



---

## SELECTING AN INTRANET SEARCH ENGINE

---

April 1<sup>st</sup>, 2004, © ISYS Search Software Inc

---

### ISYS Search Software Worldwide

---

#### **United States**

ISYS Search Software Inc  
8775 East Orchard Road  
Suite 811  
Englewood, CO 80111  
USA

Phone: +1 303 689 9998  
Email: [info-usa@isys-search.com](mailto:info-usa@isys-search.com)  
Fax: +1 303 689 9997

#### **Australia**

ISYS Search Software P/L  
Suite 102,10-12 Clarke St  
Crows Nest NSW 2065  
Australia

Phone: +61 2 9439 5800  
Email: [info-aus@isys-search.com](mailto:info-aus@isys-search.com)  
Fax: +61 2 9439 8569

#### **United Kingdom**

ISYS Search Software (UK)  
Limited  
The Steam Mill  
Steam Mill Street  
Chester CH3 5AN, UK

Phone: +44 0 1244 313 216  
Email: [info-uk@isys-search.com](mailto:info-uk@isys-search.com)  
Fax: +44 0 1244 313 003

## INTRODUCTION

An effective search tool on an intranet can make an enormous difference to its usability. In fact, usability expert Jakob Nielsen found that "Poor search was the greatest single cause of reduced usability across intranets"<sup>1</sup> A good search engine ensures that users find what they're looking for, first time, regardless of the format or location of the information. This means that a wide variety of information can be effectively dispersed and made available to staff, without the need for complex navigation systems or filing conventions.

Most intranets evolve over time, and search functionality need not be a daunting task. A search tool can be implemented quickly, and then refined as the intranet grows and the needs of the organization change. Importantly, a flexible search engine that is cost-effective and expands to suit your growing requirements can be a much better investment than a cheap product with no ability to scale or an expensive solution where the majority of its functionality is wasted.

It is important to recognize that every intranet is different, with its own objectives, requirements and environment. This document is designed to guide you through the process of selecting an intranet search engine, by addressing in turn a range of key variables that need to be considered. You can use the 12-step checklist at the end as a template for defining your own requirements and evaluating vendors. The following topics are covered:

1. The Basics - Objectives
2. Data Sources & File Types
3. Technology Environment
4. Features
5. Installation and Maintenance
6. Pricing Structure
7. Vendor Credentials

### 1. THE BASICS - OBJECTIVES

A good start to understanding what you need from your intranet search engine is to firstly understand the role of the intranet for your organization. What is the objective of the intranet? Is it to provide human resources information to staff? To provide information to your help desk? Or to support the sales team? Perhaps the intranet is designed to meet a number of different objectives. Once these are clearly defined, you can begin to determine how a search engine can best add value.

The functionality you require from a search engine will depend on your intranet objectives. Keeping these in mind, the following sections of this document outline some of the issues that you need to consider.

---

<sup>1</sup> <http://www.useit.com/alertbox/20021111.html>

## 2. DATA SOURCES & FILE TYPES

Once you have your objectives clearly defined, you can work out what type of file formats and data sources your search engine will need to support. List out every file type used in creating the information that you want to share on your intranet. These usually fall into one of three categories:

- **Unstructured formats** are file formats that contain primarily text-based information. These include text files, word processor files, PDFs, emails and formats used to create most documents. There is no real structure to these file formats and few relationships exist between elements within them.
- **Semi-structured formats** are file formats that contain a mixture of text-based and data-based information, with a basic structure. These include file types such as HTML, spreadsheets, XML. There may be relationships between elements within these files, however they are not as rigidly defined as they are in structured formats, and there may be sections of textual information where no structure exists.
- **Structured formats** are file formats where the information is contained in a well-defined structure, such as a relational database. Many enterprise systems have a structured architecture, such as ERP and CRM systems, as well as many legacy databases.

For example, you may want to start your search engine implementation by simply making available all of the HTML content on your intranet. Then you may decide to include PDF and word documents. Eventually, you may want to provide access to your customer database or to XML data created in other systems. Your intranet search engine needs to be able to support the full spectrum of file types that you require, even in multiple languages, if that is a concern for your organization.

In addition, you should consider how many files will be in the intranet data repository. This can affect search engine performance, and may impact the price you are ultimately charged, since many search engines are priced according to the number of documents indexed. Be careful in such situations, as any expansion in the size of the intranet can push the search engine into the next pricing bracket, thereby attracting additional charges from the search engine vendor.

