ENCYCLOPEDIA OF LIBRARY AND INFORMATION SCIENCE

VOLUME 50

EXECUTIVE EDITOR
Allen Kent

ENCYCLOPEDIA OF LIBRARY AND INFORMATION SCIENCE

Executive Editor

ALLEN KENT

SCHOOL OF LIBRARY AND INFORMATION SCIENCE
UNIVERSITY OF PITTSBURGH
PITTSBURGH, PENNSYLVANIA

VOLUME 50

SUPPLEMENT 13

MARCEL DEKKER, INC., New York • Basel • Hong Kong

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

LIBRARY OF CONGRESS CATALOG CARD NUMBER 68-31232

ISBN 0-8247-2277-9

Current Printing (last digit):
10 9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

CONTENTS OF VOLUME 50

CONTRIBUTORS TO VOLUME 50

ROGER ATTIAS, Ph.D., Ingénieur de Recherche, Centre National de la Recherche Scientifique, Chargé d'enseignment Université Paris 7, Paris, France: *Substructure Systems and Structural Retrieval Systems*

NORMAN DESMARAIS, Acquisition Librarian, Phillips Memorial Library, Providence College, Providence, Rhode Island; Editor-in-Chief, CD-ROM Librarian: *CD-ROM in Libraries*

ANDREW DILLON, Senior Research Associate, HUSAT Research Institute, Loughborough University, Loughborough, England: *Hypermedia*

TERENCE E. GRAY, Ph.D., University of Washington, Seattle, Washington: *Computer Viruses, Hardware/Software*

MARIA M. KING, Ph.D., Santa Monica, California: *Computer Viruses, Hardware/Software*

RICHARD KRZYS, Professor, School of Library and Information Science, University of Pittsburgh, Pittsburgh, Pennsylvania: *Encyclopedics: The Origin and Development of Encyclopedia Design; Louis Shores's Theory of Encyclopedics: A Theoretical Base for Encyclopedia Design*

SHARON CHIEN LIN, Associate Librarian, University Libraries, State University of New York at Buffalo, Buffalo, New York: *Automation of Library and Information Services in China. I. Mainland*

CLIFF MCKNIGHT, Senior Research Fellow, HUSAT Research Institute, Loughborough University, Loughborough, England: *Hypermedia*

SANDY MACMILLEN, University of Essex, Wiverhoe Park, Colchester, United Kingdom: *Vietnam: Libraries and Information Services*

O. GENE NORMAN, Head, Reference Department, Indiana State University Libraries, Terre Haute, Indiana: *Information Marketing in Libraries*

ELIZABETH L. PATTERSON, Head of Reference, Robert W. Woodruff Library, Emory University, Atlanta, Georgia: *The Bibliographic Control of Microforms*

STEPHEN C. PERRY, Manager, Management Information Systems, Cotton, Inc., Raleigh, North Carolina: *File Access Techniques*

JOHN RICHARDSON, Senior Research Associate, HUSAT Research Institute, Loughborough University, Loughborough, England: *Hypermedia*

JAMES L. SWAUGER, Curator Emeritus, Anthropology, Carnegie Museum of Natural History, Pittsburgh, Pennsylvania: *Museums, Museology, and Museography*

RENEE TJOUMAS, Associate Professor, Graduate School of Library and Information Studies, Queens College of the City University of New York, Flushing, New York: *Evaluation of Library Science Collections in the Accreditation Process*

CHIH WANG, Learning Resources, University of Guam, Mangilao, Guam: *Micronesian Area Research Center (MARC): Collections, Research, and Publications*

LEI ZENG, Ph.D., School of Library and Information Science, University of Pittsburgh, Pittsburgh, Pennsylvania: *Achieving Compatibility of Indexing Languages in Online Access Environment*

ACHIEVING COMPATIBILITY OF INDEXING LANGUAGES IN ONLINE ACCESS ENVIRONMENT

Online Subject Searching in a Multithesauri Environment

BACKGROUND

The application of computer and telecommunication technologies to information storage and retrieval systems has made large-scale integrated retrieval systems possible. Today, the majority of the world's recently generated research literature is accessible in the form of bibliographic references through computerized online search services. Vendors have shown remarkable capabilities in loading machine-readable databases of all descriptions and formats into standard computer storage and retrieval systems, and in updating and maintaining these databases. Yet some problems still are beyond the capabilities of these vendors. One major problem remaining is developing mechanisms necessary for the integration of subject retrieval in multidatabases which have been indexed on the basis of various predetermined thesauri. From an online subscriber's viewpoint, as Niehoff (*1,* pp. 2-3) pointed out, trying to match a client's information need with a thousand databases could be a very real and complex problem. It is not simply a matter of deciding which database or bases to search, or which vendor to use, but how to effectively search the resultant combination once it has been identified.

The controlled vocabularies used for subject indexing are as many and as varied as the materials they index and the audiences they serve. Most thesauri have been developed independently for specialized catalog, index, or information services since the early 1960s, when differences among their vocabularies did not appear to pose a problem for their users. For today's world of online access, however, the multiplicity of thesauri creates complications significantly affecting at least two types of information services. In the first instance, it affects online bibliography services such as DIALOG which offers over 350 diverse databases through a single interface. Second, it affects integrated library systems such as NOTIS which provides subject authority control for more than one logically separated subject vocabulary. In an automated environment, the absence of compatibility can add substantially to the costs of time, intelligence, and money, while it can bring unsatisfied search results to end users.

Unesco published its *UNISIST Study Report on the Feasibility of a World Science Information System* in 1971 (*2*). The major recommendation 4 pointed out that "The attention of scientists, learned societies and information science associations should be drawn to the need for joint efforts in developing better tools for the control and conversion of natural and indexing languages in science and technology." Several international conferences have since focused discussion on compatibility concerns (*3, 4*). The Third International Study Conference on Classification Research (held in 1975) had the theme "Ordering Systems for Global Information Networks." Themes of other conferences included "Overcoming the Language Barrier" (1977 at Luxembourg); "Unified System of Information Retrieval Languages" (1977 at Latvia); and "Compatibility between Indexing and Retrieval Languages" (1982 at Ohio).

Today, twenty years after the publication of the UNISIST report, three working hypotheses have directed the inquiry of this review of the issues relating to achieving compatibility of indexing languages in today's online access environment. First, with the increase of cross-database searching and intersystem cooperation, and with the better understanding of subject analysis and accesses, compatibility and integration of indexing languages are becoming issues of great practical impact and are experiencing renewed interest. Second, projects for the standardization of thesaurus construction led by international and national agencies* have played an important role in promoting compatibility of indexing languages. However, compatibility no longer always depends on standardized thesauri, because many more progressive organizations had developed their own standards before international and national standards were developed. Studies on achieving compatibility based on existing indexing languages have contributed significantly to positive trends that affect our online services today. Third, the development of approaches to securing compatibility of indexing languages is impacted by the following outside forces which are also influencing the whole field of library and information science: the widespread and ever-increasing use of automated systems for providing subject access, the increase of searching requests for integrated, rather than fragmented research, and the involvement of patron searching.

DEFINITIONS AND EXPLANATIONS

The subject indexing process involves two quite distinct intellectual steps: the "conceptual analysis" of a document and the "translation" of the conceptual analysis into a particular vocabulary or "indexing language." In the majority of systems this involves the use of a "controlled vocabulary," a limited set of terms that must be used to represent the subject matter of documents. Such a vocabulary might be a list of subject headings, a classification scheme, a thesaurus, or simply a list of "approved" key words or phrases (*5*).

In the 1971 Unesco UNISIST Study Report, *compatibility* was defined as "a quality of systems whose products can be used interchangeably, notwithstanding differences in notation, structure, physical carriers, etc., without any special 'conversion machinery'." Two related terms were defined as well. *Convertibility* was defined as "a quality

*This includes the publication of ISO 2788, ISO 5964, etc. For detail, see ref. *8*, pp. 29-33.

of systems whose products can be made interchangeable through 'conversion' programmes." *Conversion* was defined as "the process of transforming information records, with regard to transcription encoding, data structure, etc., so as to make them interchangeable between two or more services or systems using different conventions and media" (*2*, p. 147) The implication of these terms for indexing languages is demonstrated by Dahlberg's (*6*) definition that compatibility of an ordering system is the quality which permits the elements of one such a system to be used together interchangeably with the elements of another ordering system.

From a formal point of view, a concept, as an element of an indexing language, consists of three parts: the *referent*, the *characteristics* predicated of a referent, and the *verbal or coded expression* denoting the referent and comprising the predicated characteristics in its designation in form of a term, a descriptor, or a notation (Fig. 1).

From this perspective, the *conceptual compatibility* between elements of controlled vocabularies can be compared in relation to three measures of strength of correlation. Listed from the strongest type of correlation to the weakest these include: (a) *conceptual coincidence*—two concepts match in all their characteristics; they can also be called "equivalent"; (b) *conceptual correspondence*—two concepts match in many of their characteristics; also called "similar concepts"; (c) *conceptual correlation*—two concepts are set into "correlation" where the kind of correlation can be indicated, e.g., by mathematical symbols (*6*).

OVERVIEW: APPROACHES TO SECURING COMPATIBILITY OF INDEXING LANGUAGES

Many possible approaches to securing compatibility of indexing languages have been studied (please refer to Appendix 1. Chronology of Research Projects). In a report prepared for Library of Congress Processing Services on multiple thesauri in online library bibliographic systems, Mandel (*7*) categorized four basic approaches to providing access to databases indexed by different vocabularies: (a) *segregated files*—databases using different thesauri are searched separately; (b) *mixed vocabularies*—terms from all vocabularies are retrieved together in subject searches; (c) *integrated vocabularies*—techniques used to relate different thesauri can be used to develop syndetic structures that would aid retrieval in a multiple database environment; and (d) *front-end navigation*—features being developed in "intelligent" interfaces to online retrieval can be designed specifically to aid in searching from multiple data-

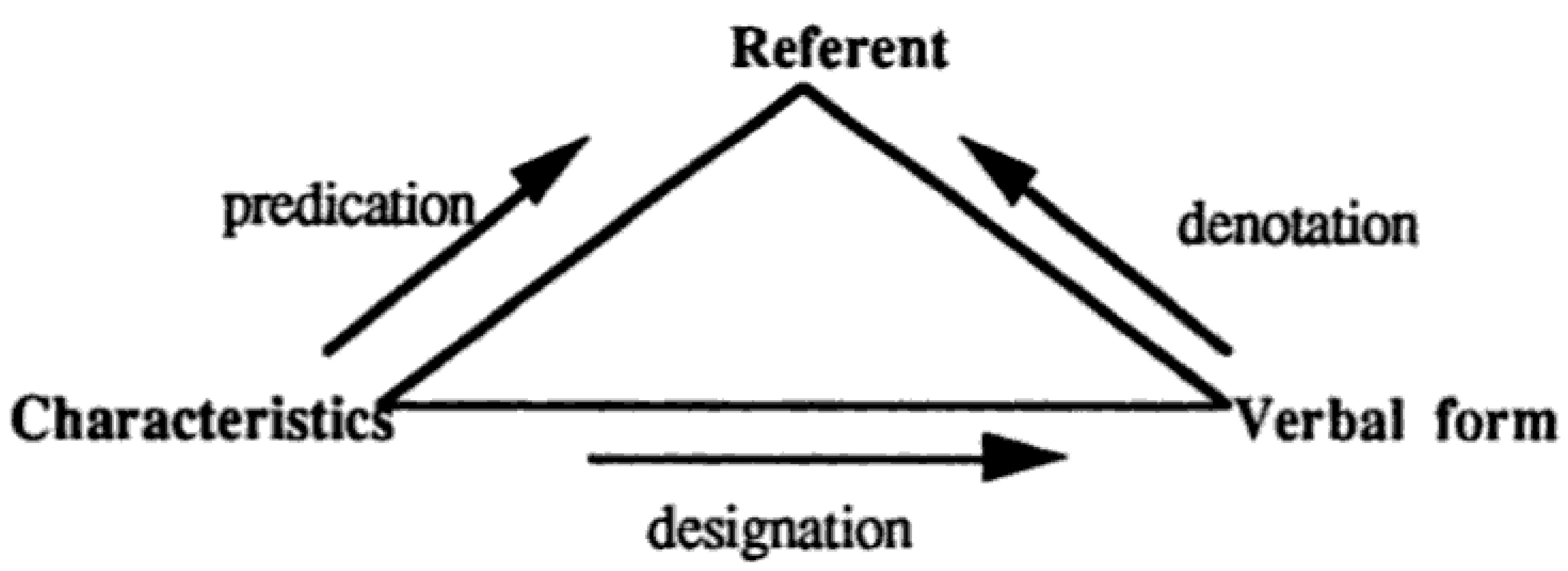

FIGURE 1.

bases. In an earlier report for UNESCO's General Information Program, Lancaster and Smith (*5*) grouped approaches to securing vocabulary compatibility into five categories: mapping, intermediate lexicon (i.e., switching language), integrated vocabulary, microthesauri, and macrovocabularies.

