



# The Future of Web Searching

**Annette van der Merwe**

Postgraduate Diploma in Information Management

Rand Afrikaans University

[annettevdm@yahoo.com](mailto:annettevdm@yahoo.com)

---

## Contents

[Introduction](#)

[Terminology](#)

[How do search engines work?](#)

[Multiple search engines](#)

[Search engine shortcomings](#)

[Emerging trends](#)

[Conclusion](#)

[References](#)

---

## Introduction

Search engines gather and organize information to match keywords or phrases provided by the user. These matches are then presented to the user in the form of lists, ranked according to what the search engine believes is most relevant. This is helpful and often very useful. However, as the flow of information on the Internet increases, so does 'noise' in the system and 'the truth' becomes more and more elusive. The user is often presented with either far too much information and suffers from information overload, or the information that is presented by the search engine, is often less useful or irrelevant.

It is necessary to clarify why these search engines provide too much information or information that is less useful: is it because search engines don't always understand the request or is it because they cannot sort the available 'matches' to be as specific as possible? Or is it the Web that is not 'search-engine-friendly' enough? Or perhaps it is the user not being specific or clear enough about the request? The answer is not a simple one.

Three components, namely the user, the Internet and search engines should be looked at to determine their influence on the future of Web searching. Search engines only deal with content, not the context or meaning of the query. Search engines retrieve blindly, without making any judgement about the relevance of the documents or the quality of links that the document contains. The Internet, on the other hand, is not indexed in such a way that search engines can read and retrieve everything that is available. Lastly, the user considers the Net both an information source and a unique means for connectedness and communication. It is therefore used for different reasons. The tools we need in future to assimilate information into useful knowledge, will have to reflect all these requirements.

Both user requirements and technological development exert influence in shaping the future of information retrieval on the Web. For search engines to make judgements, choices and recommendations, they have to become like humans.

They presently stand ready with power and speed to serve us, but they lack intelligence. Until computers can comprehend language and hold their own in a conversation, there will be a gap in their capabilities. A lack of structure surrounding the Internet, however, exacerbates the problem. Human powered indexing methods, that is the classification and description of documents by topical experts, is perhaps needed.

This research suggests that a combination of human and machine qualities might be the answer in providing an effective search tool for the future.

---

[\\_top](#)

## **Terminology**

### **1. The Internet and World Wide Web (WWW)**

The Internet (referred to hereafter as the Net) is an information system composed of a massive network of computers around the world (Griffin, 1999:1). An easy way to visualize the Net is to view it simply as a permanent physical connection of tens of thousands of computer networks scattered around the world. Every individual computer on any of these networks can, if they have permission, communicate with any other computer. This ability to communicate allows all the connected computers to share a vast resource of information. This information may have an academic, commercial or general interest content and be available for unrestricted (free) viewing, or it could involve the transfer of files between computers in a more controlled way (access by paid subscription).

There are many different components to the Net. The World Wide Web (WWW), referred to hereafter as the Web, often confused with the Net itself, is actually just another protocol for navigating the Net (Leigh & Kelmer, 1996:10). The Web delivers documents in hypertext format ([http](http://): hypertext transfer protocol) and is indeed not just a hypertext system but a hypermedia system allowing text, sound, graphics and video to be mixed together in a multimedia format. Most Web pages are currently produced in hypertext mark-up language (HTML).

### **2. Search engines**

A search engine 'is a computer program that searches for documents containing keywords or phrases of interest to the user' (Peterson, 1997:2). A search engine acts as an information robot or 'info-bot', a sort of obedient servant finding dozens or even thousands of documents quickly.

Search engines have three major elements:

The first part is the spider, also called the crawler (Kapoor, 1999:2), indexing robot (Lynch, 1997:2) or robot/bot (Chamberlain, 2000:1) as well as worm or search bot (Hermans, 1997:2) and harvester (Taylor, 1999:2). The spider visits every site (each site being a set of documents, called pages) it can identify on the Web, examines these pages and extracts indexing information that can be used to describe them. This is what it means when someone refers to a site as being 'spidered' or 'crawled'.

Everything the spider finds goes into the second part of a search engine, the index,

sometimes called the catalog. This process, which varies among search engines, may include simply locating most of the words that appear in Web pages or performing sophisticated analyses to identify key words and phrases (Lynch, 1997:2). This data is stored in the search engine's database, along with an address, termed a URL (uniform resource locator).

