An interesting paper on a database for food science and technology from Robert Belts of the Leatherhead Food Research Association made the point, which indexers find obvious, that abstracts are best written so as to contain the indexing terms that searchers will be using. (The firm avoids author-abstracts, preferring, despite the extra cost, to write their own so as to slant them towards their users as well as making them rich in index-terms.) To aid the inexperienced searcher, LFRA now have four thesauri which can be switched in to ensure that information is not lost through having used the ‘wrong’ search terms. (It continually amazes me that text retrieval people have only recently cottoned on to this problem, especially as we have had by now some three decades of experiencing online retrieval failures to spur systems-providers on to better efforts. There seems to have been an inertia effect in that searchers will not enquire too deeply into how many more they might find as naive users when it comes to information retrieval. Many of them still think that if you index every word in a text there is no further need to worry . . .)

Kevin Jones attended the rest of the conference

The latest Institute of Information Scientists’ Text Retrieval conference was a thoroughly worthwhile event and most enjoyable. The speakers were all better than competent. The New Cavendish Centre was both central and well-run. Only a large audience was lacking: this sad feature will be returned to subsequently.

Two of the papers failed to reach the high overall standard and were merely sales pitches. One introduced the Sony Electronic Book which exploits Sony’s small-sized compact disc in combination with CD-ROM technology to encapsulate reference-book type material. Currently screen resolution is fairly poor, but colour is promised eventually. Searching is menu-driven. At present it would appear to be little more than a gizmo, its small screen probably more suited to the highly compressed nature of Japanese, rather than to Western texts. The other commercial plug was for The Times on CD-ROM. This appeared an inappropriate venue for such a presentation. Furthermore, the unsuitability of the talk was emphasized by its snide asides on the unfortunate indiscretions of one British politician.

Several papers discussed multimedia systems and
access to graphical information, and this was the theme of Tony Cawkell’s keynote address which gave an excellent, sometimes disturbing, overview of the associated problems. In part the solutions being sought are technology-driven. Some are associated with forefront computational techniques, such as pattern matching and neural networks. Others appear to hark back to methods which were advocated in the early 1960s, combining microfilm technology with that of the power-sorted punched card. Kodak’s system for storing and indexing photographic images on CD-ROM has a disarming similarity to techniques such as Filmorex. On the other hand, ICONOCLASS returns to traditional classificatory techniques to provide in-depth access to fine art works. The classification employs ‘conceptual’ techniques: for instance, 31A: parts of the body. The system is highly hierarchical. It is also labour-intensive and difficult to learn, but does appear to provide the required access. It is employed at the Courtauld Institute and in about eight other major art collections.

In complete contrast to the sophistication of ICONCLASS, Professor Peter Enser described the crude empirical techniques adopted at the vast Hulton-Deutsch picture collection. Pictures are stored under very broad categories. Nevertheless, the skill of the eight picture researchers enables the collection to meet the particular demands set by its clients both for highly specific items and for very general material, much of which is needed to convey a certain mood. Enser and his co-workers had sought to establish how the requests were categorized. There was general agreement that mechanized techniques were not yet appropriate in such a situation.

John Mallett of Sentana gave a lively presentation on database design for multimedia systems. There was a powerful demonstration that icons do not have to be visual: they can be aural (summoned by bells), or verbal/textual as well. Unfortunately, this did not lead to a deeper discussion on the basic requirements for icons within a computational context.

The core papers considered automatic indexing and related areas. The two major systems discussed were SIMPR, a European Community project, and TELMI, which is being developed for the French Telecommunications Minitel service. These share much in common. Both have large lexicons containing around 60,000 items. Both employ natural language processing techniques. The respective presenters were Forbes Gibb from the University of Strathclyde, and Edmond Lassalle. Their comments were reinforced by Antoinette Renouf from the University of Birmingham, who showed how linguistics has an important role to play in the development of advanced text-retrieval systems. This was illustrated by comments on the fairly obvious topics of homography and synonymy (but her comments on the latter were illuminating) and the less obvious problem of anaphoric reference which bedevils any quasi-statistical approach.

Professor Peter Willett described the application of computerized spelling correction techniques to historical texts as an attempt to make such texts simpler to search. The most successful technique, n-gram matching, is computationally simple to implement and greatly enhanced access to a variety of texts which had been created prior to the implementation of standardized spelling.

