CODATA Referral Database

STEPHEN R. HELLER

USDA, ARS, BARC-W, Beltsville, Maryland 20705-2350

Received May 23, 1990

CODATA, The Committee on Data for Science and Technology, has just released its first computer readable product, the CODATA Referral Database (CRD).[1] CRD is an IBM PC database and retrieval system for locating sources of information, but not the data themselves, in the area of scientific numeric data. The database comes from a collection of CODATA database directories and the UNESCO Inventory of Data Referral Sources in Science and Technology. CRD allows one to rapidly search for data centers or other institutions which have said they can provide numerical data or other factual information in any specified area of science and technology. The results of a search are the names and location of each relevant organization, along with a description of its scope of coverage, holdings, outputs, and dissemination policies.

The current release of the system, Version 1, is dated January 1990. The CRD is similar in a number of ways to the Dictionary of Online Databases[2] in that it contains detailed records on what is in a given database. That is, both databases are referral sources. In the case of the Directory of Online Databases the databases are all available online, while the databases in the CRD are available in a number of forms.

The first release of the CRD database contains 1165 records. The search software is a subset of the MicroBIRD Full Text Retrieval System, developed at the Queen's University of Belfast.[3] The system comes on either 5-$^1/_4$- or 3-$^1/_2$-in. disks. I chose the 3-$^1/_2$-in. disk system and was able to install the program and database on IBM compatible PC hard disk systems in a matter of minutes using the brief, but fairly complete 19-page user manual (I installed the CRD on both an Epson and a AGI computer). The manual does have an excellent 5-page description of the CRD file structure with every field well described. This is followed by a 3-page sample file record. The manual does lack information on what error messages mean (e.g., Error 2 with file CRD.TXT), but this is a minor point. The ability to search on all fields in the 1165 records is very good, but one does pay the price of needing a little over 3.5 million bytes of storage on a hard disk in order to load the system.

Searching the database is very simple. One is able to search for a word from the entire record and have the full record or just the field in which the word appears given as the hit. Search hits can either be displayed on the computer screen, printed, or directed to a disk file. Search terms can be combined with the usual Boolean AND, OR, and NOT logic. The search term which caused the record or field to be displayed as a hit is highlighted in bold face type, which is a nice feature. In a sample search I did for the term or keyword "solubility", there were five hits, of which the first is shown in Figure 1. While the record is quite complete, it is a bit out of date: NBS, for example, changed its name to NIST (National Institute of Standards and Technology) in 1988. Also, one person listed as a contact died some two years ago. The quality control on the spelling in this record [there are two such errors in this particular record as well as the different spellings for the same word (centre and center)] and others needs to be improved. It would also be nice if the output was formatted a bit more clearly, with perhaps having one record per page, or at least have a new record start at the top of a new page. Also, on laser printers, the program does not automatically eject the last page of the printout. The information in all records, like the information in the Directory of Online Databases, is supplied by the lab or organization responsible for the database. This means that the accuracy of the information cannot be assured by CODATA. Considering the massive undertaking of this project in getting so many sources around the

```
Search #1

So far your search comprises....
   * 'SOLUBILITY '

This combination occurs 5 times .

MFN:  1161.

COUNTRY CODE:  usa.

COUNTRY:  UNITED STATES.

CATEGORY CODE:  a.

CATEGORY:  Science and Technology (In General).

SECCATEGORY:  m, Materials c, Chemistry and Chemical
Engineering oe, Other Physical Science and Engineering st,
Standards.

TITLE ORIG:  Office of Standard Reference Data.

TITLE ACR ORIG:  OSRD.

INSTITUTION:  National Bureau of Standards.

ACRONYM:  NBS.

MAIL ADDRESS:  Room A321, Physics building
               Gaithersburg MD 20899
               United States.

TELEPHONE:  +1 301 975-2200.

TELEX:  898493 GARG.

STAFF:  9, 1, 4.

DIRECTOR:  Lide, D.R.

CONTACT:  Sauerwein, J.

ACTIVITIES:  production; compilation; evaluation; referral.

INSTITUTION TYPE CODE:  Ga.

INSTITUTION TYPE:  Government Administration.

COVERAGE:  Physical and chemical properties at well-defined
substances. Emphasis given to thermodynamic and thermophysical
properties, chemical kinetics, atomic and molecular properties,
alloy and ceramic phase diagrams, and mechanical properties.

KEYWORDS:  thermochemical properties; thermophysical
properties; transport properties; spectroscopy; atomic data;
phase diagrams; corrosion; solubility; rate constants;
crystallography.

OUTPUT:  1, Publication of printed data compilations; 4,
Magnetic media containing data.

DESCR OF OUTPUT:  Evaluated data on physical, chemical, and
materials properties in printed and computer-readable formats.
10 databases now abailable in magnetic tape format.

DATABASES:  10 databases now available in magnetic tape format.

PUBLICATIONS:  J. Physical and Chemical Reference Data.

SERVICES:  2, Provision of specific data upon request; 3,
Referral of institutional or published data sources; 5,
Consultancy services on data centre activities.

DESCR OF SERVICES:  Publications and on-line services available
through various sources, magnetic tapes are distributed by OSRD.

AVAILABILITY:  2, Open to all users ; 3, Fees charged.

LANGUAGES:  eng, English.

REMARKS:  Indididual data centers coordinated by OSRD are listed
separately.

DATE:  22/08/1987.

CODE:  ks135.

TYPE: None.
```

