*Journal of Universal Knowledge Management, vol. 1, no. 1 (2006), 45-53*
*submitted: 14/5/06, accepted: 29/5/06, appeared: 28/8/06 © J.UKM*

# Exploitation of MPEG-7 Descriptions on Multi-modal Meeting Data: First Results within MISTRAL Project

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

**Abstract:** In this paper we present first results of the implementation work on the exploitation of MPEG-7 descriptors within the MISTRAL system, which aims at enhanced procedures for semantic annotation and enrichment of meeting-related multi-modal data. As the system consists of several specialised units, the focus is set on its Semantic Applications Unit (SemAU). First, we give a short overview on the MISTRAL system as well as explain the overall architecture and some functional modules of SemAU. Then, emphasis is placed on the role of MPEG-7 metadata within the MISTRAL system as standardised bridge between its units and as basic standardised input for SemAU. Finally, some insights are given into the current development stage of two user-centred applications of SemAU. These applications manage semantically enriched multi-modal data in order to enhance its retrieval and visualisation.

**Keywords:** meeting information, multi-modal data, MPEG-7
**Categories:** H.4, J.4, K.3, I.5

## 1 Introduction

At present, face-to-face and virtual meetings increasingly take place in business processes. As stated in [Romano et al. 2001], managers and knowledge workers spend 25% to 80% of their working time in meetings (the median number of meeting attendees was nine). From our literature survey we identified the following most common meeting goals: reconciliation of conflicts, decision making, problem solving, learning and training, knowledge exchange, reaching a common understanding, and exploration of new ideas [Whiteside et al. 1988]. As big financial and human efforts are invested to create and transfer knowledge among meeting attendees, we identify an economisation potential by increasing the efficiency of meetings applying smart tools, such as meeting browsers, e-conferencing or group support systems ([Antunes et al. 2003], [Lalanne et al. 2005]). Further, knowledge addressed and created in meetings should be preserved and made accessible for all company members.

Indeed, multi-modal meeting applications and meeting information systems are of emerging interest. Thus, it's not surprising to find several research projects being conducted in this context. Though, our survey has shown that there is still a lack of integration facilities into knowledge management and e-learning systems. This fact motivated us to set the focus for the first prototypes within the Semantic Applications Unit (SemAU) of MISTRAL's system on *integrative information retrieval and visualisation*. The paper gives first results of the implementation of those prototypes.

A short overview on the MISTRAL system and an introduction into architecture and functions of SemAU follow this chapter. The prototypes are explained in chapter 3.

## 2 The MISTRAL System and its Semantic Applications Unit

The MISTRAL research project aims at smart semi-automatic solutions for semantic annotation and enrichment of multi-modal data from meeting recordings and meeting-related documents [Mistral 2006]. Face-to-face and virtual meetings can be recorded using different modalities. Usually, cameras and microphones are applied for this purpose, but also other specialised modalities are conceivable, such as device-usage or click-behaviour recorders. Still, all relevant information remains hidden or beyond users' reach into those distinct multi-modal streams. Thus, in order to efficiently and contextually manage the knowledge addressed and generated in meetings, further semantic processing, enrichment and integration of multi-modal data are needed.

### 2.1 Multi-modal Data within the MISTRAL System

To manage multi-modal data, MISTRAL's system consists of *conceptual units* for data management, uni-modal stream processing, multi-modal merging of extracted semantics, semantic enrichment of concepts, and semantic applications (see Fig. 1).

The *Data Management* unit represents the main storage repository for all MISTRAL-relevant data. The *Uni-modal* unit consists of five modules: video, audio, speech-to-text, text and sensory modules. Respectively, they process image data, sound data, speech transcriptions, text documents and additional multi-modal streams. The *Multi-modal* unit merges uni-modal data as well as checks the confidence of comparable uni-modal extractions. The *Semantic Enrichment* unit is responsible for conflict detection between uni-modal annotations and for further semantic inferences. The outcome of these units is annotated into a common file following the MPEG-7 standard. Finally, the *Semantic Applications* unit (SemAU) is the 'first consumer' of this outcome and builds the front-end of the MISTRAL system for external clients. For more details about the MISTRAL system please refer to [Mistral 2006].

