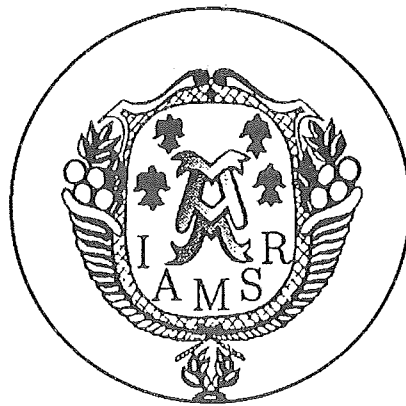


RESEARCH REPORTS

Theory and Methods on Text Summarization

Shuhua Liu

Research Report 2/2004



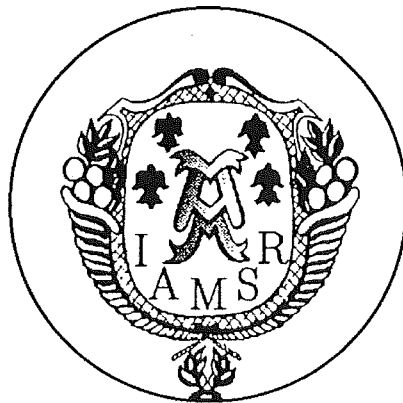
Institute for Advanced
Management Systems Research

RESEARCH REPORTS

Theory and Methods on Text Summarization

Shuhua Liu

Research Report 2/2004



Institute for Advanced
Management Systems Research

"If I had more time, I would have written a shorter letter." -- B. Pascal

Theory and Methods on Text Summarization

Shuhua Liu

*Institute for Advanced Management Systems Research IAMSAR, Åbo Akademi University
Lemminkäisenkatu 14 B, DataCity 6th floor, FIN-20520 Turku, Finland
Email: sliu@abo.fi*

1. Introduction

Summarization is an activity that we do everyday, as a way of communicating key messages from a long story or as a way of drawing up the major substances from a long document. As a matter of fact, humans often prove to be extremely capable summarizers (Mani and Maybury, 1998). Nonetheless, cognitive research on summarizing and abstracting behavior has revealed that human summarization activity is a highly context dependent, person-specific and complicated cognitive process (van Dijk and Kintsch, 1983; Endres-Niggemeyer, 1998). Summarization is even considered as an important achievement of human cognition, a work of "an art form" (Ashworth; Endres-Niggemeyer, 1998). As such, it is not difficult to understand that computer based automated abstracting and summarizing has proven to be very challenging tasks. In this paper, we set out for a preliminary investigation into the theory and methods for summarization. Our special focus here is on text summarization, especially the summarization of *natural language text*.

What is text summarization? Although there are as many different descriptions of what summarization is (or should be) as people may wish to have, there is in fact not so much disagreement in them. For example, text summarization may be described as "to reduce (long) textual information to its most essential points", "to condense information down to critical bit", or "to distill the most important information from a source or sources to produce an abridged version for a particular user (or users) and task (or tasks)" (Endres-Niggemeyer, 1998; Mani and Maybury, 1999; Spärck-Jones, 1999). These descriptions emphasize the purpose and goal of summarization.

Text summarization can also be understood from a process point of view. Humans read an entire text and understand it before summarizing it. They "take an original article, understand it, and pack it neatly into a nutshell without loss of substance or clarity" – or at least ideally so. Thus, text summarization covers both *text understanding* and *text generation*. Text understanding again is not solely a process of language processing (or text processing) based on understanding of syntax and recognition of word meanings. Rather, it is also intertwined with other cognitive activities such as thinking, or knowledge processing (reasoning), and active meaning construction. Text generation refers to the subsequent reformulation of the core content and rewriting of a new text for summary, on the basis of text understanding, by synthesizing elements from disparate

parts of the document as well as the summarizers prior knowledge (E-N, 1998). In between text understanding and text generation there is of course the step that important substance is distinguished from the unimportant ones.

A more mechanical way or computational way to look at text summarization is to see it as a *text transformation process*. For example, Spärck-Jones (1999) modeled text summarization as a three-stage text transformation activity that includes *interpretation, transformation and generation*. Interpretation refers to source text interpretation that analyzes source text and transforms it into an appropriate text representation. Transformation refers to source representation mapped into summary text representation. This involves key content recognition (as key words, key concepts, significant words and significant sentence), concept organization, synthesis of an appropriate summary output and summary text representation. Generation refers to the generation of summary text from summary representation. Such a model presents a more general concept of text summarization. Viewing text summarization as text transformation activities may or may not rely on text understanding.

What is a summary? A summary as the physical output of a summarization process is the *concise, condensed* description of the most important information or the main ideas in a text, with extraneous, details and repeated information omitted. A summary can serve as surrogate for the complete, unabridged version of a document. A summary can also serve as only a rough indication of the topics and major substance in a document but not as a substitute of the original (Salton; Jones). News headlines, article outlines, meeting minutes, preview of a movie, or review of a book, chronologies of salient events, abridgements of a book are some examples of very different types of summaries.

Although there is not much debate on what a summary is, the exact form and content of a summary vary greatly depending on its source, its nature, purpose, intended reader, and so on. In terms of structural composition, it may vary from a list of keywords to a list of independent single sentences, or a fully planned and generated coherent summary text. For example, abstracts for scientific articles give a good example of summaries that we often encounter that have a very clear structure. An abstract, as defined in the American National Standard for Writing for example, should "state the purpose, methods, results and conclusions presented in the original documents, either in that order or with an emphasis on results and conclusions". It should "use precise technical terms, use standard English; follow grammar rules; give expanded version of less known abbreviations ...". In terms of content composition, a summary of a news story may be a compact of background information or "just-the-news", or may contain both the key events or event chains and key views on the events. An abstract of a scientific article should "exclude background information and citations... should not contain any information that is not contained in the original text." (Cremmins, 1982, 1996, in Radev). In fact, there are many dimensions of variation for summaries as well as the summarization process that produce them. We will introduce the various context factors of summarization activity in the next section following the framework for text summarization environment given by Jones (1999). In section 3, we give an overview of computational methods and techniques for text summarization.

2. Context Factors in Text Summarization Activities

Context factors are the relevant “operating factors” for any particular text summarization application (Jones, 1999). Jones outlined three sets of context factors that can help us to systematically understand the forces that contribute to shaping a summarization process: *input factors*, *output factors* and *purpose factors*. They are illustrated below in Table 1. Together, these factors set up requirements and expose constraints to the summarization process, determine what is the best strategy and methods for summarization. The different combination of these factors result in the rich varieties of summarizing.

