MARTIN MARIETTA ENERGY SYSTEMS LIBRARIES
3 4456 0153586 7

ORNL/TM-10348

ornl

OAK RIDGE NATIONAL LABORATORY

MARTIN MARIETTA

An Information Retrieval System for Chemical Separations Scientists

B. B. Corey
N. B. Gove
S. L. Yount
W. J. McDowell

OAK RIDGE NATIONAL LABORATORY
CENTRAL RESEARCH LIBRARY
CIRCULATION SECTION
4500N ROOM 175
LIBRARY LOAN COPY
DO NOT TRANSFER TO ANOTHER PERSON
If you wish someone else to see this report, send in name with report and the library will arrange a loan.

OPERATED BY
MARTIN MARIETTA ENERGY SYSTEMS, INC.
FOR THE UNITED STATES
DEPARTMENT OF ENERGY

Printed in the United States of America. Available from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road, Springfield, Virginia 22161
NTIS price codes—Printed Copy: A03; Microfiche A01

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

ORNL/TM-10348

Energy Division

AN INFORMATION RETRIEVAL SYSTEM FOR CHEMICAL SEPARATIONS SCIENTISTS

B. B. Corey
N. B. Gove
S. L. Yount
W. J. McDowell*

*Chemistry Division

Date Published—January 1987

Prepared by the
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831
operated by
MARTIN MARIETTA ENERGY SYSTEMS, INC.
for the
U.S. DEPARTMENT OF ENERGY
under Contract No. DE-AC05-84OR21400

MARTIN MARIETTA ENERGY SYSTEMS LIBRARIES
3 4456 0153586 7

CONTENTS

LIST OF FIGURES v

ABSTRACT vii

ACKNOWLEDGMENTS ix

1. INTRODUCTION 1

2. SEPARATIONS SCIENCE 3
   2.1 As a Field of Study 3
   2.2 Problem Solving by Computers 3

3. DOE/RECON AS AN INFORMATION STORAGE AND RETRIEVAL SYSTEM 5
   3.1 System Overview 5
   3.2 User Access 6
   3.3 File Structures 6
   3.4 Keywording 7
   3.5 Querying the System 8
   3.6 File Updates 10

4. SEPARATIONS SCIENCE DATA BASE 15
   4.1 Unique Keywording 15
   4.2 Splitting Records 15
   4.3 Developing a Search Strategy 16

5. THE FUTURE OF SEPSYS 21

6. SUMMARY 23

7. REFERENCES 25

LIST OF FIGURES

3.1 System overview of DOE/RECON 5

3.2 RECON file structure for display of search results 6

3.3 RECON file structure for keyword search 7

3.4 SEPSYS subject approach indexes 8

3.5 Sample SEPSYS input form 9

3.6 SEPSYS other searchable fields 10

3.7 SEPSYS display of partial list of separated substances 11

3.8 SEPSYS select items from partial list 12

3.9 SEPSYS special search options 12

3.10 RECON file update process 13

4.1 Display of "split-subject indexing" records 16

4.2 SEPSYS display of portion of index 18

4.3 Typical SEPSYS bibliographic record 19

ABSTRACT

Retrieving information on chemical separations is difficult because the literature is scattered among many branches of science; methods are often buried in articles dealing mainly with other topics; existing information services, such as Chemical Abstracts, do not index articles reliably by separation method; and separation method terminology is not consistent across disciplines. Attempts to solve problems of information retrieval in separations science have led to the concept of a Separations Science Data Base. In a cooperative effort, chemical separations scientists and computer analysts have developed a novel concept for supporting separations science data on a very large on-line information retrieval system that was developed primarily for bibliographic data. The system supports unique indexing techniques that provide rapid and accurate access to information about specific separations from specific matrices and minimize false returns that result from cross coupling of unrelated terms in multisubject reports. This data base is uniquely useful to separation scientists because, by the use of appropriate combinations of terms from selected fields, information on separations under specific conditions can be retrieved. These fields include separation system, separated substance, separation agent, matrix, and type of information. The computer system design and the creative approach to structuring of the separations science data are complementary with respect to data entry and indexing, information retrieval, and the various ways in which the results can be displayed at the user's terminal or microprocessor.

ACKNOWLEDGMENTS

The authors wish to thank Candice Fraker, Corky Ives, Ed Soler, and Carole Tubbs for their technical review and assistance in preparing this report.

1. INTRODUCTION

From the beginnings of research at Oak Ridge National Laboratory (ORNL), the role of separations science has been important in supporting the development of nuclear energy. For that reason, many ORNL staff members have developed unequaled expertise in the fields of solvent extraction, ion-exchange, and electromagnetic separations. In support of research in their fields, the Separations Science Research Group began several years ago to collect references to publications that described methods primarily for analytical or hydrometallurgical separations. The documents collected became the Separations Systems Data Base (SEPSYS). In cooperative effort, the Separation Science Research Group and members of the Computer Sciences Division at ORNL developed an appropriate data base design for the reference material. Initially, the program was aimed at proving the feasibility and success of such a data base in hopes that additional funding would be made available to complete the collection of past documents and continue the collection of useful references in the future. It was anticipated that, with a complete data base, many hours of work could be saved by scientists who could use already-developed separations methods and avoid repetition of work.

Several problems had to be solved in the creation of the separations data base. One problem arose because of the need for specific retrieval. Another was related to the desire to reduce false returns.

