The Basics of Searching

The Electronic Database:

All electronic bibliographic databases, like CAB ABSTRACTS, can be thought of simply as very large electronic journals. Although you can't physically see the records and the indexes, en-mass, within the database, the two are structured in the same way. They each have a section that contains the records as well as a separate section that contains the indexes that allow the user to find or "search" for the records in which they are interested. In a typical printed abstract journal, such a Forestry Abstracts, published by CABI, the journal will have a section containing all the records as well as an Author and a Subject index at the back. In the case of the printed journal, the Subject index contains an alphabetical list of index terms that describe the key concepts within a record. These index terms, which may be single words or a descriptive phrase, are accompanied by a list of record numbers which allow the user to find all the records that have been indexed with that term. An electronic database works in exactly the same way. It has a section that contains all the records, each with its one unique record number, and a section that contains all the indexes used to search for those records. The one major difference is the size of the electronic database and the large number of indexes that can be created without the limitations that a print journal imposes.

In the case of the CAB ABSTRACTS database, for example, there is an index of Title words, an Author index, a Source index, an Address field index, a Language index and several Descriptor indexes. In addition to these individual indexes, which each correspond to an individual field within a CAB ABSTRACTS record, there is also what is called a Free-Text Index which combines many of these individual indexes into one large index and which allows the user to search for words and phrases from the complete record. The Free-Text index is the one that is most commonly used for quick and simple searches and is the index that is searched by default. Like the printed index, it contains an alphabetical list of terms with record numbers alongside, as in the example shown below:

CATTINI-C	**12, 35,57**
CATTINI-PA	**8, 11**
CATTLE	**1,2,5,7,8,11,14,17,21**
CATTLE-BREEDING	**2,5,8,17**
CATTLE-BREEDS	**2,5,11,20**
CATTLE-DISEASES	**1,5,14,21**
CATTLE-FARMERS	**7,12**
CATTLE-FARMING	**7,11**
CATTLE-FARMS	**7,14,20**
CATTLE-FEEDING	**11,14,21**
CORN	**6,9,13**
CORN-FLOUR	**6,9**
DESAI-VG	**22**
DESALINATION	**4,16**
FLOUR	**6,9**

The use of such electronic search indexes is the reason that searching of even very large databases is so quick. Words may appear many thousands of times in the database records, but they only appear once in the search index along with reference numbers to all the records in the database that contain that term. When we search for a term (keyword) in one of the search indexes, the system simply looks for the term in the index and, when it finds it, it counts the number of record numbers and returns a result to the search screen that says something like:

#1 7553 Rice

The exact format of the returned results depends upon the search interface being used. This is an example from CABCD where **#1** is the number of the search, the **7553** is the number of records that the search has found, and the word **Rice** is the term that we have searched for. The reference numbers, within the search index, are used to locate and display the records on the screen.

Keywords and their Selection

One of the most important aspects of searching is the selection of all the appropriate search terms or Keywords, as they are often known, and the building of the search profile. The steps to creating a search profile are as follows:

1. **Understand the question:**
   Write down exactly what information you or the end-user wants. If you are performing a search for someone else, make sure that you are absolutely clear about what it is they want.

2. **Select all the important concepts:**
   From the search question, select all the search terms. A typical search question might be:

   “I want 50 recent references about the irrigation of rice in South East Asia.”

   Inexperienced searchers will often try and search for the phrase “**irrigation of rice in South East Asia**”. If you do this, the system will look only for records that contain exactly that phrase and it is likely to find very few records if, indeed, it finds any at all. Instead, we need to select out the different concepts and search for each separately. In this example, the key concepts are:

   Irrigation, **Rice** and **South East Asia**

   What we now need to do is to instruct the search system to look for records that contain all three of these concepts or “Keywords”