Table 1: Context Factors of Summarization (Jones, 1999)

<i>Input Factors</i> (source context)	<i>Output Factors</i> (target context)	<i>Purpose Factors</i> (usage context)
Source form	Material (Content)	Situation
Scale	Generic	Tied
Structure	Query-specific	(<i>Structured</i>)
Media	Abstract	Floating
Genre	Extract	(<i>Unstructured</i>)
Subject type	Format	Audience
Ordinary	Length	Targeted
Specialized	Media	Untargeted
Restricted	Running text vs	Use
Unit	Broken text	Retrieving aid
Single source	Style	Previewing
Multiple source	Indicative	Promotes to read
	Informative	the source
	Evaluative	Substitute of the
	Aggregative	source

2.1. Source Context

Summarization processing needs to respond to source text properties. Input factors such as *source form*, *subject type* and *unit* will define the source context, which means that the characteristics of the source along these dimensions need to be taken into consideration in the summarization process. *Source form* refers to the *scale*, *structure*, *media* and *genre* of the source text. The relevance of these factors can be explained with concrete examples. For example, we know from our own everyday summarizing experience that summarizing a new story of 2 pages is obviously different from summarizing a business report of 20 pages or a scientific book of 200 pages. *Text structure* is usually explicitly shown with the use of headings or implicitly embedded in some rhetorical patterns (e.g.

statement followed by elaboration, first sentence of the section or paragraph followed by detailed descriptions, and so on) (Jones). Different text structure can make a big difference to the summarization process too. *Source medium*, in the case of text summarization, concerns the different languages that may be encountered. *Genre* is a rather popular term in the world of linguistics and natural language processing. Jones used the term to refer to “the type of literary form” of the text rather than the type of content. In terms of genre, a text may be a *description* or a *narrative* (Jones). And again the summarization of a description will be different from that of a narrative, whether with regard to the output or the process of summarization.

The *subject type* of source text is to indicate its coverage of a subject matter in relation to the subject knowledge of the source text reader (i.e., the summary producer, not necessarily the summary user). It is broadly characterized as *ordinary*, *specialized* and *restricted*. An *ordinary* type text covers subjects that contain many individual domains with diverse knowledge (e.g. sports, gardening). A *specialized* source text concerns a subject matter that depends on knowing what people in many different locations may be assumed to know. A *restricted* source text is then a text that covers a subject matter that is limited to some specific community. (Jones)

A summary may be generated based on a single-document or based on multiple documents. The *source unit* feature (single source vs. multiple source) distinguishes summarization over a single input source (e.g. one document) from summarization over multiple sources (e.g. document collection). It has important implication for handling information redundancy, changing data, and changing opinions over time.

2.2. Target Context

Summarization also needs to take into consideration output factors such as different *material*, *format* and *style* requirements of the summary. They pose different requirements to the summarization process. Together they form the target context.

In terms of material feature, a summary may be *generic* or it may be *query-oriented*. The *material* factor may be more precisely named as a “substance” or a “content” factor. It is to indicate whether the expected summary should capture all of the important content in the source text, or only some aspects of the key content. In the former case, it is a “covering” (generic) summarization and in the latter case it is an intentionally “partial” (query-specific) summarization. Obviously different coverage of the “material” will call for different extraction/summarization strategies.

The *format* factor is to make a distinction between summaries that are “*running text*” and those that are “*headed*” *broken text* that may be incoherent, or may violate grammar rules. A summary may be in the form of an “abstract” or an “extract” (Endres-Niggemeyer, 1998; Mani and Maybury, 1999). An “extract” summary is made up of “portions of text” (sentences or passages) extracted from the original text. It will not guarantee that the result is in sound natural language. At the other end is an “abstract”

type summary that consists of novel phrasings in coherent language describing the content of the original document. We can expect that to obtain an abstract summary pose many more significant challenges than for an extract summary.

The *style* factor refers to the alternative nature of summaries: *indicative*, *informative*, *reflective*, *critical* or *aggregative*. An *indicative summary* is only there to indicate the topic or subject of the document but not to tell what the document says about the topic. A list of keywords or broken text would be enough for such purpose. An *informative summary* is “content-laden” and it tells what the source text says about the subject. A summary presented in a neutral tone will simply state the key content delivered by the author, thus a reflective summary. A *critical summary* will then add remarks on the merits and drawbacks of the source text substance. Finally, an *aggregative summary* refers to a summary of multiple sources of the same type (Spärck-Jones, 1999).

2.3. Usage Context

Purpose factors give the reasons for summarizing and build up the usage context of a summarization process. Usage context is the context where the summary is utilized. The properties that text summaries should have are determined by purpose requirements and personal preferences. That is, the target context is actually determined by the purpose context.

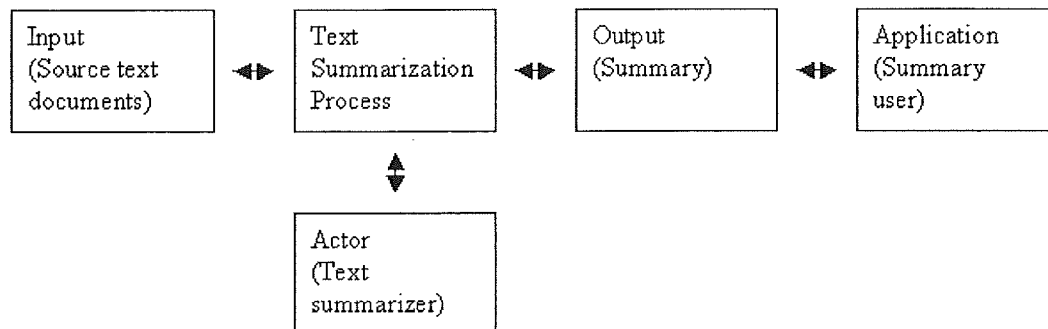
Purpose factors may be materialized as *situation factors*, *audience factors* and *use factors*. The *situation* factor accounts for the structuredness of the situation where the summary is to be used. In a “*tied*” *situation*, the particular context in which the summaries are to be used is known in advance so that summarizing can be tailored in a detailed way to meet these context requirements (e.g product description summaries adapted for the use of a company’s marketing department for a particular sales drive). This refers to the “structured” situation. In a “*floating*” *situation*, on the other hand, it is impossible to have a precise context specification in advance.

The *use* factor refers to the function or value of the summary. Possible uses for summaries could be: (1) as aids for retrieving source text; (2) as means of previewing a text about to be read (a course synopsis); (3) as information-covering substitutes for their source text; (4) as devices for refreshing the memories of an already-read source; (5) as action prompts to read their sources (Jones).

Finally, summarization is an activity that is very user-dependent as well as summarizer-dependent. The *audience factor* addresses the role of the intended summary reader (or user) of the summaries. Audience is characterized as being along a spectrum from targeted to untargeted in terms of assumed domain knowledge, expertise, language skills, and so on. The summary audience should not be taken as similar to the source audience. The summary users/readers personal preferences also exert direct influence to the summarization activity.

