

A New Approach to Information Retrieval Problems in Separations Science

W. J. MCDOWELL*

Chemical Technology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830

BARBARA B. COREY

Computer Sciences Division, Union Carbide Nuclear Division, Oak Ridge, Tennessee 37830

Received November 17, 1983

Retrieving information on specific chemical separations is among the most difficult problems in information management, although the ability to find methods to cleanly and efficiently separate chemical species is of the utmost importance to chemists and chemical engineers. Information on performing specific chemical separations is largely buried in the literature of dozens of branches of science. Most methods of indexing (both hard copy and computer) do not provide good means of retrieving information on specific separations because index terms such as extraction, leaching, chromatography, and even ion exchange have different meanings in different disciplines. Recent attempts to solve some of the problems of information retrieval in separations science have resulted in the concept of a Separations Science Data Base. This data base is designed for the chemical separations information and contains unique indexes that allow rapid and accurate retrieval of information about specific separations from specific matrices and a method of minimizing false returns that result from cross coupling of unrelated terms in multisubject reports. Although the data base is presently only about 20% complete, the success of this work has been encouraging, and further work is indicated.

DISCUSSION

Finding information on specific chemical separations is often difficult because, in its broadest sense, separations science could include all science, and separations procedures and methods are scattered in journals of various scientific disciplines with varying nomenclature used in describing them. For example, liquid-liquid extraction is called "solvent extraction" in the field of separations of inorganic ions and molecules, but it is generally called "phase transfer" in biochemistry and organic chemistry. Similarly, "extraction" means removing ore from the ground to a mining engineer, removing metal from an ore to a metallurgist, and removing an essence from an herb to an organic or pharmaceutical chemist. There are many more such examples. These differences in nomenclature make indexing and retrieving of separations information particularly difficult. In attempts to retrieve information on specific hydrometallurgical separations by either manual or computer methods, it was found that existing indexing systems either located many references, 90% of which were of no interest, or (if the search was restricted to solve that problem) did not identify much of the needed material. The situation would probably be equally bad in attempting to retrieve information on biochemical or organic chemistry separations.

The Separations Science Data Base was developed at Oak Ridge National Laboratory (ORNL) by members of the Separations Science Research Group to solve the above problems as they relate to solvent extraction and ion-exchange separations and also to index a body of early separations science literature that was poorly and incompletely indexed. Most of the now large hydrometallurgical solvent-extraction industry grew out of the separations developed in the nuclear industry. Some of the first applications for solvent extraction separations were in winning uranium from ores, in purifying uranium compounds, in reprocessing nuclear fuels, and in separating various radionuclides. Much of this literature was originally in classified internal reports. Many of those in the period 1945-1967 are poorly indexed and are now difficult to find. If the indexing of this material can be accomplished while scientists familiar with its whereabouts are still available, the repetition of considerable experimental work may be avoided.

The data base is stored (along with about 40 other data bases) on the computer system at ORNL, and the data base can be searched and the information retrieved through the Department of Energy data base management program DOE/RECON (REmote CONsole).¹ Access is through local hard-wired terminals, commercial telephone lines, TYMNET, or the Federal Telecommunications System (FTS). All Department of Energy employees or contractors and state or federal employees engaged in energy-related work may have direct access to the DOE/RECON system. In addition, special arrangements can be made for others to have direct access under certain circumstances. The cost of using RECON is similar to other data base search systems. An average search lasting 10-15 min costs \$4.00-6.00, and references printed off-line are charged at \$0.25 per page. In addition to direct searching, customers can have searches performed by

Western Regional Information Service Center
Lawrence Berkeley Laboratory
Building 50, Room 130
Berkeley, California 94720
Telephone: (415) 486-6307

The Separations Science Data Base has several unique features that make it especially useful to both users and abstractors in the separations science field, and these features may have applications to data bases in other fields. The most common question asked by, and of, a separations chemist is How do I separate A from B out of matrix M? We have constructed our data base to answer that and related questions. In order to accomplish this, the subject key terms that are usually lumped together under "subject index" are categorized as shown in the right-hand column of the table in Figure 1; that figure shows the introductory material that is displayed when the data base is accessed. Carefully selected descriptor terms within these fields describe the information contained in a report or paper. Dividing the subject index terms into categories in this way makes it possible, by means of Boolean AND-OR-NOT combinations, to retrieve specific information on specific separations under specific conditions. Searching of the bibliographic fields for author, title, data, and literature type can also be done, of course, and it is also possible to search the abstracts and titles for any chosen word or words. While

