22
7-8-76
25 cop to NTIS.

A Short Guide to SDI Profiling at ORNL

Herbert S. Pomerance

OAK RIDGE NATIONAL LABORATORY

OPERATED BY UNION CARBIDE CORPORATION FOR THE ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION

DISCLAIMER

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.
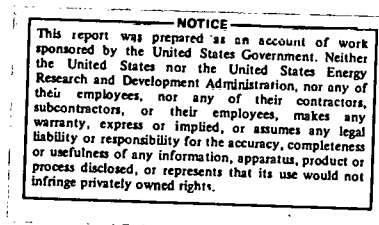
ORNL/TM-5484

Contract No. W-7405-eng-26

Information Division

A SHORT GUIDE TO SDI PROFILING AT ORNL

Herbert S. Pomerance

JUNE 1976

OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37830
operated by
UNION CARBIDE CORPORATION
for the
ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

CONTENTS

ABSTRACT

ORNL has machine-searchable data bases that correspond to printed indexes and abstracts. This guide describes the peculiarities of those several data bases and the conventions of the ORNL search system so that users can write their own queries or search profiles and can interpret the part of the output that is encoded.

INTRODUCTION

The Information Division of ORNL supports machine searching of several indexing and abstract bases that are published on magnetic tape as well as in hard copy. The tapes are leased either by ORNL or by the Technical Information Organization of ERDA, and are processed by the ORSDI group in the Computer Sciences Division. There is no direct charge for ordinary use of the search system; the cost is borne by the Information Division.

Other machine search is also supported here. Some SDI (Selective Dissemination of Information) work on tapes not held here is bought outside ORNL. Retrospective searches can be performed on our holdings, and also on several other systems: The Recon system of ERDA; the Dialog system of Lockheed Retrieval Service; the Orbit III system of System Development Corp.; the New York Times Information Bank.

The user may use this guide to write his own query, called a search profile, or join with some one from the library reference section. In either case the profile must be submitted on standard forms in a standard way so that the operator can enter the material into the computer without further editing. This guide suggests sources for identifying index terms or subject categories, it defines rules to be followed in entering terms and combining them, it describes briefly the several data bases and the fields in them that are generally used for search.

The materials here are not all original. Guides for use at other stations were written by Harry Hamilton and Hilary Burton. The local material comes from Carla Giles , who directs the ORSDI group, on the computer end of the work, and from Herbert Pomerance, who is in charge of machine-based search for the Information Division, on the information end.

DATA BASES HELD AT ORNL FOR SEARCH IN THE ORSDI SYSTEM (March 1976)

Data Base Symbol	Name	Frequency (per year)	Earliest Issue vol.(iss)	date	Records per issue, 1974
NSA	Nuclear Science Abstracts	24	25(14)	7/71	2750
	Chemical Abstracts				
CAO	odd issues	26	77(25)	12/72	6600
CAE	even issues	26	77(24)	11/72	8200
CBAC	Chemical-Biological Activities	26	13(1)	1/71	1800
BA	Biological Abstracts	24	51(1)	1/71	6000
BRI	BioResearch Index	12	71(1)	1/71	8000
GRA	Government Reports Announcements	24	72(1)	1/72	2500
CAIN*	National Agricultural Library	12		1/70	10000
MET	Metals Abstracts	12	69(1)	1/69	900
EI*	Engineering Index	12	72(1)	1/72	7500

*SDI service only, except by special arrangement.

This list is for batch search in the ORSDI system. The on-line systems (Recon, Dialog, SDC, NY Times) are not shown here nor are the systems separately operated by some of the ORNL information centers.

The large machine-searchable index bases are generally derived from the magnetic tape record of the printing operation of that base; therefore, searching the magnetic tape record is similar to searching the printed record. In changing from manual searching to machine searching, the profiler in framing his query must remember that the data base generally undergoes a conversion in format and style that may not match the format of the printed issue; therefore, for an effective search the profiler should familiarize himself with the format of the tape as it is used as well as with its content. The content of the data base is often augmented in the tape version compared with the printed version with additional index terms or subject classifications.

The query must be phrased in accord with the requirements of the search system. There are several on-line interactive systems for retrospective search in use at ORNL. This guide does not directly deal with them but only with the batch searches that are the principal source of SDI search (that searches only the last tape) and the similar retrospective search that goes through all the tapes from first to last. The on-line systems use ordered indexes of the various fields (inverted files), such as alphabetical lists of authors, of index terms, of journal names or codens, of subject categories, of locations of work. The batch search uses no indexes but only the original tape converted for search purposes. Each item or record is searched in its turn for the presence of the terms in the queries. Only when the last item has been searched can there be a statement of how many hits have been found for each query.

The text search program in computer use at ORNL was first developed with NSF money at the University of Georgia. The program converts all data bases to a common search format and uses a single computer search program for all bases. The common format calls all titles TTL, calls all authors ATI (for authors inverted), calls all journal Codens CDN, and so on. A data base may also have its unique elements. For example, Biological Abstracts has two subject codes it calls Cross Code and Biosystematic Code; Nuclear Science Abstracts uses the INIS Thesaurus terms and the INIS subject classifications. For the common

elements variations are found. Are authors' first names spelled out, are the initials spaced with periods or closed up without periods? Are chemical element symbols used in the titles or are the chemical names spelled out? There are five-letter Codens and six-letter Codens.

Because the titles in the several bases are named TTL, it is possible to write a single search profile or query that fits all the bases and that does nothing more than search the titles for the selected keywords. The same keywords can be sought in the abstract field ABS and in the tree-index-term field FIT in the bases that have such fields. By a convention in our search system, the name TXT includes the three fields TTL, FIT and ABS. Of the bases now at ORNL (1976), CBAC (Chemical-Biological Abstracts) in full and GRA (Government Reports Announcements) in part have abstracts; the others have none. Only NSA (Nuclear Science Abstracts) and its successors in 1976, Atomindex and Energy Information Data Base remain without a field called FIT. Like several other bases, NSA and EIDB use a controlled index vocabulary but for the other bases the controlled field in the ORNL conversion is also designated FIT (which may now mean free or fixed index terms) whereas for NSA the index terms are designated NST.

