A Comparison of Keyword- and Keyterm-based Methods for Automatic Web Site Summarization

Yongzheng Zhang
Evangelos Milios
Nur Zincir-Heywood

Technical Report CS-2004-11

October 2, 2004

Faculty of Computer Science
6050 University Ave., Halifax, Nova Scotia, B3H 1W5, Canada

A Comparison of Keyword- and Keyterm-based Methods for Automatic Web Site Summarization

Yongzheng Zhang, Evangelos Milios, and Nur Zincir-Heywood

Faculty of Computer Science, Dalhousie University

6050 University Ave., Halifax, NS, Canada B3H 1W5

{yongzhen, eem, zincir}@cs.dal.ca

Abstract

Automatic Web site summarization, which is based on keyword and key sentence extraction from narrative text, is an effective means of making the content of a Web site easily accessible to Web users. This work is directed towards summary generation based on multi-word terms extracted by the C-value/NC-value method. Keyterm-based summaries are compared with keyword-based summaries for a list of test Web sites. The evaluation indicates that keyterm-based summaries are significantly better than keyword-based summaries, which have previously been shown to be as informative as human-authored summaries.

keywords: automatic Web site summarization, text classification, keyword extraction, automatic term extraction, key sentence extraction

Contents

1 Introduction

The information overload problem on the World Wide Web has brought Web users great difficulty in information seeking tasks. Automatic Web site summarization is one of the effective ways to alleviate the information overload problem. An automatically generated Web site summary can help users get an idea of the key topics covered in the site without spending a lot of browsing time. However, to generate summaries as coherent as human authored summaries is a great challenge.

Web document summarization techniques are derived from traditional text summarization techniques. Existing text summarization systems generate summaries automatically either by "extraction" or "abstraction". Extraction-based systems [7, 12] analyze source documents using techniques such as frequency analysis to determine significant sentences based on features such as the density of keywords [25] and rhetorical relations [18] in the context. Kupiec et al. [14] construct a statistical classification model based on training documents with hand-selected extracts. The model estimates the probability that a given sentence is included in an extract based on a set of heuristic features. Summary generation for new documents proceeds with ranking sentences according to this probability. Abstraction [3], on the other hand, requires a thorough understanding of the source text using knowledge-based methods and is normally more difficult to achieve with current natural language processing techniques [11].

Unlike traditional documents with well-structured discourse, Web documents are often semi-structured, and have more diverse contents than narrative text, such as bullets, short sentences, emphasized text and anchor text associated with hyperlinks. Consequently, Web site summarization task is a non-trivial extension of the plain document summarization task due to the greater variety of possible feature sets. Research work in [25] has shown that identification of narrative text for summary generation is a key component of Web site summarization.

The aim of this paper is to extend the keyword-based method in our previous work [25] by using automatically extracted multi-word terms in identifying key sentences in the narrative text of a Web site. Keyterms and key sentences are selected to be part of a Web site summary. The keyterm-based summaries for a list of test Web sites are experimentally compared with the keyword-based summaries [25]. We statistically evaluate the performance of automatic Web site summarization under different feature sets, namely, keywords or keyterms.

The rest of the paper is organized as follows. Section 2 reviews previous Web document summarization approaches. Sections 3 and 4 explain how to generate keyword- and keyterm-based summaries, respectively. Section 5 discusses the design of our experiments and shows the evaluation results. Finally, Section 6 concludes our work and describes future research directions.

2 Related Work

Research on Web document summarization to date has either been *content-based* or *context-based*. Content-based systems [3, 6] analyze the contents and extract the significant sentences to construct a summary, while context-based systems [2, 8] analyze and summarize the context of a Web document (e.g. brief content descriptions from search engine results) instead of its contents.

Berger and Mittal [3] propose a system named OCELOT, which applies the Expectation Maximization (EM) algorithm to select and order words into a "gist", which serves as the summary of a Web document. Buyukkokten

et al. [6] compare alternative methods for summarizing Web pages for display on handheld devices. The *Keyword* method extracts keywords from the text units, and the *Summary* method identifies the most significant sentence of each text unit as the summary for the unit. The test indicates that the combined *Keyword/Summary* method provides the best performance.

Amitay and Paris [2] propose an approach, which generates single-sentence long coherent textual snippets for a target Web page based on the context of the Web page, which is obtained by tracing back-links, a service offered by search engines like Google. Experiments show that on average users prefer summary created by this system compared to the textual snippets provided by search engines. Delort et al. [8] address three important issues, *contextualization*, *partiality*, and *topicality* in any context-based summarizer and propose two algorithms, the efficiency of which depends on the size of the text content and the context of the target Web page.

In our previous work [25], we extend single Web document summarization to the summarization of complete Web sites. The "Keyword/Summary" idea of [6] is adopted, and the methodology is substantially enhanced and extended to Web sites by applying machine learning and natural language processing techniques. This approach generates a summary of a Web site consisting of the top 25 keywords and the top 5 key sentences. Since Web documents often contain diverse contents such as bullets and short sentences, the system applies machine learning and natural language processing techniques to extract the narrative content, defined as coherent and informative text, and then extracts keywords from the narrative text together with anchor text and special text (e.g. emphasized text). The key sentences are identified based on the density of keywords. Evaluation by users shows that the automatically generated summaries are as informative as human authored summaries (e.g. DMOZ[1] summaries). This work extends the keyword-based approach to a term-based summarization system.

3 Keyword-based Web Site Summarization

In this section we describe our content-based approach [25] to summarizing an entire Web site automatically based on keyword and key sentence extraction. The system consists of a sequence of stages as follows.

3.1 URL Extraction

In order to summarize a given Web site, Web pages within a short distance from the root of the site, which are assumed to describe the content of the site in general terms, are collected. A Web site crawler is designed to collect the top 1000 Web pages from the Web site domain via breadth-first search starting at the home page, assumed to be at level (depth) one. The choice of a limit of 1000 is based on the observation that there is an average of 1000 pages up to and including depth equal to 4 after crawling 60 Web sites (identified in DMOZ). The selected depth of 4 is based on a tradeoff between crawling cost and informativeness of Web pages. For each Web site, the crawler will stop crawling when either 1000 pages have been collected, or it has finished crawling depth 4, whichever comes first.

[1] http://dmoz.org

3.2 Plain Text Extraction

After the URLs of the Web pages have been collected, plain text is extracted from these Web pages by the text browser *Lynx*[2], which was found to outperform several alternative text extraction tools such as *HTML2TXT*[3] and *html2txt*[4]. Another advantage of Lynx is that it has a built-in mechanism to segment text extracted from a Web page into text paragraphs automatically.

3.3 Narrative Text Classification

The Web site summary is created on the basis of the text extracted by Lynx. However, due to fact that Web pages often contain tables of contents, link lists, or "service" sentences (e.g. copyright notices, webmaster information), it is important to identify rules for determining the text that should be considered for summarization. This is achieved in two steps. First, text paragraphs which are too short for summary generation are identified and discarded. Second, among the long paragraphs, narrative paragraphs are assumed to provide more coherent and meaningful content than non-narrative ones, so additional criteria are defined to classify *long* paragraphs into *narrative* or *non-narrative*. Only narrative paragraphs are used in summary generation.

3.3.1 Long Paragraph Classification

The decision tree learning program C5.0 [21] is applied to generate decision tree rules for filtering out *short* paragraphs, which are observed to be too short (in terms of number of words, number of characters, etc.) for summary generation, e.g., *This Web page is maintained by David Alex Lamb of Queen's University. Contact: dalamb@spamcop.net.*

For this purpose, a total of 700 text paragraphs are extracted from 100 Web pages (collected from 60 DMOZ Web sites). Statistics of three attributes *length of paragraph*, i.e. total number of characters including punctuation, *number of words*, and *number of characters in all words* (without punctuation), are recorded for each text paragraph. Then each text paragraph is manually labelled as *long* or *short*, and C5.0 is used to construct a classifier, *LONGSHORT*, for this task.

The training set consists of 700 instances. Each instance consists of the values of three attributes and the associated class. The resulting decision tree is simple: if the number of words in a paragraph is less than 20, then it is a *short* paragraph, otherwise it is classified as *long*. Among the 700 cases, there are 36 cases misclassified, leading to an error of 5.1%. The cross-validation of the classifier LONGSHORT shows a mean error of 5.9%, which indicates the classification accuracy of this classifier.

3.3.2 Narrative Paragraph Classification

Informally, whether a paragraph is narrative or non-narrative is determined by the coherence of its text. Analysis of part-of-speech patterns has proved to be effective in several Web-based applications such as query ambiguity reduction [1] and question answering [20]. It is hypothesized that the frequencies of the part-of-speech tags of the words in a paragraph contain sufficient information to identify the paragraph as narrative or non-narrative. To test

[2] http://lynx.isc.org
[3] http://user.tninet.se/~jyc891w/software/html2txt/
[4] http://cgi.w3.org/cgi-bin/html2txt

the hypothesis, a training set is generated as follows: First, 1000 Web pages are collected from 60 DMOZ Web sites, containing a total of 9763 text paragraphs identified by Lynx, among which 3243 paragraphs are classified as long. Then, the part-of-speech tags for all words in these paragraphs are computed using a rule-based part-of-speech tagger [4].

After part-of-speech tagging, attributes of percentage values of 32 part-of-speech tags [4] are extracted from each paragraph. Two more attributes are added to this set, *number of characters* and *number of words* in the paragraph. Then each paragraph is manually labelled as *narrative* or *non-narrative*. Finally, a C5.0 classifier *NARRATIVE* is trained on the training set of 3243 cases.

There are 5 rules in the resulting decision tree. Among the 3243 cases, 63.5% are classified using this rule: if the percentage of *Symbols* is less than 6.8%, and the percentage of *Preposition* is more than 5.2%, and the percentage of *Proper Singular Nouns* is less than 23.3%, then this paragraph is *narrative*. There are 260 cases misclassified, leading to an error of 8.0%. The cross-validation of the classifier NARRATIVE shows a mean error of 11.3%, which indicates the predictive accuracy of this classifier.

3.4 Keyword Extraction

Traditionally, keywords are extracted from the documents in order to generate a summary. In this work, single keywords are extracted via supervised learning. Based on such keywords, the most significant sentences, which best describe the document, are retrieved.

Keyword extraction from a body of text relies on an evaluation of the importance of each candidate keyword [6]. For Web site summarization, a candidate keyword is considered as a true keyword if and only if it occurs frequently in the Web pages of the site, i.e., the total frequency of occurrences is high.

As discussed before, Web pages are different from traditional documents. The existence of *anchor text* and *special text* (e.g., title, headings, italic text) contributes much to the difference. Anchor text is the text associated with hyperlinks, and it is considered to be an accurate description of the Web page linked to [5]. A supervised learning approach is applied to learn the significance of each category of keywords.

In order to produce decision tree rules for determining the keywords of a Web site, a data set of 5454 candidate keywords (at most 100 for each site) from 60 DMOZ Web sites are collected. For each site, a standard set of 425 stop words (*a, about, above, ...*) [9] is discarded. Next the frequency of each unique word in narrative text, anchor text and special text, is measured. Then the total frequency of each word over these three categories is computed, where the weight for each category is the same. If a word happens to appear in anchor text, which is also italicized, then it is counted twice. This effectively gives more weight to this word.

For each Web site, at most the top 100 candidate keywords are selected. For each candidate keyword, eight features of its frequency statistics (e.g., ratio of frequency to sum of frequency, ratio of frequency to maximum frequency in anchor text) in three text categories and the part-of-speech tag are extracted. In particular, the weight of a keyword is defined as the ratio of its frequency (over three categories of text) to the sum of frequency of all keywords.

Next, each candidate keyword is labelled manually as *keyword* or *non-keyword*. The criterion to determine if a candidate keyword is a true keyword is that the latter must provide important information about the Web site.

Based on frequency statistics and part-of-speech tags of these candidate keywords, a C5.0 classifier *KEYWORD* is constructed.

Among the total of 5454 cases, 222 cases are misclassified, leading to an error of 4.1%. The cross-validation of the classifier shows a mean error of 4.9%, which indicates the predictive accuracy of this classifier.

Once the decision tree rules for determining keywords have been built, they are applied to automatic keyword extraction from the Web pages of a new Web site. The top 25 keywords (ranked by overall frequency) for each site are kept as part of the summary. It is observed that 40% to 70% of keywords appear in the home page of a Web site.

3.5 Key Sentence Extraction

Once the keywords are identified, the most significant sentences for summary generation can be retrieved from all narrative paragraphs based on the presence of keywords [7]. The significance of a sentence is measured by calculating a weight value, which is the maximum of the weights for clusters within the sentence. A cluster is defined as a list of words which starts and ends with a keyword and less than 2 non-keywords must separate any two neighboring keywords [6]. A cluster's weight is computed by adding the weights of all keywords within the cluster, and dividing this sum by the total number of keywords within the cluster.

The weights of all sentences in all narrative text paragraphs are computed and the top five sentences (ranked according to sentence weight) are the key sentences to be included in the summary.

3.6 Summary Generation

The overall summary is formed by the top 25 keywords and the top 5 key sentences. These numbers are determined based on the fact that key sentences are more informative than keywords, and the whole summary should fit in a single page. Table 1 shows the generated summary of the Software Engineering Institute (SEI) Web site[5].

4 Keyterm-based Web Site Summarization

The keyword identification in [25] is based on word frequency analysis against three different categories of text, narrative text, anchor text, and special text. However, this method is unable to extract terms consisting of two or more component words. Since multi-word terms often convey specialized meaning, they may be more informative than single words. Based on this assumption, we aim to extract multi-word keyterms via automatic term extraction techniques and further identify key sentences based on the density of keyterms only.

