related and resurfaced issue is reclassification. It was current in the mid 1950's when the DDC was being taken as a waning, even dead classification. Now it is becoming retrospective conversion in online catalogues. It has been dealt with in the aptly named chapter „Reclassification revisited: an automated approach" by Ling Yu W (Miko) PATTIE. It is a case study of the University of Kentucky library reclassification in an automated project named NOTIS: The chapter considers methods and planning aspects of reclassification.

The last section on Information Technology comprises two chapters: the first is on information technology vis-à-vis classification by Gertrude S. KOH, and the second is on Electronic Dewey by Ross TROTTER. Dr. Koh studied the use of classification in virtual union catalogues and in virtual libraries. Classification in online systems demonstrates workability and cost effectiveness. The recent emerging consensus is on combination of classication, free text and controlled vocabulary as a more powerful retrieal method. It is a comprehensive survey of the usc of classification in online subject searches. Ross Trotter critically but comprehensively describes the features of „Electronic Dewey" (EDDC) published in 1993 in CD-ROM form by Forest Press/OCLC. It heralds the electronic age of classification providing greater flexibility to access the data.

The book comes as a whiff of fresh wind across some old and mostly new array of wide-ranging issues. The topics are not only very pertinent but immensely practical, too, as one could expect from an American book on classification. It opens new vistas to classification studies and research. An optimistic outcome of the volume is that classification, its practice, study and research are equally valid in the days of global information networks and virtual libraries. The trend seems toward making classification more socially relevant and user-friendly than to be logical. The editor is successful in achieving his declared intentions of the „set" of papers „to encourage fresh and wider choice in library and bibliographic classification decisions, the extent of choice and 'best fit' of a system to local factors" (p.2).

All contributions are marked by indepth and well documented research. References given at the end of each chapter almost make a current bibliography on classification studies. At the beginning of each chapter an abstract is given and every chapter ends with a conclusion and a summary. Most of the authors are well known while others have amply justified their selection to set with outstanding names. It is one of the fundamental books exclusive to classification to originate from the United States. One can ignore it at one's own peril of lagging behind in classification studies.

M.P.Satija

1 Comaromi, J.P., Satija, M.P.: History of the Indianization of the DDC. Libri 35(19985)No.1, p.1-35

Dr.M.P.Satija, Guru Nanak Dev University, Head, Department of Library and Information Science, Amritsar-143005, India.

LEPSKY, Klaus: **Maschinelle Indexierung von Titelaufnahmen zur Verbesserung der sachlichen Erschließung in Online-Publikumskatalogen**. (Machine indexing of bibliographic data to improve subject analysis in OPACs). Köln: Greven 1994. V,121p., ISBN 3-77743-0572-2 = Kölner Arbeiten zum Bibliotheks-und Dokumentationswesen, 18

Online public access catalogues have been with us for a couple of years now, and we are getting used to public access to the files of our own library as well as of libraries in distant parts of the world. We are convinced that OPACs are the only reasonable way to pass on the benefits of library automation to library users. Yet a somewhat uneasy feeling prevails about the way OPACs should properly be designed to allow for maximum quality of subject searching. Taking a closer look at this issue, we may find that the uneasiness is rooted either in the design of the OPAC's user interface or in the way bibliographic data are being prepared for use in OPACs. More often than most librarians would be prepared to admit, it turns out that the uneasiness results from both the design as well as the data preparation.

As for the design issue, librarians should wonder if there is any reason to pursue the distinction between subject headings and keywords taken from book titles or other fields of bibliographic description. What may sound a clear-cut distinction in English, is more likely to become mixed up in the German language which only provides two very similar looking (and sounding) words for this antinomy: Stichwort (keyword) as opposed to Schlagwort (subject heading)., German librarians have traditionally been very eager to insist on this distinction, and have consequently been relying on it in the design of OPAC user interfaces. With the benefit of hindsight, surveying some ten years of OPAC design and use in German libraries, one may safely argue that this insistence must be criticized for at least three reasons: first, it is wrong to say that the complete stocks of libraries (public and academic alike) are comprehensively subject indexed, which is bound to lead to incomplete search results; second, it is misleading to argue that there is a significant difference between the information to be derived from the keywords on the one hand and the added subject heading(s) on the other, that is to say the subject heading(s) are quite often merely redundant to the key words; third, what ever benefits of subject headings librarians may have expected for information retrieval, most library users simply ignored the eagerly pursued distinction, or, what would be nearer to the truth, could not cope with it. Which, in turn, raises the question why libraries proceed spending considerable staff resources in subject indexing presumed fit for their OPACs, which, after all, is hardly appreciated by their users. To adopt a bit more positive thinking instead, why do libraries not - as yet - devote more efforts to explore the sources of information inherent in keywords of book titles and other related fields, painstakingly recorded in the process of descriptive cataloguing, for subject retrieval?