Although a single search profile may serve for all the bases, it may not be efficient in all of them. Aside from the controlled vocabularies there is subject indexing and each base has its own subject catagories. In the bases with only one subject assignment to an item, a subject may best be used for negation (AND NOT). In the bases with multiple subject assignments, a profile with subjects alone may be more effective than one with index terms or with title keywords.

Each index journal has its own manner of presentation of its indexes. It may use subject heading plus modifier line as in Nuclear Science Abstracts until 1972; URANIUM, analysis for boron. Or it may use coordinate indexing where all terms have equal value, as in NSA after 1972: URANIUM; BORON; QUANTITATIVE CHEMICAL ANALYSIS. Or it may use KWIC indexes (permuted words) of titles, as in Biological Abstracts. Always there is an alphabetical author index. Implicit in the tape record are not only the indexes found in the printed version

but indexes of any other element: the year, the journal, the author's institution, the words of the title. And also implicit in the tape record are indexes by pairs of words and by triples.

Each data element in the text search system is classified as one of two types: left-anchored (LA) or free text (FT). The classification determines the manner and speed in which the data element is searched; therefore, the profiler must be aware of this characteristic for each data element he codes and structure his request accordingly. Left-anchored attributed to a data element implies that its format and content are precisely known. The match always begins with the left-most character. Before further character-by-character comparisons are made, the length of the search term is checked against the length of the data base term. If right-hand truncation is not specified and if the lengths are unequal, further comparisons are skipped and the terms are flagged as unequal. In other words, left-anchored search terms which are not specified to be truncated must match the data base term exactly. Word fragments from the data base cannot be retrieved unless the search term is truncated. Examples of left-anchored data elements are: Author Name (last name first); Coden; and Registry Number. Only suffix (right-hand) truncation or no truncation can be used with LA terms. The truncation asterisk (*) is interpreted to mean accept any following character or blank.

The second type of data element is free text (FT). This type specifies a character-by-character match of all terms preceded by a blank or by a punctuation mark. Truncation on either or both ends is applicable. Examples of free-text data element are: title; abstract; and free index term or keyword.

The following examples illustrate the four possible modes of truncation.

Term	Explanation
CELL	Retrieves the term when it is bounded by blank spaces or by punctuation; e.g., CELL, CELL-FREE, STEM-CELL.
CELL*	Retrieves the term when it is followed by any character including a blank space, and when it is preceded by a blank space or punctuation; e.g., CELL, CELL-FREE, STEM-CELL, CELLOPHANE, CELLS, CELLULAR.

*CELL   Retrieves the term when it is preceded by any character, including a blank space, and when it is followed by a blank space or punctuation; e.g., CELL, CELL-FREE, STEM-FREE, MICROCELL.

CELL Retrieves the term when it is preceded or followed by any character, including blank spaces or punctuation; e.g., CELL, CELL-FREE, ACELLULAR, MICELLES, CELLULOSE (the costliest in use of computer time).

Any one term for search is limited to sixty characters. Any one query (search profile) is limited to 240 terms or less depending on the length of the terms. For the search the computer treats all characters that are not alphabetic or numeric, whether in the query or in the data base, as blanks. For the printout the computer preserves all those characters, the hyphens, commas, brackets, apostrophes, etc.

The computer first seeks out all the terms that are asked for, then it tests the terms against the logical statement (Boolean expression) that was submitted with the query. If you ask for 'metabolism of lithium in man', you first establish three groups: G001 = metabolism; G002 = lithium; G003 = man. The logical statement is G001 & G002 & G003, where & is the symbol for AND. If you desire other alkali metals than lithium, then G002 = (lithium | sodium | potassium), where | is the symbol for OR. Similarly G003 may have to be (man | woman | child | human). The convention in our system is that the OR is understood to tie all the terms within one group and need not be there written out; it must be written between groups, as in G004 & (G005 | G006). If you desire metabolism of lithium in any animal except man, then the logical statement is G001 & G002 &¬ G003, where &¬ is the symbol for AND NOT.

The Venn diagrams show the relations. For two groups, where A is metabolism and B is lithium:
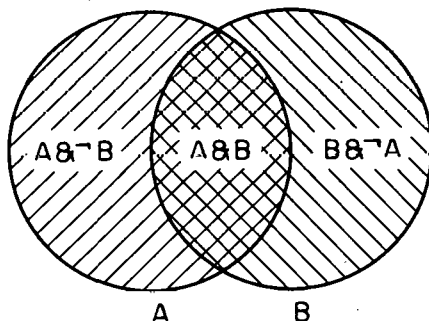
A&¬B is read "A and not B", that is metabolism but not lithium. A&B is read "A and B", that is metabolism and lithium (logically) or metabolism of lithium (in common speech). Three or four circles can be used to illustrate the various combinations of three or four groups. A|B is read "A or B" and is the combined area, everything to do with metabolism or with lithium or with both together.

The computer program performs AND before OR and so use parentheses for punctuation if the OR must come first. There is a difference between (A&B) | C and A&(B | C); when you write A&B | C without parentheses, the search program will assume the former.

The terms in the profile may be words or codes, numerals or letters that belong to the data base you search. The column DEM is for the three-letter designation of the field in which that term will be found. The 'weight' is optional; it is used for sorting or as part of a logical statement. With sorting, the biggest point score is printed at the head of the list (maximum weight is ±32,767). You can use weighting to keep track of parts of the question; for example assign 100 to lithium, 40 to sodium, and 20 to potassium.

