Conceptual Modelling in Multimedia Information Seeking

Carlo Meghini and Fabrizio Sebastiani
Consiglio Nazionale delle Ricerche
Istituto di Elaborazione dell'Informazione
Via S. Maria 46 - 56126 Pisa, Italy

E-mail: {meghini,fabrizio}@iei.pi.cnr.it

1 Motivation

Information retrieval, once considered "the Cinderella of computer science" (in the words of one of its main advocates), is now witnessing a booming interest in the light of the ever-growing amount of information repositories available on both local and, especially, distributed platforms, including the Web and the Internet at large. Other relatively new paradigms of information seeking, such as browsing, information gathering and information filtering, are just other tiles in this generalised search for tools and techniques capable of reducing information overload and of selecting the right information at the right time in very large, dynamically evolving sets of documents.

The availability of non-textual ("multimedia") documents has given a new twist to information retrieval research, unfortunately setting even farther in the future the time in which generalized, completely automatic indexing methods will be available that allow the answering of content-based queries. If a system really capable of fully automatic recognition of "Authoritative sources on the social impact of nuclear waste disposal" seem currently out of reach, even more so for needs of "piano sonatas in styles influenced by art nouveau", "cubist paintings representing violins" or "scenes on sieges from war movies".

Complex queries such as the above mentioned only reflect the complexity of the information needs of many sophisticated information seekers of today; and it is only apparently that the queries of less sophisticated users have a smaller intrinsic complexity. While in a few decades these information needs might be satisfied by fully automatic systems, for today and tomorrow we will have to make do with systems relying on a mixture of automatic, semi-automatic and manual indexing methods, as it is only through manual intervention that it is possible to inform a retrieval system of the ins and outs of a (non-textual) document.

In this respect, an important contribution to multimedia information seeking may come from the more content-minded conceptual modelling and knowledge representation communities, as these have been developing formal and semi-formal languages and methodologies for the representation of information and for the conceptualization of an application domain. It is expected that, in the mid to long term, successful information seeking systems will have to be based on the close interplay of manually (or semi-automatically) created representations of document content based on conceptual modelling technology, and the automatically created representations of document form of the information retrieval tradition.

2 Topics addressed at the workshop

The papers presented at the Workshop address several related aspects of the information seeking problem. In the following, we have categorized them in three broad classes: modeling Web documents, modeling multimedia documents, and browsing. While the last class is clearly disjoint from the previous ones, the other two classes only overlap in the abstract sense. Concretely, the papers thay containg emphasize different aspects and thus interact only to a negligible extent.

The first category has attracted most attention, due to the importance of the Web as a repository of information. Indeed, the success of the Web is due to the total freedom it leaves to information publishers with respect to document formats and content. However, the price that everybody pays everyday for this freedom is the lack of effective access to information. The papers falling under this category address the problem of providing access facilities to the Web based on a predefined model for its documents.

The section on multimedia is devoted to models for video data or for complex documents, comprised of time-dependent basic media. The papers greatly differ in style: one is entirely informal, whilst the other is (almost) entirely formal. They both make valuable proposals, taking the notion of multimedia at heart.

Finally, effective browsing is an open problem for the users of multimedia digital collections. Due to either the limited effectiveness of the underlying retrieval services, or to the imprecision and vagueness of user queries (if not both), these users often face the problem of exploring large answer sets to find out what they are looking for. The first paper presented in this section introduces generic tools for supporting effective browsing, while the second one discusses the problem of browsing through a special class of documents, namely videoconferencing records.

3 Modeling Web documents

In “Formal model for integration of data in complex structure”, **Y. Kambayashi and S. Meki** discuss the problem, of particular relevance to the field of Data Warehousing, of recognising structure in heterogeneous data obtained from distributed and heterogeneous platforms, such as the Web, and of identifying commonalities and differences in structure among different data items, which ultimately allow the classification of these items into a common scheme. Extracting structure is seen here as preliminary, but ultimately conducive, to the extraction of content, and to the recognition of common content features.

The authors develop a model of structure for these heterogeneous types of data. The basic assumption is that structure reflects semantic constraints that the data items comply with, and that structure identification may thus be conducive to inferring these constraints. The approach expoused in the paper is articulated in three main steps: (a) extracting semantic constraints from the data obtained from a given source, (b) identifying the commonalities among semantic constraints pertaining to the different sources, and (c) providing a structure that accommodates the data that complies with the constraints so identified.

Issues dealing with the visualization of structures are tackled by distinguisishing abstract structure (here called *form*, and represented by regular expressions) from the possible ways of visualizing it (called *frames*). Abstract structure is the main focus of the paper, which developes a set of rules for the translation between equivalent abstract forms and the detection of commonalities between non-equivalent ones.

In “Web document modelling and clustering” **W. Song** discusses the need for a formal model for the description and characterisation of Web documents, a model that would allow to represent the bewildering variety of Web-based information chunks under a common scheme. The benefits of this proposed scheme would consist in greater ease of Web-based document searching, grouping, classification and filtering. Song argues that keyword-based searching, as performed by many well-known search engines, yields too low precision and recall to meet the need of sophisticated information searchers, and that these needs can be met only if adequate ways of describing Web documents are introduced that allow search engines to exploit the descriptions built along these guidelines. The paper goes on to propose a model, called WDM, which has the objective of being sufficiently general to accommodate existing structures of Web documents, and to be easily applicable for a variety of document management tasks, such as classification or clustering. The model is based on the hypothesis that Web browsers should become capable of interpreting metadata labels from a *label taxonomy*; these labels describe documents under a variety of facets, e.g. access rights or content, and may be provided either by the authors themselves or by third-party rating services. As different authors or sites may exploit different labels, a WDM schema integration method is proposed that closely resembles the well-known methodologies for subschema integration in conceptual database design. The paper also proposes that a number of previously proposed similarity measures for documents described in WDM should be used for document clustering, or categorisation, so that related documents may be either stored or browsed together.

"A new approach to query the Web: Integrating Retrieval and Browsing", by **Z. Lacroix, A. Sahuguet, R. Chandrasekar** and **B. Srinivas** presents AKIRA, a system for accessing the Web via queries expressed in the PIQL language, an extension of the OQL query language with object identity and generalized path expressions. To this end, AKIRA views the Web as an object oriented database, which can be queried in the standard way, with additional predicates for matching text expressions. The result of a query is then a view of the whole database, whose structure is defined by the original query. This view, which current Web crawlers store in an apposite cache, is handled by AKIRA through an object-oriented DBMS, where the view is accumulated by Web crawling agents and accessed by the user through PIQL. Essentially, the view creation process is based on information retrieval techniques, applied to the candidate Web pages by the AKIRA agents; once a view is obtained, it is filtered and structured as specified in the user query, by relying on database techniques. The main merit of this work is its attempt to provide a more sophisticate interaction with the Web than current Web search engines. This goal is pursued from within a database perspective, by employing information retrieval techniques as well as by relying on advanced software technology, such as agent-based system engineering. Despite the fact that several aspects of AKIRA deserve a higher attention, the system is a step in the right direction, that is the development of systems that permit to study the problem of accessing the Web from a pragmatical point of view.

In "Multimodal Integration of Disparate Information Sources with Attribution", **T. Lee** and **S. Bressan** present COIN, a system providing access to a number of diverse information sources, including flat files, databases and, of course, the Web, through a relational database interface. The novelty of COIN over the many systems that offer the same, or a more general, service, is the explicit handling of attribution, that is information about the source where the data retrieved to the user come from. The authors argue in favor of attribution as a way of (a) ensuring the protection of intellectual property, (b) giving a mean to test the quality of the retrieved data, and (c) monitoring updates and revisions of data in, *e.g.* data warehousing. At the formal level, the authors rely on an extension of relational algebra to treat attribution information. At the implementation level, a key role in the COIN system is played by COIN Web Wrappers, that is software modules extending the typical functionality of wrappers to handle Web pages and attribution information.

4 Modeling multimedia data

In "A Multiple-Interpretation Framework for Modeling Video Semantics", **C. Lindley** attacks the problem of identifying a rich semantic model for the retrieval, browsing and synthesis of video data. The paper starts out with a wide review of current approaches, pointing out their various limitations in providing a satisfactory account of video content. The author observes that the kind of representations considered, while successfull from a purey computational viewpoint, performs poorly in terms of effectiveness, as judged from a user perspective. In order to overcome these limitations, the author argues in favour of a semantically richer model, able to capture aspects of a video that directly relate to the cultural context in which the video is meant to perform its information purpose. As a necessary step in this direction, the notion of interpretation paradigm is introduced. An interpretation paradigm is a set of fundamental assumptions about the world to be modelled, along with a set of beliefs following from those assumptions. The meaning of a phenomenon is then to be understood as the interpretation of that phenomenon within the selected interpretation paradigm, while modelling semantics reduces to identifying the interpretation paradigms that are suitable to the phenomenon at hand. For video data, four paradigms are identified by the author. The *diegetic* level deals with the narrative elements of a video. These include the video narration, fictional time and space dimensions, the characters, landscape, and events, all giving raise to a four-dimensional spatio-temporal world. The *connotative* level is the level of metaphorical, analogical and associative meaning that the diegetic level may have. The *subtext* level deals with the hermeneutics of video, that is the hidden and suppressed meanings, extending the immediacy of intuitive consciousness. Finally, the *cinematic* level is concerned with how formal and video techniques are used to express the meanings of a video. The author shows how typical content-based queries relate to the four levels of his framework, and discusses the implications of his model for system infrastructure. The tight links with semiotic analysis make this work really valuable; the discussion on the state of the art in the field is pregnant.

In "Modelling Multimedia and Hypermedia Applications using an E-LOTOS/MHEG-5 Approach", **P. Maia Sampaio** and **W. Lopes de Souza** apply formal models techniques to support a crucial phase of multimedia/hypermedia document authoring, namely the specification of the temporal and logical synchronization of the various components making up a document. The basic motivation underlying this work is the development of a concise, well-understood and well-founded notation to describe multimedia/hypermedia documents, thus favouring the interoperability over heterogeneous platforms of applications handling such documents. The approach followed by the authors is to combine the international standard MHEG-5 *Multimedia and Hypermedia InformationCoding Expert Group*) with E-LOTOS, an extension of LOTOS (*Language Of Temporal Ordering Specification*). In particular, the building blocks of the notation are E-LOTOS processes, which may be instantiated and composed to define a specification. There are two libraries of such processes: the presentation library, constituted by basic media objects (video, image or audio sequence), and the constraint library, containing objects for expressing non-trivial logical and temporal synchronization constraints among media objects. The paper discusses how these blocks can be modelled in MHEG-5, deriving a complete mapping of the considered primitives. The resulting model is a contribution to the retrieval of multimedia documents on the basis of their structure. In fact, the model can be understood as a formal specification of the strucutre of dynamic multimedia documents, from which a suitable representation and an associated query language may be derived.

5 Browsing

In "Browsing navigator for efficient multimedia information retrieval" **T. Kakimoto and Y. Kambayashi** tackle the problem of providing users of information retrieval systems with friendlier tools for the browsing of retrieved documents. The problem is of special importance in the non-textual case, as, due to the still low effectiveness of current multimedia retrieval systems, following a user request many documents may have to be visualised before the user has satisfied his information need, and this visualisation process may be too time-consuming for the user. A set of tools is proposed aimed at making browsing the retrieved documents easier and more intuitive. One of these tools is a clustering facility, that allows the set of retrieved documents to be partitioned into clusters of documents having a strong intra-cluster similarity; a variety of similarity measurements may be used, including system-generated or user-tailored similarity functions. Clusters are also automatically analysed in order to find, based on the above-mentioned similarity measures, documents that may be used as cluster representatives in order to spare the user the cognitive load of visualizing too many documents. Users may also give feedback to the system by marking documents relevant, and the system is able to exploit this "relevance feedback" information by individuating the documents that are similar to the documents marked relevant by at least a pre-specified threshold, thus displaying them to the user as "strong candidates" for relevance. The problem of visualising the retrieval results is also tackled, and a solution involving Kohonen's "Self-Organizing Maps" is proposed; the separate visualization of clustering results obtained through different similarity functions may thus help the user to select the similarity function that gives the results most appropriate for his needs.

In "Index generation and advanced search functions for multimedia presentation material" **Y.**

Kambayashi, K. Katayama, Y. Kamiya and O. Kagawa tackle the problem of providing sophisticated facilities for browsing recordings of presentations, possibly produced by videoconferencing software. Presentation records are usually stored as a video of the presentation complemented by the set of overhead slides that were used by the speaker; the video and the slide sequence are synchronised. This record can be searched, e.g. by people who were not able to attend the presentation, by the standard functions defined either on the video component, including "playback", "fast forward" and "fast rewind", or on the slide component, such as "go to the next slide", etc. One of the main problems of video retrieval is the location of relevant portions of a video, because current scene analysis technology is not yet ready to provide automated content-based indexing of videos, and because manual content-based indexing is impractical. The authors suggest that the slides component of a video recording of a presentation may be useful for this, as portions of the video may be located by using standard textual information retrieval techniques against this component, thus allowing to locate sections of the presentation in which the speaker was discussing a given topic of interest to the searcher. The authors present a sophisticaded method for exploiting this idea that can not only search slides by direct location, or by content, but that can also use as searching aids the recordings of important actions by the speaker, such as manually underlining (or using the pointer to locate) portions of a slide, meaning that those portions are of particular relevance to the presentation section synchronised with these actions.