### 2.2 Some Requirements for the Semantic Applications Unit

MISTRAL's SemAU focuses on the integration, retrieval and visualisation of meeting information. As stated in [Gütl and García-Barrios 2005], the main requirements for SemAU are divided in three categories: *user roles*, *information needs* and *system functionality*. These viewpoints are relevant, because SemAU is not meant to represent a single application. Rather, it embraces several modules that provide services, which can be orchestrated in order to fulfil different requirements.

To give some examples consider the following requirements:

**User roles:** *UR1* personalised views on meeting data; *UR2* context-dependent provision of data, e.g. according to specific tasks in workflows or knowledge expertise.
**Information needs:** *IN1* access to concepts extracted from uni-modal data (i.e. low-level semantics), such as object movements (video), utterances of persons during a

meeting (audio) or agenda-specific topics (text); *IN2* access to annotations from multi-modal or semantic enrichment (i.e. high-level semantics), e.g. identified spatio-temporal actions (standing up, leaving); *IN3* identification of e.g. hot topics or recurred problems in past meetings.
**System functionality***: SF1* search & retrieval, e.g. identification, pre-processing, access and recall of relevant information; *SF2* multi-model orchestration, e.g. users, knowledge domains or business processes; *SF3* multi-modal orchestration and disassembly, i.e. provision of different views on meeting data, e.g. if users are only interested in data extracted from audio stream; *SF4* explorative visualisation, e.g. enable visualisation of search results as linear lists, graphs or similarity clusters.

### 2.3 Architecture and Functionality of SemAU

As stated in the last section, SemAU aims at providing a modular, flexible and extensible architecture in order to enable the orchestration of distinct applications. For this purpose a service-oriented architecture was chosen (as shown in Fig. 1). SemAU embraces different components: MISTRAL Portal (MP), Adaptation System (AS), Modelling System (MS), Visualisation System (VS) and Retrieval System (RS).

External clients may access MISTRAL'S semantic applications via Web-based services of MP. Only RS has direct access to the MISTRAL Core Framework (MCF) via specialised modules: data from the MCF is pre-processed, indexed and managed by the xFIND Search & Retrieval Module; further, a Streaming Server enables access to audio/video streams. SemAU's adaptive behaviour is located in AS. The tasks of AS are defined by semi-transparent functions, i.e. its Context Resolver may force adaptation procedures or just pass-through a client request to RS or VS.

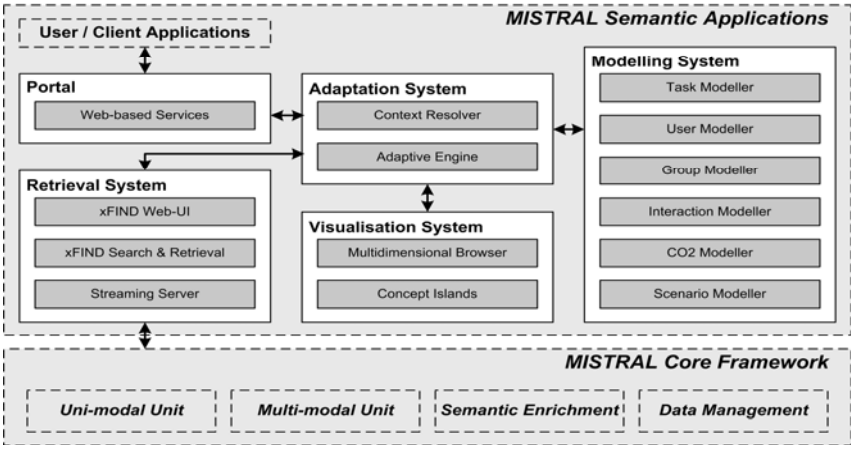


*Figure 1: Architecture of MISTRAL's Semantic Applications Unit.*

In SemAU, the orchestration of services depends on scenarios, which are defined by requirements (see section 2.2). For clarification, consider the following examples. The scenario for a personalised access to meeting information (*UR1*) depends on current traits of users, i.e. for each individual need different modellers in MS are

needed (*SF2*). Further, let's say a user is looking for information about one specific topic over many meetings. This common retrieval need, which is generally defined in *SF1*, can be easily solved by RS. On the other hand, what if the same user does not really need the topic information, but wants to contact an expert on that topic? This more complex semantic need (*UR2*) requires specific data in the user models and its relation to meeting topics. Thus, RS needs support from AS and MS (*SF1*, *SF2*). Now, consider a user needing the identification of hot topics in a specific set of meetings (*IN3*) according to specific agenda points (*IN1*), but only those hot topics addressed by meeting participants that are involved in the current user's role or task at a certain state in a workflow process (*UR2*). For this complex scenario, the AS requires knowledge from several modellers in the MS and has to adapt their outcome according to the given scenario. In fact, this last example may not be completely solved trough service orchestration, but SemAU enables also the possibility of utilising different visualisation tools (*SF4*). E.g. a multi-dimensional visualisation of search results that are enriched with specific metadata and clustered by context-similarity could help the user to self-find a solution for the requirement.

