

ISYS for WINDOWS
Information Retrieval
Version 4.0

Advanced Guide

TABLE OF CONTENTS

Introduction	Page 3
The Query Program	Page 4
The Operators	Page 4
Precedence of Operators	Page 5
Special Symbols	Page 5
Common Words	Page 6
Reserved Words	Page 6
Significant/Insignificant Characters	Page 7
Querying within the Query Results	Page 7
Other Query Functions	Page 7
Query by Concept Facility	Page 8
Establishing the Concept Tree	Page 8
Making a Query by Concept	Page 9
Plain English Query.....	Page 8
Annotations (notes, images, hyperactivities).....	Page 8
Word Processor Activation	Page 10
Database Maintenance	Page 11
Network Operation	Page 11
ISYS Auditor	Page 11
Database Publishing	Page 13
ISYS Publisher	Page 13
The Application Programming Interface	Page 14
Specifications	Page 14
Sample Applications	Page 15
Common Questions	Page 16
Intro Disk Distribution Policy	Page 17
Getting Help	Page 17

INTRODUCTION

~~~~~

This document is designed for the more technically-minded reader who wishes to more fully understand the advanced functions of ISYS for Windows. It is assumed that you have already read the "shorter" Introductory Guide, but is still a highly abbreviated document. The Reference Manuals supplied with the full ISYS product contains in-depth explanations of all functions, as well as full technical specifications, but we have attempted with this document to give you a sense of the power and versatility which ISYS provides.

## THE QUERY FUNCTION

~~~~~

In the smaller introductory guide, we have already discussed the various methods for querying the database. Listed here is an explanation of the operators and their use.

The Operators

ISYS offers the full range of Boolean (logical) and proximity operators. ISYS also provides a special type of truncation known as Conflation, which is activated by appending a tilde (~) to the search word. This feature is an intelligent form of "wildcarding" which applies standard rules of tenses and spelling to only retrieve related forms of the same "root word". By entering "POST~", for example, ISYS would match with POSTED, POSTS, POSTING, etc., but would not match with POSTGRADUATE or POSTULATE. By comparison, using the standard wildcard (*), e.g. "POST*", any word commencing with the characters "POST", including POSTULATE and POSTERN, would be matched.

The Operators are:

- AND means 'this ... and ...'
Documents retrieved will contain BOTH the first and second search terms, in either order, anywhere in the document.
- OR means 'this ... or ... '
All documents will be retrieved which contain either the first or the second search term.
- NOT means 'this ... but not ...'
Documents retrieved will contain the first search term but not contain the second.
- EXCEPT The EXCEPT operator is a more specific version of NOT and is used to locate documents in which the first search term appears, but only if the second search term is not in the same paragraph as the first.
- IN The IN operator is used to locate documents which contain the first search term in a paragraph that starts with the second search term, e.g. JONES IN AUTHOR only locates documents where the word JONES is in a paragraph starting with the word AUTHOR.
- LABEL This operator is used to find documents which contain a paragraph that starts with the search term.
- // means TERM1 in the same paragraph as TERM2
Documents retrieved will contain both search terms in the same paragraph in either order. Documents that contain both phrases, but not in the same paragraph, would not be retrieved.

- \ \ means TERM1 within a paragraph of TERM2
Documents retrieved will contain both search terms in the same paragraph as, or within a paragraph of each other, on either side.
- ... means TERM1 followed anywhere by TERM2
Documents retrieved will contain both search terms but only when the first term precedes the second term in the document.

The "/" and "\" operators can also be used far more precisely. For example "/10/" means within ten words of, "\10\" means within ten paragraphs of, "/-10,30/" means from ten words before to 30 words after, and so on.

Precedence of Operators

ISYS will assume a certain order of precedence, that is, an order for resolving portions of a query, unless told otherwise. The default precedence is:

1. Phrases
2. //, LABEL, IN, EXCEPT, TO
3. \ \
4. ..., AND, OR, NOT

Where a search expression, or portion of an expression, contains several operators of the same precedence, ISYS will process the operators from left to right.

When you enter a command-based search, you can control the order of precedence by using parentheses (brackets). Just as we use intonation in speech to clearly express a sentence that could be misinterpreted, we use parentheses to clearly express a query, overriding the ISYS defaults where necessary. For example,

(TERM1 OR TERM2 OR TERM3) IN TERM4

would retrieve all those documents where ANY of the first three terms appear in a paragraph starting with TERM4.

Special Symbols

There are a number of special symbols which change the meaning of the commands entered. They are -

- * Is similar to the DOS wildcard character.
The wildcard can be used anywhere within a word, but only once per word.
- ! This is an alternative to the "*" operator above.

