1. Introduction

This article aims at providing the relevant background information necessary for the implementation of an intranet search engine, and its first focus is therefore the context of the intranet. It is not a do-it-yourself intranet search engine implementation, but brings the theoretical framework necessary before undertaking the practical implementation of an intranet. The report touches upon the comparative aspects of different search engines in a very limited way. This is so because search engines like everything else around Internet technology dates very quickly. For example, if Alta Vista the server of choice today, this may change completely tomorrow. At the beginning of this year, for example, some of the expensive Netscape solutions became free resources overnight.

This report is not about building an intranet, but a discussion of intranets is presented in order to state its purpose and its relationship to an intranet search engine. It therefore does not address intranet issues such as a preliminary needs analysis, the design, management and maintenance of an intranet, etc. Intranet search engines and intranets are completely interconnected and so is an intranet search engine with its underlying software which is dependent on database engineering. A discussion of an intranet search engine must therefore take its two other components constantly into consideration.

This report does not claim to be exhaustive about intranet search engines either, but attempts to highlight the fundamental points and some subsidiary points in the implementation of search engines. It is also not a report about the evaluation of a catalog of intranet search engines (see Zorn et al. 1997, for example, for specific evaluations), but highlights the points to be taken into consideration before making a decision about a particular choice of search engine. Appendix A lists some URLs of search engines for further exploration and up-to-date information.

2. What is an intranet?

2.1 Definition
At its most basic level, an intranet can be defined as a secure internal private network for the access and distribution of information within a company. If the company also has Internet access, the intranet will usually be hidden behind a firewall to prevent access from outside into the intranet. The Internet is a public network of networks and a Web site is placed outside a firewall. Tittel and Stewart (1997:85) say that 'the intranet is a hybrid of traditional private networks and the Internet, as well as a new entity of its own', and they add that 'the intranet has taken the best parts of networking and combined them with the capabilities of the Internet to create an IS solution that can be better than both' (Tittel and Stewart 1997:88).

If we move further from a basic level, an intranet can be seen as an information solution for a company in its daily activities. This goes beyond storage and retrieval of information. According to Frappaolo, of the Delphi group, 'the value in a company's information assets no longer lies in the ability to store and retrieve them but in the dynamic matching of the information to specific processes and unknown situations' (PR Newswire 1998).

2.2 Origins

One tends to link an intranet to the Web, but a secure internal non Web-based network was the precursor of the Web-based intranet and can still be an organisation's intranet. A Local Area Network running on a Novell Server is an intranet and packages such as Lotus Notes are intranet solutions. When a company uses its network internally, it can be seen as running an intranet. The intranet, however, is now associated as a private network which uses the TCP/IP protocol as used by the Internet.

2.3 Web-based intranet and intranet search engine

A Web-based intranet is a private network using a Web browser as an interface to its information, and it is run by a Web server such as Apache. It is much cheaper to run and easier to implement. The technical differences between an Internet and an intranet Web is that an intranet Web is run under a secure port (a secure port can be seen as a private address not visible or accessible by the outside world). Application software running on an intranet is also different, and although an intranet as defined in this report uses Internet technology, its purpose (to be addressed later) is completely different to that of a company's public Web site.

As employees are becoming more familiar with the Web and its search engines, providing the same interface eases the learning process. Training employees in the use of intranet search engines may also have the added benefit of making them better Web searchers.

2.4 Internet and intranet: some differences

Some of the issues which differentiate an Internet and an intranet are the following: firstly, an Internet is a public information system, while an intranet is a private one. Secondly, the Internet is an insecure system while security is a crucial aspect of an intranet so that different people have access or are denied access to different parts of the intranet. Thirdly, the Internet is not controlled by a specific individual or organisation, while an intranet is controlled by a company. Fourthly, the users of the Internet are potentially as many as the inhabitants of the planet, while the intranet's population is limited to a company's population.

2.5 Purpose and uses of an intranet

The central purpose of an intranet is to meet the information needs of a company. The kind of information needed to be accessed and distributed can be in the usual electronic formats: Web pages, word processor documents, Excel worksheets, e-mail, and the ability to participate in video conferences, ftp, etc. and to interact with colleagues on joint projects. The possibilities are endless and the intranet accommodates all these possibilities.