Think about publishing scope – depending on the sources of the information you want to make available via your intranet, you may need a search engine that provides umbrella functionality over many different data sources. Some search engines are designed for specific data types, such as structured data, or are limited to single data repositories, such as search engines within content management systems. If you need to incorporate legacy data into your intranet, you need a search engine that can support this.

Finally, consider whether you want to incorporate external data sources into your intranet. Spidering functionality enables external websites to be integrated into your intranet data repository. You can include the websites of suppliers or partners, for example, into your intranet search functionality, so that staff can search that content directly from the intranet.

### 3. TECHNOLOGY ENVIRONMENT

Your options will be determined by your internal technology infrastructure. List out all of the systems and software that will interact with the intranet search engine. These include the intranet web server operating system and application software, your networking software and architecture, and your file system security environment. Think about any other systems that need to integrate with the intranet search, such as content management systems, ERP, CMS, email servers or legacy systems.

### 4. FEATURES

For most intranets, there will be a wide spectrum of users, from very basic all the way through to highly technical power users. The search function needs to cater for all of these people, with a simple yet powerful interface that provides options for advanced searching if required.

There should be three steps to the search process, and a range of features work to streamline each of these steps; the important ones are described below. But remember, you may have specific objectives for your intranet that require certain features rather than others and you should keep this in mind when discussing features with vendors – which ones do you really need for your specific application?

The three steps in the search process are (1) Entering the Query, or asking the initial question, (2) Getting the Search Results, or receiving the list of found documents back from the search engine, and (3) Finding the Right Answer, or examining and refining the search results to find the information you were looking for.

**Step 1: Entering the Query.** When a user enters their query, they should have the option to do this using a natural language approach; that is, by simply entering the question as they would ask it. Such as “What is our returns policy on refrigerators?” There should also be the option to build queries using Boolean operators, so that users who know exactly what they want can be extremely specific with their search. For example “returns~ within 10 words of refrigerator but not freezer”. Check the user interface to make sure it is intuitive for basic users, but also provides powerful advanced search functionality for more experienced users.

A good search engine should enable you to group logical chunks of information together so that searches can be conducted on specific areas of interest. For example, an engineer may only be interested in searching technical documentation and might not want to search the HR policies that are also available on the intranet. The search engine needs to be able to group information separately to enable this to happen.

There should also be tools such as thesaurus functionality, where the search engine picks up common words with similar meanings, and synonyms, a thesaurus that can be custom-defined to include terms relevant to your particular organization. Spelling errors should be catered for with a ‘sounds like’ function, which enables users to find other words that sound similar to the one that they are typing. This compensates not only for spelling errors when the user enters the query, but also for errors within the data itself, ensuring that such data is not lost forevermore.

**Step 2: Getting the Search Results.** If you have very specifically defined data, such as legal documents, a high degree of precision may be required to identify and return specific information. In other situations, however, it may be better to return a wider range of documents for a given query. The accuracy you require depends on the role of the search engine and the nature of the data.

### *Selecting an Intranet Search Engine*

If you want to make available a large volume of data on your intranet, providing a fast search engine is important. Otherwise users find it frustrating to wait for the search engine to bring back the search results. With smaller amounts of data this will be less of a concern; it all depends on the volume of data that you intend to make available on the intranet.

Any good search engine should use some form of intelligent relevancy determination. This is where the search engine, based on the query entered, makes a judgement about which results will be the most relevant, and ranks them accordingly.

**Step 3: Finding the Right Answer.** The search process doesn't stop once the user receives the list of results. They then need to refine and manipulate the results list until they find exactly what they were looking for. There are many features that can assist in this task, some of which include:

- **Document summary information** – the display of useful document attributes such as file type, file size, date last changed, relevancy rating and the number of 'hits' (key words found) in the document. The display of an extract of the document, say several lines above and below the first hit, is helpful for determining the context in which the document has been returned.
- **Re-sorting** – the ability to re-sort the results list using different criteria, such as title, number of hits, relevancy, date changed, file type or any other criteria that makes sense for your organization.
- **Hit-to-hit navigation** – the provision of navigation buttons enabling users to go directly to the first hit in the returned document, and thereon to the next or previous hit as required. This means users avoid having to read through pages and pages of document before finding the relevant section, making it much more efficient.
- **Hit highlighting** – a familiar concept from searching the web, hit highlighting is when the key words, or 'hits', in a document, are highlighted in a different colour. This feature is often not available in an intranet search engine, but it really should be, as combined with hit-to-hit navigation it enables users to immediately see the relevant sections of the document.
- **Fast preview** – the ability to preview large non-HTML documents in a basic HTML format, without the need for downloading the whole document. This function enables users to view a few lines above and below each hit, and then to expand up or down to continue reading.
- **Search within** – the ability to search within the current set of results, to further narrow them.

Although just some of the features available in intranet search engines, these are the main features required to ensure that users have the best overall experience. Others that may be relevant to your organization might include **intelligent agents** that automatically advise users when relevant content appears in the data repository, or the ability to **save or export** search results.

If you have a specific need, you should discuss it with your potential vendor. If they don't have a ready-made feature to solve your problem, they'll probably have software tools to enable custom requirements to be developed.

## 5. INSTALLATION AND MAINTENANCE

The process required to install and maintain the search engine software can have an enormous impact on the overall cost and ROI of your project.

- Does the search engine require significant data preparation and re-location, or can it index the information where it currently resides?
- Does the system have to be customized and will it require external systems integration work to install? What will all this cost?
- What happens if you want to modify or expand the system – can you do this in-house using staff with general IT skills or will you need to call in a specialist?
- What ongoing maintenance of the system is required, and what type of person is required to undertake this maintenance?

All of these are important considerations when it comes to investing in search engine software, and ideally you should look for a system that minimizes both the up-front installation cost as well as the ongoing maintenance. There are certain types of systems that legitimately require a lot of IT effort in order to maintain their usefulness, however the standard intranet search engine should not fall into this category.

## 6. PRICING STRUCTURE

Pricing structures vary between vendors. Some charge according to the number of documents indexed, while others charge according to the number of end users. Consider carefully how you think your system may expand over time, as increases in documents and/or users can incur additional charges. Some vendors use a rental type model, where the license fee paid is only valid for a year or two years. At the end of the period the fee is payable again. Others charge a once-off fee and only charge for upgrades to new versions.

Most vendors offer some type of maintenance – make sure to ask how much this will cost, and how it will change over the life of your project.

## 7. VENDOR CREDENTIALS

As with all purchases, it's critical to ensure that your vendor is reputable. Check that they have been in operation for a reasonable length of time, and find out how mature their particular offering is. Early generations of software are likely to be less tested and less featureful than software that has been refined over many years.

Establish whether the intranet search engine product range is a priority for the vendor. If the vendor is primarily focused on web search rather than enterprise search, their methodologies are likely to have a very different focus and their software will be geared towards web content (HTML, PDF etc) rather than enterprise content, which can span a diverse range of formats. Make sure the product will continue to receive ongoing development and technical support. Ensure the vendor offers a help desk for technical support.

Importantly, ask the vendor to provide case studies and examples of other intranet implementations for which their product has been used. Ask for references so you can talk to other organizations that have used the software.

*Selecting an Intranet Search Engine*

**SUMMARY**

In this document we have provided some guidelines for evaluating and selecting an intranet search engine. As with all software, knowing what you're trying to achieve and talking to a number of different vendors is a good start. The massive benefits that can be obtained from getting your intranet search functionality right from the start are worth the effort of some upfront research. You can use the checklist at the end of this document to specify your requirements and compare these with different vendor offerings. Good luck!

**TO FIND OUT MORE...**

Contact ISYS Search Software in your area for more information on intranet search solutions.

**United States**

ISYS Search Software Inc  
8775 East Orchard Road  
Suite 811  
Englewood, CO 80111  
USA

Phone: +1 303 689 9998  
Email: [info-usa@isys-search.com](mailto:info-usa@isys-search.com)  
Fax: +1 303 689 9997

**Australia**

ISYS Search Software P/L  
Suite 102,10-12 Clarke St  
Crows Nest NSW 2065  
Australia

Phone: +61 2 9439 5800  
Email: [info-aus@isys-search.com](mailto:info-aus@isys-search.com)  
Fax: +61 2 9439 8569

**United Kingdom**

ISYS Search Software (UK)  
Limited  
The Steam Mill  
Steam Mill Street  
Chester CH3 5AN, UK

Phone: +44 0 1244 313 216  
Email: [info-uk@isys-search.com](mailto:info-uk@isys-search.com)  
Fax: +44 0 1244 313 003

## INTRANET SEARCH ENGINE REQUIREMENTS & EVALUATION CHECKLIST

Complete the blanks before evaluating your selection of search engine software against the listed criteria.

