*Chapter Ten*

# Types of Legal Database

In the early days of the development of databases, most of them consisted of codified information that had a formal structure and the free-text databases were usually short abstracts, which described a document. While useful, these databases were not sufficient to satisfy the legal practitioner who required the full text of a document. Improvements in technology, particularly cheaper storage devices, enabled large documents to be processed and stored for ease of retrieval. Now, most databases are a mixture of structured information and full text.

A structured database is designed with knowledge of the questions that will be asked. This means that such databases are limited in their scope, but sufficient for a specific purpose. Examples include mailing lists, catalogues, stock lists and personnel records. In contrast, a free-text database is designed so that any question may be asked of it at any time. In countries with common law jurisdictions, such as the UK and the USA, legal databases began with free-text designs. In countries with a codified system of law, such as France, legal database design began with structured methodologies. The limitations of each of these approaches became apparent as soon as legal practitioners put them to actual use. The result, which we have today, is the combination of the two approaches into one methodology, commonly called the "hybrid" approach. CELEX is a good example of a hybrid database.

Searches can only find what has been stored. The design of the database must never discard information that might be useful to the user. In every database there is a basic component which is familiar to the user, for example, a complete case report or a section of an Act. The designer must identify this component and design the database around this logical unit of information. Within that unit there will be further subdivisions, but the whole operation of the system will revolve around that basic component.

Take the example of a legal database of case reports. Each report has a structure in which the text is subdivided into sections, including:

- Title
- Catchwords
- Headnote
- Citations
- Judgment
- Counsel

Any of these subdivisions may be used in searching, but are subordinate to the case report as a whole. The user should realize that the size of the basic component will have an effect on the efficiency of the search techniques available. Restricting the search to one type of subdivision can speed up the search, improve precision, and be more efficient, but may recall fewer documents than expected. The use of the larger components will mean the retrieval of more documents, but some may be less relevant than others. The user has to exercise judgement about which method to use and that can only come with practice.

## *Search Languages*

When we first phrase a question, it is often imprecise, so we produce several versions before the question is sufficiently accurate to produce the answer we require. Likewise, when using computer search language, the best procedure is to work on the question. Once the right question is asked, the correct answer will be returned. The successive refinement of the question is the correct methodology.

The simplest query consists of a single word which, in a large full-text database, may return many documents if it occurs frequently. Often, a single word is a sensible search query since it will reveal if it is used at all, for example, a proper name or a technical term. However, it is more common to add another word. This is done by combining the two words using logical operators, for example:

***Search 1***

- *husband AND wife*

This will find documents that contain both of these words.

***Search 2***

- *husband OR wife*

This will find documents that contain either one of these words, as well as both of them.

***Search 3***

- *husband NOT wife*

This will find only those documents that contain the word husband and will ignore those that contain both words.

Note that these words can occur anywhere in the document. The search is satisfied if the word husband occurs in the first line of a document and the word wife in the last line. Such a search may not be particularly relevant, especially in a long document such as a case report, which may be a thousand or more lines in length. In a large database containing many lengthy documents, which is usual in a legal database, such a simple search would require further refinement in order to provide a truly relevant set of documents. The simplest way to develop such a search is to use the operator **AND** to add more terms:

- *husband AND wife AND custody AND financial AND provisions*

This would reduce the number of retrieved documents considerably. The use of the operator **AND** to combine a series of words is very effective and is probably the most common technique used in searching full-text databases. This type of search may be further refined by adding a condition of proximity to the search:

- *husband WITHIN 25 of wife*
- *husband NEAR wife*

These are all more precise than the simple *husband AND wife* search. These days, most retrieval systems incorporate these facilities within their search languages.

It is also possible to search for a whole phrase (e.g., "motor vehicle licence"). This makes searching even more precise but requires greater knowledge of the terminology used within the documents. It should therefore be used with care since there is a danger of overlooking relevant documents.

The **OR** operator is useful for finding synonyms:

- *child OR infant OR juvenile*

This will find documents in which any one of the alternatives occurs.

Combining words using the **AND** operator with the **OR** operator is sufficient for the majority of searches. When a mixture of different operators is used, it is good practice to use brackets to clarify the logic:

- *(air OR atmosphere) AND pollution*

This will find documents in which air pollution is mentioned and also those referring to pollution of the atmosphere.