The illustrative profile for USER: I. Example is not one you should copy directly. It displays some features of writing a composite query, in this case the toxicology of three elements.

a) The names of the elements are spelled out in these data bases, so symbols are not needed. The fields of search TXT may be title, keywords or abstracts in GRA; title or free index terms (keywords) in BA, BRI, CAE or CAO; title in NSA. The "TXT" in column DEM does not include the fixed index terms of NSA so we must use the DEM which is unique to NSA, "NST" (for nuclear science term.) The NST terms are left-anchored and we wish mercury, for example, to encompass mercury oxide and mercury 198; thus right truncation, represented by the asterisk, is needed. Truncation is not needed for the free-text terms searched under TXT.

b) The weights are used here for convenience to keep the elements separate in the printout. An abstract with all three elements named once each will appear first with weight 1600; a paper with

OAK RIDGE SELECTIVE DISSEMINATION OF INFORMATION
PROFILE SUBMISSION
(PLEASE PRINT OR TYPE)

	DATE
	1 April 1976

INFORMATION SPECIALIST: H. Pomerance	PHONE NO. 3-1393

USER *(IF NOT THE ABOVE):*   I. Example

USER NO.:	PROFILE NO. *(IF NOT NEW):*

PROFILE STATUS:	[x] NEW	☐ REVISION	☐ DELETION	☐ INACTIVE

USER COMMENTS:

THRESHOLD WEIGHT:

BOOLEAN EXPRESSION *(NOT MORE THAN 140 CHARACTERS):*

    G001 & G002

TYPE OF SEARCH (USE FOR NEW AND FOR REVISION):

☐ RETROSPECTIVE ONLY	DATA BASE CODES:

☐ RETROSPECTIVE & CURRENT	DATA BASE CODES:

[x] CURRENT ONLY	DATA BASE CODES: N3A BA BRI GRA CAO CAE

OUTPUT FORM:	[x] INDEX TERMS	☐ ABSTRACT	☐ INDEX TERMS & ABSTRACT
OUTPUT CLASS:	[x] PAPER	☐ CARDS (6" X 4")	☐ TAPE*	☐ PAPER & TAPE*

* BY SPECIAL ARRANGEMENT ONLY

SEND TO: ORSDI
        COMPUTER SCIENCES DIVISION
        BLDG. 4500N, ROOM 218

UCN-11151
(3   5-74)

ORNL COMPUTER-BASED INFORMATION SERVICES

PROFILE SUBMISSION

User _____ I. Example _____   Page _____

Please print or type; indicate blank spaces in terms.
Maximum term length, 60 characters; maximum profile length, 240 terms.

GROUP	WEIGHT	DEM	TERM	
G001	1000	TXT	MERCURY	
	400		CADMIUM	
	200		VANADIUM	
	1000	NST	MERCURY*	
	400		CADMIUM*	
	200		VANADIUM*	
G002		TXT	TOXI*	
			POLLUT*	
			ENVIRONMENT*	
		CXC	225*	(toxicology)
			3701*	(environmental health)
		CSC	06T	(toxicology)
			06F	(environmental biology)
		NCC	48*	(life sciences)
		NST	BIOLOGICAL EFFECTS	
			TOXICITY	
			POLLUTION	
			WATER POLLUTION	
			LAND POLLUTION	
		CAS	005*	(toxicology)

cadmium and vanadium will have weight 600; a paper with mercury alone will have weight 1000.

c) In group G002 there are three word stems and some subject classes. Toxi* includes toxin, toxins, toxic, toxicology. But what if the three word stems do not appear in the title TTL or in the free index terms FIT or in the abstract ABS? For the BioAbstracts BA and BRI we can use the Cross Codes CXC; the topic (subject heading) corresponding to the code has been added in the example. For GRA we can use the Cosati fields and groups CSC. For NSA we can use the subject categories NCC or the index terms in NST. For Chemical Abstracts CAE and CAO, an abstract published in Sect. 5, Toxicology, does not also have the keyword "toxicology" added although that word may appear in the title; in this case the subject field CAS and the free-text word stem "toxi*" must both be used.

Composite profiles as this one are not always efficient. Separate profiles for each data base may be better.

The following lists are a brief statement of the likely fields of search among the large data bases now held at ORNL. The fields that appear in most data bases are given in the first list. The three-letter designation must be used for a search profile submitted to the computer department. Left-anchored terms (LA) can't suffer left-truncation; free-text terms (FT) can. Both can suffer right-truncation. For guides to truncation, use a general dictionary, or the authority list of index terms for the data base, or the printed indexes of that base such as the permuted word index in Biological Abstracts.

Data Elements Appearing in Most Bases	Designation	Type
Title	TTL	FT
Free index terms, keywords, augmentation of title	FIT	FT

Personal author
    may be Jones AB; or Jones A B; or Jones, A. B.;
    or Jones, A.B.; or Jones, Arthur B.
	ATI	LA

Corporate Author	CAN	FT
Location of work	LOW	FT

Coden (may be 5 or 6 letters)	CDN	LA
Year of Publication of the article, usually 4 digits	PDD	FT
Abstract	ABS	FT
Text This is used to search TTL and FIT and ABS without writing all three separately	TXT	FT
Volume Number of the index journal; it is found in different fields in different journals	SJV SEC SJD	LA FT FT

