Intelligent Teaching Models for STEM Related Careers Using A Service-Oriented Architecture and Management Science

Naser El-Bathy
North Carolina A&T State University
nielbath@ncat.edu

Clay Gloster
North Carolina A&T State University
cgloster@ncat.edu

Ghassan Azar
Lawrence Technological University
gazar@ltu.edu

Mohammed El-Bathy
Lawrence Technological University
melbathy@ltu.edu

Gordon Stein
Lawrence Technological University
gstein@ltu.edu

Ricky Stevenson
North Carolina A&T State University
rbsteven@ncat.edu

***Abstract*—The development of World Wide Web (WWW) a little more than a decade ago has caused an information explosion that needs an Intelligent Web (IW) for users to easily control their information and commercial needs. Therefore, engineering schools have offered a variety of IW courses to cultivate hands-on experience and training for industrial systems. In this study, Intelligent Teaching Models for STEM Related Careers Using Service-Oriented Architecture (SOA) and Management Science project course has been designed. The goal is to help students learn theoretical concepts of IW, practice advanced technical skills, and discover knowledge to solve problem. Undergraduate Science, Technology, Engineering and Mathematics (STEM) students involved in the development of innovative approaches and techniques. They are able to help solve the problems of disease misdiagnoses that medical and health care professionals experience. They co-authored and presented numerous research papers introducing the solution via scientific conferences and journals. This study provides the solution in the form of an Intelligent models using an integration of Service-Oriented Architecture and Management Science to decrease disease misdiagnosis in health care. Results show that this new course strengthens the capacity and quality of STEM undergraduate degree programs and the number of overall graduate student enrollment. It promotes a vigorous STEM academic environment and increases the number of students entering STEM careers. It expands the breadth of faculty and student involvement in research and development. It enhances and leverages the active engagement of faculty technology transfer and translational research. It improves and develops new relationships between educational institutions and research funding entities to broaden the university's research portfolio and increase funding. The proposed project course is a software engineering research methodology, an educational tool, and a teaching technique is needed in future medical and health IT fields.**

Index Terms—SOA, Management Science, Teaching Models, Disease Diagnosis.

I. INTRODUCTION

Last decade, the researchers have designed and developed several intelligent web technologies such as Web Mining (WM) and Web Services (WS). These technologies have become the major courses that provide engineering graduate students with intelligent web skills. Some schools offer these courses as elective courses in undergraduate program. Others recommend it as directed study courses for undergraduate and graduate students. The research project has been evolved as a key course at North Carolina Agricultural And Technological

978-1-4673-0818-2/12/$31.00 ©2012 IEEE

State University, and Lawrence Technological University in the State of Michigan. In the fall of 2010, the course has been offered for the first time at Lawrence Technological University as a directed study course for undergraduate program. This research project prepares students for STEM careers using the criteria of Health Informatics, SOA, Management Science, Artificial Intelligent, Bioinformatics, Intelligent Information Retrieval, Web Middleware, and Server Technologies.

El-Bathy designed the course as a software engineering research methodology, an educational tool, and a teaching technique. As a research methodology, the instructor addresses the conceptual aspects of innovation and discusses the research complications associated with the notion. The instructor also outlines a list of factors said to contribute to innovation within organizations. The course is an educational tool that the instructor uses in teaching an array of technologies. This tool is an extensive workshop in which the students learn these new technologies, implement it, and discover knowledge to solve problems using technical skills they learn. The teaching technique is a structure in which the development of the research project is formed, designed, and managed. This technique enforces the concept of software engineering. It ensures accuracy, efficiency, and high quality during the process of the research project analysis, design, assessment, implementation, test, maintenance and reengineering.

Web Information Retrieval (WIR) courses are being offered for both undergraduate and graduate students in many schools such University of Arkansas, University of Texas at Austin, New York University, and Lehigh University. Harding University offers Search Engine Development as an elective undergraduate course for sophomores, juniors, and seniors [1]. The course builds a search engine through a set of bottom-up projects. It also develops projects to modify an existing open source search engine.