BEGIN SESSION 0013--FILE 35 WAS SELECTED
 ?SEP SEPARATIONS SCIENCE DATA BASE
 (3,978 RECORDS RELOADED ON 02/07/83)

This data base is designed specifically for the storage and retrieval of information needed in chemical separation problems. Initial entries are concerned primarily with solvent extraction, ion exchange and related fields. Available indexes are:

AU-AUTHOR SS-SEP. SYSTEM
 TL-TITLE IT-SEPARATED SUBST.
 YR-DATE SA-SEPARATION AGENT
 LT-LITERATURE MA-MATRIX
 TI-TYPE OF INFO.

Each of the indexes in the right-hand column contains index terms describing information within that index name. By selective use of combinations of indexes and index terms, information on specific separations under specific conditions can be retrieved.

SEPARATIONS SCIENCE DATA BASE Page 2
 In SEPARATED SUBST. index, substance name is followed by notations giving additional information about the separation. Name only indicates substance is extracted (or retained on column or filter) or that specific info. on a separation not given.
 Name - = substance rejected
 Name , - = subst. optionally collected or rejected
 Name +- = substance splits
 Name? - info. in doubt
 Roman numerals indicate oxidation state.
 Text search avail. on title and abstr.

For further information contact:
 Separations Science Data Base
 Bldg. 4500S, Mail Stop C-256
 Oak Ridge National Laboratory
 P.O. Box X
 Oak Ridge, TN 37830
 (615) 574-6714, PTS 624-6714 • ?-

Figure 1. Introductory material that is displayed when the Separations Science Data Base is accessed on DOE/RECON.

ENTER:EXPAND COPPER (a)				ENTER:EXPAND IRON (b)			
>PROCESSING<				>PROCESSING<			
EXPAND		IT-COPPER		EXPAND		IT-IRON	
REF	CIT	RT	DESCRIPTOR	REF	CIT	RT	DESCRIPTOR
E01	74		IT-COBALT(II)	E01	2		IT-IRIDIUM(III)-
E02	17		IT-COBALT(II)-	E02	8		IT-IRIDIUM(IV)
E03	10		IT-COBALT(II),-	E03	10		IT-IRIDIUM(IV)-
E04	11		IT-COBALT(III)	E04	1		IT-IRIDIUM(VI)
E05	1		IT-COBALT(III)-	E05	3		IT-IRIDIUM(VI)-
E06	284		IT-COPPER	-E06	138		IT-IRON
E07	1		IT-COPPER +-	E07	2		IT-IRON +-
E08	54		IT-COPPER -	E08	117		IT-IRON -
E09	16		IT-COPPER ,-	E09	16		IT-IRON ,-
E10	1		IT-COPPER ?	E10	20		IT-IRON(II)
E11	9		IT-COPPER(I)	E11	35		IT-IRON(II)-
E12	1		IT-COPPER(I)-	E12	3		IT-IRON(II),-
E13	78		IT-COPPER(II)	E13	170		IT-IRON(III)
E14	11		IT-COPPER(II)-	E14	41		IT-IRON(III)-
E15	10		IT-COPPER(II),-	E15	23		IT-IRON(III),-
E16	93		IT-CURIUM	E16	1		IT-KRYPTON
E17	11		IT-CURIUM -	E17	1		IT-KRYPTON -
E18	13		IT-CURIUM ,-	E18	371		IT-LANTHANIDES
E19	29		IT-CURIUM(III)	E19	46		IT-LANTHANIDES -
E20	4		IT-CURIUM(III)-	E20	19		IT-LANTHANIDES ,-

ENTER:SELECT E6,E9,E13,E15		ENTER:SELECT E8,E9,E14,E15	
>PROCESSING<		>PROCESSING<	
1	388 E6,E9,E13,E15	2	197 E8,E9,E14,E15
		ENTER:COMBINE 1 AND 2	
		>PROCESSING<	
		3	52 1 AND 2

Figure 2. Examples of the result of the RECON "Expand" command showing how keywords for the separated substance are selected and combined.

the references placed in the data base to date are primarily concerned with solvent-extraction or ion-exchange separations, the SEPARATION SYSTEM field contains terms designating various separations systems, e.g., extraction, precipitation, distillation, etc., thus making the indexing system potentially useful for a wide range of types of separations by expanding the appropriate index term list.