Biosis Previews   BA and BRI

The Biosciences Information Services publish two journals. BA, Biological Abstracts, reports the journal and conference proceeding literatures with abstracts and indexes. BRI, BioResearch Index, has two thirds as many items but without abstract from symposia, review journals and books, trade journals, annals of science academies, institutional reports, and letters to the editor. The two have the same indexing. The titles are augmented with genus and species, with chemical names, with field of work, with method; in the printed issue the augmented title is permuted for a KWIC index. For the variant spellings and for the word breaks (such as di ethyl ether), study the printed indexes or use "A Guide to the Vocabulary of Biological Literature" which gives the frequencies of usage of 8700 words and fragments for the years 1959-73. The "Biosystematic Code," a five-digit code, is used for taxonomic orders and larger divisions; smaller divisions, genus and species, appear in the title augmentation or in the title itself. The nearly one hundred subject classifications of the printed issues of BA have been further subdivided in the "Cross Code". We use six characters, the five numerals of the code followed by a P (primary) for the section in which the printed abstract appeared, by an S (secondary) for subject fields nearly as important, by a T (tertiary) for fields of lesser relevance. Because truncation of a Cross Code also removes the P. S, and T. it may pay to enumerate all the desired Code numbers and use the P and S rather than to truncate

the Code numbers and get the unwanted T titles.

Data Elements

Title	TTL	FT
Augmentation of title	FIT	FT
Author, as Goldberg A R	ATI	LA
Biosystematic code, five digits	BTC	LA
Cross code, five digits plus P, S, or T	CXC	LA
Coden, five digits	CDN	LA
BA volume number, 59 and 60 in 1975	SJV	LA
BRI volume number, 75 in 1975	SJV	LA

SUBJECT CLASSIFICATIONS AND CROSS CODES IN BIOABSTRACTS

The Subject Classification Guide which appears as the Table of Contents in each issue of Biological Abstracts or BioResearch Index is alphabetical. The hundred subjects are further divided in the Cross Codes and assigned numbers in accord with a logical grouping. The table below in numerical order shows an abbreviated name of the subject for each numerical code. It is intended for reading output, not for writing a profile. The five-digit numbers shown here are each the first of a series but may themselves not be usable numbers.

Subject	Code	Subject	Code
general	00500	biochem-clinical	10010
methods, apparatus	01000	—methods	10050
evolution	01500	—studies	10060
cytology, cytochem	02500	replication,...	10300
genetics	03500	biophysics	10500
mathematical	04500	external effects	10600
anthropology	05000	enzymes	10800
social biology	05500	anatomy, histology	11100
aero; underwater	06000	chordate body, genl	11300
subterranean	06400	physiology, genl	12000
radiation biology	06500	movement	12100
periodic, circadian	07200	pathology	12500
ecology	07500	metabolism	13000

nutrition	13200	in vitro studies	32600
food technology	13500	virology	33500
digestive sys	14000	immunology (chem)	34500
cardiovascular sys	14500	immunology (parasit)	35000
blood, fluids	15000	allergy	35500
urinary sys, secreti	15500	microbiology, medical	36000
respiratory sys	16000	public health	37000
reproductive sys	16500	veterinary science	38000
endocrine sys	17000	chemotherapy	38500
muscle	17500	food and ind. microbio	39000
bone, connectives	18000	disinfection	39500
coelomic, mesenteries	18200	soil microbiology	40000
		paleobotany	50000
integumentary sys	18500	palynology	50100
dental, oral biology	19000	systematic botany	50500
sense organs	20000	morphology of plants	51000
nervous sys	20500	physiol, biochem plants	51500
psychiatry	21000	economic botany	52000
pharmacology	22000	agronomy	52500
immunization	22100	soil science	52800
toxicology	22500	horticulture	53000
temperature	23000	forestry	53500
neoplasms	24000	pharmaceutical botany	54000
gerontology	24500	phytopathology	54500
pediatrics	25000	pesticides, herbicides	54600
embryology	25500	economic entomology	60000
animal production	26500	parasitology	60500
poultry production	27000	sanitary entomology	61000
laboratory animals	28000	chordata, systematic	62500
microoraanisms	29500	animal distribution	62800
bacteriology, general	30000	paleozoology	63000
-morphol, cytol	30500	invertebrata, system.	63500
-physiol, biochem	31000	-compar., morphol.	64000
-genetics (also virus)	31500	-structure	64200
microbiology methods	32000	paleobiology	64500
-ultrastructure	32300	geological periods	64700
tissue culture	32500		

BIOSYSTEMATIC CODES (TAXONOMY) IN BIOABSTRACTS

The major divisions of the Biosystematic Code are summarized here as phyla and classes, and as orders within Mammalia. The full 27-page list shows the taxonomic orders in alphabetical array within each phylum and with their codes in ascending numerical order; within Mammalia the finest division is the family. This list is intended as a table of contents for the complete list and for reading output. In

a search profile you must truncate these numbers to search a whole class
(or family).

00500	Organisms, gen'l	69000	Linguatulida
01000	Microorganisms, gen'l	71000	Tardigrada
03000	Viruses	73000	Onychophora
06000	Bacteria	75000	Arthropoda, gen'l
11000	Plantae, gen'l	75100	Crustacea
12000	Cryptogamae	75200	Myriapoda
13000	Algae	75300	Insecta
15000	Fungi	75400	Chelicerata
19000	Lichenes	77000	Chaetognatha
20000	Embryophyta	79000	Hemichordata
21000	Bryophyta	81000	Pogonophora
22000	Tracheophyta	83000	Echinodermata
23000	Pteridophyta	85000	Chordata
25000	Spermatophyta	85100	Protochordata
25100	Gymnospermae	85150	Vertebrata, gen'l
25200	Monocotyledones	85200	Pisces
25500	Dicotyledones	85300	Amphibia
27000	Paleobiology	85400	Reptilia
29000	Paleobotany	85500	Aves
31000	Paleozoology	85700 to 86490	Mammalia
33000	Animalia, gen'l	85705-750	Artiodactyla
4000	Invertebrata, gen'l	85760-795	Carnivora
35000	Protozoa, gen'l	85800-840	Cetacea
36000	Metazoa	85845-930	Chiroptera
37000	Mesozoa	85940-945	Dermoptera
39000	Porifera	85955-970	Edentata
41000	Cnidaria	85980-985	Hyracoidea
43000	Ctenophora	85990-6030	Insectivora
44000	Helminthes	86035-045	Lagomorpha
45000	Platyhelminthes	86055-100	Marsupialia
47000	Rhynchocoela	86105-115	Monotremata
49000	Acanthocephala	86140-155	Perissodactyla
51000	Aschelminthes	86160-165	Pholidota
53000	Entoprocta	86170-185	Pinnipedia
55000	Phoronidae	86190-245	Primates
57000	Ectoptocta	(86215)	(Hominidae)
59000	Brachiopoda	86250-255	Proboscidae
61000	Mollusca	86265-425	Rodentia
63000	Sipunculoida	(86376)	(Muridae, rats and mice)
65000	Annelida	86430-440	Sirenia
67000	Echiuroidea	86465-470	Tubulidentata

