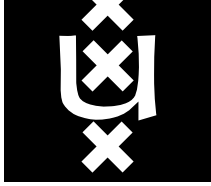


Intelligent Sensory Information Systems
University of Amsterdam
The Netherlands



Text Mining Tools on the Internet

An overview

Jan van Gemert

Intelligent Sensory Information Systems
Department of Computer Science
University of Amsterdam
The Netherlands

Thorough overview of 71 text mining tools available on the internet. In the document I give a brief description of the company and the product. Also the estimated price and the availability of a demo are mentioned. Every product has a Hyperlink to the page describing it. Also, an table summerising the results has been created.

Contents

1	Introduction	1
2	SRA International, Inc	2
2.1	NetOwl Extractor	2
3	Megaputer Intelligence	3
3.1	TextAnalyst: natural language text analysis software	3
3.2	MegaSearch	3
4	IBM Software	4
4.1	Intelligent Miner for Text	4
5	InsightSoft-M	5
5.1	ICrossReader	5
6	Department of Intelligent Systems, Slovenia	6
6.1	Yahoo Planet	6
7	Intercon Systems	7
7.1	Dataset	7
8	Conceptual Dimensions	8
8.1	Concept Map User Interface	8
9	Cartia	9
9.1	ThemeScape	9
10	Management Information Technologies, Inc. (MITi)	10
10.1	IpServer	10
10.2	Software Developer Toolkit (SDK)	10
11	Research Outlet and Integration	11
11.1	twURL	11
12	Semio	12
12.1	Taxonomy engine	12
13	STATE SCIENTIFIC AND TECHNICAL CENTRE FOR HYPERINFORMATION TECHNOLOGIES	13
13.1	System of Meaning Integrities structural Creation (SMIsC)	13
14	PARACEL INC.	14
14.1	TextFinder	14
15	ZyLAB	15
15.1	ZyIMAGE	15
16	Autonomy	16
16.1	Knowledge Server	16

17 Clearforest	17
17.1 IE Studio	17
17.2 Administrator	17
17.3 Text-o-Scope	17
18 Verity	18
18.1 Verity Developer Kit	18
19 Dataware	19
19.1 InQuery	19
20 Excalibur Technologies Corporation	20
20.1 Excalibur RetrievalWare	20
21 Open Text	21
21.1 Basis	21
22 Dataflight Software	22
22.1 Concordance	22
23 dtSearch Corp.	23
23.1 dtSearch	23
24 InfoSphere	24
24.1 ProIndex	24
25 Datagold	25
25.1 Multi-Site Search Software	25
26 America Online	26
26.1 Callable Personal Librarian (CPL)	26
27 Thunderstone	27
27.1 Taxis	27
28 Sigma	28
28.1 TEXCOVERY	28
29 Search Technology	29
29.1 VantagePoint	29
30 Hummingbird	30
30.1 Fulcrum SearchServer	30
31 ASTAware Technologies	31
31.1 SearchKey PRO	31
32 SeekWare	32
32.1 SeekIt Developer TM	32

33 Sunrizen Software	33
33.1 Searchopia	33
34 Eidetica	34
34.1 t-find, t-store, t-mining, and t-repository.	34
35 Signiform	35
35.1 ThoughtTreasure	35
36 Carnegie Mellon University	36
36.1 BOW library	36
36.2 CHILDES	36
37 Xerox	37
37.1 askOnce	37
38 Inxight	38
38.1 LinguistX	38
38.2 Inxight CategorizerTM	38
39 San Diego State University	39
39.1 ht://Dig	39
40 Etymon	40
40.1 Isearch	40
40.2 Amberfish	40
41 KE Software	41
41.1 Textpress	41
42 Index data	42
42.1 Zebra and Z'mbol Information Server	42
43 Executive Technologies	43
43.1 SearchExpress	43
44 United Nations Educational, Scientific and Cultural Organization	44
44.1 ISIS	44
45 Limit point software	45
45.1 Boolean Search 2	45
46 Delphes technologies International	46
46.1 DioWeb Search	46
47 Lextek	47
47.1 Onix Full Text Indexing and Retrieval Toolkit	47

48 Internet Research Task Force Research	48
48.1 Harvest	48
49 Oracle corp.	49
49.1 interMedia	49
50 University of Louvain	50
50.1 PROTAN	50
51 Text Analysis International	51
51.1 VisualText	51
52 NextPage	52
52.1 Folio Integrator	52
53 CHASS	53
53.1 Text Analysis Computing Tools	53
54 Zuma	54
54.1 Textpack	54
55 Oxford University Press	55
55.1 WordSmith Tools	55
56 askSam Systems	56
56.1 askSam	56
57 University of Nijmegen	57
57.1 AGFL system	57
58 Mitre	58
58.1 Alembic	58
59 Centre for Spoken Language Understanding (CSLU)	59
59.1 CSLU Toolkit	59
60 Cogilex	60
60.1 QuickTag and QuickParse	60
61 Basis Systeme netzwerk (BSn)	61
61.1 InterBasis	61
62 Pacific Software Publishing	62
62.1 Alise	62
63 AnswerLogic	63
63.1 AE1	63

64 LexiQuest	64
64.1 LexiQuest	64
65 University of Maryland	65
65.1 Translingual Information Retrieval System (TIDES)	65
66 The Features	66
67 Some general Links	67

Intelligent Sensory Information Systems
Department of Computer Science
University of Amsterdam
Kruislaan 403
1098 SJ Amsterdam
The Netherlands

tel: +31 20 525 7463

fax: +31 20 525 7490

<http://www.science.uva.nl/research/isis>

Corresponding author:

Jan van Gemert

tel: +31(20)525 7507

jvgemert@science.uva.nl

www.science.uva.nl/~jvgemert

1 Introduction

This document, is a result of many hours of searching, reading, e mailing, selecting and cut-n-pasting on the Internet. It aims at an overview of the available text mining applications on the Internet. The scope is rather wide. It ranges from complete document management systems, to software libraries for natural language processing. An Internet search on, (among others): "Text mining, Information retrieval, document analysis, tools, Natural language processing, software, data mining, content analysis" Revealed many, many hits, and even more indirect links. At a point, it was decided to stop searching. The amount of small parsers, little search engines, and other small tool would have kept us searching the web indefinitely. Here are the results, 64 companies, and 71 tools describing:

- A brief description of the company producing the product;
- An overview of the product itself, including a URL;
- The availability of a demo;
- Pricing information.

Also, a table of the features each tool implements is included. In this table, the features are put out against the tools. Answering the following questions, which were asked to the companies:

- Can I search documents for phrases/words?
- Can I parse Sentences with it?
E.g. recognize nouns, verbs, etc. in a grammar
- Does it support keyword/concept extraction of a text?
E.g. "Clinton is playing on the saxophone."
Extracts 'Clinton' and 'saxophone'.
- Does it support a fixed/non-Fixed (extendable by user) taxonomy of concepts?
E.g. a dog is a mammal, a mammal is an animal,
a hand is part of an arm, arm is part of a body, etc.
- Does it use the document structure, if available,
e.g. with XML, can it recognize an <author> tag?
- Is it multi-language supporting?
- Is it possible to have an API interface to your product?
Or complete access to the sourcecode?
- Is there a database interface possibility?
Or a database provided, or ODBC?
- Is there a pricing indication?

For a raw list of links, used to find these tools, see <http://www.science.uva.nl/~jvgemert/> under the section `some links`.

2 SRA International, Inc

SRA is an industry leader in the research, development, and application of advanced natural language processing (NLP) and knowledge discovery in databases (KDD) technologies. SRA has significant experience and a continuing commitment to advancing the state-of-the-art in a number of areas that are key to processing large volumes of structured and unstructured text. Areas of expertise include:

- Data mining
- Multilingual information extraction
- Multilingual information retrieval
- Multimedia clustering
- Text categorization
- Text mining, visualization, summarization, and categorization
- Machine translation
- Speech recognition applications

2.1 NetOwl Extractor

<http://www.textmining.com/>

Initially developed for the most demanding government intelligence applications, NetOwl Extractor is based on advanced computational linguistic and natural language processing technology. By intelligently analyzing structure and context within text, NetOwl accurately identifies key information. NetOwl Extractor is an automatic indexing system that finds and classifies key phrases in text, such as personal names, corporate names, place names, dates, and monetary expressions. NetOwl Extractor finds all mentions of a name and links names that refer to the same entity together. NetOwl Extractor combines dynamic recognition with static look-up to achieve high accuracy and coverage at very high speed.

Demo: Download is currently not available, due to an update of the demo program.

Cost: Yet, unknown.

3 Megaputer Intelligence

Megaputer is a leading manufacturer and distributor of advanced software tools for data mining and knowledge discovery in databases, semantic text analysis, and information retrieval. Our solutions help reveal knowledge hidden in your data warehouse or textbase in order to facilitate better business decisions.

3.1 TextAnalyst: natural language text analysis software

<http://www.megaputer.com/html/textanalyst.html>

The new text mining system, TextAnalyst, implements a variety of important analysis functions based on utilizing an automatically created semantic network of the investigated text. This system is built on the results of twenty years of research and development of a new paradigm by a team of mathematical linguists. The key advantage of TextAnalyst against other text analysis and information retrieval systems is that it can distill the semantic network of a text completely autonomously, without prior development of a subject-specific dictionary by a human expert. The user does not have to provide TextAnalyst with any background knowledge of the subject – the system acquires this knowledge automatically.

Demo:

60 days evaluation version available

Cost:

- Government organizations -40% . Educational institutions -70%
- Normal price: \$ 976 (-70% = \$ 292.8)
- API version: \$ 5,940 (-70% = \$ 1782)

3.2 MegaSearch

<http://www.megaputer.com./html/megasearch.html>

MegaSearch is natural language query based document retrieval tool. It runs fast and easy semantic search and retrieves relevant documents from your PC or an entire local network. MegaSearch turns documents stored on your machine or network into a personal electronic encyclopedia without any effort on your part.

Demo:

60 days evaluation version available

Cost:

- Government organizations -40% . Educational institutions -70%
- Normal price: \$ 259 (-70% = \$ 77.7)

4 IBM Software

At IBM, we strive to lead in the creation, development and manufacture of the industry's most advanced information technologies, including computer systems, software, networking systems, storage devices and microelectronics.

And our worldwide network of IBM solutions and services professionals translates these advanced technologies into business value for our customers.

