

The claim of the author that the Glossary also includes important European terms may raise hopes that cannot be filled. I found only a few terms from the German library science, they seemed to have got into the Glossary more by accident than by careful consideration. And when they were so fortunate to be included then it happened to be in the English translation (if they are not entered under their acronyms). So the Deutscher Bibliotheksverband is entered under German Library Association, but no reference is made from its original name. Mention is made of the Prussian Instructions, but I looked in vain for the Regeln für die Alphabetische Katalogisierung (RAK) which have superseded them. The same is true of the other countries of the European continent.

A problem is the alphabetization word by word. As in English orthography the spelling of compound terms often varies between spelling in two (or more) words, hyphenation and spelling in one word the user of the Glossary must not forget to look up such terms in the various possible ways of spelling if he does not succeed at the first attempt. Even the author could not wholly avoid the pitfalls of his filing order. The material pasted on the inside of the spine of a book is entered under back lining and backlining. A special problem are the hyphenated words. Sometimes the hyphen ranks as a blank, sometimes hyphenated words are treated as if they were one word. My recommendation is to change over to filing letter by letter.

Eberhard Sauppe

Prof. Dr. E. Sauppe
Fachhochschule Hannover, FB Bibliothekswesen, Information und Dokumentation
Hanomagstr. 8, D-3000 Hannover 91.

PANYR, Jiri: **Automatische Klassifikation und Information Retrieval**. Anwendung und Entwicklung komplexer Verfahren in Information-Retrieval-Systemen und ihre Evaluierung. (Automatic classification and information retrieval. Application and development of complex procedures in IR systems and their evaluation). Tübingen, FRG: M. Niemeyer Verlag 1986. 416 p., 358 refs., ISBN 3-484-31912-7; = Sprache und Information, Bd. 12

The title of this book is too broad. This is a dissertation on a specific procedure for automatic classification called STEINADLER. This procedure is a component of CONDOR (COmmunication in Natural language with Dialogue-Oriented Retrieval), a system developed by Siemens in Munich from 1973-1981. There is a very condensed review of some methods for automatic classification in Chapter 2, but the methods are simply described without giving their rationale; the interested reader should look for a good book on the subject. Chapter 11 gives a superficial literature report on retrieval system evaluation without much analysis. Chapter 12 on the role of automatic classification in combination with other techniques in a retrieval system is rather opaque.

The real contribution of the book is the description of the STEINADLER system in Chapters 3 through 10. The chapters are: 3. Principles of the STEINADLER approach; 4. Thesaurus structure; 5. The determination of the levels of hierarchy in the classification (formation of priority classes); 6. Cluster analysis within the indi-

vidual hierarchical levels of the classification; 7. Linkages between neighboring levels of hierarchy in the classification; 8. Possibilities for updating the classification (dynamic classification); 9. Interpretation of the classification results (relevance feedback).

This may very well be a good and useful system, but the book offers no evidence of its usefulness. There is no report of, or even reference to, results of retrieval experiments, even though such experiments have been carried out. There is no cost data. There is no indication of whether the system is in practical use anywhere, and if so, what experience has been gathered in such use. The reader therefore, cannot judge whether the complexities in the procedure being described really serve a purpose or whether indeed the effort needed to understand these complexities is worthwhile.

The description is by no means easy to read. It makes quite heavy use of symbols, subscripts, and superscripts, even when that is not necessary.

What is called a thesaurus in the book is a list of word stems extracted from a corpus of text with relationships among these stems derived through automatic classification. Given that synonym control is *not* guaranteed in such a thesaurus - as the author very usefully points out - it seems unwise to refer to these word stems as "thesaurus concepts" (Thesaurusbegriffe) (p.135).