This tape corresponds to the weekly issues of Chemical Abstracts (CAO in odd weeks, CAE in even weeks). Another set of tapes (CASIA and CAISF, not used by us) corresponds to the volume indexes that include the analytical subject index, the formula index, the substance index. In CA Condensates the title, the section assignment and the keywords are all that are available for subject searching. Each item is assigned to one or several of the sections, shown below, and to their subdivisions defined in "Subject Coverage and Arrangement of Abstracts by Sections." Word frequencies, phrases from the titles, and key-letters-in-context (KLIC) for the two years 1973-74 are found in the microfiche publication "CA Condensate Search Aids."

For the keywords upon which most SDI searches depend, CA makes the following statement.

> The purpose of this index is to provide quick entry into the subject content of the abstracts, regardless of the particular CA section in which the abstracts occur. One or more keyword entries are derived from the title, text, or context of the abstract. There is no specific relationship between the keyword phrases in the Keyword Index and the much more detailed index entries which appear in the subsequent CA Volume Indexes for the same abstract. No major effort has been directed toward standardization of terms in the Keyword Index. Synonyms, however, are not usually included as additional keywords for the same abstract. For example, if "ethylene oxide" is the keyword phrase chosen for an abstract concerning ethylene oxide, "oxirane" is not also used. Abbreviations and acronyms employed in keyword phrases are the same as those listed in the introduction to CA, issue No. 1 of the current volume, or in the latest Index Guide. Single-letter abbreviations, those that contain internal punctuation (for example, a.c., m.p.), and abbreviations that spell a word are not used. For complex organic compounds the basic structural unit is used as a keyword entry. The name of the complete compound may also be added, e.g., "phenanthrene" and "isopropylphenanthrene." All concepts and substance names, whether of specific substances or classes of substances, are given in the singular form. Thus, the word "phenol" may refer to the specific substance phenol or to a class of substances of which phenol is the simplest example. A chemical substance name or class name standing alone as a keyword implies preparation of that particular chemical species. The standard positional and isomeric locants are rarely used in keywords, cis-cinnamic acid appears in the Keyword Index at "cinnamic acid" and 2-butanol at "butanol."

The keywords are often abbreviated. In each volume of CA is to be found a list of 350 abbreviations and symbols that are used in the abstracts; e.g., concn for concentration. Not all of them are abbreviated when used as keywords and only inspection of the keywords in a printed issue can be depended on.

Data Elements

Title TTL, corporate author CAN, location of work LOW, year of publication PDD

Author, as Hagstrum HD	ATI	LA
Keywords	FIT	FT
Chemical Abstracts section as 0ab*, section and subsection as 0ab0cd*	CAS	LA
Coden, 5 characters plus a CA check character usually truncate after the fifth	CDN	LA
CA volume number, 082 and 083 in 1975	SJV	LA

The section numbers here are for the printed issues of Chemical Abstracts and for magnetic tape issues of CA Condensates. Many of these sections reappear in magnetic tape subset versions that carry the full abstract and all the deep indexing of the volume indexes; the subsets are indicated for 1975 although only one of them is now in use at ORNL.

CAO (odd issues)

Biochemistry Sections

1. Pharmacodynamics .... C .....
2. Hormone Pharmacology .... C .....
3. Biochemical Interactions .... C .....
4. Toxicology .... C F EE
5. Agrochemicals .... C F ...
6. General Biochemistry ; ....
7. Enzymes ....
8. Radiation Biochemistry ....
9. Biochemical Methods ....
10. Microbial Biochemistry ....
11. Plant Biochemistry ....
12. Nonmammalian Biochemistry ....
13. Mammalian Biochemistry ....
14. Mammalian Pathological Biochemistry ......
15. Immunochemistry ....

CAE (even issues)

Macromolecular Chemistry Sections

Applied Chemistry and Chemical Engineering Sections

Physical and Analytical Chemistry Sections

C = CBAC, Chemical-Biological Activities

EE = Ecology and Environment

En = Energy

F = Food and Agricultural Chemistry

M = Materials

P = POST, Polymer Science and Technology

Chemical-Biological Activities   CBAC

This name is now used by Chemical Abstracts for eight of its eighty sections: 1-5, 62-64. It differs from Chemical Condensates, which has all eighty subject sections, by carrying the full abstract, the preferred names and molecular formulas and Registry numbers of all the reported compounds, and beginning in 1975 the subject headings (fixed language) and modifier lines (free language) that are used in the analytical subject index which is published twice a year. The subject section numbers and the keyterms of each issue of Chemical Abstracts are also included in CBAC, which appears biweekly. Registry numbers are published in book form; also the "Name-Match" computer program at ORNL can be used to find the Registry number together with the preferred name and all the trivial names from any one name. Chemical symbols are not used in title or text, only full spellings; unlike BioAbstracts there are no spaces between identifiable fragments of compound names.

