call, two cornerstones of traditional retrieval assessment, are clearly described in section 2 through prose and without reference to mathematics.

In section 3 the connection between the format of information as normally perceived by people and its format in the computer is discussed. The role of records and fields in decomposing the information in databases is clearly illustrated in several examples. The significance of Boolean queries for information retrieval is explained. The trade-offs between data base costs and searching costs are discussed.

The last half of the book centers around index languages, particularly thesauri. In section 4 the function and structure of index languages are described. In section 5, the role of the indexing language in guiding indexing and searching is highlighted. The function of an index language is different depending on whether the entity-oriented or request-oriented view is taken. In the entity-oriented view a document description is desired which optimally characterizes the document in terms of capturing the content of the document precisely and completely. In the request-oriented view a description of a document is desired which is most likely to make that document accessible to the searchers who will want the document.

Soergel's book emphasizes the hierarchical nature of index languages. The hierarchy helps searchers and indexers find the concepts that they want and accordingly serves an important, specific function. Facet analysis is most useful in developing hierarchies. In facet analysis the characteristics of a concept are listed. These characteristics then serve as a guide in determining when one concept is broader or narrower than another. For instance, if one concept x inherits all the facets of another concept y, but x also has a facet which y doesn't then x may be considered narrower than y in a hierarchy. The degree of precombination and postcombination in an indexing language is also important. Soergel recommends that precombination be done to facilitate searching by users who are likely to know the precombined concepts of the thesaurus. However, the elemental concepts which are contained in the precombined concepts should also be available in the indexing language. Thus the language retains its flexibility for current and future users who might want to combine in unique ways sets of elemental concepts.

Soergel's book abounds with detailed, real-world examples. His experience with food thesauri manifests itself with frequent use of terms like frozen foods, canned foods, and distinctive distilled alcohols. On the other hand, there are also a number of systematic critiques of the components of ISAR systems. Diagrams and tables are liberally used to present these systems-theoretic analyses.

The Book is tailored towards a library-science type audience, although the preface notes that a wider audience is desired. No formulas are presented in the description of precision and recall in section 2 of the book. There ist, in fact, not a mathematical formula in the book. This contrasts markedly with some books that have a similar title. For instance, Salton and McGill's book *Intruduction to Modern Information Retrieval* (2) spends about half its pages discussing formulas that are used in characterizing documents by their frequencies of word co-occurrence. Salton and McGill's book at the same time gives almost no attention to thesauri.

Soergel's book won the American Society of Information Science's award as one of the best books of 1986. Soergel has taken a complex topic and beautifully organized and presented its subtleties. People from many disciplines and at many different levels of expertise may appreciate the richness of Soergel's understanding of the organization of thesauri-based ISAR systems.

Roy Rada

References

(1) Dagobert Soergel, *Indexing Languages and Thesauri: Construction and Maintenance* Wiley, New York, 1974.

(2) Gerard Salton and Michael McGill, *Introduction to Modern Information Retrieval*, McGraw-Hill, New York, 1983.

Dr. Roy Rada
National Library of Medicine
Bethesda, MD 20894, USA

DEGENS, P.O.; HERMES, H.-J.; OPITZ, O. (Eds.): **Die Klassifikation und ihr Umfeld (Classification and its Environment).** Proceedings der 10. Jahrestagung der Gesellschaft für Klassifikation eV, Münster/W 18.–21. Juni 1986. Frankfurt/Main: Indeks Verlag 1986. XXXII, 379 p. – ISBN 3-88672-017-9 (hardbound), 3-88672-016-0 (paperback).

The book consists of 39 papers (16 of them in English) arranged systematically into nine groups. However, they can be grouped under three main topics: (1) conceptual classification, (2) formal concept analysis, and (3) numerical classification.