This article would suggest, however, that recently it has become possible to move beyond the complications inherent in these categories by drawing a distinction in terms of focus between approaches to mechanisms that enable the establishment of compatible vocabularies (to be discussed in the next section) and approaches to automated systems that help to relate separately created thesauri to one another (to be discussed later).

Developing Mechanisms for the Establishment of Compatible Vocabularies

When integrated retrieval systems became possible in the 1970s, there were few technical problems to merging databases which had been indexed with different thesauri. However, it soon was found that these files could not be searched by an identical strategy because a single concept might be quite differently represented in the different vocabularies. The conventional way of solving this problem is to develop mechanisms which would enable the establishment of compatible vocabularies. Intermediate lexicon, or switching language, microthesauri, and macrovocabularies have been significant contributions. Other examples also exist but these at least illustrate the situation.

INTERMEDIATE LEXICON/SWITCHING LANGUAGE

Before the idea of an intermediate lexicon, some studies were concerned with automated mapping between vocabularies (to be discussed). Mapping two vocabularies on to each other requires only one mapping operation (as shown in Situation I in Fig. 2). When more and more vocabularies are to be mapped, the situation becomes increasingly complex. For example, a four-vocabulary-mapping requires twelve separate mapping operations (see Situation II in Fig. 2). The alternative is to construct an intermediate lexicon, or, a "switching" language that can be used to convert from any one vocabulary to another (see Situation III in Fig. 2, where A-F are vocabularies to be converted, and X is an intermediate lexicon).

An intermediate lexicon is intended to transfer the contents of documents expressed in terms of any index language to another without loss of information. It entails the mapping of two or more vocabularies to an intermediate or neutral language, such as a classification or a coding scheme. For example, a user can enter a term from A's vocabulary, say, TUMORS, the term will be converted to the intermediate lexicon (X) code, say 17904, and switched from this to NEOPLASMS, the equivalent term in D's vocabulary: TUMORS ← 17904 ← NEOPLASMS (*8*, pp. 189-190). This idea was developed by J. C. Gardin and his group from 1967 to 1968 in France. A matrix had been worked out previously and the elements of six thesauri available to the field of information science were correlated by the matrix (*9*). Continuing research on this project were discussed by Coates (*10*) and Horsnell and Merrett (*11*). In a related experiment, Antopolskii et al. (*12*) defined the concept of

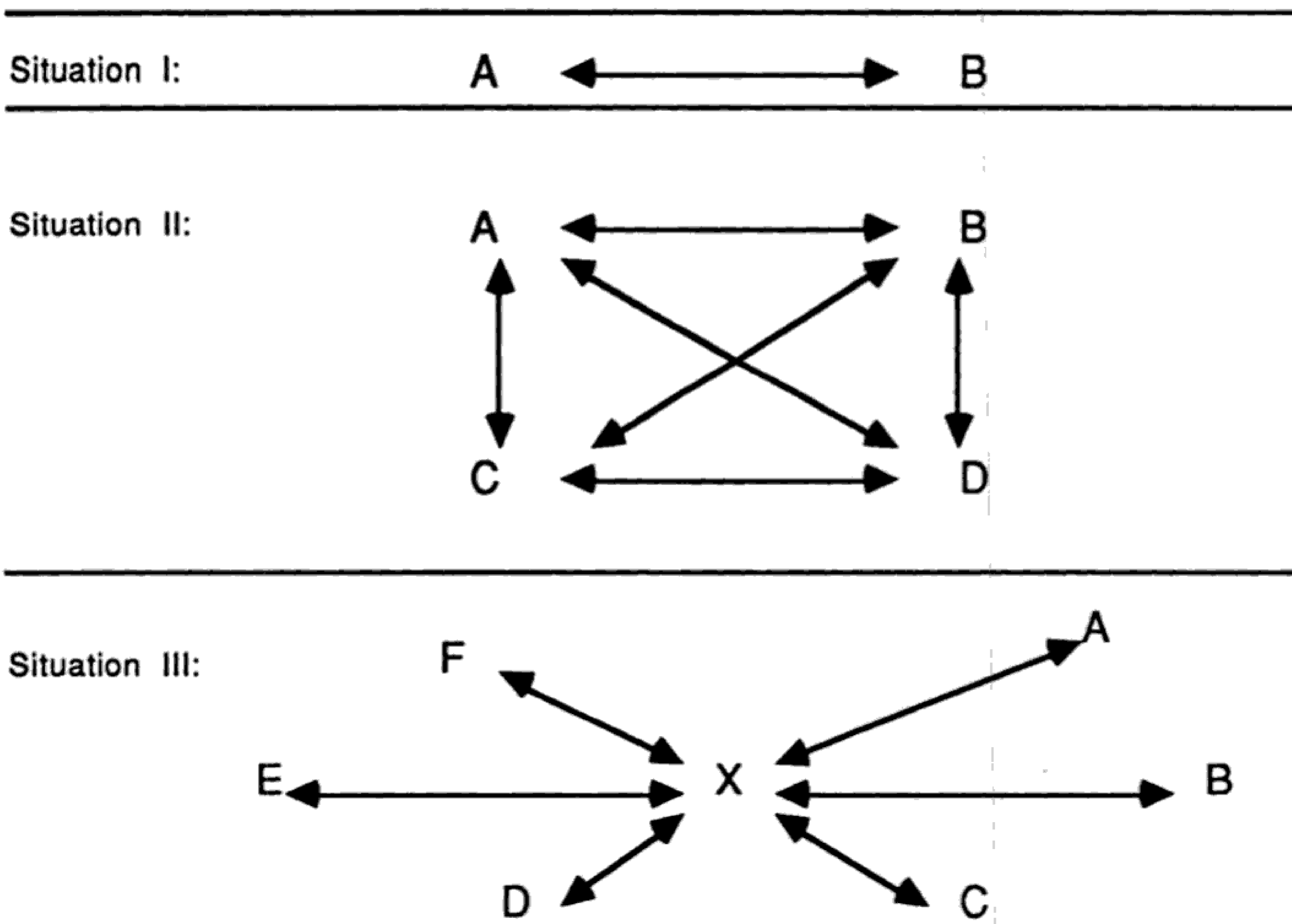

FIGURE 2. *Possible mapping Operation (adapted from Ref. 8, pp. 182, 183, and 189).*

"lexical intersections" (a set of lexical units from different thesauri found by means of formal and/or semantic identification) for use in the development of a switching mechanism. They also examined the efficiency of various correspondence criteria for retrieving lexical intersections. Other work includes Neville's idea of a "supra-thesaurus" (*13*), Beling and Wersig's "intermediary language" for the West Germany information systems (*14*), and several additional switching languages used in multi-lingual vocabularies.

The intermediate lexicon is still a theoretical approach since no such tool has been fully implemented. The reason might be that the development of such a switching tool (usually a coding system) requires a lot of intelligent work, while its effectiveness is limited in the environment that all participating vocabularies should have similar or closely related subject coverage. It is worth noting that, instead of working on an intermediate lexicon, the British Library investigated the switching of PREserved Context Index System (PRECIS) input strings from one language to another and used PRECIS in the British National Bibliography for many years. The PRECIS system expresses concepts in terms selected from natural languages encountered in the literature. A major feature is that each term is placed in a context-dependent order in such a way that each term sets the next term into its obvious context. Syndetic control is maintained through use of a thesaurus (*15*). PRECIS, as an indexing system, was suggested as a switching mechanism because it could be adapted for various European languages (*16*).

INTEGRATED VOCABULARIES AND METATHESAURI

In an online search environment, an alternative is to provide user with a composite or integrated vocabulary. This approach entails the construction of a master list of all terms and references in the target vocabularies. The terms are drawn from all of the databases processed by a particular information facility. The integrated vocabulary is different from the intermediate lexicon approach because no neutral switching language is created. Recently a term "metathesauri" is widely used. Rada (*17*) briefly stated that "a metathesaurus transcends a set of thesauri." Roulin (*18*) defined "a metathesaurus is a set of reference thesaurus and various indexing languages that are attached to it either via common concepts hierarchically structured in the same way (in the case of 'sub-thesauri' and 'associated thesauri') or only via relationships established between the indexing terms that remain proper to each indexing language (in the case of 'national vocabularies')."

A *Unified Medical Language System (UMLS)* supported by the National Library of Medicine (NLM) was reported in 1987 (*19*). Experimental linkings of *Medical Subject Headings* (MeSH) with the SNOMED (*Systematized Nomenclature of Medicine*), CMIT (*Current Medical Information and Terminology*), and PDQ (the National Cancer Institute's retrieval system) thesauri were conducted. A metathesaurus has been used to store medical concepts and terms in a canonical form to which multiple existing vocabularies and classification systems would be mapped. Each Meta-1 record contains three types of information: basic facts, relationships, and usage data (*17, 20*). However, the goal of UMLS goes beyond a single medical vocabulary, it attempts "to make the myriad of classifications of medical knowledge invisible to the user while providing a single logical path to a broad range of biomedical information sources" (*21*, pp. 4-6). In Europe, Directorate XIII-B3 of the European Economic Community has outlined a five-step procedure to create a metathesaurus which the commission calls the "Multifunction, Multilingual Thesaurus Database." A list of about 1,000 thesauri in prominent use in the world has already been compiled. Connecting the thesauri that have been cataloged by DG XIII-B3 and allowing easy access by users could be an enormous benefit to those who want to share terminology (*17*). In China, a *Chinese Classified-Thesaurus* is nearing completion. This project, sponsored by the Peking Library, the national library of China, aims at an integrated vocabulary based on the national standard classification *Chinese Library Classification* (CLC) and the recommended national standard thesaurus *Chinese Thesaurus* (CT). The new vocabulary provides for each of the classes or divisions from CLC corresponding descriptor(s) from CT and vice versa (*22*).

The creation of an integrated vocabulary still needs a lot of intelligent work. A similar, but improved approach is to use automated systems for subject authority control and for mixed vocabulary searching. This will be discussed in the third section of this article.

MICROTHESAURI

This approach treats specialized thesauri as the satellites of a superstructure. In its original application, a microthesaurus was a specialized subset of terms extracted from, and therefore compatible with, a larger thesaurus. A microthesaurus can therefore be defined as a specialized vocabulary that maps onto a broader thesaurus and is entirely included within the hierarchical structure of that thesaurus (*8,* pp. 198-199).

In the early 1970s, Soergel (*23, 24*) developed a concept of a universal source thesaurus, from which other thesauri could be derived. In 1981, The British Standards Institution (BSI) produced its *ROOT Thesaurus*. It was developed first as a comprehensive indexing and searching tool for technological applications and second as a labor-saving device in the construction of further thesauri. It conceivably could serve as the superstructure of a national information system and as the basis from which a whole series of thesauri and related products could be obtained (*25*). The entire system may comprise three, four, or five levels and thesauri could exist at each level (*8,* p. 201). Another example is the *Thesaurus of Common Topics* (TCT), developed at the Institute of Scientific, Technical and Economic Information in Warsaw. It was intended to provide particular branch thesauri or branch system of thesauri with some ready sets of terms, and thus eliminate unnecessary duplication of effort (*26*).

