

Managing Litigation Documents with Concordance

Putting Concordance into the Proper Perspective

Although it was initially introduced more than 20 years ago and has acquired a loyal user base, Concordance is still sometimes erroneously compared with traditional database management software. However, the simple fact is that Concordance cannot be compared in any respect to a typical, general purpose database management program. On the contrary, Concordance is a highly focused program that features sophisticated text search and retrieval functions specifically designed for managing the high volumes of data generated in litigation.

The primary purpose of this document is to provide the reader with greater insight into Concordance by clearly defining not only what it is, but also what it isn't. We'll start the process by looking at traditional database management and how it compares to text database management.

Traditional Database Management

As used in the computer industry, the term "database" typically refers to a computer-based storage area for a large amount of both general and specific information related to a particular topic. For example, a database of customer records can store basic information such as names, addresses, and phone numbers, but can also contain detailed information regarding past purchases payment histories, and more.

In this example, the standard database software package normally includes search and retrieval functions for locating specific customer records. In addition, it also provides the ability to cross-tabulate records as a means of acquiring more insight into the relationships between various data elements. While these database management products may possess some rudimentary ability to search for text/text strings, and can even include some file searching capabilities, they are incapable of performing detailed search and retrieval of very specific text information.

Text Database Management

Text database management differs substantially from traditional database management in that its focus is searching for, retrieving, and categorizing specific words, phrases, or combinations of words. The ability to perform detailed text searches is particularly important for individuals whose work is heavily text-concentrated, for example attorneys, scientific researchers, marketing researchers, librarians, and personnel managers for resume retrieval.

Applications

Any collection of text-based information that people want to access in an ad-hoc manner is a candidate for a full text database manager. Information that would be of value to a large number of corporate or institutional users is a great candidate for network placement. One typical application is the human resources department that not only needs to search through thousands of resumes to identify potential job candidates, but also wants to make the resource available directly to hiring managers throughout the company. Another example is the corporate legal department that needs to track litigation through large volumes of records. Drug companies require this type of system to maintain indexes to manufacturing status, lab and research notebooks, regulatory compliance, and adverse effects data. Other applications include CD-ROM publishing by corporations, foundations and agencies.

These specific examples illustrate ways in which Concordance is being used in the "real world":

- The Primate Information Center at the University of Washington uses it to publish the world's premier source of indexed scientific literature on primate research.
- The Medical Data Exchange in Mountain View, California uses it to synopsise and publish important medical articles for the layperson.
- Fabulous Footage in Toronto, Ontario, Canada uses Concordance to manage their stock footage library. They publish the library in Concordance format for their worldwide library associates.

Concordance: Sophisticated Text Search and Retrieval

In text database management, text search and retrieval is highly sophisticated in nature. Concordance provides users with an advanced, interactive language to locate information quickly and precisely. Searchers can specify individual numbers, dates, words, word fragments, phrases, specific fields, previous searches, or any combination. Each retrieved word is automatically displayed on the search screen showing the total number of occurrences and documents located. The display shows users both what was found and why it was found. Modeled on powerful mainframe text retrieval systems, Concordance combines the speed and capacity of the large systems with the ease of use and flexibility found in personal computers.

Theoretical Overview: Full Text Information Retrieval

There are two basic measurements of text-retrieval effectiveness: recall and precision. Recall is related to the total number of documents retrieved by a given search out of the total number of documents in a database that should have been found. Total recall would locate every relevant document in a database, but this may include unwanted documents as well. Precision is the percentage of documents retrieved that the searcher is actually interested in. Total recall with perfect precision is the ultimate goal.

So, the search term disk may locate all documents that refer to the phrase disk drive, but it would also locate extraneous documents that mention disk brakes. By refining the search to disk drive specifically, we would lose any documents whose authors chose to use the term floppy disk or hard disk to mean disk drive. The more accurately a user can specify a search, the greater the precision in the retrieved documents; however, recall will have been sacrificed. Generally speaking then, as precision increases recall decreases because words and phrases take on different meanings when used in different contexts, patterns, or even different fields within a document.

Overall, the effectiveness of text information retrieval systems is therefore dependent upon three things:

1. The uniformity of the language used in the database
2. The experience and familiarity of the searcher with the data
3. The quality of the searching tools provided by the search software

A uniform database and knowledgeable searchers still need a powerful retrieval system to be effective. The usefulness of any full text information retrieval package, measured in terms of recall and precision, becomes a function of the query language and the data structures supported by the software.