All positive thinking notwithstanding, this is, of course, quite a sensitive issue. It touches on the self-respect of subject librarians. What is needed, then, is a study that takes an

unprejudiced approach towards this issue. What is more, the study in question should provide some empirical evidence on the feasibility of keyword indexing for subject searchiung in OPACs. The author of the present book, slim as it may be, deserves our gratitude for having met both requirements.

The book is the published version of a study undertaken in 1993 in the course of qualification for academic library service. It is based on empirical tests on bibliographic data from the University Library of Düsseldorf and software products from Softex GmbH, Saarbrücken. The study combines the theoretical discussion of controversial issues with empirical evidence which may pave the way to a solution working in terms both of subject retrieval as part of OPAC use and, in due course, of subject indexing as part of library management.

What the author has in mind can aptly be illustrated by refering to the bibliographic as well as subject catalog data of his own book. Compare its title *Maschinelle Indexierung von Titelaufnahmen zur Verbesserung der sachlichen Erschließung in Online-Publikumskatalogen* to the subject heading to be found in the *Deutsche Nationalbibliographie*: „Automatische Indexierung", „Alphabetische Katalogisierung", „Inhaltserschließung", „On-line-Katalog". I cannot help feeling that the subject headings add very little to the information to be gained from the keywords themselves. The time spent in defining the subject headings for this particular book was apparently spent in vain; for the results of subject indexing in this case obviously do not contribute to any real improvement of information content that could be used in retrieval. So why not rely merely on keywords?

Of course, it isn't all that simple. It would be quite a different matter if the book in question were written in a language other than German. In that case, only subject headings in German would make this book accessible during retrieval in a German-based OPAC search. And, to mention another complicating matter only in passing, subject headings would be urgently required if the keywords were not as telling as in Lepsky's example: Titles such as the *Wheel of Fire* or *The Imperial Theme* certainly need subject heeadings to be identified immediately as books on Shakespearean tragedy.

These very brief remarks point towards the issues that are at stake if a more positive attitude is to be taken at the process of machine-based indexing of keywords for improving subject retrieval. The keynote is 'machine-based indexing', for the strategy Lepsky suggests has nothing in common with the awkward examples of the „keyword-in-context" type that contributed very much to the bad reputation of machine-based as opposed to intellectural indexing. while KWIC (or KWOC, for that matter) produced endless lists of limited information value, Lepsky has sophisticated linguistic programming at his disposal to provide a structure to the sundry variety of keywords in English, French, and German. Moreover, this structure includes not only keywords, but subject headings as well. And to provide a kind of finishing touch, the structure even provides translations.

But we should not hurry things. Lepsky's book divides in three main parts. His first step is a concise description of the dificiencies of subject indexing intended for OPACs and the frailties inherent in OPAC design based on the misleading retrieval approaches offered by common OPACs. Although mainly focused on the German situation, Lepsky refers also to other countries. As to Lepsky, there are two main problems in OPAC design and user: a high percentage of searches does not produce any results; too many searches produce quite an impressive number of titles, but the number of titles actually relevant to the user's query is considerably low (the remarkable 'recall' is only rarely matched by 'prccision'). It might be the librarians' fault, who may have expected the user unfamiliar with library conventions and terminology (remember *Stichwort* vs. *Schlagwort*) to walk in their path.

Lepsky then continues by pointing out the need for linguistic programming to cope with the morpholocial and semantic varieties of keywords. Methods that have proved successful to transform keywords into search terms include elimination of stop words, orthographic checking, automatic truncation, decomposition, derivation. Lepsky introduces the program IDX, which was masterminded by H.H.Zimmermann[1]. In contrast to several related programs designed mainly for analyzing texts from a particular discipline (say, Law or physics), IDX is intended for application on texts - such as bibliographic descriptions - from any discipline. This is an eminent precondition to the implementation of such a programme in the context of an OPAC accessing the large stocks of a university library. The results of Lepsky's study indicate that IDX passed this test.