In some respects, large comprehensive vocabularies developed by national libraries and information centers are likely to be treated as superstructures. The Chinese Documentation Standardization Committee suggested in 1979 that all special thesauri developed after 1980 should consider establishing compatibility with the *Chinese Thesaurus,* a universal vocabulary developed by the China Science and Technology Information Institute and the Peking Library. Most of the Chinese special thesauri developed since then have used the *Chinese Thesauri* as their basic term source and structure sample (*27*). In America, LCSH has been acting as superstructure as well. The *Legislative Indexing Vocabulary* (LIV), LC *Thesaurus of Graphic Materials, Subject Headings for Children's Literature,* and the newly developed *Art and Architecture Thesaurus* (AAT) are examples of microthesauri which were developed as satellites of LCSH in order to meet the specialized needs of particular subject fields, materials, and patrons. LCSH compatibility codes have been used in these microthesauri (*7,* pp. 4; 80–92). For instance, each LIV term is assigned an LCSH compatibility code as follows:

LC—term is the same in LCSH
LCX—term is a "see" reference in LCSH
LCC—term is similar to one in LCSH
LCD—characters match an LCSH term, but the meaning is different
LCO—no match in LCSH

Unfortunately, most existing indexing languages are developed completely independently. Although these thesauri may conform to common standards or guidelines in their structure, they are not been integrated or interconnected with a common superstructure. At the same time, few thesauri could be expected to perform such complex functions as the LCSH does.

MACROVOCABULARY

The macrovocabulary approach, as described by Lancaster, is conceptually similar to that of the microthesaurus, but the method of implementation is virtually reversed. The idea is simply to create a kind of generic superstructure of terms that will subsume a group of thesauri, or other types of vocabularies, in diverse subject fields so as to link terms in specialized vocabularies (*8*, pp. 202–203).

Historically, there are a number of examples of such efforts. Started in the late 1950s, the British Classification Research Group (CRG) pursued the idea of a sort of "Ur-Classification", a generalized scheme of terms not necessarily related to any one practical application in a library or documentation center. The origins of all these ideas were first announced to the world at the First International Study Conference on Classification for Information Retrieval in 1957. CRG's experience from the 1990 revision of Class J Education of the *Bliss Bibliographic Classification (BC)*) suggested that a good general scheme could be compiled by integrating special schemes. A general, or "Ur-Classification" will provide a reservoir of terms for special schemes, while the special schemes provide detailed analysis and enumeration by experts in each field (*28*).

In 1978, the UNISIST program developed an element of *The Broad System of Ordering* (BSO), essentially intended as a switching language, but only as a general superstructure rather than the more specific switching approach embodied in the intermediate lexicon (see also Appendix 5). It existed in the form of a classification scheme, with notations, containing 4,000 terms (*8*, pp 203–205; 29). Sager et al. (*30*) and Whitelock (*31*) described developments in another macrovocabulary project—a Unesco-funded project to create an *Integrated Multilingual Thesaurus* for social sciences. The Descriptor Bank, which lay at the heart of this project, converted a number of thesauri to a common form, included terminological and syndetic information, and resolved differences in preferred term choice, definition, hierarchy, precoordination and compounding, and language. In the preliminary study for this macrothesaurus, Meyriat (*32*) analyzed 60 existing information languages in order to identify for each the scope and depth of coverage of 41 subject fields.

A feasibility study for developing a *Universal Agricultural Thesaurus* unifying CABI, NAL and FAO thesauri was initiated in 1989. C.A.B. International (CABI), National Agricultural Library (NAL) and Food and Agriculture Organization of the United Unions (FAO) have been involved in this effort which will improve access to agricultural information in a more cost-effective way. It was stated that "without such a tool, individual institutions would continue the costly and duplicative activities associated with local subject thesaurus development and validation of taxonomic names, and researchers would continue to be uncertain as to the completeness and accuracy of their research documentation" (*33*).

Additionally, a number of projects for macrovocabularies were conducted in the United States, Soviet Union, and France. Their purposes included using the superstructure to suggest useful search terms to retrieve information on particular subject from a variety of databases, and linking various specialized information centers in order to achieve some measure of compatibility among the vocabularies of these centers (please refer to Appendix 1).

Developing Automated Systems for the Creation of Linkages Among Vocabularies

Many automated systems have been designed to aid the process of integrated subject retrieval. Most of these systems reflect an effort to map separately maintained vocabularies. These approaches tend to emphasize the phase of subject searching instead of the phase of thesaurus construction, and tend to be more dependent upon the state of the art of computer support.

AUTOMATIC MAPPING

Mapping entails the direct translation of terms in one vocabulary into corresponding terms in another. Machine support is particularly valuable for mapping terms and merging term lists, with algorithms developed for matching references, component words, stems, Boolean combinations, etc. Concepts can be mapped between two thesauri with a variety of tools. First, direct lexical matching between concept main terms can be performed; second, knowledge about the syntax or the morphosemantics of main terms can be employed; third, the knowledge in the relationships within the thesauri themselves can be the basis for sophisticated mapping of terms from one thesaurus to another (*20*). When vocabularies exist in machine-readable form, part of the conversion such as direct lexical matching can be done with common computer software tools. Automatic identification includes exact match, some variant spellings, some variations in word forms and inversion, and also, under certain conditions, it involves mapping on the basis of the cross-references and hierarchical structure of the vocabularies (see Appendix 6).

It has been convincingly demonstrated that the more alike the two vocabularies are in structure, and the higher level of subject overlap, the more conversion can be performed automatically. Wall and Barnes (*34*) found that they could automatically map 76% of a sample of *Medical Subject Headings* (MeSH) to the *Agricultural/Biological Vocabulary* because both of the vocabularies were precoordinated subject heading lists and had high degree of subject overlap. Another study was an experiment of convertibility between *ASTIA Subject Headings* and *AEC Subject Headings* by Hammond and Rosenborg (*35*). Although the two vocabularies differed with regard to their term structure principles, they were highly compatible because of their high degree of agreement in subject coverage. In a project of the European Economic Community to connect patient records with the medical literature, *International Classification of Diseases* (ICD) of the World Health Organization, and the *Systematized Nomenclature of Medicine* (SNOMED), a thesaurus for patient records, were mapped into EMTREE, the thesaurus used for indexing *Excerpta Medica.* More than 50% of SNOMED terms were directly mapped to equivalent EMTREE terms with simple matching rules. Mapping between ICD and EMTREE was performed through facets (*17*).

A related possibility is that some terms which are identified by a machine as the "equivalent" might have different meaning when they are used in the different subject fields, such as *"pressure"* in psychology and in mechanical engineering. As a result, mapping between two vocabularies which are composed of different subject topics may lead to a more or less confusing situation. Therefore, automatic mapping is encouraged for use only when two vocabularies have a similar subject coverage.

SEGREGATED FILE SEARCHING AIDS

Online vendors have made various attempts to provide users with term search or database selection aids. In a system maintaining logically separate subject indexes and subject authority files for each vocabulary, instructions usually are provided so that patrons can use different commands when searching files which were indexed by different vocabularies. As early as in the 1970s, the University of Illinois began to design a more user-oriented "transparent system," i.e., a system "containing the necessary converters or translators to help the user circumvent the need for understanding all the specific differences of databases, systems, command languages, vocabularies and access protocols" (*36,* p. 361). In such a system, the user is not distracted by many of the differences that exist among systems and within host files. Williams discussed later the wider issues involved in the design of transparent systems and described a number of approaches, which she categorized as gateways, front ends, intermediaries, and interfaces (*37,* p. 51).

A well-known experiment is the Vocabulary Switching System (VSS), an experimental automated subject switching mechanism for searching multiple databases in a single natural language, developed by Niehoff et al. at the Battelle Columbus Laboratories. It is a stand-alone, online database containing the subject descriptors and all the syndetic relationships found in the vocabularies included (*38*). It was originally generated from existing energy-related vocabularies (*39,* p. 135). All vocabularies included in VSS were reformatted into a common input format (see also Appendix 3). The unique records from each vocabulary ended up in one of five files: Concept file, Term file, Phrase file, Word file, and Stem file. As of 1985, VSS had incorporated 15 different database vocabularies in four modules (organized by discipline) and could offer various cross-database switching options and browsing capabilities. In later studies the scope of the object was expanded to service other subject areas and offered twenty-one options for "vocabulary switching." The system could compare a term entered by the user with terms in the vocabulary of a target database, and indicate the degree of match (*40*). Analysis of the results illuminated two points. First, the approach used was as relevant in a multilingual setting as it was in a monolingual setting. Second, different subject areas had varying degrees of vocabulary compatibility. For example, social science vocabularies were more compatible than those in other areas (*1, 38, 41*).

Subject authority control is vigorously developed in the integrated library systems such as ORION, WLN, NLC/DOBIS, UTLAS, Geac BPS, NOTIS, and Carlyle (*7*). These systems currently can provide authority control for more than one logically separate subject vocabulary. Generally, they have two major functions. One function is to identify a subject source list, while the other is to validate subject headings against selected vocabularies. Validation ranges from a match of headings in order to capture cross-references (e.g., Carlyle) to complex linking mechanisms that replace headings in bibliographic records with authority record I.D. numbers (e.g., WLN, UTLAS). These systems are interesting not only for their support of multiple thesauri, but for their power to maintain control over headings and subdivisions and for their support of linking relationships among terms. For example, NLC/DOBIS and UTLAS's CATSS create a special translation to link LCSH terms, the *Canadian List of English Subject Headings,* and their French counterparts in *Repertoire de vedettes-matiere* (*3, 7;* pp. iv; 11-68).

MIXED VOCABULARY SEARCHING

Mixed vocabulary searching is a commonly used approach. Terms from all vocabularies are retrieved together in subject searches. It has already been implemented by the major online service centers in the United States. An example is one of the new search features of DIALOG, OneSearch. When used with DIALINDEX, it facilitates searching in up to 20 files concurrently with one search strategy (*42*). DIALINDEX and CROSS (developed by BRS) are files of all the searchable terms appearing in the many databases of bibliographical records accessible through these services. In response to a term entered by a user online, such a tool will show in which databases the term occurs and how frequently it occurs in each (*8,* p. 194). In BRS' TERM database, terms of different thesauri are organized into concept records, which also include hierarchical information and free-text searching suggestions (*43*) (see also Appendix 6). A similar approach has been used by online integrated library systems. For example, in ORION and DOBIS, the subject authority file index is not partitioned according to subject source. Terms from different thesauri are retrieved in the same alphabetical list, while the hits, terms and subject sources are indicated on the screen (*7,* pp. 11-15; 31-42) (see also Appendix 7).

It seems that this approach is acceptable because it is easy to implement and needs little artificial work on the existing vocabularies. But one problem this approach risks is obvious vocabulary clashes (e.g., the same term is postable in one vocabulary and nonpostable in another). The seriousness of the problem depends upon the subject searching design features and the particular mix of collections and vocabularies included (*7,* pp. v; 71-72).

Factors That Influence Compatibility Among Controlled Vocabularies

A GENERAL ANALYSIS

Many factors contribute to the compatibility of thesauri. The extent of overlap in the subject matter, specificity, and vocabulary size are commonly believed to play important roles (*8, 9*). The degree of precoordination in the terms (*44*) and the extent to which the vocabularies are 'constructed' (*8,* p. 184) have also been shown to influence compatibility. Emphasis on the structure, the quality of a hierarchical display is also involved (*45*). Dahlberg (*9*) stated that the complexity of any thesaurus, which is determined by the number of precombinations of the concepts involved, the manner in which concepts are described, and the structural components of the thesaurus are the factors that cause limitations in compatible comparison. To summarize the factors that influence the compatibility of thesauri, we may deal with the factors from the following aspects (*28*).*

Principle Aspects

Principles, which have the most direct influence on compatibility, are decided at the first step of thesaurus design. They may include at least three parts, that is:

1. *Initial term-structure principle.* A term-structure based on either precoordination or postcoordination was chosen at the beginning. Meanwhile, the percentage of precombined terms in a postcoordinated vocabulary was also decided.
2. *Original orientation of vocabulary design.* Vocabularies might be designed for certain collections, or for particular projects or roles. The discipline-oriented vocabularies are likely to have great differences from collection- or role-oriented ones in their vocabulary size, specificity, and subject coverage.
3. *Initial purpose of vocabulary usage.* Thesauri are mainly developed for mechanized retrieval, while a few might take manual retrieval into account. Thus leads to differences in the degree of precombined terms.

