CA Selects—A Specialized Current Awareness Service†

J. E. BLAKE,* V. J. MATHIAS, and J. PATTON

Chemical Abstracts Service, P.O. Box 3012, Columbus, Ohio 43210

Received January 23, 1978

CAS, in cooperation with UKCIS, introduced *CA Selects* in response to many requests from the user–subscriber community for such a service. The flexibility of the CA production system, the availability of a complete computer-readable version of CA, and the advent of full-page photocomposition in 1975 allowed the production of the selects in a timely and economic manner with a quality equivalent to that of CA itself. While it is not a completely individualized, user-designed SDI service, *CA Selects* is nevertheless aimed at meeting the needs of individuals, small and large companies, and academic, government, and research institutions.

INTRODUCTION

This paper describes the development, content, and production of *CA Selects*, a series of topical current-awareness abstract bulletins published by Chemical Abstracts Service (CAS). The topics represent various interest areas of biochemistry, chemistry, and chemical engineering. The bulletins are published every other week concurrent with the even-numbered issues of *Chemical Abstracts* (CA). The abstracts contained in each *CA Selects* issue are retrieved from the full scope of the 80 topic-oriented sections appearing in two weekly issues of CA by means of a machine search with a profile custom tailored for the topic. Six topics were published when the service began in October 1976 (Table I), 16 more topics were added in July 1977 (Table II), and 14 more topics were introduced in January 1978 (Table III).

Further topics will be added in the future including popular areas of interest and newly emerging areas of importance. It is also anticipated that waning interest and declining subscriptions will result in the discontinuation of some topics.

BACKGROUND

The utility of any information system depends on whether it is possible for its users to locate effectively specific items of information contained within that system. Consequently, CA includes keyword, volume, and collective indexes to aid users to find particular chemical information contained within the abstracts and the original documents. The division of CA into 80 topic-oriented sections also permits the user to restrict a search to specific areas such as Toxicology, Physical Organic Chemistry, or Electrochemistry, etc. CAS publishes broad topical groupings of related sections including keyword indexes. The five Section Groupings, which together include all of the 80 CA Sections, are: Biochemistry, Organic Chemistry, Macromolecular Chemistry, Applied Chemistry and Chemical Engineering, and Physical and Analytical Chemistry.

However, all such printed publications, in spite of the number and range of indexes which are provided, still require extensive manual searching to locate the desired information. As a consequence, machine-searching techniques have evolved to replace manual searches for chemical information.

In the past, CAS has attempted to meet current-awareness needs in several ways. As early as 1963, CAS published bibliography bulletins on a one-time experimental basis on "Catalysis", "Platinum Metals", and "Chromatography".[1] These bulletins were compiled by searching document titles in *Chemical Titles* (CT). Another example is *Selenium Tellurium Abstracts*, which CAS has been producing since 1967 for the Selenium Tellurium Development Association.

† Presented in part to the Division of Chemical Information, 173rd National Meeting of the American Chemical Society, New Orleans, La., March 22, 1977.

Table I. Initial Six *CA Selects* Topics

	av abstracts/ issue	av pages/ issue
1. Forensic Chemistry	66	6
2. High Speed Liquid Chromatography	37	4
3. Mass Spectrometry	93	8
4. Organosilicon Chemistry	207	18
5. Photochemistry	226	19
6. Psychobiochemistry	260	26

Table II. Sixteen *CA Selects* Topics Introduced in July 1977

	av abstracts/ issue	av pages/ issue
1. Catalysis (Applied & Physical Aspects)	394	33
2. Catalysis (Organic Reactions)	206	16
3. Chemical Hazards	174	14
4. Coal Science & Process Chemistry	180	14
5. Corrosion	231	19
6. Electrochemical Reactions	119	12
7. Electron & Auger Spectroscopy	54	5
8. Electron Spin Resonance (Chemical Aspects)	82	8
9. Gas Chromatography	107	9
10. Gel Permeation Chromatography	25	3
11. Ion Exchange	88	7
12. Nuclear Magnetic Resonance (Chemical Aspects)	288	26
13. Paper & Thin-Layer Chromatography	40	4
14. Radiation Chemistry	112	12
15. Solvent Extraction	80	6
16. Surface Chemistry (Physicochemical Aspects)	161	13

More recently, topical bibliographic bulletins based on data extracted from the *CA Condensates* database have been developed by several information centers. The broadest range of these bulletins are those produced by the United Kingdom Chemical Information Service (UKCIS) which eventually published 45 different topical bulletins.[2]