Data Elements

All the data elements of CA Condensates. Additional elements are:

Molecular formulas, Hill system	MFS	FT
Chemical Abstracts preferred name	CNA	FT
CA Registry number, nine digits with leading zeros	REG	LA
CBAC volume number, use the CA volume number	SJV	LA
Abstract	ABS	FT

Engineering Index   EI

Engineering Index publishes three printed services and one magnetic tape service. They all carry the same citation record and abstract but the index terms and the serial numbers are not the same in all. The weekly Card-A-Lert service for current awareness uses a three-level subject classification. The Engineering Index Monthly (Vol. 12 appeared in 1974) and Engineering Index Annual (Vol. 73 in 1974) group the

abstracts by subject heading and subheading taken from another list.
The Annual cumulates the abstracts from the Monthly but with a revised
serial number. The magnetic tape version, Compendex (for COMputerized
ENgineering indEX), carries all the information of the three printed
issues except for the serial number in the Annual, because it is
derived from the tape used in photocomposition of the Monthly and of
Card-A-Lert. In Compendex the volume number, as in EI Monthly, is the
last two digits of the year (Vol. 74 in 1974).

Selection criteria are established both according to subject matter
and to worth. Engineering is the practical application of science.
Marginal matter, mostly in pure chemistry, pure physics, pure geology,
sociology and biology, is excluded. Articles which are purely commercial,
very elementary, editorials, announcements, news items are excluded.
The areas of acceptance change as the areas of concern to society change.
Pollution, urban problems, safety, food technology, solar energy, coal
liquefaction are among more recent areas of engineering.

In 1974 Engineering Index carried about 90,000 abstracts. There
were two thirds that many items in 1969, the first year of Compendex.
The articles were found in 2000 journals and 1200 proceedings of confer-
ences and symposiums. The journals of American engineering societies
are given express handling and complete coverage. The published list of
the 2000 journals with titles and Coden is called PIE (Publications
Indexed for Engineering).

The subject classifications for the current awareness service
(Card-A-Lert) divide all of engineering by a three-digit code into 6
disciplines (first digit), into 38 groups (first two digits), and into
167 divisions (all three digits). The guide to Card-A-Lert shows both
the codes with their definitions and an alphabetical inversion of the
9000 terms used in the definition of the codes.

The Index Terms, the second indexing scheme, come from SHE
(Subject Headings for Engineering), a controlled vocabulary with 12,000
authorized combinations of index headings and subheadings. Generally
several headings are applied to each item. The Thesaurus of Engineering
Terms, published by the Engineers Joint Council, is not directly used

by Engineering Index for vocabulary control.

Free-language index terms are occasionally assigned; if used the rules permit no more than five for one article. The terms may come from the title or abstract.

The scope of Engineering Index is shown by the titles of the six disciplines used in the Card-A-Lert service: 400, Civil- Engironmental- Geological- Bioengineering; 500, Mining- Metals- Petroleum- Fuel Engineering; 600, Mechanical- Automotive- Nuclear- Aerospace Engineering; 700, Electrical- Electronics- Control Engineering; 800, Chemical- Agricultural- Food Engineering; 900, Industrial Engineering- Management- Mathematics- Physics- Instruments.

Data Elements

Title TTL, year of publication PDD,		
Subject code (Card-A-Lert), as A522	SCC	FT
Controlled vocabulary from SHE	FIT	FT
Free-language index terms (occasional)	FIT	FT
Author, as Smith, R.S.	ATI	LA
Abstract	ABS	FT
EI volume number (tape issue), 75 in 1975	SJC	FT

Government Reports Announcements GRA

This journal was formerly U.S. Government Research and Development Reports. Indexing is pluralistic. For subject categories, the publisher NTIS must use the twenty-two Cosati fields with their groups (dates from the 60s). NTIS uses another twenty-three fields for its WGS – Weekly Government Abstracts (dates from the 70s). What NTIS calls descriptors (index terms from an authority list) are assigned by NASA from the "INIS Thesaurus" for its reports, by the DDC from its "Thesaurus of Engineering and Science Terms" for Department of Defense reports, and by NTIS for the remainder from the quite similar "Thesaurus of Engineering Terms." Additionally NTIS adds identifiers (free index terms or keywords) to all the reports. Our text search program combines

the descriptors and identifiers under FIT. Abstracts published by NASA in STAR or by the AEC in ERDA are referred to but not recorded on the tape; for other reports there is an abstract on the tape.

Data Elements

Title TTL, Corporate author CAN, Publication year PDD,

Author, as Ferguson, Edward G.	ATI	LA
Descriptors and identifiers, index terms	FIT	FT
Subject fields and groups, Cosati or NTIS (a leading zero is used for fields less than 10)	CSC	FT
Abstract	ABS	FT
Report number	RNO	LA
Contract grant number	CNO	FT
The NTIS document serial number (accession no.)	DSN	LA
Document availability	DAV	FT
Volume number of GRA, 75 in 1975	SEC	FT

Subject Classifications in GRA

In the printed issues of GRA, the Cosati field and group assignments are shown; each item has at least one. Not all items appear in WGA, Weekly Government Abstracts but those that do carry one or more NTIS field and group assignments. The tape issue carries both sets of numbers. The list below shows only the fields (divisions) and not the groups (subdivision).