Vocabulary Aspects

Vocabulary aspects are also very influential on the conversion among thesauri, but they are flexible. Vocabulary size, specificity, entry format control, number of entry points, and precision of expression are variable among thesauri and therefore make the results of conversion between any two of the thesauri variable.

Structure Aspects

Structure aspects have an indirect influence on conversion. Most thesauri use extra structure to display terms and term relationships in addition to an alphabetical list, such as hierarchical display, subject category display, graphical display, etc. It is believed that the more a thesaurus displays its vocabulary, the more helpful to conversion.

Other aspects which influence compatibility more or less come from characteristics of language. For instance, multilingual conversion brings many special problems. Other influences may come from the frequency of term occurrence in the indexing and searching process; the policy of index term assignment; and the differences of the presentation of the same concept in a thesaurus and in the indexing and searching processes. These aspects are very changeable and are influenced by some subjective factors.

GENERAL VOCABULARY VERSUS SPECIAL VOCABULARY

It has been recognized that the compatibility mechanisms are easier to develop when the indexing languages have a common knowledge structure and framework for

*The remainder of this article has the same reference if no specific indication is given.

representation of subjects (*39*). The most difficult work exists in the compatible approach between general and special thesauri. Some of the reasons are:

Identification of Equivalent Concepts

Usually, a general thesaurus could neither reach as high a specialization as a special one nor cover as much special vocabulary as in a special one. As a result, it will be very difficult or sometimes impossible to access equivalent concept transferring for some special vocabularies.

Concept Environment Design

Many scientific terms may represent different concepts or meanings when they are used in different subject fields. Therefore, it is a more complicated job to set up concept environments for each term in a compatible tool. Although BT (Broader Term) and NT (Narrower Term) may be decided according to the 'family' relationship in a hierarchy, RT (Related Term) cross references are hard to decide.

Coverage and Specificity Balancing Among Different Subject Fields

When establishing a macrovocabulary or integrating several existing thesauri (we will call them 'source thesauri') online, a balancing between coverage and specificity is hard to attain. Although we can design the coverage, average specificity, and vocabulary size for each subject field in a superstructure, we can not design or change those for the source thesauri. It is possible that the existing thesauri will not adapt to our design framework exactly. In this case, the choice of coverage and specificity for each subject fields, and the balance among different subject fields will be very troublesome.

MULTILINGUAL ISSUES

More and more bilingual and multilingual controlled vocabularies have been developed or proposed for development. UDC has been available in several languages for decades. MeSH now exists in several languages, allowing MEDLINE to be interrogated in languages other than English (*8*). However, when dealing with translingual processes, extra problems appear in all of the approaches discussed above.

Multilingual Aspects

Glaushkov et al. divide forms of compatibility into two: semantic compatibility and structural compatibility. Semantic compatibility can be reduced to lexical, paradigmatic, and syntagmatic compatibility; that is, compatibility in the representation of objects, in hierarchical relations recognized, and in nonhierarchical relations recognized, respectively. Structural compatibility can be reduced to morphological compatibility (similarity in the structure of terms) and syntactic compatibility (similarity with respect to the structure of groups of terms) (*8*, p. 186). These compatibilities vary between any two languages. It seems that the thesauri conversion between some languages, such as English with Chinese, has to be based on a full translation of meanings rather than on morphological and syntactic automatic mapping because of the extreme differences in both semantic and structural compatibilities.

Cultural Background Aspects

No matter whether a thesaurus is designed with an orientation toward a collection or discipline, it reflects its social environment. Each word has its semantic meaning in isolation, in textual content as well as in subject environment. The problems in conversion may come not only from different subject coverage, but also from different political and ethical standards of terms. In most fields of science and technology, there are fewer influences from the cultural background. However, in some fields, there are still problems of subject coverage. For example, traditional Chinese medical science and medicine are very important in China, but terms in these fields are excluded by most English medicine and related thesauri.

Cultural background aspects cannot be avoided and will influence the compatible approach. We might find rules to solve problems brought about by multilingual aspects, but are unlikely to find rules to solve problems brought about by cultural background, especially when machine automated mapping or conversion are considered.

Conclusion

The importance of compatibility of indexing languages in information systems is being given increasing emphasis in today's online multiple database access environment, in light of the many examples of experimental extent and practical use. In order to solve problems which are beyond the scope of gradually developed international and national standards, many approaches to achieving compatibility among indexing languages have been investigated. Research projects that set new trends or reflect new approaches are reviewed in this article.

In general, efforts to develop mechanisms that enable the establishment of compatible vocabularies can lead to more compatible information system performance, because attention is being given to the earlier stages of the information transfer cycle. In the long run, the earlier that standardization and compatibility take place, the more cost-effective it is likely to be. But some problems remain. Intermediate lexicon and integrated vocabulary methods require creating additional tools, while microthesaurus and macrovocabulary approaches may introduce some inflexibility and inability to tailor a system to local needs. Therefore, much intelligent work must be involved in the establishment of any of these compatible vocabularies. Sometimes, a measurement of the cost of, and effectiveness achieved from, the process of establishing a compatible thesaurus will call into question the value of such an effort and the extent to which it may leave the original objectives of earlier approaches to the problem of compatibility.

New developments in technology have made automated systems an alternative means to achieve compatibility. These have been especially useful for relating separately created thesauri to one another in an online multidatabase access environment. Where no standards or compatible vocabularies exist, or where compatible vocabularies are felt to be too difficult to establish or too restrictive to satisfy local needs, automated systems can suggest themselves as, with some qualifications, an efficient means to support unique authority files. Undoubtedly, automated search aids that link vocabularies have emerged as feasible alternatives to dealing with the incompatibility of indexing languages. However, it is just at its embryonic stage. Most studies have concentrated on how to provide access to multithesauri and multidatabases. Few have addressed user satisfaction with the search results. This is actually a very important issue because users of online bibliographic databases or online catalogs are no longer limited to professional searchers.

Notwithstanding questions which have yet to be given optimal treatment, the present work toward the establishment of compatibility among indexing languages should be seen as becoming increasingly extensive. Further improvement in compatibility will depend largely on two kinds of support. The first is the development of science and technologies such as artificial intelligence as well as the theoretical exploration in many fields such as linguistics, psychology, and information science. The second is the continuation of efforts for standardization led by international agencies such as UNESCO, IFLA, ISO as well as related national agencies of all countries.

Appendix 1: Chronology of Research Projects

Date	Reporter	Research	Subject/Field
1957	British Classification Research Group (UK)	Ur-Classification	Social science
1962	Hammond and Rosenborg (USA)	Convertibility study	ASTIA SH and AEC SH
1965	Hammond (USA)	Term profiles (superstructure)	NASA and DDC
1966	Henderson (cit.)	Dictionary of equivalents	Arms and atomic energy
1967	Bauer (FRG)	Microthesaurus	Chemistry
1969	Gardin (France)	Intermediate lexicon	Information science
1969	Hammond (USA)	Satellite thesaurus	
1969	Wall and Barnes (USA)	Automatic mapping	MeSH and ABV
1970	Coates (UK)	Intermediate lexicon	Information science
1970	Neville	"suprathesaurus"	
1972	Soergel (USA)	"Core Classification"	
1974-	Horsnell (UK)	Switched indexing	Library and information science
1974	Ducrot (France)	Intermediate lexicon	Multilingual processing
1974	Soergel (USA)	Universal Source Thesaurus	
1975	Neicu	Microthesaurus	Romanian national thesaurus

APPENDIX 1 *(continued)*

Date	Reporter	Research	Subject/Field
1976-1983	Niehoff et al. (Battelle Lab.)	Vocabulary Switching System (VSS)	10 Databases 15 Databases
1976	Durrance (USA)	Integrated subject authority file	CLISH & LCSH
1976-	Austin et al. British Library (UK)	PRECIS as switching mechanism	Translingual
1977	Beling and Wersig (FRG)	Intermediary language	Bundes-Thesaurus system
1977	De Besse	Terminology data bank	Multilingual processing
1977	Goetschalckx	Eurodicautom (Term bank)	
1977	Marcus et al. (USA)	Integrated vocabulary list	DIALOG, MEDLINE, ORBIT, Intres
1977	Plante et al. (France)	Macrothesaurus des Sciences et Techniques	Science and technology
1977	Schulz	TEAM system	Translation
1977	Unesco	INTERCONCEPT (Term bank)	Social science
1977	Williams, Preece, and Merrett (UK)	"Data base selector"	
1978	Sokolov et al.	"Universal paradigm," "Type paradigms," "Specific paradigms"	
1978	Williams (USA)	Transparent system	Chemical databases
1979	Giertz	SfB system as satellite of UDC	Building construction
1979	Toman and Lioyd (UNISIST)	Broad System of Ordering (BSO)	
1979	Wolff-Terroine (USA)	"Common subsumption scheme"	*COSATI Subject Category List*
1980	CLR, RLG and WLN (USA)	Integrated consistent authority file service	
1980	Scibor (Poland)	Thesaurus of Common Topics (TCT)	
1980	Bulter and Brandhorst and WEECN (USA)	Mixed "Vocabulary Guide" list	MeSH, ERIC, PA, SSIE, TEST
1981	BSI (UK)	ROOT Thesaurus	
1981	Dahlberg (FRG)	Compatibility matrix	
1982	Sager et al. (UNISCO)	Integrated Multilingual Thesaurus	Social sciences
1986	Schwartz and Eisenmann (USA)	BRS/Term database (KNAPP)	
1987	Library of Congress (USA)	LCSH as a superstructure	LCSH, LIV, TGM, etc.
1988	European Economic Community	Automatic mapping	ICD, SNOMED, EMTREE
1988	NLM (USA)	Metathesaurus (Switching language)	Medicine
1989	NAL (USA)	Universal Agricultural Thesaurus	Agriculture
1990	C. Roulin (Belgium)	Subthesauri as part of metathesaurus	European Education Thesaurus

Appendix 2: Acronyms and Abbreviations

AAT	*Art and Architecture Thesaurus*
ABV	*Agriculture/Biological Vocabulary*
ASTIA	Armed Service Technical Information Agency (Changed to DDC)
BC	*Bliss Bibliographic Classification*
BSI	British Standards Institution
BSO	*Broad System of Ordering*
CABI	C.A.B. International
Carlyle Systems	(TOMUS) The Online Multiple User System designed and operated by Carlyle Systems, Inc. of Berkeley, California
CATSS	The UTLAS Catalogue Support Systems
CC	*Colon Classification*
CHS	*Canadian Subject Headings*
CISTI	Canada Institute for Scientific and Technical Information
CLESH	*Canadian List of English Subject Headings*
CLR	Council on Library Resources
CRG	British Classification Research Group
DDC	Defense Documentation Center
DIALOG	DIALOG Information Services (formerly DIALOG Services of Lockheed Information System)
DOBIS	Dortmunder Bibliotheks System (Originally developed in Germany, it now supports bilingual union cataloging for NLC, CISTI, and a number of Canadian federal libraries)
EJC SH	*Engineers Joint Council Subject Headings*
EMTREE	Thesaurus used for indexing *Excerpta Medica*
FAO	Food and Agriculture Organization of the United Nations
Geac BPs	Geac Bibliographic Processing System (Developed by Geac Computers International, it was first installed at New York University and the Smithsonian Institution in 1986)
HLAS	*Handbook of Latin American Studies*
ICD	*International Classification of Diseases* (classification of the World Health Organization)
LC	Library of Congress
LCSH	*Library of Congress Subject Headings*
LIV	*Legislative Indexing Vocabulary*
MeSH	*Medical Subject Headings (NLM)*
NAL	National Agricultural Library
NASA	National Aeronautics and Space Administration
NLC	National Library of Canadian
NLM	National Library of Medicine
NOTIS	Northwestern Online Total Integrated System
NUCMC	National Union Catalog of Manuscript Collections
OPAC	Online public access catalog
ORION	Online integrated library system developed and operated at the University of California, Los Angeles
PA	*Psychological Abstracts*
PRECIS	PREserved Context Index System
RLG	Research Libraries Group
SNOMED	*Systematized Nomenclature of Medicine*
SSIE	Smithsonian Science Information Exchange
TCT	*Thesaurus of Common Topics*
TEST	*Thesaurus of Engineering and Scientific Terms*