Often, a journal or report will reference multiple separation processes in which each has different separation agents, methods, and material components. The retrieval software available (DOE/RECON) at that time could not handle retrievals of this complexity. Therefore, an analysis had to be done of the retrieval requirements and the unique solutions designed.

The purpose of this report is to highlight the method by which this complex information has been stored, the technique used to retrieve it, and what considerations must be taken into account when using these data in future retrieval systems.

2. SEPARATIONS SCIENCE

2.1 AS A FIELD OF STUDY

Although scientists were developing general concepts in separations science as early as 1954 which were applicable to the field as a whole,[1] most scientists involved in this field became highly specialized in their particular area of application. As a result, there was very little communication among separations scientists with different specialties and also very little effort to identify the common principles of separations science.

It was not until the late 1970s that formal agencies and organizations were established in an effort to endorse separations science as a unified field. Examples are a Subdivision of Separations Science and Technology within the American Chemical Society's (ACS) Division of Industrial and Engineering Chemistry, and the University of Texas' Separations Research Program. These programs emphasize the need for communication between specialists in different areas of separations science to develop a global approach to solving problems from multiple disciplines.[2]

2.2 PROBLEM SOLVING BY COMPUTERS

Since the early 1970s, computerized retrieval systems have been available to facilitate the location of specific information from large data bases. Many of these systems were not commercially available, however, until around 1975; and most of those were aimed at solving the problems of retrieval from huge legal documents or bibliographic citations. In an attempt to categorize this type of textual data for easy retrieval and use, methods were developed to assign keywords and subject categories to each document. Normally, a subset of documents was formed through a logical combination of two or more keywords.

Unfortunately, in the area of separations science, the nomenclature used in different specialties was often quite different. Words to describe the same procedures and methods varied widely, and the same word often had different meanings to the various specialists. As a result, indexing systems built on the commonly used nomenclatures often led to the wrong information.[3]

In an effort to rectify this problem, the members of the Separations Science Research Group at ORNL began developing a method of indexing chemical separations literature with a consistent nomenclature. They built a demonstration data base of about 4,000 records which became the SEPSYS on the Department of Energy's (DOE) on-line storage and retrieval system called DOE/RECON. This system runs on an ORNL IBM Model 3033. The other data bases on DOE/RECON primarily consist of bibliographic citations from energy-related research literature.

3. DOE/RECON AS AN INFORMATION STORAGE AND RETRIEVAL SYSTEM

3.1 SYSTEM OVERVIEW

The DOE/RECON (REmote CONsole) system was originally developed in the late 1960s by Lockheed for the National Aeronautics and Space Administration as a bibliographic information retrieval system for multiple on-line users. A commercially marketed service (Lockheed's DIALOG) and the DOE/RECON use similar retrieval syntax and file organization but offer generally different data bases. DOE/RECON grew from 1 data base and four users in 1971 to more than 40 data bases and hundreds of users in 1986.

A system overview of DOE/RECON is shown in Fig. 3.1. Further DOE/RECON information is available in Refs. 4 and 5.

ORNL-DWG 87-7294

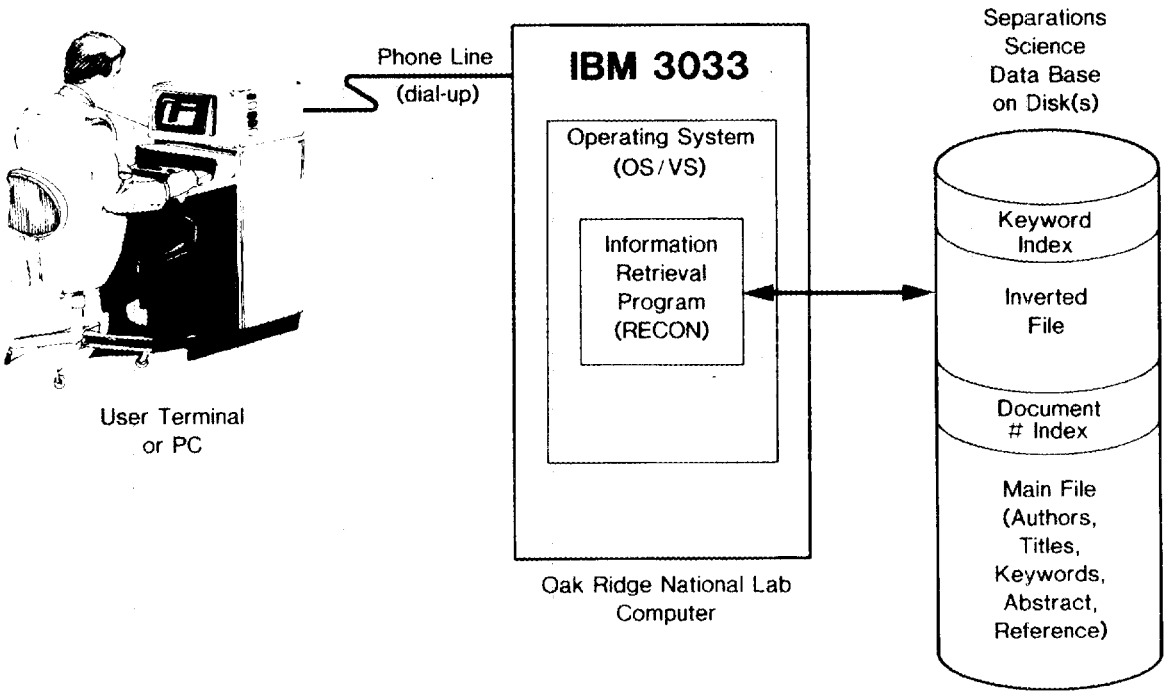

Fig. 3.1. System overview of DOE/RECON.