The third part of a search engine is the search engine software. This is the program that sifts through the millions of pages recorded in the index to find matches to a search and to rank them in order of what it believes is most relevant.

Chamberlain's (2000:1) definition of search engines as 'huge databases of Web page files that have been assembled automatically by machine' refers mainly to the index/catalog part of the search engine. Kapoor (1999: 2) however, defines a search engine as 'an HTML interface to a database of locations of information'.

There are different methods in which search engines classify/index information. Kapoor (1999:2) provides a distinction between search engines, directories and hybrid search engines.

- Search engines create their listings automatically. Search engines crawl the Web, then people search through what they have found.
- A directory depends on humans for its listings. You submit a short description to the directory for your entire site, or editors write one for the sites they review. A search engine will look for matches only in the descriptions as submitted.
- Hybrid search engines maintain an associated directory. Inclusion into the directory is however not automatic, but according to Kapoor, 'a combination of luck and quality'.

Peterson (1997:2) provides the following categories for search engines:

- robotic search engines use a Web robot to retrieve a significant number of documents;
- mega-indexes, also known as meta-indexes, do not have their own databases; instead they are linked to robotic search engines;
- simultaneous (parallel) mega-indexes, also known as multi-threaded meta-indexes; these are mega-indexes which access robotic search engines in parallel (simultaneously) and present the unified results as a single package;
- subject directories; these are manually maintained, browsable and are often searchable with robotic search engines; and
- robotic specialized search engines; these are search engines which focus on a portion of the Net.

---

[\\_top](#)

## **How do search engines work?**

Search engines gather and organize (classify/index) descriptions and locations of information available on the Net and store this information in a catalog/index, thus creating their own database of information. Whenever one searches the Web using a search engine, one asks the engine to scan its index/catalog, to sift through the recorded pages and to provide matches to keywords and phrases it is given. The query produces a list of Web resources (URLs) that can be clicked on to connect to the sites identified by the search.

Search engines use selected software programs to search their indexes, presenting the findings in some kind of ranking according to what it believes is most relevant. Although software programs may be similar, no two search engines are exactly the same in terms of size, speed and content. No two search engines use exactly the same ranking scheme and not

every search engine offers one the same search options.

Chamberlain (2000:2) says it is important to remember that when we are using a search engine, we are not searching the entire Web as it exists, but only a portion of the Web, captured in a fixed index that has been created at an earlier date. It is, she says, difficult to say how much earlier the index had been captured. Spiders regularly return to the Web pages they index to look for changes. When changes occur, the index is updated to reflect the new information, but the process of updating can take a while, depending how often the spiders do their crawling and also how promptly the information they gather, is added to the index. Until a page has been both 'spidered' and 'indexed' the new information cannot be accessed.

---

[\\_top](#)

## **Multiple search engines**

There are two broad types of search engines: individual and multiple search engines. Individual search engines compile their own searchable databases, as described above. Examples of individual search engines are AltaVista, HotBot, Excite, Infoseek and the latest, Northern Light.

Multiple search engines, also known as mega or parallel search engines (Notess, 1998:1) as well as meta (Chamberlain, 2000:1) search engines, do not crawl the Web compiling their own databases. Instead they search the databases of multiple sets of individual search engines simultaneously and then presents results. The results of these searches are presented in two ways: some display results in a single merged list, from which duplicate entries have been removed, while some do not collate results, but display them in separate (multiple) lists as they are received from each engine. Duplicate entries may thus appear.

Multiple search engines do not offer the same selection of search options as individual search engines will do and do not return all the results retrieved from the individual engines they search. However, the results they do return are drawn from the top of the search engine list, and tend to be more relevant. Multiple searchers are fast and therefore very useful when the searcher is in a hurry or wants to obtain a quick overview on a subject and/or unique term. Multiple search engines do offer timesaving features for those doing an exhaustive search on obscure topics. Because multiple search engines cannot take full advantage of the unique features of the individual search engines, the quality of results is in a sense as strong as the weakest link offered (Hubbard, 2000:4).

Inference Find, Dogpile and MetaFind are three mega search engines that offer some advantages for comprehensive searching, even while suffering from some of the problems mentioned above.

