Chemical Information Science Coverage in *Chemical Abstracts*

GARY WIGGINS

Chemistry Library, Indiana University, Bloomington, Indiana 47405

For many years *Chemical Abstracts* has included in its coverage publications on chemical documentation or chemical information science. Although the bulk of those publications can be found in section 20 of *Chemical Abstracts*, many relevant articles were found scattered among 39 other sections of *CA* in 1984–1985. In addition to the scattering of references in *CA*, the comprehensiveness of *Chemical Abstracts* as a secondary source for chemical information science is called into question. Data are provided on the journals that contributed the most references on chemical information science and on the languages of publication of relevant articles.

Any abstracting or indexing service that publishes its entries in a classified system is faced with a dilemma—where to put the main entry for an item which is clearly interdisciplinary. Articles written on the subject of chemical information science are often interdisciplinary, so Chemical Abstracts Service (CAS) must frequently deal with the question of placement of such articles in their main printed product, *Chemical Abstracts* (*CA*). Since the beginning of 1972, publications on chemical history, education, and documentation (information science) have been abstracted primarily in section 20 of *Chemical Abstracts*.[1] Some cross-references at the end of that section are given for relevant entries placed in other sections of *CA*. This study investigates the extent to which relevant articles dealing with chemical information science have been scattered in other sections of *CA* in recent years. In addition, comments on the comprehensiveness of *CA*'s coverage of chemical information science are presented.

Entries in the 1984 and 1985 issues of the *Current Awareness Profile on Chemical Information* (a computer-based standard interest profile consisting of edited entries from an SDI search using CAS's Individual Search Service) were examined.[2] The profile entries were classified into the following categories: (1) general chemical or scientific information; (2) dictionaries, nomenclature; (3) handbooks; and (4) pattern recognition, QSAR, and artificial intelligence. All of the entries in the first section of each profile issue were examined for the period 1984–1985, covering volumes 100–103 of *Chemical Abstracts*. On the basis of those publications, relevant entries were found in 40 different sections of *CA*.

The selection of *Chemical Abstracts* as the only secondary source for publications on the topic of chemical information science leads to the exclusion of a number of relevant articles, as the following will illustrate. Two journals in which significant articles on the subject of online searching in chemically related databases have been published in recent years are *Online* (*Weston, Conn.*) and *Database*. Appendix I lists several 1983 articles from those journals that, in my opinion, should have been covered by *Chemical Abstracts*. However, as of Sept 20, 1986, none of the articles had been abstracted by CAS.[3] At the very least, users of *CA* should expect coverage of relevant articles from *Database*. Since March 1979, *Database* has included a regular feature called "Chemcorner" written or edited by Dr. Robert E. Buntrock. Prior to 1979, the column appeared in *Online*. Nevertheless, the "CAS 1986 DDS; Directory of Publications", which lists in alphabetical order the nearly 10 000 journals, technical reports, and other serials currently received at Chemical Abstracts Service, does not include these well-known journals.[4,5]

Table I. Sections of *Chemical Abstracts* with Five or More References on Chemical Information Science Topics, 1984–1985 (As Listed in the *Current Awareness Profile on Chemical Information*)

section	section title	no. of ref	%
3	Biochemical Genetics	6	1.0
4	Toxicology	10	1.7
6	General Biochemistry	6	1.0
9	Biochemical Methods	16	2.8
20	History, Education, and Documentation	357	62.2
21	General Organic Chemistry	16	2.8
22	Physical Organic Chemistry	18	3.1
48	Unit Operations and Processes	15	2.6
51	Fossil Fuels, Derivatives, and Related Products	6	1.0
59	Air Pollution and Industrial Hygiene	9	1.6
65	General Physical Chemistry	8	1.4
73	Optical, Electron, and Mass Spectroscopy and Other Related Properties	15	2.6
75	Crystallography and Liquid Crystals	14	2.4
79	Inorganic Analytical Chemistry	17	3.0
80	Organic Analytical Chemistry	7	1.2
	25 other sections (with 1–4 ref each)	54	9.4
	total	574	99.8

Table II. Language of Publication of Chemical Information Science Works Abstracted in *Chemical Abstracts* Volumes 100–103 (1984–1985)

language	no. of ref	%
English	368	64.1
Japanese	72	12.5
Russian	45	7.8
German	34	5.9
Chinese	16	2.8
French	11	1.9
Spanish	10	1.7
eight other languages	18	3.1
total	574	99.8

Two papers by Ayers include data on the coverage of chemical information science by Chemical Abstracts Service.[6] Conclusions about the major journals of chemical information science and the languages of publication of relevant works were reached by examining the entries in only one section of *Chemical Abstracts*. However, Ayers's decision to limit his studies to just the main section in which chemical information science publications are abstracted in *CA* undoubtedly skewed the results, as will be shown below.

