Panel: Multimedia Database Management Systems

Chairperson: S. Christodoulakis, University of Toronto

Panelists:

D. Badal, Hewlett-Packard.

A. Cardenas, UCLA.

P. Mantey, University of California Santa Cruz (tentative).

F. Tompa, University of Waterloo.

D. Tsichritzis, Computer Research Institute of Crete (tentative).

Additional members to be announced.

Multimedia Data Base Management: Applications and Problems

A Position Paper

Stavros Christodoulakis

Database management systems have been very successful in a variety of applications in the commercial world such as air line reservations, banking, etc. These applications typically use formatted data. There are other interesting applications which require unformatted data such as text, bitmaps, graphics data, and voice.

Application environments which may potentially require the storage and manipulation of unformatted data are office information systems, geographic data bases, CAD/CAM and technical documentation systems, software reusability systems, some military applications, medical information systems, law information systems, tourist information systems, advertising catalogues, newspaper and magazine production systems, large dictionaries and encyclopaedias, systems used in education, and library automation environments.

Management of unformatted data presents a variety of new challenges for data base management researchers. In addition, new and diverse devices (like bitmap displays, voice input output devices, optical disks, etc.) may be used for the presentation and storage of unformatted data. These new devices give more dimensions to the problem. We use the term *multimedia data base management* to refer both to the problems of managing unformatted data, as well as to the problems introduced by the devices which are used for the presentation and storage of unformatted data.

There are several issues associated with the development of such systems.

1. A first issue relates to the *software architecture* of these systems. Should they be *extensions of data base management systems?* Should they be *completely different systems with a common user interface?* Should they be based on a *completely new data base management system which understands all the different data types?* It is my opinion that extentions to existing data base management systems to handle all these diverse data types in all application environments may result in complex and inefficient systems.
2. Another issue relates to the user requirements for *content addressibility* in these diverse data types. Access based on a unique identifier is clearly inadequate. *Content addressibility in text* may have to deal with spelling errors. *Content addressibility in images* may have to use *indirect techniques* (like text content addressibility), *similarity retrieval techniques, adaptive techniques,* or *spatial relationships* Automatic information extraction techniques may be needed. Content addressibility in voice data is still bound by technology limitations.
3. *Performance* will be crucial for the success of these systems. *Access methods* for these data types have to be developed. *Hardware architectures, physical data base design techniques* and *query optimization techniques* must be investigated. *Clustering of information* to facilitate the search should be investigated.
4. *User interfaces* in these environments may be different than in traditional data base environments. Images and voice may be used for *efficient* and *"active"* ways of communication. The capabilities of voice and images to increase the bandwidth of man-machine interaction should be exploited. *Information* and *presentation of information* should be separated. Presentation will eventually have to deal with the problem of *diverse devices.* Some *image processing* techniques also become of interest in this context.
5. Very large data bases may be used as repositories of information (*archivers*). Information may be *extracted* from these repositories, possibly *transformed* and *correlated* and then used together with information which is inserted interactively for the *formation* of new information. Primitives and supporting structures have to be developed. Extraction of information from *facsimile data* is also important.
6. *Concurrency control, recovery, security* and *version support* for these environments have to be investigated. They may be heavily application dependent, however. Common requirements across applications should be identified.
7. Unformatted data have very large storage capacity requirements. A single picture may require hundreds of kilobytes or, sometimes, megabytes of storage. In addition many of the applications discussed are largely archival in nature. New, *large capacity storage devices* like *optical disks* may be used for storing unformatted data. These devices and the different nature of the unformatted data types may create new performance problems. *Duplication control, compression* and *compaction* techniques may have to be developed.
8. Application of various *information retrieval* techniques seems relevant in some of these application environments. It may be desirable that they are integrated with the data base techniques.
9. *Working prototypes* should be developed and given to users for *feedback.*

Many of the above problems are not well understood. The design and implementation of integrated prototypes will help us understand better these issues and pinpoint new problems.