The problems of navigation in large hypertext systems were examined by Chris Reynolds who did not blanch from citing a comment by Catherine Hamilton that hypertext costs grow exponentially with size. Thus the contribution was characterized by a suitable lack of hype, and an accurate assessment of the difficulties in implementing hypertext. There was also an interesting observation that even taxonomic literature may be subject to difficulties in automatic manipulation: for instance, a text may be all-about something, but this is evident only from the heading. Reynolds was an illuminating speaker, and contributor to the discussion sessions.

Online Public Access Catalogues (OPACS) were examined by Shirley Anne Cousins from the University of Wales. The main problem is their origin in library card catalogues which provide insufficient subject information for satisfactory online searching.

The most surprising contribution of Graham Cornish from the British Library was his clerical dress: he is an Anglican priest. He gave an excellent overview of the impending problems of European Community legislation on copyright, especially as it may relate to databases. Perhaps his attire was a timely reminder that copyright is a moral/ethical issue. Two papers were missed as they were organized during a parallel session: they related to the special information requirements for emergency services.

The big debate:
Is the thesaurus redundant?

The conference ended with a ‘big debate’, the issue being: ‘This house believes that free-text searching has rendered the thesaurus redundant’. This was proposed by Professor Charles Oppenheim of Strathclyde University, probably the most colourful member of the British information profession, and opposed by Ann O’Brien from the University of Wales. The main problem is their origin in library card catalogues which provide insufficient subject information for satisfactory online searching.

The conference ended with a ‘big debate’, the issue being: ‘This house believes that free-text searching has rendered the thesaurus redundant’. This was proposed by Professor Charles Oppenheim of Strathclyde University, probably the most colourful member of the British information profession, and opposed by Ann O’Brien from the University of Wales. The main problem is their origin in library card catalogues which provide insufficient subject information for satisfactory online searching.

The Indexer Vol. 18 No. 3 April 1993
CONFERENCES AND SEMINARS

Ann O’Brien opposes Loughborough University. Charles appealed to the research performed (a long time ago) at Cranfield and Aberystwyth which had demonstrated the futility of controlled language for information retrieval. Unfortunately, the delegates were well aware of the current work on advanced information systems which requires the compilation of large lexicons (which might be called thesauri). Charles’ contribution was brilliant and enjoyable: nevertheless he lost the debate.

In part this failure on Charles’ part may characterize one of the reasons for the failure to attract a worthy number of delegates: most of the speakers arrived for their session and left after it. This was why Chris Reynolds was so important: speaker/contributors enrich the general ambience, and if they fail to be involved with the whole conference, the general standard suffers. The venue was excellent: the management even increased the heating when there were complaints about coldness on the second day. Recent Text Retrieval conferences, as assessed by their proceedings, have also attracted good speakers. Three days is probably too long and thus expensive to attract a large number of delegates.

KEVIN JONES

Copyright and the indexer

12 November 1992, London; Society of Indexers

Do indexers own the copyright of their indexes? If so, what are the implications? A meeting attended by several publishers as well as SI members was held to examine these much-disputed questions.

Elisabeth Ingham, a freelance indexer of legal texts, gave her personal views of indexers’ moral rights and the effect on them of the 1988 Copyright Act, which she welcomed as benefiting both indexers and publishers. She had consulted a solicitor regarding indexers’ entitlement to authors’ moral rights over their work, and been assured that in his opinion there was no doubt that an index qualifies as a literary work, and the indexer thus as the author, as the person who creates it. First ownership of copyright in the index is therefore vested in the indexer.

Indexers who regard themselves as the authors of their works cherish the integrity of their creation and are anxious to keep control of what happens to it, protecting it from the tampering of junior copyeditors—‘index abusers’, Ingham designated them. She felt strongly about the subsequent career of her indexes, after they had been published, she had been paid, and her moral rights well asserted. Revised or updated editions, or a paperback following a hardback, may require revision of the index; and if revision is to be done, she insisted, then morally it must be the original indexer who is first asked to do the work. Indexing is a subjective craft, and only the original indexer will know the reasoning underlying the index structure. Probably they may also have the index available on disk, with their own program, easy to work further on. Elisabeth Ingham addressed fellow-indexers in the audience with some passion: ‘If indexers want any professional respect, they should never dabble in other people’s indexes without permission. Please, very kindly, leave my indexes alone, and I won’t touch yours’.

Of course, if the original indexer is for any reason unable to undertake the revision, then another indexer must be sought; a publisher could put a time limit on the acceptance of the work.