Cosati fields

01. Aeronautics
02. Agriculture
03. Astronomy and Astrophysics
04. Atomspheric Sciences
05. Behavioral and Social Sciences
06. Biological and Medical
07. Chemistry
08. Earth Sciences and Oceanography
09. Electronics and Electrical Engineering

10. Energy Conversion (non-propulsive)
11. Materials
12. Mathematical Sciences
13. Mechanical, Industrial, Civil and Marine Engineering
14. Methods and Equipment
15. Military Sciences
16. Missile Technology
17. Navigation, Communications, Detection, and Countermeasures
18. Nuclear Science and Technology
19. Ordnance
20. Physics
21. Propulsion and Fuels
22. Space Technology

NTIS fields

45. Communications
46. Physics
47. Ocean Technology and Engineering
48. Natural Resources
49. Electrotechnology
50. Civil and Structural Engineering
57. Medicine and Biology
62. Computers, Control and Information Theory
68. Environmental Pollution and Control
70. Administration
71. Materials Sciences
85. Transportation
88. Library and Information Sciences
89. Building Technology
90. Government Inventions for Licensing
91. Urban Technology
92. Behavior and Society
94. Industrial and Mechanical Engineering
95. Biomedical Technology and Engineering
96. Business and Economics
97. Energy
98. Agriculture and Food
99. Chemistry

Metals Abstracts and Metals Index   MET

The American Society for Metals and the Institute of Metals (London) jointly publish a monthly index of metallurgy and metals. All articles in 100 journals are indexed and another 1000 journals are used selectively to give 28,000 entries a year, half from English-language sources and half from foreign-language sources. Ninety per cent of the entries come

from journals, five per cent from reports and five per cent from
monographs or proceedings. There are no patents.

The bibliographic citations in Metals Abstracts and the index
headings in Metals Index are produced from the magnetic tape. The
index headings are selected from a controlled vocabulary "Thesaurus
of Metallurgical Terms" and are most often heading/subheading pairs;
for example, Oxygen/sorption. On the average six index headings are
applied to one citation. In addition there are alloy designations, used
for publication of Alloys Index, that allow a finer division of the
alloy types along with a statement of compound, alloy, or system.

To allow us to distinguish between an element and its mixtures, the
index pairs are separated by a comma plus a blank space; 'cobalt*" will
retrieve all references beginning with cobalt, but 'cobalt,ɒ*' (where
ɒ means a single blank space) will disqualify the two-word phrases with
cobalt. Unfortunately, the pair 'cobalt, alloying element' or the pair
'cobalt, binary system' confuse the matter of whether cobalt followed
by a comma is pure. Only one subject classification is assigned to an
abstract; the two-digit number code for the subject is used as the
first two digits of the accession number.

SUBJECT CLASSIFICATIONS

11	Constitution	42	Extraction and Smelting
12	Crystal Properties	43	Refining and Purification
13	Lattice Defects	44	Physical Chemistry of Extraction and Refining
14	Structural Hardening		
15	Physics of Metals	45	Ferrous Alloy Production
16	Irradiation Effects	46	Non-Ferrous Alloy Production
21	Metallography	51	Foundry
22	Testing and Control	52	Working (Forming)
23	Analysis	53	Machining
		54	Powder Technology
31	Mechanical Properties	55	Joining
32	Physical Properties	56	Thermal Treatment
33	Electrical and Magnetic Phenomena	57	Fininshing
34	Chemical and Electrochemical Properties	58	Metallic Coating
35	Corrosion	61	Engineering Components and Structures
41	Ores and Raw Materials	62	Composites

63 Electronic Devices

71 General and Non-Classified
72 Special Publications

Data Elements

Title TTL, year of publication PDD,

Author, as Farrell, D.E.	ATI	LA
Index term	FIT	FT
Subject classification, as 21	3JA	LA
Journal title abbreviation; no Coden	PRC	FT
Volume number, as 1975 in 1975	SJD	FT

National Agricultural Library   CAIN

The CAIN tapes (CAtaloging and INdexing) are used monthly to print two indexes. The "Bibliography of Agriculture" (Oryx Press) is a monthly index to the periodical literature received at NAL. The "National Agricultural Library Catalog" (Rowman and Littlefield) is a monthly catalog of books, new serials, translations and other things. Neither has abstracts or index terms. Each item is assigned to one (rarely more) subject category from a list of 63 common and 7 infrequent categories. To perhaps one fourth of the items one or two words, subject phrase or a genus-species name, are added either to the title or to a separate index term list. Foreign titles are translated into English; both are searchable. The call numbers that appear on the tape record are an internal assignment in the NAL library and are different from the identification numbers in the printed issues of CAIN. Also the April CAIN tape carries the items that appear in the May "Bibliography."

The CAIN tape, but not the "Bibliography" or "Catalog," is extended to include citations compiled by the Food Nutrition Information Center and citations from the "American Bibliography of Agricultural Economics." These are indexed and abstracted, unlike the direct NAL citations. Food and Nutrition citations can be recognized by F&N following the NAL call number (e.g., HD366.R30 F&N) and can be separately searched for by use

of the field NAL. They are assigned to the CAIN categories. The Agricultural Economics citations are assigned to three CAIN categories and to six additional categories.

Data Elements

Title, including some augmentation
  For retrospective search before 7402,
  use both TTL and TFL
	TTL	FT

Title in foreign language	TFL	FT
Other augmentation	FIT	LA

Author, as Jones Arthur B
       or Jones A B
       or Jones A
	ATI	LA

Subject classification, six digit code	CCC	LA
Location of work	LOW	FT
Year of publication of article or book, as 1974	PDD	FT
Library call number, as TX341.J6 F&N RJ1.A453	NAL	LA
Abstract (for Food Nutrition, Agric. Econ.)	ABS	FT

SUBJECT CATEGORIES AND SUBJECT CODES IN 'CAIN', 1971 to date

050500	General agriculture and rural sociology
100500	General agricultural economics and land economics
101000	Agricultural administration and management
101500	Agricultural production costs and returns
102000	Agricultural production distribution
102500	Statistical data and methodology
103000	Outlook, policies, programs, and legislation
150500	Consumer protection
151000	Human nutrition
151500	Home economics
200500	Agricultural products - general
201000	Dairy products
201500	Livestock products
202000	Poultry products
202500	Field crop products
203000	Horticultural products
203500	Feed products
250500	Animal husbandry
251000	Livestock biology
251500	Livestock feeding