The IDX software is a remarkably functional tool. It is based on comprehensive English, French, and German dictionaries. The methodical approach is defining word relations, that is to say, structuring the various dictionary items according to a set of word relation types. Among these we find, to name but a few, synonym, antonym, compound/part, derivate, sec also, orthographic variants, truncation. IDX works sequentially by lemmatization (identifying basic word wfoems), elimination of stop words, decomposition of compounds (such as „Bildschirmarbeitsplatz", i.e. terminal or PC workstation), establishing of word relations according to the grid mentioned, identification of multi-word word items (such as 'machine-based translation!), translation of individual words. As already mentioned, IDX is reported to be applicable not only on German texts, but English and French ones too. The advantages of IDX, as compared to other indexing software, clearly are to be seen in the handling of compound forms and the translation facility.

About half of Lepsky's book is devoted to the third main part. It describes a feasibility study of IDX, undertaken at the University Library of Düsseldorf. The OPAC of this library has been in operation since 1987, offering subject retrieval on the basis of a 'basic index' comprising keywords as well as subject headings. The proportion of titles actually subject indexed is estimated at about 35%. As can be expected in a university library, there is a high proportion of foreign language titles (40%). It would, of course, be ideal to have the whole stock properly subject indexed, but this option can be ruled out for obvious staff reasons. This is all the more true if a library, such as Düsseldorf University Library, is engaged

in retrospective conversion of conventionally catalogued titles.

The only improvement, within acceptable staff and financially conditions, may indeed be expected if one is prepared to have a new look at the amorphous mass of keywords contained in bibliographic descriptions. Lepsky investigates the feasibility of improving subject access facilities by applying IDX to parts of the keyword material (from titles proper) and subject headings, if available. His study is based on a sample of some 12.000 titles, taken from all disciplines. 57% were titles in German, 28% in English, 4% in French, 11% in other languages; only 39% were subject-indexed. Lepsky describes all the steps of IDX indexing indicated above, leaving the reader with the impression that the indexed parts of bibliographic descriptions are indeed being subjected to a thorough indexing. A good example would be results of compound analysis. Compounds abound in German, in both everyday speech as well as academic discourse; quite consequently, the current rules for verbal subject indexing (*Regeln fiir den Schlagwortkatalog*) encourage the formation of compounds. The examples Lepsky quotes (admittedly, only a few) clearly show that the IDX dictionaries are comprehensive to allow for decomposing compounds into meaningful parts, while avoiding irrelevant stuff in most cases (for example, „Agrarmarktpoliitik" is decomposed into „Agrarmarkt", „Politik", „Marktpolitik", „Markt"). It is interesting to see that the IDX indexing does not lead to an uncontrollable mass of terms. Quite surprisingly, out of 80.000 words contained in just under 8.000 German titles not more than 93.000 index entries are produced. To give an impression of the translating results, Lepsky estimates that the 21.000 words contained in the English titles (the complete sample) were refered to 62.000 German translations. As said before, indexing of this kind is machine-based. It certainly needs intellectual support. This refers to the constant task of reviewing the dictionaries to eliminate irregular terms or nonsense, or to add terms that were not produced by the IDX routines and include the relations into the dictionaries, or to correct translation.

It must be noted that Lepsky's study is not limited to a minute description of the indexing process and results, but also outlines the consequences of large scale implementation of IDX for library administrtion. Lepsky claims the staff requirements resulting from the integration of IDX into the subject cataloguing routines of a library to be not more than about 20 working hours per 21.000 titles. If this estimate were confirmed, it would be quite disturbing for library managers (and the parent bodies of libraries) to learn that considerable staff effort devoted to subject cataloguing might, in the long run, be saved if subject indexing were left to the machine.

In 1994, a research project was launched at the University Library of Düsseldorf. Supported by *the Deutsche Forschungsgemeinschaft*, it is named MILOS („Maschinelle Indexierung zur verbesserten Literaturerschließung in Online-Publikumskatalogen"), and it will be very interesting to see the results this research project will produce[2].

Heiner Schnelling

1 For a short description of IDX see: Thönssen, K.: Automatische Indexierung und Schnittstellen zu Thesauri. Nachrichten für Dokumentation 39(1988)p.227-230

2 For a retrieval test, undertaken in the course of this project and based on 40.000 machine-indexed titles taken from all disciplines, see: Lepsky, K.; Siepmann, J.; Zimmermann, A.: Automatische Indexierung für Online-Kataloge. Zeitschrift f. Bibliothekswesen u. Bibliographie 43(1996)p.47-56

Address: Dr. H. Schnelling, Ltd. Bibliotheksdirektor, Universitätsbibliothek Gießen, D-35386 Gießen, Germany.

[1] Taylor, Robert S. 1986. *Value-added Processes in Information Systems*. Norwood, New Jersey: Ablex Publishing Corporation.