3.2 USER ACCESS

DOE/RECON is currently sponsored by the Office of Scientific and Technical Information, U.S. DOE. Access to DOE/RECON is through local hard-wired terminals, commercial telephone lines, TYMNET, or the Federal Telecommunications System. All DOE employees or contractors and state or federal employees engaged in energy-related work may have direct access to the DOE/RECON system. The software accepts input from either asynchronous terminals or personal computers using asynchronous terminal emulator software. User access to SEPSYS via DOE/RECON is expected to be available only through the remainder of 1986. Future availability is discussed in Sect. 5.

3.3 FILE STRUCTURES

The DOE/RECON file structure is illustrated in Figs. 3.2 and 3.3. The main data files are usually accessed via keywords or search terms found in one of the data base indexes. Each data

ORNL-DWG 87-7295

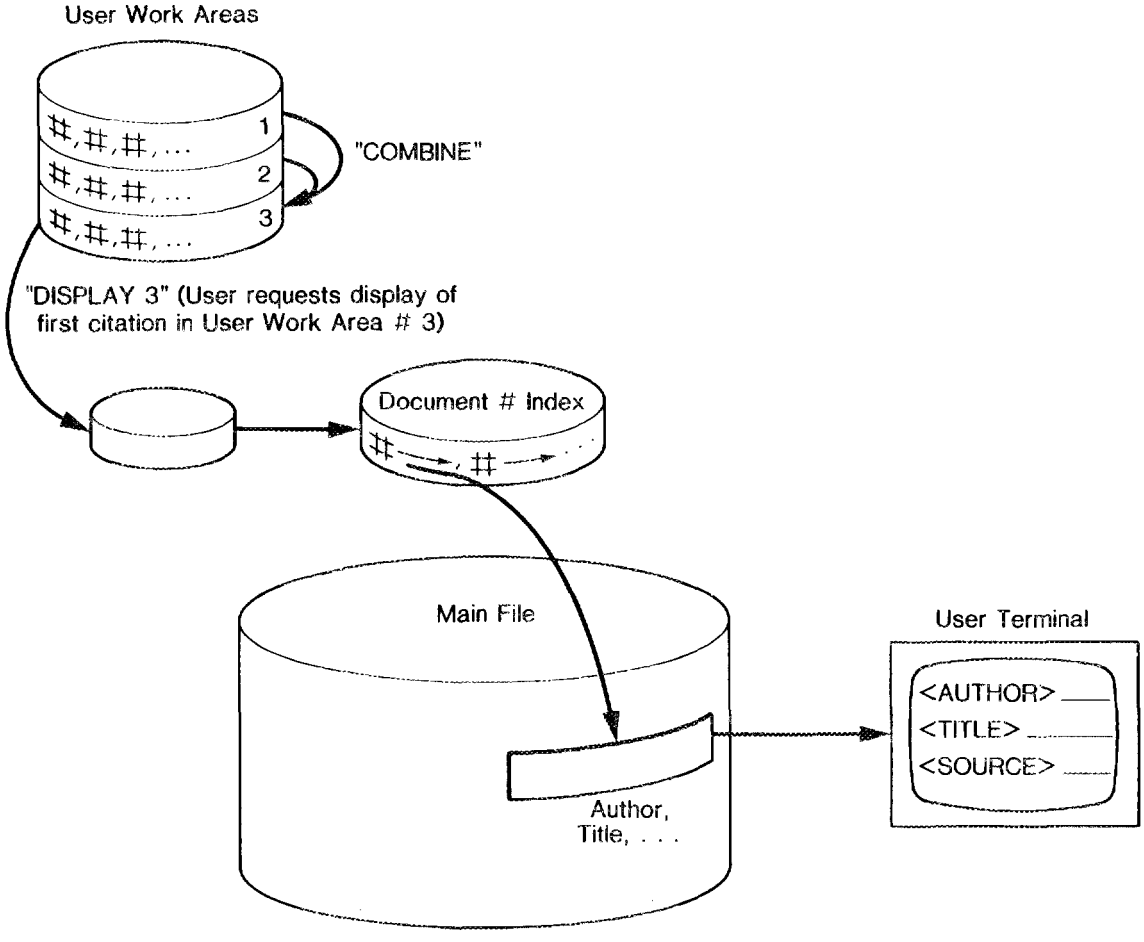

Fig. 3.2. RECON file structure for display of search results.