2.4 Text Summarization in Context

To sum up of the concept of text summarization, it is at the highest level a process of given Input Factor data (constraints), to satisfy Purpose Factor requirements, via Output Factor devices (Jones, 1999). In addition, we think that the characteristics of the summarizer (knowledge, expertise, skills) are also very relevant factors that should be reflected in the framework, as is shown below.



3. Computational Methods and Techniques for Automated Text Summarization

How is text summarization done? If we refresh our understanding about text summarization, we can say that our main task is to learn the content, to recognize, when the summarization context fully or partly known and specified, the right, critical bit of substance from the source text and, and to write it up into a preferred format (for example, a running text with appropriate length). In other words, the key issues that we have to deal with include: (1) *how to create a formal specification of the source text content?* (2) *how to identify the most important content out of the rest of the text?* (3) *how to synthesize the substance and formulate a summary text based on the identified content?*

How is text summarization done computationally? The earliest studies on automated text summarization dated back to the late 1950's with the pioneering work of Hans Peter Luhn¹ (1958), in which he demonstrated a text extraction method based on word frequency and sentence significance using a statistical approach. Luhn's work was followed by an early effort in the 60's (Climenson et al, 1961, and Edmundson, 1969),

¹ Hans Peter Luhn is known as the "father of information retrieval". He was a pioneer in information science and in engineering field. He is the person behind the development of key word in context (KWIC) indexing, selective dissemination of information (SDI), full text processing, automatic abstracting, and the first modern use of the word "thesaurus", among many other inventions (cited from <http://web.utk.edu/~jgantt/hanspeterluhn.html>). In 1958, Luhn published his article entitled "The Automatic Creation of Literature Abstracts", which marked the beginning of research work on automated text summarization.

and a relatively slow but continuous steady development period during 1970s and 1980s. A marked growth is seen since mid eighties, especially in the nineties (Mani and Maybury, 1999). Comprehensive reviews of the field can be found in Endres-Niggemeyer (1998), Spärck-Jones (1999), Mani and Maybury, (1999) and Marcu (2003). Currently, various methods and techniques exist that can be applied to deal with the different issues in text summarization. Mani and Maybury (1999) classify the various text summarization methods into three groups according to the level of text features that are used in summarizing: *surface level approach*, *entity level approach*, and *discourse level approach*.

Surface-level approaches use some *shallow text features* to represent content information. These features may include: *thematic features* based on term frequency statistics, *location features* such as position in the text, position in the paragraph, particular section, *background features* such as terms from title and headings in the text, *cue words and phrases* such as in-text summary cues “in summary”, “our investigation”, domain specific *bonus and stigma terms* such as “significant”, “impossible”, and so on. Such features can be analyzed individually or selectively combined together to form a function that is used to identify important words and significant sentences in the text. Luhn (1958) and Edmundson (1969) are the most cited work on surface-level text summarization.

Entity level approaches differ from surface level approaches in that they build an internal representation of the original text by modeling text entities (words, concept terms, value terms) and their relationships. Relationships between entities may include thesaural relationships among words (synonym, part-of relations), syntactic relations, logical relations (such as agreement, contradiction, entailment, consistency), co-occurrence (words related based on their occurring in common contexts), co-reference, and so on (Mani and Maybury, 1999). The summarization methods will make use of the pattern of connectivity among text entities to help determine what is the important content. Climenson et al (1961) presents an entity-level approach based on syntactic analysis. In addition, with the development of expert systems technology in the 1970s and 1980s, there emerged summarization systems adopting a knowledge-based approach, which is also an entity level approach. Examples of such systems can be found in E-N (1998).

Discourse-level approaches model the global structure of the text and its relation to communicative goals. This structure can include the outline of the document, the threads of topics as they are revealed in the text, as well as the rhetorical structure of the text and so on (Mani and Maybury, 1999). Such information is argued/expected to provide important information that can be used to help identify the important substance as well as to help in the formulation of a new summary text. Daniel Marcu has presented much work on the discourse analysis for text summarization (Marcu, 1996 to 2003).

In the following sections, we will look at some of the most popular summarization methods, but we will look at them very generally in two categories according to the strategies for summarization they use in approaching text summarization problems. On one hand, a summary of a text can be created as a result of fully comprehension and understanding of its content. This is *text summarization based on understanding*.

Apparently this is the way to go in order to obtain “abstract” type of summaries. So it is also *abstraction based* summarizing. On the other hand, a summary can also be merely composed of sentences or passages picked up from the original text, without having to fully understand the content. This is *text summarization without understanding*. It is also *extraction based* summarizing.

3.1. Text Summarization Based on Understanding

Ideally, an abstract must correctly and accurately reflect the contents of the original. To grasp the content of a document, its text language must first be thoroughly understood and its substance comprehended. This is how human do summarization work. While trying to learn the text content, the summarizer also interprets the content and actively reconstructs meaning according to his prior knowledge or with respect to his information gaps. The final summary output can be produced in various formats according to the circumstances (E-N, 1998).

So far, understanding based text summarization systems rely heavily on theories about human cognition, which according to E-N (E-N, 1998), are especially influenced by the theories of cognitive psychologists Kintsch and van Dijk (1983). The main propositions include: (1) in doing summarization, people need to apply necessary *literacy skills* as well as *world knowledge and domain knowledge and knowledge processing skills*. Literacy skills mainly refer to the capability to recognize word meaning, apply syntactic knowledge, reasoning, text organizing, and so on (Endres-Niggemeyer, 1998; Kintsch). (2) The use of knowledge requires the storage of knowledge. Thus human memory representation serves as a basis for text understanding and summarization. (3) The process of meaning construction is often the reasoning or inference that draws upon prior knowledge to fill the gaps of incoming information and that adapts the new knowledge to what is already in memory (E-N, 1998).

The computational solutions for understanding based text summarization are rooted very much in natural language processing and knowledge-based approach from the AI field. In Salton (1989), he proposed that a complete language analysis (or content analysis, more precisely) package for natural language understanding needs to be based on several components: *a text analysis module, a knowledge module, a search and comparison module and a problem solving module*. Such a framework conforms very well with the theoretical foundation laid in Kintsch and Dijk (1983).

The Text Analysis Module

In order to get an accurate, complete formal description of the content of a natural language text, one approach is to compare the certain elements (or entities) in the incoming text with the semantic information stored in a knowledge base to get meaning out of the text. Before this can be done, the text must be broken down into meaningful text constituents. Ideally this should be done through a deep syntactic-semantic analysis, and this is the major responsibility of the text analysis module. The text analysis module

consists of two major parts. One is a *syntactic analysis component* for syntactic structural decomposition of texts. Such syntactic decomposition is sometimes supplemented by limited, partly domain-independent semantic process. The other is a *semantic analysis component* for transforming syntactic decompositions into formal frameworks representing text meanings and content (Salton, 1989). In a complete content analysis, text sentences cannot be treated in isolation. One of the key issues here is how to ensure macro-textual *coherence* and micro-textual *cohesion*².

