Book Reviews

DIDAY, E. et al: **Optimisation en Classification Automatique.** (Optimisation in Numerical Taxonomy). (In French)
Le Chesnay, FR: Inst. National de Recherche en Informatique et en Automatique 1979. 2 vols. 896 p.. 120ff. ISBN 20-7261-0219-0

This is a book on classification and clustering methods and their use in exploratory data analysis. The book is mathematically oriented and has little overlap with other monographs on cluster analysis. Consider a set E of objects i, each one described by some observation or data vector x_i. The problem is to subdivide this set E into a certain number of homogeneous groups or classes, i.e. such that the elements of each class will be as 'similar' as possible (and near to some suitable class representative) whilst different classes are most 'dissimilar' in some sense. Hereby 'similarity' resp. 'dissimilarity' of objects will be measured by some distance of the corresponding feature vectors. It is evident that the resulting classification of objects may be useful for information retrieval or documentation sciences, for organisational purposes in economics, for pattern recognition, psychology a.s.o. In particular, the resulting groups may be interpreted as 'types' or can serve for information (data) reduction.

Diday et al describe a lot of cluster analytic methods which were developed by a French research group at INRIA and which all look for those classifications (partitions) of E which are (nearly) optimal according to some given optimality criterion W. Most algorithms are modifications of the well-known k-means method and operate by iteratively determining a minimum-distance partition and correspondingly adapting the former class representatives ('dynamic clusters method'). The principle and the convergence of this method are established in § I. The following 24 chapters show – this is the true value of this book – that by suitably defining the notions 'distance' and 'class representative' the method may be used for many practical situations or optimality criteria: The usual quadratic distances may be modified or adaptively optimized (§§ 17, 18); preference and relational data (§ 11) or mixed and missing data may be handled (§ 7, 16); kernels, centroids, regression surfaces or principal component hyperplanes may be used, alternatively, as the class representatives (§§ 2, 4, 8, 13, 16); optimal scalings or discriminant spaces can be invoked (§§ 18, 9); multiple criteria problems and additional constraints (e.g. spatial, class size) are considered (§§ 2, 3, 20); the mixture problem and some continuous case theory is investigated in §§ 12, 16; contingency tables are treated in §§ 8, 19; hierarchies are relevant to § 22. Some applications concern real data problems from quality control, linguistics, picture processing, biology, and informatics (§§ 10, 14, 15, 23–25). Bibliographical notes and computer programs have been added, too.

However, no probabilistic results, no hints for choosing a special algorithm, and no techniques are given for graphical display, for interpreting the results, or assessing their relevance.

The book could be an excellent guide to the topics mentioned above and their development in France if its reading were not hindered by a huge number of misprints, a badly arranged lay-out and false or lacking reference possibilities (e.g. no index of topics). The mathematical rigour, the depth of argumentation and the notation is rather varying over the 25 chapters of the book due to the fact that 23 different authors contributed to this volume. Often the notation is obscure or only partially defined, ambiguous or heterogeneous (e.g. § 8 or p. 529 where, in the same line, for $k = i$ J_k and J_i mean different sets). Several chapters presuppose, without introducing the subject, results and notations from other French publications (e.g. on correspondence analysis). Substantial mathematical errors occur on page/line 185/16, 205/14, 339/14 and in the proofs on p. 373 and in § 18.3.7. A more concise and dense presentation would have been desirable.

However, the monograph will be useful for statisticians, data analysts and practitioners from many fields if they have a medium level of mathematical education, an elementary knowledge of cluster analysis and are willing to overcome the difficulties described above.

H.H. Bock

Address:
Prof. Dr. H.H. Bock
Institut f. Statistik u. Wirtschaftsmathematik
der RWTH Aachen
Wüllmerstr. 3, D-5100 Aachen

RICHMOND, Phyllis A.: **Introduction to PRECIS for North American Usage.** Littleton, CO: Libraries Unlimited 1981. 321 p., ISBN 0-87287-240-8

Das Buch beschränkt sich bei der Erklärung von PRECIS auf den englischsprachigen Bereich, insbesondere auf den Gebrauch in Nordamerika.

Es sei deshalb erwähnt, daß Versuche mit PRECIS auch in anderen Ländern mit anderen Sprachen durchgeführt werden. So wird PRECIS im französischen Sprachraum von verschiedenen Bibliotheken in Frankreich und der Bibliothèque Nationale du Québec erprobt.

Bei der Deutschen Bibliothek läuft derzeit ein einjähriges Projekt, in dem die Anwendung von PRECIS für die deutsche Sprache erprobt wird.

Dies sei vorangestellt, um zu verdeutlichen, daß das besprochene Buch nur einen Ausschnitt aus der PRECIS-Forschung darstellt.

Die Autorin schlägt vor, das Buch „Introduction to PRECIS for North American Usage" gemeinsam mit dem PRECIS-Handbuch von Derek Austin zu benutzen. Die von ihr angeführten Beispiele und Erläuterungen gehen oft über das Handbuch hinaus. Die zahlreichen und übersichtlichen Verweisungen der Autorin ermöglichen es dem Leser, mit beiden Werken gleichzeitig zu arbeiten und aus den wertvollen Ergänzungen der Autorin zum Handbuch großen Nutzen zu ziehen.