- ~ This is the conflation operator, already discussed previously, and may be used at the start or end of the word
- + This operator temporarily reverses your synonym expansion setting. By default, ISYS automatically searches for synonyms, which are words or phrases with similar meaning. You tell ISYS which words and phrases are similar by establishing synonym rings (through the Preferences menu). By appending this operator to a search term, you tell ISYS to turn the synonym search off for this term. Conversely, if you have chosen to reverse the default setting so that ISYS never automatically searches for synonyms, this operator is used to instruct ISYS to search for synonyms for this term.

Common Words

Common words (often called "noise words") are words which appear frequently in documents but convey little or no information. Classic examples are "a", "the", "at", "it", and so on. ISYS does not index these words, in order to save disk space used by the indexes. There are 317 common words ISYS loads as a default on installation, however this list can be modified by you as necessary.

If you use a common word in a query, ISYS automatically reconstructs the search taking into account the proximity of the surrounding terms. If you use a common word in a phrase search, such as -

Beside the road

ISYS will turn this into a search for "Beside" followed two words later by "road".

Reserved Words

All the operators listed above are reserved words. To include them as a search term, rather than as an operator, prefix the word with an insignificant character (see below), for example -

Postage AND _label

Also, the '/' and '\' characters are operators. If these have been defined as significant, and you want to search for them, enclose them in double quotes. This prevents ISYS from interpreting the string.

Significant/Insignificant Characters

"Significant" characters are characters which convey meaning and are an important part of a word. For example, all the letters A to Z, in both upper and lower case, and the international character set, are automatically in this category. The numerals 0 to 9 are also significant, unless you change the list or choose to have pure numbers treated as common words. You may want to specify other characters to suit your purposes, for instance, the dollar sign ('\$') if you use a lot of documents that contain prices.

"Insignificant" characters are characters which are part of a word but are not regarded as important. Insignificant characters are treated as if they are invisible. For instance, if the hyphen character '-' was made insignificant, then hyphenated words would be treated as one word, e.g. "nucleic-acid" would be indexed as "nucleicacid", and a search for "acid" would not retrieve this.

Any character that is not defined in either of the above categories is regarded by ISYS as a punctuation character. These are deemed to indicate the end of one word and the beginning of the next.

Querying within the Query Results

You can make a query within the current query's results, or the previous query's results. That is, the query will only search the documents which were found in the current query's results or the previous query's results.

This feature is located within the Tools menu. ISYS uses the query method you last used to run the new query. In effect, this new query is ANDed onto the end of your previous query. The earlier query is not run again, however, so the search time will normally be less than running the whole query again.

Other Query Functions

There are a number of functions available to you from the Browse window when you select some text by blocking it with your mouse. These are:

- Making another Query
- Making a Hypertext Query
- Copying text to the Windows Clipboard
- Copying the entire document to another file
- Gathering text to the Windows Notepad
- E-mailing the selected text
- Faxing the selected text
- Printing the selected text
- Activating the document's application (discussed below)

Most of the above functions should be self-explanatory, but the first two may require some explanation. If you block a word or phrase and then click on the "Q" button, ISYS will start another instance of itself and use the blocked text as the search term for another search. Therefore, it is not wise to block too much text in this way, as the meaning of the search may be lost in the process. Conversely, the Hypertext Search function "massages" the blocked text and executes another search for "similar" portions of text in the database. It does this by Conflating (see above) each term in the block, then inserting the "within a paragraph of" operator between the terms, thereby effectively searching for sections of text where the same underlying ideas are closely juxtaposed. Once again, highlighting too much text for this type of Query can result in very confusing results, so be selective.

PLAIN ENGLISH QUERY

~~~~~

Plain English Queries save you the process of formulating a query expression using proximity operators, Boolean operators, and all the other features ISYS has to offer.

With a Plain English Query, you simply type a brief question or statement, and ISYS will attempt to satisfy your needs. Plain English Query isn't guaranteed to work perfectly 100% of the time, but you can reasonably expect it to work correctly 95% of the time. On those other rare occasions where it misinterprets your needs, you may have to reword your request, just as you would with a human assistant who did not properly understand what you were asking.

The full ISYS product contains detailed information on how you can 'teach' ISYS more about the type of information you have in your database, to make Plain English Query even more accurate and responsive.

## ANNOTATIONS

~~~~~

You can attach annotations to a document with ISYS in the IQ Browse window. Annotations can be text notes, graphic images, or hyperactivities. Hyperactivities are applications which you run from ISYS.