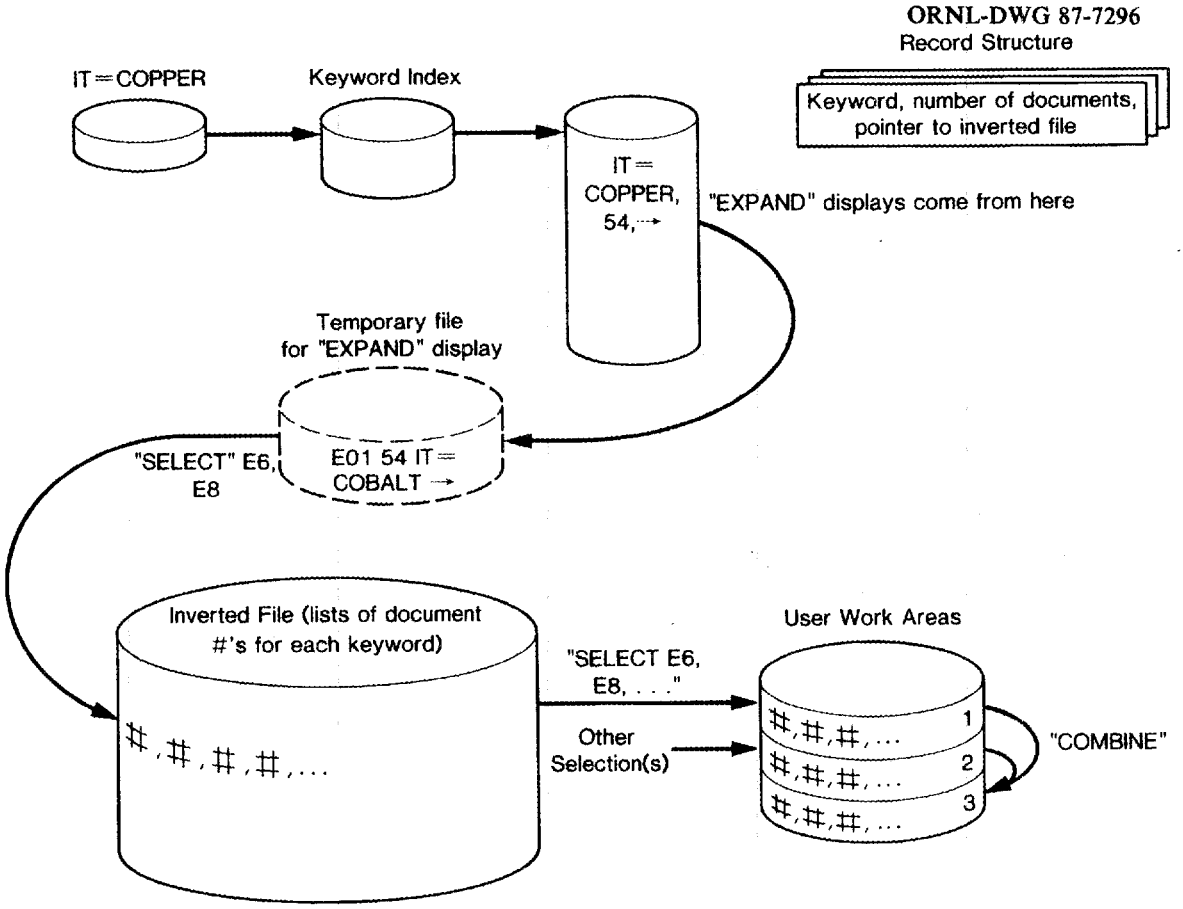

Fig. 3.3. RECON file structure for keyword search.

base is analyzed individually, and indexes are built from the fields. The indexes will best allow the user to form subsets of the data or to isolate exact documents of interest.

In all data bases on DOE/RECON, the different types of search terms (e.g., author, separation agent, matrix) are given prefixes (AU=, SA=, MA=) in the index file. During file updating (see Sect. 3.6), the new search terms are sorted by prefix and merged with existing terms so that the user can easily perform large searches using predefined index points. Indexes for SEPSYS are shown in Fig. 3.4.

3.4 KEYWORDING

The assignment of keywords (including Separation System, Information Type, and Separated Substance) is done by chemists. Instructions for keywording are available in ref. 6. Keyword choices are indicated on a special SEPSYS input form. In most instances, the appropriate keywords can be indicated by simply circling items shown on the input form. The first page of the input form is shown in Fig. 3.5.

RETRIEVAL METHODS

SUBJECT APPROACH INDEXES			
PREFIX	FIELD	AUTHORITY/ DESCRIPTION	EXAMPLE
IT=	Separated Substance	on-line; use ?SEP	IT=GOLD
MA=	Matrix	ORNL/TM-7805	MA=ORE
SA=	Separation Agent	ibid.	SA=UREA COMPOUND
SS=	Separation System	ibid.	SS=SOLID-LIQUID
TI=	Type of Information	ibid.	TI=KINETICS

Each of the indexes 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. In the SEPARATED SUBST. index, the substance name is followed by notations giving additional information about the separation. The name without punctuation indicates the substance is extracted (or retained on column or filter) or that specific information on a separation is not given.

Name - = substance rejected

Name ,- = subst. optionally collected or rejected

Name +- = substance splits

Name ? = information in doubt

Roman numerals indicate oxidation state.

Fig. 3.4. SEPSYS subject approach indexes.

3.5 QUERYING THE SYSTEM

After the dial-up and log-on steps are completed, the SEPSYS user chooses File 35 to begin a retrieval session. The user may search via one of the five types of Subject Indexes [MA, SA, SS, IT, or TI (see Fig. 3.4)] or other searchable fields [AU, TL, LT, or YR (see Fig. 3.6)].

Typically, the user will issue an EXPAND command, such as EXPAND COPPER. This leads to a display of items alphabetically near IT=COPPER, as shown in Fig. 3.7. The user may scan this list, noting the different oxidation states, extraction/rejection indicators, and citation counts. The user may then select one or more entries from this list by entering a command such as that shown in Fig. 3.8.

Methods for some of the special search options unique to SEPSYS are shown in Fig. 3.9.

Each selection leads to an internal transfer of document numbers to a User Work Area. For example, after making five selections, the user will have five User Work Areas. These may be combined with Boolean logic if the user wishes to try a specialized search strategy. Each such combination leads to a new User Work Area.