(1) In conceptual classification the problems of indexing and retrieving something (information, books, data) predominate. 'Knowledge' is an important concept in this field worth an examination. Thus, different kinds of knowledge found in the AI literature were compiled by *J. Panyr* ("Knowledge and an approach to its taxonomy in the area of Artificial Intelligence") and brought into a knowledge taxonomy thought of as a substitute for an explicit definition of 'knowledge'. It is doubtful, however, whether the rather unprecise AI terminology can serve as a good starting point for such a task. What is the reason for the occurrence of different kinds of knowledge? Is it because there actually are different kinds of knowledge, or because of the expert system tools and/or programming languages used in AI, or because of the different subjects of knowledge? Relevant to indexing and searching is the knowledge on how to organize things to be archived, which ideally should be familiar to both the indexer and the searcher. Such an ideal condition will be found only in the case of an alphabetic order, whose role in classification was investigated by *G.J.A. Riesthuis* in his contribution "Alphabetic subject indexes and classification".

Normally, there will be a lack of agreement. With respect to indexing, a simple answer to this problem is the definition of a special representation structure. *J. Gesell* reports on the "New edition of the International Patent

Classification" (IPC-4). This edition features a change from a mono-hierarchical structure to a hybrid system as the first step towards a multi-hierarchical structure. *H.U. Weidemüller* (in "RSWK-application in the German Library: automatic permutation of index-term chains") outlines the necessary steps for an automatic permutation of index-terms as a new way of indexing at the German Library. Instead of using coordinate indexing, index-term chains are now being built according to the new rules of the subject catalogue (RSWK). Syntactic relationships are expressed through the position of a term within the chain. Indicators and operators had to be introduced which, along with the rules for their handling, form a sort of grammar. However, a special representation tool in general restricts possible applications: *W. Rescheleit* and *L. Menner* (in "Comparison of the knowledge representation language FRL with UDC and a faceted classification") compare the Universal Decimal Classification (UDC) and faceted classification with a knowledge base generated by the frame-oriented programming language FRL. It allows the bulding of a poly-hierarchical knowledge structure based on the generic relation (features of a class are passed on to subclasses). Knowledge bases violating this structure cannot be transformed into frames without loss of information. This fact applies to the UDC and faceted classification systems because they use (among other 'irregularities') also the partitive relation.

Since in the field of indexing, formal instructions predominate at present over contents-related aspects in, *H. Schnelling* "The present function of the alphabetic subject catalogue in a research library") stresses these latter aspects and (neglecting the formal ones) poses operationalistic questions as a framework for constitutive rules which help to objectify decision-making during the indexing process.

But because of the interrelationship of form and contents it seems inopportune to favor the one at the expense of the other. Approaches unifying both aspects in a single theory should be more promising. Some papers point in that direction. In adapting teaching material, the question of 'How should it be taught?' turns out to be a classification problem. *H. Löckenhoff* (in "Didactics of the systems approach: Outlining knowledge") suggests the systems approach of economics and social cybernetics as a basis for structuring and representing knowledge. To improve an indexing language, *R. Fugmann* (in "The function of semantic categories in indexing languages and in indexing") proposes the use of "semantic categories" defined as "extremely general concepts" of a given domain. These categories (somewhat misleadingly called 'semantic') may be interpreted as primitives of an indexing theory, whereas the grammar of the indexing language forms its formal part. The theory appears as yet not fully developed; some unclear phrasings are found, such as "the syntax of a grammar", "syntactic tool" (p. 58), "syntactic relation" (p. 59). *K. Veltman's* classification of fine arts and visual images ("A new classification for art") can be interpreted as elements of a metatheory on this domain. *J. Hölzl's* "Recommendation for commodity description" comes closest to a theory-oriented approach. He defines at first basic concepts (like goods, products, commodities and then gives principles for their handling.

With respect to searching, a strategy for online subject access has been developed by *G. Riesthuis* and *A.-M. Colenbrander-Dijkman* ("Subject access to central catalogues: Incompatibility issues of library classification systems and subject headings in subject cataloguing") to overcome the difficulties in handling subject retrieval in central catalogues caused by the lack of a common indexing and retrieval language. The strategy is to retrieve more documents from keywords, notations, etc. of a source document. This approach reportedly reduces the number of zero-hits and increases the number of relevant items retrieved because of the (indirect) use of a controlled vocabulary.

