ask for . . .". I was really pleased about this book and enjoyed reading it very much. In my opinion it has the right mixture between informal and formal presentation. Very often it contains nice real-life examples as a motivation for the reader, and as a hint for the practicability of the various methods for related problems.

The main emphasis of the book obviously is on hashing and on bit level operations (signatures, superimposed coding) as these parts encompass almost half of the book; the other parts are devoted to B-trees and other tree structures, and especially also to approximate string searching. It contains also a nicely written section on the various sorting methods.

The following list is an extract from the contents of the book in order to characterize its emphasis rather than to quote from the table of contents. The introduction puts together basic notions on files and storage media. The first main part (1) contains binary and interpolation search and discusses self-organizing sequential search. But the main contents here is devoted to all different and relevant variants of hashing including comparisons of collisions resolution and including perfect hashing. The next part (2) is mainly devoted to signatures as binary attribute representations, superimposed coding and its application to record *and* text searches. It contains partial match retrieval with signature trees and a comparison to inverted files. A significant section is devoted to Bloom filters and it explains that this method could be applied much more often for various problems.

Part 3 contains the classical subject on binary trees, B-trees, B+-trees, B#-trees. Furthermore this part is remarkable in that in contains the new hashing techniques for explandable files. In addition it contains a section on tries, like PATRICIA. It explains how approximate string matching can be supported. It also contains a nice presentation on grid files. The last part (4) then is devoted to sorting in a convincing presentation.

The book is surprisingly complete. I made a sample on perfect hashing and found a complete list of the relevant literature on this object. Moreover and once more I was impressed of the brilliant characterization of the usefulness of perfect hashing for practical problems. Similar observations can be made on the (much huger) field of signature techniques and signature trees.

To sum up, I can recommend this book strongly to everybody who is interested in the practical representation of information e.g. in databases, expert systems or in text processing systems. H.-J. Schek

Prof. H.-J. Schek, Inst. f. Informationssysteme, ETH-Zentrum, CH-8092 Zürich, Switzerland

HUNTER, Eric J.: **Classification – Made Simple.** Aldershot, UK: Gower Publ. Co. 1988. IX, 115p. ISBN 0-566-05605-4(pbk).

Eric J. Hunter is well known to those who keep in touch with the literature on the organisation of knowledge. The range of his writings is wide, both in variety and in scholarship: textbooks on descriptive cataloguing, research papers on historical developments in classification, and now "Classification – made simple". The book under review explains the scope, basics, limitations and techniques of library classification laid thread bare. The author is a versatile and gifted teacher.

This is a unique book in its refreshing approach. It emphasises that the role of classification is to irganize items in relation to the ultimate purpose in mind, i.e. to create relations. Major stress is laid on explaining the technique of designing classification systems for information storage and retrieval. Faceted classification systems figure mostly in the text. Emphasis on faceted systems is justified as it "enables it to be more easily interpretable by both human beings and computers". However, the existing practical realities of the art have not been ignored. We find chapters such as: Use of synthesis on a basically enumerative scheme; synthesis grafted on to an enumerative scheme. Incidentally this matches well with Ranganathan's different species of library classification. The entire text is divided into 13 chapters of very small to medium length. Practical aims as the book has, it is not lacking theoretical questions: Classification as a tool of search, Classification and thesaurus, modern developments in classification in the form of switching languages, role of classification in online databases; and lastly also so-called automatic classification. Common place questions such as advantages and disadvantages of faceted and enumerative classification, qualities of a good notation, too, have not been ignored. Although treatment of topics is concise it is not superficial, and no important consideration has been left untouched.

This eminently readable book unfolds gradually. The chapters have been skilfully graded. Each chapter is divided into sections with feature headings. The book is well summarised in its last chapter. The topography is pleasing and inviting. Each concept is tangibly explained with ample illustrations drawn from various real classification systems and indexing systems both general and special, namely:

1. Classification of Machine Bolts.
2. CI/SfB Construction Indexing Manual.
3. London Classification of Business Studies.
4. Ranganathan's CC.
5. Classification for Office Organization.
6. Guildhall Classification for Local Material.
7. D.D.C.
8. L.C.C.
9. BC-1, BC-2.
10. D.D.C.
11. NATO Classification system.
12. London Education Classification
13. British Classification of Music.
14. BSI Root Thesaurus.
15. Thesaurofacet.
16. Chain Procedure.
17. PRECIS.

The examples cited are strikingly apt. Simplicity and clarity are its hallmarks. The author has the gift of presenting involved ideas in a simple language without erroding any meaning. The anologies are fresh and thought provoking and full of insights. For example, PRECIS is said to provide full statement of the subject – a kind of précis (p.93).

The book, primarily for the beginners or the non-professionals, provides a valuable opportunity to refresh oneself with the basis of library classification.

A list of 40 bibliographic items splitted into 67 references represents a cross section of the representative literature on the subject.

https://doi.org/10.5771/0943-7444-1989-1-49
Generiert durch IP '3.21.43.60', am 12.10.2024, 07:53:07.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.

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 accomodate 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 dicussion. 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 prectice 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-oeeurences 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.

https://doi.org/10.5771/0943-7444-1989-1-49
Generiert durch IP '3.21.43.60', am 12.10.2024, 07:53:07.
Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.