Two additional features enhance the accuracy, speed, and convenience of searching the data base. First, since the primary purpose of most searches will be to retrieve information on the separation of various substances, the SEPARATED SUBSTANCE index has been made the default index on the DOE/RECON system; that is, a search for information on the separation of a particular substance can begin without

specifying the SEPARATED SUBSTANCE index. Thus, the most useful kind of search is begun most easily. Second, the SEPARATED SUBSTANCE index contains notations following the substance name that give additional information about the separation, as is shown in Figure 1. For example, the use of the substance name alone and unmodified indicates that the substance is retained in the separation (extracted, held on column or filter, etc.) if that information is given. Other notations are as defined in Figure 1. Indexing in this way allows the user to include or exclude as much of the material as is desired.

The indexes for the various fields are contained in the DOE/RECON memory, and selected portions of the indexes may be viewed by the use of an "EXPAND" command. For

```

ENTER:DISPLAY 10/3/1-4
>PROCESSING<
DIS 10/3/000001-000004//1
<ACCESSION NO.> 83*0002762      *****1
<AUTHOR> Agarwal, J.C.; Barner, H.E.; Beecher, N.; Davies, D.S.; Kust, R.N.
<TITLE> Kennecott Process for Recovery of Copper, Nickel, Cobalt and Molybdenum
        from Ocean Nodules
<SOURCE> Mining Engineering V31(12), pp1704-1707
<DATE> 1979
<ORGANIZATION> Charles River Associates, Boston, Mass; Lexington Development
        Center, Kennecott Copper Corp., Lexington, Mass.

<ACCESSION NO.> 83*0002663      *****2
<AUTHOR> Sohn, H.Y.
<TITLE> Developments in Physical Chemistry and Basic Principles of Extractive
        Metallurgy in 1980
<SOURCE> J. Met. V33(4), pp18-23
<DATE> 1980
<ORGANIZATION> Univ. of Utah, Salt Lake City, UT

<ACCESSION NO.> 83*0002549      *****3
<AUTHOR> Spisak, J.F.; McClincy, R.J.
<TITLE> Solvent Extraction of Copper from Smelter Dust Treatment Liquors with
        Acorga P-5100
<SOURCE> Proceedings, ISEC '80, Int. Solvent Extr. Conf. V2, S13, Paper 99
<DATE> Sep 1980
<ORGANIZATION> Federal-American Partners, Riverton, WY; Anaconda Copper Co.,
        Denver, CO
<PAPER NUMBER> 2-13-099

<ACCESSION NO.> 83*0002542      *****4
<AUTHOR> Martin, T.P.; Davies, G.A.
<TITLE> Extraction of Copper from Aqueous Solutions Using a Liquid Membrane
        Process: A Model to Simulate the Process
<SOURCE> Proceedings, ISEC '80, Int. Solvent Extr. Conf. V2, S12, Paper 230
<DATE> Sep 1980
<ORGANIZATION> Dept. of Chemical Engineering, Univ. of Manchester, Manchester,
        England
<PAPER NUMBER> 2-12-230

```

Figure 3. Short bibliographic listings resulting from a DOE/RECON search on the separation of copper from iron described in the text.

```

a  <ACCESSION NO.> 09Z0003714      *****35
    <PARENT REF.> See accession No. 00P0003705

    <ACCESSION NO.> 09Z0003675      *****36
    <PARENT REF.> See accession No. 00P0003666

    <ACCESSION NO.> 08Z0004138      *****37
    <PARENT REF.> See accession No. 00P0004130

    <ACCESSION NO.> 08Z0003832      *****38
    <PARENT REF.> See accession No. 00P0003824

b  DISPLAY 00P0003705/3
    >PROCESSING<
    DISPLAY 00P0003705/3          PAGE    1
    <ACCESSION NO.> 00P0003705      *****
    <AUTHOR> Crouse, D.J.; Brown, K.B.; Arnold, W.D.; Moore, J.G.; Lowrie, E.S.;
        Kelmers, A.D.; Debnam, J.
    <TITLE> Progress Report on Uranium Extraction with Organonitrogen Compounds
    <SOURCE> ORNL-2099, 63pp
    <DATE> 1956; Jun
    <ORGANIZATION> Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA

```

Figure 4. (a) Cross-references to multisubject reports resulting from split-subject indexing (printout of accession numbers only; optionally, the set of index terms could also have been printed). (b) Parent record referenced by one of the split records (00P0003705) in (a) (bibliographic printout only; optionally, abstract and/or complete index terms could have been printed).

