H. (1982), 'A single quantum cannot be cloned', *Nature* 299(802), 802–803.

Ying, Z., Hanping, H. & Wenxuan, G. (2006), 'Network security transmission based on bimatrix game theory', *Wuhan University Journal of Natural Sciences* 11(3), 617–620.