In summary, the main functionality of MISTRAL's SemAU focuses on three aspects: *Search & Retrieval* (S&R), *Multi-purpose Modelling & Adaptation* (MMA) and *Multiple Visualisations* (MV). Some details about the S&R functionality were already published e.g. in [Gütl and García-Barrios 2005, 2006]. Since the MMA functions of SemAU are at present in the design phase, the reminder of this paper concentrates on current results regarding the first prototypes for S&R and MV.

## 3 First Implementation Results

This chapter summarises the implementation results of two SemAU prototypes at the current development stage. These software prototypes are "Semantic Meeting Information Application" (*SMIA*) and "Explorative Visualisations Framework" (*EVF*, in concrete its module "Multi-Dimensional Metadata Visualisation System" *MD²VS*). Still, let us first focus on the role of MPEG-7 within the MISTRAL project.

### 3.1 The Role of MPEG-7 within MISTRAL

The components of SemAU enable users an access to semantic meeting information, which is stored according to the MPEG-7 standard [Martinez 2003] and obeys basically an event-driven data structure. This means that semantic annotations from the MISTRAL units are assigned to segments on the time-line of the meeting recordings and are all collected in a single MPEG-7 document for each meeting. Thus, each annotation corresponds exactly to "the time when something happened in the meeting".

It is important to state at this point, that the role of MPEG-7 metadata within the MISTRAL project is defined as follows: (a) MPEG-7 represents a standardised bridge between the MISTRAL units, (b) MPEG-7 serves as basic metadata source for the applications in SemAU, (c) MPEG-7 enables a standardised exchangeability of semantic annotations for other systems or research groups, and (d) MPEG-7 annotations may be viewed, processed and reused by non-MISTRAL tools.

For the MISTRAL system, a common and exchangeable annotation file format is needed, because each unit independently processes distinct media from the meeting recordings and makes own annotations at different semantic levels. With MPEG-7, multimedia content can be described in a standardised way, Thus, it enables a standardised possibility to easier archive, access, locate, navigate, search and manage multimedia metadata, which in the case of MISTRAL, is extracted from different documents (mostly text) and multi-modal data streams (e.g. audio, video, click-data). Further, MPEG-7 provides the ability to annotate (1) low-level features, (2) semantics and (3) structural aspects of multimedia files. In MISTRAL, semantic enrichment of multimedia takes place along the line *(a) uni-modal features → (b) multi-modal merging → (c) further semantic inferences.* Thus, (a) can be mapped to (1), (b, c) to (2), and (3) can be e.g. defined by storage structures in the Data Management unit or by semantic categorisation of meeting topics extracted by the uni-modal text module.

### 3.2 Semantic Meeting Information Application

First basic descriptions of *Semantic Meeting Information Application* (SMIA) can be found e.g. in [Gütl and García-Barrios 2005, 2006]. The main functionalities of SMIA are contained within the Retrieval System (see Fig. 1), for which the open source search system xFIND represents its core module [xFIND 2006]. The left side of Fig. 2 shows one search result for a *SMIA Simple Search* (SSS). Also, Real Player's streaming output is shown (after clicking on the corresponding hyperlinked elements).