4.1 Intelligent Miner for Text

<http://www-4.ibm.com/software/data/iminer/fortext/>

Provides a comprehensive suite of text analysis and text search tools :

- The Language Identification tool. This tool automatically discovers the language in which a document is written. You can also train the tool to cover additional languages;
- The Feature Extraction tool. This tool recognizes significant vocabulary items in text, automatically, and without requiring you to predefine a domain-dependent vocabulary;
- The Summarizer tool. Analyzes the words and sentences in a document to produce a summary of the document;
- The Topic Categorization tool. This tool automatically assigns documents to categories, topics, or themes that you have previously defined;
- The Clustering tools. These tools divide up a set of documents into groups, or *clusters*. The members of each cluster are similar to each other because they share common features. The clusters are not predefined; they are derived from the document collection automatically.

Turns unstructured information extracted from workgroup applications and large corporate solutions into business knowledge Includes components for building scalable knowledge management, text mining and text search applications.

Demo:

60 day trail version available

Cost:

The commercial package costs Dfl. 73.000.- The 'academic' version requires Dfl 30.000.-

5 InsightSoft-M

InsightSoft-M is a software developing company. Our leading specialists have a 15-20 years background in approaching the problem of automatic generation of coherent texts.

5.1 ICrossReader

<http://www.insight.com.ru/>

- Hunt only highly relevant documents across WWW.
- Screen texts within an unstructured database and perform semantic clusterization of information.
- Digest the documents to sift paragraphs or sentences relevant to your subject.
- Compose original text surveys on the fly.

Demo:

Free evaluation version available.

Cost:

Normal: \$ 149 Semantic filters: 'Professional' set - \$ 449

6 Department of Intelligent Systems, Slovenia

The principal goals of the Department of Intelligent Systems are to develop a computational theory of intelligence and to develop high-impact practical applications in areas such as intelligent information systems, data analysis, decision making, intelligent agents, medicine, ecology, chemistry, intelligent manufacturing, and economy.

The Department of Intelligent Systems of J. Stefan Institute is one of the established European computer science research groups with a 20 year tradition in R& D in artificial intelligence, intelligent systems, information systems, medical informatics, natural language processing, and cognitive sciences.

6.1 Yahoo Planet

<http://www-ai.ijs.si/DunjaMladenic/yplanet.html>

Yahoo Planet is a project where we use the Yahoo hierarchy of Web documents as a base for automatic document categorization. Several top categories are taken as separate problems, and for each an automatic document classifier is generated.

Demo:

Only a demo version is available, the real product should be ready by fall.

Cost:

Unknown

7 Intercon Systems

Intercon Systems has no company information available.

7.1 Dataset

<http://www.ds-dataset.com/default.htm>

DataSet combines Relational-Database (RDB) paradigm with Focused Information Retrieval paradigm. RDB technology is supplemented with DataSet's unique capabilities to manage text. DataSet provides comprehensive search and retrieval tools, that can locate Items almost instantly, by words, phrases and much more; Interrelationships between stored items are identified, providing tools that allow navigation through text, with unprecedented ease and accuracy.

Demo:

Free evaluation version is available. Including textual test data.

Cost: \$ 49.–

8 Conceptual Dimensions

Conceptual Dimensions, Inc., is a supplier of advanced text analysis, text retrieval, and text visualization products and services for both the pharmaceutical research industry and the Internet and Intranet text publishing markets. We specialize in user interfaces and automated text analysis software enabling your users to search, organize, and navigate your text databases.

8.1 Concept Map User Interface

<http://www.cddimensions.com/products.html>

The Concept Map User Interface is based on a unique new method for accessing and comprehending large groups of documents. This patented technology automatically organizes and groups the results list of documents returned by a text search engine, allowing the user to quickly find useful documents and to view only those documents that meet their needs. The technology has three main components: Client-resident Visualization User Interface Client or server-resident Taxonomy Analysis Sub-System Server-resident Classification Sub-System (optional)

Demo: Downloadable Beta version available.

Cost: No pricing information is available.

9 Cartia

Founded in 1996, Cartia, Inc. produces software able to automatically organize document collections based on the information they contain. The result is a visual landscape of information that shows the actual contents of documents and web pages. The company sells a suite of products based on the technology.

9.1 ThemeScape

<http://www.cartia.com/products/index.html>

ThemeScape is a software application that automatically organizes documents based on the information they contain. In minutes, ThemeScape creates a visual landscape of information - a topographical map - that actually shows you what's inside large collections of documents and web pages. ThemeScape maps convey a tremendous burst of information in just a few seconds - many times faster than reading. With a quick scan of the landscape, you know the major topics within thousands of documents, and how different topics relate to one another.

Demo:

No downloadable version is available. However, a demo is present at the web site.

Cost:

Prices for a ThemeScape publishing system vary depending on configuration, beginning at about \$ 20,000.

10 Management Information Technologies, Inc. (MITi)

MITi is a privately held Research and Development Corporation organized under the Statutes of the State of Delaware. The Company was founded in 1985 to find, fund, and develop more advanced methods for textual information processing. Since 1987 we have been improving our semiotic theory, our theory of textual association, our concept of ontology, and our taxonomy of human knowledge and information.

10.1 IpServer

<http://www.readware.com>

The Readware Information Processor (IpServer) is the first intelligent text analysis server available for use over the Internet. MITi's IpServer, now in its second release, applies Readware technology to systems taking advantage of public access through the Internet or in wide area networks. It is designed for TCP/IP networks, intranets and the Internet. The IpServer is delivered as a software engineering resource kit.

Demo: Demo version is available, after registration.

Cost: More information requested

10.2 Software Developer Toolkit (SDK)

<http://www.readware.com/Readware.htm>

Readware is a system of advanced text analysis by the computer. Information product developers and content providers can use the Readware SDK to custom program products which will: automatically establish fully-indexed, deeply-analyzed collections or archives of digital messages, text files and/or textual records from a database. give your clients a superb automatic information service consisting of customizable intuitive queries that return fast responses focused on the subjects, topics and issues they are interested in.

Demo: Demo version is available, after registration.

Cost: More information requested

11 Research Outlet and Integration

ROI addresses the needs of WWW Information Professionals, competitive analysts, marketing/sales managers, search engine promotion specialists, educators, journalists, librarians, and others who make intensive use of the Web in their professions.

11.1 twURL

<http://www.twurl.com/>

A power tool for the WWW Information Professional who needs to process web content for high quality, relevance, and rapid insight. twURL takes up where search engines leave off and brings order to collection, analysis, selection, and presentation of 1000s of URLs. twURL has been designed for decision-making using linguistic classification of web pages, e.g. to filter out irrelevant, duplicate, or vacuous pages or, more advanced, to select and rank URLs. twURL has only rudimentary, but still useful, classification by user-defined keyword sets.

Demo: Demo version is available, after registration.

Cost: More information requested

12 Semio

Semio is positioned to be a major player in the emerging market for corporate portals, an architectural approach that allows a single point of access to all corporate information. The Delphi Group projects that this market will grow to \$ 5B within five years. For designers of corporate portals, the single greatest challenge is helping users find the information they need. Semio solves this problem. Offered through a monthly service to its customers, Semio's one-of-a-kind taxonomy technology builds and maintains a browsable, searchable directory of concepts.

12.1 Taxonomy engine

<http://www.semio.com/>

Semio's Taxonomy engine creates a multi-level directory structure that includes thesaurus-like links. These links cross-reference related parts of the directory to help match thought patterns both hierarchically and across concepts. After defining the top-level categories appropriate to the content and interests of various users, Semio's Taxonomy engine automatically generates the lower levels of the directory, eliminating the labor-intensive process of creating rules and assigning metatags.

Demo: On the site is some demonstration. Not very detailed. For a sample view of the taxonomy: <http://demo.semio.com/public/taxonomy.cgi>

On appointment, a demonstration can be given.

Cost: Starting at Dfl. 10.000. -

13 STATE SCIENTIFIC AND TECHNICAL CENTRE FOR HYPERINFORMATION TECHNOLOGIES

Solution of diverse analytic tasks on the base of structuralisation of the preliminary selected textual information - is the niche which is occupied by the worked out in Russia the Structural Analytic Technologies, in the meanwhile without any serious competition. The State Scientific and technical Centre for Hyperinformation Technologies (SSTC "HINTECH") has put on the market a complex of instrumental means for realization of these technologies.

13.1 System of Meaning Integrities structural Creation (SMIsC)

<http://www.hintech.ru/smisce/index.htm>

SMIsC is a text-mining computer software tool enabling the user to integrate large unstructured masses of text into browsable networks of local coherence, to uncover in them unexpected patterns of meaning thematically organized text clusters transformable into discourses or narratives, and to analyze these patterns in their statics and dynamics.

Demo: A demo is in development.

Cost: Not for sale, but perhaps in the future free.

14 PARACEL INC.

Paracel develops high-performance genomic data and text analysis systems for the pharmaceutical, biotechnology, information services and government markets. In industries such as bio-informatics where searches are complex and time-critical, Paracel's massively parallel, FDF-based systems, GeneMatcher and TextFinder, deliver data throughput unrivaled by ordinary general-purpose computers.

14.1 TextFinder

<http://www.paracel.com/html/textfinder.html>

TextFinder is the fastest, most accurate, adaptive information-filtering system in the world. It is a massively parallel computer with associated software designed to search, filter, categorize, and disseminate massive quantities of free text. TextFinder is particularly well suited to time-critical problems where spelling errors, foreign languages or non-standard data formats are involved. A typical TextFinder application may involve trillions of bytes of text and thousands of online users, or gigabytes of live data stream per day that are filtered against tens of thousands of complex interest profiles. TextFinder is designed to allow real-time text searching without pre-processing. More than 12,000 processors in each TextFinder make it possible to filter in excess of 40M bytes of text per second. That's equivalent to more than 50,000 pages of text every second.

Demo: Not relevant, because it's a hardware system.

Cost: \$ 120.000

15 ZyLAB

ZyLAB develops complete systems to archive, search, find, organize, share and reproduce all your documents with unparalleled ease of use. Sharing and re-using knowledge starts by easily accessing all your documents. With ZyLAB's archiving systems one can access huge amounts of paper and over 250 different electronic file formats with advanced full-text search techniques. In addition, end-users can organize these unstructured data collections into more structured data collections at very little costs and share them with third parties by using LAN/WAN networks, the Internet, a Intranet or CD-ROMs.