In 1971, Ayers reported that 1228 journal articles were published on chemical information science topics in the period

Table III. Source Publications for English-Language Works on Chemical Information Science Abstracted in *Chemical Abstracts* Volumes 100–103 (1984–1985)

abbrevd journal title or type of publ	no. of ref	%
J. Chem. Inf. Comput. Sci.	113	30.7
Comput. Phys. Commun.	13	3.5
Nucleic Acids Res.[7]	8	2.2
TrAC, Trends Anal. Chem. (Pers. Ed.)	5	1.4
J. Chem. Educ.[8]	5	1.4
Chem. Br.	5	1.4
CHEMTECH	5	1.4
79 other journals	106	28.8
conference proceedings articles	71	19.3
books	20	5.4
technical reports	17	4.6
total	368	100.1

1962–1969, an average of 154 per year. Increasing numbers of chemical information science publications were found in Ayers's second study (1981), which was expanded to include conference proceedings. In volumes 70 and 71 (1969) and 90 and 91 (1979) of *Chemical Abstracts*, he noted virtually the same number of works published on the topic each year (221 in 1969; 220 in 1979). However, in the present study, for the 2-year period 1984–1985, an average of only 179 relevant publications were found in section 20. In addition, an average of 109 publications considered relevant to chemical information science by the author was found in 39 sections other than section 20. Whether this represents a decrease in the number of works published on chemical information science topics·or a tendency by CAS to scatter the more recent works throughout sections of *CA* was not determined, since no attempt to verify Ayers's results was made. In all, 40 sections of *CA* including section 20 contributed relevant publications in 1984–1985. While section 20 is obviously the area of greatest concentration of publications on chemical information science, a thorough analysis of the literature on this subject in *Chemical Abstracts* should be more broadly based than on just that one section.

In order to support this contention, a sampling of the 217 references that appeared in 1984–1985 in sections of *CA* other than section 20 plus one such 1986 *CA* reference are included as Appendix II. While some might argue with the relevance of certain of the publications found outside section 20, it cannot be denied that many chemical information scientists would find a substantial number of them to be of interest. Only one of the nine publications in Appendix II (the paper on PHYTO-TOX published in the *Journal of Chemical Information and Computer Sciences*) was cross-referenced from section 20.

On the basis of the 574 references from 1984–1985 that were included in this study, certain statements can be made about the literature of chemical information science as covered by *Chemical Abstracts*: (1) English continues to be the dominant language of publication for chemical information science. (2) *Journal of Chemical Information and Computer Sciences* is the leading journal in which English-language articles relevant to chemical information science are published.

There is a large amount of scattering of chemical information science publications in many journals and other works. This fact alone makes it understandable that Chemical Abstracts Service might overlook some relevant publications. *Chemical Abstracts* is, after all, an abstracting journal that is devoted more to areas central to the discipline of chemistry than to supportive areas like chemical information science. Nevertheless, as the major English-language chemistry abstracting journal, which has elected to include in its coverage chemical information science, one might expect better coverage of relevant chemical information science English-language publications. Until CAS improves its coverage in that area,

chemical information scientists must use other sources to achieve truly comprehensive coverage of the relevant literature. Those who choose to utilize only *CA* should be aware of the scattering of relevant references in many different sections of *CA* and the seemingly sporadic use of cross-references from section 20 to relevant articles in other sections. It is my opinion that Chemical Abstracts Service should reexamine its editorial policy for the area of chemical information science to include more informative (not necessarily research) articles in English. Furthermore, there should be more consistency in the use of cross-references from section 20 to relevant chemical information science articles in other sections of *CA*.