APPENDIX 2 *(continued)*

TGM	*Library of Congress Thesaurus of Graphic Materials*
UDC	*Universal Decimal Classification*
UTLAS	University of Toronto Library Automation System (it is now a large computer service wholly owned subsidiary of International Thomson)
VSS	*Vocabulary Switching System*
WEECN	Women's Educational Equity Communications Network
WLN	Washington Library Network (now Western Library Network)

Appendix 3: Examples of Integrated Vocabulary

3-A VSS INPUT FORMAT

RECORD LAYOUT

Position	
1	Relationship Code (Alphanumeric, see below)
2	Blank
3-4	Line Sequence Number—for Continuation Cards (e.g., 01, 02)
5	Blank
6	Vocabulary Source Code (Alphabetic; Arbitrarily Assigned)
7	Blank
8-67	Term
68-80	Blank

Relationship *Code*	*Relationship*
0	LEAD (OR MAIN) TERM
1	SCOPE NOTE
2	USE
3	
4	USED-FOR (UF)
5	UF+
6	SPECIAL SCOPE NOTE
7	BROADER TERM (BT)
8	NARROWER TERM (NT)
9	
A	SEEN FROM (SF)
B	RELATED TERM (RT) as designated by supplier
C	Subject Category as designated by supplier
D	Suggested NT as designated by supplier
E	Suggested BT as designated by supplier
F	Suggested RT as designated by supplier
G	Array NT as designated by supplier
H	Array BT as designated by supplier
I	Array RT as designated by supplier
J	Top terms
K	Frequency count

Source: Ref. *1*, p. 26.

3-C WEECN VOCABULARY GUIDE

FEAR (MESH)		
	RT	CONFLICT (ERIC)
		CONFLICT (PA)
		CONFLICT (SSIE)
		CONFLICT (TEST)
		FEAR (ERIC)
		FEAR (SSIE)
		FEAR (TEST)
		PSYCHOLOGICAL PATTERNS (ERIC)
		STRESS (PSYCHOLOGY) (TEST)
		STRESS-BEHAVIORAL ASPECTS (SSIE)

(For each entry, it shows synonyms, near-synonyms, and otherwise-related terms from all vocabularies.)

Source: Ref. *8*.

3-B VSS TERM DISPLAY

PLEASE ENTER SEARCH TERM OR VSS COMMAND
? memory
SWITCH SUCCESSFUL

TERM TYPE	VOCAB	TERM
YOUR TERM	ERIC	MEMORY
YOUR TERM	PSYCH	MEMORY
REL PHRAS	ERIC	MEMORIZING
RELATED	ERIC	COGNITIVE PROCESSES
RELATED	PSYCH	COGNITIVE PROCESSES
RELATED	ERIC	CUES
RELATED	PSYCH	CUES
RELATED	PSYCH	FORGETTING
RELATED	PSYCH	HUMAN INFORMATION STORAGE
RELATED	PSYCH	LEARNING/
RELATED	ERIC	LEARNING
RELATED	PSYCH	MEMORY DISORDERS
RELATED	PSYCH	RELEARNING
RELATED	ERIC	RETENTION
RELATED	PSYCH	RETENTION
RELATED	ERIC	ROTE LEARNING
RELATED	PSYCH	ROTE LEARNING
RELATED	ERIC	LEARNING PROCESS
RELATED	ERIC	MNEMONICS
RELATED	ERIC	RECALL (PSYCHOLOGICAL)
RELATED	ERIC	RECOGNITION
RELATED	ERIC	VISUALIZATION
WD MATCH	ERIC	COMPUTER STORAGE DEVICES*
WD MATCH	ERIC	KINESTHETIC PERCEPTION*
WD MATCH	PSYCH	KINESTHETIC PERCEPTION*
WD MATCH	PSYCH	MEMORY FOR DESIGNS TEST
WD MATCH	PSYCH	SHORT TERM MEMORY
WD MATCH	PSYCH	MEMORY TRACE
WD MATCH	PSYCH	MEMORY DECAY

(When the term *memory* is input, for two databases, ERIC and PSYCH, exact matches, word variants, related terms, and world matches ("memory" as a component of other term) are shown. The terms marked * are brought in through thesaurus cross-references from "memory".)

Source: Ref. *8*.

Appendix 4: Broad System of Ordering (BSO) as a Basis for an Integrated Social Science Thesaurus

CORRELATION OF BSO SOCIAL SCIENCE FIELDS WITH EQUIVALENTS IN OTHER SYSTEMS

BSO		DDC		BBC	
530/588	*Social Sciences*	*300*	*Social Sciences*	*K*	*Social Sciences*
533	Cultural Anthropology	306	Cultural & institutions	KC	Cultural, social anthropology
535	Sociology	30	1/7 Sociology	KA	Sociology
537	Demography	304.6	Population (Demogr.)	—	
540	*Political sci., polit.*	320	Political science	R	Political science
542	Political inst. & org.	306.2	Political institutions	—	
543	Pol. org. patterns &	321	Kinds of governments	RJ	Political systems
544	Political history	320.9	History of political institutions	RJS	History by subject, political
545	Pol. of part. groupings	—		—	
546	Pol. of part. states &	—		RK	Pol. by place
550	*Public administration*	350-359	Public administr.	RO	Public administration
554	Central admin. & gov.	351	Central governments	RR	Central
556	Devolved admin.	352	Local governments	RS	Regional and Local
560	*Law. Juridical sciences*	340	Law	S	Law
562	Civil Law	346	Private Law	SBF	Private and public law
563	Public Law, Const. etc.	343	Public law, Constitutional, Criminal	SBF	Private and public law
565	International Law	341	International Law	SDD	International Law
567	Systems of Law	—		SE	Common law systems
568	Law of particular countries	349	Law of indiv. states &	—	
570/575	*Social welfare*	361	Social probl. & social welfare	Q	Social welfare
580	*Economics*	330	Economics	T	Economics
581, 89	Microeconomics	338.5	General Product economics	—	
582	Macroeconomics	339	Macroeconomics & rel. t.	—	
584	Economic organization	330.1	Economics system & theory	—	
586	Sectorial economics	—		TT	Economic systems, sect.
588	Management of enterpr.	658.2	Management of plants	TX	Management of enterprise

Source: Ref. *30*.

Appendix 5: Examples in Automatic Mapping

1. Exact match	CALORIMETRY *to* CALORIMETRY
2. Variant spellings	DATA BASE *to* DATABASE CATALOGUE *to* CATALOG
3. Word forms	ELECTRIC MOTERS *to* ELECTRICAL MOTERS
4. Inversions	DISEASE, HUMAN *to* HUMAN DISEASES
5. Via cross-references	VERY HIGH FREQUENCY *to* VHF (where one vocabulary contains a reference: VHF *use* VERY HIGH FREQUENCY)
6. Via hierarchy	CAROTID SINUS *to* ARTERIES (where the hierarchy in the mapped-from vocabulary indicates that CAROTID SINUS *BT* ARTERIES)

Source: Ref. *8,* p. 188.

Appendix 6: Examples of Mixed Vocabulary Searching in Online Bibliographical Databases

BRS/TERM record for the concept of "depression" in its psychological sense.

TI	DEPRESSION (PSYCHOLOGY)
ER	DEPRESSION-PSYCHOLOGY (1978+).
ME	DEPRESSION (OF MODERATE INTENSITY). CONSIDER ALSO: ANTIDEPRESSIVE-AGENTS. DEPRESSIVE-DISORDER+ (FOR PROMINENT OR PERSISTENT DEPRESSION).
NM	DEPRESSION.
PS	DEPRESSION-EMOTION+.
SO	CONSIDER: DEPRESSIVE (129010).
MN	NEURASTHENIA. DEPRESSION-INVOLUTIONAL.
PN	ANACLITIC-DEPRESSION. ENDOGENOUS-DEPRESSION. NEUROTIC-DEPRESSIVE-REACTION. POSTPARTUM-DEPRESSIVE. PSYCHOTIC-DEPRESSIVE-REACTION. REACTIVE-DEPRESSION.
FT	AGITATED DEPRESSION. DYSTHYMIA. MELANCHOLIA. DEPRESSION. DEPRESSIVE. DEPRESSED. DESPONDENT. WEEPINESS. DESPONDENCY. DEJECTION. GLOOMY. SADNESS. DESPAIR. DISCOURAGED. LOWERED MOOD TONE. DISCOURAGEMENT. DEPRESSIVE SYMPTOMS. MODERATE DEPRESSION. ANTIDEPRESSANTS. DEPRESSANTS. DEPRESSING. DEFEATIST. HOPELESS. WITHOUT FAITH. WITHOUT HOPE. CONSIDER ALSO: SELF DENIAL. SELF ABNEGATION. SELF NEGLECT. SELF ABUSE. WITHDRAWAL. SELF DEPRECATION. FEELINGS OF WORTHLESSNESS. LEARNED HELPLESSNESS. LONELINESS. GUILT. LISTLESSNESS. UNHAPPINESS.SELF DEPRECIATION. SELF ACCUSATION. HYPOCHONDRIA. SELF-DESTRUCTIVE TENDENCIES. STUPOR.

(Controlled terms are shown from the thesauri used by the ERIC (ER), MEDLINE (ME), NIMH (NM), PSYCINFO (PS), AND SOCIOLOGICAL ABSTRACTS (SO) databases. Narrower terms are also shown for MEDLINE (MN) and PSYCINFO (PS). Terms or phrases are suggested for free-text searching of this concept (FT)).

Source: Ref. *40,* p. 414.

Appendix 7: Examples of Mixed Vocabulary Searching in Integrated Library Systems

ORION: (Terms from different thesauri are retrieved in the same alphabetical list.)

Hits	***Term***	***Source***
25	Clinical psychology	(LCSH)
	Clinical psychology *SEE* Psychology Clinical	(Medical)
12	Psychology, Clinical	(Medical)
	Psychology, Clinical *SEE* Clinical Psychology	(LCSH)

DOBIS: (Subjects from all sources are displayed in an integrated alphabetical list, with the subject source clearly indicated. The number of bibliographic records linked to each heading is also shown.)

Canada—History—War of 1812	LCSH	10
Canada—History—War of 1812	CSH	50
Canada—History—War of 1812	CUT	2

Source: Ref. 7, pp. 13; 33.

REFERENCES

1. R. Niehoff, et al., *The Design and Evaluation of a Vocabulary Switching System for Use in Multi-base Search Environments,* Battelle Columbus Laboratories, Columbus, OH, 1980.
2. Unesco, *UNISIST Study Report on the Feasibility of a World Science Information System.* Unesco, Paris, 1971, p. 97.
3. E. Svenonius, "Compatibility of Retrieval Languages: Introduction to a Forum," *Int. Class.,* **10**(1), 2-4 (1983).
4. L. C. Smith, "UNISIST Revisited: Compatibility in the Context of Collaboratories," paper presented at the 5th International Study Conference on Classification Research, Toronto, Canada, June 24-28, 1991.
5. F. W. Lancaster and L. C. Smith, *Compatibility Issues Affecting Information Systems and Services,* UNESCO General Information Program, Paris, 1983, p. 6.
6. I. Dahlberg, "Conceptual Compatibility of Ordering Systems," *Int. Class.,* **10**(1), 5-8 (1983).
7. C. A. Mandel, *Multiple Thesauri in Online Library Bibliographic Systems: A Report Prepared for Library of Congress Processing Services,* Library of Congress, Washington, D. C., 1987.
8. F. W. Lancaster, "Compatibility and Convertibility," in *Vocabulary Control for Information Retrieval,* 2nd ed; Information Resources Press, Arlington, VA, 1986, Chap. 14.
9. I. Dahlberg, "Towards Establishment of Compatibility Between Indexing Languages," *Int. Class.,* **8**(2), 86-91 (1981).
10. E. J. Coates, "Switching Languages for Indexing," *J. Doc.,* **26**(2), 102-110 (1970).
11. V. Horsnell, and A. Merrett, *Intermediate Lexicon Research Project Phase 2,* Polytechnic of North London, School of Librarianship, London, 1978.
12. A. B. Antopolskii, A. D. Salii, and V. I. Fedosimov, "Automatic Retrieval Methods for Lexical Intersections of Information Retrieval Thesaurus," *Auto. Doc. Math. Linguistics,* **13**(1), 85-91 (1979).
13. H. H. Neville, "Feasibility Study of a Scheme for Reconciling Thesauri Covering a Common Subject," *J. Doc.,* **26**(4), 313-336 (1970).