This work introduces a keyterm-based approach which applies the same process as the keyword-based approach except in the keyword extraction phase. In the keyterm-based method, multi-word terms are extracted from narrative text automatically and the top 25 keyterms are used to identify the top 5 key sentences in the narrative text for summary generation.

[5] http://www.sei.cmu.edu

Part I. top 25 keywords
system, product, information, organization, institute, architecture, program, course, research, carnegie, defense, development, team, department, term, component, sponsor, process, design, management, education, method, technology, service, acquisition
Part II. top 5 key sentences
1. Explore the topics listed on the left for more information about software engineering practices, SEI projects, and software engineering.
2. The Software Engineering Institute (SEI) is a federally funded research and development center sponsored by the U.S. Department of Defense and operated by Carnegie Mellon University.
3. The Software Engineering Institute offers a number of courses and training opportunities.
4. The Software Engineering Institute (SEI) helps organizations and individuals to improve their software engineering management practices.
5. The Software Engineering Institute (SEI) sponsors, co-sponsors, and is otherwise involved in many events throughout the year.

Table 1: An example of keyword-based summary.

4.1 Automatic Term Extraction

Terms are known to be linguistic descriptors of documents. Automatic term extraction is a useful tool for many text related applications such as text clustering and document similarity analysis [19]. Traditional approaches to automatic term extraction are focused on information-theoretic approaches based on mutual information in detecting collocations [17]. Recently more effective systems have been developed. Krulwich and Burkey use heuristic rules such as the use of acronyms and the use of italics to extract key phrases from a document for use as features of automatic document classification [13]. Turney proposes a key phrase extraction system GenEx which consists of a set of parameterized heuristic rules that are tuned to the training documents by a genetic program [22]. Witten et al. propose a system called KEA which builds a Naive Bayes classifier using training documents with known key phrases, and then uses the classifier to find key phrases in new documents [24]. Both GenEx and KEA generalize well across domains. However, they are aimed towards extracting key phrases from a single document rather than a whole document collection.

In this work, we apply a state-of-the-art method *C-value/NC-value* [10] to extract multi-word terms from a Web site automatically. The *C-value* is a domain-independent method used to automatically extract multi-word terms. It aims to get more accurate terms than those obtained by the pure frequency of occurrence method, especially terms that may appear as nested within longer terms. The *NC-value* is an extension to *C-value*, which incorporates context word information into term extraction. Context words are those that appear in the vicinity of candidate terms, i.e. nouns, verbs and adjectives that either precede or follow the candidate term. This term extraction approach uses linguistic analysis (linguistic filter, part-of-speech tagging [4]) to generate candidate terms, and statistical analysis to sort them by a frequency-based measure of termhood (C-value/NC-value). A linguistic filter is used to extract word sequences likely to be terms, such as noun phrases and adjective phrases.

Experiments in [10, 19] show that *C-value/NC-value* method performs well on a variety of special text corpora. In particular, with linguistic filter 2 `(Adjective|Noun)`$^+$`Noun` (one or more adjectives or nouns followed by one noun), *C-value/NC-value* method extracts more terms than with linguistic filter 1 `Noun`$^+$`Noun` (one or more nouns followed by one noun) without much precision loss. For example, terms such as *artificial intelligence* and *natural language processing* will be extracted by linguistic filter 2. Hence, in our work, we experiment with both linguistic filters to extract terms from a Web site. Finally, the resulting keyterms from each linguistic filter are used to extract key sentences to summarize the Web site as described in Section 3.5.

4.2 Keyterm Identification

The candidate term list C (ranked by *NC-value*) of a Web site contains certain noun phrases (e.g. *Web page*, *Web site*, *home page*, *credit card*, *privacy statement*, ...), which appear frequently in Web sites. These noun phrases are not relevant to the content of the Web sites and hence must be treated as stop words. We experimented with 60 DMOZ Web sites and identified a stop list, L, of 51 noun phrases. The candidate term list C is filtered through the noun phrase stop list L, and the top 25 terms are selected as keyterms.

5 Experiments and Evaluation

In this section, we discuss how to evaluate and compare the quality of keyword-based and keyterm-based summaries.

5.1 KWB and KTB Summaries

In our work, both keyword-based (KWB) and keyterm-based (KTB) approaches are used to generate summaries for 20 DMOZ Web sites (in four subdirectories), which are selected randomly and are of varying size and focus [25]. Table 2 lists the URLs of these test Web sites.

We denote KTB summaries based on terms extracted by linguistic filter 1 as KTB_1 and KTB summaries based on terms extracted by linguistic filter 2 as KTB_2. Each KWB summary consists of the top 25 keywords and the top 5 key sentences as shown in Table 1. Each KTB (KTB_1 or KTB_2) summary consists of the top 25 keyterms and the top 5 key sentences. Table 3 gives an example of KTB_1 and KTB_2 summaries for the Software Engineering Institute Web site.

As we can see in Tables 1 and 3, there are 19 matches out of 25 keyterms in the KTB_1 and KTB_2 summaries of the same Web site, but the order of these keyterms is significantly different. Among the top 5 key sentences generated based on these key phrases, there are 4 matches out of 5 between any two of the three summaries, but again the order of sentences is different.

5.2 Summary Evaluation

In this subsection, we describe how to compare the quality of KWB summaries with that of KTB summaries. Evaluation of automatically generated summaries often proceeds in either of two main modes, *intrinsic* and *extrinsic*. Intrinsic evaluation compares automatically generated summaries against a gold standard (ideal summaries), which

Subdirectory	Site URL
Software/ Software Engineering	1. http://www.ispras.ru/groups/case/case.html 2. http://www.ifpug.org 3. http://www.mapfree.com/sbf 4. http://www.cs.queensu.ca/Software-Engineering 5. http://www.sei.cmu.edu
Artificial Intelligence/ Academic Departments	6. http://www.cs.ualberta.ca/∼ai 7. http://www.ai.mit.edu 8. http://www.aiai.ed.ac.uk 9. http://www.ai.uga.edu 10. http://ai.uwaterloo.ca
Major Companies/ Publicly Traded	11. http://www.aircanada.ca 12. http://www.cisco.com 13. http://www.microsoft.com 14. http://www.nortelnetworks.com 15. http://www.oracle.com
E-Commerce/ Technology Vendors	16. http://www.adhesiontech.com 17. http://www.asti-global.com 18. http://www.commerceone.com 19. http://www.getgamma.com 20. http://www.rdmcorp.com

Table 2: URLs of the 20 DMOZ Web sites used in experiments.

is very expensive to construct. Extrinsic evaluation measures the utility of automatically generated summaries in performing a particular task (e.g., classification) [16]. In this work, however, we evaluate the quality of KWB and KTB summaries in a different way which has been extensively used in related work [15, 19, 23]. Four human subjects are asked to read and evaluate summaries. In order to avoid bias towards any type of summaries, each subject reads 5 KWB, 5 KTB_1 and 5 KTB_2 summaries, which are different from the summaries assigned to other subjects. And the order of summaries is random, so that the subjects do not know which method is used to produce a specific summary. Then they judge the relatedness of key phrases and key sentences to the essential topics covered in the Web site as follows:

1. Browse the Web site for a sufficient time in order to extract two essential topics from each test Web site.
2. Read KWB and KTB summaries and rank each **summary item** (i.e. keyword, keyterm, or key sentence) into *good*, *fair* or *bad* using the following rules:
   - If it is pertinent to both of the two topics of the Web site, rank it *good*.
   - If it is strongly pertinent to one of the two topics, rank it *good*.

Part I. KTB_1 top 25 keyterms	**Part I. KTB_2 top 25 keyterms**
engineering institute, software engineering, software engineering institute, product line, carnegie mellon, development center, software architecture, software development, software product, software product line, software process, system component, process improvement, design decision, coordination pattern, reference architecture, software system, coordination protocol, infrastructure capability, application developer, whiteboard course attendees stryker infantry carrier vehicle, system architecture, capability maturity, target system, risk management	engineering institute, software engineering institute, software engineering, product line, software architecture, carnegie mellon university, capability maturity, capability maturity model, carnegie mellon, maturity model, software process, mellon university, process improvement, development center, system component, software development, software system, reference architecture, personal software process, software product line, capability maturity model integration, target system, design decision, software product, team software process
Part II. KTB_1 top 5 key sentences	**Part II. KTB_2 top 5 key sentences**
1. The Software Engineering Institute (SEI) is a federally funded research and development center sponsored by the U.S. Department of Defense and operated by Carnegie Mellon University. 2. 2002 SEI Annual Report Published The online version of the Annual Report of the Software Engineering Institute (SEI), reporting on fiscal year 2002, is available at http://www.sei.cmu.edu/annual-report/. 3. The Software Engineering Institute (SEI) helps organizations and individuals to improve their software engineering management practices. 4. The Software Engineering Institute (SEI) sponsors, co-sponsors, and is otherwise involved in many events throughout the year. 5. The Software Engineering Institute offers a number of courses and training opportunities.	1. The Software Engineering Institute (SEI) is a federally funded research and development center sponsored by the U.S. Department of Defense and operated by Carnegie Mellon University. 2. The Software Engineering Institute (SEI) sponsors, co-sponsors, and is otherwise involved in many events throughout the year. 3. The Software Engineering Institute offers a number of courses and training opportunities. 4. The Software Engineering Institute (SEI) helps organizations and individuals to improve their software engineering management practices. 5. The SEI provides the technical leadership to advance the practice of software engineering so the DoD can acquire and sustain its software-intensive systems with predictable and improved cost, schedule, and quality.

Table 3: An example of KTB_1 and KTB_2 summaries of the SEI Web site.

- If it is pertinent to one of the two topics, rank it *fair*.
- If it is not pertinent to any of the two topics at all, rank it *bad*.

3. Count the number of *good*/*fair*/*bad* items in each summary.

Let n_g, n_f, and n_b be the number of good, fair, and bad summary items, respectively. For example, in the summary shown above, the two essential topics for the Web site could be: 1) Software Engineering Institute at Carnegie Mellon University, and 2) software engineering management and practice. In the KWB summary, there are 13 good, 10 fair, and 2 bad keywords; and 4 good, 1 fair, and 0 bad key sentences. Details of KWB, KTB_1, and KTB_2 summary item numbers are listed in Table 4.

The average number of good, fair, and bad summary items per Web site summary is listed in the bottom line of Table 4. Related research in [23] defines *acceptable* terms as good and fair terms. The percentage of acceptable key phrases (keywords or keyterms), p, is formally represented by Equation 1.

$$p = \frac{n_g + n_f}{n_g + n_f + n_b}. \tag{1}$$

The values of p in KWB, KTB_1, and KTB_2 summaries above are
$\frac{13+10}{25} = 92.0\%$, $\frac{20+4}{25} = 96.0\%$, and $\frac{23+2}{25} = 100.0\%$, respectively.

Further we assign weights 1.0, 0.5 and 0 to good, fair, and bad summary items, respectively. Let kp be the quality value of key phrases and ks be the quality value of key sentences in KWB and KTB summaries, respectively. These values are formally represented by Equation 2.

$$kp, ks = \frac{1.0 \times n_g + 0.5 \times n_f + 0.0 \times n_b}{n_g + n_f + n_b}. \tag{2}$$

For example, the key phrase quality value kp for the KWB summary above is calculated as $\frac{1.0\times13+0.5\times10}{13+10+2} = 0.72$, and the key sentence quality value is $\frac{1.0\times4+0.5\times1}{4+1} = 0.90$.

Finally let s be the quality value of KWB and KTB summaries. We give equal weights to key phrases and key sentences when calculating the summary value, which is formally represented by Equation 3.

$$s = 0.5 \times kp + 0.5 \times ks. \tag{3}$$

Table 5 summarizes the quality values of 20 Web site summaries.

Figure 1 shows the quality values of key phrases from three different approaches. As we can see, key phrases in KTB_1 summaries achieve higher scores than those in KWB summaries in 11 out of 20 Web sites. Key phrases in KTB_2 summaries achieve higher scores than those in KTB_1 summaries in 12 out of 20 Web sites. This indicates that key phrases in KTB_2 summary are generally better than those in KTB_1 summary, which are further better than those in KWB summary.

Figure 2 shows that key sentences in KTB_1 summaries outperform those in KWB summaries with 9 wins, 9 ties and only 2 losses, and that key sentences in KTB_2 summaries outperform those in KTB_1 summaries with 9 wins, 5 ties and 6 losses.

Figure 3 indicates that KTB_1 summaries are generally better than KWB summaries with 15 wins, 1 tie, and only 4 losses, and that KTB_2 summaries are generally better than KTB_1 summaries with 13 wins, 1 tie, and 6 losses.