---

[\\_top](#)

## **Search engine shortcomings**

The Net, also referred to as 'the information highway', enables all types of users to access and share information with one another. Users are therefore exposed to large volumes of dynamic information, retrieved from thousands of computers spread all over the world. However, as the flow of information increases, so does 'noise' in the system, and 'truth' becomes ever more elusive. Users are beginning to suffer from severe information overload. Paul Saffo, director with the Institute for the Future in Menlo Park, California says 'information overload is not a function of the volume of information out there, it is a gap between the volume of information and the tools we have to assimilate that information into useful

knowledge' (Foley, 1995:1).

### 1. **Restricted intelligence**

Search engines and services like Yahoo! and Alta Vista are currently embedded in Netscape, MS Explorer and AOL and aid us in keyword searching. These engines are Boolean word matching engines and only deal with content, not the context or meaning of the query. Search engines therefore blindly retrieve documents available on the Net, without making any judgement about the relevance of the documents or the quality of links that they contain (Kapoor, 1999:1). In some cases they do return a list that is ranked according to relevance, but only relevance according to the keywords it was given, for example Northern Light. Many of these engines offer no filtering function or intelligent search strategy and return hundreds and sometimes thousands of responses to a query, thus adding to the problem of information overload.

Information on the Net is dynamic. Often search engines refer to information that has moved to an unknown location or has disappeared. They do not learn from these searches and will look for these locations every time it is asked. Search engines allow a user to search for information in many different ways, but seldom 'remembers' what was done previously. It does not, for example store any query results, so the same query will be repeated over and over again starting from scratch.

Taylor (1999:3) provides the following summary of why search engines are not always the solution to finding relevant information at present:

- Relevant information can be missed because sites contain types of resource in addition to html text, for example images, databases, PDF documents.
- The search engines frequently do not harvest every page on a site, but often only the top two or three hierarchical levels, thus missing significant documents which, on larger and more complex sites, may be located in lower levels of the hierarchy.
- Search engines, especially the more comprehensive ones, may index sites on an infrequent basis and may therefore not contain the most current data.
- Irrelevant information can be retrieved because the search engine has no (or very few) means of distinguishing between important and incidental words in the document text.

Some intelligence is thus required from these engines in order to assist in finding only that information which is useful and relevant: the right information available to the right people at the right time.

### 2. **Automated indexing**

The Web lacks standards to facilitate automated indexing. There is virtually no bibliographic control on the Net (Hubbard, 2000:2). As a result, documents are not structured in such a way that programs can reliably extract routine information that a human indexer might find through a cursory inspection of author, date of publication, length of text or subject matter. In contrast to human indexers, automated programs have difficulty identifying characteristics of a document such as its overall theme or genre – whether it is a poem or a play or even an advertisement. The professional indexer can describe the components of individual pages of all sorts (from video to text) and can clarify how those parts fit together into a database of information. Analyses of a site's purpose, history and policies are beyond the capabilities of a crawler program.

Another drawback of automated indexing is that most search engines recognize text only although there exists an intense interest in the Web's ability to display images. Some research

has however moved forward to find colour or patterns within images, but no program can deduce the underlying meaning and cultural significance of an image.

Automated spiders that crawl sites on the Web can only read open text formats, such as HTML files and cannot record any more than the basic file attributes of non-text format files, including PDF, sound, image and video files. Furthermore, most search engines cannot survey frame-based sites or dynamic pages and have problems with pages in XML format (Hubbard, 2000:5).

There are insurmountable difficulties in unleashing computers to comprehend language. Language is replete with, for example, synonyms and spelling variations and is full of variable contextual meanings and linguistic nuances, making full-text databases rather blunt tools in their over reaching attempts to process natural language (Hubbard, 2000:3).

The great promise of automated indexing tools is that they provide a level of detail greater than any humanly powered method of indexing. Automated searching aids are necessary to keep up with the millions of pages being added to the Net daily. Finding dead links and maintaining information on the Net are needs that can be met, very effectively, by existing search tools. When it comes to sorting through all of this data, however, and making best sense of what is available online, the power and assurance of human understanding and editorial control must be called upon (Hubbard, 2000:6).