SEPARATIONS SCIENCE DATA BASE
(SEPSYS)

Input Form

<ACCESSNO> _____

<SPLIT> See accession No. _____

<AUTHOR> _____

<TITLE> _____

<SOURCE> _____

<DATE> _____ **<ABST ID> (Abstract ID)** _____

<ORGANIZ> (Organization) _____

[NOTE TO COMPUTER PERSONNEL]
(If more than one term per field has been selected by the abstractor, separate the terms with a semicolon followed by a space.)

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

<LIT TYPE>
(Literature Type)

1 Bibliography	3 Conference	7 Other	6 Report
2 Book	4 Journal	5 Patent	

UCN-14367
(3 3-83)

Fig. 3.5. Sample SEPSYS input form.

OTHER SEARCHABLE FIELDS*			
PREFIX	FIELD	AUTHORITY/ DESCRIPTION	EXAMPLE
AU=	Author		AU=WAN, C.C.
TL=	Title		TL=LEACHING
LT=	Literature Type		LT=JOURNAL
YR=	Date		YR=1982

*Note that the LOOK command may be used to retrieve title (T) and abstract (A) words, e.g., LK2/a/t/'thiourea'.

Fig. 3.6. SEPSYS other searchable fields.

A special search option (the LOOK command) is also available on DOE/RECON to allow retrieval from User Work Areas by searching for any specified word in the title or abstract. When material of interest is located in the data base, it can be displayed at the user's terminal and/or printed on a high-speed printer in several formats ranging from a listing of accession numbers only, to brief bibliographic listings, to complete listings containing all of the index terms and the abstract.

3.6 FILE UPDATES

A simplified scheme of the DOE/RECON file update process is shown in Fig. 3.10. For the main data file (Linear File or LF), the new material is appended; for the other three files (LX, IX, IF), the old versions are replaced by the new versions. Some extra steps to produce backup files are not shown.

Updating occurs primarily when new records are to be added to the data base. Each data base consists of the four files (LF, LX, IF, IX) and is updated separately as required. Updating is not practical during the day because of the DOE/RECON software; hence, it is performed at night or on weekends. Changes to existing records can be done by full replacement of the record; however, space used to store the old data cannot be reused until a data base reorganization is completed.

The SEPSYS data base is currently not being updated. See Sect. 5 for a discussion of the future use of this data base.

ORNL-DWG 87-7297

User enters: Expand Copper

Screen display: (List of items alphabetically near COPPER)

Expand **IT=Copper**

REF	CIT	RT	Descriptor
E01	74		IT=Cobalt(II)
E02	17		IT=Cobalt(II)-
E03	10		IT=Cobalt(II),-
E04	11		IT=Cobalt(III)
E05	1		IT=Cobalt(III)-
E06	284		IT=Copper
E07	1		IT=Copper +-
E08	54		IT=Copper -
E09	16		IT=Copper, -
E10	1		IT=Copper ?
E11	9		IT=Copper(I)
E12	1		IT=Copper(I)-
E13	78		IT=Copper(II)
E14	11		IT=Copper(II)-
E15	10		IT=Copper(II), -
E16	93		IT=Curium
E17	11		IT=Curium -
E18	13		IT=Curium, -
E19	29		IT=Curium(III)
E20	4		IT=Curium(III)-

Temporary reference for use in SELECT command

Number of citations

Separated Substance

Fig. 3.7. SEPSYS display of partial list of separated substances.

ORNL-DWG 87-7298

User enters: SELECT E6, E9, E11, E13, E15

Screen display: 1 388 E6, E9, E11, E13, E15
(388 references on COPPER EXTRACTION are now in User Work Area #1)

Fig. 3.8. SEPSYS select items from partial list.

ORNL-DWG 87-2799

SPECIAL FEATURES	METHOD
Combine results of two or more searches (inclusive or exclusive)	Boolean logic (and, or, not) on user work areas
Distinguish between extraction and rejection of separated substance	Addition of "-" to substance keyword to signify rejection e.g., IT = IRON -
Select specific oxidation state(s)	Addition of Roman numeral to substance keyword e.g., IT = IRON(III)
Select or restrict search to a particular matrix, separation system, or separation agent	Special prefix: "MA=" for matrix keywords, "SS=" for separation system, "SA=" for separation agent
Avoid cross coupling of unrelated terms in multisubject reports	"Split-subject Indexing" Parent record contains full bibliographic data and abstract. "Split" records for each separate subject contain reference to parent record

Fig. 3.9. SEPSYS special search options.