Inserting an annotation is done by choosing one of the Annotations buttons in the Browse window, then clicking on the place in the document where you want to insert the annotation. From left to right they are -

- Text Note
- Graphic Image
- Hyperactivity
- Query link
- Hypertext link to another document

Annotations are stored in a separate file from the document, with the same document name and an ".ANN" extension, in the same directory as the document.

To insert a text annotation, select the Text Note Annotation button, position your cursor to the desired location and enter your text. Once entered, text annotations can be edited, moved or deleted by first selecting the Annotation button, then clicking on an existing annotation symbol. To view an annotation, simply click on the annotation symbol in the text of the document.

To insert a graphic image, choose the Graphic Image Annotation button. Click on the place in the document where you want to insert the graphic image. ISYS shows you the Create Graphic/Image Annotation. Select the directory where the graphic file is, and select the file you want. ISYS shows you the selected graphic in the small box on the right. Choose the OK button and ISYS returns you to the Browse window.

If you insert a 256 color graphic, and you only have a 16 color display, ISYS will display the graphic in 16 colors. To display 256 colors, you may need to update your Windows video card driver, or the video card itself, or both. To view an image, simply click on the graphic annotation symbol in the text of the document.

You can also associate a single image with an entire document by including an appropriate image file in the same directory as the document, and with the same DOS filename as the document, but with an appropriate image file extension.

For example, you may have a document C:\NEWBLDG\PLAN4.DOC. If you include a graphic file in the same directory with the same name, e.g. C:\NEWBLDG\PLAN4.BMP, ISYS will include the View Image item on the Browse menu bar when you view the document. Click on the View Image item to see the image.

File formats supported -

- BMP (Windows bitmap)
- TIF (Tagged Image File format)
- GIF (CompuServe's Graphic Image Format)
- TGA (TARGA files)
- EPSF (Encapsulated Postscript format)
- WMF (Windows metafile)
- PCX (PC Paintbrush)

A hyperactivity is an instruction to ISYS to activate a program you define, when the hyperactivity symbol is clicked in Browse.

Hyperactivities can be used to activate just about any program that can be run under Windows. In a document that discusses a spreadsheet, for instance, you might use a hyperactivity to activate the spreadsheet application, and load the spreadsheet.

To insert a hyperactivity, choose the Hyperactivity button, click on the place in the document where you want to insert the hyperactivity link. ISYS shows you the Create Hyperactivity Annotation window. Enter the settings you want, and choose the OK button to save the changes.

The settings are -

"Command to execute..." - You can either put the name of a program, or the name of a data file. If you enter a data file name, Windows will activate whatever the default application is associated for a file of that extension type.

"Directory to make current" - The current directory is the working directory when the application is opened. This field is optional, however some programs may require a particular directory to be the current directory.

If you want to pass any keystrokes to the program automatically when it starts, enter these into the "Keystrokes to emulate" field. These keystrokes will be passed to the program as if you had pressed them.

You can send any keyboard key (e.g. A...Z, 0...9, etc.), and you can send function keys by enclosing them in curly braces, e.g. {F10}.

You can also send the value of DOS environment variables. Send the environment variables by enclosing their name in curly braces, e.g. {UID}.

WORD PROCESSOR ACTIVATION

~~~~~

You can activate the document's application (e.g. the word processor which created it) with the document automatically loaded - ready for viewing, editing, or printing. You can activate a document either from the Document List window or the Browse window. The "WP" button is active in both cases. From the Document List window, you will need to highlight the appropriate document name before choosing the "WP" button, whereas from within the Browse window the current document will be activated when you choose "WP".

If ISYS cannot open the application for the document, it uses an activation fallback strategy to activate another application that can display the document. Further details on the activation fallback strategy are in the ISYS Reference Manual.

ISYS can automatically install a special macro and button which allows ISYS to be activated from within Word for Windows. When you first run Word for Windows from ISYS, ISYS will ask you if you want to install the macro. Answer Yes or No accordingly. If you choose yes, ISYS will activate Word for Windows, and install the macro into NORMAL.DOT. A query button will be added to the Word toolbar. To use the macro, select some text, then choose the Query button on the Word toolbar, or select the Edit menu, then the ISYS Search item. ISYS will automatically conduct a search on the word or phrase you selected. Word macro queries are run as hypertext queries.

## DATABASE MAINTENANCE

~~~~~

The database (the ISYS index files) is the most important part of ISYS, and its correct maintenance is very important.