14. G. Beling, and G. Wersig, "The New Concept of an Intermediary Languages System for Information Networks," in *EURIM II: A European Conference on the Application of Research in Information Services and Libraries,* W. E. Batten (ed.), Aslib, London, 1977, pp. 117-121
15. F. E. DeHart, and J. Glazier, "Computer Searching on PRECIS: An Exploration of Measuring Comparative Retrieval Effectiveness," *Int. Class.,* **11**(1), 3-8 (1984).
16. D. Austin, and J. A. Digger, "PRECIS: The Preserved Context Index System," *Libr. Res. Tech. Serv.,* **21**(1), 13-30 (1977).
17. R. Rada, "Maintaining Thesauri and Metathesauri," *Int. Class.,* **17**(3/4), 158-164 (1990).
18. C. Roulin, "Sub-thesauri as Part of a Metathesaurus;" paper presented at the 5th International Study Conference on Classification Research, Toronto, Canada, June 24-28, 1991.
19. B. L. Humphreys, "Unified Medical Language System: Progress Report," *Int. Class.,* **15**(2), 85-86 (1988).
20. R. Rada, "Connecting and Evaluating Thesauri: Issues and Cases," *Int. Class.,* **14**(2), 63-69 (1987).
21. *National Library of Medicine News,* October 1987, pp. 4-6.
22. L. Zeng, "Research and Development of Classification and Thesauri in China," paper presented at the 5th International Study Conference on Classification Research, Toronto, Canada, June 24-28, 1991.
23. D. Soergel, "A General Model for Indexing Languages: The Basis for Compatibility and Integration," in *Subject Retrieval in the Seventies: New Directions,* H. Wellisch and T. D. Wilson (eds.), Greenwood, Westport, CT, 1972, pp. 36-61.
24. D. Soergel, *Indexing Languages and Thesauri: Construction and Maintenance,* Melville Pub. Co., Los Angeles, CA, 1974.
25. S. G. Dextre, and T. M. Clarke, "A System for Machine-Aided Thesaurus Construction," *ASLIB Proc.,* **33**(3), 102-112 (1981).
26. E. Scibor, and E. Jabrzemska. "Thesaurus of Common Topics," in *Encyclopedia of Library and Information Science,* A. Kent (ed.), 95. Marcel Dekker, Inc., New York, 1989, vol. 44, pp. 388-395.
27. L. Zeng, "Establishing a Compatible General Vocabulary in China: The Capability," *Int. Class.* **17**(2), 91-98 (1990).
28. D. J. Foskett, "Concerning General and Special Classifications," *Int. Class.,* **18**(2), 87-91 (1991).
29. I. Dahlberg, "The Broad System for Ordering (BSO) as a Basis for an Integrated Social Sciences Thesaurus," *Int. Class.,* **7**(2), 66-72 (1980).
30. J. C. Sager, et al., "Thesaurus Integration in the Social Sciences, Part II: Stages Towards Integration; Part III: Guidelines for the Integration of Thesauri," *Int. Class.,* **9**(1), 19-26; **9**(2), 64-70 (1982).
31. P. J. Whitelock, "A Descriptor Bank of Social Science Terms," *Int. Class.,* **9**(3), 145-151 (1982).
32. J. Meyriat, "Social Science Information Languages: A Comparative Analysis," *Int. Class.,* **7**(2), 60-65 (1980).
33. P. Q. J. Andre, "Universal Agricultural Thesaurus Discussed," *Q. Bull. Int. Assoc. Agricul. Libr. Doc.,* **34**(3), 150-151 (1989).
34. E. Wall, and J. M. Barnes, *Intersystem Compatibility and Convertibility of Subject Vocabularies,* Auerbach, Philadelphia, 1969.
35. W. Hammond, and S. Rosenberg, *Experimental Study of Convertibility Between Large Technical Indexing Vocabularies,* Datatrol Corp., Silver Spring, MD, 1962.
36. M. E. Williams, "Online Retrieval—Today and Tomorrow," *Online Rev.,* **2**, 353-366 (1978).
37. F. W. Lancaster, C. Elliker, and T. H. Connell, "Subject Analysis," in *Annual Review of Information Science and Technology,* M. E. Williams (ed.), American Society for Information Science, Washington, D. C., 1989, Vol. 24, pp. 35-84.
38. R. Niehoff, and G. Mack, "The Vocabulary Switching System: Description of Evaluation Studies," *Int. Class.,* **12**(1), 2-6 (1985).
39. I. L. Travis, and R. Fidel, "Subject Analysis," in *Annual Review of Information Science and Technology,* M. E. Williams (ed.), American Society for Information Science, Washington, D. C., 1982, Vol. 17, pp. 123-157.
40. A. B. Piternick, "Vocabularies for Online Subject Searching," in *Encyclopedia of Library and Information Science,* A. Kent, (ed.), Marcel Dekker, Inc., New York, 1990, Vol. 45, pp. 399-420.
41. R. Niehoff, and S. Kwasny, "The Role of Automated Subject Switching in a Distributed Information Network," *Online Rev.,* **3**(2), 181-192 (1979).

42. *DIALOG Database Catalog* 1990.
43. C. Schwartz, and L. M. Eisenmann, "Subject Analysis," in *Annual Review of Information Science and Technology,* M. E. Williams (ed.), American Society for Information Science, Washington, D. C., 1986, Vol. 1, pp. 37-61.
44. F. Levy, "Compatibility Between Classifications and Thesauri: Evaluation of a First Study in the Field of Information Storage and Retrieval," paper presented at the conference of the International Federation for Documentation, Tokyo, 1967.
45. F. Qiu, "The Relationship Between General Thesauri and Special Thesauri," *Inform. Sci.,* **2,** 1982 (text in Chinese).

LEI ZENG

AUTOMATION OF LIBRARY AND INFORMATION SERVICES IN CHINA. I. MAINLAND

Introduction

China has had a recorded history for more than 4,000 years, while its civilization dates back much earlier. Collections of the official records of each dynasty may be traced as early as the Shang Dynasty (1766–1123 B.C.). Access to the collections at the imperial libraries of these dynasties was restricted to an extremely limited class of people: literati, scholarly aristocracy, and high government officials. The general public were not able to take advantage of the rich treasures. This situation changed, however, at the turn of the century.

In 1896, Li Tuan-fen proposed that libraries be established in Beijing and in the provincial capitals. In 1902, Lo Chen-yu advocated the establishment of libraries throughout the country (*1*). From 1905 to 1911, a series of imperial edicts brought about the establishment of thousands of new schools and many provincial libraries. By 1925, the number of public libraries was estimated at 290 (*2*). All types of libraries in China for the same year totaled 552. This figure grew rapidly and reached 1,527 in December 1931, and 2,818 in 1935 (*3*).

Another forceful factor that helped foster the realization of a modern library system in China came in the person of Mary Elizabeth Wood, an American librarian. Miss Wood came to China in 1899 and founded the first Chinese school library in 1902. Through her continuous efforts, she secured funds and sent two college graduates to the United States for library training. She also established, in 1920, the Boone Library School, the first library school in China (*4*). The graduates from Boone and the library schools in the United States introduced American library practice to most of the Chinese libraries. Consequently, American influence has played an important role in developing the concept of modern Chinese librarianship. An exception to this was under Soviet influence between 1949 and 1976.

The rapid growth of modern libraries came to an abrupt halt with the outbreak of the Sino–Japanese War in 1937 and did not recover until 1949. Following almost 15 years of continuous war, first with Japan, and then the civil war, the new government of the People's Republic of China (PRC) came into power.

According to a 1950 nationwide survey, there were only 391 libraries in operation at that time (*5*). The new government soon launched its vast program of library expansion. In 1958, the total number of different types of libraries jumped to 321,618, though many of these fell far short of adequate library standards (*6*). The number increased to 388,900 in 1987, and the quality had also improved greatly (*7*). In Taiwan, when the Nationalist Government evacuated there in 1949, only about 100 libraries were in operation. These included 17 public libraries, 6 college libraries, and a small number of special libraries (*8*). The government of the Republic of China, and the library profession spared no efforts in improving library services. As of 1967, there were a total of 695 libraries (*9*). This number increased to 3,672 in 1987 according to a study by the National Central Library (*10*).

General Background

Library and information services in the People's Republic of China are under the jurisdiction of two separate government agencies, although there have been discussions on combining them into one operation. The first Five-Year Plan (1952–1956) implemented by the new government stimulated much scientific and technological research in order to reach a national goal of building China into an industrialized country. A long-term development project in science and technology for 1957–1968 called for the creation of a scientific and technological information organization. The Institute of Scientific and Technological Information (ISTI) was established in October 1956 under the auspices of the Chinese Academy of Sciences (CAS) to serve the research needs of numerous institutes under the academy. In May 1958, ISTI was expanded and elevated to a new status as a scientific and technological information center as well as a coordinating organization for documentation, information retrieval and information analysis for the entire country. The result was the Institute of Scientific and Technological Information of China (ISTIC), now a government agency under the State Commission of Science and Technology (SCST) (*11*). Since then, ISTIs were established under every ministry, and most state commissions. Similar agencies usually were also created at the provincial level. Additionally, regional information centers were installed in seven major cities, and regional and national information exchange networks were set up in the country. Thus was spawned an enormous national information system parallel to the nation's library system, and information seekers in China are still served by units under these two separate systems.

The Chinese library system can be grouped into four major categories: (1) public libraries headed by the National Library of China (NLC, formerly the National Library of Beijing) and including the provincial, municipal, prefectural, and county libraries, etc., which are under the jurisdiction of the Administrative Bureau of Library Service (ABLS) of the Ministry of Culture; (2) the education libraries under the control of the Ministry of Education and including the university and college libraries and middle and elementary school libraries; (3) the libraries of CAS and a large number of libraries affiliated with other research institutes throughout the country, constituting the science and research libraries; and (4) the great number of factory and workshop libraries organized by the trade unions within factories. Though different types of libraries are under the administration of different central authorities, they are all coordinated by ABLS of the Ministry of Culture (*12*).

While Chinese library and information services belong to two parallel systems, information scientists, to a certain extent, have higher social status than librarians. Agency support for information centers is generally greater than that for libraries. It was only after the nation adopted its open policy that China's government officials began to realize that library and information sciences, especially in an academic environment, have many attributes in common; there is no real distinction between them.

The First Working Conference of Library and Information Science, held in 1978, was considered an important milestone in the development of the contemporary Chinese library system. It was at this meeting that the idea of integrating the library and information sciences was introduced for the first time in the new society (*13*). Although some progress has been made, and perceptions have gradually changed in the past ten years, libraries and information centers remain under the jurisdiction of completely independent government agencies.

Despite all these differences, the development of Chinese computerization is so intertwined between the two systems that this discussion will treat them as an integrated entity.