On page 142 one finds the following assertion: "The descriptors which occur infrequently in the collection of documents when used in retrieval lead to higher precision (with accordingly smaller recall), while retrieval with general terms may lead to higher recall with a large number of unwanted documents". This generalization, however widespread it may be and however attractive it may seem to mathematicians, is not tenable; a narrow search with a narrow term and a broad search with a broad term will both have acceptable recall and acceptable precision. Recall and precision are not defined in terms of the large or small number of documents found but in terms of the relationship between what is found and what is wanted. At least the author recognizes the value of broad terms for further narrowing a request that uses more thematically focussed terms (p.146).

Dagobert Soergel

Prof. Dr. D. Soergel
College of Library and Information Services
University of Maryland, College Park, MD 20742, USA

CAPURRO, Rafael: **Hermeneutik der Fachinformation** (Hermeneutics of special-subject information). Freiburg-München: Karl Alber 1986. 239p., ISBN 3-495-47593-1

Naturally, not every factman finds himself confronted with it understandable by its very nature. Man's "being-in-the-world" has always been conditioned by a "pre-understanding" which he necessarily introduces into his every concrete understanding action.

To Rafael Capurro, any such pre-understanding has its roots in what, following Hannah Arendt, he defines as the fundamental characteristic of "human existence": the communication among and actions by people in their common world. Denying thereby the existence of an "encapsulated psyche" (Medard Boss), Capurro's hermeneutics - despite a terminology indebted to Heideggerian thinking - approaches Marxist dialectics (cf.

note 5, p.14) (- which the self-styled exclusive proponents thereof will, however, dismiss as a bourgeois attempt at "convergence").

In the first part of his book, Capurro topicalizes, as it were, the pre-understanding arrived at so far of the concepts of "hermeneutics" and "special-subject information" as well as first attempts at linking both fields together as undertaken in the theories of Langefors, Diemer, Henrichs and others.

In the second part, Capurro takes a critical look at existing recognition models by pointing out their reductionistic character. Thus, in his opinion, the widely supported image theory neglects the dialectical relationship between reality and consciousness, thus leading to a metaphysical separation of both; the "transmitter/channel/receiver model" tears human consciousness away from its contingency upon the given situation, while finally "Platonistic" models -e.g. by Popper or Brookes - entirely loosen themselves, through postulating a "third", intelligible world of objective knowledge, from the existence of acting human beings in a "communicated openness to the world" (p.124). Within the latter, the segment concerned with a special subject, and as such of particular interest to us here, constitutes an enclave whose inhabitants, through provisionally assuming a common pre-understanding or paradigm, aspire to a singular depth of insight. Through the publication of special-subject information within an information-retrieval framework, such information is potentially separated from the horizon, conditioning it, of the given pre-understanding, which may both be regarded as a potential (information-deforming) evil and as an (information-generating) opportunity.

Taking this into consideration, Capurro, in his third part, investigates the various information-retrieval stages (from the conceiving, as it were, of information to the retrieving thereof), whose sequence bears the imprint of the successive pre-understanding stages at their distinct updating levels. To Capurro, therefore, the ultimately decisive yardstick for assessing a retrieval result is, besides relevance, the "pertinence" as oriented to the practical needs of the given searcher. This pertinence, however, is circumscribed by temporal as well as cultural limits. Capurro cautions, for example, against the danger of an "information colonialism" as resulting from the necessarily ideological character of documentation systems. In the end, thesauri, classification, and retrieval languages reproduce, in his view, objectified pre-understandings, thus assuming on the part of the user a minimum degree of adaptation amounting, in the extreme case, to cultural alienation.

In his outstandingly systematic argumentation, Capurro's hermeneutics of special-subject information offers valuable assistance both to all those data-bank practitioners who entertain a shadow of a doubt concerning the "threefold Platonism" expected by them (namely of "the absolute textuality, (...) the absolute identity of author, indexer and searcher, and (...) the absolute universality and unambiguousness of a classification system") (p.112), as well as to all critical theoreticians, to whom Capurro's philosophical-historical localization of current problems may be recommended as exemplary.

