

This book is a laudable contribution. The author has admirably succeeded in presenting the high science of classification in simple words and in meaningful examples. It is a valuable service to the classification discipline and may provoke some good minds to learn more of it.

Mohinder Partap Satija

M.P. Satija, School of Library and Information Science
Guru Nanak Dev University, Amritsar, India

LUSTIG, Gerhard (Ed.): **Automatische Indexierung zwischen Forschung und Anwendung**. (Automatic indexing between research and application). Hildesheim: Olms 1986. 182p.

In information science or information retrieval, we actually find three main lines of research – the system-driven approach, the user modelling view and the cognitive paradigm (1,p.150). Among them, the system-driven approach is the eldest. The research and development work which is reported by Gerhard Lustig and his collaborators is an example of this system-driven approach. After some 25 years of scientific efforts, the time of applicable results in automatic indexing has matured. In the past years, the AIR (Automatic Indexing and Retrieval) system by Gerhard Lustig and his colleagues has made its way to practical application. At the Karlsruhe information center for Energy, Physics and Mathematics (FIZ 4) it is applied to index Physics Abstracts.

The authors concentrate on a documentation of their work for readers interested in technical information about the AIR system. They do not so much aim at a larger public with a more general interest in the problem of automatic indexing (or content analysis) and its possible solutions. To accommodate readers of this type, a broader perspective and a picture of the whole research field around the AIR approach would have been helpful.

The description of the AIR system with the research to bring it about is divided into four parts. First, the AIR concept is explained. The following chapters treat the dictionaries of the system, the indexing method itself and the work which was necessary to test the system and to transfer it into its working environment at FIZ 4.

Gerhard Lustig specifies the actual indexing task: AIR indexes English abstracts with concepts from a thesaurus. Then, he expounds on the basic concepts of the system. They inspire respect because they are the foundations of a system that really works. AIR mimics the work and results of a human indexer. Its central instrument is a huge indexing dictionary. It contains pairs of text terms (words or word groups) and descriptors. The text terms point to descriptors with weighted relations. The strength of the relations has been calculated from the co-occurrence of text terms and descriptors in manually indexed documents. In the indexing process, all hints at a descriptor are collected and evaluated. If a descriptor reaches a sufficient score, it is assigned. The use of multiple information in every indexing decision makes the system tolerant against erroneous data. This feature is important because there is no way to avoid faulty data, for instance, in the dictionary. The dictionary size calls for automated acquisition methods. Inevitably, they will

yield a certain quantity of false entries. The quality of the indexing results is evaluated by a large retrieval test.

The idea configuration built up by Gerhard Lustig is, of course, a subject for discussion. In order to illustrate the kind of possible questions, I discuss two points in some detail.

It is a quite common practice to index documents from their abstract, referring to the original only in case of doubt. Like this, the indexing step becomes much quicker and easier, because the abstractor has already cleared away most of the information which is not to be used for indexing. An automatic indexing system that indexes abstracts does just the same as the indexer when working from abstracts – it performs only the last and comparatively easy part of the whole process. It relies on the intellectual effort of the human abstractor who actually did most of the work while (s)he reduced the information size of the original document to the small size of the abstract. Further research efforts must mention this point if they aim at methods for a real full-text indexing. Consequently, they cannot avoid to conceive the problem of automatic indexing as a part of the more general problem to represent a document's meaning.

While the AIR group was still busy to develop a system along the main lines presented above, the situation in information practice and research has changed. Concepts defined for the AIR system may no longer be taken for granted in present and future work. This is, the case, for instance, with the strict submission of indexing quality to retrieval success as advocated by Lustig. It seems much less compulsory in a more articulated model of the process that transmits information from an author to the user (1,p.161), and when the user gets, in an advanced information industry, an information product rather than a retrieval result. Today, Lustig's principle of probabilistic modelling and of a system which tolerates erroneous data would find its counterpart in a new effort to acquire more positive knowledge and to obtain a more informed model of the process performed by the system.

What follows Gerhard Lustig's outline of AIR in this book is dedicated to more or less technical problems of the system. Readers interested only in an overall appreciation may safely get straight down to the end of the book and concentrate on the retrieval test and the transition of the system to its working environment at FIZ 4.