Summary	**KWB**						**KTB$_1$**						**KTB$_2$**					
Item	25 keywords			5 key sentences			25 keyterms			5 key sentences			25 keyterms			5 key sentences		
Site	n_g	n_f	n_b	n_g	n_f	n_b	n_g	n_f	n_b	n_g	n_f	n_b	n_g	n_f	n_b	n_g	n_f	n_b
1	19	3	3	4	1	0	20	3	2	4	1	0	19	5	1	4	1	0
2	12	10	3	2	2	1	15	5	5	4	1	0	16	4	5	4	1	0
3	14	7	4	2	1	2	13	7	5	2	2	1	15	9	1	3	1	1
4	16	5	4	1	3	1	14	6	5	2	2	1	14	7	4	2	3	0
5	13	10	2	4	1	0	20	4	1	4	1	0	23	2	0	5	0	0
6	13	9	3	1	3	1	14	7	4	1	4	0	18	3	4	1	4	0
7	11	9	5	1	3	1	15	5	5	1	3	1	17	5	3	3	2	0
8	13	10	2	3	2	0	12	8	5	1	3	1	16	5	4	3	2	0
9	11	9	5	3	1	1	13	8	4	3	1	1	14	7	4	4	1	0
10	12	9	4	1	3	1	14	6	5	1	3	1	12	7	6	0	5	0
11	14	9	2	2	2	1	12	9	4	2	2	1	14	7	4	2	3	0
12	12	4	4	3	1	1	13	8	4	3	1	1	14	10	1	2	3	0
13	13	7	5	1	3	1	13	7	5	1	4	0	20	4	1	0	5	0
14	11	10	4	1	2	2	12	9	4	4	1	0	10	7	8	2	2	1
15	10	6	2	1	2	2	12	10	3	2	2	1	12	7	6	1	3	1
16	11	8	6	2	3	0	16	7	2	1	3	1	16	3	6	3	2	0
17	18	4	3	4	1	0	18	4	3	4	1	0	17	5	3	5	0	0
18	10	9	6	1	3	1	10	12	3	4	1	0	17	5	3	3	1	1
19	13	8	4	0	5	0	15	9	1	0	5	0	13	8	4	0	4	1
20	8	7	6	1	3	1	14	9	2	2	3	0	14	7	4	2	2	1
Average	13	8	4	1.9	2.2	0.9	14	7	4	2.3	2.2	0.5	16	6	3	2.4	2.3	0.3

Table 4: Details of good, fair, and bad key phrase or key sentences, denoted as n_g, n_f, and n_b, respectively, in each KWB, KTB$_1$, and KTB$_2$ summary for a list 20 Web sites.

Site	kp_w	kp_{t_1}	kp_{t_2}	ks_w	ks_{t_1}	ks_{t_2}	s_w	s_{t_1}	s_{t_2}
1	0.82	0.86	0.86	0.90	0.90	0.90	0.86	0.88	0.88
2	0.68	0.70	0.72	0.60	0.90	0.90	0.64	0.80	0.81
3	0.70	0.66	0.78	0.50	0.60	0.70	0.60	0.63	0.74
4	0.74	0.68	0.70	0.50	0.60	0.70	0.62	0.64	0.70
5	0.72	0.88	0.96	0.90	0.90	1.00	0.81	0.89	0.98
6	0.70	0.70	0.78	0.50	0.60	0.60	0.60	0.65	0.69
7	0.62	0.70	0.78	0.50	0.50	0.80	0.56	0.60	0.79
8	0.72	0.64	0.74	0.80	0.50	0.80	0.76	0.57	0.77
9	0.62	0.68	0.70	0.70	0.70	0.90	0.66	0.69	0.80
10	0.66	0.68	0.62	0.50	0.50	0.50	0.58	0.59	0.56
11	0.74	0.66	0.70	0.60	0.60	0.70	0.67	0.63	0.70
12	0.70	0.68	0.76	0.70	0.70	0.70	0.70	0.69	0.73
13	0.66	0.66	0.88	0.50	0.60	0.50	0.58	0.63	0.69
14	0.64	0.66	0.54	0.40	0.90	0.60	0.52	0.78	0.57
15	0.72	0.68	0.62	0.40	0.60	0.50	0.56	0.64	0.56
16	0.60	0.78	0.70	0.70	0.50	0.80	0.65	0.64	0.75
17	0.80	0.80	0.78	0.90	0.90	1.00	0.85	0.85	0.89
18	0.58	0.64	0.78	0.50	0.90	0.70	0.54	0.77	0.74
19	0.68	0.78	0.68	0.50	0.50	0.40	0.59	0.64	0.54
20	0.55	0.74	0.70	0.50	0.70	0.60	0.52	0.72	0.65
Average	0.682	0.713	0.739	0.605	0.680	0.715	0.644	0.697	0.727

Table 5: Quality values of key phrases, key sentences, and overall summaries, denoted as $kp_{w/t_1/t_2}$, $ks_{w/t_1/t_2}$, and $s_{w/t_1/t_2}$, respectively.

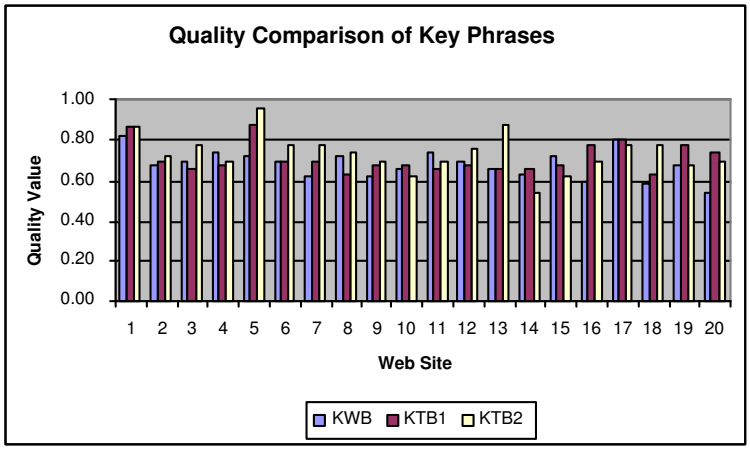

Figure 1: Comparison of quality values of key phrases in KWB summaries and KTB summaries of 20 test Web sites.

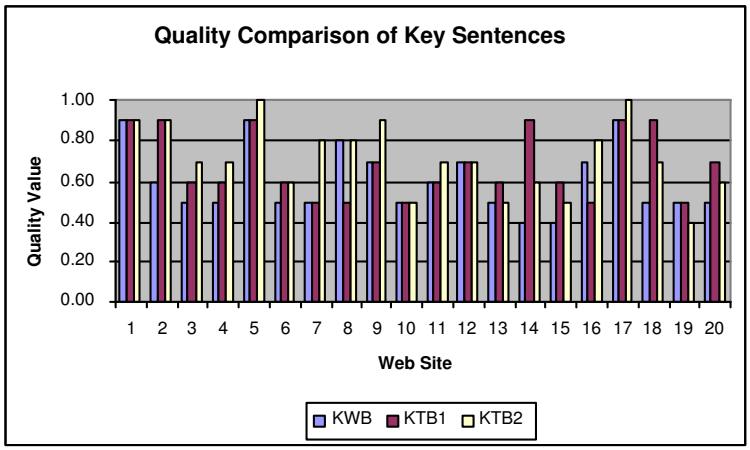

Figure 2: Comparison of quality values of key sentences in KWB summaries and KTB summaries of 20 test Web sites.

In order to statistically measure if the differences between summaries created by three methods are significant, we apply two-tail paired t-tests, which generally compares two different methods used for experiments carried in pairs. It is the difference between each pair of measurements which is of interest.

For example, when comparing the quality of KTB_1 summaries and KWB summaries, we have 20 pairs of quality values of summaries, s_{t1_i}, s_{w_i} (i = 1, 2, ..., 20), which are independent observations from the two samples in KTB_1 approach and KWB approach, respectively. Then the differences $d_i = s_{t1_i} - s_{w_i}$ (i = 1, 2, ..., 20) will be a sample of size n ($n = 20$) from a population with mean zero. Furthermore, if the populations, where the above two samples are drawn from, are approximately normally distributed, then the differences will also be approximately normally distributed. If the observed average difference is denoted by $\overline{d}$, the standard deviation of the observed differences by s_d, and the t-test statistic by t, then we have the following equations:

$$\overline{d} = \frac{\sum_{i=1}^{n} d_i}{n} \tag{4}$$

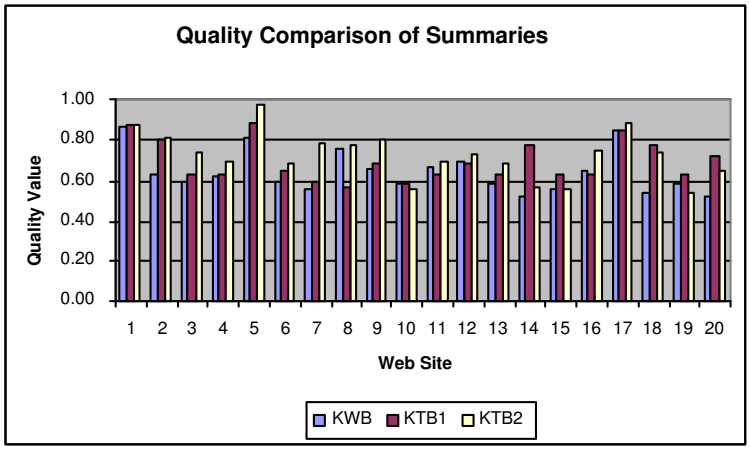

Figure 3: Comparison of quality values of KWB summaries and KTB summaries of 20 test Web sites.

$$s_d^2 = \frac{\sum_{i=1}^{n} (d_i - \overline{d})^2}{n-1} \quad (5)$$

$$t = \frac{\overline{d}}{s_d/\sqrt{n}}. \quad (6)$$

The null hypothesis H_0 and the alternative hypothesis H_1 are given by:

$H_0 : d = 0$ (KTB_1 and KWB have the same performance), and

$H_1 : d > 0$ (KTB_1 is significantly better than KWB).

If H_0 is true, then the distribution of t will be a t-distribution with $n-1$ degrees of freedom, as the estimate s_d is calculated from n differences.

From Table 5, we have: $\overline{d} = 0.053$, $s_d^2 = 0.011$, $s_d = 0.105$, $t = 2.238$.

By checking the t-table, we have $t_{0.05,19} = 2.093$. Since $t > t_{0.05,19}$, it is reasonable to reject the null hypothesis H_0, i.e., there is a significant difference between the quality values of summaries obtained from the two methods. More precisely, the KTB_1 approach performs significantly better than the KWB approach.

Comparisons of the three methods via t-tests are summarized in Table 6, which shows that both KTB_1 and KTB_2 methods are significantly better than KWB method, and that there is no significant difference between KTB_1 method and KTB_2 method.

Method	KWB	KTB_1
KTB_1	$t_0 = 2.238$ $Pvalue < 0.040$	
KTB_2	$t_0 = 4.951$ $Pvalue < 0.001$	$t_0 = 1.378$ $Pvalue = 0.184$

Table 6: Pairwise t-test results for the three methods.

6 Conclusion and Discussion

In this paper, we apply automatic term extraction techniques in a keyterm-based approach to automatic Web site summarization. Our approach relies on a Web crawler that collects shallow Web pages from a Web site and summarizes them off-line. It applies machine learning and natural language processing techniques to extract and classify narrative paragraphs from the Web site, from which keyterms are then extracted. Keyterms are in turn used to extract key sentences from the narrative paragraphs that form the summary, together with the top keyterms. We demonstrate that keyterm-based summaries are significantly better than former keyword-based summaries.

Future research involves several directions.

- Use of machine learning in setting the relative weights for keywords from narrative, anchor and special text.
- Hierarchical summarization of complex Web sites that may include a multitude of topics, for example Web sites of large organizations (e.g., government, university).
- Application of the keyterm-based approach to summarizing the Web pages returned by a query to a search engine, after clustering the returned pages.
- Integration of keyword- and keyterm-based methods in Web document corpus summarization.
- Refinement of the evaluation process, including extrinsic evaluation.

Acknowledgements

This research has been supported by grants from the Natural Sciences and Engineering Research Council of Canada.

References

[1] J. Allan and H. Raghavan. Using Part-of-speech Patterns to Reduce Query Ambiguity. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 307–314, Tampere, Finland, August 11–15, 2002.

[2] E. Amitay and C. Paris. Automatically Summarising Web sites: Is There a Way Around It? In *Proceedings of the Ninth ACM International Conference on Information and Knowledge Management*, pages 173–179, McLean, VA, USA, November 6–11, 2000.

[3] A. Berger and V. Mittal. OCELOT: A System for Summarizing Web Pages. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 144–151, Athens, Greece, July 24–28 2000.

[4] E. Brill. A Simple Rule-Based Part of Speech Tagger. In *Proceedings of the Third Conference on Applied Natural Language Processing*, pages 152–155, Trento, Italy, March 31–April 3 1992.

[5] S. Brin and L. Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. In *Proceedings of the Seventh International World Wide Web Conference*, pages 107–117, Brisbane, Australia, April 14–18, 1998.

[6] O. Buyukkokten, H. Garcia-Molina, and A. Paepcke. Seeing the Whole in Parts: Text Summarization for Web Browsing on Handheld Devices. In *Proceedings of Tenth International World Wide Web Conference*, pages 652–662, Hong Kong, China, May 01–05, 2001.

[7] W. Chuang and J. Yang. Extracting Sentence Segments for Text Summarization: A Machine Learning Approach. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 152–159, Athens, Greece, July 24–28, 2000.

[8] J. Delort, B. Bouchon-Meunier, and M. Rifqi. Enhanced Web Document Summarization using Hyperlinks. In *Proceedings of the Fourteenth ACM Conference on Hypertext and Hypermedia*, pages 208–215, Nottingham, UK, August 26–30, 2003.

[9] C. Fox. Lexical Analysis and Stoplists. In W. Frakes and R. Baeza-Yates, editors, *Information Retrieval: Data Structures and Algorithms*, pages 102–130, 1992.

[10] K. Frantzi, S. Ananiadou, and H. Mima. Automatic Recognition of Multi-word Terms: the *C-value/NC-value* Method. *International Journal on Digital Libraries*, 3(2):115–130, August 2000.

[11] J. Goldstein, M. Kantrowitz, V. Mittal, and J. Carbonell. Summarizing Text Documents: Sentence Selection and Evaluation Metrics. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 121–128, Berkeley, CA, USA, August 15–19, 1999.