The remaining material of this paper is structured as follow: In Section II, the motivation, the course philosophy, and the course project are identified. In Section III, Management Science Models are developed. In Section IV, Service-Oriented Architecture principles are planned. In Section V, a new clustering genetic algorithm and intelligent agents are built. In Section VI, the prototype is created. Finally, the results and conclusion are given in Sections VII and VIII.

II. MOTIVATION AND PHILOSOPHY

A. Motivation

Researchers have often studied open source software engineering solutions for health care information technology including OSCAR, FreeMed, TORCH, and OpenEMR. These solutions have provided high-quality electronic medical records, practice management systems, simpler prescription writing, scheduling, and billing [2], [3]. However, the authors believe that these solutions cannot entirely solve the problems of disease misdiagnosis because of its incapability to check diagnoses with symptoms. Motivated by these problems, the authors propose "Intelligent Teaching Models for STEM Related Careers Using SOA and Management Science". The proposed theory is an automated solution to capture the challenge of disease misdiagnosis while students learn theoretical concepts and technical skills. The causes for this challenge involve four main factors: absence of open software systems' integrity, inefficient information retrieval processes, poor quality of clustering different diseases' relevant information, and lack of information that analysts require to strategically plan medical and healthcare industries.

B. Course Philosophy

The philosophy of this course project is based on its level. In an undergraduate program, an introduction to intelligent web development course is designed and structured. The course is highly motivated forward looking students in computer science, engineering, education, instructional technology, medical science, and management. After completing this course the student are acquainted with fundamentals of Service SOA, Extensible Markup Language (XML) schema, fundamentals of Semantic Web, introduction to Artificial Intelligence, Search Methodologies, Service Orchestrations with Business Process Execution Language (BPEL), Introduction to Web Applications development, and Introduction to IT Research Methodology.

In a graduate level, advanced intelligent web development course is designed and structured. The course is of interest to graduate students in computer science, engineering, education, instructional technology, medical science, and science management. Students master new technologies such as: Business Process Execution Language, Java Server Faces (JSF), Web Services, SOAP, WSDL, UDDI, APIs and XML. In this course, we use major platforms for web application and web services development such as Oracle Server Application (OSA) and Java EE Application server, along with IDEs such as JDeveloper. All background material related to HTML, XML, JavaScript, Java SE/EE, and client/server architecture are developed within the course itself from scratch. The course is for students who prefer hands on experience of advanced IT Applications and research methodologies and like the thought of using real tools. It is also for students who want to be graded based on what they can do as well as what they know and the students who are interested in writing, publishing, and presenting papers in scientific conferences and journals.

Thus, the Intelligent Teaching Models for STEM Related Careers Using SOA and Management Science project course is an integration of theory and practice approaches. This paper focuses on the discussion of these approaches by providing a technical solution that can help in solving the problems of disease misdiagnosis in healthcare.

The instructor introduced concepts and approaches of technologies, techniques, and software tools that are needed to complete the project. The objective is to get students to be familiar with these concepts to develop the course project. The instructor divided the class into teams. Each team member had a primary task with his/her team and a secondary task with other teams. Each team selected a team leader. The role of team leaders was assigning a task to each team member, clarifying the procedures of each task, solving problems, and providing a weekly progress report to the project manager, the instructor. The tasks are based on Software Development Life-Cycle

(SDLC) phases. These phases are planning, implementation, testing, documenting, Deployment, and maintenance. The students trained on each of these phases.

At the same time, the instructor initiated IT Research Methodology that the students followed during the development of the project. The instructor presented research concepts and approaches. These include research purpose and process, research classifications, Institutional Review Board (IRB), scientific research approach, innovation, research process model, research methodology, and research criteria.

C. Course Project

Automated clustering of information relies on the ability to programmatically adapt over time to find new methodologies necessary to break data into meaningful clusters. With data constantly changing, it is desired to develop an algorithm capable of clustering in a way that is relevant to the data that is being clustered. In order to tackle this problem, the algorithm must have the ability to try numerous ways of clustering a particular data set [4-7].