---

[\\_top](#)

## **Emerging trends**

The Web has become the largest, most complex search space ever. As the Web continues to grow, search tools must evolve and adapt to remain useful. Both user requirements and technological developments will exert influence in shaping the future of information retrieval on the Net. East Stroudsburg University (1999) and the University at Albany (2000) talk about second generation search engine services on the Web. They suggest watching the coming trends with second generation services. These trends all include the human elements of:

- concept processing. Second generation services such as Ask Jeeves apply different kinds of concept processing to a search statement to determine the probable intent of the search. This is often accomplished by the use of human generated indexes. With these services, the burden of coming up with precise or extensive terminology is shifted from the user to the engine. These services are therefore taking on the role of thesauri;
- collective judgement. Search services such as Google and Direct Hit derive their results from the behaviour of millions of Web users; and
- directories. First generation search services have partnered with second generation services and/or include content from human gathered directories with their search results to supplement documents retrieved from the spider-indexed Web.

### **1. Agent/filter technology**

There are two main approaches for sifting through information (Omar, 1997:1): The first method puts the responsibility in the hands of the user who searches for what he/she needs. As part of this approach, the user PULLS the information, using search engines, which search by index, key words and subject. With the second method, the responsibility of providing information is shifted to special software systems on the Net. Based on predefined criteria about the information, this method PUSHES the information to the user. It can

however push information the user does not want, that is junk mail, and requires a next level of sophistication: agent/filter technology. Agent/filter technology underpin in a push-based solution allows both the sender and recipient to filter and select information. This technology provides personalized information automatically, based on the changing preferences of the user as a result of a two-way dialogue. (Omar, 1997:2). Although both approaches talk about agent/filter technology, this article focuses only on PULL strategies using search engines or 'smart spiders' (Kapoor, 1999:1).

Many vendors refer to their filters as 'agents' but, says Berst (2000:1), we should not be confused. Real agent software makes decisions and takes action, while filters merely screen out information. Harper (1997:4) agrees with this statement when he describes an intelligent agent as 'a software entity that assists people and acts on their behalf'. He provides the following attributes for intelligent agents:

- Agency; the degree of independence the agent exhibits; at least, he says, the agent must be able to operate on the Net while the user is disconnected or not in the process of interacting with the Web. Feldman (1999:3) says agency is 'basically autonomy plus a little social ability or interactivity'.
- Intelligence; the amount of learned behaviour and the degree of reasoning an agent may have; at a minimum, he says, there must be a statement of preferences or a set of rules that have been pre-defined by the user to follow, with an inference mechanism to interpret and act on these rules; higher levels of intelligence will include the ability of the agent to learn/adapt to its environment.
- Mobility; this concept refers to the ability of agents to migrate in a self-directed way from one host to another on a network in order to perform their assigned duties. Ketsavapitak (1997:5) provides the following characteristics of intelligent agents:
  - Agents are autonomous; that is, an agent has control over its own actions. Feldman (1999:2) says autonomy is the first and foremost common criterion for agents. She continues to say that autonomous agents use the knowledge of their owner's needs and interests to undertake those tasks that their owner does repeatedly.
  - Agents are goal-driven; they have a purpose and act in accordance with that purpose.
  - Agents are reactive; that is, an agent senses changes in its environment and responds timeously.
  - All agents continue to run, even when the user is gone.

Rather than retrieving documents blindly, smart spiders (Kapoor, 1999:1) will thus be able to make some judgements over the relevance of documents and the quality of links that they contain. Hermans (1997:4) speaks about intelligent software agents that will be able to search information based on context. He says these agents will deduce this context from user information or by using tools such as a thesaurus, enabling them to search on related terms, or even on concepts. Coleman (1997:3) mentions Verity as a good example of a product providing information in context. This information/text retrieval technology improves somewhat on basic search engines. Having capabilities well beyond basic search, the Verity Knowledge Suite product family addresses the key problems of retrieving and sharing knowledge: automated classification and categorization; intelligent group dissemination and information profiling and organizing content for corporate re-use. This approach, Demers (1997:1) says, increases the relevancy of returned results, but fails to search across all data types, thus limiting effective exploitation of all the information assets in the organization and contributing only partially to comprehensive and robust knowledge management.

## 2. Knowledge retrieval

Knowledge retrieval is the next wave in information search and retrieval technology. This technology uses advanced algorithms and sophisticated processes to enhance queries,

accesses all data formats and then accurately sifts the information it finds to return only the most relevant results (Demers, 1997:1).

Coleman (1997:5) mentions Excalibur as an example of a product that 'cannot only do a multi-media search but also uses the information it finds to focus the query or interact with the user to refine the query'. Excalibur is a leading developer of high performance software products for search and retrieval of knowledge assets over all media types: paper documents, text, images, video and other multimedia data types, throughout internets, LANs/WANs, extranets and the Internet (Demers, 1997:4).