ORNL-DWG 87M-6241

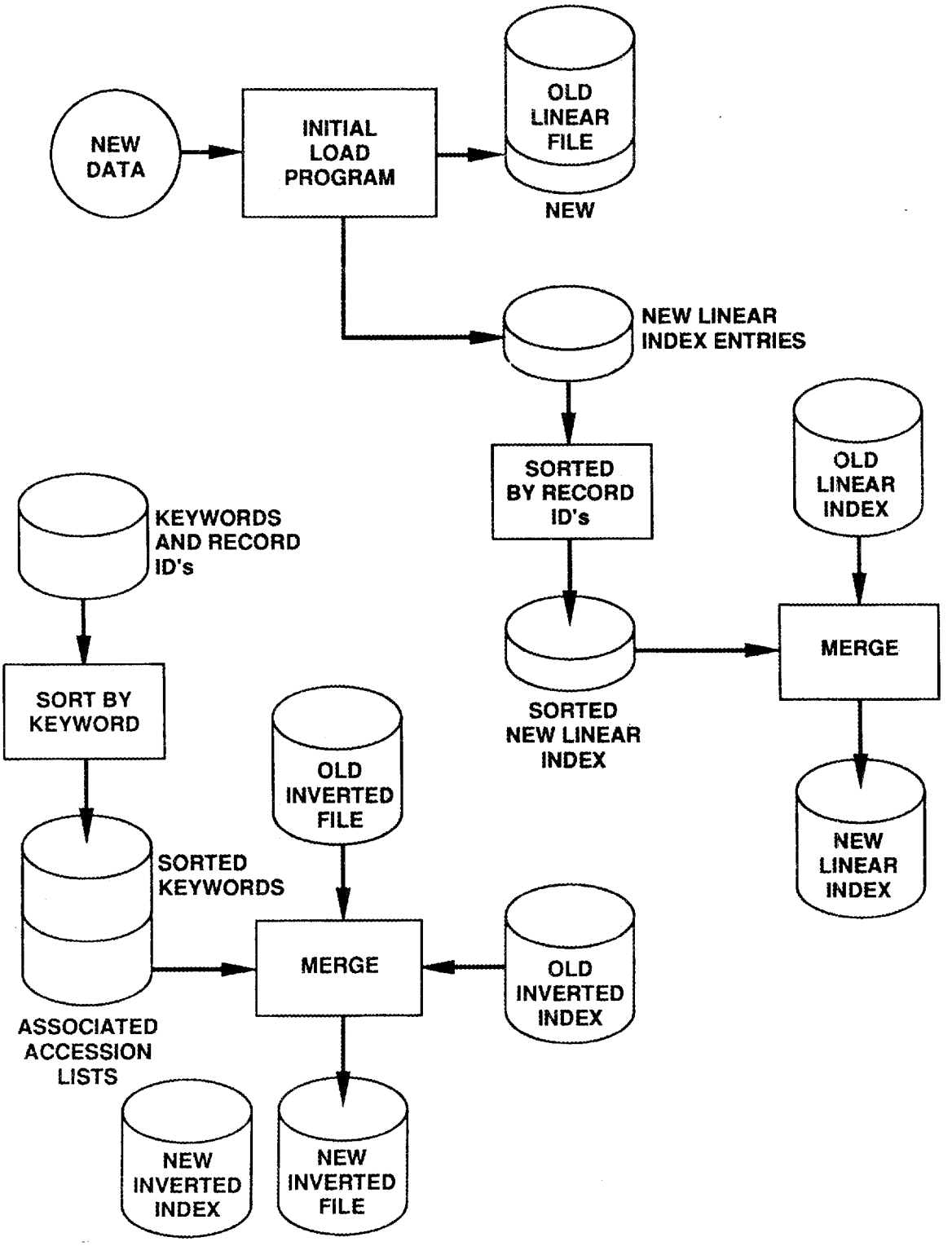

Fig. 3.10. RECON file update process.

4. SEPARATIONS SCIENCE DATA BASE

4.1 UNIQUE KEYWORDING

The design of a keyword-based retrieval system for separations science presented a challenge and an opportunity to information specialists. Traditional keyword retrieval systems simply combine all the keywords into a single file, ignoring relationships and roles that are important in separations science. For example, the most common type of question asked by a separations chemist is "How do I separate A from B out of matrix M?" With the usual data base index structure, searches intended to answer this type of question are limited to simple keywords (specific examples of A, B, and M) and will typically retrieve far too many irrelevant documents (80 to 90%) while missing essential documents.[7]

To improve retrieval in SEPSYS, separate indexes were devised (see Figs. 3.4 and 3.6) for each distinct searchable field. Internally, this is implemented by adding a prefix to each keyword to indicate its role, function, or position, such as matrix (MA), separation agent (SA), or separation system (SS). In addition, at the option of the indexer, a suffix may be added to the name of the separated substance to indicate two attributes of interest to separation chemists—namely, oxidation state and the behavior of the substance in the separation system (i.e., whether the substance is rejected, split, optionally rejected and collected, or in doubt). However, retrieval by assigned suffix is optional because, with the SELECT command in DOE/RECON, one can easily lump together all suffixes of a given keyword or select specific suffixes.

Separating the matrix, agent, and system fields into unique indexes in this manner makes it possible, by means of Boolean AND-OR-NOT combinations, to retrieve specific information on specific separations under specific conditions. Of course, the bibliographic fields can also be searched for author, title, data, and literature type, and it is also possible to search the abstracts and titles for any chosen word or words.

4.2 SPLITTING RECORDS

In addition to the problem of distinguishing separated substances from separation agents, another serious problem in computer retrieval of information is the cross coupling of unrelated terms in multisubject reports, which results in false returns that contain none of the desired information. This problem has been solved in SEPSYS 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 does not contain the bibliographic information and abstract: for those, the split simply refers back to the parent record. Cross coupling of unrelated index terms is thus effectively eliminated. An example of how such references appear in the output is shown in Fig. 4.1. For this illustration, the index terms "URANIUM" and "IRON -" were used. Most "split" records are references to progress reports that contain information on many separations.