In an attempt to allow for this capability, the use of an intelligent clustering based extended genetic algorithm has been put in place to provide a way of clustering data that is relevant to the type of data being clustered, with the ability to adapt over time to changes in subjects of topics of desired data. By developing such algorithm, data can evolve into information in a way that produces robust flexibility.

Researchers have often studied general algorithms and technical types of information systems which cannot entirely solve these problems. Therefore, the authors claim that the industries' organizations still face severe obstacles mainly in clustering relevant information that have adapted over time. This claim is derived from the observation of the results of disease misdiagnosis in medical and healthcare industries. Such results include unnecessary treatments and testing, long term stay for the patient, high costs and major health risks, useless resources, lateness, and unreliability[8].

In this paper, the problem of clustering intelligent web search engine using K-means algorithm has been analyzed and the need for a new data clustering algorithm such as Intelligent Clustering Based Extended Genetic Algorithm (ICEGA) is justified to improve the process of disease diagnosis. While K-means is useful and efficient when it comes to clustering data, it lacks the ability to intelligently evolve over time to user browsing patterns and collected data topics. In this paper, the concept of genetic algorithm based clustering has been modified and applied to provide better diseases clustering results in a more efficient manner [9].

The goal of ICEGA is to address the applicability of potential extended genetic algorithm to solve the efficiency and limitation problems in data clustering. To achieve this goal, this course project integrated concepts and approaches of management science, search methodologies, information extraction, intelligent information retrieval, clustering, extended genetic algorithm, and data warehousing. This project is designed and developed in a SOA environment to enable an intelligent architecture.

A prototype model is created and examined in order to validate the concepts. This project involves collaboration with domain scientist and students to evaluate ICEGA on important scientific computing application. Also, the authors collaborate with the Children's Hospital of Philadelphia to increase the number of students and underrepresented cultural minorities in undergraduate research.

III. MANAGEMENT SCIENCE

Management science is an approach of decision making that is based on a scientific methodology by significantly applying quantitative concepts to decision making [10][11].

The analysis phase of decision making process may be qualitative and quantitative [11]. Qualitative analysis is based on the medical doctor's judgment and experience with the diagnosis of a disease. However, if the doctor has little experience, then a quantitative analysis of disease diagnosis can be an especially important consideration in the doctor's final decision. When using quantitative approach, the doctor concentrates on quantitative data associated with the disease. Then, a mathematical model is developed to represent the diseases and symptoms by a system of symbols and expressions[11].

A. Data Preparation

An important step in quantitative analysis is the preparation of data required for the model. All data are calculated before the model assessment and the optimal solution development to validate the accuracy of disease diagnosis [10][11]. Since each disease is associated with a set of symptoms, a search methodology is implemented to search for diseases that match patients' symptoms. The search results return diseases from local and web databases. Web results are source code of web documents and local results are set of data records. Then, information of these results are extracted, retrieved, and processed. A parser and intelligent information retrieval techniques ensure uniqueness of the results, top-ranked diseases, and high relevancy between symptoms and diseases. These symptoms and associated diseases are stored in near-real-time database along with treatments of each symptom. The accuracy of disease diagnosis are computed according to the results using the following equations:

If all symptoms of a disease exist, let A be the accuracy and

$$A = 100\% \quad (1)$$

Let P be the positive treatments, T be the total number of treatments, and

$$A = P / T \quad (2)$$

Once the data are prepared, the models are developed.

B. Model Development

In this paper, the accuracy of disease diagnosis has been computed in Management Science by developing a combination of Network and Linear Programming models. A network model is a graphical representation of the problem. It generates particular solution actions for the accuracy of disease diagnosis problem [11]. A network model has been developed to determine the accuracy of checking symptoms with set of

diseases. It minimizes disease misdiagnosis problem by minimizing symptoms' treatment time. Then a linear programming model has been developed to determine the accuracy of positive treatments. It maximizes the accuracy of these treatments. The linear programming model is the concept of problem-solving developed for maximization or minimization of a linear objective function that is subject to linear constraints. These constraints restrict the degree of the objective function continuation [11].

