

## ISI's Experiences with ASCA—A Selective Dissemination System \*

EUGENE GARFIELD and IRVING H. SHER  
Institute for Scientific Information, 325 Chestnut  
Street, Philadelphia, Pennsylvania 19106

Received June 17, 1967

**ASCA (Automatic Subject Citation Alert) is a commercially available SDI system covering the journal literature. The repertoire of questions which ASCA can utilize includes cited references, words from titles, authors, organizations, etc., and allows for logical combinations of these questions. This paper discusses differences and similarities between "citations" and "words" in retrieving and disseminating information. The problem of user-system interaction is explored, and some techniques for developing effective interest profiles are described. Although ASCA is a multi-disciplinary system, examples from fields like synthetic chemistry and biochemistry are provided.**

During the past three years, the Institute for Scientific Information has been testing and operating the first large-scale selective dissemination system commercially available to individual scientists. During this time, researchers in

almost every discipline have been utilizing the ASCA (Automatic Subject Citation Alert) system. More than 500 scientists have been involved in the tests of ASCA files that cover approximately 300,000 current articles each year, requiring about 10 million indexing terms.

Many systems for the selective dissemination of information (SDI) have been reported. Most of these systems are designed to provide information to individuals in

\* Presented before the Division of Chemical Literature Symposium on Selective Dissemination of Information, 153rd National Meeting of the American Chemical Society, Miami, Fla., April 10, 1967.

specific large organizations like IBM, NASA, etc., which ordinarily already possess large-scale computers. For this reason, until now most academic, industrial, and governmental scientists and administrators who could benefit from SDI systems have not. ASCA has changed that situation and puts SDI at the disposal of anyone in large or small organizations.

ASCA has substantial coverage of chemistry and allied fields and has much to offer chemists who want to "keep up" with the literature. Of the 1600 journals covered by ASCA, perhaps the largest segment, beside physics and biomedicine, deals predominantly with chemistry.

Each journal covered is completely indexed as it is ISI's policy to cover journals comprehensively. Selective segmentation of journals or parts of journals into *a priori* categories, as is done in most conventional discipline-oriented systems, usually results in loss of information for the user. One of the significant advantages of the ASCA multidisciplinary approach is that it integrates the literature of seemingly diverse disciplines. Rapid technological advances in many cases involve mixtures of disciplines. ASCA enables the scientist to be aware of and take advantage of this cross fertilization.

In 1965 ISI launched its ASCA system (ASCA I). For the past two years we have had operating experience with ASCA I, ASCA II, and now ASCA III. The main differences between each succeeding generation of systems have been (1) new searching capabilities, (2) lower costs, and (3) added conveniences for the user. Simultaneously, journal coverage has been rapidly expanded.

ISI publishes the *Science Citation Index* (1), which covers about 1600 journals, including most of the significant chemical journals of the world. The input to the ASCA system is exactly the same as the input to the *Science Citation Index* (SCI). However, there are significant differences in the manipulations and outputs. For ASCA, output is a series of individual weekly reports of data selected on any of a vast repertoire of indexing criteria. For SCI, the output is the camera-ready copy of the large alphabetized Citation, Source, Corporate, and Permuterm Indexes.

ASCA was developed after 10 years of research on a retrieval system called "citation indexing" (2). The citation indexing method of retrieval of subject matter is based on the fundamental idea that when a scientist cites an earlier work, he specifies a conceptual—that is, subject—relationship between the two papers. In brief, he indexes his paper by use of citation terms. Later, when the user of a citation index or ASCA expresses his subject interest in citations (cited references) rather than words, the semantic problems frequently associated with changing nomenclatures or complex descriptors are avoided completely.

In all SDI systems it is necessary for the user to construct a "profile" of interests (3). He must "ask" the system what he is interested in. The computer, in turn, will "answer" the user by stating which current articles have appeared that should be of interest. The computer does this by comparing the user's profile with the attributes of the individual source documents.

In the ASCA system, the user has several advantages not found in other SDI systems (4, 5). He can construct his profile in many different ways and can use, among

others, citation and/or word profiles in building his interest profile.

As examples, questions to the ASCA system can include requests for articles which:

1. contain in their titles any specific words, initial parts of words, or phrases, either alone or in any conceivable combinations.
2. are written by a given author.
3. describe work done at a given organization.
4. are published in a given journal.
5. cite any given paper, book, thesis, patent, etc.
6. cite a given first author.
7. conform with any combinations of the above.