ACKNOWLEDGMENT

APPENDIX I: SELECTED ARTICLES FROM 1983 ISSUES OF *ONLINE* AND *DATABASE* RELEVANT TO CHEMICAL INFORMATION SCIENCE BUT NOT ABSTRACTED IN *CHEMICAL ABSTRACTS* AS OF SEPT 20, 1986

Kaback, Stuart M. "Online Patent Searching: The Realities". *Online (Weston, Conn.)* **1983**, *7*(4), 22–31.

Hartwell, Ieva O. "Using the New Features in the DIALOG Chemical Information System". *Database* **1983**, *6*(1), 11–27.

Buntrock, Robert E. "Database Search Aids Review: Guides to Chemical Searching Online Chemical Abstracts". *Database* **1983**, *6*(1), 65–68.

Dueltgen, Ronald R. "Chemcorner: Chemical Information in Non-Chemical Databases. IV. Forest and Paper Databases". *Database* **1983**, *6*(2), 80–83.

Antony, Arthur "Chemcorner: Registry Numbers on CA SEARCH: Four Vendor Interpretations". *Database* **1983**, *6*(3), 76–83.

APPENDIX II: SAMPLING OF CHEMICAL INFORMATION SCIENCE CITATIONS IN SECTIONS OF *CHEMICAL ABSTRACTS* OTHER THAN SECTION 20, 1984–1986

104(22):190364h CASCT 55

Balasubramanian, V.; Bhattacharyya, S."Iron and Steel Online Search System". *Online Rev.* **1985**, *9*(3), 251–52.

103(20):162859j CASCT 51

Fries, J. R.; Kennedy, F. E. "Bibliographic Databases in Tribology". *J. Tribol.* **1985**, *107*(3), 285–295.

103(4):26390x CASCT 59

Conkling, Thomas W. "On-line Information Systems for the Environmental Sciences". *J. Environ. Sci.* **1985**, *28*(2), 70–73.

102(25):220149v CASCT 21

Rhodes, Peter "Chemical Structures Online". *Chem. Br.* **1985**, *21*(1), 53, 55, 58.

102(7):61538c CASCT 22

Morffew, Andy "Databases in Molecular Graphics". *J. Mol. Graphics* **1984**, *2*(3), 66–69.

101(16):133257q CASCT 48

Selover, T. B., Jr. "The Engineering Societies Library: An Untapped Resource for Chemical Engineers". *AIChE Symp. Ser.* **1984**, *80*(237), 12–15.

101(4):26683d CASCT 54

McGarr, M. "IMMAGE—the Institution of Mining and Metallurgy Information Service for the International Minerals Industry". *Appl. Comput. Math. Miner. Ind., Pap. Int. Symp., 18th* **1984**, 703–712.

100(15):116098a CASCT 04[9]

Royce, Christopher L.; Fletcher, John S.; Risser, Paul G.; McFarlane, James C.; Benenati, Frank E. "PHYTOTOX: A Database Dealing with the Effect of Organic Chemicals on Terrestrial Vascular Plants". *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 7–10.
100(9):62823n CASCT 03

Chen, H. R.; Dayhoff, M. O.; Barker, W. C.; Hunt, L. T.; Yeh, L. S.; George, D. G.; Orcutt, B. C. "Nucleic Acid Sequence Database. V: Completely Sequenced Genomes". *DNA* **1983**, *2*(4), 275–280.