Barbara KIENITZ-VOLLMER and Johannes REICHARDT treat the problem of dictionary construction. English compound terms are isolated in abstract texts by boundary items like function words. Additional selection criteria improve the efficiency of the method. It is necessary to speed up the dictionary construction process by pragmatic measures, e.g. by dropping rare compounds.

Formulas present a special problem for the indexing of physics abstracts. Ingeborg MUHLER and Jürgen REPP explain how the formulas are identified in the text, analysed, classified and used as descriptors. The contribution of formulas to the indexing result is positive.

Norbert FUHR, Roswitha JÄGER-BECK and Michael SCHWANTNER show how the statistical relations between text terms and descriptors (z-values) are obtained. As very many co-occurrences of terms and descriptors enter into the calculation, pragmatic shortcuts of the computation are necessary.

Peter BIEBRICHER, Norbert FUHR, Gerhard KNORZ and Michael SCHWANTNER give an account of the technical structure of the system's dictionaries as they were used in different stages of the system development.

The AIR system proposes two indexing functions. The Boolean solution is explained by Ute BEINKE-GEISER, Gerhard LUSTIG and Gudrun PUTZE-MEIER. Gerhard KNORZ describes his polynomial approach to the problem.

Peter BIEBRICHER, Norbert FUHR and Bernd NIEWELT report on the AIR retrieval test. For possible users of automated indexing methods, the results are quite interesting. The test was based on 309 queries and 15.000 documents. Surprisingly, the automatic indexation retrieved relevant documents that were not found by the manual indexation. The manual indexation attained better precision values, the automatic one made more recall. On the whole, the manual indexation performed better, but there was no deep quality gap between manual and automatic indexing.

Because the AIR method yields a weighted indexing, the indexing weights may be used to rank the query result. Norbert FUHR explains his related experiments.

In the final chapter, Gerhard KNORZ describes some of the very practical organization problems that had to be solved to make the system fit its application environment. They concern the whole input processing framework at FIZ 4. Consequently, the author broadens the horizon from the special problem of automatic indexing to a comprehensive reorganization of the input to bibliographic data bases, including descriptive cataloguing. For descriptive cataloguing, he passes to a knowledge-based approach. Like this, he opens a view at research under the cognitive paradigm in information science.

Brigitte Endres-Niggemeyer

(1) Ingwersen, Peter: Towards a new research paradigm in information retrieval. In: Wormell, Irene (Ed.): Knowledge engineering. Expert systems and information retrieval. London: Taylor Graham 1986. 182p. (herep. 150-168).

Prof. Dr. B. Endres-Niggemeyer, FHS Hannover, FB BID
Hanomagstr. 8, D-3000 Hannover 91

Hoffmann, Lothar: **Kommunikationsmittel Fachsprache** (Special language as a communication medium). Eine Einführung. 2. völlig Neubearb. Aufl. Tübingen: G. Narr Verl. 1985. 307 p. = Forum für Fachsprachen-Forschung, Bd. I. ISBN 3-87808-771-3 (geb.), -875-2 (kt.)

Seit dem ersten Erscheinen von "Kommunikationsmittel Fachsprache" ist mehr als ein Jahrzehnt vergangen – und das Interesse an der darin behandelten Thematik ist noch immer im Zunehmen begriffen. Es war daher die erklärte Absicht des Autors, in der nun vorliegenden zweiten und völlig neu bearbeiteten Auflage des Bandes wenigstens einen Teil der diesbezüglich neueren Entwicklungen und Erkenntnisse aufzunehmen; eine Absicht, die teilweise aus Raumgründen scheitern mußte, die aber dennoch in bezug auf dieses Grundanliegen als gelungen bezeichnet werden kann.

