



ISYS Search Software Resource Consumption and Performance Characteristics

Contents

What Software Salespeople Can't Tell You	1
ISYS and Your CPU	2
Indexing.....	2
Querying.....	2
Disk Space and Growth Characteristics	3
Structure of index files.....	3
Document index	4
Vocabulary Index	4
Index References	5
Memory	6
Data format	6
Presentation method and quality.....	6
Number of users.....	6
Input/Output (I/O) Activity	7
Indexing.....	7
Querying.....	7
In Conclusion... ..	8

What Software Salespeople Can't Tell You

If you're looking to buy or upgrade a piece of software, you're probably asking all sorts of questions, such as:

- Does the software perform the functions I want it to?
- Can I get adequate after-sales support?
- Is it within my company's budget?
- How will it impact the resources that are already in place?

Most of these questions are easily answered – all a software company has to do is look at your business needs and match the right product. But the last question, 'How will it impact the resources that are already in place?' – is not such a simple one to answer.

How a piece of software impacts on your IT infrastructure depends on a lot of factors – some are foreseeable, some are not. You may know what the load on your network is now, and how many people will be using the software to start off with, but it can be very difficult to predict how things will happen in the future. How will your network expand? Will you employ more people? What other software will the company use? What processes will be running on different machines? How will your users' habits change?

For example, some companies who invest in setting up an intranet will spend time and money designing it, implementing it and training staff to use it, only to find that an initial boom of interest wanes in short time. On the other hand, a new piece of software that is used initially by a handful of people may quickly be in high demand when word gets out that it is easy to use and increases productivity.

This paper aims to put some concerns to rest about how ISYS Search Software consumes resources in a typical environment. We can't predict what your company will do tomorrow, but we can certainly give you an idea of how ISYS can fit almost seamlessly into your established environment.

ISYS and Your CPU

The amount of CPU that ISYS uses depends on which of the ISYS products you are using, and what functions you are performing.

Indexing

When indexing your documents, CPU usage can be as low as 2% during the scanning stage (when ISYS is looking for new, changed and deleted documents to index), and as high as 90% during the document reading stage (when ISYS is extracting each word from your documents and adding them to your index) in extreme cases. An 'extreme' case may involve large, complex documents using RichHTML. Whatever *your* case, ISYS will not use CPU that is required by other priority tasks. ISYS will only use as much CPU as your operating system can spare at the time.

Querying

Querying an index takes up very little of your CPU. Likewise, if you're using the ISYS:web server, CPU usage is minimal.

Disk Space and Growth Characteristics

All files associated with an index will grow in size as the number of documents increases. However, ISYS index files have been designed to be efficient in both speed of retrieval and disk space usage over a wide range of situations. The actual size of disk files is dependent on a number of factors, such as spread of vocabulary, average size of documents, use of special characters and common words.

Structure of index files

To illustrate the way in which ISYS uses disk space, it's helpful to look at how ISYS index files are structured. An ISYS index consists of three interdependent files. ISYS uses a proprietary index access method called V-Tree. This is similar to a B+ Tree but faster and more space efficient. The index files are described below:

ISYS.IXA

This is a block index into the ISYS.IXB file.

ISYS.IXB

This is a collection of different types of records. It contains:

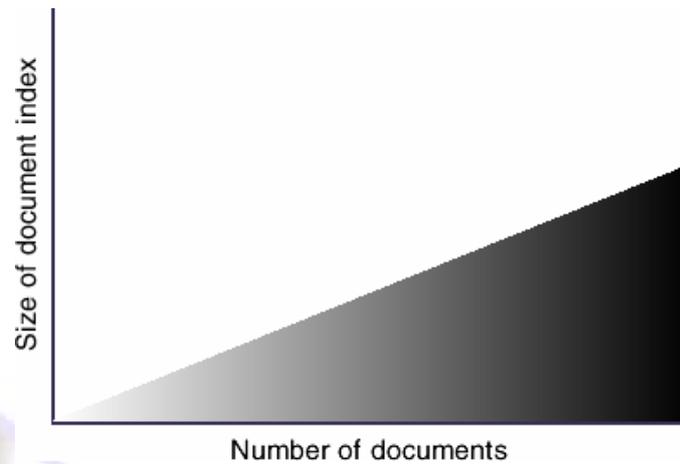
- V-Tree entries for documents, one per indexed document
- V-Tree entries for words, one per unique word
- Blocks of word references that contain document, paragraph, word and number information
- Document bookmarks or documents over a page in length

ISYS.IXC

This is an index of the documents by unique number, containing statistics such as number of words and the date/time stamp of the file at time of last index.