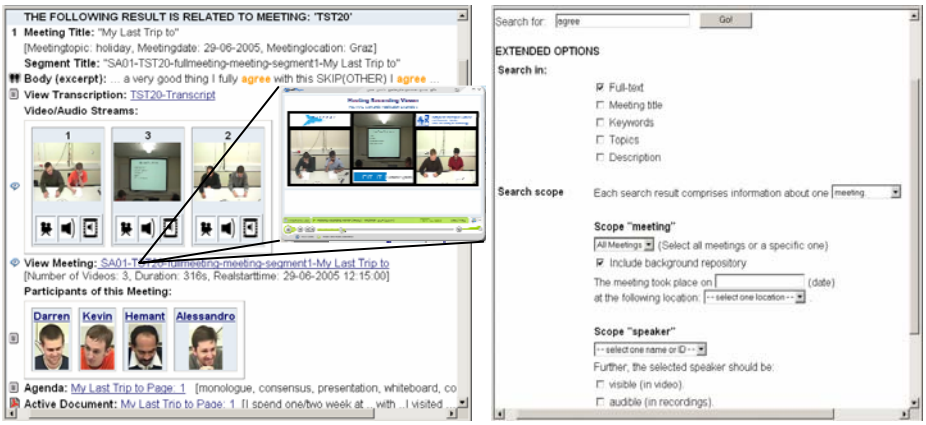


*Figure 2: User Interfaces of SMIA (left: Search Result; right: Extended Search).*

Unlike common search engines, where one result corresponds to one document, each result of SSS (see Fig. 2, left side) visualises the most relevant information about an entire meeting corpus. Thus, SSS is not meant to find single documents or modalities, rather it matches search queries in order to find relevant meetings. To identify the document where the keywords were found, SMIA highlights them in excerpts of indexed content (see *agree* in *Body (excerpt)* of Fig. 2, left side). In addition, visual information about meeting recordings is provided through thumbnails. Hyperlinked titles and icons (for audio, video or audio&video) enable users to access

these specific video recordings via SemAU's Streaming Server. This enhanced linear visualisation of search results is made possible, because of the MPEG-7 descriptions from MCF. Thus, an MPEG-7 file represents the semantic envelope for each meeting, enabling an easy and compact exploration of semantically enriched multi-modal data.

SMIA's *Extended Search* UI (Fig. 2, right side) supports users to specify special needs by simply typing keywords and selecting options in a search form. Yet, as the usage of extended features in search engines is hard to learn for novice users, SMIA places two special interfaces at the disposal of users: *Special Search: Speaker* (Fig. 3, left side) and *Special Search: Agenda*. More special searches are under development.

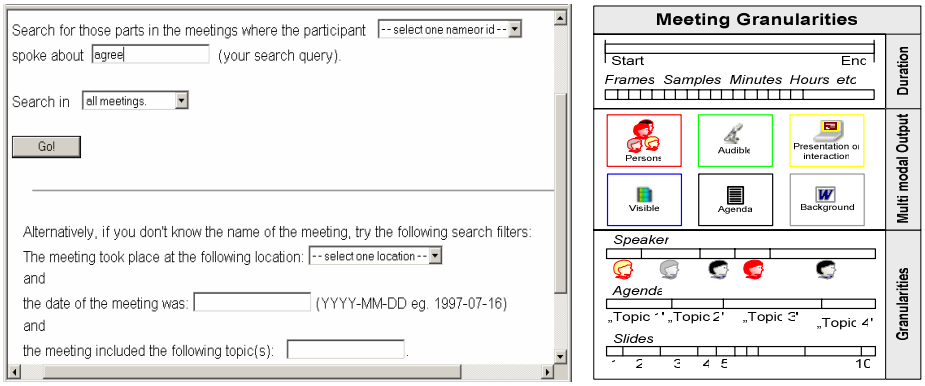


*Figure 3: Speaker Search UI in SMIA (left) and Meeting Granularities (right).*

These two special UIs are examples for *special instances* of the Extended Search, where specific granularities of meetings are given. These granularities are explained as follows: meeting recordings can be explored from different perspectives, which describe distinct divisions of the meeting duration (Fig. 3, right side). Thus, one meeting can be not only divided in *time periods* but also depending on the output of multi-modal analysis, i.e. event-driven MPEG-7 annotations: e.g. *speakers* (visible or audible), *agenda topics* (extracted from text documents) or *presentation slides* (from click-data streams). Within the context of SMIA and taking as example the *speaker* granularity (Fig. 3, left side), each search result may contain audio or video segments where meeting participants have spoken continuously or are visible in continuous video scenes, and SMIA will add for each segment all related semantic annotations.

### 3.3 Explorative Visualisations Framework

This section gives a brief introduction into the *Explorative Visualisations Framework* (EVF) of SemAU, which aims at placing distinct visualisation possibilities at the disposal of users in order to extend the UI-possibilities of xFIND and to enable multi-dimensional perspectives on the semantic spaces of meeting corpora.