3. **Add synonyms, singulars and plurals:**
   Once you have selected the keywords from the search question, you should consider if there are any alternative words, or spellings of those words, that you should include in the search. Remember, the system only looks for what you ask it to. Typical examples are singular words or plurals, American versus British spelling, common names like Rice or scientific names like Oryza. The more variations that you can add to the search profile, the more records you are likely to retrieve, particularly when performing a FREE-TEXT search. Many databases will use a controlled indexing vocabulary like the CAB Thesaurus and, if this is available, it can be a valuable source of Keywords and alternative terms. For singulars and plurals, most search engines offer a technique known as truncation. You simply enter the singular term followed by a truncation symbol (usually an *, a $ or ?) and this symbol tells the search engine to look for all words that start with the term entered. As an example, searching **Nutri***, on CAB Direct, would retrieve all the records containing the words **Nutria, Nutrient, Nutrients, Nutrition** and **Nutritional**. On CAB Direct, he Question Mark (**?**) can also be used for truncation where each **?** represents up to one character. For example, searching for **CAT?** Would retrieve records containing the word **CAT** or the word **CATS** but not the word **CATTLE**. The **?** may also be used for character "masking" within a word, e.g. **Int??net** would search for **Inernet** or **Intranet.**

Creating the Search Profile

When searching, each time a term or group of terms is searched, the system creates and stores a set of the records that it finds. In the very simplest of searches, the user may only create one set of records and then print or display them. However, in most cases, the search will need to be a little more complex than a single keyword. Searches for more complex subjects that contain many keywords will require the creation of a more complex search profile, or search strategy, and will usually result in the creation of many sets of records, where these sets themselves are then combined to obtain the final result. Let's look at our example of Irrigation of Rice in South East Asia

We have already selected our three keywords as:

Irrigation, **Rice** and **South East Asia**

In a simple search, we would search for each term separately. This would result in the creation of three separate sets as in the following example:

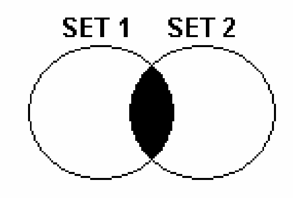
#1	**20972**	**irrigation**
#2	**27332**	**rice**
#3	**20072**	**south east Asia**

This is the "Search History" and shows the three sets of records, the numbers of records found for each of the terms searched and the terms themselves. In order to complete the search to find the records that contain all three of these search terms, we now need to combine these three sets together.

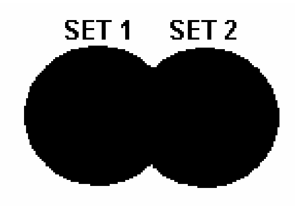
Combining Sets of Records:

Sets of records are combined together using one or more of the three search operators **AND**, **OR** and **NOT**. These three words are often referred to as "Boolean Operators". So what do these words do?

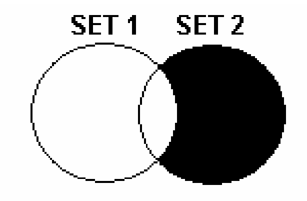
AND is used to instruct the search system to retrieve only records that contain all the terms chosen. In our example, we want only records that contain all three search terms **Irrigation AND Rice AND South East Asia**. We don't want records that only contain **Rice** and **Irrigation**, for example. The **AND** operator narrows a search. It reduces the number of records but increases the relevance of those records.



OR is used to broaden a search; to make the set of records bigger. An example might be a search for all records about sheep or goats. Here the user wants all records that are just about goats plus any records that are just about sheep as well as all the records that might be about both sheep and goats. Here the search would be **Sheep OR Goats**



NOT is used to exclude certain terms from a search as in a search for records about Rice where the user is not interested in papers about the diseases of rice. Such a search would be performed as **Rice NOT Disease***. Note here the use of the truncation symbol, mentioned earlier, which selects several words that start with Disease (e.g. Disease, Diseases and Diseased).



These Boolean Operators can be used directly with keywords, as in these examples, as well as set numbers. If we go back to our example, the final search would be to combine the three sets with the **AND** operator:

#1 AND #2 AND #3

Note: In the case of a search on CABCD (Ovid's version using the SilverPlatter search interface) you must include the **#** symbol, otherwise the system would simply search for records containing the number.

This final search creates a set of records, **#4**, which contains all the records that include all three of our search terms. So, to finish, let's take a look at the complete search history:

#1	**20972**	**irrigation**
#2	**27332**	**rice**
#3	**20072**	**south east asia**
#4	**197**	**#1 and #2 and #3**

We have found 197 records about the irrigation of rice in South East Asia. Once the search has been performed and you can see the records, it is a good idea to check the records for other possible keywords that you may have missed in your original profile. Because you are able to create sets when searching, it is perfectly possible to continue to modify a search by adding in new Keywords, and creating and combining new sets of records until you are satisfied with the results that you have obtained.