Figure 1. Result of searching for records with the term "solubility". Only the first hit is displayed.

world to contribute, these few deficiencies are minor, especially since the database is in the process of being revised and updated; a new version will be available in the middle of 1991. The cost of this tool for searching for sources of scientific data is quite modest, only $175, and should make a useful addition to every science library throughout the world, in both industrialized and third world countries where information of this nature is not easy to obtain.

REFERENCES AND NOTES

(1) The CODATA Referral Database (CRD) is available from the CODATA Secretariat, 51 Boulevard de Montmorency, 75016 Paris, France. Telephone: 33-1-4525-0496; FAX: 33-1-4288-1466. The price of the CRD is $175.00.
(2) The Directory of Online Databases is available from Cuadra/Elsevier, 655 6th Avenue, New York, NY 10010. Telephone: 1-212-633-3980.
(3) The MicroBIRD Full Text Retrieval System is available from Professor F. J. Smith, Department of Computer Science, The Queen's University of Belfast, Belfast, BT7 1NN, Northern Ireland.

BOOK REVIEWS

Handbook of Polycyclic Hydrocarbons. Part B: Polycyclic Isomers and Heteroatom Analogs of Benzenoid Hydrocarbons. By J. R. Dias. Elsevier Science Publishers: New York. 1988.

All known benzenoid compounds, their polycyclic isomers, and heteroatom analogues are organized into isomer and isoskeletal groups in this handbook. The compounds are organized according to a formula periodic table which is based on the graph theoretical principles. Part A deals with hydrocarbons with fused hexanogal rings only. Part B presents compilations of experimental and theoretical properties of 150 polycyclic ring isomers and 250 heteroatom analogues of benzenoid hydrocarbons. Updated and additional data for Part A is included in Part B.

Part B begins by introducing the reader to the molecular graph theory and the isomer enumeration, which are fundamental for the organizational format. The first chapter gives a brief, but informative, introduction of the rules that form the basis of the nomenclature for the two groups; polycyclic conjugated hydrocarbons and heterocyclic compounds. This introduction leads to the chemical graph theory in the next chapter. The author has given a thorough presentation of the theory, the graph theoretical notations, and the terminology used for molecular graphs. However, a list explaining the mathematical symbols used in this chapter would have been useful for readers with less background in this field. This chapter includes periodic tables of all possible formulas for benzenoids, as well as of a selected number of totally fused polycyclic conjugated hydrocarbons and substituted benzenoid hydrocarbons. Methods for enumeration are described in Chapter 3. A summary of the characteristics of the UV, PE, IR, MS, and NMR spectra of polycyclic conjugated hydrocarbons and of aza and related analogues of benzenoid hydrocarbons is given in Chapter 4. This chapter also demonstrates a facile computation of molecular orbital eigenvalues and presents a compact compendium of select eigenvalues for more than 1000 different structures.