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

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, 63 pp
    <DATE> 1956; Jun
    <ORGANIZATION> Oak Ridge National Laboratory,
    Oak Ridge, TN  37831, USA
```

This is a partial result from a search for reports on separation of copper from iron. The first display (a) appears on the screen with an artificial document number for the "split" record and a reference to the "parent" document number. The user must then key in the parent document number to see display (b). This extra step could be eliminated by using display (a) automatically as a pointer to (b) or by modifying the document number index to point directly to the parent document.

Fig. 4.1. Display of "split-subject indexing" records.

4.3 DEVELOPING A SEARCH STRATEGY

The following example demonstrates the methodology used in searching SEPSYS. In separating copper from iron, the command EXPAND COPPER would allow the user to view the portion of the SEPARATED SUBSTANCE index shown in Fig. 3.7. From this index, we would then select E06, E09, E11, E13, E14, and E15, or the terms:

IT=COPPER
IT=COPPER,-
IT=COPPER(I)
IT=COPPER(I),-
IT=COPPER(II)
IT=COPPER(II),-

The substance names, such as COPPER, are (1) followed by no punctuation, meaning that the substance is retained in the separation; (2) retained on column or filter; or (3) followed by other punctuation as follows:

Name -	(The substance is rejected.)
Name ,-	(The substance is optionally collected or rejected.)
Name +-	(The substance splits.)
Name?	(The information is in doubt.)

From the portion of the index displayed in response to the EXPAND IRON command (Fig. 4.2), we would select the entries in which trivalent iron is rejected—E08, E09, E14, and E15, or the terms

IT=IRON -
IT=IRON ,-
IT=IRON(III)-
IT=IRON(III),-

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 4.3 shows a typical bibliographic listing resulting from such a search. Note that the search could have been restricted to a single valence for copper and a single type of information. Many similar useful combinations of fields and terms will quickly become obvious to the separations chemist, once he or she has become familiar with the basic features of the DOE/RECON system and the special features added for SEPSYS.

>PROCESSING<

EXPAND IT=IRON

REF	CIT	RT	DESCRIPTOR
E01	2		IT=IRIDIUM(III)-
E02	8		IT=IRIDIUM(IV)
E03	10		IT=IRIDIUM(IV)-
E04	1		IT=IRIDIUM(VI)
E05	3		IT=IRIDIUM(VI)-
−E06	142		IT=IRON
E07	2		IT=IRON+-
E08	118		IT=IRON-
E09	17		IT=IRON,-
E10	23		IT=IRON(II)
E11	37		IT=IRON(II)-
E12	3		IT=IRON(II),-
E13	174		IT=IRON(III)
E14	45		IT=IRON(III)-
E15	23		IT=IRON(III),-
E16	1		IT=KRYPTON
E17	1		IT=KRYPTON-
E18	379		IT=LANTHANIDES
E19	47		IT=LANTHANIDES-
E20	21		IT=LANTHANIDES,-

-MORE-

ENTER:

Fig. 4.2. SEPSYS display of portion of index.

ENTER:DISPLAY 8/5/3
>PROCESSING<
DISPLAY 8/5/3
DIS 8/5/000003-000003//3
<ACCESSION NO.> 85*0002549
<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
<ABSTRACT ID> MCD
<PAPER NUMBER> 2-13-099
<ABSTRACT> The development of Imperial Chemicals' ACORGA P-5100 solvent extraction reagent provides the capability to extract copper from hydrometallurgical process streams resulting from the treatment of smelter dusts accumulated from roasters, furnaces and converters through the use of equipment required by expanding environmental control. Laboratory studies demonstrate the effective loading ability of the extractant to transfer copper from 10–15 gpl Cu liquors in the pH range of 1.5–2.0 while achieving 96–98% recoveries, high Cu/Fe selectivity ratios, and excellent kinetics and phase disengagements. Small commercial scale (0.90 metric tons/day) pilot plant studies indicate solvent stability and reliability and verify results achieved in the laboratory.
<LITERATURE TYPE> Conference
<SEP. SYSTEM> Extraction; Liquid-Liquid
<SEPARATED SUBST.> Copper; Iron –
<SEPARATION AGENT> Hydroxyoxime; Oxime
<MATRIX> Acid; Aqueous; Residuals; Sulfate
<TYPE OF INFO.> Capacity; Flowsheet; pH effect; Process; Sepn. degree; Agent stability

Fig. 4.3. Typical SEPSYS bibliographic record.

5. THE FUTURE OF SEPSYS

During January 1986, the Department of Energy announced its intention to discontinue the operation of DOE/RECON at ORNL effective January 1, 1987. Since that time, alternate search software systems have been evaluated as to their appropriateness, availability, and cost-effectiveness. Work is currently in progress to load the SEPSYS data base on the INQUIRE Data Base Management System available on the IBM 3033 system at ORNL. Anyone interested in having access to this data base should contact W. J. McDowell at ORNL for more information. Because of a shortage of funding, the data base may in the future be transferred to a standalone IBM personal-computer-compatible data base management system, as that hardware is readily available to the sponsors. Unfortunately, this approach might eliminate outside users from accessing the data base; but, unless further funding becomes available, there will be little alternative.

The SEPSYS data base may be used in another project currently being evaluated within ORNL, the purpose of which is to develop a center for separation research that would integrate the theory and application of traditional and newer separation techniques. Part of the proposal is to build an expert system that would provide access to existing expert knowledge and established data bases in the separations science area, as well as to create new expert systems and data bases where none exist. One of the data bases that will be used as a source of reference to existing research documentation is the SEPSYS data base developed at ORNL.