15.1 ZyIMAGE

<http://www.zylab.nl/zylab1/Products/products.htm>

ZyIMAGE is a robust imaging system that lets you retrieve documents based on their content with great flexibility and power. It is a fully integrated solution, which provides unrivalled data capture, full-text retrieval and document management. Provides an API to functions to index, fast (fuzzy and proximity) searching of many documents.

Demo: Some online demonstrations are available. For further testing, it is possible to send files to them, and let them handle them. It's also possible to request a demo CD.

Cost: Fl. 10.000

16 Autonomy

Autonomy, Inc. was founded in March 1996 by Dr. Michael Lynch, a world-renowned expert in the field of adaptive pattern recognition. Headquartered in San Francisco, CA., Autonomy has offices in Boston, Dallas, New York and Washington D.C., Cambridge, UK, Paris, France and Oslo, Norway. Autonomy maintains close ties with Cambridge University in Cambridge, England, where Dr. Lynch originally developed Autonomy's technology.

16.1 Knowledge Server

<http://www.autonomy.com/knowledge/ksfeatures.html>

Automates the accurate categorization and tagging of large volumes of both internal and external information. All that is required is to identify examples of documents within specified categories. Provides powerful natural language search facilities, including relevancy ranking, term highlighting, clustering, automatic summarization and query-by-example. Allow cross-language search. Users can describe their interests using natural language or showing the system an example of what they are interested in. Using neural networks and pattern matching.

Demo:

No demo version available on the web site. More information requested

Cost: From £50.000 to £3.000.000

17 Clearforest

We are a well-established start-up company in the process of massive entry into the Internet market. The search engine we have developed scans huge text banks from various data sources, analyzes the search results and presents the information in a simple, intuitive graphic form.

<http://www.clearforest.com/>

Designed to meet the ever growing challenge of data management in the Internet age, it provides the user with unrivalled leverage in extracting the essential information embedded in large collections of news, patent abstracts, corporate databases and other text-based sources. It does so by means of a uniquely powerful information extraction language known as DIAL (Data Information Analysis Language), capable of assimilating text data of indefinite size, identifying key terms, assigning them to meaningful categories (a taxonomy), and noting their inter-relationships. The result is a highly structured, "intelligent" body of information that the user can "slice and dice" at will, discovering within seconds insightful patterns based on keywords of their choosing in a variety of charts and graphs, while maintaining full access to the source documents at any time. The software is structured around three basic modules, which may be purchased together or separately to suit the user's requirements:

17.1 IE Studio

Information Extraction Studio For the writing or customizing of information extraction rules based on the DIAL language to be used during text-mining operations.

17.2 Administrator

A suite of utilities, for finding and downloading source texts from Internet sources or converting existing local collections (as applicable) into suitable proprietary format, information extraction and/or term extraction, building a thesaurus (if required), creating a taxonomy, and other operations necessary for the construction of a database repository file (.drf) ready for interrogation by the user.

17.3 Text-o-Scope

For the aforesaid interrogation of the structured .drf and display of results in a variety of charts and graphs.

Demo: Not yet available, site is under construction.

Cost: More information requested

18 Verity

Verity's mission is to lead the market for knowledge retrieval solutions by turning unstructured text-intensive information into usable and shareable knowledge. Verity's products and services are designed to power the next generation of content rich corporate, OEM and e-commerce applications and are used by more than 1000 market leading organizations worldwide.

18.1 Verity Developer Kit

<http://www.verity.com/products/devokit/index.html>

Provides unparalleled high-speed search precision. Queries are expanded automatically using thesaurus, word stemming and linguistic analysis. A comprehensive set of query functions support Boolean, proximity, zone, density, typo, frequency, field, concept and fuzzy searching operations. Includes advanced search capabilities combining metadata, concept (Topics) and full text searching. Advanced search features include relevancy ranking, term highlighting, clustering, summarization and query-by-example.

Demo: More information requested

Cost: More information requested

19 Dataware

Dataware is a leading provider of e-Business solutions to Global 2000 and "Dot-Com" companies worldwide. We use a unique combination of our proven Xcellera Methodology, cutting-edge skills, superior technology components, and significant experience to deliver innovative, Internet-based solutions at Net Speed. Dataware solutions extend the Internet's potential to provide our customers with key competitive and operative advantages.

19.1 InQuery

<http://www.sovereign-hill.com/>

Dataware's InQuery is the most advanced knowledge discovery tool in the industry. InQuery couples advanced intelligent search technology with intuitive concept mining to provide transparent and extensive access to information in over 200 file formats. InQuery's highly scalable, distributed component architecture enables you, the knowledge worker, to easily pinpoint the most relevant information in any source on your enterprise-wide network.

Demo:

A downloadable demo version is available, after registration.

Cost:

<i>InQuery Pricelist</i>						
<i>£ STERLING</i>			<i>Version 2.1</i>		<i>Jan 25 2000</i>	
<i>Price Band</i>	100-160	161-320	321-640	641-1280	1281-2560	> 2560
<i>Change Number of User Seats :</i>	100	320	500	1280	2560	10000
	£17,688	£37,000	£48,250	£75,000	£91,000	£160,750
<i>Total Software</i>	£17,688	£37,000	£48,250	£75,000	£91,000	£160,750
<i>Maintenance at 18%</i>	£3,184	£6,660	£8,685	£13,500	£16,380	£28,935
<i>Total, incl. Maintenance</i>	£20,872	£43,660	£56,935	£88,500	£107,380	£189,685

This pricelist applies to all platforms and operating systems

20 Excalibur Technologies Corporation

Excalibur's mission is to provide access to the precise information people require despite its form or location. Our industry-leading, innovative Internet e-Commerce and Intranet Corporate Portal and Media Management search solutions help capture, index, catalog, access, navigate, retrieve, publish and share any data, anywhere - text, image, video, audio and paper.

20.1 Excalibur RetrievalWare

<http://www.excalib.com/products/rw/rwadvantages.shtml>

Advanced Search and Retrieval

Highly accurate, natural language concept searching based on Excalibur's unique semantic network (available in English, French, Spanish and German; inquire about other languages). Range of search features including concept and keyword searching, idiom recognition, fielded searching, query-by-example and more. Adaptive Pattern Recognition Processing (APRPTM) technology allows fault-tolerant search against error-prone text from OCR processes, misspelled words, and irregular names. Proper name recognition and summarization which allows users to request assistance in viewing and locating people, places and things. Custom Search Templates can be easily created for specific libraries for faster searching and field validation. Excalibur RetrievalWare adds key KM features to version 6.7 Document Summarization provides a fast and efficient way to determine a document's subject and key themes. Advanced Web-based interface for "power-users," as well as optional, simplified Document Explorer.

Demo: Several demonstrations are available on the website.

Cost: Unknown

21 Open Text

Open Text is the leading provider of intranet, extranet and e-business applications. We have achieved this position through our core technology assets; deep technical expertise in scalable intranet and extranet platforms and more than a decade of related experience in workflow and document management applications.

21.1 Basis

<http://www.opentext.com/basis/>

BASIS puts all of your corporate knowledge at your fingertips with a robust document collection and a high performance search and retrieval solution. Designed for comprehensive library control, BASIS provides an ideal solution for companies who need access to hybrid document collections consisting of both documents and metadata. Used by major commercial and government information centers, BASIS provides library automation, research and records management, technical document publishing, legislative tracking, content archives, litigation support, and competitive intelligence. The system also supports a range of powerful search and retrieval functions, including traditional relational queries as well as text queries against structured and unstructured text. One of the strengths of BASIS is the precision of searching large document collections. Full-text queries, nested queries, proximity searches and Boolean operators allow anyone to quickly locate the documents or information they require.

Demo:

A demo CD can be ordered on the site.

Cost: More information requested

22 Dataflight Software

No company information available, on the site.

22.1 Concordance

<http://www.Dataflight.com/Index.html>

Dataflight produces Concordance, a full text database management package for the retrieval of resumes, depositions, research, video tape and film archives, and World Wide Web server publishing. Concordance can retrieve full text and fixed field data quickly and easily, whether it is stored locally, on a LAN, at remote locations through a wide area network, very remotely over the Internet, or stored on CD-ROM.

Demo: A Demo is available, on the page.

Cost: More information requested

23 dtSearch Corp.

dtSearch Corp. develops, manufactures and sells the award-winning dtSearch(R) line of text search and retrieval products. The product line is known for its "industrial-strength" (PC Magazine) ability to instantly search gigabytes of text.

23.1 dtSearch

<http://www.dtsearch.com/dtsoftware.html>

dtSearch's proprietary indexing and searching algorithms allow for fast indexing and searching performance even over extremely large databases and other diverse collections of documents. The algorithms are engineered to maintain consistent indexing speeds regardless of the size of the document set. Indexed search speed is generally less than a second, even through multiple gigabytes of text. dtSearch products also provide unindexed search options.

Demo:

A downloadable Beta version is available.

Cost: \$ 999 See also, <http://www.textinfo.nl/> for a Dutch reseller.

24 InfoSphere

InfoSphere is the full-text indexing and retrieval technology specialist. With over Ten years experience developing customized full-text indexing and retrieval solutions for the information technology industry, InfoSphere has the experience, stability, and reliability to help your development team add this technology to your applications. ProIndex, InfoSphere's indexing and retrieval development package, was released on June 1, 1995 and is available on over ten platforms and multiple programming environments.

24.1 ProIndex

<http://www.proindex.com/proindex.htm>

ProIndex is a software development toolkit that allows software developers to add full-text searching to their applications. It consists of a linkable library or DLL of related functions for indexing and searching any type of textual data. ProIndex is available on a wide variety of development platforms for use with C/C++ compilers along with other development environments such as Visual Basic and Delphi. ProIndex is the perfect solution for distribution of information across a wide variety of mediums including CD-ROM, network, and Internet applications.

Demo: Downloadable demo version is available.

Cost: \$ 3000 - \$ 12000

25 Datagold

Datagold is a search engine developer and builder. As well as licensing our technology, we offer a full search engine building and hosting service.

25.1 Multi-Site Search Software

<http://www.datagold.com/uk-tech.htm>

The Datagold multi-site search system allows you to input the URLs of any Web sites that interest you, and then fully index every page they contain. This creates a fully featured search engine dedicated to the content of those sites, which can then be published on the Internet or a corporate Intranet. The search "front end" software is driven by any standard Web browser. It supports all the most important features found in the major search engines, including natural language searching, Boolean operators (AND, OR, NOT), phrase searching, word proximity and relevancy ranking. Its appearance and operation can be fully customized as required.

