FIDEL, Raya: **Database Design for Information Retrieval: a Conceptual Approach.** New York: Wiley 1987. xv, 232p. ISBN 0-471-82786-X

Databases have become part and parcel of information business in recent years; moreover, they have almost become a common feature of privately based computing. Very much like system software, database management systems can be purchased off the shelf at tremendously attractive prices. Yet the brave new world of decreasing prices and rapidly improving technical performance appears to be but one aspect of database construction, leaving important questions to be answered by individual database designers themselves: what about the contents of the database to be designed; indeed, what about the concepts?

While a mass of literature has been devoted to technical matters of database design, Dr Fidel's book focuses explicitly on the conceptual approach which may easily be underrated if primary attention is drawn to technicalities. Her approach is to be understood as a general method applicable to the design of databases of any kind. Although distinguished internal, external, as well as conceptual levels of database design, Fidel only mentions the first and second in passing, if at all. Matters of entry format, indexes information display and so on are taken to be basically independent from conceptual ones. Fidel's objectives, then, are different. She sets out to demonstrate, in minute detail, the very first step among four major ones of database design which is crucial to every kind of conceptual approach: data requirements analysis. Subsequent steps, namely hardware and software considerations, coding (actual implementation of the design), and testing of performance are only mentioned in introducing the book.

The book, intended for novices and advanced readers alike, is written in a clear and jargon free style. It is well illustrated, referencing both to graphic displays as well as to the basic example which is repetitively drawn upon to demonstrate the various steps of an intricate process called conceptual design in a non-structural information environment. Most helpful, a glossary and an index are provided. Reading is not impaired by a mass of references. The example mentioned, the design of a database on restaurants, may be felt, at times, to divert some readers' interests from abstract conceptual considerations to rather more fundamental attractions. Yet before readers turn to recipes and conventionally published restaurant guides, Fidel manages to direct the attention back to more information related business.

Following the introduction, the process of database design is displayed in four stages, dealing with "study of the problem", "representing the data in formal terms", "selection of rules", and "evaluation" respectively. It is most welcome that from the very beginning the information needs of potential database users are taken into account. Elaborating her very telling example, Fidel distinguishes, first of all, two basic concepts of database design, namely "enterprise" (restaurants) and "environments" (possibly varying from administrative to culinary ones). As a consequence, a strong plea is made to start user-oriented database design by finding out about users' attitudes and likely information needs.

The recommended method is rooted in a thorough problem study: survey of environments and interviewing potential database users (considering, of course, varying environments), followed by a complete documentation of their statements. Although these statements include information on all levels of design, conceptual ones are most important for the present study. The interview data are then subjected to linguistic analysis (elimination of redundancy, ambiguity) and to interpretation of information provided only implicitly. If successful, all this will result in a data dictionary, listing precise definitions of all data required to describe the enterprise of the database, accompanied by an operations dictionary, comprising information on actions involved with these data. At this stage, environments are still to be kept separately.

In order to facilitate the transformation of that information into a database, a formal representation of the elements in the data dictionary is required. It has to display all entity types contained in the data dictionary, the relationships among them as well as the attributes found out necessary to describe both entities as well as relationships. Due to a surprisingly straightforward method, generating a clear mini-diagram from each data dictionary entry, complex structures can be built up step-by-step and vividly displayed in an "entity-relationship diagram" (as outlined by P.P.-S. Shen, "The Entity-Relationship Model: Towards a Unified View of Data", *ACM Transactions on Database Systems,* 1(1976)No.1, p.9-36). In its final version, such a diagram will include all environments specific to that database. As Fidel explains, integration of schemata will involve the resolution of conflicts arising from inconsistent names or incompatible or missing diagram structures. Modifications of formal structures must be met by correspondent alterations of the data dictionary.

Before embarking on data collection, however, some stringent rules have to be dealt with. These pertain to the mandators or optional character of relationships and attributes, their degrees of cardinality (e.g. one-to-one or one-to-many), specifications of attributes, the selection of authorized sources, dealing with fuzzy or borderline cases etc. Further to the rules, it is necessary to control the rigorous design and the quality of the database concept as far as that can be achieved before the database is actually in operation. The list of control features is long and quite likely to point to a somewhat cyclical character of database design, for all weaknesses outlined in that evaluation process prior to database implementation will unavoidably refer the designer back to the diagram and probably to the data dictionary to amend the conceptual scheme itself.                     Heiner Schnelling

Dr. H. Schnelling, Universitätsbibliothek
PF 5560, D-7750 Konstanz 1

THARP, A.L.: **File Organization and Processing.** New York: John Wiley 1988. ISBN 0-471-60521-2

The book by Alan L. Tharp covers a very interesting subject which – for some – may not be expected under this title. An alternate title might be "All about Signatures, Bloom Filters, Hashing and other magic representations of information you wanted to know but were afraid to

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. Schck, 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.