### 1. INTRANET OBJECTIVES - what are the main objectives of your intranet?

*For example:*

- *Share office policies & procedures with staff*
- *Accelerating the sales process*
- *Providing a tool for the help desk*
- *Helping management make decisions*

How do you expect the intranet to grow and expand over time?

- Inclusion of additional file types
- Inclusion of more data
- Incorporation of other parts of the organization
- Integration with other enterprise systems

How do you see the search engine best adding value to your intranet?

*For example:*

- *Providing access to 10 years of legacy data*
- *Searching previous sales proposals*
- *Quick access to technical manuals for the help desk*
- *Searching minutes of board meetings*

### 2. ENVIRONMENT

Upon which technology is your intranet based?

- |                                  |                                    |   |
|----------------------------------|------------------------------------|---|
| <input type="checkbox"/> Windows | <input type="checkbox"/> Unix      | <input type="checkbox"/> Other (which?) |
| <input type="checkbox"/> Linux   | <input type="checkbox"/> Macintosh |   |

### 3. FILE FORMATS

Which file formats do you need to support? *Check all that apply*

#### Unstructured formats

- |  |   |  |
|--|---|--|
| <input type="checkbox"/> Text (ANSI/ASCII)   | <input type="checkbox"/> Microsoft PowerPoint   | <input type="checkbox"/> Email – which format? |
| <input type="checkbox"/> Adobe Acrobat (PDF) | <input type="checkbox"/> Rich text format (RTF) | <input type="checkbox"/> Flash                 |
| <input type="checkbox"/> Microsoft Word      | <input type="checkbox"/> ZIP                    | <input type="checkbox"/> Wordperfect           |

#### Semi-structured formats

- |  |   |  |
|--|---|--|
| <input type="checkbox"/> Dbase (dbf)     | <input type="checkbox"/> Lotus Notes                  | <input type="checkbox"/> External websites |
| <input type="checkbox"/> Microsoft Excel | <input type="checkbox"/> Lotus 1-2-3                  | <input type="checkbox"/> HTML              |
| <input type="checkbox"/> AutoCAD         | <input type="checkbox"/> Comma separated values (csv) |  |

#### Structured formats

- |   |                               |  |
|---|-------------------------------|--|
| <input type="checkbox"/> Microsoft Access     | <input type="checkbox"/> XML  | <input type="checkbox"/> Oracle              |
| <input type="checkbox"/> Microsoft SQL Server | <input type="checkbox"/> ACT! | <input type="checkbox"/> Other ODBC-database |
| <input type="checkbox"/> mySQL                |                               |  |

#### Other formats (custom, legacy information or other)

- 1.
- 2.
- 3.

### 4. FILE VOLUME & SIZE

How many files will be included in the searchable data repository? How large will the data repository be?

#### Number of files

- <1,000
- 1,000 – 10,000
- 10,000 – 100,000
- 100,000 – 500,000
- > 500,000

#### File storage

- < 10 MB
- 10 – 500MB
- 500MB – 1GB
- 1GB – 10GB
- >10GB

### 5. SECURITY

What system is used for file-level security?

- NT
- Netware
- .htaccess

### 6. DATA SEGMENTATION

What groupings will be required for your data?

- |                                |   |   |
|--------------------------------|---|---|
| <input type="checkbox"/> None  | <input type="checkbox"/> Geographical         | <input type="checkbox"/> Functional<br>(human resources, engineering,<br>marketing, customer service etc) |
| <input type="checkbox"/> Other | <input type="checkbox"/> By business division |   |

Selecting an Intranet Search Engine

## 7. FOREIGN LANGUAGES

What foreign languages must the system support? (list)

- 1.
- 2.
- 3.
- 4.
- 5.