At the same time that UKCIS was developing these bulletins, CAS itself started to consider the possibility of a series of printed alerting bulletins based on specialty topics. In early 1974, the first in a series of projects was begun to evaluate the feasibility of such bulletins. The utility for both on-line and batch search systems, techniques for selecting abstracts for bulletin topics, and the economics of producing such bulletins were all examined. The various topics investigated required substance-oriented as well as concept-oriented

0095-2338/78/1618-0187$01.00/0

Table III. Fourteen *CA Selects* Topics Introduced in January 1978

	av abstracts/ issue	av pages/ issue
1. Anti-Inflammatory Agents & Arthritis	87	9
2. Antitumor Agents	118	12
3. Atherosclerosis & Heart Disease	187	19
4. Carcinogens, Mutagens, & Teratogens	105	10
5. Flavors & Fragrances	73	6
6. Infrared Spectroscopy	267	23
7. β-Lactam Antibiotics	93	10
8. Liquid Crystals	63	6
9. New Books in Chemistry	311	8
10. Photobiochemistry	170	17
11. Prostaglandins	103	10
12. Raman Spectroscopy	92	8
13. Silver Chemistry	236	20
14. Solar Energy	18	2

searches. Profiles and sample bulletins with various formats were developed for energy, nucleic acids, textiles, solid wastes, safety, forensic chemistry, and organosilicon chemistry. Some of these topics were chosen as a result of discussions with users who were seeking ways of eliminating manual searches of CA. Three topics—Safety, Solid Wastes, and Forensic Chemistry—were published on an experimental basis as special bibliographies in *Chemical Titles* from May 7, 1975 until October 18, 1976, when *CA Selects* began routine publication.

Another CAS project investigated the feasibility of more than 100 additional topics not then being published as bibliographic bulletins by CAS, UKCIS, or other information centers. Complete profiles were developed for 15 of the more promising of these topics; however, no effort was made toward actual publication of the 15 at that time. A portion of these topics have since been incorporated into the *CA Selects* series.

In 1975 when the full abstract text from all 80 CA sections became available in computer-readable form and machine composition of structure diagrams became a reality, full-page photocomposition of topical bulletins which would include full CA abstract text then became feasible. With all these tools available, CAS began planning for the publication of a series of topical current-awareness bulletins for the fall of 1976.

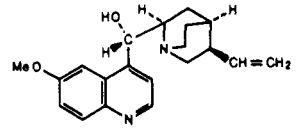
UKCIS, whose cooperation helped bring about *CA Selects*, offered the search profiles for their two most popular Macroprofiles topics: Photochemistry and High Speed Liquid Chromatography. In addition, Forensic Chemistry, the most popular bibliography being published in CT, and Organosilicon Chemistry, the profile for which was developed as a result of discussions with Union Carbide Corporation and Dow Corning Corporation, were chosen as two of the initial *CA Selects* topics. Two further topics, Psychobiochemistry, one of the 15 topics developed earlier at CAS, and Mass Spectrometry, were also selected for inclusion in the set of six topics which began publication as *CA Selects*.

Continued cooperation with UKCIS, further CAS development, and favorable user response to the initial topics resulted in the addition of 16 new topics in July 1977 and 14 further topics in January 1978. More topics will be added in the future.

FEATURES OF *CA SELECTS*

The most important feature of *CA Selects* is that the full abstract text appears with each citation, which eliminates the need to consult the regular CA issues as a second step in determining the relevancy of retrieved citations. In addition, the bulletin is timely, appearing every two weeks on a parallel schedule with the full CA. Its reasonable cost makes possible its acquisition by individuals and small companies and institutions who may have more limited information budgets.

87: 128219s **A novel quinidine metabolism in a suicide case with quinidine sulfate detected by gas chromatography-mass spectrometry.** Leferink, J. G.; Maes, R. A. A.; Sunshine, I.; Forney, R. B., Jr. (Cent. Hum. Toxicol., State Univ. Utrecht, Utrecht, Neth.). *J. Anal. Toxicol.* **1977**, 1(2), 62-5 (Eng). Lactate conjugates of *quinidine* (I) **[56-54-2]** and



I

its 3-hydroxy metabolite **[53467-23-5]** were detected by gas chromatog. and mass spectrometry in forensic tissue and fluid samples from a I overdose suicide patient. Mass spectral fragmentation of lactate conjugates is discussed.

Figure 1.