Another approach aims at an expert system-like search support. From the point of view of statistical data documentation, *M. Domokos-Gombosi* explains (in "Metaorder in statistical data documentation") how to develop a metainformation system for the handling of information about different information systems. It helps to find out whether or not data on a given domain exist, and where these data can possibly be found. *A Vasiljev* (in "Online subject access to library holdings") looks at the subject searching capabilities of online catalogues and discusses some research problems encountered in striving for their further improvement. Online catalogues are considered an extendable part of a larger library information system. The planned integration of two subject catalogues into one online catalogue is described.

Suitable ordering principles can also serve to stimulate new ideas. *H. Gasthuber* (in "Application of ordering principles in industrial innovation") discusses the exploitation of the systematic power inherent in commodity ordering systems serving industrial innovations. In detecting gaps occuring in the ordering systems, new products might be found like new elements were found in the past from gaps in the periodic system of elements. Systematics can also help to generalize known manufacturing procedures to new fields of application, and new methods can be derived from a systematically ordered list of all possible solutions.

After presenting a survey of the annual meetings of the German Classification Society of the past ten years, *I. Dahlberg* presents "Future perspectives for classification and indexing". Among other things, she emphasizes the need to work out a general order of knowledge and makes a plea for concept analysis as an independent discipline. Particularly in Germany there is a lack of suitable textbooks and introductory texts. She also suggested the offering of further educational programs particularly for all those who work with classification and indexing. This seems necessary because too many thesauri (and classification systems) are inadequately designed and used, as *W. Krumholz* reports form experiences of the conference on the "Use and Mis-Use of Thesauri" in Brussels, March '86, which conference aimed at setting up an inventory of problems encountered in compiling, using, and managing a thesaurus and to make recommendations to overcome them or to avoid their occurrence.

(2) The new approach of formal concept analysis can be regarded as an attempt to compute order and to make it transparent. The set of all concepts ordered by the hierarchical relation 'super/subconcept' forms – with respect to a given context – a complete lattice so that the methods of lattice theory can be utilized for concept analysis. As an example, *U. Kipke* and *R. Wille* discuss (in "Concept lattice schemes for the determination of objects") a subject classification within a concept lattice. Unlike treestructered systematics, this approach no longer requires that a question be answered before the next determination step can be taken. The super/subconcept relation involves the heredity of attributes from superconcept to subconcept which allows the introduction of an implication between one-valued attributes on concept lattices. *G. Ganter* and R. Wille (in "implications and dependences of attributes") show how the dependences between many-valued attributes can be translated into implications between one-valued attributes so that concept lattice techniques also will make functional dependences transparent. Closely related to the field of formal concept analysis is a heuristic proof technique proposed by *D. Gernert* (in "Classification as a proof technique and a heuristic tool in graph theory"). The idea is to subdivide a complex problem (an extremely difficult graph-theoretical proof) into numerous less complex problems (the proof concerning smaller and smaller subclasses of graphs). The problem of how to select suitable subgraphs still remains; this problem could be a challenge to formal concept analysis, where it might be considered a non-trivial application.

(3) The first group of papers belonging to numerical classification deals with discrimination and ordering of data in a wide sense. *J. Krauth* presents a "Modification of the nonparametric discrimination method of Kendall" which is extendable to more than two populations and which does not lose its selectivity with increasing numbers of data. *S. Krolak-Schwerdt* ("A graph-theoretical allocation criterion for single linkage") describes a graph-theoretical extension of single linkage using a maximum spanning tree and a homomorphic function for clustering conditions. The transitivity condition guarantees disjunct solutions but is very time-consuming. To save computer time, *O. Opitz* and *T. Bausch* (in "Nondisjunct classification with qualitative data") first look for the non-disjunct solutions doing without transitivity; thereafter they look in a second step for a method to stretch the non-disjunct classes in an appropriate way. *H. Späth* (in "Maximizing partitioning cluster criteria for quantitative data") discusses a maximizing partitioning cluster criterion for quantitative data where, in deviation from the usual way, the objects have to be subdivided into groups of maximum possible similarity formed by objects varying as widely as possible. It was shown from eight examples that from a formal point of view the exchange method, using the maximizing criterion, works better. Based on a trend-following procedure, *S. Stank* (in "The decomposition of time-varying mixtures") gives a decomposition approach for mixtures of time varying distributions. In order to make agglomerative hierarchical cluster methods robust, *W. Stach* and *P.O. Degens* (in "Starting more robust estimation of ultrametrics") suggest trimming, i.e. the neglecting of so many observations that the remaining observations fit the assumed model. The authors recommend their somewhat arbitrary method only if outliers are expected; they found a deterioration in case of normally-distributed errors.