Demo: Downloadable demo, for 60 days, is available.

Cost: \$ 750 - \$ 3,850

26 America Online

Founded in 1985, America Online, Inc., based in Dulles, Virginia, is the world's leader in interactive services, Web brands, Internet technologies, and e-commerce services.

26.1 Callable Personal Librarian (CPL)

<http://www.pls.com/cpl.htm>

CPL enables information providers to develop custom retrieval systems to manage full text, structured data, hypertext, forms-based searching and multimedia applications. CPL is highly extensible and customizable, and its open architecture allows for a wide variety of applications. CPL is intended for software programmers familiar with the C language. CPL's object-based API enables developers quickly and easily to construct indexing, search and retrieval applications. CPL combines natural language and Boolean queries with the most effective document relevance ranking in the industry. CPL is best known for its unique ability to discover concepts on the fly with no extra effort by users or administrators. This industrial-strength tool provides concurrent search and administration with transaction integrity. CPL stands up to 24 x 7 operation and is available on most major platforms. CPL's robust performance is illustrated by its adoption in a wide spectrum of applications. For example, it is the backbone of PLWeb, it supports major online systems and it is used in distributed applications such as CD-ROM publications.

Cost: The source code and the product are Free.

27 Thunderstone

Thunderstone is an independent R& D company that has been providing state-of-the-art solutions to intelligent information retrieval and management problems for over 19 years. More Internet searches are conducted by our software on a daily basis than any other available package.

27.1 Taxis

<http://www.thunderstone.com/jump/taxisdetail.html>

TEXIS is the only fully integrated SQL RDBMS that intelligently queries and manages databases containing natural language text, standard data types, images, video, audio, and other payload data. In TEXIS you can store text of any size, and you're able to query that information in natural language for just about anything you can imagine. We took our powerful Metamorph concept based text engine and built a specialized relational database server around it.

Demo: Online demo is available

Cost: > \$ 10,000

28 Sigma

Sigma is a professional services firm focused on: Providing its customers with exceptional quality information systems solutions, Development of innovative and market competitive technologies through a broad spectrum of research and development efforts, Fostering commercialization and market introduction of information system products

28.1 TEXCOVERY

<http://www.sigma-sys.com/>

The recent explosion of on-line information has generated an urgent need for more effective and more “intelligent” tools for information access. The main objective of the TEXTCOVERY system is to help answer some of today’s information access needs by introducing an intelligent tool for text document classification. TEXCOVERY utilizes advanced machine learning techniques that have already proven to be effective in text classification applications. These learning techniques are integrated together in a unique way to yield classification results that are superior to those of the individual techniques. TEXCOVERY also utilizes advanced HCI and visualization techniques in order to help its users in interpreting the outputs of the different system modules. A number of closely related information access tasks can directly benefit from TEXTCOVERY. These include retrieval, routing, filtering, categorization and browsing.

Demo: More information requested

Cost: More information requested

29 Search Technology

Search Technology performs research and development in human-relevant aspects of technology for complex systems. We employ human-centered approaches to analysis, design and implementation that insure technology will support aiding and training of humans in systems

29.1 VantagePoint

<http://www.TheVantagePoint.com/>

VantagePoint provides Competitive Technical Intelligence professionals and Technology Managers with new, powerful, and unique capabilities to help extract knowledge from text databases. VantagePoint works with text data from bibliographic databases. The user first conducts a search using a search engine provided by the database provider. The raw data is downloaded to the user's computer, and if it is delivered in "chunks" of records, it is merged into one file before processing. VantagePoint is most useful when the user's search strategy returns more than you want to read. VantagePoint can provide great benefit when working with only a few hundred records, but it is most helpful with several hundred or even several thousand records. VantagePoint can also extract meaningful words and phrases from the abstracts using Natural Language Processing techniques.

Demo:

Information about how and why, is available at the site.

Cost: \$ 6,500

30 Hummingbird

Hummingbird is a leader in the development of enterprise software solutions that provide access to all business-critical information and resources.

30.1 Fulcrum SearchServer

<http://www.hummingbird.com/products/dkm/km/searchserver/index.html>

SearchServer is an advanced information retrieval solution for high-volume, line-of-business applications, such as electronic publishing, e-commerce, customer care and on-line technical support. It provides enhanced indexing and data management, plus flexible, powerful search options. And its search kernel draws the most from a server, delivering tremendous speed and reliability.

Natural-language searching: Thanks to PC DOCS/Fulcrum WordSense, linguistic technology based on a semantic network, SearchServer users can derive accurate results from natural-language queries. PC DOCS/Fulcrum WordSense analyzes queries—to reduce ambiguities—before beginning a search. Users can tailor queries to their needs through semantic network control options.

Multilingual document searching: SearchServer supports all main European languages, plus Japanese and Korean. It supports the Japanese and Korean versions of Microsoft Office 97 and other multi-byte document formats.

Demo: Not available on the web site. More information requested

Cost:

- SearchServer server on NT: 7,800 USD
- SearchServer server on UNIX: 15,600 USD
- SearchServer server on NT additional CPU: 3,900 USD
- SearchServer server on UNIX additional CPU: 7,800 USD
- SearchServer server on NT and UNIX HTML rendering option: 3,900 USD
- SearchServer client access license Read Only: 254 USD
- SearchServer client access license Read & Write: 384 USD

Yearly maintenance fee is 15% of the license fee.

31 ASTAware Technologies

ASTAware is a software developer using Java-based technology to build high-speed search and retrieval software tools to help businesses navigate in a web-based, e-commerce environment. Additionally, our technology is readily used by software developers to build search and retrieval capabilities into their own web-based and other applications and is easily integrated and interfaced with other technologies and products.

31.1 SearchKey PRO

http://www.astaware.com/r_prod_info.html

ASTAware SearchKey PRO is a JavaTM-based, search engine application. SearchKey PRO offers full-text indexing and searching for HTML, PDF and Text documents. It operates on any platform and enables end users to search multiple indexes, domains or servers, concurrently or individually. SearchKey PRO will add value to your intranet or ISP business.

Demo: Both an online demo as a downloadable version is available.

Cost: \$ 10,000

32 SeekWare

SeekWare Inc. evolved in 1997 as a software products company which focuses on providing specialized text search and document retrieval toolkits. The SeekWare product line is built upon time tested core search technology that has been in use within government and industry since 1991. The new generation of tools we have developed are targeted to the needs of Web site developers, system administrators, systems integrators, and software developers. SeekWare also offers access to our years of experience in the development of document-centered solutions via consulting, software development, and training services.

32.1 SeekIt Developer TM

<http://www.seekware.com/>

SeekIt Developer TM provides the software developer with a toolbox of functions that support the creation of real-time text search and document routing systems. The core of the product is a programming library that provides a high performance sequential text search engine and support software. The library serves as a base for building real-time text search applications on a variety of platforms. The search engine allows numerous queries to simultaneously search every word and phrase of an incoming text stream, automating the content based routing of textual information to large populations of users. The unique capabilities of SeekIt Developer allow it to perform this function in a timely manner that is not feasible for traditional indexed retrieval systems.

Demo: No demo available on the site.

Cost: \$ 9,999 first seat, \$ 4,999 per additional seat, \$ 4,999 per additional platform, \$ 24,999 binary distribution rights, \$ 24,999 source code

33 Sunrizen Software

No company information available

33.1 Searchopia

<http://members.aol.com/sunrizen/index.html>

Searchopia integrates multiple file searching with file viewing. Searchopia looks into the contents of files for words and phrases using the logical boolean operators - And, Or, & Near (proximity searches). Files can also be found by their dates with or without the content search. A search can be as vast as an entire disk or as small as a single file. You select the file types and folder to search, including subfolders. Searchopia displays search times in seconds and makes information retrieval easy by highlighting all the matching text.

Demo: A shareware version is available

Cost: \$ 10 (U.S.) and \$ 5.- extra for each additional person per site.

34 Eidetica

Founded in October 1998 by scientists of the Dutch national research centre for Mathematics and Computer Science CWI, Eidetica today is a small and specialist team of skilful experts providing proven, highly focused knowledge management solutions on a hosted basis, also known as Application Service Providing (ASP). Backed by investments of CWI and Twinning

34.1 t-find, t-store, t-mining, and t-repository.

<http://www.eidetica.com/site/index.html>

Eidetica's hosted software suite consists of a search engine t-find, a text analysis and indexing system t-store, and t-mining, an advanced text mining and automatic classification system. The three tightly integrated products and services are based on the enabling software core t-repository: a textual repository system with term recognition and automated thesaurus construction that tightly integrates features of a database and an information retrieval system.

Demo:

Online demo available on the site.

Cost: Dfl.100.000 a year

35 Signiform

Signiform was founded in 1997 by an MIT and UCLA-trained computer scientist, to bring natural language and commonsense capabilities to computers and devices.

35.1 ThoughtTreasure

<http://www.signiform.com/tt/htm/tt.htm>

ThoughtTreasure contains a database of 25,000 concepts organized into a hierarchy. For example, Evian is a type of flat-water, which is a type of drinking-water, which is a type of beverage, which is a type of food, and so on.

Each concept has one or more English and French approximate synonyms, for a total of 55,000 words and phrases. For example, associated with the food concept are the English words food and foodstuffs and the French words aliment and nourriture (and others).

ThoughtTreasure contains 50,000 assertions about concepts, such as: a green-pea is a seed-vegetable, a green-pea is green, a green-pea is part of a pod-of-peas, and a pod-of-peas is found in a typical grocery store. ThoughtTreasure contains about 100 scripts, or computer-understandable descriptions of typical activities such as going to a restaurant or birthday party.

ThoughtTreasure contains 70,000 lines of C code implementing:

- Text agents for recognizing words, phrases, names, and phone numbers;
- Mechanisms for learning new words;
- A syntactic parser;
- A natural language generator;
- A semantic parser for producing a surface-level understanding of a sentence;
- An anaphoric parser for resolving pronouns;
- Planning agents for achieving goals on behalf of simulated actors ;
- Understanding agents for producing a more detailed understanding of a discourse.

ThoughtTreasure will not be supported, as the author is taking a full-time job elsewhere.