Additionally, unlike the Internet, the information on an intranet can be trusted, and can be managed.

2.6 Internet, intranet and extranet
The terms Internet, intranet and extranet have become buzz words which can create confusion. If we look at the kind of information which is suited to each of these networks, we can see where the intranet fits in.

Internet information is often available to serve as a PR function of the organisation publishing the information, to provide information in a very general way, often of cosmetic value, and as a way of attracting business. The emphasis is on public information. Information available on the Internet is potentially available to the whole world.

Intranet information, on the other hand, has its main purpose in easing the daily exchange of information between departments within an organisation, or within a department itself, as well as providing a platform to access and provide information from one department to another and to provide information which is private to the organisation or to individuals within the organisation, such as staff benefits and records. The issue of levels of access to information becomes of prime importance, and the intranet deals with its outside users through its firewall and with its inside users through its search engine.

Extranets can be seen as an extension of an intranet outside a company. Private information involving customers can be made available to them outside the company. The information is passed over the Internet but the data is encrypted and secured so that the rest of the Internet cannot tap into it. When a bank advertises itself as rendering services such as inter-account payments, balance enquiries, etc. over the Internet, this is not technically true: they are using Internet technology to provide the information over an extranet.

2.7 Value of an intranet to a company

Corcoran (1997) says that 'Intranets are becoming collaborative workspaces rivalling Lotus Notes in scope. The potential benefits include better and more efficient communication and faster decision making. In some cases, an intranet better integrates existing applications and eases the workload of IS.' The following give some examples of the value of an intranet to a company:

1. Timely information

   An internal phone book searchable on an intranet, for example, would provide up-to-date information which could overcome the problem of staff turnover and the resulting inaccuracy of information in the annually produced phone book.

2. Cost savings

   The example above also illustrates that the costs involved in the compiling and printing of a staff internal directory, searchable through departments and first names, can run into a lot of money.

3. Collaboration and teaching

   An intranet can be of enormous value for teaching within an organisation. The same course, tutorial, or workshop can be simultaneously accessed by employees, or rather by learners. As the kind of information needed by a particular company's employees can be very specific, this is best implemented on an intranet rather than on the Internet, with the possibility of moving to an extranet, depending on the nature and scope of the organisation's business. Information can also be shared immediately and work on interdisciplinary and collaborative projects can be shared. In the discussion on intranet search engines, it will be seen that something else than the rudimentary HTML index or table of contents is required, especially when it comes to full-text searching, and where an intranet search engine comes into its own.

2.8 Implementation of an intranet

The implementation of an intranet starts with a needs analysis. According to Zorn et al. (1997), 'intranets are scalable – they can start out small with just a few links or home pages in place, and can then grow easily over time to include a huge variety of information with little or no additional investment in infrastructure.' I must confess that I find this statement extremely
problematic. The intranet literature endlessly denies this. The best way to implement an intranet successfully, it says, is to first do one's homework properly. One has to do a needs analysis, to anticipate growth, to cater for extra disc space, etc. Even on a basic intranet, a search engine must be in place. In addition, one's first impression of a mediocre site is not conducive to the extensive utilisation of it. To my mind, although it is tempting to start small, other factors come into consideration which will determine the success or failure of an intranet. Just in terms of user interface only, an identity which keeps changing does not create a sense of trust in the use of the product and hence the continued use of an intranet may be jeopardised.

'To realize an intranet's strategic potential', Corcoran (1997) writes, 'companies must have a vision of what they want to accomplish and a central plan to carry it out. Once they have an objective, companies can then decide how to logically organize information so users can find it efficiently. They can also choose an architecture that will be able to handle anticipated increases in traffic and new types of applications and data down the line.'

Corcoran (1997), citing Cronin, says that he 'suggests asking what the critical functions of each department are and how people find the bits of information they use to do their jobs'.

The process involved in the implementation of an intranet and an intranet search engine are intimately related. As Fleenor says, 'the harder a piece of information is to find, the more time people will spend tinkering and finding other things to distract them' (Corcoran, 1997).

