Text management and indexing

1 April 1992, London; British Computer Society Electronic Publishing Specialist Group/Society of Indexers

Genevieve Hibbs' subtitle to her paper, 'Good indexing: makes searching so easy even an adult can do it', deserved to place her in pole position at this joint BCS EPG/SI meeting. Hibbs argued that it is important to standardize the vocabulary within texts, as this both makes the text easier to understand and leads to improved indexes. She also observed that good indexing should display the key themes within the text. Through the introduction of the neologism, 'asnegist' (as negative existence), Hibbs suggested that humans have a serious problem of overlooking obvious sources of information: a human failing that manifests itself in manually produced indexes, although Hibbs failed to make this important connection. The paper led to a considerable amount of discussion which included the doubtful value of thesauri within word-processing packages, the desirability for the automatic display of cross-references in retrieval systems, and the highly structured nature of newspaper indexes (due to the need to link stories which may extend over a lengthy time span).

Professional indexers were represented by Richard Raper and by Hilary and Drusilla Calvert. The first began with a number of illustrations from how not to through to how to in terms of physical presentation. The how not to was illustrated by a reference work to ballet which had vast blocks of page-references without subheadings. Raper argued that indexing is a mapping process which enables the searcher to move from the known to the unknown. Concepts need to be expressed in phrases of two to three words. Raper noted that it was virtually impossible to produce a good index if the publisher set unrealistically severe restrictions to length. The one delegate who appeared to have a background in hypertext was particularly eager to establish a measure for the density of indexing.

Drusilla and Hilary Calvert, wife and husband team of indexer and programmer, both described their MACREX system and outlined the primary criteria (clarity, concision, comprehensiveness and consistency) for good indexes. MACREX had grown over ten years from a simple system running on a do-it-yourself-kit computer to being a product able to support a wide range of options (many of which reflect the specialist needs of their clients). Drusilla thought indexers 'witty people': a remark well deserved in their delightful double-act presentation.

The Oxford Text Archive is a depository for electronic texts, the bulk of which appear to relate to the major classical literary texts. The aim of the archive is to conserve effort and materials, to enable research to be repeated and resources to be integrated. The policies have to be ad hoc, and there was a lack of standardization and bibliographical control, which the Proud Report from the British Library's Research & Development Department sought to improve. The achievements and problems of this key academic resource were ably discussed by Alan Morrison. Unfortunately, the discussion was barren as it concentrated on the mainly non-existent problems of copyright which resembled that of a group of physicians being presented with a folk cure for warts.

The remaining papers, by Ian Kempster of Promco Monitor Ltd and by David Crystal, were the most radical presentations as they both introduced the notion of the database publication in which the source text is housed in a database or retrieval system and is output either at regular intervals or when needed. The former paper described the application of a relational database management system which also incorporates an inverted file. This software is being used to maintain the highly dynamic Guinness Book of Records. It appears to have two key advantages over traditional publishing methods: final publication is relatively simple as the text is closely monitored at the input stage, and the editors have rapid access to their copy. Simple Americanization is also attainable. The indexing system employs a fierce stoplist including 'can'. Ever since, one has contemplated the fate of the biggest can of beans, or even the longest running can-can.

David Crystal has employed INMAGIC software to house the database for the Cambridge Encyclopedia. The ease with which material could be found and up-dated was its main advantage. It was also possible to offer purchasers the ability to have searches performed on the database. This had led to some surprises as everyman (or woman or child) tended to seek information on broad themes like single-parent families or cruelty to children. An enquiry on yawning had defeated the system. There is a need for a thesaurus for the encyclopedia, and possibly for cross-references of the type: Greeks see also Romans. David Crystal must be a good indexer as he is very witty: how, he asked, could the machine produce such marvels as the deadness of Beethoven (sic) or the Chanel (sic) Tunnel?

The meeting was very well organized and was both informative and enjoyable.

Kevin Jones
The Council of the Association of Jewish Libraries unanimously approved at its Midwinter Meeting, held in New York, 9 Dec., 1991, the following resolution on unindexed Judaica facsimile collections, prepared by Bella Hass Weinberg:

The Association of Jewish Libraries demonstrates its concern for the quality of Judaica reference books by administering two awards in this field: a bibliography award and an award for other categories of reference works. A major criterion in the evaluation of candidates for these awards is the quality of indexes.

The Association also shows its interest in the publication of high-quality information sources through its active membership in the National Information Standards Organization. One of the important standards issued by this organization is Basic criteria for indexes.

The Council of the Association has become aware that major Judaica facsimile collections and anthologies of journal articles are being issued, without indexes, by commercial publishers. These publications draw on the collections of American Judaica libraries and are expensively priced, but the publisher has added little value to them by merely reproducing the materials without providing indexes to facilitate access to the documents.

These facsimile reprint collections have no value as reference tools because their classified arrangement is not predictable by the user, and a very-time consuming search is required to identify a particular document or article.

Judaica libraries are the primary consumers of such published collections, and we urge publishers in the field of Jewish Studies to adhere to information standards, in particular through the inclusion of professionally compiled indexes in reference books and facsimile/reprint collections.

I must say . . .

At a seminar of the United Kingdom Serials group in June 1992 it was suggested that there are coming to be 'keywords of the month' favourably selected by abstracting and indexing services, which lead in turn to the deliberate insertion of these terms by authors into their articles, so that these will then be optimally accessed and distributed.