When considering the accuracy of disease diagnosis problem, we usually find that the problem phase leads to a specific objective. Such objective may be maximization of accuracy of diagnosis or minimization of misdiagnosis. Also, a set of restrictions or constraints, such as positive treatment or treatment time are defined. The mathematical model and quantitative approach succeed if the objective and constraints can be precisely articulated in terms of mathematical equations or relationships [10][11].

The data preparation and model development steps are followed by the model solution step. Such step identifies decision variables values that provide optimal solution or accurate disease diagnosis. The authors built a network model as three symptoms and three diseases, and nine treatments alternatives. The time required to treat each symptom depends on the severity of each symptom corresponding to each disease.

IV. SERVICE-ORIENTED ARCHITECTURE (SOA)

Service-Oriented Architecture allows developing a flexible and adaptable infrastructure for information technology of medical organization. SOA is a growing successful paradigm that enables development of the solution of this research as smoothly integrated and reused web services [12-15]. Fig. 1 shows the Intelligent Optimization Models Architecture.

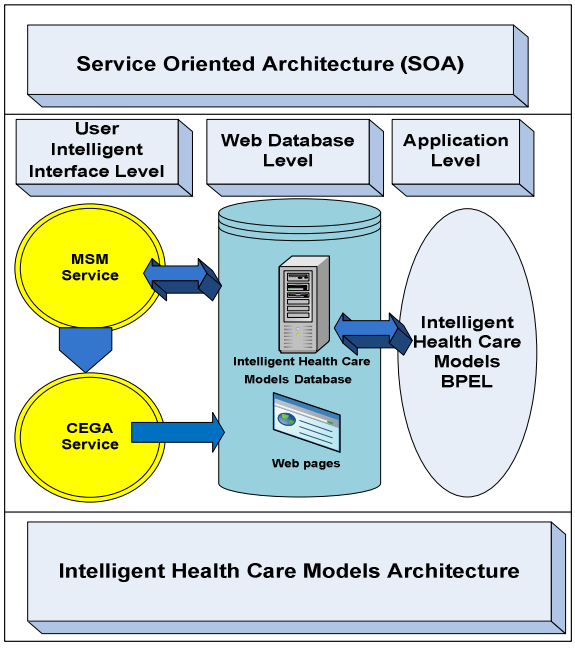

Fig. 1. Intelligent Optimization Models Architecture

Web service is a technology that enables programs to communicate through Hypertext Transfer Protocol (HTTP) on the Internet [13]. The orchestration of web services is supported by Business Process Execution Language (BPEL) [14]. BPEL is end-to-end business flows [15]. The study solution of this research mainly publishes and consumes two web services to perform operations that are required for developing the solution. They mutually exchange information with each other. Exchange of information includes simple data passing from a browser to a web server [15]. Data that are passed from a browser are symptoms. These symptoms are the input of the first web service. The output of first web service is a disease ID. This output is input of the second web service. Output of second web service is sets of clustered disease. Fig. 1 describes the intelligent optimization models architecture.

This architecture allows a new approach to the design and compilation of service-based solutions and environments [15]. The study solution of this research has been deployed, executed and tested using Oracle Application Server (OAS). The services' operations include:

- The first service is Management Science Models (MSM) web service. It performs four operations:
  1. Search Engine operation searches web and/or health care organization's databases for a set of symptoms. This operation returns only unique addresses and unique contents.
  2. Information Extraction operation extracts text from the source code of web documents.
  3. Intelligent Information Retrieval operation retrieves top ranked diseases that are relevant to a set of symptoms [16]. This operation involves diseases/symptoms representation, diseases ranking, retrieval modeling, and retrieval quality evaluation.
  4. Linear Programming and Network models operation determine the accuracy of checking symptoms with set of diseases and determine the accuracy of positive treatments of disease.
- The second service is Clustering based Extended Genetic Algorithm (CEGA) web service. It performs operations that are needed for clustering top ranked diseases. The operations of this service are described in details in Section V.