The specific advantages that Excalibur brings to the problem of knowledge retrieval are two-fold: semantic networks and pattern recognition (Demers, 1997:1).

Semantic networks are built-in knowledge bases, derived from published dictionaries, thesauri and other lexical resources. The semantic network automatically identifies words and concepts that are related to the content of the query. The network even helps users to distinguish among the multiple meanings of words. The ability to paraphrase the query frees the user from having to define the words which an author might have used in discussing an idea.

Pattern recognition (also known as Adaptive Pattern recognition Processing, or APRP) provides robust 'fuzzy spelling' to recognize query terms even if hit words within a document are misspelled due to errors made by either the author or the searcher or even errors made when transcribing foreign names (Demers, 1997:2).

Pattern recognition and semantic networks can make users more productive, no matter what the nature or quality of the text they are working with. These technologies give users direct access to relevant information and minimize the time they spend filtering through irrelevant or false results. Equipped with the information they need and given navigation channels through captured organizational expertise, people work more knowledgeably, apply their individual skills and make better decisions.

The differentiating technology here is that it cannot only search all data types but also applies semantic networks and APRP technologies to filter results to provide higher levels of both accuracy and relevance to the user. In other words, it can search all types of data and do sophisticated pattern matching (intelligent search) to help identify knowledge (Coleman, 1997:5). The objective of each technology is to maximize the accuracy of search and retrieval, while minimizing the costs of system set-up and maintenance. With these unique techniques users are able to find documents that would be impossible to find otherwise.

Knowledge retrieval solutions are differentiated by two primary factors:

- Comprehensiveness to provide access to all information assets, beyond the traditional text collection; and
- accuracy to retrieve and rank the most relevant responses to best serve the user's needs for precise information, or knowledge (Demers, 1997:2).  
He continues to say that knowledge retrieval technology is distinguished by the components of the acronym ASSETS:
- Accuracy is the critical variable that determines effectiveness of a knowledge retrieval product. Accuracy is measured with two statistics: precision (what percentage of the documents retrieved is useful, or relevant to the request) and recall (how many of the useful or relevant documents in the system are found).
- Scalability is more than just fast searching. It is the ability to maintain search performance even when the demands on the system rise by orders of magnitude.

- Security and protection of corporate information assets on Intranets is critical.
- Extensibility refers to the ability to start with text on paper assets and grow to multiple media (images, videos).
- Transparency refers to the invisibility of the incorporation of knowledge retrieval to the end user. Simple screens or even microphones may serve as the way people take advantage of these helpful techniques.
- Simplicity refers to how easy this solution is by using plain English and visual queries.

### 3. Access to the Invisible Web

Despite its uniform interface and seamless linked integration, the Web is not a single coherent element. There are two distinct elements: the visible and the invisible Web. The visible Web consists of manually produced, static pages. It provides the same generic information to everyone and is therefore available for indexing to all search engines. The invisible Web consists of computer generated, dynamic pages and provides customized information according to specific requirements (Green, 2000).

Andrews (1997:1) says software developers of search engines are seeking to exploit the thorny problem of invisible Web databases that search engines cannot 'see'. The opportunity exists, he says, because Web pages that are generated dynamically via databases, are different from what are generally known as 'flat html' pages. The latter are generated, one at a time by people using authoring tools or coding by hand, and then left on a server until someone requests them. Dynamically generated Web pages do not exist as separate files, so spiders from the major search engines do not generally discern them. The problem is intensifying because of the proliferation of off-the-shelf tools to link databases to the Web, whether as whole sites or as site components. This means that proportionately less and less is available for search engines to see.

One response to this problem has been to divide the Web into vertical sections intended to appeal to specific interests. Kapoor (1999:1) predicts that there will be an explosion of vertical search sites, providing access to deep, tightly focused databases.

Another benefit to search precision (Hubbard, 2000:3) is narrowing search domains to specific subjects, accomplished by honing the scope of what information is searched, perhaps by limiting searches to certain domains or languages, or conducting specialized searches in subject oriented search engines.

Andrews (1997:2) predicts a change in how people will use the Web in future. Instead of wandering around bookmarking what looks interesting, he says, people are already activating their Internet connections with a specific goal in mind. He continues to say that databases are listed in categories, and users choose which to search based on brief descriptions, instead of searching through them all at once.