National Efforts

Initial interest in computers in the People's Republic of China began as early as 1956 with the formation of the Institute of Computing Technology (ICT) in CAS. The Institute produced in 1958 a first-generation vacuum-tube computer capable of 1,000 operations per second (*14*). In early 1965, ISTIC set out to study the possibilities of information retrieval and imported some basic equipment. Unfortunately, for about two decades, the automation of information and library processes achieved little accountable progress due to historic, economic, and political reasons as well as the problem of written Chinese. Toward the end of the Cultural Revolution (1966–1976), however, China's leaders decided that it was important to build a modern system of scientific communication, and directed their attention toward that goal.

Efforts began in August 1974 when a national plan known as the '748 Project' gave high priority to the computer processing of Chinese scripts, including automatic typesetting and information storage and retrieval. The major tasks for the plan included: (1) the creation of Chinese nomenclature and a thesaurus for science, (2) experimentation with foreign databases, using available, domestically made computers to develop operational information services for the Chinese scientific and technological communities, and (3) the development of computer techniques to process Chinese text (*15*).

In 1975, the fourth People's Congress raised the issue of the four modernization plans. At the 1978 National Science Conference, Vice-Premier for science and technology Fang I designated 'automation of key research centers' and 'national networking' as top priority goals of the modernization plan: to build within eight years enough databases and terminals to form an online bibliographic and information retrieval network. In order to reach this goal, the following steps were mapped out according to the information automation program formulated after the Conference:

1. To further strengthen the information organizations of the entire country so that the structure becomes more systematic, distribution of resources more rational, and coordination and cooperation more effective in sharing information resources with all.
2. To enhance the acquisition, reporting, and analysis of information as well as the training of personnel in the information profession in appropriate techniques so that they are prepared for automation.
3. To equip in a systematic way institutions with computers and other advanced facilities in order to perform machine retrieval, machine translation, and typesetting.
4. To establish a computer retrieval network among major information organizations.

In response, many activities were stimulated in the fields of librarianship, scientific and technological information work, and documentation among others. Many information delivery organizations began to study the application of computers and engaged in the development of computer-based information retrieval. By 1980, there were more than 20 information retrieval groups in the Beijing area alone (*16*).

Preparation for Chinese Database Creation

We begin our discussion with an account of the efforts in paving ways for the database creation. Standardization and related problems involving Chinese character processing will also be addressed.

EFFORTS ON STANDARDIZATION

Establishing National Standards

The development of standards regarding library and information services is one of the most important tasks before any actual steps can be taken in the creation and conversion of Chinese bibliographic records into machine-readable form. This is particularly important with regard to cooperation among libraries in an online environment. In response to this need, the Technical Committee on National Standardization of Documentation (TCNSD) was established on November 29, 1979 under the Bureau of National Standards (BNS) (*17*). TCNSD is a member of the International Standards Organization (ISO), TC46. The Committee's objective is to formulate, re-examine, and revise the national professional standards for documentation work. Initially, it focused on the analysis of the international standards relating to library and information science, the study of conditions within China, and the promotion of standardization. Under the supervision of TCNSD, NLC and several other institutions drafted many standards. Between 1978 and September 1990, 31 national standards in the related fields were issued, all compatible with international standards. They relate to codes for document types and media, Chinese character sets for information interchange, and standards for editing and abstracting journals, subject indexing, format for bibliographic records exchange on magnetic tapes, and cataloging rules for various types of publication (*18*). Another 13 national standards are waiting to be confirmed.

In October 1981, after accepting the recommendation of the National Conference on Standardization of Classification, Subject Heading on Retrieval System, the Bureau of National Standards advised the libraries around the nation to use the *Chinese Library Classification* and *Chinese Thesaurus of Subject Terms* on a trial basis. Meanwhile, they planned to develop them further for possible adoption as national standards (*19*). The issuance of these national rules greatly facilitated the data processing of both Chinese and foreign materials, and boosted cooperation and interchange of information on computers.

Adoption of International Standards

Policies and guidelines relating to international standards that represent radical breaks from the Chinese tradition were also established, and recommendations were made by the 1983 Conference on Standardization and Automation for Cataloging Western-Language Books. It was sponsored jointly by the TCNSD and the Working Committee of National Libraries of Higher Education (WCNLHE), a unit under the Ministry of Education. In the Conference, there was unanimous agreement for the adoption of the *Anglo-American Cataloguing Rules,* 2d edition (AACR2), the *International Standard Bibliographic Description* (ISBD), and the *Library of Congress Subject Headings* for cataloging Western-language books with modifications to suit local practice (*20*). This consensus not only solved practical problems encountered in developing a machine-readable catalog for Western-language books, but also provided a very favorable ground for the formulation and promotion of other standards for library use. Participants at this meeting also explored topics such as MARC (Machine-Readable Cataloging) format, the format of computerized bibliographic files, searching strategy, authority files, and networking techniques (*21*).

CHINESE CHARACTER PROCESSING

While the development may have taken different routes and speeds on either side of the Taiwan Strait, the fundamental problem of automating Chinese libraries and information services is the same, namely, the Chinese character processing by computers. Since the main collections in practically all Chinese libraries are in the Chinese language, the development of computerized Chinese character processing is therefore an important prerequisite for library automation and online information retrieval. In what follows, we discuss the technology of the Chinese character-processing system, the encoding methods, character pattern file organizations, necessary software, input/output methods, and the corresponding equipment for dealing with Chinese characters.

The technique of Chinese character processing is the essential foundation for developing Chinese databases. At the end of the Cultural Revolution, academic organizations and national institutes, such as the Computer Society and the Instruments Society, began to set up groups to study Chinese character processing (*22*). Emphasis during the first stage was on developing good Chinese character-encoding schemes and appropriate input equipment. To date, more than 400 encoding schemes have been designed. In April 1981, TCNSD presented the Code for Chinese Character Sets for Information Interchange (CCCSII) (Primary set, GB2312-80) to ISO/TC 46 Subcommittee 4. The code, adopted as a national standard on mainland China in May 1981, is a counterpart of the American Standard Code for Information Interchange (ASCII), which it subsumes. This not only enables the different Chinese processing systems to communicate with one another, but also facilitates communication between a Chinese system and a Western-manufactured system such as those in the network (*23*).

Focus then was shifted to developing a complete Chinese Character-processing system including both hardware and software. More than 100 facilities for Chinese character processing have been developed by 80 factories and research institutes (*24*). In 1978, ISTIC, with the help of two other institutes, conducted pilot tests with satisfactory results in translating 20 titles from English into Chinese (*25*). Beijing University and Beijing Normal University carried out Chinese character-processing trials successfully on the mainframe and microcomputer; Hangzhou Automation Institute developed a Chinese character terminal using an MC6800 microcomputer. Other units have also developed Chinese character software on microcomputers and can execute simple Chinese character processing (*26*). A chinese script online processing system using a 2D-2000 Chinese character terminal connected to UNIVAC 1100/10, IBM 4341, and M180 machines has been developed but needs further improvement (*27*). Additionally, many experiments were conducted on the automatic division of Chinese characters in text to expedite automatic indexing and abstracting. A prototype developed by Shanghai Jiao Tong University (SJTU) was tested in 1985 (*28*). Though the technique still needs refinement, it has laid a foundation for automatic indexing and retrieval of Chinese text.

Other institutions chose to purchase computers with Chinese character-processing capabilities from abroad. In early 1975, ISTIC imported a TK70 minicomputer and a T4100 computer with the Chinese character information processing system and developed a program to experiment with Chinese character typesetting and retrieval (*29*).

Since Chinese databases inevitably contain Roman alphabets as well as Chinese ideographs, it is necessary to use bilingual database management software. Most institutions elect to purchase existing software packages and modify them to handle the Chinese scripts. For instance, UNIDAS/BDS is the Beijing Documentation Service's version of a UNIDAS/DBMS running on UNIVAC machines but modified to handle Chinese script. The retrieval software MINISIS has also been adapted for processing Chinese characters. Another institution has leased the BRS/Search information management software that supports numerous table-driven, user-selectable facilities such as alternate character sets (*30*).

Another important development is a *Chinese Character Attribute Dictionary* and its software system (*31*). It was developed by NLC and passed its appraisal test on

December 28, 1985. The system consists of several Chinese character codes and is capable of converting between the simplified and traditional forms of Chinese characters. This makes it possible for libraries in one area to take advantage of the bibliographic records in Chinese created online by their counterparts in other areas of the world. The system also has the interchange information for Chinese and Pinyin/Wade-Giles romanization systems, and has therefore the ability of automatically transliterating Chinese into English alphabets. Not only can this attribute dictionary carry out tasks of processing Chinese characters for library and information services, it also meets the needs of general office management and library automation. It is a welcome product that fills the gap in the peripheral software for computer processing of information in Chinese. At present, more than 30 institutions have adopted this system.

The Ministry of Electronic Industry has specified this system as the basis for formulating national standards of Chinese character attributes. On November 3, 1987, the Computer Information Development Research Center of the Ministry and NLC signed an agreement to formulate its national standard. The completion of this project and its popularization and adoption will certainly enhance the creation of software for Chinese character processing. This will ultimately facilitate international cooperation in sharing each other's database.

By 1987, China already had made substantial progress in the field of Chinese character-processing techniques. Many microcomputers were equipped with Chinese character-processing systems, and many software tools were adopted to handle both English and Chinese languages. Some local area networks (LANs) were in popular service with Chinese character-processing capability. Softwares for Chinese character input, word processing and database creation with powerful functions also became available.

COMPILATION OF THESAURUS

Chinese Thesaurus of Subject Terms (*32*), the first national standard thesaurus in China, was designed as a general working vocabulary in all fields of science and technology covering broad subject topics for the Chinese library and information network. It can be used for indexing, storing, and retrieving Chinese documents, and is suitable for manual as well as computerized indexing and searching. Over 30 subject thesauri for use in the more specialized information centers are being developed in line with this thesaurus so that they will eventually form a single uniform work. Some of these subject thesauri as the *Chinese Chemical Engineering Thesaurus* already have been completed. As subject approach has not been traditionally provided in the Chinese library catalogs, the thesaurus is indeed a very important step toward computer applications in developing bibliographic databases and online retrieval.

At present, there have been over 60 subject lists compiled, of which over half were machine prepared. These systems normally have functions in editing and printing. Recent lists indicate further improvements. Beijing Information Engineering College has developed a subject control system that can recognize relations between keywords and produce references (*33*).

Four Phases of Development

The development of the computerization of Chinese library and information services can be roughly divided into four phases: (1) preparation, (2) experimentation, (3) development, and (4) implementation. Each phase has its own characteristics, although many of the projects overlap. It is noted that there are cases in which projects were initiated but never completed, or were realized much later than originally planned. There are also cases in which follow-up information simply is not available.

PREPARATION (1974–1979)

Early in this period, China was still isolated from the rest of the world. General library automation efforts started with studying the professional literature of the advanced countries. The most important venture involved machine-readable tapes produced by the Library of Congress (LC). Gradually, experiments with foreign bibliographic tapes on domestically made computers, and with the access to international databases also took place.

Introduction of LC MARC and Automation Technology

In 1975, Liu Guojun published "Introduction of MARC project" (*34*). In this article, he introduced, in detail, the concept and function of MARC developed by the Library of Congress to the Chinese professional field for the first time. He also discussed the importance of computer technology, and translated the LC MARC into Chinese. During 1975-1976, Liu published more than 10 papers on library automation. In 1976, he wrote another article on the compilation of catalogs by computers (*35*). Liu is regarded as a pioneer in introducing automation to Chinese libraries. In later years, many others also contributed by providing translations, comments, and visiting reports about library automation abroad. A typical example was "Application of Computer in Information Works," translated by Zeng and Gao (*36*). The feasibility of importing and using LC MARC tapes to compile Western-language catalogs in card and printed form was a topic of much interest.

Efforts of the National Library of China

The NLC is the counterpart of the Library of Congress in the People's Republic of China. Both are depository libraries of their nation's publications and make available their cataloging to other libraries through cards, floppy disks, and magnetic tapes.