V. GENETIC ALGORITHM

The structure of genetic algorithm is extended to hold multiple populations in the population space. The Intelligent Clustering Extended Genetic Algorithm (ICEGA) is designed using artificial intelligence methodologies, not geometric approaches, to the clustering problem [17].

Our proposed method extends a genetic algorithm to find an ideal clustering solution instead of a more mathematical method such as the k-means algorithm.

This key difference allows for more adaptive behavior within our clustering method.

This paper builds a utility-based intelligent agent that implements a faster extended genetic algorithm with greater efficiency than the original algorithm.

Chromosomes are encoded to represent an extended genetic algorithm and to be parsed into tree structures, which prevents syntax crossovers and allows for mutation stages.

Once proper algorithms are put in place, the desired service item from the web part can be requested. Upon this initial request, the first generation of information retrieval is randomly generated, which can lead to a slight decrease of efficiency.

What makes up for this initial sacrifice in performance is that as the workflow processes information, the algorithm creates a new generation of logic and the results are assessed based on goodness of fit to results.

As new logic workflows are developed, they can be selected and mutated to produce better results. As this process continues, eventually the service can be provided in such a way to enable increased efficiencies over time.

Upon delivery of the user request, the generation cycle is terminated. The proposed algorithm does not have a normalization step as it does not use centroids to define the clusters like the traditional clustering method.

The ICEGA algorithm is tested on set of sample data. The data is based on 50 generations/iterations of the ICEGA or K-means respectively, using the same random sample set of 15 documents with 600 words each.

Fig. 2 describes Intelligent Clustering Based Extended Genetic Algorithm (ICEGA).

1. Create initial random population P of N individuals
2. $i \leftarrow 0$
3. If i is equal to the number of desired generations,
   return the best individual of the most recent generatic
4. $P_{i+1} \leftarrow$ empty set
5. B ← the most fit individual of the previous generatic
6. Add B to P_{i+1}
7. Insert into P_{i+1} mutate(B)
8. Repeat 7 until P_{i+1} has N individuals
9. Evaluate the fitness for all individuals from P
10. $i \leftarrow i + 1$
11. Goto 3

Fig. 2. The Algorithm

VI. RESEARCH PROTOTYPE

The study of this research implements Architected Rapid Application Development (ARAD) prototype model [18-20]. In this research, the prototype intelligent processes are Information Retrieval (IR) and Clustering based Extended Genetic Algorithm (CEGA). The prototype system is Intelligent Optimization Models Using Service-Oriented Architecture and Management Science for Disease Diagnosis. It receives patients' symptoms. Then, it returns sets of distinctive diseases with the highest score of similarity between the symptoms and diseases.

The system consists of three types of projects:

- Projects that provide services. These are MSM (Management Science Models) and CEGA (Clustering Based Extended Genetic Algorithm).
- Project that defines flow of action in SOA application. It is MSMBPEL project. It invokes projects that provide services.
- A web front-end application called the MSMInterface is provided such that the prototype system can be invoked by the users.

VII. EXPERIMENTAL RESULTS

The techniques of walkthrough are approved as experimental assessment approaches to evaluate system application usability. The walkthrough method evaluates the different phases of the research process. During the system evaluation phase, the examiners evaluated the interfaces that are related to real roles and real users [20].

The walkthrough examiners of this study are professors, researchers, and SOA engineers in North Carolina and Michigan States in the United States of America. They identified different types of problems. These types include design, development, testing, usability, and maintenance problems. They verified that the prototype satisfies requirements of this research. Also, the prototype is evidence that proves the new concept is valid, the solution is conceptualized, and the findings answer the research question and solve the research problem.