Marduk Buscher

Hauptstr. 77, D-7570 Baden-Baden

BALLMER, Th.T.; BRENNENSTUHL, Waltraud: *Deutsche Verben* (German verbs). Tübingen: G.Narr Verl. 1986. XII,413 p., ISBN 387808-319-X

Bereits in Heft 1983-3 von *Int.Classif.* war eine Besprechung zur Untersuchung englischer Verben erschienen. Die Grundidee, Verben nach dem Gesichtspunkt der Sprachhandlung (speech act) zu systematisieren, entspricht im vorliegenden Buch dem englischsprachigen Vorgänger ebenso wie entscheidende theoretische Begründungen. Allerdings sind Theoriegebäude wie Wortschatz nun wesentlich ausgebaut und nahezu abgerundet.

In einer Zeitschrift für Klassifikation dürfen wir die linguistischen Aspekte des Buches vernachlässigen zugunsten des gewählten Ordnungssystems unter dem Blickwinkel von Informationsaufgaben. Es bleibt ohne hin noch genug an schwerer sprachwissenschaftlicher und methodologischer Kost zu bewältigen; denn dies ist eines derjenigen Werke, aus denen man kaum einen Satz, aus den Sätzen kaum ein Wort tilgen könnte, ohne den gedanklichen Zusammenhang einer Aussagenfolge zu beeinträchtigen.

Bereits aus den theoretischen Grundlagen der "Deutschen Verben" lassen sich Mängel traditioneller Indexierverfahren ableiten, z.B.: Einzelne Wörter (Deskriptoren) können zwar als Submerkmale fungieren, aber die Elementaraussage als die kleinste Informationseinheit fordert die Subjekt-Prädikat-Beziehung; wo aber findet man Indexierung nach Aussagen? Oder: Substantive bezeichnen gegenständliche oder abstrakte Dinge, Adjektive deren Eigenschaften, und nur Verben können Vorgänge (Prozesse) über einer Zeiteinheit abbilden. Aber Thesauri enthalten fast nur Substantive, kaum Adjektive, fast nie Verben, allenfalls substantivierte. Man kann das im ältesten Thesaurus (Roget 1952) nachprüfen wie auch in dessen deutscher Fassung (WEHRLE/EGGERS 1961): "Sein" (Sachgruppe 1), "Leben" (359), "Zischen, Schrillen, Schreien" (409-411), "Lernen" (539) und "Fasten" (956) sind wohl die einzigen verbalen Sachgruppen-Namen, aber auch sie nur in substantivierter Form. Wird denn bei der Informationsrecherche nie nach Prozessen, nach Funktionen, nach Handlungsabläufen gefragt; wenn doch, wie will man sie ohne Verben aussagefähig darstellen?

Ballmer/Brennenstuhl ordnen die Verben semantisch nach Klassen von Handlungsabläufen, morpho-syntaktisch nach den Mitspielern solcher Handlungen. Teilgebiete dieser Herangehensweise sind bekannt als Valenzen, als Kollokationen oder, in Aktionsfolge, als Frames. Auf die letztgenannten zielen die Verfasser auch hin und wagen sich damit in jenes Gebiet, das unscharf formuliert als Künstliche Intelligenz bezeichnet wird. Hier liegen auch, unabhängig von linguistischer Betrachtungsweise, die wichtigsten Anwendungsmöglichkeiten der Ergebnisse für die Informatik. Besonders die maschinelle Sprachverarbeitung vermag sich ja nur schwer von der Vorstellung zu trennen, man könne allein auf der Ebene der Lexik operieren, indem man ein Wort oder Wortfolgen vom Rechner durch andere Zeichenketten ersetzt und so, quasi additiv, ganze Texte übersetzen, analysieren, komprimieren oder selektieren könne.

Hauptanliegen des Buches ist es also, sinnsemantische Zusammenhänge in ihrer Beziehung zum Verb als Folge von Abläufen darzustellen. Allein diesem Ziel dient das