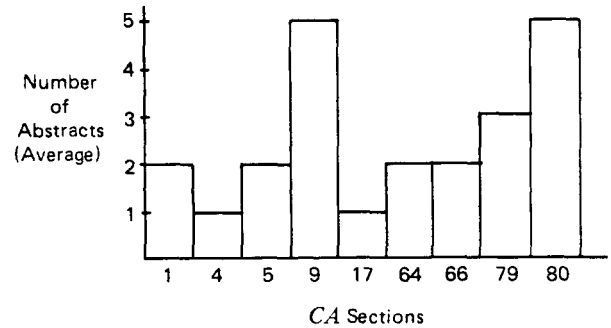


Figure 2. CA section distribution of HSLC over 8 issues.

The average size of an issue of *CA Selects* is 136 abstracts on 12 pages, with individual topics ranging from 25 to 394 abstracts and 3 to 33 pages per issue. Three of the topics—Electrochemical Reactions, Radiation Chemistry, and New Books in Chemistry—are subdivided to provide separation of different subject areas within the topic.

Although an abstract may appear in only one of the 80 topical sections into which CA is divided, it may be included in several of the 36 Selects topics. It is quite often the case that a document is relevant to more than one *CA Selects* topic. Figure 1 illustrates an abstract retrieved for three *CA Selects* topics.

This abstract appeared in its CA weekly issue in Section 4: Toxicology. However, it is pertinent to, and was selected by, the search profiles for three *CA Selects* topics: Forensic Chemistry, Gas Chromatography, and Mass Spectrometry. The profile terms for retrieval of each topic are respectively "suicide", "gas chromatography", and "mass spectrometry", all found in the title. Thus, the individual topic profiles permit the retrieval of pertinent abstracts, even though the main interest of the paper is in another subject area.

For most of the topics, although the majority of the abstracts do come from two or three CA sections, several are often retrieved from many other sections as well. Figure 2 shows the average number of sections from which abstracts were retrieved for eight issues of *CA Selects*: High Speed Liquid Chromatography (HSLC). For the eight issues, there was an average of one or more abstracts per issue retrieved from each of the nine sections shown. There were also abstracts from other sections, but in those cases, the average was less than one abstract per issue.

In addition to the analytical and biochemical methods sections (9, 64, 79, 80) where retrieval on HSLC would be expected, articles of interest were found in Section 1 (Pharmacodynamics), Section 4 (Toxicology), Section 5 (Agrochemicals), Section 17 (Foods), and Section 66 (Surface Chemistry and Colloids). These nine sections are found within

three of the five CA Section Groupings; 48% of the abstracts are from the Biochemistry sections, 9% from the Applied Chemistry and Chemical Engineering sections, and 43% from the Physical and Analytical Chemistry sections.

PROFILE DEVELOPMENT AND EVALUATION

The types of searchable information in the CA database include:

Bibliographic information
Keyword phrases
Abstract text
Volume and Collective General Subject Index entries
Volume and Collective Chemical Substance Index entries
CAS Registry Numbers
Molecular formulas

Comparisons of searching bibliographic information and keywords with searching volume index entries did confirm the conclusions of Dayton et al.[3] that, for general subject type profiles, the recall from one is approximately equivalent to the other, while substance-oriented topics usually yield better results from the volume index entries including CAS Registry Numbers. The CAS information system, however, is sufficiently flexible to allow profiles to be specifically written to search that particular subset of the whole information base which will give the best results for a given topic. Often, the profile will search a combination of two or more of the information types listed above. For example, the profiles for 19 of the current topics search bibliographic information and keywords only. In addition to the bibliographic information and keywords, the abstract text is searched for 11 topics and the volume index entries for six topics.

Potential *CA Selects* topics have been suggested by users, information center staff, UKCIS staff, and CAS advisory and editorial staff. Candidate topics are reviewed and experimental search profiles developed for those which seem most promising. The same editorial staff who generate the abstracts and index entries develop the profiles. Their expertise and the flexibility of the CAS information system make possible profiles tailor-made for the subset of the information base which will give the best results for a given topic. The test runs of the experimental profiles are evaluated in terms of size, precision, recall, etc., by editorial staff and information specialists. In some cases outside experts, perhaps the user who suggested the topic, are asked to evaluate both the profile and the set of abstracts retrieved. Those topics previously published by UKCIS as "Macroprofiles" are based primarily on the profiles already developed by UKCIS.

Once a topic is determined to be suitable in terms of the number of abstracts retrieved per issue, precision, and recall, then the potential markets for the topic must be evaluated. The final decision on any topic must, of course, be an economic one—the potential market must be deemed large enough to support the publication of the topic.