[12] J. Goldstein, V. Mittal, J. Carbonell, and J. Callan. Creating and Evaluating Multi-document Sentence Extract Summaries. In *Proceedings of the Ninth ACM International Conference on Information and Knowledge Management*, pages 165–172, McLean, VA, USA, November 6–11, 2000.

[13] B. Krulwich and C. Burkey. Learning User Information Interests through the Extraction of Semantically Significant Phrases. In M. Hearst and H. Hirsh, editors, *AAAI Spring Symposium Technical Report SS-96-05: Machine Learning in Information Access*, pages 110–112, 1996.

[14] J. Kupiec, J. Pedersen, and F. Chen. A Trainable Document Summarizer. In *Proceedings of the Eighteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 68–73, Seattle, WA, USA, July 09–13, 1995.

[15] W. Lu, J. Janssen, E. Milios, and N. Japkowicz. Node Similarity in Networked Information Spaces. Technical Report CS-2001-03, Faculty of Computer Science, Dalhousie University, Halifax, NS, Canada, September 26, 2001.

[16] I. Mani, T. Firmin, D. House, G. Klein, B. Sundheim, and L. Hirschman. The TIPSTER SUMMAC Text Summarization Evaluation. In *Proceedings of the Ninth Conference of the European Chapter of the Association for Computational Linguistics*, pages 77–85, Bergen, Norway, June 8–12, 1999.

[17] C. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing.* MIT Press, Cambridge, MA, USA, June 18 1999.

[18] D. Marcu. From Discourse Structures to Text Summaries. In I. Mani and M. Maybury, editors, *Proceedings of the ACL/EACL'97 Workshop on Intelligent Scalable Text Summarization*, pages 82–88, Madrid, Spain, July 1997.

[19] E. Milios, Y. Zhang, B. He, and L. Dong. Automatic Term Extraction and Document Similarity in Special Text Corpora. In V. Kešelj and T. Endo, editors, *Proceedings of the Sixth Conference of the Pacific Association for Computational Linguistics*, pages 275–284, Halifax, NS, Canada, August 22–25, 2003.

[20] D. Radev, W. Fan, H. Qi, H. Wu, and A. Grewal. Probabilistic Question Answering on the Web. In *Proceedings of the Eleventh International World Wide Web Conference*, pages 408–419, Honolulu, Hawaii, USA, May 7–11, 2002.

[21] RULEQUEST RESEARCH. C5.0: An Informal Tutorial. Available at `http://www.rulequest.com/see5-unix.html`, last visited on March 26, 2004.

[22] P. Turney. Learning Algorithms for Keyphrase Extraction. *Information Retrieval*, 2(4):303–336, May 2000.

[23] P. Turney. Coherent Keyphrase Extraction via Web Mining. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, pages 434–439, Acapulco, Mexico, August 9–15, 2003.

[24] I. Witten, G. Paynter, E. Frank, C. Gutwin, and C. Nevill-Manning. KEA: Practical Automatic Keyphrase Extraction. In *Proceedings of the Fourth ACM Conference on Digital Libraries*, pages 254–256, Berkeley, CA, USA, August 11–14, 1999.

[25] Y. Zhang, N. Zincir-Heywood, and E. Milios. Summarizing Web Sites Automatically. In Y. Xiang and B. Chaib-draa, editors, *Advances in Artificial Intelligence, Proceedings of the Sixteenth Conference of the Canadian Society for Computational Studies of Intelligence*, pages 283–296, Halifax, NS, Canada, June 11–13, 2003.

A Noun Phrase Stop List

bay area
canada site
content area
content index
content type
corporate logo
costa rica
credit card
credit information
customer service
detail information
document owner
form list
home page
java script
los angeles
mcgraw hill
meta data
meta http
meta name
prentice hall
privacy act
privacy statement
robot txt
script language
service claim
site index
site map
site search
springer verlag
style sheet
subject index
table content
text html
top home
united states
us site

web browser

web content

web crawler

web document

web education

web magazine

web master

web page

web robot

web server

web site

web space

web spider

web worm

B KWB Summaries of the 20 Test Web Sites

B.1 Department for CASE tools of the Institute for System Programming

`http://www.ispras.ru/groups/case/case.html`

Part I. top 25 keywords
software, scenario, sdl, system, mansurov, klocwork, code, tools, case, architecture, development, requirements, specification, model, msc, language, department, engineering, proc, notation, existing, telecommunication, validation, generation, design
Part II. top 5 key sentences
1. Since 1994 the Department for CASE tools performed several joint research and development projects, subcontracted to the Institute for System Programming by Nortel Networks (Canada) and Telelogic AB (Sweden). 2. Department for CASE tools of the Institute for System Programming (short for Computer-Aided Software Engineering tools) performs research and development in the area of modern tool support for software development. 3. Department of CASE tools performs world-leading research in building next-generation model-based tools that accelerate software development through the use of formal modeling techniques, automated transformations, including advanced code generation techniques, managed software architectures, validation and verification, especially at the early phases of the development process. 4. Current members of research staff of the Department for CASE tools actively participate in international standardization on formal software development techniques in OMG and ITU-T. 5. Department for CASE tools is involved in MDA standardization though Klocwork, Inc., who is a permanent Domain-Member of OMG.

B.2 International Function Point Users Group

`http://www.ifpug.org`

Part I. top 25 keywords
workshop, software, project, bulletblue, program, member, counting, measurement, conference, tool, cfps, information, course, cpm, certification, participant, box, committee, process, download, language, isbsg, analysis, selection, article
Part II. top 5 key sentences
1. The International Software Benchmarking Standards Group (ISBSG) Benchmark data can be used to demonstrate the many advantages of function point analysis. 2. An application baseline is the size of the current functions provided to the user by the system. 3. The International Standards Organization (ISO) Task Group is working on the behalf of IFPUG members to advance Function Points as an international standard through ISO. 4. In support of this, IFPUG maintains the Function Point Counting Practices Manual, the recognized industry standard for FPA. 5. Also, through industry and academic relationships, IFPUG sponsors and supports projects for applied research on software measurement issues, and conducts studies in support of advancing the Function Point Counting Standards.

B.3 Software Build and Fix

`http://www.mapfree.com/sbf`

Part I. top 25 keywords
tcl, tk, name, widget, command, procedure, section, window, example, book, exercise, programmer, code, object, solution, value, arg, author, return, character, option, programming, language, argument, scripts
Part II. top 5 key sentences
1. Tcl was meant to be portable and Tcl/Tk has been ported to versions of Microsoft Windows and to the MacIntosh. 2. John Ousterhout, the creator of Tcl/Tk, is chief technology officer of Ajuba Solutions. 3. This company maintains a Tcl/Tk web site where you can download Tcl/Tk. 4. This is the "main" Tcl/Tk FAQ with links to several others. 5. The current version, 8.3, of Tcl/Tk adds to, rather than alters, earlier versions.

B.4 Software Engineering Archives

`http://www.cs.queensu.ca/Software-Engineering`

Part I. top 25 keywords
tool, software, system, type, object, change, contact, entry, process, level, case, ltd, group, point, design, model, environment, project, version, management, example, development, product, name, code
Part II. top 5 key sentences
1. The Empirical Software Engineering Research Group at Bournemouth University maintains a bibliography on OO metrics, originally maintained by Robin Whitty of South Bank University. 2. There is a reference model of end-user services for software engineering environments (e.g., requirements, design, code, test, tracing, planning, publications, plus about 50 others) called the Project Support Environment Reference Model that was developed by the PSESWG (Project Support Environment Standards Working Group). 3. A tightly integrated, interactive software engineering environment for total lifecycle control of Ada projects. 4. Brad Myers (Brad.Myers@cs.cmu.edu) maintains a list of user interface software tools, which are tools that can help to create the user interface part of the software. 5. This file is an index of CASE tools, sorted by category.

B.5 Software Engineering Institute

`http://www.sei.cmu.edu`

Part I. top 25 keywords
system, product, information, organization, architecture, institute, program, course, research, carnegie, defense, development, team, department, term, component, sponsor, process, design, management, education, method, technology, service, acquisition
Part II. top 5 key sentences
1. Explore the topics listed on the left for more information about software engineering practices, SEI projects, and software engineering. 2. The Software Engineering Institute (SEI) is a federally funded research and development center sponsored by the U.S. Department of Defense and operated by Carnegie Mellon University. 3. The Software Engineering Institute offers a number of courses and training opportunities. 4. The Software Engineering Institute (SEI) helps organizations and individuals to improve their software engineering management practices. 5. The Software Engineering Institute (SEI) sponsors, co-sponsors, and is otherwise involved in many events throughout the year.

B.6 Artificial Intelligence Research Group

`http://www.cs.ualberta.ca/~ai`

Part I. top 25 keywords
problem, system, search, model, research, method, information, approach, algorithm, technique, agent, computer, constraint, knowledge, result, network, space, recognition, computation, class, graph, image, distribution, task, number
Part II. top 5 key sentences
1. Smodels has been successfully used to solve hard satisfiability problems and planning problems. 2. Recommender Systems Recommender systems are becoming increasingly important for E-commerce websites. 3. In practice, a "good" model of the problem is as important as a fast algorithm, as OR people have concluded that a "good" model of a problem is crucial to it efficiently. 4. As there may be many different models for a problem, the questions then are, what is a "good" model and how can we obtain a "good" model? Our work on the conversion of non-binary CSPs to binary CSPs considers two general binary representations of a non-binary CSP, namely, the dual representation and the hidden-variable representation. 5. It has been recently recognized that both choosing the right solving algorithm and the right problem model are crucial for efficient problem solving.

B.7 MIT Artificial Intelligence Laboratory

`http://www.ai.mit.edu`

Part I. top 25 keywords
lab, research, system, page, information, model, people, project, paper, group, image, design, language, learning, problem, computer, course, robot, students, world, computation, vision, publication, need, number
Part II. top 5 key sentences
1. The STatistical AI Reading group (STAIR) meets weekly to host speakers and to read and discuss current and ongoing research in statistical methods in artificial intelligence and machine learning. 2. The MIT Laboratory for Computer Science (LCS) is an interdepartmental laboratory whose principal goal is research in computer science and engineering. 3. The MIT Museum is currently showing an ongoing exhibition entitled "Robots and Beyond: Exploring Artificial Intelligence at MIT" – an interactive excursion into the world of artificial intelligence taking us behind the scenes and into the research labs of MIT. 4. To find out more, visit our home page, peruse our research projects, look at our research abstracts or our publications, or browse the home pages of our people. 5. Artificial intelligence: basic research on learning, problem solving and programming.

B.8 Artificial Intelligence Applications Institute

`http://www.aiai.ed.ac.uk`

Part I. top 25 keywords
plan, activity, spar, process, aiai, model, constraint, project, planning, issues, system, detail, knowledge, representation, agent, entity, domain, page, activities, level, value, ontology, object, arpi, specification
Part II. top 5 key sentences
1. In recent years, for example, AIAI has been involved in projects to build systems to plan and schedule spacecraft assembly, to design the layout of telephone directory pages, to detect possible financial fraud and to help diabetics monitor their current physical state. 2. The key research areas of AIAI are: Planning and Activity Management: AIAI continues to be a world-leading center for planning and workflow; Knowledge Systems and Knowledge Modelling: the formal side of intelligent systems, concerned with models, ontologies and the methods for acquiring knowledge. 3. Applications of genetic algorithms include timetabling, scheduling, and models of complex systems to fit new data. 4. AIAI has extensive experience in providing formal support for knowledge and process modelling that enables rapid system prototyping and automatic verification that shortens system development cycle and enhances system quality. 5. AIAI can play a role in Intelligent System development projects during any of these phases, but most typically would be involved during the earlier stages of analysis, modelling and design, and proof-of-concept implementation where AIAI's experience of working with specifically AI concepts and technologies most telling.

B.9 Artificial Intelligence Center

`http://www.ai.uga.edu`

Part I. top 25 keywords
program, uga, system, option, code, library, paper, number, ftplib, function, microcontroller, version, information, computer, language, value, name, application, address, software, problem, gen, misc, internet, students
Part II. top 5 key sentences
1. Located in the same building as the Artificial Intelligence Center, the library has millions of books and bound periodicals and also provides online access to many indexes and journals. 2. Areas of specialization include automated reasoning, cognitive modeling, neural networks, genetic algorithms, expert databases, expert systems, knowledge representation, logic programming, and natural-language processing. 3. The AI Center currently offers a two-year interdisciplinary masters degree program in Artificial Intelligence and houses the undergraduate Cognitive Science program for UGA. 4. The interdisciplinary Artificial Intelligence Program was established in 1986. 5. Artificial intelligence is the computer modeling of intelligent behavior, including but not limited to modeling the human mind.

B.10 Waterloo Artificial Intelligence Group

`http://ai.uwaterloo.ca`

Part I. top 25 keywords
system, problem, information, algorithm, research, constraint, text, paper, document, uwaterloo, conference, topic, people, patient, sentence, group, model, kerr, task, result, question, beek, meeting, technique, language
Part II. top 5 key sentences
1. The HealthDoc project is developing natural language software systems for producing, on demand, health-information and patient-education brochures that are tailored to the medical condition and personal requirements of the individual patient. 2. Most of the papers employed machine learning (especially unsupervised or semi-unsupervised learning) for natural language learning tasks such as POS tagging, Name entity extraction, grammar induction, text categorization, and etc. 3. This approach combines techniques from natural language processing and knowledge representation with a penalty-based technique for relevance passage retrieval. 4. A preliminary version of the paper appears in Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, Montreal, Quebec, 541-547, August, 1995. 5. HealthDoc is funded by Technology Ontario and the Information Technology Research Centre, and is centred at the University of Waterloo, with partners at the University of Toronto and the University of Southern California.

