Book Reviews

Universal Classification. Subject Analysis and Ordering Systems.

Proceedings 4th International Study Conference on Classification Research, 6th Annual Conference of Gesellschaft für Klassifikation, Augsburg, 28 June – 2 July 1982. 2 vols. Ed.by I.Dahlberg and J.M.Perreault; FID/CR and Ges.f.Klassif.eV. Frankfurt: INDEKS Verl. 1982/83. 360+216p. FID Publ.615. DM 92.40. ISBN 3-88672-010-1, -011-X

(Translation from text published in DK-Mitt 28(1984) No.1, p.3–4)

Towards the end of the 4th Study Conference on Classification Research, Eric de GROLIER summarised the papers presented to the Conference in five main fields of discussion (II,p.163–7):

- the fight to cope with new technologies,
- the fights for classification science to acquire a truly scientific status
- the fight against self-conceit
- the fight for universality
- the fight against dogmatism

To these points he added, with a little less emphasis, the evaluation of classification systems as a central topic of discussion of the Conference. According to this, there were more discussions about the attitudes of classification researchers than about the single problems of classification research. At Augsburg, de Grolier's evaluation immediately appeared to me to be plausible, and a subsequent study of the conference volumes has not led me to reverse my opinion.

Classification research did not present itself as self-satisfied and bound up in itself at Augsburg. On the contrary, a tentative willingness to tackle new research subjects was noticeable. New impulses came from the fields of cognitive science, data bank technology, logics, linguistics, epistemology and the social sciences. How tentative this willingness to tackle other fields of knowledge still is, is revealed in the small number of bibliographical references, which go beyond the limits of classification research, in the proceedings.

It is certainly not easy to incorporate ideas from other disciplines into one's own conception of one's own subject, especially when wider conceptual gaps are to be bridged.

The scientific status of a discipline - and therefore also of classification research - rests, inter alia, on the ability and the willingness to discuss with other disciplines. Starting points for more intensive discussions with other disciplines are to be found in the root problems of classification research.

Complex indexing languages and the representation of knowledge

One starting point for discussion with other disciplines was elaborated on by Robert FUGMANN ("The complementarity of natural and indexing languages"), II, p.86–101).

Fugmann compares the functional efficiency of natural language and indexing languages. The main advantage of indexing languages is that they make the representation of subjects in the retrieval system predictable. The great merit of natural languages lies in their specificity, the accuracy -adaptable to any situation - with which the subjects are given. Both accuracy and predictability of representation are important for retrieval performance. Indexing languages are structurally too primitive to be a match for the accuracy of natural languages.

The realisation that the structural primitiveness of indexing languages is avoidable prepares the way for a more intensive examination of the more complex indexing languages.

Faceted classification

Facet classification is a more complex classification method and indexing language. In recent years, it has come more and more under discussion, but it is not yet the common property of each and every documentalist and classification researcher, which is all the more regrettable.

Irene TRAVIS ("Faceted classification in an online environment", I, p.269–76) describes lucidly that faceted classification, despite its suitability as a solution to the problem, may meet with problems of acceptance: in the U.S.Energy Information Administration an online retrieval system with ADABAS was developmed on the basis of a faceted classification. It fulfills its purpose well. The user interface was formed on the same lines as that of a retrieval system with a thesaurus:

"We not only tried to make the DRD no more difficult to use than a system indexed by conventional alphabetical thesaurus, but, indeed, even to make it look like such a system." I, p.273).

Faceted classification concepts were even systematically replaced by thesaurus concepts in the system description.

Faceted classification and knowledge representation

The methodical principles of faceted classification further lead to procedures of knowledge representation approximately as they are used in Artificial Intelligence.

Tim CRAVEN ("The representation of facets in a general concept network for index display generation", I, p.157–65) presents this relationship most clearly. Craven's aim is a general index generator on a network basis. As usual the network consists of nodes, which represent the concepts, and of lines, which give the relationships between the concepts. The main question is now: How can the representation of subjects formulated by means of facet classification be shown as a semantic net? First of all, there is no generally accepted definition of "facet". Craven's suggestion is to derive facets with a function from the primary concepts (nodes) of the network. As the data base must be extendable, the facet deriving functions must be formulated as rules.

The relationship between faceted classification and general techniques of knowledge representation has also been recognized by Marek CIGANIK ("Fuzzy faceted thesaurus construction". I, p.166–72). He sees content analysis as a method of text processing which leads to a

structured knowledge basis. It can be considered as a thesaurus which contains complex data structures instead of concepts.

Dynamics in classification

A dynamic classification theory is not something new. Prithvi N.KAULA ("Rethinking on concepts in the study of classification", II, p.61-7) pointed out that as early as 1916, James Duff Brown characterised classification as a "mental process". Ranganathan, too, propounded the idea of a dynamic classification concept: every "universe of subjects" had, for him, his own peculiarities, was dynamic, infinite, multidimensional, multidirectional, and continuous.