The Knowledge Module

To comprehend the content of a text, substantial knowledge is needed about the discourse area under consideration, and about the world at large. They are called domain knowledge and world knowledge respectively. A knowledge module will consist of a knowledge base with stored domain knowledge and world knowledge, in the form of entities (terms, concepts) and predicates that characterize and relate entities. It classifies the principle entities or concepts of interest and specifies certain relationships between them thus define the semantics of the discourse area, which form the basis for semantic interpretations, that is, resolve word meanings. The knowledge base may also contain a system of inference rules that are used to infer new facts and relations from incoming information and already existing information and knowledge (Salton, 1989). It offers a mental representation that will join elements of text information with each other and with elements of prior knowledge content.

For our human beings, cognitive schemata are recognized as the central instrument of understanding and subsequently also summarizing (Kintsch and Dijk, 1983; E-N, 1998). A schemata includes e.g. prior knowledge about the meaning structure of incoming events in a specific discourse area. For example, a reader who understands an article about a political convention uses knowledge about what goes on in political convention because he knows a cognitive schema that stores prior knowledge about that event type (E-N, pp. 310). A cognitive schemata allows us to examine the input text with regard to the most important aspects of an event or subject. They are represented computationally very often as frames and scripts (Salton, 1989; E-N, 1998).

Scripts are good at representing the cognitive structure that organizes our (sequential) knowledge about commands, common events or event chains such as events regularly reported in news articles: earthquakes, political conventions, basketball matches, etc. (E-N, 1998). For all these events, the news-readers have appropriate scripts that are formed

² Macro-textual coherence refers to focus identification and theme progression in the text, and the consistence of information presented in the text with a larger world. Micro-textual cohesion refers to the recognition of ellipsis, pronoun referents, and the proper interpretation of coordination in the text. The main themes of a text can be used to build indicative abstracts that provide information about the topic area under discussion. Additional theme aspects may be included in an expanded summary providing more detailed information about text contents. Even related facts not directly mentioned in the text can also be added sometimes to produce even more detailed text summary (Salton, 1989).

based on prior experiences and that allow them to interpret a new event. By these scripts people establish a pre-established structure of interesting or important information items and their relations. For instance, they expect political conventions to have delegates debating and voting (E-N, 1998). For an earthquake they can set up a script structure that accounts for the time and place of the quake, its strength, the injuries, the damages. So what is important to the user thus need to be included in the summary is explicitly specified by what is captured in the script or frame structures. Such knowledge base also forms the foundation for fact synthesis, concept/topic fusion and generalization, and eventually formulating a compact concise representation of the input text.

Some examples of understanding-based summarization systems can be found in E-N (1998). Those systems are developed following the advent of cognitive science and knowledge processing. The FRUMP (1982-85) system is one of the early abstraction systems based on script representation of the knowledge of situations in the world. It tries to find in the input text instances of expected events in a script, and then predict general events that are likely to be reported.

Search and Comparison, Problem Solving Module

Although the above prescription for a text understanding and summarization system seems to be a rather natural design, there are considerable difficulties in implementing such a complete analysis strategy. The biggest obstacle comes from the need for a matching (search and comparison) system to relate arbitrary input texts to a pre-constructed knowledge/semantics base (Salton, 1989). First and foremost, it is next to impossible to construct a large enough knowledge base to resolve word meaning clearly. There will always be uncertain results from semantic analysis; and it will give uncertain result, even when the semantic contexts is properly specified. Second, even for one specific discourse area there already involves a large number of facts and entities that are not so easily mapped into an unambiguous concept and meaning. Not to mention that it will be almost equally time consuming to extend the system to cover a substantial variety of texts in different subject areas.

To cope with the difficulties in deep semantic analysis, *case grammar* approach presents one of the best-known, relatively shallow, domain-independent approaches to semantic processing. Case grammar is based on *frame* knowledge representation. In case grammar analysis a number of *case frames* that are attached to certain *head concepts* are first defined. These case frames or templates in fact function as the specific application requirement. Head concepts are there to effectively represent the topic, subject area or specific meaning group of interest. Case frames all have *case identifiers*, which are associated with certain syntactic elements of a sentence or text segment³, and help to recognize from the text segments the content that can fill in the frame slots.

When case frames are properly specified, a case grammar analysis of an input text is then a process of comparing the expectations (syntactic or words) embodied in case frames

³ Text segments are the original text after syntactic analysis. A text segment may be a full sentence or not, depending on the methods used for segmenting the text.

with the case identifiers in the incoming text segments (original text after syntactic analysis) (Salton, 1989). When a match occurs, sentence elements will be assigned to the corresponding slots of the case frames thus fill in the slots with values and create a new concept instance that can be used later on in formulating summary text.

Case grammar analysis can help avoid some of the tedious work in building up semantics and knowledge base for an application domain. Specialized grammar containing necessarily complete syntactic and semantic information, and capable of analyzing certain kinds of texts in restricted subject area can be specified with reasonably small amount of effort. However, it is still not possible to deal with the ambiguities where specialized subject knowledge and commonsense information are needed for language understanding. The system will still become unmanageable when the number of rules (syntactic and semantic) in the case frames gets bigger, not to mention it still has the vulnerability and substantial inflexibility in adapting to new domains, new classes and new objects (Salton, 1989). Perhaps exactly due to these reasons, so far this approach is not often seen in the literature of text summarization studies (except a very brief mention in Spärck-Jones, 1999), although the application of case grammar analysis approach seems to have important seat in the very closely related “information extraction” school of natural language understanding studies (Soderland et al).

To complement the knowledge-based approach is the “machine learning” approach, using large text corpus as the source to draw *case frame definitions* (i.e. the *case grammars*, or in some applications called *concept nodes*) from them. Learning systems start with simple/partial structures and try to capture knowledge more completely as the system operates and gains experience in repeated text analysis process. This reduces the need for the construction of a complete semantics, to only a number of primitive semantics. A number of information extraction systems developed by the research group at the Natural Language Processing Laboratory at the University of Massachusetts (<http://www-nlp.cs.umass.edu/>) are based on case grammar approach and has achieved some very interesting results by enhancing the method with corpus based machine learning mechanisms, although it seems that learning systems have also only worked well in relatively simple and structured text environments (CRYSTAL, Soderland, 1997).

