CASurveyor[1]

Y. Wolman

Department of Organic Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

Received February 14, 1994

An important and interesting addition to the growing number of chemistry related CD-ROMs is the CASurveyor set. At present, the set is composed of five CDs, one disc for each of the following subject areas: Cancer Chemical Research, Chromatography, Food and Feed Chemistry, Magnetic Resonance, and Organometallic Chemistry. CASurveyor is available for both Microsoft Windows and Apple Macintosh platforms. I examined the Windows version, and although the system requirements to this version (according to the manual) are stated as "IBM compatible 286 or later, 10Mhz or faster, 80386 or better is preferred", I would not recommend any machine with less than a 33-Mhz, 386 processor, with a faster and higher CPU being better.

Each CD-ROM contains between 30 000 and 80 000 records from the Chemical Abstracts (CA) database, covering the present year and the past 2 years. The discs are updated every 3 months; thus, the fourth-quarter disc contains the complete data for the past 3 years. The coverage of each of the CASurveyor discs does not correspond to any of the other Chemical Abstracts Service (CAS) products. One can view each of them as a "super"electronic *CA Selects*, covering a wide, clearly defined subject area and using a well refined standard interest profile. Each of the records consists of a complete bibliographic citation, CA abstract, CA indexing terms, and CA chemical substance information data (CA index name, other chemical names, molecular formula, and chemical structure). All of the information (except the chemical structure) is text searchable, and hypertext links are incorporated between substances and document records.

The user interface is very simple and very friendly, enabling a user to build a complicated search strategy without any prior knowledge about Boolean and/or proximity operators, just by using a dialog box. I would recommend using parts or even all of the dialog box features in the Guided Search module in STN Express (in my opinion, this module was[2] and has remained the weakest module in STN Express). As a matter of fact, some of our faculty and students got so excited searching Surveyor that they even started searching the CA file on STN by themselves.

Searching Surveyor may be very similar to searching CA on STN using the Messenger language; however, there are a few differences in the structure of the database and in the searching language which greatly simplified the searching mechanism. The greatest difference is that all the text records (except the CAS Registry Number (RN) and the author name) are included in the Word Index (the equivalent of the Basic Index in the CA database); this includes words from the corporate source, corporate source location, patent classification, coden, publication year, etc. Specific indexes are also available: Atom, element count; Author, author names; CAS RN; CAS year, the years the record appeared in CA; Component MF, component molecular formulas from multicomponent substances; Compound, chemical names; Journal Name, abbreviated journal titles; Mol Formula, substance molecular formula; Organization, organizations; Patent Assignees; Corporate Author; Corporate Sources; Segment, substance name segments; Pub Year, publication years of the source documents.

There are two options for searching Surveyor: the Word Search and the Index Browse. Word Search enables the user to build a search strategy (a simple or a complicated one), by entering the various search term(s), their corresponding fields, and their relationship to each other using the dialog box. Index Browse enables one to retrieve terms using the index of a search field. Although it resembles the Expand command of Messenger, one cannot combine two or more index terms, as is done by using E numbers. Very few commands are used while searching Surveyor; they are entered by clicking the appropriate box in the main menu bar or the tool bar.

For some reason the wild cards of Surveyor and of Messenger are completely different. Personally, I would like to see and use the same wild cards when searching Surveyor and/or the CA file on STN: after all, we have one organization, the same database, or parts of it, and the same users; why not try to build systems as similar as possible?

The manual is easy to read and use; it is written in such a way that once you start reading it, you want to start searching even before you have finished reading. Unfortunately, there are a few typographical errors in the User Guide, as well as some minor bugs in the software. They do not interfere with learning and using the system, and I am sure that they will be fixed in the very near future. Besides the manual, there is an excellent online Help module that helps the novice, as well as the expert user, in each step of his search.

However, Surveyor has two weak points which I do hope will be taken care of as soon as possible. While the User Guide discusses, explains, and even lists the stop words (Appendix B) (there are 15 stop words defined by the Messenger software), nothing is mentioned about the CAS standard abbreviations. Only expert searchers of the CA file are aware of the existence, the importance, and the usage of these abbreviations. Just a few examples can show the role and place of the abbreviations in searching the various Surveyor discs. The term "liquid chromatography" appeared 3635 times in the Chromatography CD-ROM; there are 6766 hits for the term "liq chromatog" (liq and chromatog are the CAS standard abbreviations of liquid and chromatography, respectively). Furthermore there are two hits for the liq chromatog of hexamethylbenzene (CAS RN 87–85–4) but none for liquid chromatography of the above compound. There are 385 hits for decompn (it is the CAS standard abbreviation for decomposition), 161 hits for decomposition, and 465 hits for decompn or decomposition in the Food and Feed Chemistry disc. While there are two hits for the decompn of ethylene in the above disc, no hits could be located by searching for the decomposition of ethylene. There is an urgent need to explain the usage and application of the CAS standard abbreviations in the User Guide, as well as to list the various abbreviations in an appendix or add the CAS standard abbreviations booklet to the User Guide.

A second weak point is the system response when one of the search terms has zero postings (due to a misspelling, for