

Structuralizing Multimedia Data

When you store something—data or physical things—the goal is to be able to access them later. All storage techniques follow organization principles to let us retrieve information rapidly and efficiently. Computer scientists were the first to develop powerful data models to organize and store data so that we could retrieve it quickly at the right time with a minimal amount of computing effort.

When data started to grow, we developed database techniques to deal with structured data. Relational databases became popular because they let us map data on storage devices and articulate queries in a unified manner. As a result, developers created applications that took advantage of relational databases. Although object-oriented databases offered a more flexible data model, they weren't successful because of their poor efficiency and efficacy compared to the relational model. To work around this problem, people learned to represent object-oriented models on relational databases.

The Internet changed things by allowing millions of people to create Web sites. These Web sites mostly contained text documents. When the Internet started to grow rapidly, portals appeared on the scene with their arsenal of keyword-based searches and taxonomies inspired by library science approaches. This worked initially, but we realized that keywords couldn't capture the meaning of documents. So, techniques based on artificial intelligence approaches to natural-language understanding or on neural networks approaches to classify documents started cropping up. It became clear that all these approaches would have only limited success and a limited lifespan.

The wonder of XML

Then, a time-proven technique came forward: Extensible Markup Language (XML). XML brings us to the basics by asking document creators to introduce enough clues, or structure, in the document so that an automatic process can read what the document or a section of it is about. This

metadata approach will let more advanced systems know more about the document much better than today's automatic techniques can. It also has the ability to work gracefully with more automation. So what's XML's basic contribution? It introduces structure in otherwise unstructured documents. That is, it structuralizes text.

Multimedia data, like other data, must be stored using organization principles. We should organize multimedia data more carefully because of its time-serial nature and its enormous size. Another difficulty is that current metadata for audio, video, images, and other similar sources is more about the data than about its semantic content.

The tags in XML introduce semantic partitioning of text. We need to develop techniques for introducing the semantic partitioning of video, audio, and images. We've spent considerable effort in the last few years on developing automatic techniques for video and audio segmentation and for indexing images based on some basic characteristics such as color and texture. These techniques are very useful and will revolutionize how we'll organize multimedia data in the future. However, we need to organize multimedia data today. The current automatic techniques for semantic partitioning are even more infantile than those for text. The only solution may be to develop powerful approaches for structuralizing multimedia data, which could prove to be as revolutionary as the introduction of XML.

Where do we go from here?

So, how do we start structuralizing video? How can we start introducing some semantic metadata while creating video? What dictionary will we use for this VXML, or Visual Extensible Markup Language? How far will emerging standards like MPEG-7 go in this direction? Unfortunately, I don't have the answers, but I think this could be a very interesting research direction. I leave it in your hands to guide us through this maze. **MM**