REFERENCES AND NOTES

(1) Prior to 1972, section 1 of *Chemical Abstracts* covered history, education, and documentation.
(2) *Current Awareness Profile on Chemical Information*; ISSN 0276-8712; Wiggins, Gary, Ed.; Chemical Information Center: Bloomington, IN, 1981–1985. The publication was issued quarterly in all years but the first and was based on a CAS Individual Search Service SDI profile covering vol. 94–103 (1981–1985) of *CA*.
(3) Appendix I originally included two 1984 articles from *Online Review* and a 1984 *Online* article. However, these were removed when it was discovered that Chemical Abstracts Service added those references to the *CA* database in the summer of 1986.
(4) "CAS 1986 DDS: Directory of Publications; Serials Currently Received at CAS". Chemical Abstracts Service: Columbus, OH, 1986. It should be noted that *Online* was previously included in *CA*'s coverage. Prior to 1982, there are eight citations to the journal in the STN International CA File. Neither *Database* nor *Online Review* had any entries in the CA File until the summer of 1986 when four references to *Database*, two to *Online*, and six to *Online Review* were added.
(5) It cannot be argued that CAS's editorial policy excludes articles of the type listed in Appendix I. One can, for example, find the following reference in *CA* section 20, vol. 100, issue 9, abstract number 67418u: Revesz, Gabrielle S.; Cassidy, Phoebe A. "Modern Literature Searching; the Case of C₄". *CHEMTECH* **1984**, *14*, 18–25. That article discusses online literature searching, including the steps and logic involved, as well as the search costs and the databases used. These are topics typically found in articles published in *Online*, *Online Review*, or *Database*.
(6) Ayers, Jerry B. "Journal and Conference Proceedings of Chemical History, Education, and Documentation". *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 71–72. Ayers, Jerry B. "The Journals of Chemical History, Education, and Documentation". *J. Chem. Doc.* **1971**, *11*, 12–13.
(7) The fact that *Nucleic Acids Research* appears among the top six journals of chemical information science which publish articles in English points out one of the pitfalls in such ranked lists. All of the references are taken from a special issue (vol. 12, issue 1, parts 1–2) that was devoted to the applications of computers to research on nucleic acids.
(8) Curiously, Chemical Abstracts Service neglects to include some relevant articles from *Journal of Chemical Education* in its chemical information science coverage. For example, none of the following articles had been abstracted by CAS as of Sept 20, 1986: Gorin, George "An Approach to Teaching Chemical Information Retrieval". *J. Chem. Educ.* **1982**, *59*, 991–994. Wiggins, Gary "The Indiana University Chemical Information Center Program of Chemical Literature Instruction". *J. Chem. Educ.* **1982**, *59*, 994–997. Hendrickson, W. A. "Library Searching; An Industrial User's Viewpoint". *J. Chem. Educ.* **1982**, *59*, 997–999. Allen, Ferne C. "Instruction in Chemical Literature; Industrial Librarian Viewpoint". *J. Chem. Educ.* **1982**, *59*, 999–1002.
(9) Only this paper of the nine listed in Appendix II has a cross-reference from section 20.

Computer as a Versatile Research Assistant

V. ZITKO

Fisheries and Environmental Sciences, Biological Station, Department of Fisheries and Oceans, St. Andrews, N. B., Canada E0G 2X0

Simple uses of a computer in routine research tasks are described. They include extraction of information from texts, programming by using a library of routines, combined data and word processing, and reprint indexing. The applications require only moderate experience in programming. Listings of programs and a library of routines, written mostly in a fairly standard BASIC, are included as supplementary material.

INTRODUCTION

Word processing is now an established and valuable tool in research. However, there are many specialized operations the word-processing programs were not designed to perform. Similarly, the commercial statistical and mathematical packages may not meet specific needs, and commercial database software may be too cumbersome for a relatively straightforward task of reprint indexing.

This paper describes the use of a computer in the daily activities of a researcher. It deals with entry and manipulation of text in a computer. The manipulation includes the extraction of specific information such as names and mass spectra of chemicals from a database output and assorted calculations. A library of routines is discussed, and a system for indexing of reprints is described. The applications require some knowledge of programming BASIC and general familiarity with the computer available. An HP 3000 computer (Hewlett-Packard Co., Cupertino, CA) was used, but the approaches and procedures are generally applicable and most of the programs and routines described are implementable after minor modifications on other systems.

The emphasis is on the use of the computer as a flexible tool rather than on sophisticated programming. The objective is to make versatile use of the computer by simple commands and programs and to be able to perform assorted tasks on short notice. It is often better to write a simple program to perform a specific task than to spend a lot of time developing a complex program that can handle all eventualities or working with a poorly documented commercial package.

DATA INPUT

Data discussed here consist of either text or numbers or both. They are stored in the computer in the form of text (ASCII) files (even numbers are stored as "text"). Text files are usually easier to edit and to communicate between computers than files containing numbers in a "numeric" format. Files consist of records.

With interactive applications and screen displays, it is advantageous to use text files with records 72–80 characters long. "Editor" programs, present on most computers, can perform considerable editing of such files. In this way one makes good use of the software existing on the computer.