A dynamic classification concept opens the way to a dialogue with the cognitive sciences, an idea which was expressed several times during the Conference.

A contribution to the development of the classification theoretical paradigm (in Augsburg, Thomas Kuhn was also referred to more than once) was made by Erhard OESER ("The two systems of knowledge organization: On the characteristics and foundations of a universal background system", II, p.41–51). Oeser remarks on the diversity of existing classification systems. He postulates a more standardised background system from which the various universal classification systems may be derived. The background system represents -which is quite in keeping with the linguistic sense - the structure in depth of the organisation of knowledge on which the documents are ultimately based which are ordered with the surface classification systems. The surface systems are static; the background system, on the other hand, is oriented toward the process of the acquisition of knowledge and the assimilation of information: "The mechanism of cognitive-method-change represents .., the basis for concept-change and term-change, as well as for change in the ordering system of our knowledge", (II, p.49). "What library, information and documentation science look for – i.e. the Universal Classification Systems (UCS) with a primarily static and stabilizing function – corresponds the better to reality the more it includes the dynamic aspects of acquisition and devaluation of human knowledge. This implies not only a critique of existing UCSs but also a positive support of all attempts at dynamizing this area which have helped to supplant the original monohierarchical, static structure of classification with a greater variety of systems. ... the system which forms the background of the classification system is itself not a classification system. It refers neither to concepts nor to their contents but to the methods of obtaining and eliminating knowledge, thereby forming a nonhierarchical typology". II, p.46).

That classification research which lays more stress on the cognitive process offers a better basis for a universal view of classification.

What does "universality" in classification mean?

Derek AUSTIN ("Basis concept classes and primitive relations", I, p.86–94) postulates: "Classification, in the broadest sense, can be described as the expression of relationships between concepts (represented by terms) such that the concepts and relationships together comprise a statement or proposition. This definition encompasses not only taxonomic systems, bibliographic classifications and controlled indexing languages, but also logical propositions ... and declarative statements in natural language; in fact, any set of terms related in a meaningful way." (I, p.86).

Such a wide interpretation of classification closely resembles the idea that mathematics is a general structural science. Such universality only makes sense if it aims at co-operation with other disciplines. Under these conditions, it will provide classification research with fruitful prospects for research.

Right at the beginning of the conference, Helmut ARNTZ ("Universality of classification?", II, P.31–40), on the other hand, questioned the traditional claim of classification systems to universality: Classification systems subdivide fields of knowledge, but they do not provide inter-disciplinary concepts which would allow the integration of knowledge split up in specialised disciplines. They are general, because they provide the concepts of all fields, but they are not universal, for they do not offer an integral view of the knowledge shown.

De GROLIER, in his summary of the results of the Conference, sees universality above all as "transculturality". He would like to see the "Eurocentricity" of classification discarded in favour of other culture groups. Nasser M.SWAYDAN's plea for the needs of developing countries concerning classification was most impressive. ("The universal classification and the needs of libraries in developing countries", I, p.321–8). He demands, from Europeans and Americans alike, support in the adaptation of classifications to the needs of libraries in Africa and Arab countries.

His closing words on the subject of universality were: "If these needs are met to entire satisfaction, the existing classification schemes will be considered universal classification schemes. Without achieving this goal, none of them is considered a universal classification in its scope." (I, p.328). In these three diverging opinions, the problem of universality as understood by classification research is clearly stated.

Classification in new application areas

Within our European culture, there is new ground for classification to break. This was demonstrated in a public library by Anneliese Mark PEJTERSEN ("A new approach to the classification of fiction", I, p.207–15) and Jutta AUSTIN ("The AMP classification system for fiction: Trial applications and retrieval tests", I, p.216–26). The content analysis of novels cannot be achieved using the procedure which has been developed for scientific documents. A reader-oriented description of narrative literature requires complementary depictive elements to record the plot and its backgrounds, the experience value of the novel, its availability, and its information value. The faceted depiction of the individual novel comes very close to the form of a prestructured abstract (Positionsreferat). First of all, a systematic catalogue with an alphabetic index served as a retrieval system, then a computer-aided retrieval with free text search was used.

New Technologies

Two speakers dealt with videotex. Nancy J.WILLIAM-

SON ("Videotex information retrieval systems: The logical development and optimization of tree structures in general online interactive systems", I, p.277–84) presented video retrieval as classification done up in a menu. The video search tree, despite the data processing system, corresponds in its structure to a linear, at most a two-dimensional manual retrieval system. This can be remedied to some extent by using references and alphabetic indexes.