The intranet search engine is a solution to the management of information over the intranet. As DeVoney (1996) says, 'if information is money, the Internet and the corporate intranet are gold mines. The problem is letting your customers and users mine that data.' Or, in Kalin's (1997) words: 'As intranets extend their reach into more and more corporate data, using intranets without a search engine may become as frustrating as seeking a library book without an electric card catalog.'

3. Intranet search engines

Prior to the intranet, Internet search engines were (and still are) software solutions to retrieve information from the Web. The usual Internet search engine indexes the Web and builds a database, which is then searchable by means of keywords. Different degrees of sophistication of searching are offered by the various search engines such as Boolean searching and proximity searching or retrieval by date or URL.

As the database of the Web consists mainly of HTML documents, text and graphics, the Internet search engines have been developed to primarily search these formats of information.

With the intranet, however, the variety of documents is limited to the number of software packages available in creating them. Therefore, a different type of search engine must be used to retrieve documents which could be WordPerfect documents, Excel spreadsheets, Dbase files, etc. Furthermore, whereas the Web is a public place where documents are available to anyone, the intranet is the private database of a company and different people or sections need to have access to different documents. Therefore, a different kind of search engine is necessary to search different types of documents and to provide different levels of access to these documents.

3.1 How intranet search engines work

Most intranet search engines work in a similar way to Internet search engines. There is a robot wandering through documents to build an index and a database of documents, and an agent then proxies between the end-user and the database to retrieve the required document. Intranet search engines, however, index private documents behind a firewall instead of public documents over the Internet. They are administered locally and mostly through an Internet browser, and the majority of search engines index HTML and plain text files. At a cost, however, different search engines are available with varying degrees of sophistication.
In terms of the administration of search engines, there are varying degrees of control over how the engines index. For example, Wits uses the Infoseek search engine, which is used both as a search tool for the Wits Internet site and for its intranet. Unfortunately, little configuration can be done. Most of the engine is already compiled and what can be controlled is not much more than how frequently the search engine should revisit the site, and which directories can be excluded from the search engine. In addition, one cannot control which directory to index: if a new department goes live, one can direct the search engine to revisit the whole site but not specific directories – a great handicap considering that the search engine slows down the site considerably while re-visiting every single of the thousands of URLs. This search engine, however, is an Internet search engine used in an intranet environment and not an intranet search engine as such.

3.2 Why an intranet search engine?

An intranet search engine is used to categorise and make the intranet’s data accessible. The intranet, unlike the Internet, is also the place where trade secrets and strategic plans, for example, can be made accessible within a secure environment.

Given the limitations mentioned above, an Internet search engine used in an intranet environment is often a solution determined by financial necessity as opposed to an ideal solution. An intranet search engine should be able to index a variety of information formats and provide a high level of security in determining user access to data.

Corcoran (1997) says that 'companies designing corporate intranets should take care to ensure that data is carefully categorized and easy to navigate in order to ensure that the intranet enhances productivity. Many firms find that intranets waste time instead of saving it; much of the information available on intranets is poorly organized because the technology makes posting very easy.' It is the intranet search engine which will determine whether an organisation's intranet is a success or not.

Zorn et al. (1997) also say that, 'Internet search engines are often measured by their ability to provide access to the largest Web indexes, and retrieval from many Internet searches can be overwhelming. Intranet search engines are usually designed to provide more precise data filtering and retrieval, limiting the amount of information the user is required to sift through.' To do this, the actual indexing process of an intranet search engine is probably deeper than its Internet counterpart.

According to Balderston (1996) in InfoWorld, 'the companies that can configure their search engines for better relevance in search results will be the winners in the intranet field. That difference will come from how their search engines house information.'

3.3 Levels of intranets

Different types of intranets can be implemented. Thus, one could have an intranet being used as an information server where staff can check leave forms, salary scales, internal vacancies, medical aid information, etc. On the other hand, an intranet can be a very powerful interactive information system to enable people in a company to perform their actual duties.

4. Factors affecting the choice of an intranet search engine

4.1 Server dependability

When choosing an intranet search engine, one must take into consideration the available platform (such as Windows NT or Unix) and the type of Web server under which it will run. For example, Netscape's Catalog Server will work under any server but will run optimally under Netscape's Commerce or Enterprise servers (DeVorney, 1996).