To sum up, understanding based text summarization requires that a reasonably complete formal specification of document content. Given the formal specification of document content available, a knowledge-rich approach can usually deliver a quality summary text in terms of both the substance and the presentation. However, the amount of work required in creating such systems can be huge and the resulted system usually “allow only one or a few of views of what is important in a source through pre-specified lens, regardless of whether this is a view showing what the original author would regard as significant, to be presented in the summary” (Spärck-Jones, 1999). Such an approach is only suitable for situations where information needs can be pre-defined, and when useful knowledge can be constructed in advance. It trades quality with less general applicability and less flexibility.

3.2. Text Summarization without Understanding

As we have seen, there are inherent difficulties relating to analyzing unrestricted natural language texts unambiguously using a knowledge-rich approach. When a sufficiently complete understanding and specification of the text content is impossible to obtain, some short-cut abstracting methods must be used. In fact, the earliest, and also the most widely used text summarization methods nowadays are not understanding-based. Rather, in practice, machine summarization is more often selection-based.

Selection-based means that a summary is created based on sentence extraction and then the sequential re-organization of the extracted sentences. As such, a summary is composed of certain important sentences or certain critical passages picked from the original document, without re-written. So it is also often referred to as an extraction based (vs abstraction based) approach. Extract summaries are what many of today's summarization tools can offer only.

Extraction based summarization is conceptually a very simple approach. It uses simple, well-understood statistical methodologies to replace missing language understanding. Summaries are formulated based on statistical analysis of individual or mixed surface level features such as word/phrase frequency, location, cue words, and so on, to identify key text segments (sentences or passages)⁴. The basic assumption is that, if the text content is important, it will be emphasized and reiterated. It thus treats "most important" as "most frequent".

The most widely adopted extracting method consists of first computing a significance value or weight for each content words and then computing a significance value for each sentence of the original document, based on some clustering analysis of the words and phrases included in the sentences, and then use the most highly weighted sentences to generate an abstract (Luhn, 1958; Edmundson, 1969; Salton, 1989). (For example, sentences that are concentrates of high score words (significant words) are the target sentences). A summary is generated as a body of sentences selected from the original text and organized in sequence of their appearance in the original text, sometimes followed by smoothing to certain degree (for example, by taking source sentences preceding the key sentences containing anaphoric references). An extracted summary will not guarantee that the result is in sound natural language; it will not guarantee that it will not lose important substance; and more importantly, it may cause misunderstanding when context information is ignored in formulating a summary.

The first text summarization algorithm is a selection-based method, and it was invented by Luhn in 1958. When words are input to the algorithm from the text, common/non-substantive words are deleted through table look-up. Only "content words" are retained, sorted alphabetically and stored, along with their position in the text, as well as any punctuation that is located immediately to the left and/or right of the word. Similar spellings are consolidated into word types. Any token with less than seven letter non-

⁴ New extraction methods may incorporate discourse level, or corpus scale features as explained in the following paragraphs.

matches are considered to be of the same word type (for example, frequently and frequent, 10 letters, 8 match, 2 non-match; The frequencies of word types are compared; low frequencies are deleted). The remaining words were considered significant. Remaining word types are sorted into location order. Sentence significance is determined by dividing sentences into sub-strings defined by distances between significant words. For each sub-string a significance (or representativeness) value was calculated by dividing the number of representative tokens in the cluster by the total number of tokens in the cluster. Sentences reaching a value above a preset threshold were selected for inclusion. Since it is based on word counting, the method cannot be applied directly to very long documents. However, long texts can be easily broken down into shorter sections.

Edmundson (1969) enriched Luhn's method with a number of other automatic extracting methods by making use of more surface level text features: *cue method*, *title method*, *location method*, *key method*, and the *linear combination of the four methods*. Key method is similar to Luhn's method, where sentence scores are computed based on the evaluation of word significance. Cue method uses certain words or phrases in the text as cues for important substance. For instance *bonus words* such as significant, extremely, *stigma words* such as hardly, impossible, may indicate a sentence's importance regardless of the actual sentence or word scores thus sentences with such cue words may be automatically used as part of the summary. Title method uses words in the title and headings as cues for identifying important content. It adds certain weights for words or phrases that occur in the title or subtitles of a document. Location method gives different weights to words and sentences occurring in the different parts of the document. It uses the location of a sentence in text to adjust the normal sentence score. For example, sentences under headings, sentences near the beginning and end of a document or a paragraph may be given higher weights than those in the middle, or may simply be selected automatically. The LEAD system is solely based on the "location technique", in which sentences are added to the summary based on their position in the source articles alone. All these methods can be used individually or combinatively. For example, in a summarization task, it is possible to apply the linear combination of the four features: $\alpha_1 C + \alpha_2 K + \alpha_3 T + \alpha_4 L$. However, Edmundson (1969) shows that the key method was not so important. In recent years, the new development in selection-based methods is to incorporate text features at discourse level (Marcu, 1996 to 2003), as well as make use of text features of a collection of documents (text corpus).

To sum up, in a selection-based approach to summarization, "source text is taken as its own representation without any interpretation, and this representation is then subject to a transformation stage that is simply extractive" (Spärch-Jones, 1999). The summarization methods are simple, easy to implement, and generally applicable to any text or text genres, old and new; and they can usually identify many useful sentences for inclusion in a summary. However, this advantage comes at the expense of the summary text quality. The largest drawback of the approach is that whole sentences are extracted, not rewritten. A substantial lack of coherence, lack of balance and lack of cohesion is usually evident in the output summary text due to the presence of dangling references. The output summary is of poor readability, may cause possibly incorrect or distorted comprehension, even

after smoothing (Paice, 1990; Salton, 1989; Hovy and Lin; Mani and Maybury; Spärck-Jones). The very reason is the weakness of the methods used in identifying important material and in the re-organization of selected sentences: it uses simplified “classification” techniques in “understanding” tasks.

There have been some efforts in smoothing extractions. For example, to keep syntactic-cohesion, it is suggested that sentences referring to earlier passage already used in the abstract should be included in the summary. McInvoy (1978) proposed another rather different method. He suggested that a summary could be constructed by choosing one or more complete paragraphs from the original texts. The most important paragraphs are those closest on average to all other paragraphs in the document. The output summary will consist of a particular number of high scoring paragraphs whose total length equals a pre-set threshold length for the extract. Or, the paragraphs in a text can be clustered according to their importance values. A complete cluster of high importance paragraphs can be chosen for inclusion in the extract, or one important paragraph can be chosen from each cluster (in Salton, 1989). This is a rarely mentioned approach to summarization. Still, the reality seems to be such that, as long as it is the extraction strategy that is employed, the nature of the resulted summary remains unsatisfactory to many no matter what further improvements over the extraction methods are added. All the effects of the improvement seem to remain only minor (Salton, 1989; Paice, 1990; Jones, 1999).