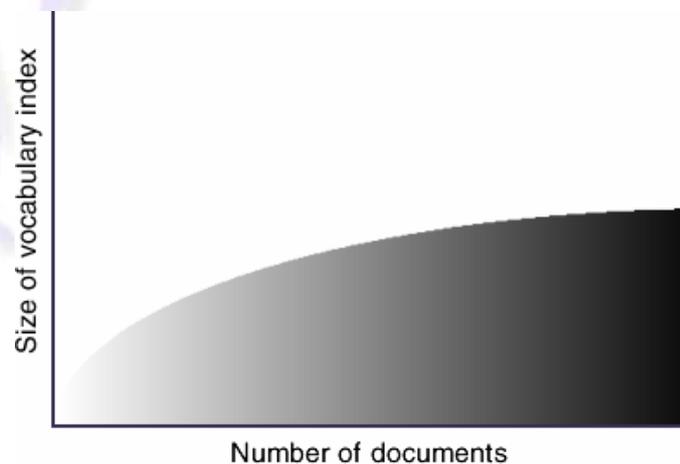
Document index

The document index file, ISYS.IXC and entries in the ISYS.IXB document V-Tree will grow linearly as the number of indexed documents increases. The browse bookmarks will increase as the number of indexed documents increases. A bookmark is made for each page and for every document that is greater than a page in length.



Vocabulary Index

The vocabulary index entries in the ISYS.IXB file have a very different growth curve to the document index and index references. Initially, new words will be added to the vocabulary regularly. Once the amount of text indexed reaches a critical mass, the number of new words encountered will be very small. The diagram below illustrates the expected growth of the vocabulary entries.

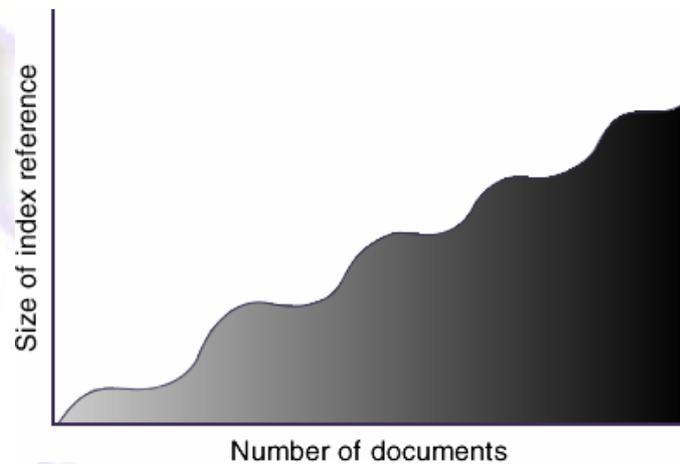


Index References

The size of each index entry is variable, so that small indexes do not carry a large overhead for pointer space that may not be used. ISYS automatically calculates the best size for index entries.

The references to words increase as the amount of indexed text increases. ISYS employs variable size entries in the index reference blocks in the ISYS.IXB file to limit wasted pointer space, while allowing a free space overhead to improve subsequent indexing speed.

ISYS dynamically and automatically adjusts the pointer size and the allocation of divisions within the pointer to give optimal performance and use of disk space. As the index grows, a mixture of pointer sizes will evolve within the ISYS.IXB index reference blocks. Pointers vary in size between 1.5 and 20 bytes.



Memory

The amount of memory that ISYS uses varies according to a number of factors. Some of these factors are outlined below.

Data format

Different data formats will have a different impact on your available memory. For example, if your indexed data consists largely of SQL records and Act! Contact Management files, you will use slightly more memory than if your data is mostly text files and Word documents.

Presentation method and quality

ISYS offers you several options for how you would like to view your documents. The basic rule is that the simpler the presentation, the lower the memory usage. That is, if you choose to view your SQL data using HTML templates and your Word Documents using the WYSIWYG viewing option, you will need a little more memory than if you view all your files in a draft format.

Number of users

The number of users has very little impact on the memory consumption of ISYS. However, if you have many users performing multiple queries at the same time, you may notice the impact this has on available memory.

Input/Output (I/O) Activity

Like memory and CPU usage, the I/O levels on your computer vary depending on what ISYS functions you are performing.

Indexing

During indexing there is a lot of activity with the ISYS index files (IXA, IXB and IXC) and temporary files associated with the indexing process.

As ISYS scans and reads your documents, there is significant activity associated with the document files. During the scanning process, ISYS identifies the documents that are to be added to, or updated in, your index. Activity levels are higher during the reading phase, as ISYS opens and reads each document sequentially.

Querying

During query time the file activity is kept to a minimum. I/O activity is only slightly higher when documents are opened and viewed.

In Conclusion...

Many people in the market for a search tool assume that a product with the capabilities of ISYS must consume resources at a great rate. The fact is, in the large majority of cases, little or no adjustment to a company's IT infrastructure has to be made to accommodate ISYS.

Some companies will stock up on memory and new machines in preparation for the installation of a new software or server system, when it's not really necessary to do so. The best practice with ISYS is to start with what you've got. As long as you meet the minimum requirements for installation, chances are that ISYS will meet your needs without having to spend extra money, time and effort.

For optimum performance and efficiency, we recommend the following:

- Store your indexes on the same machine as your ISYS software. This will improve the performance of indexing and updating.
- Take the time to customize ISYS for your needs. ISYS provides many options for indexing, querying and viewing your files – use them to your best advantage. For example, keep your Common Words List up to date; customize your Significant and Insignificant characters lists, and use draft view presentation if WYSIWYG isn't necessary.
- Update indexes regularly to remove old data and maximize the accuracy of your search results.