4.2 Characteristics of a good intranet search engine

Many of the characteristics of an intranet search engine also apply to Internet search engines,
but issues of security are much more relevant to intranet search engines. Some of these issues are listed below. What follows relate more to an ideal intranet search engine. As companies differ, the type and level of access to their information will differ accordingly, and different search engines will be appropriate to different situations.

4.3 Type of documents to be indexed

Unlike Web documents that consist mostly of HTML files, an intranet can consist of documents in a variety of formats, ranging from an Excel Worksheet to an MS Word document. A search engine, therefore, must be able to index the contents of these different document formats. E-mail must also be able to be indexed so that one can search through information in past e-mail easily and efficiently.

4.4 Searchability

It may be stating the obvious, but to come back to the internal phone directory, the searchability of a search engine's database can be much more powerful than the present hard copy in that it can allow many more field searching. One could search for phone numbers, for example, and one update of the database would change the data which would be searchable from the different fields. In a paper-based copy this would require multiple updates, with the internal directory ending up a looseleaf publication with its associated problems of distribution and removal of redundant pages and replacement of new ones. Intranet and intranet search engines are entirely dependent upon advances in database technology. Without sophisticated database software, search engines would not be in existence.

4.5 Types of searches

The types of searches vary across search engines. Among the offerings are keyword and literal searches, concept searches using a thesaurus, advanced searches using Boolean and proximity operators as well as numeric searches and searches performed to obtain results within a specific time frame. In his survey, Devenoy (1996) also finds that some search engines support searches according to file size and to specific programs such as an Excel worksheet. In some companies there is also a need for multiple language searches, and in South Africa with its eleven official languages, such a search engine could one day become the norm, where an additional function of the search engine would be to become an intelligent translator.

It is also important when looking at search engines, to evaluate its specifications, how it searches, if full-text indexing is provided, what types of files are indexed, etc.

4.6 Security

Security is perhaps the biggest issue that an intranet search engine must address. The intranet needs to be secured from the outside world and the intranet search engine needs to be further secured within the organisation so that different levels of access are provided to different departments or individuals within the organisation. Different intranet search engines rely on different security protocols. For example, Microsoft's Index Server relies on NT security.

In an organisation like a university, security could be needed to cover the following examples: the finance department where employees do not usually have access to each other's salary; exam office with exam papers and the possibility of tampering with exam results; graduation office and the possibility of awarding oneself a degree. These are high profile security areas. However, within any department, access to information is strictly controlled, whether in the form of the privacy of e-mail of members within departments or employee departmental personnel records, etc. Even in areas that do not need high levels of secrecy, there must be effective ways of making the information available only within a restricted group. To illustrate this point, the following situation developed at the WITS Computer Centre: The Computer Centre runs short courses and workshops on the different application software available over the network. When the statistics were compared with previous years, there was a marked increase in the overall number of people attending the courses, but some courses were overbooked for some days while identical courses had to be cancelled on other days or were
run at a lost. Upon investigation, it was found that people using the intranet and the intranet search engine (since the file was not linked to the Computer Centre's page) had been able to access the teaching timetable. As a result, they could see who was teaching which course, and who were attending the course when, and decided not to book for a course according to the first date it was run, but according to who was teaching the course. Although the Computer Centre does not consider its course schedule as confidential information, prompt action had to be taken so that a) the search engine could not find the information, and b) authentication access had to be implemented on some files and directories.

If organisations are to use the intranet as an electronic workplace for the more effective management of its different tasks and activities, the data must be easily accessible but at the same time completely secure. Ultimately, no data can be completely secure, as the IT people who control the machines can access the data. This, however, is a problem which has existed all along. Management at Standard Bank, for example, changes the root password of all system administrators on a daily basis, and the first duty of any system administrator every morning is to report to their line manager for their passwords. Logs of every single activity on the machines are kept and scrutinised. However, there need to be a level at which some trust prevails. I personally prefer that the user be informed of all security issues, and then make a decision and consider the options about his/her data. The balance between paranoia and legitimate concern about security can be a fine one and intranet search engines need to be programmed to address these issues.