B.11 Air Canada

`http://www.aircanada.ca`

Part I. top 25 keywords
service, flight, travel, airline, canada, custom, air, check, information, airport, baggage, fare, canadian, cargo, business, aircraft, carrier, passenger, aeroplan, cent, world, industry, america, cost, departure
Part II. top 5 key sentences
1. Welcome to the Air Canada Site Index! You can get to anywhere in our site from here, and it's easy to find the information you're looking for. 2. Air Canada is not liable for non-delivery of messages and cannot guarantee the timeliness or reliability of an e-mail sent to your pager network. 3. However, Air Canada will do everything practicable to ensure we provide the accurate flight status information at all times. 4. Air Canada Introduces New Media Centre This site is designed specifically to provide convenient, quick access to information about Air Canada. 5. I believe that Air Canada has taken a leadership position in terms of what we say and what we do for customers.

B.12 Cisco Systems, Inc.

`http://www.cisco.com`

Part I. top 25 keywords
cisco, solutions, service, product, search, networking, technologies, partner, events, network, content, contact, log, feedback, map, software, series, end, security, solution
Part II. top 5 key sentences
1. Cisco offers a portfolio of Technical Support Services to ensure that your Cisco products operate efficiently, remain highly available, and benefit from the most up-to-date system software. 2. Access information on technologies developed by the global data and telecommunications industry, defined by standards organizations, and used throughout Cisco solutions. 3. The Cisco cable solutions portfolio represents the industry's most comprehensive, innovative, and proven set of products, services, architectures, and professional services, optimized for evolving the communications-grade cable IP infrastructure. 4. Today, over 200 cable operators are powered by the Cisco cable solutions portfolio that delivers voice, video, and broadband data services to millions of subscribers around the globe. 5. Cisco delivers products and services focused on fault, configuration, accounting, performance, and security (FCAPS) systems to ensure the planning, provisioning, monitoring, and assurance of diverse network applications.

B.13 Microsoft Corporation

`http://www.microsoft.com`

Part I. top 25 keywords
microsoft, server, office, windows, support, product, information, service, business, search, exchange, page, system, security, internet, outlook, contact, solution, software, version, program, access, guide, tool, computer
Part II. top 5 key sentences
1. Microsoft Office Professional Edition 2003 is designed to improve how companies can further take advantage of XML. 2. Microsoft Office InfoPath(TM) 2003, also available in Office Professional 2003, also takes advantage of XML. 3. After using OneNote to gather information, you can organize it into an outline or easily reuse it in Microsoft Office Word 2003 documents or Microsoft Office PowerPoint 2003 presentations. 4. Windows 2000 SP4 is a recommended update that includes the updates contained in previous Windows 2000 service packs. 5. Windows 2000 Service Pack 4 (SP4) provides the latest updates to the Windows 2000 operating systems.

B.14 Nortel Networks

`http://www.nortelnetworks.com`

Part I. top 25 keywords
service, product, solutions, network, support, training, events, certification, contact, map, communities, region, country, career, enterprise, alteon, portfolio, switch, customer, partner, program, information, eacute, advance, business
Part II. top 5 key sentences
1. Armed with an unrivaled suite of IP processing capabilities, Nortel Networks Shasta 5000 Broadband Service Node (BSN) makes it possible to deliver a range of value-added, revenue-generating services on top of a multitude of subscriber aggregation technologies. 2. Nortel Networks industry-leading products are transforming the current maze of multiple networks into one, easy to manage packet-based network where data, voice and video speed to their destinations quickly and more efficiently. 3. Nortel Networks Premium Partners Participants with annual forecasted purchases of Nortel Networks products greater than $1 million* are eligible to register as a Premium Partner-the highest level within the plan. 4. Nortel Networks Resellers Resellers with forecasted annual purchases of Nortel Networks products between $250,000 and $1 million are eligible to participate as Nortel Networks Resellers. 5. Nortel Networks Security Solutions are designed with the total network in mind - for resilience, optimum performance, ease of management and scalability.

B.15 Oracle Corporation

`http://www.oracle.com`

Part I. top 25 keywords
oracle, information, customer, service, product, application, support, system, cost, database, management, arrow, business, technology, feature, page, process, browser
Part II. top 5 key sentences
1. In order to provide service in a timely, cost effective and efficient manner, information provided to Oracle may be made available to authorized Oracle users across our global offices that require access to the information for business purposes. 2. Oracle stores this information for the sole purpose of sending this one-time email and tracking the success of our referral program. 3. Oracle only shares information with selected companies after providing users with an ability to either opt-out of or opt-in to the sharing. 4. In this area, you will find information about your use of the Oracle Web site, using Oracle trademarks and logos, notifying Oracle of possible copyright infringement, reporting piracy of Oracle products, buying Oracle products on auction sites, Oracle's statement on privacy, and other issues. 5. It's a forum that enables existing customers and technology professionals to meet with Oracle executives, developers, and other key customers to get first-hand information on Oracle's product strategy and technology updates learning how to get the most out of their existing products.

B.16 Adhesion Technologies

`http://www.adhesiontech.com`

Part I. top 25 keywords
adhesion, solution, ea2, service, custom, technologies, aggregation, management, advisor, clients, spacer, market, software, client, technology, value, account, product, bbh, consolidation, partner, overview, bank, wealth, platform
Part II. top 5 key sentences
1. Adhesion has made its EA2 solution fully compatible with IBM WebSphere* and has been working closely with both the IBM financial services solutions sales team on joint opportunities in the U.S. and U.K. 2. Adhesion has entered into an agreement with IBM Corporation to jointly market Adhesion's EA2 (Enhanced Account Aggregation) solution to IBM's new and existing financial services clients in the USA, and the UK. 3. Adhesion's EA2 Financial Consolidation Platform, ranked #1 in wealth management data consolidation by Celent Communications and a leader in aggregation for wealth management by Forrester Research, enables financial institutions and services providers (FSP's) to create value-added and differentiated wealth management services for their customers. 4. Adhesion's offices are located in Charlotte's South Park district, an area of strong growth and interest in technology and financial services. 5. About Adhesion Technologies Adhesion is a leading provider of software solutions for financial account consolidation, both for internal accounts and away.

B.17 Application Solutions & Technologies, Inc.

http://www.asti-global.com

Part I. top 25 keywords
business, system, service, development, quality, customer, design, solution, tools, software, multimedia, need, application, expertise, outsourcing, experience, process, consulting, huifeng, products, shanghai, video, requirement, productivity, years
Part II. top 5 key sentences
1. As a Software Services Provider, ASTI is constantly exposed to the challenges businesses face to design, develop, and deploy e-business application systems. 2. Our expertise also includes integrating existing systems with newly developed systems, migrating legacy systems wherever necessary. 3. ASTI's expertise and experience in software tool development enables ASTI to adopt the prototyping approach to designing and developing a software solution for conducting services. 4. We will work closely with our customer to provide the necessary services, ranging from requirements generation, spec analysis, design and development, to delivering and deploying a solution that will, not only provide an effective e-business solution, but can also improve our customer's business operations. 5. System Integration ASTI is committed to delivering the technologies and services required to design, develop, and deploy software solutions that will help each business, their employees, customers and partners collaborate and this fast-paced, net-centric environment.

B.18 Commerce One

http://www.commerceone.com

Part I. top 25 keywords
class, frames, commerceone, xdk, package, overview, index, tree, commerce, method, void, document, qname, cast, type, soap, parameter, value, tostring, constr, boolean, message, object, return, interface
Part II. top 5 key sentences
1. The Commerce One Conductor platform is an infrastructure for the creation, deployment, integration and management of Composite Process Management. 2. It's this deep and proven experience in creating flexible business connections and leveraging standards that makes Commerce One uniquely qualified to deliver the industry's first standards-based Composite Process Management Platform: Commerce One Conductor. 3. Commerce One is committed to the development of standards-based solutions that deliver value to business users. 4. Commerce One Source and Commerce One Procure allow companies to achieve visibility and control across the entire source-to-pay process. 5. In addition to Conductor training, Commerce One University also provides complete education for solutions in the Commerce One Source-to-Pay applications suite.

B.19 Gamma Enterprise Technologies, Inc.

http://www.getgamma.com

Part I. top 25 keywords
infoshuttle, system, gamma, object, production, application, development, business, training, environment, technologies, transfer, test, client, ale, product, move, shuttle, need, type, testing, configuration, getgamma, support, master
Part II. top 5 key sentences
1. Gammas InfoShuttle is the only tool available that can transfer subsets of Production data to your Training environments without the need for ALE expertise. 2. Currently, InfoShuttles development team is designing the platform architecture to support the ability to transfer datasets among the various mySAP.com solutions and components, such as mySAP Customer Relationship Management (mySAP CRM), SAP Advanced Planner and Optimizer (SAP APO), and the SAP Business Information Warehouse (SAP BW). 3. InfoShuttle 2.1 leverages its Business Connector Adapter to move standard IDOCs out of SAP in industry-standard XML, giving you the capability to send business data easily from your SAP R/3 system to your partners in an XML format. 4. By incorporating SAP Business Connector functionality, InfoShuttle gives users an easy-to-use extraction tool for representing SAP data in an XML format. 5. InfoShuttle transfers only a subset of data objects (such as Material or Customer) as specified by the user.

B.20 RDM Corporation - Electronic & Paper Payment Solutions.

http://www.rdmcorp.com

Part I. top 25 keywords
rdm, micr, solution, rdmcorp, product, support, paper, company, image, privacy, document, investors, check, print, control, payment, processing, information, process, sale, service
Part II. top 5 key sentences
1. RDM's Bill Payment solution for PC based and terminal based applications provide a great degree of flexibility and compatibility for the end user. 2. The robust functionality of Synergy positions it as the leading stand-alone image payment solution in the industry today. 3. As well as magnetically reading the MICR line, OCR MICR assist will optically read the MICR line improving EC6000i's read rate to virtually 100%. 4. RDM Corporation, the RDM Corporation logo, and most other brand names in this Website are registered trademarks of RDM Corporation. 5. RDM has a full line of products that address all MICR and image quality control issues, including MICR line signal level, line intrusions, character dimensions, as well as image issues such as legibility and print contrast signal.

C KTB_1 Summaries of the 20 Test Web Sites

C.1 Department for CASE tools of the Institute for System Programming

`http://www.ispras.ru/groups/case/case.html`

Part I. top 25 keyterms
description language, case tool, message sequence chart, software development, description language model, description language specification, software architecture, message sequence, sequence chart, development process, test case, software engineering, telelogic tau, system programming, tool support, engineering tool, nn www, language specification, validation model, telecommunications software, klocwork message sequence chart, software system, data description language abstract syntax notation, development methodology, moscow state university
Part II. top 5 key sentences
1. Department for CASE tools of the Institute for System Programming (short for Computer-Aided Software Engineering tools) performs research and development in the area of modern tool support for software development. 2. Since 1994 the Department for CASE tools performed several joint research and development projects, subcontracted to the Institute for System Programming by Nortel Networks (Canada) and Telelogic AB (Sweden). 3. Department of CASE tools performs world-leading research in building next-generation model-based tools that accelerate software development through the use of formal modeling techniques, automated transformations, including advanced code generation techniques, managed software architectures, validation and verification, especially at the early phases of the development process. 4. Current members of research staff of the Department for CASE tools actively participate in international standardization on formal software development techniques in OMG and ITU-T. 5. Department for CASE tools develops methodologies and tools in the following two main directions: * Tool support for Managed Software Architectures (inSight project).

C.2 International Function Point Users Group

http://www.ifpug.org

Part I. top 25 keyterms
target audience, function point, workshop length, prerequisites participant, workshop title, workshop number, process improvement, software measurement, software project, software development, workshop description, company software maturity, cmm level, metrics plan, quality standard, business relevance, capability maturity, capability maturity model, software process, maturity model, software workshop length, none workshop number, measurement program, days workshop description, development life
Part II. top 5 key sentences
1. The International Function Point Users' Group (IFPUG) is a non-profit, member governed organization. 2. The mission of IFPUG is to be a recognized leader in promoting and encouraging the effective management of application software development and maintenance activities through the use of Function Point Analysis and other software measurement techniques. 3. In support of this, IFPUG maintains the Function Point Counting Practices Manual, the recognized industry standard for FPA. 4. Also, through industry and academic relationships, IFPUG sponsors and supports projects for applied research on software measurement issues, and conducts studies in support of advancing the Function Point Counting Standards. 5. IFPUG is the steward of the Function Point standard, the most widely recognized method used today to size software.

C.3 Software Build and Fix

`http://www.mapfree.com/sbf`

Part I. top 25 keyterms
command line, root window, event handler, window manager, security policy, command window, backslash substitution, square bracket, programming language, dollar sign, widget name, action family, tcl script, book link, caps example, tcl interpreter, policy name, entry widget, configuration file, usage documentation, markup language, object action, source code, input file, glob pattern
Part II. top 5 key sentences
1. Traditional programming languages have two principle ways of handling input. 2. The programmer makes use of the modularity features in the underlying programming language (or simply makes use of comments and conventions) to enforce a rule that only specifically approved procedures access regional variables in any way at all. 3. Controlled variable documention is consistent with all programming languages: object-oriented, scripting, assembler, whatever. 4. In another well-known example, today's programming languages often lack the go-to statement that caused problems for an earlier generation of programmers. 5. This research has been aimed at formal specification of programming languages and proofs of correctness for small programs.