### 4. Newsgroup searching

Thomas (1998: 1) says there are two different ways to look at the Net as a tool: one is to consider it as an information source and the other is to consider it as a unique means for connectedness and communication.

As a source of information, the Net makes facts more readily available and it allows research to be done more quickly.

As a tool for greater connectedness and communication, the Net has, according to Thomas (1998:2), great potential for promoting higher order literacy. He says, the Net possesses the potential for a free exchange of ideas as a result of many connections and the corresponding communication that would not have been possible before.

Viewed in this way, he continues, the Net provides four key things that have the potential to further higher-order literacy and critical/creative/constructive thinking more dynamically than ever before:

- Great amounts of information are available quickly; the information-side of the Net is viewed here as a starting point for the use thereof, rather than as the final product/benefit of using it.
- Many minds from many different places have the opportunity to learn from each other (knowledge sharing).
- The way in which people communicate on the Net is one of the more 'pure' forms of communication, weeding out factors such as race, gender, disabilities that might hinder deep and meaningful exchange of ideas.
- It produces synergy; information, knowledge, learning, and understanding can be shared interactively and dynamically on the Net and when one is connected to the Net as a communication and connectedness tool, you are not just connected to information, but to the minds of other people.

Green (2000:8) predicts that newsgroup searching will become more important, as individuals use the Net to seek out experts to help with problems. He says there are literally thousands of newsgroups, covering different topics, and a large number of specialized newsgroup search engines have emerged as a result. He mentions Deja News and says it is probably the most widely known newsgroup search engine. It contains, he says, a directory of selected newsgroups which users can browse through or search for a particular group, topic or posting.

## 5. XML vs HTML

HTML is dead (Green, 2000:10). While HTML's ease of use fuelled its widespread adoption, it is somewhat limited in that it is primarily concerned with the layout/design of a Web page, rather than the information that actually appears on that page. Considering that the primary use of the Web is for information retrieval, this design is something of a drawback.

XML is an open technology that offers tremendous possibilities for electronic publishing, e-commerce, information retrieval and data exchange, for it consists of rules that enable anyone to create their own mark-up language. It also not only enables explicit description of Web page content, but also describes the rules for manipulating each data set contained within the information (Green, 2000:11).

However, while XML will deliver great benefits for searching, publishing and exchanging information, Green continues, these benefits will not be realised without some effort. Standards for the tags describing information in different disciplines will have to be agreed to by each industry. Web publishers will require greater sophistication than simple knowledge of HTML, graphics and a few other applications. New XML tools are needed and computer programmers and information scientists who will be able to interpret the content of the information being published, will be required. Lastly, search engines will need to learn the standard tag structure that has been agreed by each industry/interest group.

## 6. Natural language interface

As already discussed, second generation search engines will not only look at the content, but will consider the context/meaning of the terms provided. Green (2000:6) mentions two natural language search tools, both adopting different philosophies in developing their solutions:

- AskJeeves operates by matching a user's query against a database of template questions. If there is no match, then the user is presented with the nearest alternatives from the database and asked to select the most appropriate. However, artificial intelligence experts have criticized the company's natural language claims.
- The Electric Monk conducts a syntactical analysis of the query using natural algorithms. These algorithms also make use of thesauri to consider alternative related words.

## 7. A free service?

To provide quality information with the help of sophisticated search tools, requires serious resources. Research suggests that users will be charged in future: Green (2000:12) talks about 'micro-payments' for search. Second generation search tool developers will probably also focus on providing the best solution for the moment. The possibility of search outsourcing, which will be in line with current business process trends, is also an option for the future.

---

[\\_top](#)

## Conclusion

We have more information available on the Net than we can possibly digest. Search engines presently stand ready with power and speed to provide hundreds and even thousands of documents quickly and effortlessly. What we need, however, are intelligent agents providing relevant information every time. The Net, however, has to be structured and indexed to make it more 'search friendly'. What is needed is a combination of human and machine qualities to ensure an effective search tool for future use.