## 8. FEATURES

**STEP 1** – Entering the Query – what query-related functionality do you require?

- |  |  |
|--|--|
| <input type="checkbox"/> Natural language query  | <input type="checkbox"/> Thesaurus (standard)  |
| <input type="checkbox"/> Boolean query   | <input type="checkbox"/> Synonyms (specific to customer)                                     |
| <input type="checkbox"/> Options for intermediate & advanced queries   | <input type="checkbox"/> 'Sounds like'<br>(eg. "litergayshun" = "litigation")                |
| <input type="checkbox"/> Query builder (menu of Boolean commands that users can click on to construct their query) | <input type="checkbox"/> Spell checker/'did you mean'?                                       |
| <input type="checkbox"/> Taxonomies  | <input type="checkbox"/> 'Starts with'   |
| <input type="checkbox"/> Fielded and meta-data searching (eg. "author = Jenkin")                                   | <input type="checkbox"/> Conflation (eg. "Manage~" returning Manager, Managed etc)           |
| <input type="checkbox"/> Search separate data groupings<br>(eg. Search HR documents separately from engineering)   | <input type="checkbox"/> Proximity searching<br>(eg. Within 10 words of)                     |
| <input type="checkbox"/> Search date ranges<br>(within files, not just on file save dates)                         | <input type="checkbox"/> Restrict search by file type<br>(eg. Only text files)               |
| <input type="checkbox"/> Intelligent date & number recognition<br>(eg. 01/01/02 = 1 <sup>st</sup> Jan 2002)        | <input type="checkbox"/> Intelligent agent – advise user when new data appears in repository |

**STEP 2** – Getting the right search results – what results-related functionality do you require?

- |  |  |
|--|--|
| <input type="checkbox"/> Relevance ranking methodology | <input type="checkbox"/> High speed return of results          |
| <input type="checkbox"/> User-level security           | <input type="checkbox"/> Save or export results list to a file |
| <input type="checkbox"/> Accuracy                      |  |

**STEP 3** – Finding the right answer – what functionality do you require to help users get the answer they are looking for?

- |  |  |
|--|--|
| <input type="checkbox"/> Document statistics displayed in results (eg. File size, file path, save date, # of hits)                           | <input type="checkbox"/> Hit highlighting in document view, including PDFs                                       |
| <input type="checkbox"/> Fast document preview – viewing relevant sections of documents without downloading or scrolling through large files | <input type="checkbox"/> Hit-to-hit navigation<br>(eg. Go to first hit, go to next hit, go to next document etc) |
| <input type="checkbox"/> WYSIWYG document view<br>(eg. View in PDF or Word layout)   | <input type="checkbox"/> Hit highlighting in results list  |
| <input type="checkbox"/> Ability to use HTML templates to display results from databases (SQL, XML, other)                                   | <input type="checkbox"/> Automatic generation of Table of Contents for documents                                 |
| <input type="checkbox"/> Re-sort results list (eg. By best match, # of hits, size, file type, date)  | <input type="checkbox"/> Filter results list (eg. By file type or date range)                                    |
|  | <input type="checkbox"/> Search within results   |

## 9. OTHER

What functionality do you require that is specific to your organization or environment? (list)

- 1.
- 2.
- 3.
- 4.
- 5.

*Selecting an Intranet Search Engine*

**10. INSTALLATION & MAINTENANCE**

How long will it take to implement the system and how much will it cost to implement and maintain?

	<u>Time Taken</u>	<u>Cost</u>
Initial piloting of software	_____	\$ _____
Customization of software to suit your needs	_____	\$ _____
Installation of software across enterprise	_____	\$ _____
Ongoing maintenance requirements (per year)	_____	\$ _____

Does the vendor offer a help desk for technical and support queries?

**11. PRICING**

What is the vendor's price for providing your desired functionality?

One-off software licenses	\$ _____
Ongoing software/hardware rental fees	\$ _____
Ongoing maintenance fees	\$ _____
Cost for adding users	\$ _____
Cost for adding data	\$ _____
Cost for adding file types	\$ _____
Cost for any additional software modules	\$ _____

**12. VENDOR CREDENTIALS**

What are the vendor's credentials?

- Have they been in business for more than five years?
- Are they profitable?
- Are they focused on enterprise search technology?
- Do they have an established customer base?
- Can they provide case studies to demonstrate that they have successfully implemented systems with similar characteristics to yours?
- Can they provide referees?

**END OF CHECKLIST**