Demo: A non-commercial version is free available for compilation.

Cost: For MIA, More information is requested.

36 Carnegie Mellon University

School of Computer Science

36.1 BOW library

<http://www.cs.cmu.edu/~simsmccallum/bow/>

Bow: A Toolkit for Statistical Language Modeling, Text Retrieval, Classification and Clustering Bow (or libbow) is a library of C code useful for writing statistical text analysis, language modeling and information retrieval programs. The current distribution includes the library, as well as front-ends for document classification (rainbow), document retrieval (arrow) and document clustering (crossbow).

Demo: The code conforms to the GNU coding standards. It is released under the Library GNU Public License (LGPL). You are welcome to use the code under the terms of the license for research or commercial purposes.

Cost: Free.

36.2 CHILDES

<http://childes.psy.cmu.edu/>

The CHILDES system provides tools for studying conversational interactions. These tools include a database of transcripts, programs for computer analysis of transcripts, methods for linguistic coding, and systems for linking transcripts to digitized audio and video. The Child Language Data Exchange System (CHILDES) is an international facility centered at Carnegie Mellon University. The director is Brian MacWhinney and the co-director is Catherine Snow. The system is dedicated to facilitating the study of language learning by both children and adults. It provides a large database of conversational interactions, a transcription system for these interactions, and a series of programs for analyzing these data. The CLAN programs permit the linking of transcript to digitized video and audio data. They also provide methods for the study of phonological, narrative, discourse, syntactic, morphological development in language learners.

Demo: The data, programs, and manuals are freely accessible over the net and can also be obtained on CD-ROM.

Cost: You have to become a member of the project.

37 Xerox

The world's leading copier maker. The big green start button has made sharing documents the lifeblood of business for the last 40 years. But we're also leading the way in areas where our name isn't the first one you think of.

37.1 askOnce

<http://www.xerox.com/go/xrx/software/overview.jsp?id=askOnce&cat=%2fSoftware%2fDocument+Management>

askOnce (Meta-Search Software) "See Beyond the Search Engine" askOnce is a web-based meta-search application that allows you to search multiple repositories and data-types with a single query. askOnce provides a uniform query interface to existing e-mail, database, document repository, internet or Intranet installations, allowing you to locate content via your web browser, without modifying the location or attributes of the information itself. askOnce is highly secure, works with any repository or database via easily-customized "Wrappers," and does not require the installation of client-side application software.

Demo:

Free evaluation copy available.

Cost: More information requested

38 Inxight

Inxight develops software products that improve the way users navigate, preview, find and analyze information on the web. The products allow companies to enhance the productivity and the quality of their users' experiences as they navigate, assess, and understand available information. The products are packaged as building blocks that can be combined with commonly available databases, search engines, document management applications and collaboration platforms to create world-class portals.

38.1 LinguistX

http://www.inxight.com/products_sp/linguistx/index.html

The LinguistX Platform is a fast, comprehensive suite of multilingual text services. It includes automatic language identification, tokenization, stemming, part-of-speech tagging, and noun/phrase extraction. It is based on flexible, high-performance, finite-state technology invented at Xerox Palo Alto Research Center (Xerox PARC), and powers many of the web's premier search engines.

38.2 Inxight CategorizerTM

A premier-class knowledge management tool that automates the process of assigning electronic documents to a taxonomy of pre-defined subject categories (e.g. a Yahoo!-like hierarchy).

Demo: More information requested

Cost:

LingusitX Platform - includes one language \$ 150K one-time advance payment against a 3% royalty \$ 20K annual maintenance (includes tech support, bug fixes, major and minor upgrades)

Summarizer - one language \$ 75K one-time advance payment against a 2% royalty \$ 10K annual maintenance (includes tech support, bug fixes, major and minor upgrades)

Additional languages (same pricing for both products) \$ 25K one-time advance payment per each additional language \$ 5K annual maintenance for each additional language

39 San Diego State University

39.1 ht://Dig

<http://dev.htdig.org/htdig-3.2/>

The ht://Dig system is a complete world wide web indexing and searching system for a small domain or intranet. This system is not meant to replace the need for powerful internet-wide search systems like Lycos, Infoseek, Webcrawler and AltaVista. Instead it is meant to cover the search needs for a single company, campus, or even a particular sub section of a web site. As opposed to some WAIS-based or web-server based search engines, ht://Dig can span several web servers at a site. The type of these different web servers doesn't matter as long as they understand the HTTP 1.0 protocol.

Demo: Source freely available.

Cost: Free

40 Etymon

An etymon is a word from which other words derive; the original, true form of a word. The meaning of an etymon can be preserved through many centuries and languages. In the 21st century, the quality of information will depend on the quality of information systems. Etymon's goal is to develop software tools with an emphasis on creative design and engineering, to provide a powerful base for building your information architecture.

40.1 Isearch

<http://www.etymon.com/Isearch/>

Isearch is an open source software package for indexing and searching text documents. It supports full text and field based searching, relevance ranked results, and Boolean queries, and it can index many kinds of documents, including HTML, mailfolders, list digests, and text with SGML-style mark up. Additional document types can be added as C++ modules. Isearch also includes CGI tools to provide a web interface for searching.

40.2 Amberfish

Amberfish is the commercial version of Isearch. More information requested.

Demo: The Isearch source is freely available.

Cost: Isearch is free. And upon contacting the supplier, it was likely that Amberfish will also be free, however not open source.

41 KE Software

KE Software Inc. is a private company with extensive commercial experience in Australia, Asia and North America. Owned by the principals and key staff, its primary focus is the information retrieval market, particularly where the data consists of a mixture of structured and textual information and multimedia data, and where high speed of retrieval and potential for data growth are important factors. Consulting services, research and development, and software training also constitute significant components of the company's business.

41.1 Textpress

<http://www.kesoftware.com/kesoft/texintro.html>

KE Texpress is a multi-user database management system which provides a rich suite of tools and utilities for flexible and cost-effective data management. Use KE Texpress' object-oriented features to design databases with any combination of fully indexed relational data, free text and multimedia objects.

Demo: An online demo is on the site.

Cost: A 2 user license costs \$ 4,000 USD, a 5 user license costs \$ 8,800 USD and a 10 user license costs \$ 16,000 USD / \$ 20,000 CAD.

42 Index data

We are a consulting and software-development enterprise working out of Copenhagen in Denmark. We specialize in solutions for distributed information management and discovery.

42.1 Zebra and Z'mbol Information Server

<http://www.indexdata.dk/zebra/>

Zebra is a fielded free-text indexing and retrieval engine with a Z39.50 frontend. You can use any compatible, commercial or freeware Z39.50 client to access data stored in Zebra. Zebra may be used free-of-charge in non-profit applications by non-commercial organisations. If you're interested in large-scale databases (up to tens of millions of records, gigabytes of data), or if you have a commercial application, you might be interested in the Z'mbol Information Server which offers additional functionality and performance with the same powerful Z39.50 protocol integration. Z'mbol is designed to meet your requirements to structure and manage information, and to share that information with others in new ways. It combines a powerful structured text search engine with standards-based interfaces for input and output of data. Z'mbol is designed to act as the core of an information sharing strategy, or to form a piece of a larger puzzle - in a consortium or community of organisations who share the same abstract data models.

Demo: Zebra is free downloadable,

Cost: £32.000 or £10.000 a year

43 Executive Technologies

Executive Technologies, Inc., was founded in 1984 to develop SearchExpress. In the modern economy, companies and organizations need to constantly create better solutions, products, and services for their clients. We believe information is the engine of creation, and, to that end, we provide solutions and tools to make information easier to find and capture.

43.1 SearchExpress

<http://www.searchexpress.com/index.htm>

The SearchExpress Ranked Search retrieves documents in order of relevance. This saves valuable time by showing you the most relevant documents first, and is particularly important when you have large numbers of documents. The SearchExpress Boolean Search gives you the power of boolean searching, including AND, OR, NOT, PROXIMITY, and RANGE searches. Other search features include word rooting and wild-carding. See specifications. The database field search capability includes range searching of numeric and date fields. For example, you can restrict a search to all faxes sent or received between 6/1/96 and 1/31/97. The SearchExpress Concept Search allows you to retrieve documents based on concepts. For example, a search for "GPS" would also automatically retrieve documents containing the phrase "Global Positioning System" The SearchExpress Fuzzy Search lets you find documents, even if the words being searched are misspelled due to optical character recognition (OCR) errors..

Demo: A downloadable version for windows is available on the site.

Cost: \$ 995

44 United Nations Educational, Scientific and Cultural Organization

The main objective of UNESCO is to contribute to peace and security in the world by promoting collaboration among nations through education, science, culture and communication in order to further universal respect for justice, for the rule of law and for the human rights and fundamental freedoms which are affirmed for the peoples of the world, without distinction of race, sex, language or religion, by the Charter of the United Nations.

44.1 ISIS

<http://www.unesco.org/webworld/isis/index.html>

The ISIS software family is a set of textual database softwares based on the same data formats to which a very strong retrieval engine (for searching information) and an even more powerful formatting language (for displaying/output of information) is applied. The data format allows for textual information of the 'semi-structured' type to be stored in a highly economical but still very organised way : all texts are stored in 'fields', and fields can have an undefined number of occurrences (zero, one or more), an undefined length (between the limits of - depending on the ISIS-member - about 8 to 64 Kb) and, if desired, subdivided into subfields. By these capabilities, ISIS software can accommodate for the international information formats MARC, CCF and other ISO-2709 based formats. These formats are most widely used in libraries, documentation- and information centres, but as long as the information fits within this semi-structured textual frame (alphanumeric data which can be grouped into logical, structured records) ISIS delivers the right approach. Applications for factual information (directories of organisations, projects, experts, scientists...) abound as well as more catalogue-like, bibliographic databases.

Demo: free

Cost: Free for non-commercial use.

45 Limit point software

No company information available

45.1 Boolean Search 2

<http://www.limit-point.com/BooleanSearch.html>

Boolean Search 2 is a powerful server-side search solution that offers lightning-fast access to information on web and file servers. We're proud of this upgrade and believe it introduces a new class of search software. It delivers a quantum performance leap over previous versions and a robust new feature set which includes high-level search methodology unique to Macintosh search engines. Boolean Search 2 is sophisticated enough to find even the most obscure information on large internet and intranet sites, yet flexible and rich-featured to accommodate the specialized needs of smaller sites. The cornerstone of the product's uniqueness is its dedication to "intelligent search" and to "search engineering" methods which fine-tune the relevancy and precision of search results.