CAREFULLY CIRCLE OR HIGHLIGHT THE APPROPRIATE TERM OR TERMS FROM EACH TABLE

<SEP SYS>

(Separation System)

1 Adsorption	8 Electrolysis	27 Gradient	21 Sedimentation
2 Centrifugation	9 Electrophoresis	15 Ion Exchange	22 Solid-Gas
3 Chromatography	10 Electrostatic Ppt.	16 Liquid-Liquid	23 Solid-Liquid
4 Crystallization	11 Extraction	17 Magnetism	24 Solid-Solid
5 Diffusion	12 Filtration	18 Osmosis	25 Sublimation
6 Distillation	13 Flotation	19 Photolysis	26 Zone Refining
7 Electrodeposition	14 Gas-Liquid	20 Precipitation	

<INFOTYPE>

(Type of Information)

1 Acidity effect	9 Constant	17 Mathematical modeling	25 Seprn. degree
2 Agent compn. effect	10 Decontamination factor	18 Matrix compn. effect	26 Sep. sub. concn. effect
3 Agent concn. effect	11 Device	19 Matrix concn. effect	27 Solubility
40 Agent loss	12 Diluent compn. effect	20 Mechanism	48 Species
41 Agent preparation	42 Elution curve	44 Particle size effect	28 Stoichiometry
47 Agent stability	38 Flow rate effect	21 pH effect	49 Stripping
analytical method	13 Flowsheet	45 Phase bet.	29 Synergism
ism	14 Intensive properties	46 Ph	Temp. effect
	36 Interfacial phenomena	22	mododynamics
	15 Isotherm		
	16 Kinetics		
	mass transfer		

<SEP SUB>

(Separated Substance)

110 Acetate	79 Gold	145 Phosphate
111 Actinides	72 Hafnium	15 Phosphorus
89 Actinium	2 Helium	78 Platinum
112 Aerosol	67 Holmium	94 Plutonium
113 Alcohol	127 Hydrocarbon	84 Polonium
114 Alkali	1 Hydrogen	146 Polycyclic aromatic hydrocarbons
13 Aluminum	49 Indium	19 Potassium
95 Americium	128 Inorganic acid	59 Praseodymium
115 Ammonia	129 Inorganic base	61 Promethium
116 Ammonium	130 Iodate	91 Protactinium
117 Anion	131 Iodide	88 Radium
51 Antimony	53 Iodine	86 Radon
18 Argon	77 Lanthanum	147 Reagent
33 Arsenic	26 Lithium	75 Rhenium
Astatine	36 Manganese	45 Rhodium
Barium	132 Molybdenum	37 Ruthenium
Berkelium		44 Rutherfordium
Beryllium		62 Seaborgium
biological molecules		21 Tennessine

Figure 5. Portions of the keywording form for the Separations Science Data Base showing the method of indicating index terms.

example, suppose we are interested in separating copper from iron. The command EXPAND COPPER would allow the user to view the portion of the SEPARATED SUBSTANCE index shown in Figure 2a. From this index we would then select E6, E9, E13, and E15. From the portion of the index displayed in response to the EXPAND IRON command (Figure 2b), we would select the entries in which trivalent iron is rejected, E8, E9, E14, and E15. We now have two sets of references. Combining these sets with the command COMBINE 1 AND 2 will produce a set of references in which copper is retained and iron is rejected. Similarly, we can choose from the SEPARATION AGENT index the term

"hydroxyoxime" and from the matrix index the terms "aqueous", "acid", and "sulfate" and, by combining all of these sets with the appropriate Boolean logic, obtain a set of references in which copper is extracted and iron is rejected out of an aqueous acid-sulfate system by the use of a hydroxyoxime. Figure 3 shows some typical bibliographic listings resulting from such a search. Many similar useful combinations of fields and terms will quickly become obvious to the user.

Another serious problem in computer retrieval of information is the cross coupling of unrelated terms in multisubject reports, resulting in false returns that contain none of the

desired information. This problem has been solved in the Separations Systems Data Base by splitting the input of such reports into two or more entries, each of which contains only the index terms relating to a single subject. Each of the "splits" is a complete and separate record as far as index terms are concerned, but it does not contain the bibliographic information or abstract and simply refers back to the parent record for those. Cross coupling of unrelated index terms is thus effectively eliminated. An example of how such references appear in the output is seen in Figure 4. For this illustration, the index terms "URANIUM" and "IRON -" were used. Many of the references shown here are in progress reports that contain information on many other separations.