Each week the newly published literature of science and technology is screened by ASCA and matched against each individual profile submitted by each scientist. To get an idea of the magnitude and complexity of the ASCA system, consider that in an average week, the computer examines about 6000 newly published items characterized in detail by some 180,000 indexing terms, including among others 60,000 cited references, 68,000 cited and publishing authors, 37,000 keywords in titles, and countless thousands of word stems, word phrases, and combinations thereof.

An important factor in the success of any SDI system is that it must allow the user to communicate with the system readily. This is true of ASCA. Users are provided periodically with cumulated statistical tabulations of their weekly ASCA reports. The user can add to, delete from, or reshape his profile on a weekly basis. This interaction between man and machine enables ASCA not only to be efficient, but constantly self-correcting.

Another factor which affects the success of an SDI system is the ease of access to the documents that are retrieved—that is, listed in the weekly reports. The ASCA subscriber can receive, together with his weekly ASCA report, tear sheets of articles listed on his report. This is called ASCAmatic service and is purely optional. Alternately, the user may order any article on his ASCA report by using ISI's Original Article Tear Sheet (OATS) service. The specified tear sheets will be mailed within 24 hours of receipt of the "Return" copy of his ASCA report. ISI's unique OATS library consists of multiple copies of journals from which articles are literally torn—hence, the name tear sheets. Tear sheets of all but the longest articles are obtained quickly for \$2.00 per article.

ASCA computer reports are sent by first class or air mail each week. Even if the literature does not provide any "hits" in a particular week, the subscriber still receives a report assuring him that the computer file was searched but that no pertinent items were found that answered any of the questions in his profile.

The frequency of ASCA reports makes it possible for any subscriber to read his report in a matter of minutes, an important factor to the busy scientist. The number of "hits" will vary from week to week and from interest to interest. Each ASCA report is quite individualized. Even two people working on similar projects generally have specific approaches which may differ considerably.

While words are inherently ambiguous in any system, and especially in so-called thesaurus systems, citations provide a means for defining subject interests relatively free of ambiguity (6). An example which illustrates the advantages of citation indexing over word indexing is the







ASCA			
AUTOMATIC SUBJECT CITATION ALERT			
A SERVICE OF THE INSTITUTE FOR SCIENTIFIC INFORMATION			
DR. GORDON ALLEN	3 Account Number		
1837 PARROTT DRIVE	81 3 UNIT 16		
BRIMSDON, HARTFORD 20014	19 2 UNIT 16		
REF AUTHOR GORDON AL	LIFE NATURE ORIGIN O	61	
REF AUTHOR POMPERLEBA C	NATURE LIND	201	331 64
REF AUTHOR POMPERLEBA C	NATURE LIND	199	222 63
CITED BY GORDON AL			
ORIG LIND J	45 66	176	142 71557
ORIG LIFE EVOLV			
REF AUTHOR GORDON AL	ORIGIN LIFE LARTH	50	
CITED BY SMITH AE	ORANGE INC		
	109	72 66	1 68 40118 7121
	CONCENTRATE BEHAVIOR IN AN ALTERNATING ELECTRIC FIELD		
THE ITEM BY MILLER SH	J AMER CHEM SOC	77	235 55
REF AUTHOR GORD J	NATURE	197	362 63
REF AUTHOR GORD J	NATURE	197	362 63
CITED BY GORDON AL			
	PREPARED BY		
	J AM CHEM SOC	82	1019 63
		108	112 13607
	HIGH-VOLTAGE ELECTRICAL DISCHARGE REACTIONS OF FATS AND RELATED COMPOUNDS		
THE BOOK BY CALVIN H	CHEMICAL EVOLUTION	24	61
REF AUTHOR GORD J	CHEMICAL EVOLUTION	137	63
THE ITEM BY LIND CU	NATURE	199	219 63
REF AUTHOR GORD J	NATURE	197	362 63
REF AUTHOR GORD J	NATURE	190	442 61
REF AUTHOR GORD J	ARCH BIOCHEM BIOPHYS	94	215 62
REF AUTHOR GORD J	J AM CHEM SOC	3	8 61
REF AUTHOR GORD J	ARCH BIOCHEM BIOPHYS	86	293 62
THE ITEM BY MILLER SH	J AM CHEM SOC	84	215 62
REF AUTHOR POMPERLEBA C	J NATL ACAD SCI US	49	133 63
CITED BY POMPERLEBA C			
	J AM CHEM SOC	81	4076 63
		128	1421 64622
	5-N-HYDROXYLATION OF 5-METHYL-2-CYANOPYRIDAZOLE IN HYDROGEN CYANIDE POLYMERIZATION AND NUCLEIC SYNTHESIS		
SOURCE ALVIN CALVIN H			
THE ITEM BY BELSET T			
CITED BY BELSET T	MADE P	238	40042
	OCURRENCE OF BIOSGENIC STRAINES AND PRECYCLOGIC REACTIONS IN AN OGDOSSE MEDIUM 112 MILLION YEARS AND IN AN EARLY PRECAMBRIAN SMALL 12.7 BILLION YEARS - 2 PRELIMINARY REPORT		