To prepare for automation, the first step NLC took was to set up the automation development department in 1975. In the next three years, a dozen young staff members were sent to American universities and colleges to study computer science, both hardware and software. In the meantime, several staff members experienced in the application of computer technology in library service were transferred to the department. An examination and translation of literature on the subject, especially information dealing with the MARC tapes of the Library of Congress, and a study of the development of library automation in major countries were carried out when the exchange students returned to China. The functions required for the library's manual cataloging of Western-language materials and the feasibility of their computerization were investigated carefully. State-of-the-art hardware and software and Chinese character-processing techniques were also surveyed (*37*).

Development of Softwares

Softwares developed during this period were mostly experimented on domestic computers using assembly language for program design. There were many technical problems including limited storage capacity and lack of an advanced language for editing and translation. Additionally, these programs could not process bibliographic magnetic tapes with the international standard communication format.

Following the implementation of the Chinese character-processing project, the computer field began to experiment on softwares for information retrieval. As early as 1974, a team for computerized information retrieval was set up by the Documentation Center of the First Ministry of Machine Building. Bibliographic records for 500 English articles on casting were edited and stored in machine-readable form. An experiment in retrieving these records was carried out in November 1975 in cooperation with the Computing Center of the same Ministry using a domestic DJS-C4 computer. Again in December 1977, an online retrieval experiment was conducted with encouraging results on the terminal connected through a telephone line with the Computing Center, which was located in a different building (*38*). In 1975, BDS experimented with the imported *Government Reports Announcements* (GRA) tapes for searching strategy, while the Beijing Aeronautical College experimented with an English-language database of technical reports obtained from the National Technical Information Service (NTIS), both using a Felix-50 computer (*39*). The CAS ICT designed a QJ-111 information retrieval system on the domestic 111 computer. In 1976, in collaboration with the library of the Academy, 5,000 articles were input into the computer. Selective Dissemination of Information (SDI) service was offered in 1977. More than 10,000 articles were in the database in 1979. The system had 31 subscribers. Access to the database was via classification number, subject headings, and so on. A printed index also was available (*40*).

In the following years, many organizations, in conjunction with the impetus of the 748 Project, engaged in computer-based information retrieval experiments using either domestically or foreign-made machines. Similar experiments were performed at such institutions as ISTI of the Ministry of Mechanical Industry, the Information Retrieval Research Unit of the Mathematics Department of Nanjing University, and the Library Science Department of Wuhan University (*41*).

Importation of Foreign Software Technology

In 1975 Zeng and Gao translated a Japanese article on computer applications in information service (*42*). This work introduced the structure and operation of computers. Emphasis was given to computer application in information management, system design, retrieval methods, and using the LC MARC file as illustration for discussion. The NDTS-78 software in the COBOL language, developed by the Information Retrieval Research Unit of the Mathematics Department of Nanjing University, was based on the two retrieval methods (Kikuchi and Fukushima) described later. A few hundred bibliographic records for Western-language materials were input into the NDTS-78 System. This is the first program that used an advanced language. This software has been a very useful teaching example in the training classes of the COBOL language given by Nanjing University. Toward the end of the 1970s and in the early 1980s, about a dozen foreign databases were imported for SDI services. Almost all software developed for them used the retrieval techniques introduced by Zeng and Gao with some improvements.

In 1978, Beijing Computing Center was created with the help of the United Nations Development Program. A B6800 computer system and five HP-3000 minicomputers were imported along with three software packages: Birds, IMAGE, and MINISIS. Training was provided in China and in the United States. Experts on system analysis and information retrieval from the Library of Congress were also invited to give training sessions for Chinese professionals in the field. Chinese programmers thus learned a sophisticated computer language for writing programs and developing automated retrieval systems (*43*).

Personnel Training

Many workshops were conducted to teach library automation and information retrieval. The latter was also included in many regular college programs at the undergraduate or graduate level. Among these workshops, the "Training School of Computer Information Retrieval" has provided specialists who have become the catalysts of the important task of library and information automation in China. This training session was conducted jointly by the CAS Library and the Department of Library Science, Beijing University from March 16 to September 4, 1979. A total of 77 professionals, all of whom were already involved in developing automated systems for their own institutions, were selected from 45 different information units. Ten courses were offered covering important topics on computer applications and library automation (*44*).

Technological exchange programs with other countries also became very active. Foreign experts were invited to Chinese institutions to share their expertise in automation. Chinese professionals were sent on study tours or special training sessions at foreign institutions, especially those in the United States.

Formation of MARC Coordination Group

A seminar on Information Automation and Online Retrieval Systems held in Beijing from September 3–28, 1979 (*45*) and sponsored jointly by UNESCO, the Bureau National de l'Information Scientifique et Technique (BNIST, Paris), and ISTIC, provided momentum for the development of automation in Chinese library and information institutions. During this period, participants from major libraries in the Beijing area exchanged ideas about a future Chinese MARC. They acknowledged the complexity of Chinese character processing, the lack of equipment, and the urgent need for foreign scientific and technological information to carry out the national modernization policy. Subsequently, it was concluded that libraries should pool their efforts to create a MARC database for Western-language books in the Beijing area instead of starting with machine-readable cataloging for Chinese materials. The first step along this project path would be experimenting with LC MARC tapes on imported computers. On the basis of that experience, research and experiments with a Chinese MARC would then take place. This policy resulted in the formation of the Beijing Area Coordination Group for Research and Experiment on Machine-Readable Cataloging of Books in Western Languages (MARC Coordination Group).

EXPERIMENTATION (1979–1983)

Activities of the MARC Coordination Group

Organized in 1980, the MARC Coordination Group consists of six units: NLC, the library of CAS, the Beijing University, Qinghua University, the Chinese People's University, and the China National Publications Import & Export Corporation. It was the first systematic effort in China to experiment with computerization in library processes. The creation of this Group provided a new environment for Chinese libraries in the development of machine-readable catalogs. Its tasks were to analyze the format and contents of LC MARC II, to develop the cataloging and retrieval program, and to compile the Beijing-area union bulletin of new books in Western languages. All three tasks were completed successfully (*46*).

In 1980, NLC began to subscribe to MARC tapes from the Library of Congress. The Coordination Group built a simulated system in early 1981 and conducted experiments with MARC tapes on a Felix 3512 computer. The experiments were conducted on the library's acquisitions and cataloging processes to test the system's feasibility. Eighteen application programs were written for the system, enabling all the functions required to do online cataloging for Western-language books (*47*). Unfortunately, due to practical limitations, this simulated system was never put into use. In 1981, five library members of the MARC group input the data from their catalog cards for Western-language books and created the *Union Bulletin of Western-Language Books in Beijing Area* (*48*). In all, 8065 records were entered into this domestic database, from which two sample copies were compiled, a giant step forward in the management of computer data.

Due to the lack of uniform standards for cataloging and classification among participating libraries, many problems were encountered in processing the large amount of data. The experience of dealing with these problems helped professional workers to better understand the MARC format, however.

The *MARC Format Working Manual* (First Draft) is a direct product of this learning process. This project also resulted in a push for standardization in many areas affecting cooperation in an online environment. The project unfortunately was suspended until 1983 due to a shortage of funds after the completion of the two sample copies.

With LC MARC tapes, the National Library of China began in 1981 to provide SDI and subject retrieval services to units both within and outside of the library. MARC tapes were also used for interlibrary loan services.

In the latter part of 1982, NLC decided to experiment with a machine-readable catalog of serials in Western languages. The rationale was that: (1) more than 80 percent of foreign references cited by the researchers were from periodicals, but the union listing of foreign periodicals lagged far behind; (2) the publication of foreign periodicals subscribed to by Chinese libraries was generally more stable than the Chinese journals. Thus, once prepared, the list could be used for a long time with only periodic updating; (3) most of the foreign journals were concentrated in a few major libraries; hence the efforts of a small group could be shared by all libraries in the nation (*49*).

The long-term goals for this project included: (1) creating a list of social science journals in Western languages held by NLC. Such a list would facilitate library work in acquiring missing issues, preparing various catalogs and indexes, establishing a database for union lists, and providing various statistics; (2) establishing a national union list of Western-language journals, with separate files for scientific and technological journals and social science journals; (3) expanding the database and, eventually, establishing a network system to better serve the public.

Preparation for this project was well underway in 6 months, which included the draft of the cataloging rules for Western-language serial publications, data input, application programs, etc. One-fifth of the data was input by mid-1983. Though a more sophisticated system still needs to be developed, the project served as a useful testing ground for future endeavors.

Development of Computer Systems

Information retrieval in the PRC received more initial attention than library automation. This was evidenced by more support for equipment as well as personnel. Due to the difficulty in Chinese character processing and the lack of equipment, the utilization of computers in information services during the early stage was focused primarily on information retrieval services that were made possible by mounting foreign magnetic tapes on Chinese-made computers. Between 1975 and 1980, 25 foreign databases were imported by 12 units for developing SDI services (*50*).

In 1975, BDS purchased the GRA tapes and started a series of experiments. In 1981, it imported a UNIVAC 1100/70 computer and, in 1982, it developed the BDS Information Retrieval System (BDSIRS) using UNIDAS information retrieval software. In June 1982, a database of 480,000 articles was built from the GRA tapes covering a period of 1974 to 1981. By 1983, the entire GRA inventory of tapes was loaded (*51*). Remote retrieval terminals set up in Shanghai, Xian, and Chengdu for this system provided retrieval services to several databases including GRA. BDS later developed its own software to replace UNIDAS, which improved the system as well as increased its storage capacity (*52*).

Shanghai Jiao Tong University and Tongji University both installed terminals on their campuses to process and circulate library materials as well as to store and retrieve information online. These experiments included the indexing of several small files, developing retrieval software for them [primarily for SDI services], and creating several experimental offline and online information retrieval systems.

In 1981, the Computing Center of Gansu Province began working to develop the Library of Academia Sinica (the Latin name of CAS) Information Retrieval System

(LASIRS) on a Wang VS-80 minicomputer. The system was improved in 1984 for adoption by the Changcun Institute of Optical Precision Instruments on its M-160H computer (*53*). LASIRS, a domestic multiuser online information retrieval system, offers major functions similar to those of large-scale foreign online retrieval systems and is suitable for commercial database application, as well as for creating databases of individual libraries and other institutions. This is probably the first software system of its kind developed in China that has reached a level suitable for practical use. It was certainly a significant development toward a nationwide computer retrieval system.

Database Creation

In 1978, the CAS library created a database and a related retrieval system for laser literature. This was one of the first experimental online information retrieval projects in China. Since then, many more libraries and information centers have created their own information databases. Major database projects have started in both English and Chinese languages.

BDS has created several small factual databases for experimental purposes. A few projects have reached limited production status. The *Chinese Pharmacology Abstracts* database, for instance, was created in collaboration with the State Bureau of Medicine by ISTIC, on the TK70 minicomputer. This database offers products in three forms: abstracts, indexes, and card catalog. Wuhan University built a full-text database of Chinese modern literature consisting of more than 10,000 characters (*54*). Institutes within CAS have built a number of databases for specialized fields; they are mostly experimental in nature, however, and are not available for online access.

International Online Retrieval

Efforts to access foreign online databases began in 1980 when nine institutions in Beijing formed a cooperative network and arranged to use the databases of the DIALOG Information Services, Inc., and the System Development Corporation (SDC) via a DTC-382 terminal in Hong Kong (*55*). In 1982, the first international online retrieval was established when the Northern China Scientific Information Institute and the Petroleum Industry Information Institute set up China's first telex terminal in Beijing to access DIALOG and SDC (*56*). It greatly reduced the retrieval time, though at the expense of high telecommunication costs.

In the following years, six information institutions and libraries joined the cooperative network. Among them were the libraries of Beijing University and Beijing Agriculture University. The next step took place, with the support of UNESCO; in April 1983, a data communication line between Beijing and Rome, Italy was opened to eight 300-baud terminals (*57*). This helped to provide even better conditions for linking China to the international network. In July 1983, by means of this communication line, users in Beijing were able to search the European Space Agency's database ESA/IRS online. ISTIC, for example, set up in October 1983 the International Online Information Network Retrieval Service (IOINRS) and provided its users with access to the ESA system.

In March 1984, the communication line was further extended to link with Tymnet

ISBN: 0-8247-2277-9