C.4 Software Engineering Archives

`http://www.cs.queensu.ca/Software-Engineering`

Part I. top 25 keyterms
base type, rw type, procedure name, equivalent xto, pascal type string, extraction error fi fi echo, base type name, microprocessor assembly language, c definition, reserve h, assembler constraint, integer type, bit mask, return value, arguments pointer, term constructor, array mapping, type xof, ascii format, applications xfor, instance name, naming conventions, doug klunder, strings constant, hungaria n xwill
Part II. top 5 key sentences
1. All variable names are composed of three elements: prefixes, base type, and qualifier. 2. Not all elements are present in all variable names; the only part that is always present is the base type. 3. Many people prefer to consider it a true indivisible base type. 4. The term constructor is used because a new type is constructed from the combination of the operation and the base type. 5. These operations are represented in Hungarian by prefixes; the combination of the prefixes and base type represent the complete type of an entity.

C.5 Software Engineering Institute

http://www.sei.cmu.edu

Part I. top 25 keyterms
engineering institute, software engineering, software engineering institute, product line, carnegie mellon, development center, software architecture, software development, software product, software product line, software process, system component, process improvement, design decision, coordination pattern, reference architecture, software system, coordination protocol, infrastructure capability, application developer, whiteboard course attendees stryker infantry carrier vehicle, system architecture, capability maturity, target system, risk management
Part II. top 5 key sentences
1. The Software Engineering Institute (SEI) is a federally funded research and development center sponsored by the U.S. Department of Defense and operated by Carnegie Mellon University. 2. 2002 SEI Annual Report Published The online version of the Annual Report of the Software Engineering Institute (SEI), reporting on fiscal year 2002, is available at http://www.sei.cmu.edu/annual-report/. 3. The Software Engineering Institute (SEI) helps organizations and individuals to improve their software engineering management practices. 4. The Software Engineering Institute (SEI) sponsors, co-sponsors, and is otherwise involved in many events throughout the year. 5. The Software Engineering Institute offers a number of courses and training opportunities.

C.6 Artificial Intelligence Research Group

http://www.cs.ualberta.ca/∼ai

Part I. top 25 keyterms
computer games, computer science, petri net, computer vision, electrical engineering, image processing, constraint satisfaction, probability distribution, pattern recognition, expert system, quantum property, search space, disease mutation, logic programming, vision system, building ai system, soar architecture, air mission, constraint satisfaction problem, monte carlo, backtracking algorithm, bayesian network, belief net, atp system, computer vision system
Part II. top 5 key sentences
1. We also present analysis methods facilitating verification and refinement of Petri Net models. 2. Two machine learning algorithms are developed to synthesize Petri Net models automatically or semi-automatically. 3. This interdisciplinary research has made the following five main contributions: A Petri Nets based approach to decision-making scheduling through environment modeling is presented. 4. Our work in developing characters for computer games using the Soar architecture tries to push even further to human-like behavior. 5. The research and development behind the system are built upon advanced methods from the areas of Artificial Intelligence, Machine Learning and Data Mining, Computer Vision, Remote Sensing, and Silviculture.

C.7 MIT Artificial Intelligence Laboratory

`http://www.ai.mit.edu`

Part I. top 25 keyterms
ai lab, computer science, graduate student, computer vision, attraction model, language text, research proposal, girl scout, machine translation, vision system, intelligence laboratory, mit ai, lexical conceptual structure, experiment kirby, project proposal, ai research, floor playroom, ai researcher, research assignment, area exam, chris crick language learning, vision stub project proposal, edward faulkner hiearchical classification, anthony kim bootstrapping syntax, jackendoff lexical conceptual structure
Part II. top 5 key sentences
1. The AI Lab Fiscal Office is here to help you navigate through all the various MIT administrative forms and procedures. 2. The AI Lab is home to many graduate students. 3. Want to know more about the research areas of the AI Lab Faculty? Check here to see results of the recent AI Lab Faculty Survey and for a short synopsis of Faculty Research Interests along with links to their individual web pages. 4. If you are an MIT undergraduate and are interested in some of the other research projects at the AI Lab, please contact the UROP Office to search their listed positions. 5. The great strength of the AI Lab has always been a willingness to put together large scale systems in ways that others have either not dared or for which they have not been able to marshal the required resources.

C.8 Artificial Intelligence Applications Institute

`http://www.aiai.ed.ac.uk`

Part I. top 25 keyterms
time unit, plan representation, spar model, activity representation, core group, time point, plan case, plan ontology, object model, agent relationship, workflow management, entity annotation, ai planning, description language, world state, workflow management coalition, enterprise ontology, nist psl, air campaign, research community, partial shared view, austin tate, core plan representation, spar structure, shared view
Part II. top 5 key sentences
1. The initial application of the CPR is in addressing plan interchange requirements of several military planning systems, but the model goes beyond military planning and presents a more general plan representation. 2. AIAI is working with Cycorp to develop process and plan representations. 3. Participants have been chosen to be familiar with the issues of plan representation and to be able to reflect specific concerns for their "constituency". 4. The working group members represent some of the most experienced people worldwide who have been involved in creating shared plan, process and activity representations or standards for some years. 5. Workflow Management Systems (WfMS) operate in dynamic environments, where they are expected to ensure that users are supported in performing flexible and creative tasks while maintaining organisational norms.

C.9 Artificial Intelligence Center

http://www.ai.uga.edu

Part I. top 25 keyterms
default value, sparse ii, source code, program option, cost function, command line, selection function, fitness function, fitness value, genetic algorithm, selection method, baud rate, mating pool, selection function name, copyright notice, object code, page address, grammar rule, intelligence center, survival selection method, blank line, genome individual, survivor function, files content, options file
Part II. top 5 key sentences
1. Strengths include logic programming, expert systems, neural nets, genetic algorithms, and natural language processing. 2. Areas of specialization include automated reasoning, cognitive modeling, neural networks, genetic algorithms, expert databases, expert systems, knowledge representation, logic programming, and natural-language processing. 3. The Artificial Intelligence Center is an interdepartmental research and instructional center within the Franklin College of Arts and Sciences of the University of Georgia. 4. The Artificial Intelligence Center also houses the undergraduate degree program in Cognitive Science. 5. It was officially established as the Artificial Intelligence Center in 1995.

C.10 Waterloo Artificial Intelligence Group

http://ai.uwaterloo.ca

Part I. top 25 keyterms
constraint programming, backtracking algorithm, co op student, steve kerr, software developer, constraint satisfaction, bell lab, constraint network, constraint satisfaction problem, sentence planning, healthdoc project, approximation algorithm, information extraction, master document, data mining, vlado keselj, darlington nuclear power station, language processing, satisfaction problem, interval algebra, electrical engineering section, microsoft windows application, text retrieval component, software development conflict, emergency communication system
Part II. top 5 key sentences
1. Constraint programming is a methodology for solving difficult combinatorial problems. 2. Constraint programming has shown that a general purpose search algorithm based on constraint propagation combined with an emphasis on modeling can solve large, practical scheduling problems. 3. Our current results show that constraint programming does indeed work well and has the advantage in terms of time and space efficiency over the current state-of-the-art planners. 4. The HealthDoc project is developing natural language software systems for producing, on demand, health-information and patient-education brochures that are tailored to the medical condition and personal requirements of the individual patient. 5. There will be an overview of the concept and utility of information extraction, what the IE process entails, what sort of technologies are involved and how IE can be applied to industry.

C.11 Air Canada

`http://www.aircanada.ca`

Part I. top 25 keyterms
air canada, star alliance, north america, airline industry, airline alliance, open sky, route network, law school, business traveler, air transportation, seniority integration, north american, market share, market opportunity, revenue synergy, load factor, business travel, air transport, close eye, airline business, labour cost, yield management, work rule, route map, air canada title slide
Part II. top 5 key sentences
1. After the combination of Canadian Airlines with Air Canada, some customers became concerned about our ability to provide services in both official languages. 2. Welcome to the Air Canada Site Index! You can get to anywhere in our site from here, and it's easy to find the information you're looking for. 3. Our interactive route map displays destinations served and routes from a departure city, or for a particular carrier from our Air Canada family. 4. Air Canada Introduces New Media Centre This site is designed specifically to provide convenient, quick access to information about Air Canada. 5. Welcome to the world of Air Canada - a world of diversity and opportunity.

C.12 Cisco Systems, Inc.

`http://www.cisco.com`

Part I. top 25 keyterms
customer satisfaction, service provider, cisco sales expert exam, cisco sales expert, cisco product, sales expert, cisco sale, software maintenance, cisco ios, cisco sales expert badge, sales expert role, cisco ip, service contract, replacement part, account manager, customer satisfaction objective, cisco channel partner, ip telephony, silver partner, cisco ios software, system software, sales professional, cisco channel, software maintenance support, cisco sales essentials course
Part II. top 5 key sentences
1. Cisco solutions enable service providers to deliver metro bandwidth and to deploy new revenue-generating services for their customers. 2. We offer best-of-breed routing platforms for service provider core and edge, enterprise, and small and medium-sized businesses. 3. With Cisco, service providers can create new, high-margin services that will help them attract and retain customers and deliver a sustainable competitive advantage. 4. Cisco enjoys a unique position as the market share leader for Internet-related networking equipment, and delivers networking solutions to 90% of the Top 25 service providers. 5. Many of Cisco's Content Services benefit both service providers and enterprises alike.

C.13 Microsoft Corporation

http://www.microsoft.com

Part I. top 25 keyterms
office xp, microsoft office, windows xp, service pack, microsoft windows, windows server, sql server, microsoft office system, distribution folder, exchange server, commerce server, office system, microsoft product, contact us, page contact us, statement accessibility, business intelligence, page contact us free newsletter, page contact, windows installer, contact us free, biztalk server, management server, application center, contact us free newsletter
Part II. top 5 key sentences
1. Office XP Professional integrates productivity innovations throughout its programs to transform the traditional Office suite into a smarter overall work experience. 2. The Office Update service covers all Microsoft Office products, including Microsoft Word, Microsoft PowerPoint, Microsoft Excel, and Microsoft Outlook. 3. Microsoft Office Professional Edition 2003 is designed to improve how companies can further take advantage of XML. 4. For added protection, Windows XP users should set Automatic Updates to download and install important updates. 5. Windows XP Professional Enjoy great digital media tools for interactive learning and more.

C.14 Nortel Networks

http://www.nortelnetworks.com

Part I. top 25 keyterms
nortel networks, service provider, contact us, nortel networks product, enterprise customer, data network, press release, networks product, media center, nortel networks advanced, developer product, market condition, multimedia service, nortel networks passport, alteon ssl, nortel networks product including without limitation, series nortel networks passport, industry leader, currency exchange rate, solutions overview, nortel networks further shall have, multiservice switches, communications technology, customer financing, nature whatsoever
Part II. top 5 key sentences
1. Nortel Networks has emerged as the leading company that is delivering value to customers around the world through Unified Networks solutions, spanning mission-critical telephony and IP-optimized solutions. 2. Nortel Networks works closely with customers in more than 150 countries and territories around the world to help speed their success. 3. Nortel Networks sells its Enterprise and Small/Medium business solutions through a worldwide network of channel partners, who can deliver everything from smaller voice systems to global integrated solutions for the international business. 4. Nortel Networks industry-leading products are transforming the current maze of multiple networks into one, easy to manage packet-based network where data, voice and video speed to their destinations quickly and more efficiently. 5. Nortel Networks does business in more than 150 countries.

C.15 Oracle Corporation

`http://www.oracle.com`

Part I. top 25 keyterms
e business suite, oracle application, oracle e business, oracle e business suite, application server, oracle applications user group, oracle product, sql server2000, oracle applications user, supply chain, user group, applications user group, applications user, business intelligence, real application clusters, oracle workflow, real application, business processe, sql server, collaboration suite, oracle process manufacturing, process manufacturing, oracle project, oracle customer, oracle process
Part II. top 5 key sentences
1. Oracle9i Application Server has consistently ranked first in the ECperf and SPECj industry-standard 'Performance' and 'Price/Performance' benchmarks. 2. The Fastest Application Server Is Also the Least Expensive to Run That's the finding of the J2EE community's new Enterprise JavaBeans benchmark run by the industry organization ECperf. 3. Oracle9i Application Server turned in 63 percent higher performance than BEA WebLogic, 39 percent higher performance than IBM WebSphere, and price/performance that shows those competitors to be more than twice the cost. 4. This website is an ever-changing collection of information on Oracle products and technology, and we've either removed or relocated the document you requested. 5. In this area, you will find information about your use of the Oracle Web site, using Oracle trademarks and logos, notifying Oracle of possible copyright infringement, reporting piracy of Oracle products, buying Oracle products on auction sites, Oracle's statement on privacy, and other issues.

C.16 Adhesion Technologies

`http://www.adhesiontech.com`

Part I. top 25 keyterms
wealth management, adhesion ea2, account aggregation, services provider, software solution, ea2 financial consolidation, adhesion technologies, ea2 financial, financial consolidation, asset management client, wealth management service, brown brothers, adhesion ea2 financial consolidation, aggregation solution, data collection, adhesion ea2 financial, account data, adhesion technologies adhesion, aggregation service, adhesion enhanced account, management service, service provider, enhanced account, account aggregation solution, aggregation software
Part II. top 5 key sentences
1. Finaplex is the leading provider of enterprise software solutions for the wealth management industry. 2. The solution is well-suited to enable financial institutions to deliver new, value-added wealth management services due to Adhesion's focus on post-aggregation data normalization and enrichment. 3. "We have stayed very focused on our product objectives of delivering high quality financial data consolidation for the powering of wealth management services," said Michael Stier, Chief Operating Officer at Adhesion. 4. Adhesion's EA2 Financial Consolidation Platform enables financial institutions and services providers (FSP's) to create value-added and differentiated wealth management services for their customers. 5. At clients' request, Adhesion can now seamlessly integrate RiskMetrics' WealthBench(TM) portfolio analysis, optimization, multi-goal planning and monitoring solution into the EA2 Financial Consolidation Platform.