Figure 6. Some papers retrieved by ASCA concerning chemical evolution (origin of life).

indexing will not be completely adequate for getting at certain other types of information.

Suppose a chemist is trying to synthesize a particular chemical compound. If this compound has never been reported in the literature before, it may not be practicable for the chemist to identify starting references that would be necessarily identified with that compound. If the title of a paper were, for example, "The First Synthesis of Pentafluoropyridine," one might assume that ASCA could only report this paper through a term like pentafluoropyridine. However, the chemist could learn of such a paper through citation indexing, if, in fact, he could anticipate the possible procedures of synthesis that might be employed in order to produce such a compound. Thus, when this particular chemical was indeed first reported, the authors of the paper cited a well-known method for synthesizing fluoropyridines. This earlier paper would have been a logical starting reference in a citation profile. Nevertheless, it is not unreasonable to assume that some other completely unique method for synthesizing the compound might also have been cited, which our chemist might not have anticipated. That papers published in chemical journals cite from 10 to 20 references, and thereby provide a considerable redundancy, does not alter the basic problem that in this type of situation word profiles may be preferred to citation profiles.

Thus, the introduction of word profiles in the ASCA III system, by which it is possible for the chemist to stipulate that he should receive alerts to papers containing in their titles a particular word, word stem, or word phrase, adds considerable flexibility to the ASCA system. By adding this feature to ASCA, however, we must be alert

to the danger that inherent inertia on the part of the scientist to avoid the self-disciplining process in preparing profiles may result in the omission from his profile of the less ambiguous "Cited Reference" questions which could make his over-all profile far more effective.

One reason why a citation profile can be more precise than a word profile is the ability of the "Cited Reference" question to delineate clearly a relatively small part of the total literature. Consider a topic as broad as DNA. There were more than 1000 papers published in 1966 that contained this term. In order to identify a subset of the information on this particular topic, one has a choice of specifying, by means of a citation profile, a group of papers on a particular aspect of DNA research. Alternatively, one can stipulate that "DNA" must appear in the title in combination with some other term or terms. Thus, if one is doing work on the DNA content of the brain, the ASCA search could retrieve only papers containing both the terms "DNA" and "brain."

This leads to a discussion of the capabilities of the ASCA III system by which it is possible to phrase ques-

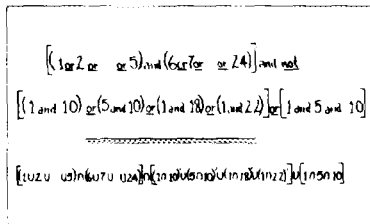


Figure 7. A Boolean statement illustrating one of the ways that logical relationships may be expressed in ASCA.

ASCA WORD PROFILE	
GENERAL TOPIC	CHEMICAL ASTRONOMY
TERM	TYPE QUESTION
01. DUSMRY	1
02. QUASTELLY	
03. QUASTELLY	
04. SUPMRY	
05. HADLUPPEN	
06. HADLUPPEN	
07. HADLUPPEN	
TYPE 1 - (1 OR 2 OR 3 OR ... OR 7)	
08. HADLUPPEN	2
09. SOURCE	
10. EMISSION	
11. EMATIVISTIC	NEGATIVE
TYPE 2 - (18 AND 9) OR (18 AND 10) OR (18 AND 10) AND (NOT 11)	

Figure 8. Use of simple type indicator instead of Boolean statements.

tions involving any complexity of Boolean expressions. Thus, Figure 7 shows how a British user expressed his ASCA question. The question is also expressed in the more familiar notation of the logicians. However, most ASCA word questions are quite simple.