Another key issue in extraction based methods is, for a specific summarization task, how to determine the relative importance or weights of the different text features: cue words, location, key words, words in title and subheadings. The answer always differs for different text genres. One viable way that are applied by many researchers is the corpus-based approach, to draw empirically based language (text) models, to learn empirical rules for automated text summarization from a collection of original document as well as manually created text summarizations.

Currently there are many extract-based summarization tools available. Examples of commercialized text summarization tool include *AutoSummarize* in Microsoft Word and *ProSum* (or *NetSum*) from British Telecom. The drawbacks of such tools are very obvious with a few test of MSWord AutoSummarize.

There is then a long list of text summarization tools that are freely accessible from various research groups. For example, SweSum is a summarization tool from the Royal Institute of Technology, Sweden (<http://swesum.nada.kth.se/index-eng.html>). SweSum extracts sentences to product an extract type summary. It can be hooked up to search engine results. It can be applied to news texts or academic texts in Swedish or English. It uses a corpus based sentence extraction methods in creating the summaries. Sentences are extracted by ranking then according to weighted word level features. The weights were trained on a tagged Swedish news corpus. Lexical resources (WordNet) are also used in summarization. Another example is MEAD (<http://www.summarization.com/mead/>), a public domain multi-lingual multi-document summarization system developed by the research group of Dragomir Radev (<http://www.clsp.jhu.edu/ws2001/groups/asmd/>). It

includes multiple summarization algorithms applied to the summarization of a document collection.

3.3. Text Summarization with Combined Strategies and Techniques

Selection-based text summarization so far has been the dominating approach adopted in the practice of text summarization. Although there has been efforts in exploring understanding based summarization in the 1970s and 1980s, recent year's work has almost exclusively focused on extracts rather than abstracts (Mani and Maybury, 1999). Although an extraction-based approach has its fatal drawbacks, it offers methods eligible for immediate applications.

On the other hand, when polished, comprehensive abstract are needed, an extracting method will almost surely perform unsatisfactorily. In that case, it is necessary to analyze the texts more thoroughly and capture its content as completely as possible, thus to produce better quality summaries. However, since a deep text analysis cannot be undertaken in an unrestricted semantic environment, So far, understanding based text summarization methods based on AI approaches can only be made to work in very restricted domains (Paice, 1990; Salton, 1989). The applicability of such an approach is highly vulnerable the change of area of discourse, which is its fatal disadvantage.

The trend is thus to seek for the integrated application of both understanding based and selection-based methods, and to take advantage of the new developments in related fields such as natural language processing (NLP), machine learning, general purpose ontology and semantics development. For example, some newer summarization systems are characterized by integration of corpus linguistics and lexical semantics knowledge through use of generally available lexical resources such as WordNet. The SUMMARIST system by Hovy and Lin (1999) at the Information Science Institute, University of Southern California (<http://www.isi.edu/natural-language/projects/SUMMARIST.html>), is an example of combining NLP methods with symbolic world knowledge (WordNet). Summarist can produce summaries of web documents. It first identifies the main topics of the document using statistical techniques on features such as position, and word counts. These concepts must be interpreted so that of a chain of lexically connected sentences, the sentence with the most general concept is selected and extracted. Subsequent work will take these extracted sentences to construct a more coherent summary. When hooked up with a machine translation system (Systran), it can provide a summarizing tool for news articles in any language.

Earlier sentence extraction systems have used only text features such as word occurrences and co-occurrences, cue expressions, sentence locations that are more direct content significance indicators. However, texts and topics in the texts have other structure features that may be captured and conveyed in a more sophisticated and explicit way that would benefit the summarization work. For example, there are now methods that take into account the role of "discourse structure" as signals for important content (Daniel Marcu). Although discourse structure may be less direct marker of the content, they are not necessarily less accurate than the direct markers. How to automatically capture this

structure from the source text and use it in the condensing/synthesis transformation as well as in summary generation are the key issues that need to be addressed and again demand non-trivial efforts (Jones; Mani and Maybury,). Currently, the research group at the Information Science Institute, University of Southern California (Hovy, Marcu, Lin) is also working on the use of cue phrases and discourse structure for Summarist.

Another important recent development is in the area of multi document summarization (MDS). The CLAIR project headed by Dr. Radev at the University of Michigan is the forerunner in MDS, and MEAD is the system from his research group. It uses a collection of documents on the same subject as the source. A cluster centroid, which is a set of the most important words/most frequent words from the whole cluster, is used to pick up sentences from individual documents that are regarded as the best description of the entire cluster. MEAD relies on a Cross-Document Structure Theory, which is a theory that enables multi-document summarization through identification of cross-document rhetorical relationships within a cluster of related documents. Relationships among documents are characterized as, for example, source agreement, contradiction, subsumption, elaboration, change of perspective, change of modality, and fulfillment. How to automatically identify these relationships in order to utilize CST in the creation of personalized, dynamic multi-document summaries is still an unresolved issue.

In the AI field, there are many new developments, especially soft computing methods and tools, that can be used to extend the capability of knowledge based language understanding and processing. For example, in the field of fuzzy research, fuzzy logic based methods for representing meanings and perceptions and fuzzy reasoning methods for manipulating/computing with such representations (Zadeh, 1965, 1986, 1996), fuzzy computational linguistics (Rieger, 2001), have been proposed and studied, which may benefit the automated summarization of text. Useful mechanisms can also be borrowed from the field of machine translation, for example, The Core Language Engine from SRI International (Hiyan Alshawi, 1992; Alshawi and van Eijck) proposed a Logical Form as a target representation and a Quasi Logical Form as an intermediate level of representation for formally representing the literal meaning of natural language English sentences (also adapted for Swedish language, French, Spanish) (Rayner et al, 1996). There are also many works on the use of artificial neural networks in place of statistical methods in NLP as well as in text summarization. Honkela (1997a, 1997b) and Visa et al (2001) have studied how to use self-organizing maps in natural language processing and understanding. We will have a detailed look at these developments in another paper.

4. Summary

Summarization is always about emphasizing certain information over some others. Different people can abstract out different substance from the same source, and even the same person can formulate different summary information from the same source at different time, in different situations. It is very important to understand the context factors or the environment for summarization task before devising the methods and tools for the task.

Computational methods for text summarization can choose to imitate human summarization process thus are based on an understanding of text content. Or they can choose to make use of text features at surface level, discourse level, or corpus level in creating the output summary, thus avoid any efforts on text understanding. In the first case, the system may produce good quality summary text (cohesive, running text, good readability), but only in a very restricted field, when information needs can be specified in advance. The bottleneck of sufficiently large semantics and ontology development has prohibited it from a wide applicability. In the latter case, the methods and systems are much widely applicable, but they suffer greatly from the poor quality of summary text due to lack of cohesion among the extracted sentences or paragraphs. In either way it is impossible for the summaries produced by a computer system to be satisfactory in many cases, although useful text products can indeed be supplied. One trend is thus to explore the integrated use of different summarization approaches or techniques. At the same time, we should probably realize that, no matter what improvements over the current systems there is to come, to have computational systems that can produce equally good summaries as human summarizers, or to reach human cognitive performance, is going to take a very long time.