---

[\\_top](#)

## References

- Andrews, W. 1997. Challenge for spiders: searching invisible Web. <http://www.internetworld.com/print/1997/02/03/industry/spiders.html>
- Berst, J. 2000. Intelligent filters to the rescue! (sort of...). AnchorDesk. [http://www.zdnet.com/anchordesk/story/story\\_1035.html](http://www.zdnet.com/anchordesk/story/story_1035.html)
- Chamberlain, E. 2000. Search engines. University of South Carolina. Bare Bones 101. <http://www.sc.edu/beaufort/library/lesson1.html>
- Coleman, D. 1997. Knowledge retrieval. [http://collaborate.com/hot\\_tip/tip0697.html](http://collaborate.com/hot_tip/tip0697.html)
- Demers, M. 1997. Knowledge retrieval: the first step to knowledge management. <http://www.excalib.com/news/tradepress/kmworld.nov04.97.html>
- East Stroudsburg University. 1999. Search engine types: first and second generation. <http://www.esu.edu/library/Enginehome.htm>
- Feldman, S. 1999. Intelligent agents: a primer. Searcher, 7(9). <http://www.infoday.com/searcher/oct99/feldman+yu.htm>
- Foley, J. 1995. Managing information: infoglut. <http://www.iweek.com/551/51mtinf.htm>

- Griffin, M.G. 1999. Computers in psychology. PSY 302.  
[http://www.umsl.edu/~mgriffin/psy302/WWW\\_Definition.html](http://www.umsl.edu/~mgriffin/psy302/WWW_Definition.html)
- Green, D. 2000. The evolution of Web searching. Online information review, 2(2):124-137.  
[http://general.rau.ac.za/infosci/information/studyguide/First/unit\\_5/green.htm](http://general.rau.ac.za/infosci/information/studyguide/First/unit_5/green.htm)
- Harper, N. 1997. Intelligent agents and the Internet.  
<http://osiris.sunderland.ac.uk/cbowww/AI/TEXTS/AGENTS3/agents.htm>
- Hermans, B . 1997. Intelligent software agents on the Internet: an inventory of currently offered functionality in the information society and a prediction of (near) future developments. [http://www.broadcatch.com/agent\\_thesis/h12.htm](http://www.broadcatch.com/agent_thesis/h12.htm)
- Hubbard, J. 2000. Indexing the Internet.  
wysiwyg://47/http://www.tk421.net/essays/babel.shtml
- Kapoor, J. 1999. Web search engines. <http://yallara.cs.rmit.edu.au/~achattar/search-engines/future.htm>
- Ketsavapitak, D. 1997. Intelligent agents. <http://www.uis.edu/~ketsavap/paper.html>
- Leigh, D. & Kelmer, M. 1996. Learning on the Net. In: Forrest, E.J. (ed.). Issues in interactive communication: the impact of the new technologies on society. Florida State University. <http://www.fsu.edu/~ic-prog/issuesbook/chapter5.html>
- Lynch, C. 1997. Searching the Internet. Scientific American , March 1997.  
<http://www.sciam.com/0397issue/0397lynch.html>
- Notess, G.R. 1998. On the Net; toward more comprehensive Web searching: single searching versus mega searching. Online, March 1998.  
<http://www.onlineinc.com/onlinemag/OL1998/net3.html>
- Omar, M. 1997. How to deal with information overload on the Internet?  
<http://star.arabia.com/971127/te2.html>
- Peterson, R.E. 1997. Eight Internet search engines compared.  
[http://www.firstmonday.dk/issues/issue2\\_2/peterson/index.html](http://www.firstmonday.dk/issues/issue2_2/peterson/index.html)
- Thomas, M.M. 1998. Furthering higher-order literacy using the Internet.  
<http://members.aol.com/csholumkc/mtholtranscript.html>
- Taylor, C. 1999. An ntroduction to metadata. <http://www.library.uq.edu.au/iad/ctmeta4.html>
- University at Albany. 2000. Second generation searching on the Web.  
<http://www.albany.edu/library/internet/second.html>

## **Disclaimer**

Articles published in SAJIM are the opinions of the authors and do not necessarily reflect the opinion of the Editor, Board, Publisher, Webmaster or the Rand Afrikaans University. The user hereby waives any claim

he/she/they may have or acquire against the publisher, its suppliers, licensees and sub licensees and indemnifies all said persons from any claims, lawsuits, proceedings, costs, special, incidental, consequential or indirect damages, including damages for loss of profits, loss of business or downtime arising out of or relating to the user's use of the Website.

ISSN 1560-683

Published by [InterWord Communications](#) for the Centre for Research in Web-based Applications,  
Rand Afrikaans University