Demo: 30-days downloadable demo available on the web site.

Cost: \$ 247

46 Delphes technologies International

The scientific and technological perspective of Delphes technologies International, related to natural language-based software development, is that, for every human, language structure is the essence of knowledge acquisition. In that perspective, software development must favor a human approach by first examining the underlying structures of language, or the mechanisms, that govern language organization. Every sentence uttered, regardless of language, is based on mechanisms common to and at the origin of all spoken languages.

46.1 DioWeb Search

<http://www.dioweb.com/01English/DioWeb/dioweb.html>

DioWeb Search is a true natural language search tool featuring powerful morphological functions that brings a new dimension to information search in Web sites and corporate Intranets. In the near future, DioWeb Search will include a multilingual “cross-language” operator that will enable users to carry out search requests in their own language and extract all relevant information regardless of the native language of original documents within the visited Web site.

Demo:

Downloadable demo available on the site.

Cost: Number of languages Number of Documents License Cost \$ 1 1000 495 2
1000 795 3 1000 995 1 5000 1,495 2 5000 1,795 3 5000 1,995 1 10,000 2,995 2 10,000
3,295 3 10,000 3,495

47 Lextek

Founded in 1993, Lextek International was created to provide advanced full-text search capabilities and linguistic technologies. Since that time, our clients have ranged from small start-up companies to large corporations such as, Apple Computer, Prodigy, Glaxo-Wellcome, and Motorola, just to name a few. Our text indexing technologies have been used in document mangement systems and web based applications, indexing millions of pages of text with ease.

47.1 Onix Full Text Indexing and Retrieval Toolkit

<http://www.lextek.com/onix/>

C++ code library for indexing and searching data, including very small and fast indexing, quick retrieval, flexible queries including proximity operators and wild-cards, stemming, multiple languages and Unicode support. Handles data from a few hundred K to hundreds of Gigabyte, and includes unlimited technical support.

Demo: An evaluation version is available for download at the web site.

Cost: More information requested

48 Internet Research Task Force Research

Harvest was designed and built by the Internet Research Task Force Research Group on Resource Discovery (IRTF-RD). IRTF-RD is supported primarily by the Advanced Research Projects Agency (contract number DABT63-93-C-0052), with additional support from the Air Force Office of Scientific Research (award number F49620-93-1-0082), the National Science Foundation (grant numbers CCR-9002351, CCR-9301129, CDA-8914587, CDA-8914587AO2, NCR-9105372, and NCR-9204853), Hughes Aircraft Company (under NASA EOSDIS project subcontract number ECS-00009), two equipment grants from Sun Microsystems' Collaborative Research Program, and the University of Colorado's Office of the Vice Chancellor for Academic Affairs.

48.1 Harvest

<http://www.tardis.ed.ac.uk/harvest/>

Harvest is an integrated set of tools to gather, extract, organize, search, cache, and replicate relevant information across the Internet. With modest effort users can tailor Harvest to digest information in many different formats from many different machines, and offer custom search services on the web.

Demo: The source code is free available

Cost: The Harvest software may be used and re-distributed without charge, provided that the software origin and research team are cited in any use of the system.

49 Oracle corp.

Leading supplier of software for information management, and the world's second largest independent software company. With annual revenues of more than \$ 9.7 billion, the company offers its database, tools and application products, along with related consulting, education, and support services, in more than 145 countries around the world.

49.1 interMedia

<http://technet.oracle.com/products/intermedia/>

interMedia Text extends Oracle8i by indexing any text or documents stored in Oracle8i, in operating system flat files or URLs. It also enables content-based queries, (e.g., find text and documents which contain this word) using familiar, standard SQL. It enables file-based text applications to use Oracle8i to manage text and documents in an integrated fashion with traditional relational information.

Demo:

none.

Cost: More information requested

50 University of Louvain

50.1 PROTAN

<http://www.upso.ucl.ac.be/protanae.html>

PROTAN (for PROTOcol ANalyzer) is a computer-aided content analysis system. Being aided by the computer means, in the present case, that PROTAN does the many tedious tasks of textual analysis that a human being can do but generally avoids doing, like counting words. Not infrequently, without further notice, PROTAN will do its job "by default", that is, by assuming that parameters have the values given initially to the system. For instance, some system's tasks require as little information as a semicolon, picking in its memory for the rest of the information required. Never, however, PROTAN does automatic content analysis.

Demo: More information requested

Cost: More information requested

51 Text Analysis International

Text Analysis International, Inc. (TextAI) was founded in 1998 to bring to market a new and pragmatic approach for analyzing electronic text. TextAI is a privately held software development company poised to take advantage of the surging demand for effective text analysis solutions with its groundbreaking VisualText™ technology. VisualText™ is the culmination of years of research and development in the field of language processing. The Company's products are text analyzers that automatically deliver critical content to match customer needs

51.1 VisualText

<http://www.textanalysis.com/Products/products.html>

VisualText is a unique integrated development environment (IDE) for developing text analyzers. It includes our revolutionary NLP++ programming language for rapid analyzer building, Conceptual Grammar KBMS for representing linguistic, conceptual, and domain knowledge, a rule generation system that learns from samples, and a runtime analyzer engine. Analyzers can run stand-alone or embedded in larger applications.

Demo: More information requested

Cost: \$ 50K to \$ 100K

52 NextPage

NextPage connects you with all the business-critical information you need, right from your own corporate portal. As the leader in content delivery solutions, NextPage provides real-time access to your must-have content, regardless of where it's from – your own files, your partner's files, even commercially-published offerings.

52.1 Folio Integrator

http://www.nextpage.com/products/folioproducts/folio_integrator.html

Folio Integrator 4 is a software developer's kit providing powerful technology for using and managing electronic reference information. Folio's search functionality makes mission-critical information instantly available to end-users. Search types include natural language, relevance ranking, wild card searches, term-proximity, field searches, and fast phrase searching.

Demo: More information requested

Cost: More information requested

53 CHASS

CHASS is a computing facility within the Faculty of Arts & Science, University of Toronto. It was created on May 1, 1996 by merging two computing centres: the EPAS Computing Facility and the Centre for Computing in the Humanities, which - for the previous decade - have been separately serving social sciences and humanities disciplines at this University. The Facility's broad mandate is to promote computing in research and teaching within the humanities and social sciences departments.

53.1 Text Analysis Computing Tools

<http://www.chass.utoronto.ca/cch/tact.html>

TACT (Text Analysis Computing Tools), a system of 15 programs for MS-DOS, is designed to do text-retrieval and analysis on literary works. Typically, researchers use TACT to retrieve occurrences of a word, word pattern, or word combination. Output takes the form of a concordance, a list, or a table. Programs also can do simple kinds of analysis, such as sorted frequencies of

Program is free downloadable.

Cost: May be freely copied by anyone for research.

54 Zuma

Research in text analysis at ZUMA involves the broadening of computer-assisted content analysis by means of integrating linguistic theories and computational linguistic methods. The aim is to define and test such procedures which the different analysis requirements support and integrate them in text analysis for the social sciences

54.1 Textpack

<http://zuma-mannheim.de/software/en/textpack/index.htm>

The majority of the procedures in TEXTPACK assist the user in exploring and editing texts; two procedures are designed for purely quantitative analysis. In the area of computer aided content analysis, TEXTPACK is suitable both for analysis with a strict dictionary approach and for empirical approaches. In addition, and as a major advantage, TEXTPACK contains special-purpose procedures which allow you to categorize/classify/tag any kind of text according to so called "content analytic dictionaries". The process of text classification is the goal of TEXTPACK and, thus, all other procedures function as service procedures which help to browse, to print a text file, to develop and validate a dictionary, and, most importantly, to help you to use the final numeric output, for example frequencies of categories, with statistical packages like SPSS or SAS for further quantitative or logical analyses. Of course, TEXTPACK can also be used simply for word frequencies, concordances, vocabulary comparison - without the goal of automatically categorizing the text.

Demo: A demo version of Textpack is downloadable on the site.

Cost: non-profit users commercial users single user copy DM 200.00 DM 1000.00
site license DM 1000.00 DM 5000.00

55 Oxford University Press

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford University Press (OUP) is the world's largest university press, publishing over 4,000 new titles a year. It has become familiar to millions through a diverse publishing programme that includes scholarly works in all academic disciplines, bibles, music, school and college textbooks, children's books, materials for teaching English as a foreign language, business books, dictionaries and reference books, and journals.

55.1 WordSmith Tools

<http://www.liv.ac.uk/~simms2928/index.htm>

WordSmith Tools is an integrated suite of programs for looking at how words behave in texts. You will be able to use tools to find out how words are used in your own texts, or those of others. The **Wordlist** tool lets you see a list of all the words or word-clusters in a text, set out in alphabetical or frequency order. The concordancer, **Concord**, gives you a chance to see any word or phrase in context – so that you can see what sort of company it keeps. With **KeyWords** you can find the key words in a text. The tools are used by Oxford University Press for their own lexicographic work in preparing dictionaries, by language teachers and students, and by researchers investigating language patterns in lots of different languages in many countries world-wide.

Demo: Program free available

Cost: Tools have not been tested and approved by Oxford University Press, and that you therefore download from here on an as-is basis, at your own risk

56 askSam Systems

Since 1985 askSam Systems has been developing some of the worlds most popular text- database and publishing tools.

56.1 askSam

<http://www.asksam.com/>

askSam is a different kind of database - a database for information. In a world where ever- increasing amounts of information are easily accessible, you need a tool flexible enough to manage this information.

askSam is a tool to organize the information that crosses your desk. It doesn't matter whether you deal with sticky notes, memos, phone messages, faxes, resumes, depositions, newspaper articles, or e-mail.

All contain information that you need to track. With askSam we provide a solution. Simply type, scan, or import your information into askSam. Straight away you can search and retrieve any word or phrase in the text. Finding your information is easy. askSam offers the widest variety of searches found in any database program. These searches include full-text, fuzzy, wildcard, Boolean, proximity, date and numeric.

Demo: A free demo is downloadable.