4.7 Platform dependability

An intranet search engine must also be able to view the data across multiple platforms. This is also addressed differently by different search engines and the choice of search engine will depend on one's computer environment. For example, speaking about the Microsoft's Index Server, an intranet search engine, DeVorney (1996) says that 'although the results can be viewed by any client with an Internet browser, Index Server only works using NT machines with Microsoft's IIS. If running your site from an NT server isn't a problem, Index Server's capability and inexpensive price (it's part of NT) make it an excellent choice.'

4.8 Search interface

The search interface of an intranet search engine must also be easy and intuitive to use. Admittedly, owing to the nature of design, one will not be able to please everyone. Navigation tools must also be displayed to ease the search process.

The search engine must be able to display the results in the original format or else to create an HTML document on the fly. Thus, an Excel worksheet should ideally be displayed as is or, as second best, in HTML format. This is a difficult issue, as all users don't possess all applications and their different versions.

4.9 Speed

Literature about the intranet repeatedly points out that the use of an intranet will far surpass the use of the Internet. This means that the search engines providing information retrieval tools will also need to cope with the volume of traffic on the intranet and provide fast search access in order to justify the use of an intranet.

4.10 Costs and a brief comparison of some intranet search engines

The costs of implementing an intranet can be relatively low if existing Internet software is used. The increased costs will be related to additional discs and to costs of additional processors to cope with increasing demands on a server.

An intranet search engine, however, is extremely expensive. On average, it ranges between $5,000 and $70,000, with different levels of searching provided across the price range. Free search engines are available, but they are often limited in one way or another. For example, the Inmagic/Lycos solution is platform dependent and it will therefore be limited to some environments.
Excite for Web Servers is also obtainable for free. It is not platform dependent, it offers concept searches and some Boolean searching, but some of its limitations are that it cannot search by number or date and it can only search HTML and text documents and the level of security is limited to the ability to decide whether a directory should be indexed. This puts Excite more in the category of an average Internet search engine.

On the other side of the spectrum, Open Text Livelink Search is extremely expensive, is also a multiple platform engine, offers extended searching facilities and can index most document types.

In deciding on which search engine is the best, factors such as the costs is an obvious one, but also whether a specific search engine is appropriate to a specific intranet environment. OpenText, for example, could be offering much more than is actually required.

4.11 Indexing and timeliness

For an intranet search engine to serve one of its purposes, namely the provision of timely information, the search engine must be able to index documents continuously. In this way, information will be available to the company as it is published.

Furthermore, an intranet search engine which can provide predetermined user profiles so that information relevant to particular users is pushed to them as it is published, would prove an additional benefit. Automatically updating a 'What's New' section together with an alert to the users would enhance the information's value even further.

These functions are not necessarily the functions of a search engine, but the functions of an intranet and its search engine are interconnected.

4.12 Accuracy and relevance

Different Internet search engines provide different degrees of accuracy and relevancy in the documents they return to the user when a search is performed, and this is dependent on both the type of search engine and the focus of its database. In an intranet environment, it is even more vital that when a search is performed, relevant documents are returned, so that a lot of time is not wasted sifting through useless information that does not contribute to decision making.

4.13 Degrees of control

An intranet search engine administrator must be able to have a high level of control over the administration of the systems, so that he/she can include or exclude directories which need to be indexed and also manipulate the way the search engine deals with matters such as duplication, dead links and relevancy.

4.14 Ease of implementation/skills required

When looking for a search engine, the expertise available is also important. Some search engines are literally 'turnkey', but others, especially the Unix-based ones, require CGI scripting, PERL and C programming.

Other considerations would be how much programming will be required to make the search engine work. The SWISH search engine at http://www.eit.com/software/swish/, for example, requires local CGI programming. Some other search engines such as Fulcrum, at http://fultech.com/, require a knowledge of the SQL language.

The level of support provided is also important. As a general rule, the cheap or free search engines tend to lack solid support.

4.15 Disc space and directory consideration

The amount of disc space available will also need to be considered when choosing an intranet
search engine. As an intranet grows, so does the size of its database, and if the search engine involves full-text searches, the space required becomes quite considerable.