When the topic is approved for publication, the profile is further refined to yield about 85–95% precision, and in some cases precision is close to 100%. The profile's recall is evaluated by the editorial staff, but exact recall studies are not routinely performed. The only long-range, in-depth recall study carried out was for Organosilicon Chemistry where a user had been doing routine manual searches. The comparison showed well over 90% recall for that profile and turned up several relevant abstracts which the manual search had missed. In general, the recall level is believed to be well in excess of 90%.

Subscriber feedback studies were also carried out for the first six topics. The overall subscriber reaction was extremely favorable and several subscribers contributed valid suggestions for improving the profiles.

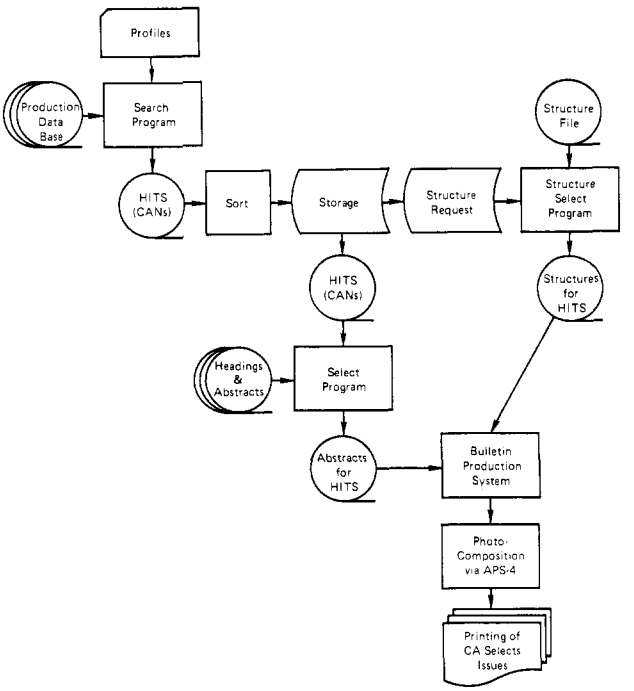


Figure 3. *CA Selects* production system.

1916j Plastics Production US
At the three-quarter mark, production and sales of plastics are up 9% and 11%, respectively, compared with same-period 1976. Latest data from the Society of the Plastics Industry, however, show some differences among plastic classes. Sales of low-density polyethylene totaled 4.9 billion lbs., up 12.7%, and output 4.8 billion, up 10.2%, in the first nine months, compared with comparable 1976 levels. High-density PE production, 2.8 billion lbs., was up 19.2% from 1976, and it outpaced consumption, which is 2.7 billion lbs., up 13.9% from 1976. Polypropylene and styrenic plastics are experiencing slower growth. PP output, up 1.9%, to 2.1 billion lbs. in January–September, is trailing sales growth, which advanced 5.8% over 1976, to 2.1 billion lbs., suggesting some inventory liquidation.
Chem. Week Nov. 23, 1977 p22

Figure 4. CIN extract.

PRODUCTION SYSTEM (FIGURE 3)

One of the factors contributing to the success of *CA Selects* is the ease with which the needed information can be searched and retrieved from the CA publication system on an on-going basis without interfering with the publication of CA itself.

The search profiles are run against the production database each week, and the CA abstract numbers (CANs) of the hits are collected on tape. The CANs retrieved from the odd-numbered issue are sorted and stored for a week until the even-numbered issue is searched. Every two weeks, these stored CANs are used to retrieve the corresponding complete abstracts and abstract headings from the appropriate file. At the same time, those abstracts containing structure diagrams are identified and the structural information is retrieved from the structure files.[4]

The complete set of information for each abstract, i.e., abstract heading, abstract text, and structures, is formatted by the *CA Selects* production system in ascending abstract number order. The output is then photocomposed and the output pages are printed.

FUTURE DEVELOPMENTS

CA Selects is presently restricted to information identified by searching the CA database. Future topics will also be derived from *Chemical Industry Notes* (CIN), either wholly or in combination with CA. CIN contains information extracted from chemical industry and chemical trade periodicals. These periodicals deal primarily with production, pricing, sales,

plant facilities, products, processes, corporate and government activities, and people within the chemical industry. A sample extract is shown in Figure 4.

Of course, while new topics will continue to be added to the *CA Selects* series, some topics will become of less and less interest to most chemists and subscriptions will decline as a result. Such topics will be deleted from the series when they are no longer able to maintain sufficient subscribers to be economically viable. All new topics will be published for a minimum of one year; thereafter, each one will stand on its own merits.