The compilations of polycyclic conjugated hydrocarbons and of aza and related analogues of benzenoids are given in Chapters 5 and 6, respectively. The tables include chemical and physical properties of the compounds together with the number of isomers. The organization format and tabulation guidelines are well described and illustrated by examples prior to each table. The literature references cited in the compilations are those regarded by the author as being most pertinent, significant, and recent. A comparison is given between the theoretical number of azabenzenoid isomers and the number that has been synthesized or isolated.

The handbook represents a systematic overview of the isomeric and heterocyclic groups of benzenoid hydrocarbons. The organizational approach makes the information on the individual compounds easily available and allows intercomparison of molecules that are topologically and graph theoretically related. The handbook is well written, clear, and with few printing errors. A glossary of terms is given for readers not familiar with the concepts. It is an excellent reference source for chemists working within many different fields, and will also be useful for students performing research in theoretical physical chemistry. Knowledge of physical and theoretical chemistry is necessary in order to fully understand the theories leading up to the formula periodic tables. However, the many examples, illustrations, and explanations throughout the book make these parts more easily available for those with a weaker theoretical background.

K. E. Thrane, *Federation of the Norwegian Metallurgical Industries*

Computer-Aided Molecular Design. Edited by W. Graham Richards. VCH: New York. 1989. 266 pp. (34 color plates).

This book is a multigraph with origins in the European Conferences on Computer-Aided Molecular Design, sponsored by IBC Technical Services (London). The 21 chapters are updated versions of talks originally presented at six of the conferences held in the 1980s.

In Chapter 1, Peter Murray-Rust gives an overview of the Databases of Molecular Structure. However, of the four main databases in this area (Cambridge Structural Database for organic and organometallic compounds, Inorganic Crystal Structure Database, the National Research Council of Canada's Metals Crystallographic Data File, and the Protein Database for macromolecules), only the CSDB and PDB are discussed in detail. Instead, references to other literature are provided for more thorough coverage. Murray-Rust also gives examples of techniques which have been developed to enhance the data in the primary databases: transformation of the data, fragment libraries, derived data, etc.

The remaining chapters in the book cover the techniques upon which computer-aided molecular designs are based: energy calculations from molecular, quantum, and statistical mechanics; newer theoretical methods of distance geometry and artificial intelligence; and the means of displaying the results with molecular graphics. The 34 color plates included at the end of the work admirably illustrate the last technique.

Gary Wiggins, *Indiana University*

Computer Graphics and Chemical Structures. By Stanley V. Kasparek. John Wiley & Sons: New York. 1990. 798 pp.

From the main title of this book, one might expect a comprehensive, comparative treatment of the systems available to manipulate graphic chemical structures on various types of computer systems. Despite the fact that the author refers to it as a treatise, the book deals with the computer products and systems available from just two organizations: Chemical Abstracts Service and Molecular Design Limited. (Chemical structure searching on Questel is not even mentioned in the work.) In fact *Computer Graphics and Chemical Structures* could more accurately be termed a user manual, since Kasparek presents in great detail techniques for using the CAS Registry File and MDL's ChemBase, MACCS-II, REACCS, and ChemTalk products. For the CAS Registry File, the author concentrates on building structures with the menu option, but provides the equivalent keyboard commands with the menu examples.

Two questions kept coming to mind as I examined the book: Are adequate documentation and training manuals available from the two organizations, and for whom is this book intended? The author states in the preface that the intended audience encompasses bench chemists, biologists, pharmacologists, research physicians, information and library specialists, students, computer specialists, and personal computer enthusiasts. (Candlestick makers are not included!)

There is a detailed table of contents (18 pages) and a 16-page index. Prepared as camera-ready text using MDL's ChemText wordprocessor, the final manuscript was printed on an Apple LaserWriter. The book is certainly attractive enough. However, it suffers from the lack of a good editor. See, for example: "folllowing" (page 20), "Endless loops...does not happen." (page 23), "nostereo" (page 29), "arrow-marked bond" (page 199), "specication" (page 112), and the sentence fragment "are selected and placed on nodes (Figure 1/92)." (page 119). I personally found the lack of indentation of paragraphs throughout the book to be somewhat distracting. A final criticism is the use of a single-type font