252000	Livestock breeding
300500	Veterinary medicine
301000	Infectious and parasitic diseases
301500	Non-infectious diseases
302000	Miscellaneous diseases
350500	Forestry - general
351000	Forest management
351500	Silviculture
352000	Forest industries
400500	General plant science
401000	Plant taxonomy and geography
401500	Plant ecology
402000	Plant morphology, anatomy and cytology
402500	Plant genetics and breeding
403000	Plant physiology and biochemistry - general
403500	Physiology and biochemistry of field crops
404000	Physiology and biochemistry of horticultural crops
404500	Physiology and biochemistry of forest trees
405000	Field crops - culture
405500	Horticultural crops - culture
406000	Miscellaneous economic plants - culture
450500	Plant fungus diseases and control
451000	Plant bacterial diseases and control
451500	Plant virus diseases and control
452000	Miscellaneous plant diseases, injuries, and control
452500	Weeds and herbicides
453000	Insect pests and controls - general and miscellaneous plants
453500	Insect pests and controls - field crops
454000	Insect pests and controls - horticultural crops
454500	Insect pests and controls - forest trees, products
455000	Insect pests and controls - products
455500	Insect pests and controls - animal and man
456000	Pesticides - general
500500	General entomology
501000	Taxonomic entomology
501500	Apiculture and sericulture
550500	Agricultural engineering
551000	Farm equipment
600500	Soil science
601000	Soil improvement materials
601500	Soil resources and management
602000	Water resources and management
650500	General natural resources and environmental pollution
700500	*Life sciences
750500	*Physical sciences and mathematics
800500	*Chemistry
850500	*Technology
900500	*Economics and administration
950500	*Soil sciences and humanities
970500	*Information science

*These categories are infrequently used.

Supplementary Subjects for "Agricultural Economics"

1010*	Agricultural marketing
1020*	Agricultural policies and programs
1030*	Agricultural products, demand, supply and prices
1040*	Food and consumer economics
1050*	Foreign development
1060*	Production economics and farm management
1070*	Regional and human development
1080*	Resource economics
1090*	General

Nuclear Science Abstracts   NSA

All but the text of the abstract appears on the magnetic tape version of this semimonthly journal. The subject categories, a five-digit code, come from the "Extended Subject Arrangement of Material in NSA (TID-26500)"; two or three subjects are frequently assigned to an item but only from 1975. The coordinate index terms come from the controlled list in the "INIS Thesaurus-INIS-13". But be careful; in progress reports the division by chapters in the indexing (splits) is not recognized by the computer and false drops come from the first profile term in one chapter, the second term in another.

<u>Data Elements</u>

Title TTL, Location of work LOW, Corporate author CAN, year of
   publication PDD

Author, as Parsons, R. A. (from 1974)	ATI	LA
Report number	RNO	LA
Index terms (use right truncation before 1974)	NST	LA
NSA subject codes, five digits (four before 1974)	NCC	LA
Contract grant number	CNO	FT
Coden, five digits	CDN	LA
NSA volume number, 31 and 32 in 1975	SJV	LA

NSA will not be published after June 1976. The nuclear science already appears in "Atomindex" published by the International Atomic Energy Agency with the same standards as NSA. The general field of energy, and including nuclear power, appears in EIDB (Energy Information Data Base) published by ERDA with the same standards as NSA.

THIS PAGE

WAS INTENTIONALLY

LEFT BLANK

INTERNAL DISTRIBUTION

1.	A. A. Brooks		30.	C. R. Nook
2.	J. R. Buchanan		31.	K. M. Olsen
3.	M. M. Cardwell		32.	J. H. Owings
4.	Z. Combs		33.	L. J. Peck
5.	T. F. Connolly		34.	H. Pfuderer
6.	R. R. Dickison		35-56.	H. S. Pomerance
7.	W. B. Ewbank		57.	C. E. Price
8.	J. R. Ferguson - K-25		58.	L. W. Rickert
9.	J. M. Fielden		59.	M. D. Shelby
10.	H. B. Gerstner		60.	V. A. Singletary
11-15.	C. A. Giles		61.	C. S. Strickler
16.	J. S. Gillen		62.	D. K. Trubey
17.	J. C. Gillespie (PAD)		63.	G. U. Ulrikson
18.	M. P. Guthrie		64.	L. H. Veach
19.	A. S. Hammons		65.	H. S. Warren
20.	C. F. Holoway		66.	J. S. Wassom
21.	J. K. Huffstetler		67.	J. A. Watts
22.	H. T. Kemp		68.	D. A. Wilkes
23.	H. F. Kuhns		69.	S. L. Yount
24.	L. R. Lewis		70-71.	Central Research Library
25.	A. S. Loebl		72-91.	CSD Library
26.	W. B. Malthouse		92.	Document Reference Section-Y-12
27.	B. F. Maskewitz		93-94.	Laboratory Records
28.	H. J. McLain		95.	Laboratory Records - RC
29.	H. F. McDuffie		96.	ORNL Patent Office

EXTERNAL DISTRIBUTION

97. Research and Technical Support Division, Oak Ridge Operations, Oak Ridge, Tennessee 37830
98. D. H. Moneyhun, Technical Information Center, Oak Ridge, Tennessee 37830
99. D. E. Bost, Technical Information Center, Oak Ridge, Tennessee 37830
100. T. E. Hughes, CXXI A15204, Office of Public Affairs, Energy Research and Development Administration, Washington, D.C. 20545
101-127. Technical Information Center, Oak Ridge, Tennessee 37830