C.17 Application Solutions & Technologies, Inc.

http://www.asti-global.com

Part I. top 25 keyterms
software solution, software tool, software service, tool development expertise, consulting service, legacy system, software development, tool development, medical bureau, application system, e business solution, application development tool, kiosk system, business operation, system integration, capability maturity model, quality software service, software engineering, business problem, data warehousing, email asti software development asti expertise, shanghai health, software product, maximize software development productivity, targets capability maturity model
Part II. top 5 key sentences
1. (ASTI) is an established software solution provider to clients, ranging from the federal government to commercial and private industries. 2. We recognize the need to help our customers meet the challenges and demands of implementing software solutions and can leverage our customer's investment by recognizing the need and benefit of developing software tools that will increase productivity and maintain the quality of the software solution. 3. ASTI's expertise and experience in software tool development enables ASTI to adopt the prototyping approach to designing and developing a software solution for conducting services. 4. System Integration ASTI is committed to delivering the technologies and services required to design, develop, and deploy software solutions that will help each business, their employees, customers and partners collaborate and succeed in this fast-paced, net-centric environment. 5. As a Software Services Provider, ASTI is constantly exposed to the challenges businesses face to design, develop, and deploy e-business application systems.

C.18 Commerce One

http://www.commerceone.com

Part I. top 25 keyterms
content event, attribute event, business process, process management, commerce one, press release, method summary, white paper, industry standard, time zone, system identifier, cpg company, networked srm white paper, supply chain, licensed product, conductor technical brochure, conductor white paper, interoperability engine, xml document, related links, solution brochure, downloads commerce, commerce one filing, namespace declaration, risk factor
Part II. top 5 key sentences
1. Commerce One's Conductor platform is one of the first complete solutions we have seen that offers a standards-based approach to enable extended business processes and application capabilities across such a complex environment. 2. Commerce One has been helping the world's leading companies - Siemens, Boeing, Daimler Chrysler, General Motors, and others - improve their bottom line and business processes. 3. Through integration of these business processes, Commerce One customers advance the business value of their systems, innovate and automate solutions, and increase the return on their investment in their technology infrastructure. 4. The Commerce One Conductor platform and industry-specific Process Accelerators represent the leading Composite Process Management solution that enable enterprises to combine and connect features, functionality, data and/or business context from existing applications to create new business functionality. 5. Commerce One intends to use the net proceeds from the financing for working capital purposes and to further its efforts in support of its Composite Process Management strategy and the recent launch of the Commerce One Conductor(TM) platform.

C.19 Gamma Enterprise Technologies, Inc.

http://www.getgamma.com

Part I. top 25 keyterms
infoshuttle move, data object, user exit, sap client, sap system, shuttle data, 3 system, development environment, sap version, rfc destination, ale configuration, 3 source, nn client, object type, test system, sales order, sap client copy, infoshuttle user, client copy, source system, user interface, data transfer, data type, infoshuttle checks sap authorization profile, sap business connector functionality
Part II. top 5 key sentences
1. Relational Object Repository InfoShuttle moves discrete, selected master data objects as well as stand-alone transactional data objects. 2. Does InfoShuttle move all types of master data? The current version of InfoShuttle moves many of the most common types of master data. 3. Does InfoShuttle move related objects? InfoShuttle has the ability to gather the multiple levels of related objects which are required for its posting. 4. Does InfoShuttle move transactional data between R/3 systems? The current version of InfoShuttle moves a variety of transactional data types between R/3 systems and from R/3 to external systems. 5. InfoShuttle easily transfers just the data objects that are required to support current events in those environments, leaving intact the unique configurations and other specially created data.

C.20 RDM Corporation - Electronic & Paper Payment Solutions.

http://www.rdmcorp.com

Part I. top 25 keyterms
quality control, payment archive, micr line, rdm corporation, image management, image quality control, image quality, payment archive service, print quality, micr qualifier gtx, rdm image, archive service, bill payment, micr qualifier, rdm micr, rdm image management, image qualifier, micr verifier mh, industry leader, image management solution, qualifier gtx, management solution, press releases, micr verifier, transaction management system
Part II. top 5 key sentences
1. RDM has a full line of products that address all MICR and image quality control issues, including MICR line signal level, line intrusions, character dimensions, as well as image issues such as legibility and print contrast signal. 2. We specialize in enabling and supporting the electronic conversion of checks and bill payments, providing complete image management, MICR/OCR and image quality control. 3. Our technologies such as the EC5000 imager and the Payment Archive Service reduce the cost burden associated with the handling, processing and storage of small paper documents. 4. RDM's new ITMS solution which includes RDM's proven Payment Archive Service (PAS) provides the tools and technology necessary for the successful launch or continued success of any electronic check conversion program. 5. In any application or environment, RDM's PAS image management solution can improve transaction processing efficiency, decrease dispute resolution times, and virtually eliminate paper document handling and storage; without the worry of system integration or a large investment of funds and resources.

D KTB_2 Summaries of the 20 Test Web Sites

D.1 Department for CASE tools of the Institute for System Programming

`http://www.ispras.ru/groups/case/case.html`

Part I. top 25 keyterms
description language, message sequence chart, case tool, software development, message sequence, sequence chart, description language specification, description language model, software architecture, formal method, development process, tabular combined notation, telelogic tau, formal model, early phase, test case, system programming, formal language, nn www, accelerated development methodology, language specification, software engineering, scenario description language called message sequence chart, language model, data description language abstract syntax notation
Part II. top 5 key sentences
1. Since 1994 the Department for CASE tools performed several joint research and development projects, subcontracted to the Institute for System Programming by Nortel Networks (Canada) and Telelogic AB (Sweden). 2. Department of CASE tools performs world-leading research in building next-generation model-based tools that accelerate software development through the use of formal modeling techniques, automated transformations, including advanced code generation techniques, managed software architectures, validation and verification, especially at the early phases of the development process. 3. Current members of research staff of the Department for CASE tools actively participate in international standardization on formal software development techniques in OMG and ITU-T. 4. Department for CASE tools of the Institute for System Programming (short for Computer-Aided Software Engineering tools) performs research and development in the area of modern tool support for software development. 5. Department for CASE tools develops methodologies and tools in the following two main directions: * Tool support for Managed Software Architectures (inSight project).

D.2 International Function Point Users Group

http://www.ifpug.org

Part I. top 25 keyterms
function point, target audience, capability maturity model integrated, project manager, prerequisites participant, capability maturity model, workshop length, workshop title, software measurement, project leader, key process area, senior analyst, workshop number, process improvement, software development, software benchmarking standards group, software project, benchmarking standards group, software benchmarking standards, managing software project, certified function point, quality assurance manager, standards group, capability maturity model level, benchmarking standards
Part II. top 5 key sentences
1. The International Function Point Users' Group (IFPUG) is a non-profit, member governed organization. 2. The mission of IFPUG is to be a recognized leader in promoting and encouraging the effective management of application software development and maintenance activities through the use of Function Point Analysis and other software measurement techniques. 3. In support of this, IFPUG maintains the Function Point Counting Practices Manual, the recognized industry standard for FPA. 4. Also, through industry and academic relationships, IFPUG sponsors and supports projects for applied research on software measurement issues, and conducts studies in support of advancing the Function Point Counting Standards. 5. IFPUG is the steward of the Function Point standard, the most widely recognized method used today to size software.

D.3 Software Build and Fix

`http://www.mapfree.com/sbf`

Part I. top 25 keyterms
curly bracket, root window, command line, regular expression, backslash substitution, variable substitution, variable name, event handler, command window, entry widget, window manager, security policy, command substitution, tcl interpreter, attached cavity, square bracket, programming language, configuration file, given string, empty string, special character, geometry manager, naming convention, tcl script, vertical slice
Part II. top 5 key sentences
1. These are regular expressions, GUI programming, TCP/IP client/server programming, and security policies for the Tcl browser. 2. In spite of the software engineering community's experience, there is no question that short variable names are valuable in mathematics. 3. The example proof follows mathematical practice and uses short variable names. 4. Traditional programming languages have two principle ways of handling input. 5. The programmer makes use of the modularity features in the underlying programming language (or simply makes use of comments a nd conventions) to enforce a rule that only specifically approved procedures access regional variables in any way at all.

D.4 Software Engineering Archives

`http://www.cs.queensu.ca/Software-Engineering`

Part I. top 25 keyterms
base type, integrated emergency management information system, variable name, procedure name, pascal type string, null terminated string, operating system, naming conventions, far pointer, data type, near pointer, character array, standard qualifiers, character run, single bit, generic subroutine, transfer operation, generic term, common loop, ascii format, common type, true function, actual first element, actual character, command dispatch
Part II. top 5 key sentences
1. Both methods (full type and base type) are accepted. 2. These operations are represented in Hungarian by prefixes; the combination of the prefixes and base type represent the complete type of an entity. 3. All variable names are composed of three elements: prefixes, base type, and qualifier. 4. Not all elements are present in all variable names; the only part that is always present is the base type. 5. Many people prefer to consider it a true indivisible base type.

D.5 Software Engineering Institute

http://www.sei.cmu.edu

Part I. top 25 keyterms
engineering institute, software engineering institute, software engineering, product line, software architecture, carnegie mellon university, capability maturity, capability maturity model, carnegie mellon, maturity model, software process, mellon university, process improvement, development center, system component, software development, software system, reference architecture, personal software process, software product line, capability maturity model integration, target system, design decision, software product, team software process
Part II. top 5 key sentences
1. The Software Engineering Institute (SEI) is a federally funded research and development center sponsored by the U.S. Department of Defense and operated by Carnegie Mellon University. 2. The Software Engineering Institute (SEI) sponsors, co-sponsors, and is otherwise involved in many events throughout the year. 3. The Software Engineering Institute offers a number of courses and training opportunities. 4. The Software Engineering Institute (SEI) helps organizations and individuals to improve their software engineering management practices. 5. The SEI provides the technical leadership to advance the practice of software engineering so the DoD can acquire and sustain its software-intensive systems with predictable and improved cost, schedule, and quality.

D.6 Artificial Intelligence Research Group

http://www.cs.ualberta.ca/∼ai

Part I. top 25 keyterms
artificial intelligence, computer games, knowledge based engineering, automated theorem proving, petri net, problem solving, search space, learning algorithm, quantum property, bayesian network, constraint satisfaction, computer science, pattern recognition, theorem proving, automated theorem proving system, automated theorem, electrical engineering, machine learning, australian pattern recognition society, execution step, surface mount technology, belief net, neural network, vision system, time interval petri net
Part II. top 5 key sentences
1. These agent systems combine techniques from database and distributed systems research and artificial intelligence techniques of learning and planning. 2. An intelligent mining system is simply applied research that promotes the use of advanced artificial intelligence, machine vision and robotics to the problems found in mining oil sand. 3. Artificial Intelligence Research at UofAlberta Abstract: For this kick-off "AI Seminar", I will first present a very short overview of the current research in UofA's "AI Lab", then various AI professors will briefly discuss their specific research agenda. 4. The research and development behind the system are built upon advanced methods from the areas of Artificial Intelligence, Machine Learning and Data Mining, Computer Vision, Remote Sensing, and Silviculture. 5. Our work in developing characters for computer games using the Soar architecture tries to push even further to human-like behavior.

D.7 MIT Artificial Intelligence Laboratory

`http://www.ai.mit.edu`

Part I. top 25 keyterms
artificial intelligence, computer science, lexical conceptual semantic, ai lab, human mind, visual routine, floor playroom, object recognition, artificial intelligence laboratory, artificial intelligence lab, computer vision, intermediate complexity, machine learning, intelligence laboratory, physical world, intelligence lab, vision system, computer vision system, lexical attraction, intermediate complexity feature, lexical attraction model, natural language text, semantic domain, thread memory, related term
Part II. top 5 key sentences
1. Artificial intelligence: basic research on learning, problem solving and programming. 2. The MIT Laboratory for Computer Science (LCS) is an interdepartmental laboratory whose principal goal is research in computer science and engineering. 3. Gates Building housing the Laboratory for Computer Science, the Alexander Dreyfoos Building housing the Artificial Intelligence Laboratory, the Laboratory for Information Decision Systems, and the Department of Linguistics and Philosophy, a below grade service facility and 2 levels of below grade parking. 4. Over the last year we have joined forces with the Laboratory for Computer Science in MIT Project Oxygen. 5. The great strength of the AI Lab has always been a willingness to put together large scale systems in ways that others have either not dared or for which they have not been able to marshal the required resources.

D.8 Artificial Intelligence Applications Institute

`http://www.aiai.ed.ac.uk`

Part I. top 25 keyterms
artificial intelligence, process interchange format, interchange format, adanan liberator, plan representation, intelligence applications institute, artificial intelligence applications institute, process interchange, applications institute, intelligence application, core plan representation, core plan, software engineering, minimum requirement, defense advanced research projects agency, artificial intelligence application, process specification language, shared view, working group, advanced research projects agency, spar model, time point, projects agency, activity representation, research projects agency
Part II. top 5 key sentences
1. AIAI is a technology transfer organisation that promotes the application of Artificial Intelligence research for t he benefit of commercial, industrial, and government clients. 2. AIAI's Knowledge Systems Research concentrates on those areas of Artificial Intelligence that are concerned with explicit rep resentations of knowledge. 3. The goal of the PIF (Process Interchange Format) project is to support the exchange of business process models across different formats and schemas. 4. AIAI is working with Cycorp to develop process and plan representations. 5. Participants have been chosen to be familiar with the issues of plan representation and to be able to reflect specific concerns for their "constituency".