Database maintenance is carried out with the ISYS Utilities program. An Update scans the documents on your hard disk and determines which ones have been changed, added, or deleted since the last update. It then updates the indexes so that they accurately reflect the content of your documents.

ISYS does not index all documents every time, it just makes whatever changes are necessary to bring the indexes into line with document changes.

A Reindex completely reconstructs the database files, and then performs an update to the index. You should perform a reindex if the database becomes damaged or corrupt, or if you change the document title options, the common words, or significant or insignificant character settings.

A Rebuild is a special type of reindex that optimizes the database for size and speed of access. This function is of most use to database publishers who need to squeeze every spare byte out of a database, but is not recommended for the average user with volatile data, as it takes as long (and sometimes longer) to Rebuild the indexes as it does to Reindex, and the gains are usually marginal.

NETWORK OPERATION

~~~~~

ISYS is developed to be LAN-ready and LAN-independent. Its network licensing is based on the principle of "concurrency", which means you simply license for the maximum number of simultaneous Query users, and ISYS takes care of the metering of the licenses internally. In other words, if you anticipate that 10 users are likely to be using the ISYS Query program at the same time, you simply purchase a 10-license "pack". As each user activates the Query program, ISYS first checks that there is an "unused" license available, then reserves that license for as long as the user has the program loaded. If the user exits the Query program, ISYS releases that license into the "pool" for another user's use.

Your network must support the extended modes of the DOS Open call, which allows files to be opened for sharing. Most popular LAN products support this call.

### ISYS Auditor

While normal LAN security does a fine job of limiting access to particular files and directories, sometimes you may want to let users have access to information, but also monitor what information they actually look at. This is what ISYS Auditor does.

Using Auditor, you can nominate that various events should be logged, either to a single central log file, or to a separate log file per user. You can log any or all of the following:

- Queries
- Results of queries
- Documents browsed
- Errors
- Annotations
- Activations
- Starts, stops, loads and unloads

This facility can let you monitor such things as who is looking at what, who your key users are, who needs additional training, and how often ISYS is used. It can also provide a convenient way of checking whether your license count is sufficient, by monitoring how frequently users are being denied access due to all licenses being in use.

The information captured into the log file is recorded in industry standard ASCII format which can be imported into virtually any spreadsheet or database product.

## DATABASE PUBLISHING

~~~~~

You may want to use ISYS for database publishing purposes, building an ISYS database centrally, then distributing the index files and text files to remote sites or other clients who have a licensed copy of ISYS.

To distribute a database, you have a number of choices as to what is to be sent out -

- all the documents and complete database index files;
- all the documents, and let the remote user create the index;
- only changed documents, and new database index files;
- only changed documents, and let the remote user update their index.

So long as the remote sites have a full copy of ISYS, the database index files do not necessarily need to be distributed, as the remote sites can create the database and update it when they receive the new set of documents.

ISYS Publisher

ISYS Publisher is a separate utility product which provides you with facilities to help you distribute your document database, with or without indexes. You can purchase the Publisher software from Odyssey.

The Publisher package consists of two programs -

IPUBLISH - This is used at your central site. It parcels up the changed and new documents into a single compressed maintenance file that you can distribute to your remote users. ISYS Publisher automatically detects which documents have been added or changed since your last publication run, so you do not need to manually track your changes.

IUNPACK - This is used at the remote sites. This takes the maintenance file and unpacks the new and changed documents. It will also remove obsolete documents as required.

Note: This product is not a replacement for LAN licenses. It is designed to assist in synchronizing central and remote copies of the same database, and is typically used by commercial publishers or corporations with multiple locations, each with their own ISYS installation.

THE APPLICATION PROGRAMMING INTERFACE

~~~~~

An ISYS Application Programming Interface (API) Toolkit is available. This enables you to incorporate ISYS features in your own programs. It embodies the complete text indexing and retrieval engine of ISYS. It is the same engine that Odyssey Development has used in its DOS product for more than four years, and therefore is a mature product. The API is implemented as a DLL under Windows, and an LIB under DOS.

If you are interested in licensing the API for inclusion in your products, please contact Odyssey Development for further information.

## SPECIFICATIONS

~~~~~

Maximum No. of documents	1,048,575
Maximum No. of paragraphs per document	65,535
Maximum No. of words per paragraph	65,535
Maximum No. of words per document	2,000 Million
Maximum No. of different words	2,000 Million
Maximum No. of indexed word references in total	2,000 Million
Maximum No. of retrieved references per Query	2,000 Million
Maximum No. of MS-DOS directories	1,048,575
Maximum significant word length (bytes)	20
Maximum file name length (bytes)	64
File formats supported	32
Search operators	10
Synonyms	Yes
Conflation	Yes
Truncation	Yes
Boolean operators	Yes
Proximity operators	Yes
Precedence operators	Yes
WP Activation	Yes
Dynamic Hypertext	Yes
Graphic Image Links	Yes
Annotations	Yes
ZIP file support	Yes

* With this abridged version, only the first 2,000 words in each of 100 documents can be indexed.