CAS actively solicits reader suggestions for improving present topics, and especially welcomes ideas concerning new topics for addition to the *CA Selects* series.

REFERENCES AND NOTES

(1) R. R. Freeman, J. T. Godfrey, R. E. Maizell, C. N. Rice, and W. H. Shepard, *J. Chem. Doc.*, **4**, 107 (1964).
(2) F. H. Barker, "A Study of Computer-Based Current Awareness Bulletins—UKCIS Macroprofiles", *J. Chem. Inf. Comput. Sci.*, **18**, 91 (1978).
(3) D. L. Dayton, M. J. Fletcher, C. W. Moulton, J. J. Pollock, and A. Zamora, "Comparison of the Retrieval Effectiveness of *CA Condensates* (CACon) and *CA Subject Index Alert* (CASIA)", *J. Chem. Inf. Comput. Sci.*, **17**, 20 (1977).
(4) J. E. Blake, N. A. Farmer, and R. C. Haines, "An Interactive Computer Graphics System for Processing Chemical Structure Diagrams", *J. Chem. Inf. Comput. Sci.*, **17**, 223 (1977).

The Literature of Noble Gas Compounds

DONALD T. HAWKINS

Bell Laboratories, Murray Hill, New Jersey 07974

Received March 14, 1978

Some bibliometric characteristics of the literature on noble gas compounds have been studied. The period considered is from the initial discovery of $XePtF_6$ in 1962, through early 1977. A total of 1192 papers were published on noble gas compounds, 1127 of them since 1962. There was a spurt of interest following the initial discovery, resulting in a large number of publications. In 1963, 139 publications (12% of the total) appeared. The annual number of papers has been substantially lower ever since. The growth of papers on Xe compounds follows this general pattern, but the growth pattern for Kr compounds shows a sharp increase in 1975 and 1976. The literature on noble gas compounds follows Bradford's law, with the "core" comprising ten journals containing 40% of the literature; 33 journals contributed more than five papers each. A total of 1123 authors published in this field; over 60% of them contributed only one paper. The 22 most prolific authors (2% of the total) contributed a total of 497 papers (42% of the total). Most authors remained active in this field for only a short time, with only 5% active for 10 years or more. These findings are discussed in terms of an epidemic model.

INTRODUCTION

Noble gas chemistry burst upon the chemical world in 1962, when Bartlett reported the synthesis of $XePtF_6$.[1] A surge of interest rapidly followed, with the preparation of many compounds of the noble gases, particularly Xe. A long-established chemical principle, the inertness of the noble gases, was overturned, so that noble gas chemistry is a well-recognized subfield today, and interest in these compounds remains high. We have already observed its influence on other areas of chemistry.[2]

The preparation of a comprehensive bibliography[3] on noble gas compounds provides a convenient opportunity to study some of the bibliometric characteristics of the literature on noble gas chemistry. In contrast to many areas, the field of noble gas chemistry has a clearly defined beginning with Bartlett's discovery in 1962. Although there were some unsuccessful attempts to prepare noble gas compounds before 1962, interest in these substances was not high before then, and few papers appeared. In this paper, therefore, we will usually ignore the literature prior to Bartlett's paper.

DATA

The data for this paper were obtained from the recently compiled comprehensive bibliography on noble gas compounds.[3] The 1192 references in the bibliography were keyboarded into machine-readable form. The Bell Laboratories BELDEX indexing system was used to manipulate them and to prepare indexes.[4]

The bibliography is organized into chapters on a substance basis. References pertinent to more than one chapter are listed in all relevant chapters. In this paper, these additional listings are referred to as "cross references". Although there are 1192 unique references in the bibliography, the cross references accounted for an additional 441 entries. Of the 1192 references listed in the bibliography, 65 of them were published before 1962. Of the 441 cross references, 27 were published before 1962. Throughout the remainder of this paper, "references" will refer to the 1127 unique references published after 1962, and "entries" will refer to the total of 1541 references plus cross references published after 1962. Table I lists the chapter titles and the numbers of references and cross references in each.

LITERATURE GROWTH

Figure 1 shows the number of references published in each year from 1962 to 1977. (Data for 1977, and possibly 1976, are incomplete because of the time lag in indexing and abstracting services.) After Bartlett's paper, there was an almost immediate surge of interest in noble gas chemistry. In 1963, 139 papers (12% of the total) were published—more than in any year since. The number of papers decreased to 97 in 1964 and then remained nearly constant at 65 (±6) papers per year from 1965 to 1973. There was an increase in 1974 to 86

0095-2338/78/1618-0190$01.00/0