Separations science is becoming increasingly important and complicated as new techniques are developed and new applications appear in such diverse fields as recovery of strategic materials and detoxification of food. Improvements are being made in the understanding of separations processes, and attempts at unification of separation theories are in progress.

There will very likely be increased need for information retrieval services in this area and increased opportunity for effective collaboration among separations chemists and information specialists.

6. SUMMARY

The SEPSYS project is an interesting and instructive example of applying and tailoring a conventional broad-spectrum information retrieval system, DOE/RECON, to an important multidiscipline research area whose information retrieval needs could not be met by existing methods or services. It is also a successful example of collaboration between separations chemists and information specialists.

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 chemicals separation is largely buried in the literature of dozens of branches of science. Most methods of indexing, both hard copy and computer, do not provide a 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. Attempts to solve some of the problems with information retrieval in separations science resulted in the concept of a separations science data base. The SEPSYS data base is designed for chemical separations information and contains unique indexes that allow rapid and accurate retrieval of information about specific separations from specific matrices; it is also a method of minimizing false returns that result from cross coupling of unrelated terms in multisubject reports. Whereas conventional bibliographic science data bases return only about 10% useful references on a separation, SEPSYS has a proven record of 90 to 95% useful returns on the same kind of search.

Tests by separations chemists indicate that the special features added to DOE/RECON for SEPSYS are easy to learn and enable the searcher to find the relevant references with a high precision of recall, a capability not available for separation science in any other information service.

Other information systems for separations scientists will surely be designed and may benefit from the methods developed in this project. The successful development and proven usefulness of SEPSYS has demonstrated the effectiveness of collaboration between research scientists and information systems specialists in combining an existing information retrieval system with a specialized research-oriented indexing and retrieval strategy.

7. REFERENCES

1. H. H. Strain, T. R. Sato, and J. Engelke, "Chromatography and Analogous Differential Migration Methods," *Anal. Chem.* **26**, 90 (1954).

2. D. J. Pruett, *Separation Science and Technology—An ORNL Perspective*, ORNL/TM-10003, Oak Ridge National Laboratory, Oak Ridge, Tenn., May 1986.

3. W. J. McDowell and B. B. Corey, "A New Approach to Information Retrieval Problems in Separations Science," *J. Chem. Info. and Comp. Sci.* **24**, 108 (1984).

4. *Energinfo/on-line, The DOE/RECON System*, U.S. Department of Energy, Oak Ridge Operations, Technical Information Center, DOE/TIC 4609, Oak Ridge, Tenn., 1981.

5. *The AEC/RECON Project at ORNL*, ORNL-4913, Oak Ridge National Laboratory, Oak Ridge, Tenn., 1973.

6. J. W. Roddy, W. J. McDowell, and D. C. Michelson, *Separations Science Data Base: An Abstractors Manual*, ORNL/TM-7805, Oak Ridge National Laboratory, Oak Ridge, Tenn., 1981.

7. W. J. McDowell, "New Indexing Methods Improve the Computer Retrieval of Information on Chemical Separations," *Trends in Anal. Chem.* **4**(6), x (1985).

INTERNAL DISTRIBUTION

1. C. F. Baes
2. S. A. Bryan
3. C. F. Coleman (consultant)
4. F. C. Coleman
5. B. B. Corey
6. G. A. Dailey
7. W. Fulkerson
8. N. B. Gove
9. M. R. Ives
10. D. Y. Kelly
11. W. J. McDowell
12. F. C. Maienschein
13. B. F. Maskewitz
14. D. C. Michelson
15. B. A. Moyer
16. D. J. Pruett
17. E. E. Soler
18. C. S. Tubbs
19. S. L. Yount
20. Central Research Library
21. Document Reference Section

22–23. Laboratory Records Department

24. Laboratory Records (RC)
25. ORNL Patent Section

EXTERNAL DISTRIBUTION

26. Office of Assistant Manager for Energy Research and Development, Department of Energy, Oak Ridge Operations Office, Oak Ridge, TN 37831
27. J. C. Carbonell, Associate Professor of Computer Science, Carnegie-Mellon University, Pittsburgh, PA 15213
28. J. R. Fair, Prof., Department of Chemical Engineering, University of Texas at Austin, Austin, TX 78712
29. S. M. Gillis, Dean, Graduate School, Duke University, 4875 Duke Station, Durham NC 27706
30. F. R. Kalhammer, Vice President, Electric Power Research Institute, P.O. Box 10412, Palo Alto, CA 94303
31. R. E. Kasperson, Professor, Government and Geography, Graduate School of Geography, Clark University, Worcester, MA 01610
32. E. Kauer, E. I. duPont deNemours & Co., Savannah River Laboratory, Aiken, SC 29808-0001
33. A. S. Kertes, The Hebrew University, Jerusalem, Israel
34. M. Lessen, Consulting Engineer, 12 Country Club Drive, Rochester, NY 14618
35. H. K. Lonsdate, editor, Journal of Membrane Science, 64550 Research Road, Bend, OR 97701
36. M. V. Metanomski, Manager, Editorial Department, Chemical Abstracts Service, 2540 Olentangy River Road, P.O. Box 3012, Columbus, OH 43210

37. E. M. Salonen, Manager, Research Library, Arthur D. Little, Inc., 15 Acorn Park, Cambridge, MA 02140
38. H. J. White, Office of Standard Reference Data, United States Department of Commerce, National Bureau of Standards, Washington, DC 20234
39. M. E. Williams, Prof., Information Retrieval Research Laboratory, University of Illinois at Urbana-Champaign, 1101 West Springfield Ave., Urbana, IL 61801

40–69. Technical Information Center, P.O. Box 62, Oak Ridge, TN 37831