SAMPLE APPLICATIONS

~~~~~

If you use a word processor, you must have noticed that there is no effective way of finding information stored in your files.

No effective way until ISYS.

The most common usage of ISYS is among professionals who create valuable information, and have a need to access that information some time in the future, rather than just print it out and forget about it.

This includes Government Agencies, Law Firms, Authors, Communications Bureaus, Newspapers, Academics, Tech Writers, Journalists, Managers and Executives, Publishers, Analysts, Ministers of Religion, Accountants and PR firms. In fact, ISYS is targeted squarely at people who use word processors, and so the more you use a word processor, the more you need ISYS. Here's a brief sample of what some ISYS users are doing:

- \* Lawyers and Judges index daily court transcripts and use ISYS in the courtroom to instantly find evidence from previous days of the trial.
- \* Organizations replace or augment their hardcopy corporate procedures manuals with an ISYS electronic version, so all employees can immediately find answers.
- \* Law enforcement agencies index their "call sheets" and other notes, and use ISYS to find connections.
- \* Accounting firms index industry regulations, and use ISYS to quickly determine which apply to the job at hand.
- \* Librarians index abstracts, and no longer have to worry about keywords.
- \* Technical departments index their support records, bug reports, news items and activity reports, and use ISYS to share knowledge and experience between staff.
- \* Organizations publish their internal documents electronically, thereby giving users the ability to access the documents efficiently.
- \* Placement agencies index personnel resumes so that people may quickly be found to meet the combination of skills and experience required by the client.



## COMMON QUESTIONS

~~~~~

- Q. How much memory does ISYS need?
A. We recommend that ISYS have access to at least 1MB of Windows workspace, over and above that required by Windows itself.
- Q. Does ISYS run on a Local Area Network?
A. Yes. More than one person can use ISYS Query on a network. Call Odyssey Development for details.
- Q. How does ISYS alter my documents?
A. It doesn't. It simply builds indexes that refer to your documents. ISYS will NEVER change one of your original documents.
- Q. Does ISYS support more than one database?
A. Yes. Full details of setting up multiple databases are given in the ISYS manual, and it also possible to query across multiple databases simultaneously.
- Q. Can ISYS handle more than one word processor?
A. Yes. ISYS allows any and all of the supported Word Processor formats to be intermixed in a set of documents.
- Q. How much disk space does ISYS consume?
A. About 1.5 Mb for its program files, and the indexes normally occupy 20-25% of the space used by the original documents.
- Q. How can ISYS index structured dBASE III files?
A. ISYS indexes either the memo fields, or the full content of each record. More details are available from Odyssey Development. ISYS can also index Microsoft Access.
- Q. Can ISYS index documents spread over multiple disk drives?
A. Yes, you just tell it which disk drives.
- Q. How does ISYS link to graphics?
A. ISYS can be set up so that each document file corresponds to a single related image file, or multiple images can be specifically linked to documents at chosen points (as annotations).
- Q. Why do I see some strange words in my spreadsheet files?
A. When ISYS indexes spreadsheets, it just looks for and indexes the "wordy" bits of the file. Sometimes this will pick up parts of formulas and other unwanted portions. If you use the ISYS DDA feature to "activate" the spreadsheet, it will appear perfectly.
- Q. Is there a DOS version of ISYS?
A. Yes, and the two versions are fully compatible, even sharing the same indexes. The DOS version can optionally be run as a TSR, providing pop-up access from within other applications.

INTRO DISK DISTRIBUTION POLICY

~~~~~

Odyssey Development grants the freedom to duplicate this disk for the purpose of enabling others to also learn about ISYS. However, it is required that the entire content of this disk is distributed intact and unmodified, including this documentation file.

ISYS, the contents of the Intro disk, and this documentation are all (c) Copyright 1995, Odyssey Development.

## GETTING HELP

~~~~~

Help is available from within ISYS by two methods. You can choose the Help index from the Help menu, or use the context-sensitive help function - just press the F1 key at any time to see information about the current feature being used.

Free technical support is available 9:00 a.m. to 5:00 p.m. Call:

Sabre Imaging

01932 252782