D.9 Artificial Intelligence Center

`http://www.ai.uga.edu`

Part I. top 25 keyterms
artificial intelligence, random access memory, cost function, source code, user module, natural language processing, random access, access memory, genetic algorithm, artificial intelligence center, adaptive simulated annealing module, natural language, visual studio, program options, operating system, forest management advisory system, embedded controller, grammar rule, intelligence center, language processing, definite clause grammar, baud rate, adaptive simulated annealing code, object code, serial port
Part II. top 5 key sentences
1. Artificial intelligence is the computer modeling of intelligent behavior, including but not limited to modeling the human mind. 2. The interdisciplinary Artificial Intelligence Program was established in 1986. 3. The Artificial Intelligence Center is an interdepartmental research and instructional center within the Franklin College of Arts and Sciences of the University of Georgia. 4. It was officially established as the Artificial Intelligence Center in 1995. 5. Areas of specialization include automated reasoning, cognitive modeling, neural networks, genetic algorithms, expert databases, expert systems, knowledge representation, logic programming, and natural-language processing.

D.10 Waterloo Artificial Intelligence Group

`http://ai.uwaterloo.ca`

Part I. top 25 keyterms
local consistency, constraint satisfaction problem, backtracking algorithm, natural language, constraint programming, constraint network, constraint satisfaction, bell lab, satisfaction problem, probabilistic context free grammars, conflict directed backjumping, master document, great scientist, information extraction, arc consistency, steve kerr, important problem, healthdoc project, generalized arc consistency, em algorithm, text specification language, feasible relationship, constraint programming approach, interval algebra, phrase structure grammar
Part II. top 5 key sentences
1. This approach combines techniques from natural language processing and knowledge representation with a penalty-based technique for relevance estimation and passage retrieval. 2. Most of the papers employed machine learning (especially unsupervised or semi-unsupervised learning) for natural language learning tasks such as POS tagging, Name entity extraction, grammar induction, text categorization, and etc. 3. Constraint programming is a methodology for solving difficult combinatorial problems. 4. Constraint programming has shown that a general purpose search algorithm based on constraint propagation combined with an emphasis on modeling can solve large, practical scheduling problems. 5. Our current results show that constraint programming does indeed work well and has the advantage in terms of time and space efficiency over the current state-of-the-art planners.

D.11 Air Canada

http://www.aircanada.ca

Part I. top 25 keyterms
air canada, star alliance, north america, domestic market, open sky, airline industry, airline alliance, canadian airline, airline business, transborder market, business travel, load factor, customer service, global economy, open skies plus, domestic competition, open skies agreement, revenue synergy, market share, trade agreement, slowing economy, revenue opportunity, business traveler, regional carrier, passenger mile
Part II. top 5 key sentences
1. Welcome to the Air Canada Site Index! You can get to anywhere in our site from here, and it's easy to find the information you're looking for. 2. Air Canada is not liable for non-delivery of messages and cannot guarantee the timeliness or reliability of an e-mail sent to your pager network. 3. However, Air Canada will do everything practicable to ensure we provide the most accurate flight status information at all times. 4. Air Canada Introduces New Media Centre This site is designed specifically to provide convenient, quick access to information about Air Canada. 5. Welcome to the world of Air Canada - a world of diversity and opportunity.

D.12 Cisco Systems, Inc.

http://www.cisco.com

Part I. top 25 keyterms
service provider, cisco product solutions essentials, customer satisfaction, virtual private network, operating system software, operating system, technical assistance center, cisco sales expert badge, internetworking operating system, cisco sales expert exam, area network, cisco internetworking operating system, cisco sales expert, service contract, local area network, simple network management protocol, cisco internetworking operating, customer satisfaction objective, local area, protocol telephony, network management, hypertext transport protocol, cisco product, system software, silver partner
Part II. top 5 key sentences
1. Cisco enjoys a unique position as the market share leader for Internet-related networking equipment, and delivers networking solutions to 90service providers. 2. With customer contact solutions designed for every organization – from small and medium businesses to large enterprises and service providers – Cisco has the right combination of Customer Contact Software to enable your business to meet these goals and provide your customers with the exceptional service levels they expect. 3. With Cisco, visionary service providers get a business partner well-versed in the needs of today's hosting environment. 4. Cisco solutions enable service providers to deliver metro bandwidth and to deploy new revenue-generating services for their customers. 5. With Cisco, service providers can create new, high-margin services that will help them attract and retain customers and deliver a sustainable competitive advantage.

D.13 Microsoft Corporation

http://www.microsoft.com

Part I. top 25 keyterms
graphics interchange format, component object model, markup language, office xp, net framework, service pack, windows xp, microsoft office, windows server, front end server, component object, sql server, microsoft windows, object model, short message service, systems management server, back end server, microsoft corporation, microsoft office system, hypertext transport protocol, application center, commerce server, operating system, technet webcast, microsoft product
Part II. top 5 key sentences
1. Many companies, and even whole industries, have benefited by using Extensible Markup Language (XML) to help organize and mana ge text-based information. 2. Office XP Professional integrates productivity innovations throughout its programs to transform the traditional Office suite into a smarter overall work experience. 3. Microsoft Office Professional Edition 2003 is designed to improve how companies can further take advantage of XML. 4. With Office Professional 2003, some of the most popular desktop productivity tools including Microsoft Office Word 2003, Micro soft Office Excel 2003, and Microsoft Office Access 2003now support XML. 5. The Office Update service covers all Microsoft Office products, including Microsoft Word, Microsoft PowerPoint, Microsoft Excel, and Microsoft Outlook.

D.14 Nortel Networks

http://www.nortelnetworks.com

Part I. top 25 keyterms
nortel networks, service provider, nortel networks product, enterprise customer, enterprise network, sockets layer, networks product, virtual private network, second quarter, optical networks, actual result, press release, current expectation, important risk, developer product, certain information, wireless networks, juniper networks, additional information, multiservice switches, nortel networks passport, wide area network, growth rate, qualified employee, industry leader
Part II. top 5 key sentences
1. Nortel Networks has emerged as the leading company that is delivering value to customers around the world through Unified Networks solutions, spanning mission-critical telephony and IP-optimized solutions. 2. At Nortel Networks, we are convinced that in the new millennium our most valued suppliers will play an increasingly important role in helping us achieve timely success with market-leading solutions. 3. Nortel Networks works closely with customers in more than 150 countries and territories around the world to help speed their success. 4. Nortel Networks sells its Enterprise and Small/Medium business solutions through a worldwide network of channel partners, who can deliver everything from smaller voice systems to global integrated solutions for the international business. 5. Nortel Networks industry-leading products are transforming the current maze of multiple networks into one, easy to manage packet-based network where data, voice and video speed to their destinations quickly and more efficiently.

D.15 Oracle Corporation

http://www.oracle.com

Part I. top 25 keyterms
e business suite, oracle e business suite, oracle e business, oracle application, application server, business intelligence, customer relationship management, oracle applications user group, cliff godwin, supply chain, relationship management, javascript enabled browser, oracle product, customer relationship, guru series, oracle process manufacturing, oracle applications user, ever changing collection, real application clusters, business processe, user group, sql server2000, applications user group, oracle contract, applications user
Part II. top 5 key sentences
1. Oracle Applications strongly recommends that our customers convert to Multi-Org as soon as possible. 2. Analysts find Oracle9i Application Server beats competitors for wireless infrastructure: Giga report, Gartner report. 3. Oracle9i Application Server has consistently ranked first in the ECperf and SPECj industry-standard 'Performance' and 'Price/Performance' benchmarks. 4. Oracle9i Application Server turned in 63 percent higher performance than BEA WebLogic, 39 percent higher performance than IBM WebSphere, and price/performance that shows those competitors to be more than twice the cost. 5. Oracle9i Reports, an award-winning business-intelligence component of Oracle9i Application Server, lets you create enterprise-level reports for more informed business decisions.

D.16 Adhesion Technologies

http://www.adhesiontech.com

Part I. top 25 keyterms
financial institution, enhanced account, adhesion technologies, wealth management, account aggregation, financial service, adhesion enhanced account, enhanced account aggregation, software solution, practitioner publishing company, financial account consolidation, north carolina, services provider, financial services provider, adhesion enhanced, leading provider, financial account, financial consolidation, personal financial center, financial planning tool, wealth management service, whole product, financial planning, data collection, aggregation solution
Part II. top 5 key sentences
1. The company's flagship offering, the Finaplex Wealth(TM) solution, enables financial institutions to strengthen client relationships, increase revenue and leverage operating efficiencies - thereby achieving competitive advantage. 2. Financial institutions and their customers continue to benefit from our leading edge data normalization, presentation, and delivery capabilities while operating within their business and cultural environments. 3. The solution is well-suited to enable financial institutions to deliver new, value-added wealth management services due to Adhesion's focus on post-aggregation data normalization and enrichment. 4. EA2 delivers to financial institutions truly usable data that can be acted upon by planning and analysis tools. 5. Adhesion can deploy for you an EA2 based service, with the user interface customized to your standards, and support single sign on and data integration with your other applications all from within our financial institution class facility.

D.17 Application Solutions & Technologies, Inc.

http://www.asti-global.com

Part I. top 25 keyterms
software tool, software solution, software service, migrating legacy system, tool development, software development, asti software service, consulting service, existing system, various technology, implementing software solution, tool development expertise, business operation, system integration, capability maturity model, quality software service, specific need, data warehousing, medical bureau, asti outsourcing service, email asti software development asti expertise, software product, application system, establishing off shore development center, maximize software development productivity
Part II. top 5 key sentences
1. (ASTI) is an established software solution provider to clients, ranging from the federal government to commercial and private industries. 2. ASTI's expertise and experience in software tool development enables ASTI to adopt the prototyping approach to designing and developing a software solution for conducting services. 3. System Integration ASTI is committed to delivering the technologies and services required to design, develop, and deploy software solutions that will help each business, their employees, customers and partners collaborate and succeed in this fast-paced, net-centric environment. 4. ASTI's years of experience in software tool development, enables us to design and develop a "custom" tool for the customer's specific needs wherever necessary. 5. As a Software Services Provider, ASTI is constantly exposed to the challenges businesses face to design, develop, and deploy e-business application systems.

D.18 Commerce One

http://www.commerceone.com

Part I. top 25 keyterms
xml development kit, normalized content event, attribute event, named location, forward looking statement, business process, development kit, xml development, supplier relationship management, containing element, licensed product, consumer packaged goods, specified character, eventcontroller source, process management, consumer packaged goods company, commerce one, supplier relationship, relationship management, supply chain, xml document, interoperability engine, developer community, networked supplier relationship management white paper, object value
Part II. top 5 key sentences
1. This Open-Source, Royalty Free, Java Web Services SOAP and XML Development Kit provides an effective and efficient way of developing Web Services solutions where the emphasis is on making it easy to do document style SOAP. 2. Commerce One Xpress is an integrated suite of implementation tools, training, and business process consulting services that helps CPG companies quickly integrate into the UCCnet GLOBALregistry. 3. Conductor enables businesses to bridge the gaps between disparate applications and automate complex business processes, regardless of the underlying technology. 4. Commerce One's Conductor platform is one of the first complete solutions we have seen that offers a standards-based approach to enable extended business processes and application capabilities across such a complex environment. 5. Commerce One has been helping the world's leading companies - Siemens, Boeing, Daimler Chrysler, General Motors, and others - improve their bottom line and business processes.

D.19 Gamma Enterprise Technologies, Inc.

http://www.getgamma.com

Part I. top 25 keyterms
data object, user exit, infoshuttle move, sap client copy, sales order, master data, application link enabling, external system, related object, development environment, shuttle data, sap system, sap client, posting function, sap release, source system, application link enabling configuration, infoshuttle move transactional data, different sap version, records using infoshuttle, current production data, infoshuttle processing, multiple destination, internal block, multiple client
Part II. top 5 key sentences
1. The shortcoming of Application Link Enabling (ALE) is that it transfers selected data objects, but does not group related objects together, or post objects in the right order. 2. InfoShuttle easily transfers just the data objects that are required to support current events in those environments, leaving intact the unique configurations and other specially created data. 3. InfoShuttle transfers only a subset of data objects (such as Material or Customer) as specified by the user. 4. JSAP lets the developer simply 'point and click' the desired SAP data objects while building highly scalable mission critical applications. 5. What major advantages does InfoShuttle offer over standard ALE? InfoShuttle moves transactional objects and some master objects not supported by ALE.

D.20 RDM Corporation - Electronic & Paper Payment Solutions.

http://www.rdmcorp.com

Part I. top 25 keyterms
quality control, micr line, payment archive, rdm corporation, financial institution, image management, image quality control, image quality, rdm image, payment archive service, print quality, micr qualifier gtx, rdm image management, numerical convenience amount field, micr qualifier, bill payment, image management solution, rdm micr, archive service, image qualifier, management solution, rdm payment archive, including micr line, micr line signal, qualifier gtx
Part II. top 5 key sentences
1. As new low cost MICR document printing methods become increasingly popular, the need for rigorous quality control also increases. 2. The Progressive MICR Method is an RDM innovation that allows our images to scan the MICR line and capture the image in a single pass with unsurpassed accuracy. 3. It is critical for printers, corporations and financial institutions to establish quality control methods in order to bring print quality within the industry standards. 4. We specialize in enabling and supporting the electronic conversion of checks and bill payments, providing complete image management, MICR/OCR and image quality control. 5. RDM Corporation may make improvements and/or changes in the products and/or the programs described in this material at any ti me.