The preliminary results show that the proposed algorithm outperforms K-means algorithm. The proposed concept ensures high level of accuracy and efficiency due to removal of irrelevant information. The Intelligent Clustering based Extended Genetic Algorithm (ICEGA) enhances an organization's ability to collect information faster at lower cost and to make accurate decisions.

The orchestrations of clustering extended genetic algorithm by applying SOA principles and concepts allow flexible service workflows to be immediately adjusted to modifications and make systems smarter.

Preliminary results also show that ICEGA can discover related diseases to doctors' original diagnosis and automatically reassesses the situation if their diagnosis is incorrect. The proposed algorithm solution markedly increase

the success of disease clustering and relevancy between patient's symptoms and diseases.

The Intelligent Clustering Based Extended Genetic Algorithm (ICEGA) enhances an organization's ability to collect information faster at lower cost and to make accurate decisions. Preliminary results also show that ICEGA markedly increase the success of disease clustering and relevancy between patient's symptoms and diseases.

The preliminary results show the optimal solution of disease diagnosis accuracy problem in Fig. 3. It provides an accuracy score of 80 points for treating 8 symptoms within 40 hours or less.

In addition, the instructor asked the students to complete a job survey and return it once they obtain a job in any of the areas that they worked on during the course the project. Fig. 4. shows the number of jobs that offered to students in each of the skills learned in course project in the past two years.

In web service technology, 160 students received job offer. In SQL and XML, 150 students received job offer. In SOA and BPEL, 140 students received job offer. In Java, 120 students received job offer.

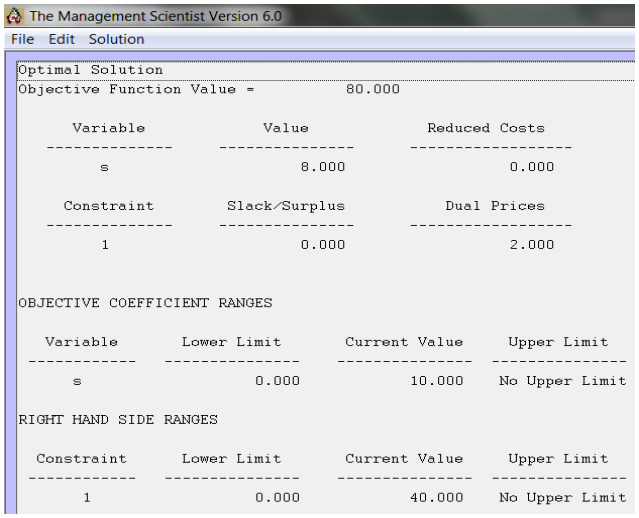

Fig. 3. Optimal Solution of Disease Diagnosis Accuracy

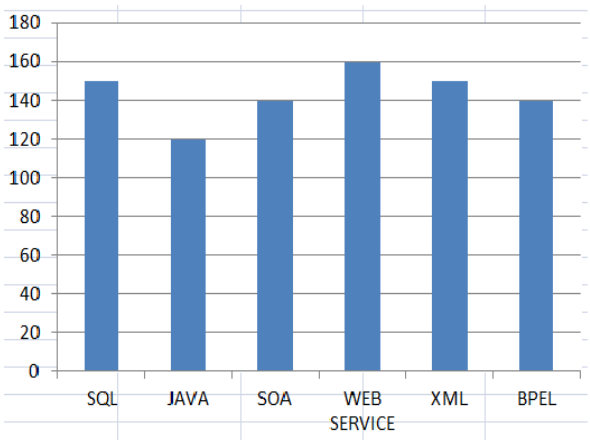

Fig. 4. Number of Jobs

VIII. CONCLUSION

Intelligent Teaching Models for STEM Related Careers Using SOA and Management Science project course is a software engineering research methodology, an educational tool, and a teaching technique. It also helps students learn theoretical concepts, practice advanced technical skills, and discover knowledge to solve problem. The course satisfies the needs of undergraduate and graduate students in computer science, engineering, education, instructional technology, medical science, and management. This new course strengthens the capacity and quality of STEM undergraduate degree programs and the number of overall graduate student enrollment. It promotes a dynamic STEM academic environment and increases the number of students entering STEM careers.