Material that is located in the data base can be displayed and/or printed in several formats ranging from a listing of accession numbers only, through brief bibliographic listings, to complete listings containing all of the index terms and the abstract.

The indexing of documents has been made simple and relatively painless by the use of an input form that contains the complete thesaurus for each field in addition to space for the abstract and bibliographic material (Figure 5). The indexer has only to circle the appropriate word or words in each section. An abstractor's manual that gives complete instructions for input to the data base is available.²

The data base contains approximately 4000 entries that include progress reports from a separations research group at

ORNL (1945-1967). Also covered are about 20 journals pertinent to solvent-extraction and ion-exchange research in the period 1978-1982. This is a small but useful fraction of the available solvent-extraction and ion-exchange information. The data base presently has no funding to expand; however, the indexing structure of the data base has demonstrated definite advantages for those wishing to retrieve information on chemical separations and contains some features potentially applicable to data bases in other subject areas. Additional information on the data base can be obtained by contacting the authors.

ACKNOWLEDGMENT

We acknowledge the contributions of C. F. Coleman, B. A. Moyer, J. W. Roddy, and R. R. Shoun of the Chemical Technology Division and D. C. Michelson of the Information Division for their contributions in designing the data base and in abstracting and keywording. This research was sponsored by the Division of Chemical Sciences, U.S. Department of Energy, under Contract W-7405-eng-26 with Union Carbide Corp.

REFERENCES AND NOTES

- (1) "Energyinfo/on-line, The DOE/RECON SYSTEM". U.S. Dep. Energy, Tech. Inf. Cent., [Tech. Rep.] DOE/TIC 1981, DOE/TIC-4609.
- (2) Roddy, J. W.; McDowell, W. J.; Michelson, D. C. "Separations Science Data Base: An Abstractor's Manual". Oak Ridge Natl. Lab., [Rep.] ORNL-TM (U.S.) 1981, ORNL/TM-7805.

Unresolved Problems and Opportunities in Chemical Literature Teaching[†]

SAMUEL H. WILEN

Department of Chemistry, The City University of New York, City College, New York, New York 10031

Received December 30, 1983

The experience of an instructor in teaching a formal chemical literature course in academia is described. The effects of unresolved problems associated with such a course are assessed, including lack of brief yet effective exercises, time limitations, and unavailability of key search tools in libraries. In the larger context, the causes and consequences of widespread lack of formal instruction on the chemical literature are discussed, as well as factors such as unavailability of user aids in libraries, professor-librarian interaction, the need for improvement in key search tools, overemphasis on instruction in the use of machine-readable files, and the issue of informal vs. formal instruction of chemical literature usage.

INTRODUCTION

This analysis of the teaching of chemical literature usage is based largely on experience gained through the delivery of a formal chemical literature course over the past 4 years.¹ This course is, in turn, generally organized along the lines of the American Chemical Society Audio Course "Use of the Chemical Literature", whose genesis and content have been described previously.^{2,3} An important feature of this course is the requirement that students carry out a set of exercises dealing with all facets of the chemical literature.

COURSE PROBLEMS

Problems associated with the course are identified here in the belief that they are of general interest in the conduct of

any chemical literature course, and as introduction to the following section.

The unavailability of appropriate instructional materials for the course has been a problem since the beginning. Textbooks in this field⁴ are very rapidly outdated and too often contain insufficient material to permit students to evaluate, and to learn, when and how to use the major secondary literature search tools. The problem of obsolescence can be overcome to some extent by use of specific user (search) aids (described in footnote 2). Without doubt, a combination of reading assignments would solve this problem if this could be done without overloading the students. The compromise achieved is admittedly unsatisfactory.⁵

The key to an effective chemical literature course is the assignment of realistic exercises involving the use of all major secondary search tools. The principal difficulty associated with these exercises is the time required to work them. Students have uniformly agreed that the exercises are essential; yet, most have felt that the time required to carry them out was ex-

[†] Presented to the Divisions of Chemical Education and Chemical Information, Symposium on Chemical Literature and Information Retrieval in the Chemistry Curriculum, 184th National Meeting of the American Chemical Society, Kansas City, MO, Sept 15, 1982.