Cost: \$ 385.00

57 University of Nijmegen

57.1 AGFL system

<http://www.cs.kun.nl/agfl/>

The AGFL system is a collection of software systems for Natural Language Processing, based on the AGFL formalism. The AGFL formalism has been developed by the Computing Science Institute at the University of Nijmegen. It is a formalism in which natural languages can be described by means of two-level grammars: a context free grammar is augmented with a feature level for expressing agreement between syntactic categories. The formalism is suited for specifying both morphological and syntactic analysis. Grammar rules can be extended with transduction parts, which specify the output. The default output consists of parse trees. Furthermore, mechanisms are provided to handle open classes of words, which enables the construction of robust parsers.

Demo: free

Cost: free

58 Mitre

In partnership with government clients, MITRE is a not-for-profit corporation working in the public interest. It addresses issues of critical national importance, combining systems engineering and information technology to develop innovative solutions that make a difference.

58.1 Alembic

<http://www.mitre.org/technology/alembic-workbench/>

The Alembic Workbench project has as its goal the creation of a natural language engineering environment for the development of tagged corpora. To enhance this process, the workbench incorporates a suite of tools for the analysis of a corpus, along with the Alembic system to enable the automatic acquisition of domain-specific tagging heuristics. The Alembic system, like others, makes use of specialized heuristics that recognize and extract the relevant facts from texts. What is different and important about the Alembic system is its ability to mix hand-coded heuristic rules with machine-learned heuristic rules.

Cost: Available without charge to use for internal purposes only

59 Centre for Spoken Language Understanding (CSLU)

CSLU is an Oregon Graduate Institute of Science and Technology Research Center that focuses on Spoken Language Technologies. These technologies are used for helping people with visual, speech, or learning problems, for dictating letters directly into the computer, for accessing the internet over the phone, for using voice commands to control devices like your mobile phone, for education, and for many other applications.

59.1 CSLU Toolkit

<http://cslu.cse.ogi.edu/tools.htm>

The CSLU Toolkit was created to provide the basic framework and tools for people to build, investigate and use interactive language systems. These systems incorporate leading-edge speech recognition, natural language understanding, speech synthesis and facial animation technologies. The toolkit provides a comprehensive, powerful and flexible environment for building interactive language systems that use these technologies, and for conducting research to improve them.

Demo: Beta version available for download.

Cost: Free for noncommercial use.

60 Cogilex

Cogilex is a team of computational linguists and knowledge engineers with a long experience in the research and development of powerful NLP techniques.

60.1 QuickTag and QuickParse

<http://www.cogilex.com/products.htm>

Quicktag is a COM component for Win32 that can efficiently tag (identify the possible grammatical categories of words), lemmatize (identify the root form of words), disambiguate (indicate the actual grammatical category of words) and extract noun phrases from English text. QuickParse is used internally by Cogilex to efficiently parse english text (i.e. identify grammatical role of words and grammatical links between words) and extract semantic content of different types. Cogilex can customize a version of QuickParse for your specific needs or can develop a complete solution for you.

Demo: Online demo available on the site, and downloadable.

Cost: The basic price for QuickTag 2.0 is US\$ 1500. The cost of a royalty-free perpetual distribution license for QuickTag is US\$ 25000. The cost for obtaining the data files for QuickTag is US\$ 30000.

61 Basis Systeme netzwerk (BSn)

BSn is a tiny R& D system house in Munich/Germany. Despite its size BSn is a full service, one-stop, source of "state-of-the-art" Open Information Management solutions. Our customers are Government Agencies, Research Institutes, Management Consultants and Financial Institutions. BSn was founded in 1992 with the aim of developing and implementing Wide Area Information Management Solutions "As Easy to Use As The Telephone" to help manage the vast resources of information stored in corporations, government bodies, libraries and other institutions.

61.1 InterBasis

<http://www.bsn.com/Z39.50/WEBCAT.html>

BSn's interBasis is an emerging family of multi-platform, multi-protocol, information management products for Inter/Intranet applications. Z39.50 Target Features: Init, Search, Present, Delete, Scan, Access Control, fully configurable attribute sets, GRS and advanced search. Explain is planned. Access and database maintenance is available via a WWW interface. Data is converted "live" via object oriented technology to any of a number of presentation formats, including HTML and XML for use in the World Wide Web. The interBasis server includes a high-speed structured full-text retrieval engine. It supports a wide range of native document types (including SGML, XML, HTML, MARC and numerous bibliographic, metadata and Internet standards), "on-the-fly" format conversion, synthetic meta-records, advanced search, database segmentation, concurrent update (Workflow capable), national language and international character set support, access control and a WWW interface. A CORBA interface to integrate legacy and RDBMS services is in development. Support for IPv6 is currently being tested.

Demo: Downloadable

Cost: ; \$ 1000 see: <http://www.bsn.com/Z39.50/Products/prices.html>

62 Pacific Software Publishing

PSPINC has been developing the system and service over 5,000 web sites at its facility in Washington State. Since 1996, PSPINC has focused on developing complete corporate Internet outsourcing services based on Windows Server platform. From the email server outsourcing to complete electronic commerce deployment, PSPINC is your one stop solution provider.

62.1 Alise

<http://www.pspinc.com/ntsg/products/alise/index.htm>

Developed for C/C++ and Visual Basic programmers, ALISE is one of the fastest index keyword search engines today. Running on Windows NT Server, it can perform "and" and "or" operant search out of upto 500,000 database in less than a blink of an eye. ALISE is also one of the smallest program you find anywhere. The program fits in a single floppy disk. As a matter of fact, the install program takes up more space on the diskette than the program itself.

Demo: A 60-day evaluation version is available on request.

Cost: \$ 1900

63 AnswerLogic

AnswerLogic was founded on the belief that computers should be made to adapt to people, not the other way around. Advances in the fields of technology and natural language processing now make it possible to radically change the way people interact with computers. AnswerLogic was founded in 1998 to build on these advances. Based in Washington, DC, AnswerLogic employs a bright and growing staff to develop innovative question-answering solutions for businesses.

63.1 AE1

http://answerlogic.com/sol_poverview.html

The AnswerLogic AE1TM is the world's first answer engine. The AE1TM uses natural language processing techniques to interpret the contents of documents and other available sources, as well as to interpret the questions that users ask. At the core of the AE1TM lies the Language Processor, which analyzes text in plain English. After identifying parts of speech, word relationships and the meanings of words in context, the Language Processor represents a text as a series of related facts. It stores these as concepts in the IdeaMap. The AE1TM's Language Processor works by drawing from an extensive linguistic knowledge base, or core lexicon. It also draws from a domain lexicon containing industry-specific information and a customer lexicon created for a specific business. Thus, product names and trade marked terms may be understood, as can industry-speak and everyday jargon. To facilitate the input of text through the AE1TM, there is a Document Manager which handles all functions related to source documents, including the identification of the document sets from which answers will be retrieved. The Answer Manager performs a similar function by passing questions to the Language Processor for in-context interpretation. To get an answer, the Language Processor compares the conceptual representation of a question against the concepts stored in the IdeaMap, retrieving only those concepts that closely match.

Demo: More information requested

Cost: \$ 250,000 - \$ 400,000

64 LexiQuest

LexiQuest is a software solution for accurate information retrieval and sharing knowledge from corporate databases and the Internet, using advanced computational linguistics and natural language processing (NLP). That means LexiQuest actually uses the power of our spoken and written language — the words we use every day — to define and find the right information. But LexiQuest isn't a search engine; it's an application that dramatically boosts the accuracy of information retrieval using search engines, overcoming their frustrating limits. Search engines require traditional and often inaccurate key word searches and use statistical methodology. LexiQuest, remember, uses language and understands the concepts of what we want when we use language. Long before most people had ever heard of the Internet, LexiQuest was born in Paris under the original name Erli SA, and today is the outcome of two decades of landmark research into the technology of computational linguistics and natural language processing.

64.1 LexiQuest

<http://www.lexiquest.com/>

LexiGuide Expert is a natural language processing and computational linguistic-driven software solution for sophisticated and expert research on the Intranets of specialized industry sectors, such as information technology, consulting, financial services and medicine. As in other fields, finding and retrieving the right information — and not wasting time getting the wrong information — is essential. LexiGuide Expert eliminates documents that have the right words but the wrong meaning. Best of all, it can be added to existing information systems without having to re-tool. LexiGuide Expert is designed to drill deep for sophisticated expert research when what you want is precise information. By automatically using contextual and grammatical parsing, word sense disambiguation and semantic networks, LexiGuide Expert makes obsolete the task of writing complex Boolean queries.

Demo: More information requested

Cost: More information requested

65 University of Maryland

The University of Maryland Institute for Advanced Computer Studies (UMIACS) is a research unit whose mission is to foster interdisciplinary research in computing at the College Park campus. The Institute's research programs are led by distinguished faculty with joint appointments in the departments of Computer Science (32), Electrical Engineering (7), Geography (2), Linguistics (2), Business and Management (1), Philosophy (2), and Education (1). Major research activities are conducted through the Center of Automation Research (CfAR), and a number of state-of-the-art laboratories including the Laboratory for Parallel Computation, the Human-Computer Interaction Laboratory (HCIL), Language and Media Processing (LAMP), and the Multimedia Systems Laboratory. The Institute's environment is enriched by a strong outreach program that includes partnerships with industry and government laboratories on focused research projects and the organization of seminars and workshops in emerging technology areas

65.1 Translingual Information Retrieval System (TIDES)

<http://www.umiacs.umd.edu/labs/CLIP/tides/>

We propose to develop and demonstrate an integrated set of techniques for building rapidly retargetable systems to support translingual access to character-coded electronic texts, images of printed documents, and recorded news broadcasts. Our techniques will enable the creation of working systems for previously unforeseen language requirements within one to two weeks using readily available resources. We will demonstrate a useful capability to search document images in a new language by analysts that are not familiar with that language based on a limited quantity of print or electronic resources and the use of a language specialist for a brief time. We also plan to demonstrate a capability to significantly increase the productivity of analysts who are familiar with the language of interest when searching recorded news broadcasts based on the above resources and transcription of a limited quantity of similar news broadcasts. The techniques we develop will be extensible, making it possible to exploit richer resources in languages for which they are available. Our approach is highly leveraged, tightly coupling document image processing with lexical acquisition and lexical acquisition with speech recognition to achieve effective translingual information access with a minimal reliance on external resources or the developer's time during retargeting to a new language. Our view of translingual information access is inherently interactive, combining advanced interfaces for translingual query formulation and translingual document selection with recently discovered translingual search techniques.