Understanding the Mechanics of Text Retrieval Systems

Until recently most micro-based retrieval systems have supported a rudimentary set of search operators, typically the Boolean operators and, and not, and a little more. The Boolean operators, named after British Mathematician George Boole, allow retrieval systems to combine documents into sets based on the combinations of the words they contain. For example, the set of documents that contain both Ferrari and Coupe, or the set of documents that contain disk not brakes.

Boolean logic provides a precise way to specify the target documents based on syntactic content. Concordance supplies an important enhancement in this respect, as it allows searchers to build a set of queries, each automatically assigned a number as they are entered into the system. The individual searches can then be combined to create a still higher level search, giving additional flexibility and precision.

There has also been a tendency to view textual data as a single unit. That is, all text was stored in a single field of a one field database and no accommodation was made for the natural structure of language. Concordance was designed to accommodate text as it exists, in any format. It also supplies a rich set of search operators that provide unparalleled flexibility in its search language.

Concordance: Search Feature Overview

Fuzzy Logic

The Fuzzy logic feature enables to perform what is termed a "fuzzy" search. This type of search enables the user to look for words that sound alike, or are spelled like a particular word that the user has highlighted. Concordance then displays a list of best guesses.

Thesaurus with Concept Searching

The expanded thesaurus includes both synonym substitutions or expansions that are found in other programs, but also offers a more powerful concept searching feature. Concept searching will perform multiple searches, or can be programmed to execute multiple queries based on any keyword found in a particular search. For example a simple synonym search for the word "rainbow" would retrieve words like red, green and blue. With a concept search, however, the user could conduct a search under the conditions forecast = rain and time = daytime to uncover the potential for a rainbow.

Pop-up Dictionary

With the click of a mouse, a dictionary dialog box appears on the screen that allows users easy access to the words in their text database without having to conduct a full search.

Search Operators

Concordance has a total of twenty search operators, grouped into four functional areas: context, proximity, Boolean and relational. The proximity operators, adj and near, allow the searcher to find words that occur near each other, from a simply adjacency to a distance of ninety-nine intervening words. The searcher can specify that the words must occur in a specific order, adj, or in any order near. The context operators will exclude or restrict a search to a specific field in the database, such as the author or address field. Using these operators, Concordance operators can attain total recall with perfect precision in many instances.

Exclusive XOR Operator

Concordance provides four Boolean operators, or, and, not, and xor. Or, and, and not are seen in most packages in the market. However, the exclusive-or operator, xor, is rarely seen in even mainframe retrieval systems. The search wolf xor sheep would locate all documents where wolf and sheep occur singly, that is in documents that contain one of the two words but not both. It is an operator that has been overlooked by most developers, yet it is the easiest way for searchers to separate the wolf from the sheep and the sheep from the wolf - not always the same thing.

Relational Operators

Relational operators, such as =, <, and >, are provided to aid in searching formatted fields: date, numeric, and text. Concordance is virtually alone in providing formatted field search capabilities in a micro-based information retrieval system. With Concordance, a searcher could locate all documents written at a particular company, restrict them to a specific date range, sort them by author and date in ascending order, and print out a complete chronology in minutes.

Implementing Concordance: A Hypothetical Case

An excellent application opportunity for Concordance would be a situation in which a law firm, in preparation for a deposition, needed to search thousands of pages of corporate data for each time that a former employee's name appeared on either an internal memo, or some form of external correspondence.

In this instance, the law firm could choose to assign a paralegal employee to physically search through every hardcopy record provided to the firm. The time required to complete such a process, as well as the resulting cost would be substantial. To optimize time and minimize cost, the firm could opt to utilize the extensive capabilities provided by Concordance.

Specifically, the firm could choose to contact a VAR that specializes in document imaging in order to scan all pertinent documents as TIFF images into a computer. The VAR could then transform these images into text files using specialized conversion software, known as OCR. Once the scanned documents were converted to text files, Concordance could then be used to complete the detailed search and retrieval process in considerably less time, and for significantly less money. This type of scenario is acted out every day in many law firms throughout the world that have standardized on Concordance.

Concordance, The Ultimate Full Text Database for Litigation

Concordance has all the features needed so that you can find facts fast and analyze with precision; an extended search language with twenty query operators, support for fixed and free text fields, automatically numbered and recallable query sets, the ability to load existing databases, fully integrated word processing, and a powerful programming language for customized applications. Moreover, it also supports fuzzy logic, a unique and very valuable feature which enables searching techniques that locate key words even when they are misspelled or mistyped. When you add to that its industry-leading database capacity, fast and accurate text retrieval, and easy to use working environment; you can see why Concordance is the discovery software of choice.