REFERENCES

[1] F. McCown. (2010). Teaching Web Information Retrieval to Undergraduates. Proceedings of the 41st ACM technical symposium on Computer science education. Milwaukee, Wisconsin.

[2] OpenEMR. Retrieved January 15, 2012, from http://www.oemr.org.

[3] M. Goulde, E. Brown. (2006). Open Source Software: A Primer for Health Care Leader. Unpublished manuscript, Brigham Young University, Provo, UT, A. (1998). Teaching vocabulary: Evidence from research in Pig Latin. Forrester Consulting.

[4] N. El-Bathy, P. Chang, G. Azar, and R. Abrahiem , "An Intelligent Search of Lifecycle Architecture for Modern Publishing and Newspaper Industries Using SOA," IEEE, Electro/Information Technology, Normal, IL, United States, pp. 1-7, 2010.

[5] N. El-Bathy and G. Azar, "Intelligent Information Retrieval and Web Mining Architecture Applying Service-Oriented Architecture," KG. Saarbrücken, Germany: LAP LAMBERT Academic Publishing AG & Co. 2010.

[6] N. El-Bathy, G. Azar, M. El-Bathy, and G. Stein, "Intelligent Extended Clustering Genetic Algorithm," unpublished.

[7] B. Coppin, Artificial intelligence illuminated. Sudbury, Massachusetts: John and Bartlett Publishers, 2004.

[8] How Common is Misdiagnosis?. Retrieved January 10, 2012, from http://www.rightdiagnosis.com/intro/common.htm.

[9] A.K. Jain, M.N. Murty, P.J. Flynn. (1999). Data Clustering: A Review. ACM Computing Surveys. Vol. 31, No. 3.

[10] M. S. Shukoor, "Fault Detection and Diagnostic Test Set Minimization", Thesis, Auburn University, Alabama, 2009.

[11] D. Anderson, D. J. Sweeney, T. A. Williams, J. D. Camm, R. K. Martin. "An Introduction to Management Science: Quantitative Approaches to Decision Making", South-Western Cengage Lerning, 2012.

[12] D. Krafzig, K. Banke, and D. Slame, Enterprise SOA Service-Oriented Architecture Best Practices. NJ: Prentice Hall, 2005

[13] A. Dan, R. Johnson, and A. Arsanjani, "Information as a service: modeling and realization," International Workshop on Systems Development in SOA Environments, IEEE, 2007.

[14] D. Steiner, "Oracle SOA Suite Quick Start Guide 10g (10.1.3.1.0)," Oracle, pp. 7 – 11,2006.

[15] D. Vladimir, "Development of applications with service-oriented architecture for grid," ACM New York, Vol. 374, 2008.

[16] R. Akerkar and P. Lingras, Building an Intelligent Web – Theory and Practice. Sudbury, Massachusetts: Jones and Bartlett Publishers, 2008.

[17] N. Chaiyaratiia and A. M. S . Zalzala, "Recent developments in evolutionary and genetic algorithms: theory and applications," Genetic Algorithms in Engineering Systems: Innovations and Applications, GALESIA 97. Second International Conference On (Conf. Publ. No. 446), pp. 270 – 277, 1997.

[18] M. Doernhoefer, “Surfing the net for software engineering notes,” ACM SIGSOFT Software Engineering Notes, Volume 30 Number 6, Nov. 2005.

[19] D. E. Rowley and D. G. Rhoades, “The cognitive jogthrough: a fast-paced user interface evaluation procedure,” CHI '92 Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, New York, 1992.

[20] D. Pinelle and C. Gutwin, “Groupware walkthrough: Adding context to groupware usability evaluation,” CHI '02 Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves, ACM, New York, 2002.