The best way to demonstrate how ASCA employs word profiles is to show a few examples (Figure 8). Type 1 terms are those which cause an alert when at least one of such terms occurs in the title. The Boolean expression is, therefore, 1 or 2 or 3 or 4 or...

Type 2 terms are those where two terms involved must co-occur in a title but not necessarily consecutively. The Boolean expression is (1 and 2) or (1 and 3) or (2 and 3) or... This, of course, can be extended to three or more terms. A term may be either a specific word, a word phrase, or an initial stem of a word or word phrase.

Negation is also possible. A reader, for example, may read a particular journal regularly and will specify that alerts shall be suppressed for items published in that journal. More frequently negation will be applied to a word to help reduce the amount of fallout and/or "noise" which may result from employing other ambiguous or more gen-

eric terms. For example, an information scientist might use the term "information" but specify that papers on "genetic information" shall not be retrieved.

A convenient device is the "Chinese menu" type of question (Figure 9). The Boolean expression for such a question is (1 or 2 or 3) and (4 or 5 or 6 or... or 16).

The ASCA III system is able to handle numerous other types of questions and combinations of questions. Our problem is not, "How can ASCA III meet an unusual requirement?" but more often rather, "How many of the almost endless variety of questions available in the ASCA III repertoire dare we try to describe without hopelessly confusing the user?" Only a small number of users are vitally interested in Boolean algebra. All they want is the pertinent information with the least effort.

LITERATURE CITED

- (1) Garfield, E., "Science Citation Index—A New Dimension in Indexing," *Science* 144, 649-54 (1964).
- (2) Garfield, E., "Citation Indexes for Science," *ibid.* 122, 108-11 (1955).
- (3) Luhn, H. P., "Selective Dissemination of New Scientific Information with the Aid of Electronic Processing Equipment," *Am Document* 12 (2), 131-38 (1961).
- (4) Sage, C. R., Anderson, R. R., and Fitzwater, D. R., "Adaptive Information Dissemination," *ibid.* 16 (3), 185-200 (1965).
- (5) Rice, C. N., "A Computer-Based Alerting System for Chemical Titles," *J. CHEM. DOC.* 5, 163-65 (1965).
- (6) Gorn, S., "The Treatment of Ambiguity and Paradox in Mechanical Languages," *Proc. Symp on Pure Mathematics* 5, Recursive Function Theory, p. 201-18 (1962).
- (7) Rogers, F. B., "MEDLARS Operating Experience at the University of Colorado," *Bull. Med. Library Assn.* 54 (1), 1-10 (1966).
- (8) Garfield, E., "World Brain or Memex? Mechanical and Intellectual Requirements for Universal Bibliographic Control," Paper presented at the Eighth Annual Summer Symposium on "The Foundations of Access to Knowledge," Syracuse University School of Library Sciences, Syracuse, N. Y., July 30, 1965; in press.
- (9) Keen, E. M., "Citation Indexes," *Aslib Proc.* 16 (8), 246-51 (1964).
- (10) Simkins, M. A., "Information Retrieval for Chemists," Lecture delivered at meeting of Fine Chemicals Group, Society of Chemical Industry, Belgrave Square, London, England, December 2, 1966. See *Chem Ind (London)* 3, 122 (1967).

ASCA WORD PROFILE					
"CHINESE MENU" COMBINATION TYPE QUESTION					
GENERAL TOPIC: ROLE OF ENZYMES AND DNA-RNA SYNTHESIS IN RAT LIVER					
CATEGORY A <sub>1</sub>			CATEGORY A <sub>2</sub>		
TERM NO.	TERM	TYPE QUESTION	TERM NO.	TERM	TYPE QUESTION
-01	REGEN RAT/	1A	-04	ADAPT/	1B2
-02	HEPAT/	"	-05	INDUC/	"
-05	LIVER	"	-06	REPRESS/	"
			-07	HISTONE	"
			-08	ACIDIC PROTEIN	"
			-09	RIBUL/	"
			-10	RIBOSOM/	"
			-11	PHOSPH/	"
			-12	CYTOL/	"
			-13	URIDIN/	"
			-14	GENE/	"
			-15	DELETION/	"
			-16	DELETION/	"

TYPE 1A, AND TYPE 1A<sub>2</sub> + (1) OR 2 OR 3) AND 14 OR 5 OR... OR 16)

Figure 9. Example of a "Chinese Menu" combination type question.