TEXTUAL DATA BASE DESIGN AND THE AUTOMATED OFFICE CONCEPT

Sally Grande
Information and Computer Services
Shell Canada Ltd.,
Head Office
Box 400, Terminal "A"
Toronto, Ontario
M5W 1E1

ABSTRACT

The storage and design requirements unique to the sophisticated handling of textual material have led data processing departments to consider special data base management packages. Along with this initial investment, large organizations have begun to explore integration of existing office equipment and internal systems with commercial data bases, electronic mail and personal files. The implementation of textual data bases in such environments also demands a long-term commitment to user support, to the unusually rapid growth rate of text files and to supporting staff with expertise in systematic information gathering and organization.

LA BUREAUTIQUE ET LA CONCEPTION DES BANQUES DE DOCUMENTS

RESUME

Les services informatiques des entreprises ont dû considérer de nouveaux systèmes de gestion de banques de données devant les contraintes uniques de stockage et de conceptions propres au traitement spécialisé des textes. Suite à cet investissement initial, les grandes entreprises se sont penchées sur l'intégration de l'équipement de bureau et des systèmes informatiques déjà en place avec les banques de données commerciales, le courrier électronique et la gestion des dossiers personnels. La mise en place de banques de documents dans un tel contexte demande un engagement à long terme pour assurer les services aux usagers, pour faire face à la croissance particulièrement rapide des fichiers de textes et pour apporter un soutien au personnel qui a acquis l'expérience dans l'organisation et la cueillette systématique de l'information.

TEXT DATA BASE MANAGEMENT AND THE INTEGRATED OFFICE

Sally Grande

The "automated office concept" represents the vision of a work environment where people can exchange, create, digest and disseminate information using a variety of electronic media.

Since much of any contemporary office worker's time is spent attempting to retrieve, create, distribute and digest text, it would appear that the text handling techniques long used and approved by the information industry for large, commercial data bases would fit naturally.

The fact is, however, that the traditional focus on text within office automation has not been retrieval but text-creation oriented. Integrated office planners have typically focused on word processing and electronic mail as means of text creation and distribution. Less well defined are the areas of text storage and retrieval for the large central files (or distributed data bases) that undoubtedly will emerge.

Office technology gurus have already identified five activities or features which can be considered as text oriented:[1]

- electronic mail
- word processing
- document storage and retrieval
- calendar management
- personal data bases

A few consultants have added to this list:[2]

- central file management by text DBMS
- access to remote, commercial data bases of published material

The Diebold Group has taken these fundamental realizations to the extent of recommending that an information specialist play an integral role in the future office to assist in the use of such external (as well as internal) text-based systems.

While there seems to be some awareness of the large research data bases, this awareness has not induced office systems planners

to explore the text handling procedures and problems long ago documented by the pioneers of such archival systems. Perhaps this lack of attraction deserves more attention.

When working with large textual data bases, it becomes clear that: the subject content of a document is distinguished emphatically from the "non-subject" related content. This amounts to making a fundamental distinction between when the item is "about" and for example, who wrote it or the date on which it was composed which, although valuable access points, have no direct relationship to the subject nature of the page of text at hand.

The key thought here is that any page of text contains values which can be fitted into certain slots or "fields". In office information systems[3] this concept is applied to the management of business forms. The activity of creating or using forms is viewed by office planners as an isolated activity having little or no bearing on the general text management.

In addition to emphasis on field structure and content, text data base specialists concentrate on the index or "directory" to the data and the control of a portion of the text so as to insure that some degree of retrieval precision is built into the maintenance of the data base.

In seeking to understanding the history contributing to the current state of affairs, one realizes that the subject approach has always caused administrative problems in the office environment. Thus, large files of business documents (correspondence, memos, working papers, etc.) began to accumulate and be organized according to alternative schemes.

Admittedly, the heading under which a document is filed in a paper file represents a single, and possibly least obvious, access point. Existing electronic mail systems, with few exceptions, have been designed to accommodate only one subject heading and typically organize documents according to their function in the workplace. The established online data bases, however, advocate the subject approach as central with a variety of peripheral, and simultaneous access points as aids to precise retrieval

Since online file development, both for central records management and personal data bases, is featured in most future office projects, the problems created by the lack of a reliable subject approach will manifest themselves after a brief period of incubation.

After all, office automation sprang forth from a union between data processing and office management aided by the persistent matchmaking of word processing. Neither group has ever embraced the document handling practices widely utilized in libraries. Since the early days of automated photocomposition, it was the

library/research market that envisioned the online data base as a natural extension of the subject sensitive card catalogue and alerting services.

From the viewpoint of the information specialist, searching data bases of text (especially inverted free text) with a subject approach combines the challenge of a crossword puzzle and the creativity of word-association. Even in the most well indexed data bases, let alone those with little or no built-in subject consistency, there are retrieval problems posed by language, its ambiguities and its dependence on context for meaning. As long as there is more than one way to express an idea and as long as English maintains its plasticity, designing and searching text effectively will remain a specialized and often frustrating endeavour.

It is at this point that, technically speaking, the problems go beyond hardware and software. It is here that so many office systems share with text DBMS a bridge between technology and user behaviour. Here the efficiency of the software is inextricably bound to the content. Perhaps now, more than at any previous point, data managers cannot afford to compromise software integrity by delegating text data base design and administration to those unfamiliar with indexing, thesauri and online searching of bibliographic and full text files.

In this light, where software integrity and quality of content merge, like medium and message, office system suppliers and designers should openly encourage more formatting and subject control of text destined for retention in online central files.

Shell Canada Limited is one of the few companies in Canada which maintains a major text management software package with the objective of providing an in-house, online file service for large text applications. This service, undertaken by the Office Systems Development Department can be integrated with other office tools currently under evaluation.

The purchase of Inquire, a product of Infodata Inc., and Shell Canada's subsequent leap into data base design have led Shell managers to tackle many of the issues raised in this paper.

Although Inquire became operational as a pilot only last fall, several objectives set forth by the integrated office have already begun to take shape. Perhaps most obvious and expected, there appeared to be an immediate centralizing of records and library applications bringing together departments which had not previously shared much assocation. Another objective constantly reiterated in the office automation literature is remedying the fragmentation of the workplace. The applications of several departments seeking to use Inquire will be subjected to rigorous evaluation and the data structured for Inquire according to standards set by the industry wherever a parallel exists. The candidate Inquire users have

discovered new worlds of vocabulary monitoring, authority files and procedures through which their document creation will be more streamlined and accurate from the onset. In addition, negotiations are taking place for the installation and maintenance of a company-wide subject thesaurus that may some day be used to index the daily output of word processing and I/O from electronic mail by each department with a need for a multi-disciplinary but business-oriented vocabulary.

While these events are localized and have only affected those within Shell who have already recognized a serious dependence on quiet and accurate access to information, they represent a beginning.

BIBLIOGRAPHY

1. Office Technology Research Group, Office of the Future, January 1981. A set of twelve binders.

2. The Diebold Group. "Tools Profiles Series: Data bases and filing tools." Document No. 10, 1980?

3. Tsichritzis, D., "Form Management." Paper presented at Office '80, University of Toronto, September 8-10, 1980.