In our next step work, we will look at the context factors for the summarization of business texts (news and reports), and evaluate the usefulness of existing summarization methods and systems in the context of preparing industry foresight reports. We will also discuss the potentiality of soft computing methods such as artificial neural networks and fuzzy logic based computing with words (CW) in natural language understanding, and further in text summarization.

REFERENCES

- Alshaw H. (ed), *The Core Language Engine*, Cambridge, MA: The MIT Press, the ACL-MIT Press Series in Natural Language Processing, 1992, 322pp
- Alshaw H. and J. van Eijck, *Logical Forms in the Core Language Engine*
- Brigitte Endres-Niggemeyer, *Summarizing Information*, Springer-Verlag Berlin Heidelberg, September 1998, Hardcover, 374 pp
- Edmundson, *New Methods in Automatic Extracting*, 1968, in Mani and Maybury (eds), 1999
- Honkela T. (1997): *Learning to understand - general aspects of using Self-Organizing Maps in natural language processing*. Proceedings of the CASYS'97, Computing Anticipatory Systems, Liège, Belgium, August, 1997, in press.
- Honkela T., *Self-Organizing Maps in Natural Language Processing*, Doctoral dissertation, Helsinki University of Technology, Neural Networks Research Centre, 1997
- Hovy E. and C. Lin, *Automated Text Summarization in SUMMARIST*, in Mani and Maybury (eds), 1999

- Luhn H. P., The Automatic Creation of Literature Abstracts , 1958, in Mani and Maybury (eds), 1999
- Mani I. and M. T. Maybury (Eds.), Advances in Automatic Text Summarization, MIT Press; 1999, Hardcover: 442 pages
- Marcu D. and L. Gerber (2001). An Inquiry into the Nature of Multidocument Abstracts, Extracts, and Their Evaluation. Proceedings of the NAACL-2001 Workshop on Automatic Summarization, Pittsburgh, PA, June 3, 2001.
- Marcu D., 1999. The automatic construction of large-scale corpora for summarization research. In Proceedings of SIGIR'99, University of Berkeley, CA, August.
- Marcu D., Discourse Trees Are Good Indicators of Importance in Text
- Maybury M., Generating Summaries from Event Data, in Mani and Maybury (eds), 1999
- McKeown K. and D. R. Radev, Generating Summaries of Multiple News Articles, in Mani and Maybury (eds), 1999
- McKeown K., J. Robin, and K. Kukich, Generating Concise Natural Language Summaries, in Mani and Maybury (eds), 1999
- Pollock J. J. and Zamora A., Automatic Abstracting Research at Chemical Abstracts Service, in Mani and Maybury (eds), 1999
- Rieger B. B., Computing granular word meanings: a fuzzy linguistic approach in computational semiotics, in P. Wang (ed.), Computing with Words, New York: John Wiley & Sons, 2001, pp.147-208
- Salton G., 1989. Automatic text processing: the transformation, analysis, and retrieval of Information by Computer. Reading, MA: Addison-Wesley.
- Spärck-Jones K. Automatic Summarizing : Factors and Directions, in Mani and Maybury (eds), 1999
- van Dijk T. A. and W. Kintsch, Strategies of Discourse Comprehension, Academic Press: New York, 1983.
- Visa A., J. Toivonen, H. Vanharanta and B. Back, Prototype Matching: Finding Meaning in the Books of Bible, Proceedings of the 34th Hawaii International Conference on System Sciences, 2001, IEEE
- Zadeh L A., Fuzzy sets, Inf. Control 8, 338-353, 1965.
- Zadeh L A., Test-score semantics as a basis for a computational approach to the representation of meaning, Literary and Linguistic Computing 1, 24-35, 1986.
- Zadeh L A., Fuzzy Logic = Computing with Words, IEEE Transactions on Fuzzy Systems, 2, 103-111, 1996

Research Reports - Institute for Advanced Management Systems Research

1992

1. *Carlsson, C.*, On Optimization with Interdependent Multiple Criteria
2. *Walden, P., Carlsson, C.*, Enhancing Strategic Market Management with Knowledge-Based Systems
3. *Brännback, M.*, An Expert System Approach to Strategic Market Management in the Pharmaceutical Industry
4. *Sandström, T.*, A Test of the Arbitrage Pricing Theory on the Finnish Forests Industry Using Product Market Prices

1993

1. *Östermark, R.*, Introducing Economic Friction in Option Pricing
2. *Östermark, R.*, Separating Trend and Cyclical Dynamics in State Space Models with Exogenous Inputs
3. *Brännback, M.*, DSS: Ranking Competitors and Critical Success Factors
4. *Carlsson, C.*, Knowledge Formation in Strategic Management: an integrated Hyperknowledge and Expert Systems Approach
5. *Vanharanta, H.*, Competitive Financial Benchmarking with Hyperknowledge
6. *Walden, P., Carlsson, C.*, Strategic Management with a Hyperknowledge Support System
7. *Brännback, M.*, Expert Systems: Sustainable Competitive Advantages in the Drug Industry
8. *Waxlax, J.*, An Object-Oriented DSS Putting Porter to Work
9. *Carlsson, C., Walden, P.*, AHP in Political Group Decisions: A Study in the Art of Possibilities
10. *Carlsson, C., Fullér, R.*, Multiple Criteria Decision Making: The Case for Interdependence
11. *Waxlax, J.*, A Comparison: Traditional vs. Object-Oriented Knowledge-Based DSSs in Strategic Management
12. *Sandström, T.*, Measuring the Association between Cash-based and Accrual-based Financial Ratios: a Canonical Correlation Approach

1994

1. *Brännback, M.*, Decision Support for Strategic Management
2. *Brännback, M., Rantanen, P.*, Improving Managerial Decision Making
3. *Walden, P., Kokkonen, O., Carlsson, C.*, Woodstrat: A Support System for Strategic Management
4. *Carlsson, C.*, Hyperknowledge and Approximate Reasoning: A Support System for Cognitive Maps in Strategic Management
5. *Carlsson, C., Walden, P.*, Cognitive maps and a Hyperknowledge Support System in Strategic Management
6. *Carlsson, C., Fullér, R.*, Interdependence in Multiple Criteria Decision Making
7. *Carlsson, C., Walden, P.*, Strategic Planning is Strategy Formation - with a Knowledge Based Support System