Ingham introduced the next speaker as representing ‘a publishing house where I feel valued; I thoroughly commend them’. This was law publishers Sweet & Maxwell (S&M), whose Publishing Director, Carol Tullo, gave their views of indexing and indexers. These were music to the ears of SI members in the audience: S&M do not dispute that the author retains the copyright of the work; know they need the professional skills of indexers—fulltime ones, since their projects are major ones; have ‘a voracious appetite for indexers’, indeed, and had ‘put a lot of thought into how they handle indexing arrangements’; and hold it axiomatic to return to the original indexer for revision of an index—‘why lose a good indexer’? This necessitates keeping careful, detailed records of work on all publications.

Tullo thought the value of S&M’s products was improved by combining a diversity of skills. She regarded the text-authors as most unsuitable to index their works, as they would lack the necessary impartiality, assume specialist knowledge on the part of all readers, and provide too thin indexes. They could not edit their own work, either; a different pair of eyes was needed. It was unfortunate that most contracts made the authors of the texts...
responsible for provision, or payment, for the index; they might then decide to redo it, and tended to be dismissive of indexes. Carol Tullo declared the index to be the proper responsibility of the publisher.

However, administration was simpler if rights were waived, with so many different inputs into a single work. Many contributors were willing to assign their copyright to the text author, or the publisher, or the packaging house, to make a single copyright-holding package. With a proper professional relationship, as part of the team contributing to the work, with trust, respect and loyalty, elements of control were retained; otherwise, there could be real difficulties. If the publishing house had not obtained copyright, then for instance, the death of the holder could lead to the rights passing to an inappropriate estate.

The sad truth was acknowledged that there are duff indexes and duff indexers. When found, make a note of, and do not use again. It was very sad to have the general standard of the work lowered by a bad index. Revision in such cases was usually not practicable; it was better, claimed Dick Greener of S&M, to scrap the rotten index (though duly paying the indexer for the work done), and start again from scratch.

In this wicked world, publishers also might transgress. Some indexers were paid by cheques with receipts attached, to be signed and returned—with copyright assignment in the small print above the signature. The chairman, Elizabeth Wallis, suggested that the respect for indexing, and awareness of the mechanics and the commercial value of indexes shown by Sweet & Maxwell were all too rare among their colleagues in the publishing world.

In discussion, the usual spectres of electronic publishing control and database manipulation were raised, and as usual shuddered at but not settled.

Hazel K. Bell

Thesaurus creator

Automatic verification of index terms in thesauri is what Term Manager claims to offer when it is run simultaneously with Cardbox-Plus. Both are available from Business Simulations Ltd, who point out, 'A database system is only as good as the way it is indexed'. Whole hierarchies, as well as individual words, can be entered in a full-screen template, or an existing thesaurus can be imported from an ASCII text file. Terms and the links between them can be edited online to show synonyms, broader and narrower terms, and preferred and non-preferred ones. The software costs £250. (from Information World Review, July-August 1992.)

Is it an index? Is it a spoof?

We reproduce below, exactly sic, the triple-column, single-page (on p. 339) index to Landscape painted with tea by Milorad Pavić, translated from the Serbo-Croatian by Christina Pribićević-Zorić, published by Hamish Hamilton, 1991. From the pen of Pavić came also Dictionary of the Khazars: a lexicon novel in 100,000 words, the 'extravagant fiction' with 'mechanisms of organizing historical knowledge . . . used in a spoofing way' reviewed in The Indexer 17 (1) April 1990, page 75.

Index

(For practical reasons, all the words in this index are given in the form in which they are used in the book.)

a. passim  
And. p. 178  
and, passim  
appear, p. 176  
are, passim  
armschair, p. 251  
as, passim  
at, passim  
book, p. 272  
bucket, p. 160  
caught, p. 154  
companion, p. 205  
completely, p. 257  
desk, p. 115  
does, passim  
desk. p. 115  
face, p. 169  
full moon, p. 157  
game, p. 195  
gripping, p. 280  
her, passim  
holding, p. 219  
in, passim  
knife, p. 201  
leaded over, p. 162  
lines, p. 131  
mother, p. 280  
murderer, p. 248  
of, passim  
or, passim  
out, passim  
pencil, p. 282  
reading, p. 225  
safe, p. 262  
saw, p. 164  
sec, p. 167  
sh, passim  
sit, p. 111  
spoon, p. 198  
that, passim  
the, passim  
Then, p. 334  
these, passim  
thinking, p. 230  
this, passim  
to, passim  
upside down, p. 164  
Vitacha Razin, p. 151  
water, p. 176  
what, passim  
who, passim  
would, passim  
you, passim  
your, passim

There's a place for us

The Online Journal of Clinical Trials, pioneer electronically-published journal, was reported at the 1992 meeting of the American Society for Information Science as being not inundated with manuscripts, and finding print indexes and abstracts important means of publicizing its contents.