As databases of search engines also grow very fast, care must also be taken in which directory these databases are allowed to grow. Some file systems on the Unix system, for example, bring the machine to a standstill as soon as they are full even if other directories on the machine are relatively empty and available.

4.16 Size of organisation and expected importance of intranet

The purpose of an intranet will vary according to the size of an organisation. Thus a small company of ten people will have very different information needs to a large organisation consisting of thousands of employees. The intranet will also be affected by the type of organisation. With a highly decentralised organisation, administrative procedures can become a nightmare. At Wits, for example, the Computer Centre runs over 55,000 account codes for the exchange of real and funny (internal requisitions) money within departments and between departments and the outside world. Implementing all these individual accounts, searchable within an intranet with the proper level of security, could prove a daunting task.

4.17 Administration

Care must be taken not to overkill an intranet site by using machines which are more powerful than the requirements of the organisation.

Ease of administration is also important. In typical organisations, new technologies are usually taken as additional tasks within a specific area of work, rather than as a category in itself. The search engine must be able to be easily maintained and administered.

4.18 Reporting functions

Depending on the level of transparency or control or 'big brother watching' culture in different organisations, this feature might have varying degrees of importance. For example, Fulcrum, a commercial intranet search engine, can produce reports on employees’ usage of the intranet.

4.19 Employee education

The successful implementation of an intranet search engine also involves the training of employees in becoming conversant with the searching of intranet organisational information. Documentation must therefore be produced as well as training workshops. This would empower the employees to find the information they need to perform their work, and this can 'save time and foster employee satisfaction – two things of value to most organisations' (Zorn et al 1997).

4.20 Convincing management

Management problems encountered in the implementation of an intranet and consequentially the implementation of an intranet search engine can be enormous. An intranet is often rejected as a latest fad, without first seeing that everything latest is not necessarily a fad. The issues surrounding acceptance by the powers that be can be real and should not be underestimated: whether it is due to hidden agendas or through sheer ignorance, the acceptance and therefore implementation of an intranet can be a very difficult process. When costs are involved, such as those which possibly have to be incurred in the acquisition of an intranet search engine, the issue is even more difficult to resolve.

There is also the opposite problem of wanting to have an intranet on the fly, not realising that this is an issue with a lot of ramifications and a hasty decision can be more detrimental in the long run. This ties up with intranet search engines, as the issues surrounding an intranet are intimately linked to the ways that the data will be searchable, accessible, protected and maintained, so that the intranet becomes the tool of choice to search for timely and accurate information which enhances the daily activities of an organisation. An intranet on the fly is just a collection of information, which serves mostly the people who are involved with the publishing of the information and therefore know where the information is located. To the end-
user, it may be a massive waste of time. In the phone book example given above, it may very well be less time consuming to dial the switchboard and wait in a queue to enquire the number from the operator.

5. Conclusion

Many enterprises and companies embark on the implementation of an intranet without realising all its implications and its full potential. The implementation of an intranet is closely and intrinsically linked to making the data accessible and is therefore inseparable from the methods of searching and hence to an intranet search engine.

An intranet search engine must be able to minimise the amount of time which is used in searching for information, so that using information becomes possible. New developments in the field of intranet search engines occur continuously. As DeVoney (1996) says, ‘expect the engines to work faster, offer even more search features and index more types of documents to better suit your corporate Web or intranet needs’.

Having an intranet with a plethora of inaccessible information will defeat the whole purpose of the intranet. Although an intranet search engine is no panacea to a company's management of information, it provides an information management solution which is extremely sophisticated as well as endless possibilities of exploitation to reach the company's targets and goals.

6. References


7. Appendix: List of some intranet search engines sites

- Lycos Site Spider and Inmagic, at http://www.lycos.com/software/software-intranet.html
- Verity, at http://www.verity.com
- Excalibur, at www.excalib.com
- AltaVista intranet, at http://altavista.software.digital.com/
- ht://Dig, at http://htdig.sdsu.edu/
- Harvest/Glimpse, at http://glimpse.cs.arizona.edu/webglimpse/
- Fullcrum Surfboard, at http://www.fultech.com/

A comprehensive site about the intranet claiming to be the complete intranet resource can also
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.