1995

1. *Carlsson, C., Walden, P.*, On Fuzzy Hyperknowledge Support Systems
2. *Brännback, M.*, DSS - A New Foundation for Strategic Thinking
3. *Carlsson, C., Walden, P.*, More Effective Strategic Management with Hyperknowledge: The Woodstrat Case
4. *Finne, T.*, Creating the Framework for an Information Security Analysis. Part I
5. *Finne, T.*, Investigating the Information Security in Organisations: Crucial Factors in the Information Security. Part II
6. *Finne, T.*, CBISA - a DSS for Information Security Analysis. Part III
7. *Brännback, M., Spronk, J.*, Managing Critical Success Factors as Value-Focused Strategic Decision Making
8. *Brännkärr, S., Carlsson, C., Westerholm, G. (eds)*, Institute for Advanced Management Systems Research. Three-Year Report 1992-1994
9. *Carlsson, C.*, Hyperknowledge Support Systems as Instruments for Corporate Success

1996

1. *Carlsson, C., Westerholm, G. (eds)*, Institute for Advanced Management Systems Research, Annual Report 1995
2. *Brännback, M.*, Measuring Intelligence in Group Decision Making
3. *Brännback, M., Tetard, F., Beijar, T.*, Structuring unknown realities using group support systems
4. *Carlsson, C.*, Fuzzy Logic and Hyperknowledge. A New, Effective Paradigm for Active DSS
5. *Finne, T.*, Computer-Based Information Security Analysis: A Tool in a Company's Risk Management

6. *Finne, T.*, The Information Security Chain in a Company
7. *Finne, T.*, Do Bank and Insurance Companies Emphasise Personnel Security More than Other Business?
8. *Lindgren, N.*, A Structural Analysis of the Spread between an Interest Rate Swap and the Yield Curve Derived from Finnish Government Bonds
9. *Finne, T.*, Passwords - a Crucial Link in the Information Security Chain

1997

1. *Brännback, M.*, The Meaning of Globalisation of Business - on Knowledge and Customers
2. *Carlsson, C., Nummila, S. (eds)*, Institute for Advanced Management Systems Research, Annual Report 1996
3. *Carlsson, C.*, Aktive DSS: Pursuing some limits in decision support
4. *Brännback, M.*, Is there a New Dominant Logic of Marketing in the Internet?

1998

1. *Carlsson, C., Fullér, R., Nummila, S. (eds)*, Institute for Advanced Management Systems Research, Annual Report 1997
2. *Walden, P., Turban, E.*, Working Anywhere, Anytime and with Anyone (Looks closer than it is)

1999

1. *Carlsson, C., Fullér, R., Lahtinen, M., (eds)*, Institute for Advanced Management Systems Research, Annual Report 1998
2. *Kvassov, V.*, A Case Study of Business Transformation for Telecommunications of St. Petersburg
3. *Tétard, F.*, Reengineering a Project Management Process: a Case Study
4. *Tétard, F.*, The Changing Nature of Work and its Implications for Knowledge Management and Information Systems
5. *Anckar, B.*, Virtual Travel Agencies - Tourist Value Through Travel Information Systems
6. *Anckar, B., Walden, P.*, Destination Maui? Online Booking is no Bed of Roses
7. *Carlsson, C., Fullér, R.*, Soft Computing and the Bullwhip Effect
8. *Carlsson, C., Fullér, R.*, On Mean Value and Variance of Fuzzy Numbers
9. *Tétard, F.*, On Fragmentation of Working Time: A Study of Causes and Effects of Work Interruptions
10. *Walden, P., Turban, E.*, Virtual Team Work – Working Together in Abo and Long Beach
11. *Walden, P., Carlsson C., Liu, S.*, Industry Foresight with Intelligent Agents
12. *Carlsson C., Walden P.*, Intelligent Support Systems – The Next Few DSS Steps.

2000

1. *Carlsson, C., Fullér, R., Isoniemi, H.,(eds.)* Institute for Advanced Management Systems Research. Annual Report 1999
2. *Carlsson C., Fullér R.*, Reducing the bullwhip effect by means of intelligent, soft computing methods
3. *Anckar, B., Walden, P.*, Designing Internet Reservation and Management Software Systems for Small Peripheral Hospitality Organizations: The *HotMot* Solution
4. *Carlsson, C.*, Mobile Commerce and Value-Added Services. The Impact of Intelligent IT
5. *Anckar, B., Olofsson, S., Walden, P.*, Agents as Agents: A Virtual Assistant for Self-bookings in Travel

2001

- *Carlsson, C., Fullér, R., Svännblad, M.,(eds.)* Institute for Advanced Management Systems Research. Annual Report 2000
1. *Kvassov, V.*, Personality Types of Managers in Development of Intelligent Systems
2. *Anckar, B.*, An Empirical Assesment of Consumers' Perceived Benefits of and Barriers to Internet Commerce: Implications in Terms of Offering *e-Suitability*
3. *Anckar, B., Walden, P., Jelassi, V.*, The Relationship Between Business Models and Customer Value in Electronic Grocery Shopping: The Case of *Nettimarket.com*
4. *Tetard, F.*, Improving Project Management with Web-based Information Systems: a Case Study
5. *Bozóki, S.*, An Analysis of the Hierarchy Process
6. *Carlsson, C.*, Decision Support in Virtual Organizations: The Case for Multi-Agent Support
7. *Turban, E., Lee, M., Walden, P., Westfall R.*, On Global Electronic Commerce: Hello California, Finland and Hong Kong

2002

1. *Carlsson, C., Fullér, R., Kataja, K.,(eds.)* Institute for Advanced Management Systems Research. Annual Report 2001

2. *Carlsson, C., Kjellman, A., Industry Foresight – Management beyond Forecasting and Scenario Planning*
- 3 *Collan, M., Liu, S., Fuzzy Logic and Intelligent Agents: Towards the Next Step of Capital Budgeting Decision Support*
- 4 *Rossi de Mio, R., A Method for a Quantitative Measure of Emotional Intelligence in Face-to Face and Computer-Mediated Communication*

2003

1. *Lindholm, C.K., Liu, S, Fuzzy Clustering Analysis of the Early Warning Signs of Financial Crisis*
2. *Carlsson, C., Fullér, R., Störöling-Sarkkila, S., (eds.) Institute for Advanced Management Systems Research. Annual Report 2002*
3. *Sell, A., Patokorpi, E., Anckar, B., Walden, P., Attitudes towards Online Grocery Shopping Service among the Elderly and Disabled*
4. *Rossi de Mio, R., Emotional Intelligence in Virtual and Face-to-Face Communication: A Pilot Laboratory Experiment*
5. *Sell, A., Patokorpi, E., Anckar, B., Walden, P., Outsourcing the Grocery Shopping Service to an Online Grocery Retailer*

2004

1. *Carlsson, C., Fullér, R., Störöling-Sarkkila, S., (eds.) Institute for Advanced Management Systems Research. Annual Report 2003*
2. *Liu, S., Theory and Methods on Text Summarization*