In multivariate data analysis, *H.H. Bock* (in "Multidimensional scaling in the framework of cluster analysis") presents a method for a simultaneous representation of objects and classes whose distances optimally reflect the given object-to-object, object-to-class, and class-to-class dissimilarities. An analogous case of the Young-Householder method is described. *H. Feger* (in "Estimating relational structures from ordinal data") discusses the estimation of relational structures from ordinal data. A categorial approach introduced by *G. Herder* ("Dissimilarity coefficients for ordinally scaled data") leads to two uniquely determined dissimilarity measures for ordinal data. *R. Mathar* (in "Metric scaling in a class of constrained configurations") considers the scaling in a prescribed maximum number of dimensions for configurations whose points are grouped not too closely together. *L. Schubert* (in "A point-vector-model for different data structures in external analysis") provides another multi-dimensional scaling model for handling hierarchical and lattice-ordered data structures. Along with data ascertained by students and faculty members on problems of higher education, *B. Miebach* discusses "Subgroup comparisons in linear structural equation models based on cluster analysis".

Four papers are devoted to the reconstructing of phylogenetic trees and problems of biological taxonomy. In simulating phylogenetic trees, *A. Dress, A. v. Hasseler* and *M. Krueger* (in "Reconstructing phylogenetic trees using variants of the 'four-point-condition'") show empirically that the 'four-point-condition' as a reconstruction principle leads to very satisfying results. *B. Lausen* and *P.O. Degens* (in "Variance estimation and the reconstruction of phylogenies") deal with the problem of evaluating reconstructed phylogenies. They propose a three-object-variance-estimator which can be computed knowledge about the underlying true topology of the tree. The core of *M.M. Patzlaff's* graph-theoretical approach (in "Reconstruction of phylogenetic relations by graph-theoretical methods: Results and problems in the chemical systematics of cytochrome C") is the reconstruction of networks and trees with minimized total weight using Steiner corners. Results and problems are studied with the chemical systematics of cytochrome C. Some methodical remarks are made concerning the interplay between biological problems and their mathematical treatment. Results should be interpreted with care; computation may lead at best to a phylogenetic tree of cytochrome not necessarily identical with the phylogenetical tree of the organism having that cytochrome. This result comes very close to that obtained by *S. Scherer* and *H. Binder* ("Comparison of biological classifications based on amino acid sequences and traditional taxonomy"). In comparing the traditional biological taxonomy with a classification of amino acid sequences, they conclude that the molecular clock

hypothesis is not valid as a common measure in evolutionary biology.

The volume closes with applications in social sciences and medicine. *I. Balderjahn* (in "Cross validation of covariance structures in one and multi group analysis: The case of ecologically concerned consumers") discusses the cross validation of covariance structures in multigroup analysis with data from ecology-oriented consumer behavior. In cognitive psychology, probands are invited to subdivide objects into classes with respect to their similarity. *T. Eckes* (in "The sorting procedure for obtaining proximity data in multivariate psychological research") stresses the necessity to analyse such experimental results with suitable multivariate methods. Finally, *H.P. Schmidt* and *C. Oberwittler* (in "Numerical taxonomy of brain tumors: A challenge to contemporary mathematical classification") discuss problems and preliminary numerical approaches to the malignancy classification of astrocytomas and mixed gliomas.

The very different contributions are clearly subdivided with respect to their contents; they give a good overview on the present state of the art so that the reader will find many stimulating ideas. The book has a detailed index.