A.H.SCHABAS ("Videotex information systems. Complements to the tree structure", I, p.285–91) also proposed complements to inadequate videotex retrieval. She called for additional orientation aids for the user, who only receives a small quantity of data on the screen, and, at least, for the possibility of a heading search with an AND-link to limit the number of responses.

According to Pauline A.COCHRANE ("Classification as a user's tool in online public access catalogues", I, p.260–8), the narrowness of the field of vision on the screen poses problems for the users of online library catalogues. Cochrane suggests a classification based retrieval system as an alternative to systems which allow a search in a natural language with automatic clustering and relevance feedback. She favours the values of traditional library classification, particularly in the transition from the manual to the computer-aided catalogue.

Harold BORKO ("The role of classification in online retrieval systems and automated libraries", I, p.235–45) offers a completely different point of view. He provokes his classification colleagues with a vision of an automated library in which reference search on a screen is a matter of course. In this dynamic library, library classification is replaced by dynamic clustering in the reference data bank which is oriented to the users' requirements. For Borko, the classification schema corresponds to the static requirements of systematic shelving and not to the requirements of information retrieval and the logical ordering of the index store independent of the physical object "book". He demands a radical, new approach to library classification which makes determined and consistent use of the possibilities of computer processing. His argumentation, too, amounts to a plea for a more dynamic conception of classification.

Brigitte Endres-Niggemeyer

Dr.B.Endres-Niggemeyer
Fachhochschule Hannover, FB BID
Hanomagstr. 8, 3000 Hannover 91

NIEMANN, Heinrich: **Klassifikation von Mustern.** (In German) (Pattern classification). Berlin–Heidelberg: Springer Verlag 1983. 340 p. ISBN 3-540-12642-2, 0-387-12642-2.

This is a book on pattern recognition and perception methods providing a thorough survey on the many facets of this topic for readers with an intermediate mathematical knowledge. It is written in a pleasant and often in an informal style and its large bibliography is a valuable guide to many specialized or practical applications or to the mathematical derivations which, as a rule, are only sketched in the text.

Pattern recognition is concerned with the mathematical and technical aspects of the automatic processing and evaluation of patterns of all kinds (texts, signals, pictures or objects described by multivariate numerical feature vectors or syntactic representations by symbols). In particular, this embraces the classification (= construction of and/or assignment to classes) of simple patterns as well as the analysis of very complex patterns (like photos, scenes, circuit diagrams etc.). Classes are formed mainly on the basis of training or learning samples (supervised or nonsupervised strategies, thus cluster analysis is invoked, too), and classification systems are assessed either by mathematical or statistical optimality criteria or by empirical simulation studies. The author partitions the subject (as well as the text) into several sections: Recording of information, preprocessing of data, feature extraction, and classification (distinguishing numerical from syntactic classification methods).

Chapter 1 gives a general introduction into the subject. Chapter 2 presents methods for preprocessing the data (numerical coding of general and specialized informations, threshold operations, reduction of noise by linear or Fourier transformations, normalization for strengthening contrasts, parallel and sequential processing of discrete patterns). Chapter 3 discusses the several methods of feature extraction (orthogonal expansions, discrete Fourier or Walsh transformations, filtering, linear approximation, problem-specific transformations, optimality criteria, extraction of symbols from pictures, applied case studies). Numerical classification methods are described in chapter 4 (Bayesian and likelihood methods, minimum-distance criteria, error probabilities, nonparametric and distribution free classification using separating hyperplanes or nearest-neighbour criteria, specialized methods for nominal data accounting for context, learning classificators, the influence of dimensionality). Chapter 5 presents a concise introduction to syntactic classification methods, mainly designed for the processing of structural properties of objects (grammars, e.g. string, tree, graph, or stochastic grammars, parsing, classification of strings by regular or context-free grammars, automatic construction of grammars). The last chapter 6 describes, as a practical application, the implementation of the methods in an automatic address reading machine and supplements the many other examples scattered through the text, from biology, engineering, photogrammetry etc.

The monograph is to be highly recommended for people which are accustomed to mathematical notations or probabilistic and statistical arguments (e.g. engineers, biometricians, computer specialists, mathematicians and statisticians). It parallels the two other books of the author: 'Methoden der Mustererkennung' (Akademische Verlagsgesellschaft, Frankfurt/Main, 1974) and 'Pattern analysis' (Springer, Berlin, 1981). Readers without a sound mathematical background may profit, however, from the informal introductions into the single chapters and subsections where the purpose and essence of the formal methods are clearly presented.

H.H. Bock

Prof. Dr. H.H. Bock
Inst. f. Statistik u. Wirtschaftsmathematik
RWTH Aachen, Wüllnerstr. 3, D-5100 Aachen