Der Schwerpunkt liegt hierbei auf den syntaktischen Fragen, doch wird auch soweit auf die Semantik eingegangen, wie es für das Verständnis des Verweisungssystems in PRECIS erforderlich ist.

https://doi.org/10.5771/0943-7444-1982-2-105-1 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Für Anwender von PRECIS dürfte der kurze Überblick über die englische Kasus-Grammatik besonders nützlich sein. Da die Erläuterungen auf den Gebrauch von PRECIS in Nordamerika ausgerichtet sind, werden in den Erläuterungen immer auch die Anglo-American Rules for Cataloguing berücksichtigt. Desweiteren geht die Autorin auch auf Unterschiede in der Terminologie des Nordamerikanischen gegenüber dem Englischen ein.

Diese Besonderheit schmälert aber keinesfalls den Nutzen für anderssprachige Benutzer, da es möglich sein dürfte, die Anweisungen auf andere Regelwerke zu übertragen.

Besonders hilfreich sind die Übersichten über typographische Codes und Term Codes, sowie die Zusammenstellung der neuen numerischen Differenzierungscodes in Matrixform.

Eine weitere Entscheidungshilfe bietet das Buch für die Ansetzung von geographischen Namen als Schlüsselsystem. Geographische Namen werden entweder als Lokalität mit Operator (0) oder als Schlüsselsystem mit Operator (1) kodiert. Die Zuordnung der Operatoren richtet sich nach der jeweiligen Funktion des geographischen Namens. So bezeichnet Operator (0) den rein geographischen Namen, während Operator (1) wirtschaftliche, soziale und politische Aspekte mit einbezieht.

Die eindeutige Abgrenzung beider Funktionen ist nicht immer einfach. Die Autorin stellt eine Liste von Begriffen zusammen, die eine wertvolle Entscheidungshilfe für die Kodierung des geographischen Namens als Schlüsselsystem (1) bieten.

Sehr gut gelungen ist die gründliche und ausführliche Darstellung der Prädikatstransformationen – eine ganz wesentliche Ergänzung zum PRECIS-Handbuch.

Insgesamt läßt sich sagen, daß „Introduction to PRECIS" eine empfehlenswerte Arbeitshilfe in Verbindung mit dem PRECIS-Handbuch ist.

Ingrid Schäfer-Link

Address:
Mrs. I. Schäfer-Link
Deutsche Bibliothek
Zeppelin-Allee 8, D-6000 Frankfurt

ROWLEY, Jennifer E.: **Abstracting and Indexing.** London: Bingley 1982. 155 p. = Outlines of Modern Librarianship. ISBN 0-85157-336-3

"A title indicates the subject content of a document." The title of this book does not, and thus proves the falsehood of the author's terse statement (p. 116), somewhat modified only three pages later, when the damage to the unsuspecting student has already been done. Although the book does indeed treat abstracting, it deals with indexing only in a limited sense, namely that of papers and articles for an indexing service (published or internal). The indexing of books, reports or periodical runs is not considered at all, and the names of almost all authors on indexing are therefore conspicuously absent from the index and bibliography. Only a fleeting reference to indexes of books says that they are "usually constructed by the author" although "a professional indexer may be employed" (p. 127), while periodical indexing is not even mentioned. Thus, students (for whom this book is specifically intended) may be left with the impression that indexing for A & I services is the only kind of indexing that exists, or that the principles of indexing for a current information service can be applied without any change to the indexing of monographic documents.

The first part of the book deals with abstracts, and is rather apt to confuse the novice by presenting no less than seven types of abstracts, among which are "indicative", "informative", and "indicative-informative" (the latter, in an example covering the same article, being just two lines longer than the merely "indicative" but otherwise pretty much the same). The difficult and intricate problems arising from the various forms and types of personal names are briefly listed in the chapter on "Bibliographic references", but not even one actual example is shown, nor is any practical advice given for the solution of such problems. The book's own index does not even have an entry for "name indexes" or "author indexes" which, as the author asserts in another passage, "present relatively few problems" (p. 46)! When we come to "Indexing", some other amazing statements can be found, e.g. "Homographs have the same spelling as each other" (where the grammer of the English language has been mangled), "All nouns have plural and singular form", (which will come as a surprise for linguists), and in the chapter on "Indexing languages" we are told that "A thesaurus summarizes an indexing language" (which is not quite true even for post-coordinate systems, much less for other types of indexing languages). In another chapter, the author says that "pre-coordinate indexing principles have also found *some applications* in subject indexes to library catalogues and the *shelf arrangement* of book stock" (p. 95) (my emphases). The first part of this sentence is an understatement, to put it mildly, since all catalogs based on subject headings have used pre-coordination for more than a century, and classified catalogs based on the Universal Decimal Classification rely on the same principle, while those based on Dewey use pre-coordination every time they "build" a number; the latter part of the sentence is simply nonsense.

Quite apart from such ill-considered and misleading statements, most topics are treated rather superficially, with the possible exception of the sections on pre-coordinate indexing, the one on the construction of a thesaurus, and the presentation of PRECIS which manages to cover the salient features of that system in a limited space.

The topics of "Editing and proofreading" are relegated to an Appendix, although they would evidently fit better into the earlier part of the book that deals with technical matters of abstracting and indexing. But then, the author thinks that these are rather ephemeral matters: "Checking and proofreading require . . . less knowledgeable staff [which] can satisfactorily complete these tasks", while admitting that "even the best abstractors and indexers may be subject to sloppy practices and grammatical indiscretions" (p. 143). Quite so. The proofreading of this book seems to have been left to badly trained chimpanzees, but some of the misprints may be due not so much to the alleged sloppiness of the typesetters but to the author's own ignorance of spelling (e.g. "superceded", p. 78, "preceed", p. 80), or semantics, as

https://doi.org/10.5771/0943-7444-1982-2-105-1 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY