Smart Multimedia Meeting Information Retrieval for Teaching and Learning Activities

Christian Guetl and Victor Manuel García-Barrios
Institute for Information Systems and Computer Media (IICM)
Faculty of Computer Science at Graz University of Technology
Austria
{cguetl,vgarcia}@iicm.edu

Abstract: Based on the fact that face-to-face and virtual meetings in organizational units increasingly take place in today's business processes, it is not surprising that a lot of valuable information is generated during such events. Further, meeting information can be a valuable knowledge source for teaching and learning activities, e.g. it can support trainers and teachers in identifying emerging topics for vocational trainings and for the provision of illustrative real-life examples. Indeed, staff members and learners get access to this helpful information in order to solve problems and improve their skills. However, there is still a gap in order to integrate meeting information into teaching and learning activities. In order to overcome this problem, this paper focuses on smart meeting information retrieval supporting teaching and learning purposes in companies and research institutions.

Introduction

It is well documented that the amount of information continuously grows in almost any area and subject. Organizational units, such as companies, research organizations and educational institutions, are affected in several ways. On the one hand, they have to keep pace with the dramatic amount of information flow outside and inside their organizations in order to gather, manage, store and retrieve relevant knowledge. On the other hand, to be competitive staff members must continuously adapt and improve their knowledge, skills and expertise order to be competitive.

In the context of this paper, we want to concentrate on the valuable information addressed in face-to-face or virtual meetings and workshops in organizational units. As stated in (Romano & Nunamaker 2001), managers and knowledge workers spend between 25% and 80% of their working time in meetings and the median number of participants in the analyzed meetings is nine (9). Such meeting-based information can be a valuable source for teaching and learning activities. On the one hand, trainers can identify emerging topics and problems addressed in meetings to plan new courses and vocational trainings. Or they can enrich their course material by providing illustrative real-life examples and case studies from archived meetings. On the other hand, staff members get access to helpful information in order to solve problems and improve their skills as well as get support by learning-on-the-job activities. (Guetl & García-Barrios 2005)

An extensive recherché has shown that there exist a number of tools and applications to support planning activities of meetings and their executions as well as to support meeting recordings and their archiving. However, there is still a gap in order to integrate meeting information in teaching and learning activities (for details see Guetl & García-Barrios 2005). This fact motivated us to design and implement the Semantic Meeting Information Application (SMIA) within the MISTRAL research project, (MISTRAL 2005).

This paper is focused on smart meeting information retrieval supporting teaching and learning purposes in companies and research institutions. The reminder of the paper is organized as follows: a brief overview is given about the MISTRAL research project. Then, we focus on the MISTRAL Semantic Application unit and enlist a set of requirements for teaching and learning activities. Based on these requirements, a first implementation for Smart Multimedia Meeting Information Retrieval and its visualization is described in more detail.

The MISTRAL Research Project at a Glance

In general, the MISTRAL project aims at storing, processing and annotating of multi-modal data streams in meeting scenarios (Guetl & García-Barrios 2005, MISTRAL 2005). The data consists of video and audio data and click-data streams of presentation interactions. In addition, meeting relevant documents in digital formats are also processed and used. This workflow is built on several conceptual units (see also Fig. 1 below).

In order to exploit semantic annotations from the multi-modal meeting recordings, four uni-modal units process the meeting data: (1) The *Video Processing Unit* provides information about meeting participants (participant identification and spatial location, gesture and facial expression) and information about trained meeting-relevant objects (object identification and spatial location). (2) The *Audio Processing Unit* exploits acoustic-based information about the meeting participants (participant identification and spatial location, voice characteristics) and trained sounds (object identification and spatial location). (3) The *Speech-to-Text Processing Unit* extracts textual information from oral talks. (4) The *Text Processing Unit* processes speech-to-text data and meeting-relevant documents and provides additional information, such as topic classifications, summaries, concepts, and the like. All the information added from these units is called *features*.

Following, the *Multi-modal Merging Unit* is applied in order to merge the extracted features of the uni-modal units and to gain further information, such as confidence boosting or contradiction detection. Finally, the *Semantic Enrichment Unit* adds additional semantic information inferred by means of a knowledge base.

The semantic information annotated by the core MISTRAL system is stored in MPEG-7 data format by applying the Detailed Audio Video Profile (DAVP), as described e.g. in (Bailer et al. 2005, Martinez et al. 2002). Thus, annotated data as well as the recorded meeting data and meeting-relevant documents provide the basis for further usage by the *Mistral Semantic Application* Unit, which is discussed in the following chapter.

The MISTRAL Semantic Application and its Requirements for Teaching and Learning Purposes

The *MISTRAL Semantic Application* focuses on knowledge transfer processes for educational purposes. It bridges the gap between the useful and important information addressed in meetings and the information needs for training and learning activities in organizational units. Based on example scenarios and concluded requirements covered in detail in (Guetl & García-Barrios 2005), the first implementation of the semantic application is driven by a subset of those requirements, which are summarized in the following.

Requirements from the Viewpoint of User Roles

(1) *Supporting trainers, tutors and staff members* responsible for vocational training, e.g. identifying interesting and emerging topics or recurred problems.
(2) *Access for learners* to practice-oriented knowledge corresponding to lesson topics and course material, e.g. best practices, FAQs, problems, and model solutions.

Requirements from the Viewpoint of Information Needs

(1) *Retrieval and combination of basic meeting information*: (a) organizational information (meeting data and project relevant information), (b) participant information (attendees, roles and activities of participants, spatio-temporal information), and (c) meeting content and time-specific information (speech-to-text and related documents, topics and content abstracts).
(2) *Retrieval of meta-knowledge* inferred from the basic meeting information: (a) discovering knowledge assets addressed in meetings (either following the topics in the agenda or representing 'off-subject' knowledge for further reuse), (b) identification of questions/answers and problems/solutions, (c) clustering similar topics as well as their frequency-based co-occurrences, (d) compiling 'pros' and 'cons' according to a particular term or subject, and (e) discovering and providing best practices and specific problem solutions.
(3) *Access to statistical information* derived from (1) & (2): (a) showing number and frequency of topics related to a meeting, a project or a participant, and (b) determining temporal occurrences of features and concepts.

Requirements from the Functional Viewpoint

(1) *Meeting Information Retrieval* embraces, on the one hand, the definition of functionalities in order to identify, to pre-process, to access and to recall relevant meeting information form the Meeting Data Management Repository. On the other hand, the system has to manage representations of the meeting data and build data structures in order to support users' information needs.
(2) In particular, *information visualization* is very important for a useful information access according to the

users' information needs. Based on the information needs stated so far, the system has to provide search result presentation as a linear list in table form, information-structure-dependent visualization by metaphors for hierarchies, graphs, similarities, and time series. In addition, meeting recordings and meeting recording annotations has to be properly visualized by the system.

Smart Multimedia Meeting Information Retrieval for Teaching and Learning Activities

To follow the lines of the MISTRAL overall objectives and the requirements in the scope of knowledge transfer stated so far, smart multimedia information retrieval features have to be provided to trainers and learners in order to support their activities. The focus of this chapter is to outline our first solution approach and implementation details (see also Figure 1).

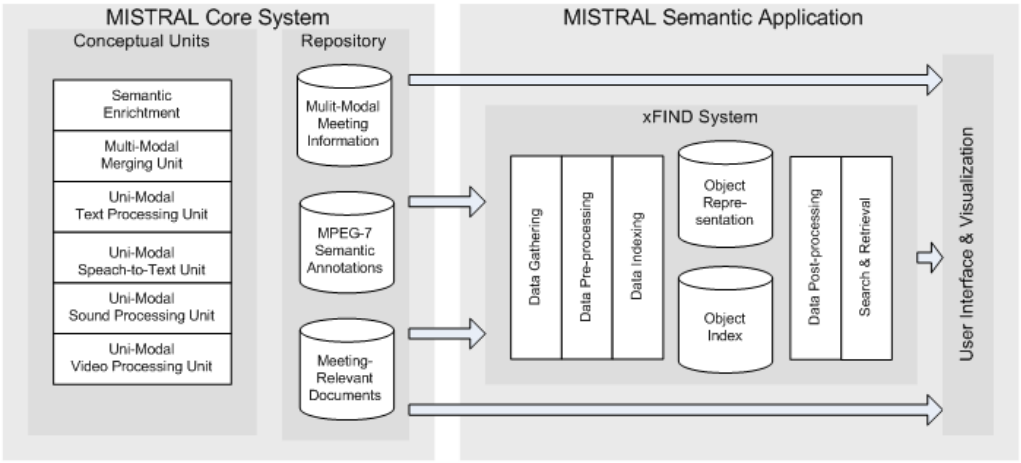

Figure 1: Conceptual architecture of the Smart Multimedia Meeting Information Retrieval solution approach

The left side of Figure 1 outlines the *MISTRAL Core System*, which can be logically split up into the *Conceptual Units* and the *Meeting Repository*. The Conceptual Units process the multi-modal meeting recordings to gain semantic information to be used for describing and searching meeting content and scenarios. The Meeting Repository stores the multi-modal meeting information and can deliver meeting recordings (video and sound) in various qualities for different retrieval or visualization purposes. Furthermore, the Meeting Repository manages also meeting-relevant documents (meeting agenda, project reports, and the like) and semantic annotations about the multi-modal meeting recordings in XML-styled MPEG-7 data structures. This information is also accessible for further processing and visualization purposes.

On the right side of Figure 1, the *MISTRAL Semantic Application* is illustrated. It can be seen as the mediator between trainers' and learners' information needs and the semantically annotated meeting information. Thus, the back-end is represented by the repository of the MISTRAL core system. The front-end on the users` side is given by tailored user interface and information visualization approaches. In order to support finding relevant information and discovering relations and coherences from the underlying meeting information, information retrieval procedures in terms of data pre-processing, data indices and data post-processing must be applied. We have decided to build on the xFIND system, because of the specific requirements and due to our familiarity with the system. The meeting-specific information retrieval solution based on the xFIND system and visualization approaches are discussed in the reminder of the chapter.

xFIND (extended Framework for Information Discovery) has been developed in a former research project at IICM. The source code is available under the LGPL license and is under continuously development by the xFIND community. (xFIND 2005, Guetl 2002) In general, the architectural design of the system follows the requirements of being a platform-independent, flexible and extensible system. xFIND is implemented in pure Java and its core modules are the *xFIND Gatherer*, the *xFIND Indexer* and the *xFIND Broker*. Each module can even run on physical separated computers, the information exchange between xFIND modules is managed by a specific protocol, the xFIND query communication format (xQCF).

The xFIND Gatherer

The Gatherer module handles the following workflow: In the first step, *Data Gathering*, it crawls through documents in arbitrary file formats according to predefined rule sets as well as through specified sources and parsed link references. The next step in the data processing chain, *Data Pre-processing*, is split into two subtasks. First, the Gatherer calls various so called 'filters' in order to process different MIME types or file types. As a result, content and structure information as well as metadata are extracted, which are made available in an internal format for further processing. A novel feature in this context is that embedded media objects are also processed and made available for further usage, for example, embedded images are extracted and thumbnails are generated, which may form part of the search results. The second subtask handles file type independent tasks, such as language guessing, topic classification, and the like. These tasks may also enrich metadata of the documents. For both subtasks specific functionality could be added by plug-in mechanism. In the third and last step of the workflow, extracted data are mapped into an internal representation for storage and further usage by the xFIND Indexer. This mapping allows for instance to compose metadata fields into one data field-group in the internal representation.

The workflow chain described so far can be processed by means of xFIND standard functionality (see [xFIND]) for various instances of meeting relevant documents in the file formats HTML, Plain Text, PDF, Windows Word, Excel and PowerPoint. In order to describe the purpose of the document (e.g. agenda, project report, and the like), the set of attributes is expanded by the attribute document type.

For the purpose of handling meeting information stored in MPEG-7 metadata files, we have decided to implement a flexible MPEG-7 document filter as a further plug-in for the xFIND system. Its functionality is focused on the necessary pre-processing tasks for extracting data structures to be indexed by the xFIND system. The MPEG-7 document filter is highly customizable in terms of the following perspectives: (a) partitioning of the meeting scenes, (b) extraction of specific sets of metadata according to the requirements for training and learning activities, and (c) mapping of metadata (as addressed in point (2) of 'Requirements from the Viewpoint of Information Needs' in the previous chapter) into the internal data representation of the xFIND system.

The need for partitioning of entire meetings in smaller scenes is given by the requirements stated so far and the underlying application scenarios. In general, it seems to be obvious that long meeting sessions need to be segmented into logical units of contents. For our first implementation we apply the following partitioning rule. A meeting scene starts at the beginning of the oral talk of a meeting participant and ends either by stopping the talk or by changing the presentation slide. Using this simple rule, we can handle dialogs as well as presentations activities.

For our first implementation we start from the substantial set of available semantic information annotated by the MISTRAL core system, and then focus on some important *groups of metadata* for each scene in order to fulfill our main requirements within the scope of training and learning activities. These semantic information groups are *meeting participants* (participants visible in the scene, active participant = speaker), *meeting content* (speech-to-text extraction), and *meeting content-based metadata* (subject classification, concepts, keywords, linkage to agenda items and/or presentation slides, content type addressing questions or answers, start-point and end-point of the scene). Like the other filters, also the MPEG-7 filter generates thumbnails from the identified scene in the meeting recordings. The number of thumbnails is given by the duration of the scene and can be used for search result representation and visualization. Furthermore, a set of administrative metadata is generated in accordance to the xFIND specification, such as document creation time, last update time, URI, and the like.

The xFIND Indexer

The Indexer module manages indices and processes search requests in order to support efficient information retrieval and to provide relevant search results. The latest xFIND version enables user-defined attributes of various types (text, numbers, and list elements) to be (a) indexed, (b) applied in search queries, (c) provided in the search results and (d) taken into consideration in the ranking procedure. Thus, the meeting relevant set of attributes provided by the Gatherer module can be easily processed by standard functionality.

Furthermore, the xFIND system enables the application of additional metadata, the *xFIND Quality Metadata (xQM)*, which is physically and organizational separated from indexed documents. The intention of the xQM idea is to describe document sources (entire document repositories, subsets of repositories, and even documents) by quality attributes that can be managed by other users or even by other organizational units. The latest xFIND version also enables user-defined attributes for xFIND quality metadata sets.

In the scope of meeting information retrieval for teaching and learning purposes, project managers and trainers can annotate parts of the meeting repository by means of educational relevant xQM sets. One of the

main advantages is given by the fact that several documents can be described by one instance of xQM, and therefore causes less annotation effort. For example, one instance of xQM can describe a particular meeting or project and any relevant indexed document can be linked to this metadata. The indexer enables either to search in xQM sets or to combine a search request using xQM attributes and the meeting document repository.

In order to prepare suitable data structures for search results representation and visualization as well as to provide relevant chunks of information for describing the search results, xFIND applies *Data Post-processing* procedures. Like pre-processing at the Gatherer module also post-processing features are added by plug-in mechanisms. In the current implementation, one post-processing feature delivers sections of full-text according to the search query and highlights query terms properly. Another post-processing plug-in compiles attributes from meeting document representations and xQM in order to describe search results by means of metadata sets.

The xFIND Broker

The Broker module is the interface between the meeting information retrieval system and end-users (trainers and learners). On the one hand, the Broker module handles the interaction with users providing proper interfaces for the information retrieval process and for the visualization of search results. On the other hand, it manages the communication with the Indexer module in terms of preparing proper search queries from user interaction, sending search requests to the Indexer and finally handling the search results according to users information needs.

The xFIND Broker architecture is designed to be flexible in terms of three aspects: (1) It can be applied as a generic search front-end for all users in an application domain or it can be tailored for a specific user group for particular purposes and tasks. (2) Because of its platform-independency, Broker instances can either run on the server side accessible by Web clients or as a kind of a thin client on the client side. It is worth to mention that search results are cached at the Broker side, reducing the network traffic between Indexer and Broker; this enables efficient representations of cached search results, where views can be easily switched by the users according to their training and learning activities. (3) Because of its extensibility, the xFIND architecture supports the integration of graphical visualization tools.

Figure 2 depicts the server-side standard xFIND search interface and the linear search result list visualization adapted for meeting information retrieval. The 'simple search' user interface (see left side of Figure 2) is designed to support untrained users to specify their information needs by simple typing one or several keywords in a search form. Complex search request options (e.g. data fields to be searched, ranking procedures, and the like) are predefined for the users. However, two advanced search interfaces are available for power users in order to specify their particular information needs. Search queries can also be stored as 'predefined queries' for further usage, e.g. by learners who are not familiar with information retrieval systems.

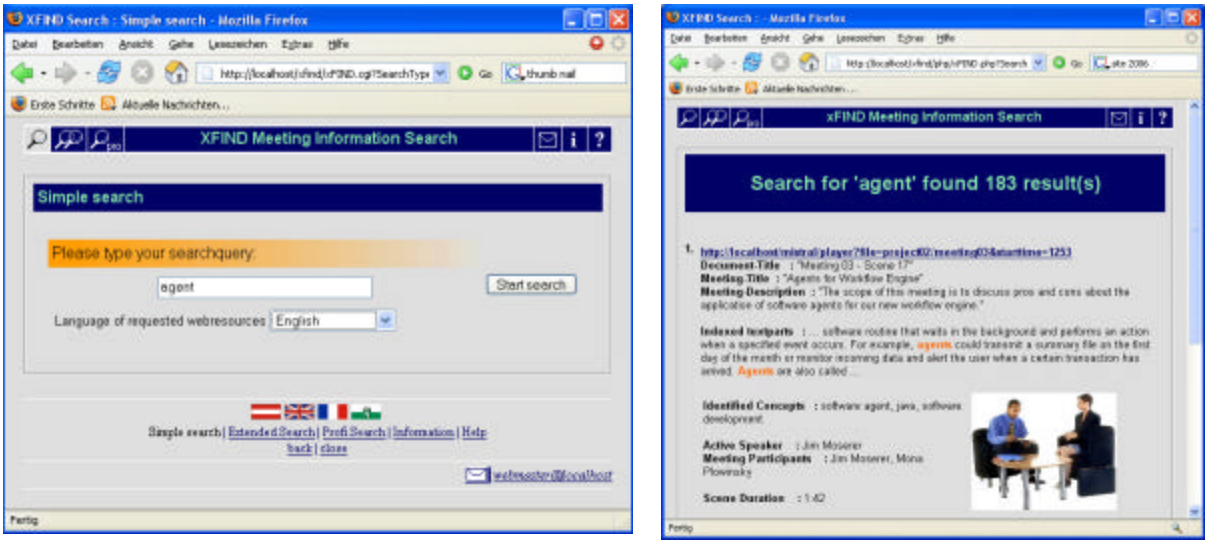

Figure 2: User interface (left side) and search result visualization (right side).

The linear search result list (see right side of Figure 2) visualizes information about relevant meeting information according to the search query and the specified set of result information in a ranked order corresponding to the

specified relevance model. Thus, users are supported to grasp relevant information related to their specific needs, e.g. the screenshot example shows content information (content excerpt in the context of the search query and concepts identified in the text). Furthermore, general information about the meeting and meeting participants are delivered to the user. In addition, a thumbnail provides visual information about the meeting scene. Hyperlinks attached to the title and to the thumbnail of search result items enable users to send requests in order to receive the video scenes.

Conclusions and Future Work

In this paper we have shown that meeting information can be valuable and helpful for teaching and learning purposes, but there is still a gap in the integration of meeting information into teaching and learning activities. The MISTRAL Core System processes meeting recordings in order to annotate and semantically enrich meeting information. The MISTRAL Semantic Application focuses on knowledge transfer procedures for educational purposes. Its first implementation using the xFIND system bridges the gap between the useful and relevant information addressed in meetings and the information needs for training and learning activities in organizational units.

As future work, we will set up a test environment at IICM for the Web Application Group and for the lecture 'Information Search and Retrieval' in order to conduct a user survey. In addition, we want to integrate our smart meeting information retrieval system in knowledge transfer environments by following basic ideas of our *Dynamic Background Library (DBL)*, as stated in (García-Barrios et al. 2002). Furthermore, we will add personalization features for tailoring information to the users' information needs. Finally, we will complement search result visualization by various instances of visualization methods.

References

Bailer, W., Schallauer, P., Hausenblas, M., Thallinger, G. (2005). 'MPEG-7 Based Description Infrastructure for an Audiovisual Content Analysis and Retrieval System'. *Conference on Storage and Retrieval Methods and Applications for Multimedia*, USA.

García-Barrios, V.M., Guetl, C., Pivec, M. (2002). Semantic Knowledge Factory: A New Way of Cognition Improvement for the Knowledge Management Process, In *Procedings of SITE 2002* (pp. 168-172), Nashville, Tennessee, USA.

Guetl, C., García-Barrios, V.M. (2005). Semantic Meeting Information Application: A Contribution for Enhanced Knowledge Transfer and Learning in Companies. In *Proceedings of ICL 2005*.

Guetl (2002). Approaches to Modern Knowledge Discovery for the Internet - An Approach to the Information Gathering and Organizing System xFIND (Extended Framework for INformation Discovery). *Ph.D. Work*, Graz University of Technology, Austria.

Martinez, J.M., Koenen, R., Pereira, F. (2002). 'MPEG-7: The Generic Multimedia Content Description Standard, Part 1' (pp. 78-87). In *IEEE Multimedia*.

MISTRAL (2005). MISTRAL Project. *Official Website*, http://mistral-project.at

Romano, N.C., Nunamaker, J.F. (2001). Meeting Analysis: Findings from Research and Practice. In *Proceedings 34th Annual Hawaii International Conference on System Sciences*.

Whiteside, J., Wixon, D. (1988). Contextualism as a world view for the reformation of meetings. In *Proceedings of the 1988 ACM conference on Computer-supported cooperative work*.

xFIND (2005). xFIND Homepage – The Bridge to Knowledge. *Official Website*, http://xfind.iicm.edu

Acknowledgement

The project results have been partly developed in the MISTRAL project (Measurable intelligent and secure semantic extraction and retrieval of multimedia data - http://www.mistral-project.at). MISTRAL is financed by the Austrian Research Promotion Agency (http://www.ffg.at) within the strategic objective FIT-IT under the project contract number 809264/9338.