Demo: Available on the website

Cost: This is a basic research project, no product.

66 The Features

	package:	Searching	Parsing	keywords	Taxonomy	Structure	MultiLanguage	API	source	Database	Price	Comment
1	TNO, Quetzal	+	-	-	-	-	+	+	-	-	free	TNO's search engine.
2	netowl extractor	+	-	+	0	+	+	+	-	+	\$70,000	Extracts names and phrases.
3	textanalyst	+	?	+	nf	-	0	+	-	-	\$1782	Toolkit for textmining.
4	megasearch	+	-	-	-	-	+	?	?	-	\$77.7	Searching in Natural language.
5	intelligent miner for text	+	?	+	nf	-	+	+	-	-	\$30,000	Toolkit for textmining.
6	icrossreader	+	0	0	f	-	+	0	-	+	\$149	Text Mining and Intelligent Search Engine.
7	yahoo planet	+	?	+	f	?	?	?	?	?	?	Classifies documents to Yahoo hierarchy.
8	dataset	+	?	?	?	?	?	?	?	+	\$49	Database with text search and retrieval tools.
9	concept map	?	?	?	?	?	?	?	?	?	?	site down.
10	thescap	+	-	+	-	-	0	0	-	-	>\$20,000	Graphical ConceptMap creator.
11	readware, ipserver, SDK	+	-	+	nf	+	+	+	-	+	> 42.304,29	intelligent text analysis server.
12	twurl	+	-	-	+	+	+	0	-	0	?	Internet searcher.
13	taxonomy engine	+	-	+	nf	0	+	+	-	+	>Dfl.10.000	Text classifier.
14	smisc	-	-	?	?	?	?	?	?	?	?	Ready in the fall. Discourse mapping.
15	textcenter, textfinder	+	+	-	-	-	+	+	-	-	\$120,000	Hardware implementation with textmining software.
16	zymage	+	-	-	-	-	+	+	-	-	Dfl. 10.000	index, fast (fuzzy and proximity) searching of docs.
17	knowledge server	+	-	+	nf	+	+	+	-	+	F10K-£3.000K	pattern-matching, contextual analysis and concept extraction.
18	ie studio, text-o-scope	+	0	+	nf	+	-	+	-	+	?	Information Extraction Studio.
19	verity developer kit	+	?	?	nf	+	+	+	?	+	?	Provides unparalleled high-speed search precision.
20	inquiry	+	0	+	nf	+	-	+	-	-	> £17,688	probabilistic text retrieval system.
21	excalibur retrievalware	+	-	0	nf	+	+	+	-	+	> eur. 25K	Advanced Search and Retrieval.
22	basis	+	-	+	nf	+	+	+	-	+	>Dfl76.000	search and retrieval solution.
23	concordance	+	-	-	nf	-	-	+	-	+	\$995 - \$5000	full text database management package.
24	dtsearch	+	-	-	f	+	+	+	-	+	\$999	fast indexing and searching.
25	proindex	+	-	-	-	-	+	+	-	+	\$3K - \$12K	toolkit for full-text searching.
26	multi-site search software	+	-	-	-	-	-	-	-	-	\$750-\$3,850	Internet indexing and searching.
27	callable personal librarian	+	?	?	nf	+	+	+	+	+	free	tools for searching and multimedia applications.
28	texis	+	-	+	nf	+	+	+	-	+	> \$ 10,000	Text Retrieval RDBMS.
29	texcovery	-	-	+	nf	-	-	-	-	-	?	Prototype text categorization with 3D visualization.
30	vantagepoint	+	0	+	+	+	-	-	-	+	\$6,500	extract knowledge from text databases.
31	fulcrum searchserver	+	0	+	nf	+	+	+	-	+	\$ 7,800	advanced information retrieval solution.
32	searchkey	+	-	-	-	+	+	+	+	+	\$10,000	Enterprise wide indexing and searching.
33	seekit developer	+	-	-	-	+	+	+	+	+	> \$9,999	programming lib for developing text search software.
34	searchopia	+	-	-	-	-	-	+	-	-	\$10	Library for text searching in files.
35	tfind, tstore, tmining,t-repository	+	-	+	nf	+	+	+	-	-	Dfl.100.000 a year	advanced text mining and automatic classification system.
36	thoughttreasure	-	+	+	nf	-	0	+	+	-	free	Database of 25,000 concepts with NLP tools.
37	bow library	+	-	-	-	-	+	+	+	-	free	Lib for stat text analysis, language modeling, IR.
38	childes	+	+	-	-	-	+	+	+	-	free	tools for studying conversational interactions.
39	Xerox, askonce	+	-	?	?	?	?	?	?	?	?	web-based meta-search application.
40	linguistx, categorizier	+	+	+	nf	+	+	+	-	+	> \$150K	fast, comprehensive multilingual text services.
41	ht://dig	+	-	-	-	0	+	+	+	+	free	www indexing and searching system for intranet.
42	isearch	+	-	-	-	+	-	+	+	-	free	indexing and searching text documents.
43	amberfish	+	-	-	-	+	-	+	-	-	free	indexing and searching text documents.
44	textpress	+	-	-	-	-	+	+	-	-	> \$4,000	DB of relational data, free text and multimedia.
45	zebra	+	-	-	-	-	0	+	+	0	free	fielded free-text indexing and retrieval engine.
46	z'mbol	+	-	-	-	+	0	0	-	0	£32.000	high performance, information retrieval tool.
47	searchexpress	+	-	-	-	+	-	+	-	+	\$995	powerful full-text searching.
48	isis	+	-	-	-	-	+	+	-	+	free	set of textual database softwares.
49	boolean search 2	-	-	-	-	-	-	-	-	-	\$247	Only on the Macintosh.
50	dioweb search	+	+	+	f	+	+	-	-	+	>\$495	true natural language search tool.
51	onix text toolkit	+	-	-	-	+	+	+	0	0	\$2,000	C++ code library for indexing and searching.
52	harvest	+	?	?	?	?	?	+	+	?	free	tools for information retrieval.
53	Oracle, intermedia	+	-	+	nf	+	+	+	-	+	?	extends Oracle8i by indexing any text or documents.
54	protan	+	0	0	nf	-	+	-	+	-	free	content-analytic engine.
55	visualext	+	+	+	nf	+	+	+	-	-	\$50K-\$100K	IDE for developing text analyzers.
56	folio integrator	+	?	?	?	?	?	+	?	?	?	devkit for using and managing reference information.
57	TACT	+	-	+	-	+	+	0	-	-	free	text-retrieval and analysis on literary works.
58	textpack	+	-	+	-	-	+	-	-	-	>DM 200	computer-assisted content analysis.
59	wordsmith tools	+	-	-	-	+	+	-	-	-	£50	tools for looking at how words behave in texts.
60	asksam	+	?	?	?	?	?	?	?	+	\$385.00	flexible database for textual information.
61	agfl system	-	+	-	-	+	+	+	+	-	free	software systems for Natural Language Processing.
62	alembic	+	?	+	?	+	?	?	?	0	free	natural language engineering environment.
63	cslu toolkit	?	?	?	?	?	?	?	?	?	free	tools to build, interactive language systems.
64	quicktag and quickparse	-	+	+	-	-	-	+	+	-	>\$1500	tools for tagging and parsing.
65	interbasis	+	-	0	nf	+	+	+	-	+	> \$1000	multi-platform, information management.
66	alise	+	-	-	-	-	+	+	-	0	\$1,999	one of the fastest index keyword search engines.
67	acl	+	+	+	nf	-	0	+	-	-	\$250K-\$400K	interpret contents of documents as questions of users.
68	LexiQuest	+	?	?	?	?	?	?	?	?	?	accurate information retrieval using CL and NLP.
69	Tides	-	-	-	-	-	-	-	-	-	-	This is a basic research project, no product.

67 Some general Links

For an overview of software:

<http://www.linguistlist.org/software.html>

<http://registry.dfki.de/>

See <http://allen.comm.virginia.edu/jtl15t/index.htm>

For evaluation of some products and links to products.

Also Rank Xerox, has created some tools at:

<http://www.xrce.xerox.com/research/mltt/toolhome.html>

The Katholic University Brabant has some tools: <http://ilk.kub.nl/>

ISIS reports

This report is in the series of ISIS technical reports. The series editor is Rein van den Boomgaard (reports-isis@science.uva.nl). Within this series the following titles are available:

References

- [1] J.M. Geusebroek, F. Cornelissen, A.W.M. Smeulders, and H. Geerts. Robust autofocusing in microscopy. Technical Report 17, Intelligent Sensory Information Systems Group, University of Amsterdam, 2000.
- [2] J.M. Geusebroek, R. van den Boomgaard, A.W.M. Smeulders, and A. Dev. Color and scale: The spatial structure of color images. Technical Report 18, Intelligent Sensory Information Systems Group, University of Amsterdam, 2000.
- [3] J.M. Geusebroek. A physical basis for color constancy. Technical Report 19, Intelligent Sensory Information Systems Group, University of Amsterdam, 2000.
- [4] J.M. Geusebroek, A.W.M. Smeulders, and R. van den Boomgaard. Color invariance. Technical Report 20, Intelligent Sensory Information Systems Group, University of Amsterdam, 2000.
- [5] L. Todoran and M. Worring. Segmentation of color document images. Technical Report 21, Intelligent Sensory Information Systems Group, University of Amsterdam, 2000.
- [6] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content based image retrieval at the end of the early years. Technical Report 22, Intelligent Sensory Information Systems Group, University of Amsterdam, 2000.
- [7] C.B.J. Bergsma, G.J. Streekstra, A.W.M. Smeulders, and E.M.M. Manders. Velocity estimation of spots in 3d confocal image sequences of living cells. Technical Report 23, Intelligent Sensory Information Systems Group, University of Amsterdam, 2000.
- [8] J.C. van Gemert. Practical tutorial for using corba. Technical Report 24, Intelligent Sensory Information Systems Group, University of Amsterdam, 2000.
- [9] J.C. van Gemert. Overview of natural language processing tools. Technical Report 25, Intelligent Sensory Information Systems Group, University of Amsterdam, 2000.

You may order copies of the ISIS technical reports from the corresponding author or the series editor. Most of the reports can also be found on the web pages of the ISIS group (<http://www.science.uva.nl/research/isis>).



Intelligent Sensory Information Systems
University of Amsterdam
The Netherlands