Im ersten Hauptabschnitt über das Wesen der Fachsprachen werden verschiedene Grundströmungen der Fachsprachenforschung zusammenfassend erläutert, wobei es gegenüber der ersten Auflage sowohl zu einer Straffung als auch Ausweitung der behandelten Themen kommt. Zu den bisherigen Schwerpunkten, wie den lexikologisch-terminologischen und funktionslinguistischen Ansätzen, kommen neuerdings z.B. auch Betrachtungen zur sog. Wirtschaftslinguistik. Zu der klassischen Definition von Fachsprache als der Gesamtheit der sprachlichen Mittel, die in einem fachlich begrenzten Kommunikationsbereich verwendet werden, um die Verständigung zwischen den in diesem Bereich tätigen Menschen zu gewährleisten, sowie der bekannten Konzeption von einer horizontalen und vertikalen Gliederung der Fachsprachen, kommt nun als neuartiges Konzept die Lehre von den Subsprachen, die nach der Gemein-

und Fremdsprache als Teil einer Gesamtsprache angesehen werden müssen.

Der zweite und Hauptteil des Bandes beschäftigt sich mit der Darstellung der spezifischen Merkmale des Fachsprachen auf den einzelnen sprachlichen Ebenen und Zwischenebenen, d.h. auf der Ebene der Grapheme und Phoneme, der Ebene der Morpheme und der grammatischen Kategorien, der Ebene der Lexeme und der Wortformen und schließlich der Ebene der Syntagmen, Phrasen und Sätze. In der vorliegenden Neuauflage wird versucht, über die formale Struktur des Satzes hinausgehend auch die Funktionalität der Sätze und ihrer Konstituenten besser zu erfassen. Dabei wurde sowohl der aktuellen Satzgliederung und – im Gefolge eines neuerdings verstärkten Interesses der Fachsprachenforschung am Verbum und seiner Umgebung – auch valenztheoretischen Ansätzen erhöhte Aufmerksamkeit geschenkt. Völlig neu ist schließlich die Aufnahme eines eigenen Abschnittes über die Ebene der Texte, worin die textlinguistischen Betrachtungen im Rahmen der Fachsprachenforschung, denen ein neueres Werk des Autors gewidmet ist, bereits ansatzweise vorweggenommen werden.

Im dritten Teil des Bandes, der den Methoden zur Ermittlung der Spezifik der Fachsprache gewidmet ist, bleibt der für die fachsprachliche Forschung in Leipzig typische statistische Grundansatz weiterhin dominant, wobei die Weiterentwicklung dieses Ansatzes in den neuen Partien des Buches dokumentiert wird. Beispiele für weitere strukturelle, semantische, funktionale und vor allem auch komparative Verfahren, die auch diesmal aus Raumgründen nicht aufgenommen werden konnten, wären sicherlich eine Bereicherung gewesen. Es steht zu hoffen, daß dies in absehbarer Zeit möglich sein wird, um hier eine gewisse Einseitigkeit zu vermeiden.

Gegenüber der ersten Auflage kommen jedoch noch eine Reihe von Verbesserungen zu den bereits bekannten Vorzügen des Werkes hinzu. Es bietet mit z.T. wesentlichen Erweiterungen in Umrissen einen weiterhin gültigen Rahmen für die sowohl theoretische als auch praxisbezogene Beschäftigung mit den Fachsprachen, an dem sich weitere Einzeluntersuchungen orientieren können. Das Inhaltsverzeichnis mit seiner stärkeren Durchgliederung mit Hilfe der Dezimalklassifikation und einer größeren Zahl von Zwischentiteln erleichtert die Benutzung des Bandes als Nachschlagewerk. Denselben Zweck erfüllt die Erweiterung des Sach- und die Neuaufnahme eines Personenregisters. Auch das Literaturverzeichnis wurde aktualisiert und durch weitere Verweisungen ergänzt. Vollständigkeit wurde dabei nicht angestrebt und wäre im vorgegebenen Rahmen auch kaum realisierbar; es vermittelt jedoch einen guten Überblick über den letzten Stand der Forschung, besonders in den sozialistischen Ländern.

Der nun vorliegende Band bietet sich daher, vor allem auch im didaktischen Kontext – wie mehrfach festgestellt werden konnte – in hervorragender Weise als Einführungs- und Nachschlagewerk an.

Hildegund Bühler

Prof. Dr. H. Bühler, Institut für Übersetzer- und Dolmetscherausbildung an der Universität Wien
Gymnasiumstr. 50, A-1190 Wien, Austria