Peter Jaenecke

Dr. P. Jaenecke
Forschungszentrum FB, Standard Elektrik Lorenz AG
Ostendstr. 3, D-7530 Pforzheim

WALNE, Peter (Ed.): **Dictionary of Archival Terminology; Dictionnaire de terminologie archivistique.** Compiled by Frank B.Evans, Francois-J. Himly, P.Walne. München, FRG: K.S.Saur 1984. 226p. ISBN 3-598-20275-X

This convenient guide to archival terminology in seven languages will interest not only archivists - its primary audience - but also information specialists, terminologists and lexicographers. The work supersedes an earlier LEXICON OF ARCHIVAL TERMINOLOGY (Elsevier, 1964). Both projects were sponsored by the International Council on Archives (ICA), which started work on the first version in 1954. The ICA Working Party responsible for this revised text labored, under UNESCO sponsorship, from 1977 until 1984. It was chaired by Peter Walne, U.K., and included Charles Kecskeméti, executive secretary of ICA. Other members were François-J.Himly and Michel Duchein, France; Eckhart G.Franz, FRG; Antonio Arago, Spain; Filip J.Dolgih, USSR; and Frank B.Evans, USA and UNESCO representative.

Each entry contains two definitions, in English and French, arranged alphabetically according to the English terms. Equivalent terms in Dutch, German, Italian, Russian and Spanish are listed subsequently, without definitions. 503 numbered terms are defined and unnumbered cross-references from synonyms to the entry terms are interspersed in alphabetical order. The cross-references and first entries are all in English, alphabetically, so no index in English is needed. However, indexes coded to entry numbers are given for each of the six other languages. Consequently users can go directly to the defining entry from any synonym for a concept, as listed in these indexes.

Normally French equivalents are available for each concept, but we are not informed about French terms that lack English equivalents. When a term has two or more senses, the equivalent may not have identical senses. Translations can then be problematical. For example, CHARGE-OUT may mean the act of recording the removal of a document from storage in an archive, or the document used to record this action. In French, there is no term for the first of these senses. Accordingly the concept is defined by a direct translation of the English definition and dashes, in brackets - "(---)" - substitute for the missing term. When French terms do exist, however - as they usually do - they are defined in French, and users must judge by comparing the definitions whether the concepts are indeed identical.

There are two French terms for the second sense of "charge-out": *fiche de déplacement* and *fantôme*. However, "fantôme" has another sense, equivalent to the English, *dummy*, meaning a card or sheet used to replace the borrowed item in its normal storage place. Comparison with terms in the other languages may reveal additional terms for the concepts that lack unequivocal terms in English or French. For example, the first sense of "charge-out" may be called, in German, *Aushebung* or *Ausleihe*; the second sense, *Leihzettel*, or *Bestellzettel*; and the idea of a "dummy" - the second sense of "fantome" - can be represented by *Stellvertreter* or *Retent*.

This example shows how terminology varies between languages and it also illustrates a problem common to all technical glossaries that rely on the alphabetical arrangement of entries. If the concepts used in archival work had been classified according to their definitions, all the terms used for them in each of the seven languages could have been listed after each concept. (See (a-d) below). Those that are equivocal, i.e. used to represent more than one archival concept, could have been marked as such, and users would more easily find the unambiguous terms for each concept. Whenever a useful concept lacks an unequivocal term, the editor or users might feel encouraged to suggest new ones that would not be ambiguous. A classified glossary, moreover, would bring related concepts together for easy comparison, thereby enabling users to grasp their logical relationships very quickly.

A common complaint against a classified format is that it requires two steps in searching, to go from the index to the numbered record. However, only English users of this dictionary can go directly to the entry terms without using the index. The users of all six other languages have to consult an index first in order to find a concept's entry.

Moreover, because of the arbitrariness of any alphabetical arrangement, users easily miss logically related concepts. In the example just given, users will be reminded under the CHARGE-OUT entry to "see also PRODUCTION TICKET", but they will not be directed to the entry for DUMMY. We discover that a "production ticket" is a document signed by users when requesting a loan (or charge-out). Four related concepts that are involved here are displayed in the following:

(a) a document signed by the user requesting an item: PRODUCTION TICKET, BULLETIN DE DEMANDE, *BESTELLZETTEL*, -SCHEIN