MODELLING MULTIMEDIA

Carlo Meghini
Consiglio Nazionale delle Ricerche
Istituto di Elaborazione della Informazione (IEI-CNR)

THE PROBLEM

Multimedia databases are a new generation of information systems accessed by an increasing number of users to satisfy their information needs. The essential difference between a traditional database and one of the new generation is that the latter contains views of reality, the basic components of multimedia documents, which are obtained by electronic devices, while the content of the former is the product of a mind that abstracts from reality relevant facts and encodes them in a language that a computer is prepared to process in a meaningful way.

As a consequence, retrieving information in a multimedia database is a much more complex task, which requires support from many different areas of computer science: digital signal processing, for extracting information from "raw" digital objects such as images and audios; information retrieval, for matching automatically derived profiles of objects with queries, typically in similarity-based retrieval; databases, for supporting the efficient storage and retrieval of large volumes of uninterpreted blobs.

This interdisciplinarity of multimedia information retrieval is gaining ground in the scientific community, and cross-fertilization from the involved areas is now happening. However, the state of the art in this field is far from being satisfactory, as one important aspect of the problem is still mostly ignored, namely the exploitation of semantic information in retrieval. A multimedia document should be seen not only as a combination of a number of signals of a different nature, but also as an information carrier. As such, it should be modelled, and consequently retrieved, also on the basis of its semantic content.

There are two important problems that need be addressed to proceed along this direction: first, the representation of document semantics and its proper use in retrieval; second, the integration of semantic-based retrieval with "traditional", similarity-based retrieval into a unique framework.

THE APPROACH

In order to properly deal with the contents of the documents in a multimedia collection, one has to set up a conceptual space, that is, an abstract model of some slice of reality, which the documents are supposed to refer to. Explicit representation and principled usage of such concepts will enable users to make their way into the realm of pixels and voxels, which means that they have an ontology on the basis of which they can understand documents and, therefore, retrieve them.

The overall goal of injecting semantics into multimedia information retrieval is to design retrieval systems that are more effective than current systems. By effectiveness, we mean the ability of an information system to carry on its task in a way that is adequate to the users' expectations. This refers to intelligence more than to efficiency, that is, the ability of exhibiting some sort of understanding of the underlying concepts.

THE PROGRAMME

In order to import semantics into information retrieval, one has to define an information retrieval model providing:

1. a multifaceted representation of documents taking into account their structure(s), content(s) and layout(s)
2. an explicit representation of relevant domain knowledge
3. a powerful language for expressing users' information needs
4. a matching function associating queries to the documents that satisfy them in a way that captures a relevance notion.

THE TOOL

From a conceptual point of view, information retrieval is a question answering task that obeys some logic, in the sense that for the answer to make some sense with respect to the question, there has to be a logical connection between the two. Therefore the system in question can be formalized in terms of a logic.

This, by now obvious, idea goes back to the 50s and to John McCarthy, who launched it as an effective tool for Artificial Intelligence; it has been well accepted in the Database field since the

70s, and was realized in Information Retrieval (IR) around 10 years ago, when logical IR modelling was called for in order to (a) make IR models open to mathematical scrutiny, and, more importantly, (b) to devise new models able to overcome the poor performance of existing models.

Happily, logic is also the most natural candidate to set up a conceptual schema, as it has been devised to represent and reason about the facts of everyday life, even though its massive usage for establishing the foundations of mathematics may have hindered its original goal.

Logic is well established both from the philosophical and the mathematical point of view. The keyword here is semantics: We know pretty well the connection between natural language and logic, and logic has thorough mathematical foundations.

A number of logics have been devised to formalize different forms of reasoning. An enormous variety of logics exist which model different aspects of reasoning: from the logics of relevance, to those of probability, time, space, etc.

A non-negligible subset has been studied from the computational point of view, and many have been implemented and tested on real cases. For many of these logics there are also computational results, relating their decision problems to computational classes such as undecidability, or the many hierarchies of structural complexity. There are numerous implementations, ranging from Prolog to modal systems, and, lately, to terminological logics, which have also been tested in real applications.

THE SPECIFICITY OF MULTIMEDIA

As it happens, logic is also a most viable tool for solving the second problem mentioned in the introduction, that is, the integration between two complementary worlds: signal processing and symbolic processing.

The degree of interaction between these two souls is a crucial factor for the adequacy of the model. Typically, current multimedia models, for instance image models, either fall into one or another of these categories, or can be decomposed into two independent sub-models, each belonging to a different category.

In order to accomplish full integration, the signal (form) and symbolic (content) components have to be related by the model, so that colors and shapes can be addressed from within the same expressions used to address object identities and properties. This amounts to defining a conceptual space able to deal with content as well as form concepts.

WHERE ARE WE

In the context of a basic research action of the EU (FERMI, n. 8134), we have defined a document model, based on a relevance terminological logic and individual and predicate closures to capture closed-world reasoning. The model deals with the structure of multimedia documents as well as with the form and the content of its basic components, texts and images. Uncertainty is dealt with by using fuzzy reasoning. The logic comes equipped with a calculus for proving assertions, thus allowing effective IR.