In EVF, the core software module is called Multi Dimensional Metadata Visualisation System ($MD^2VS$), placed at client-side and implemented in Java. $MD^2VS$ is a smart framework that is able to control and show different visualisations, such as a multi-dimensional explorer or a data clustering tool. The dimensions in

MD$^2$VS correspond to metadata definitions in search results of SMIA, i.e. also to MPEG-7 metadata. Further, in order to make metadata values comparable, the dimensions have to be normalised to proper representations: e.g. *numbers* and *dates*, where the distance between the different values is implicit, or *enumerators*, which represent equidistant textual values, such as [“Martin”, “Helmut”] for the dimension “Speaker”. Thus, the data input for MD$^2$VS represents a set of search results with comparable metadata values.

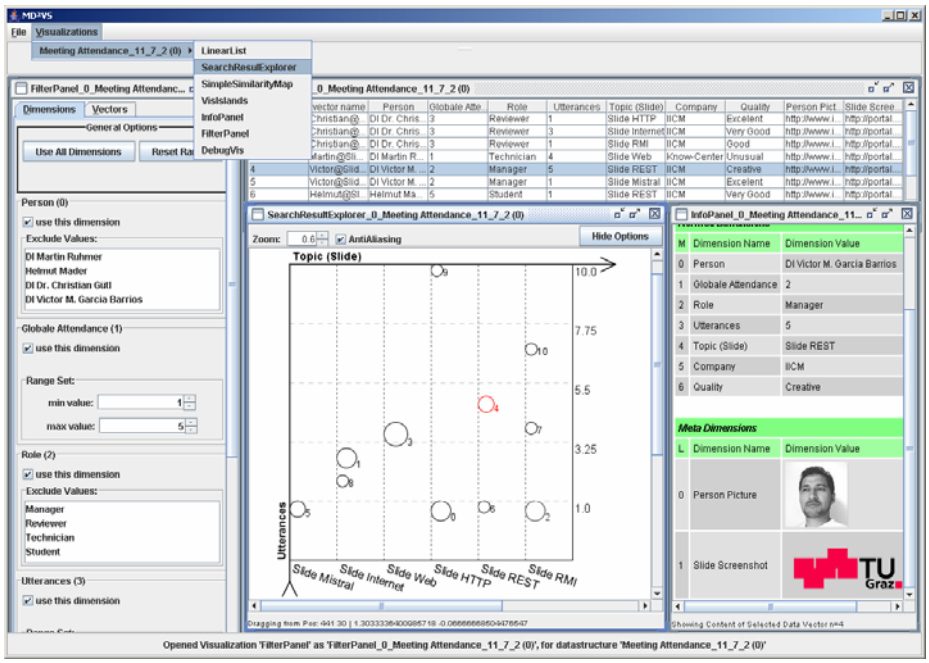


*Figure 4: Multi-Dimensional Metadata Visualisation System of EVF.*

The first prototype implementation of MD$^2$VS supports the following tools (see Fig. 4): *Linear List* (upper box), *Filter Panel* (left box), *Info Panel* (right box), and *Search Result Explorer* (middle). Further visualisations are still under development: *Similarity Map*, *Visualisation Islands* and *Speaker Communication Maps*.

With MD$^2$VS users may explore several semantic coherences among metadata defined in specific search queries. E.g. users may find “hot topics” by identifying concepts that increasingly occur in meetings. Following the example shown in Fig. 4, a user may also infer from the Search Result Explorer which meeting participants with a certain role have given most utterances related to specific topics of presentation slides. This can be useful to identify topic-dependent discussions and dialogs. One innovative value of this system is given by the fact that it integrates distinct visualisations for metadata in multi-modal corpora. The next relevant added-value relies on the fact that the visualisations are synchronised, i.e. user interactions or modifications in one visualisation propagate simultaneously to the others.

## 4 Conclusions and Future Work

The current development stage of some prototype solutions within the Semantic Applications Unit of the MISTRAL system has been presented in this paper. The results so far show clearly how the semantic gap between multi-modal data might be filled, especially within the field of Information Retrieval and Visualisation. For this purpose, modern meeting information systems should not only be highly flexible and reusable, they should also cope with the dynamic requirements of companies along time as well as with the ever-changing needs of users. We see one opportunity to enhance the efficiency of company workers in the application of modular and scalable service-oriented approaches, as for the solutions introduced in this paper.

The MISTRAL project also aims at personalised access to multi-modal meeting data. At present, we are working on the implementation of MISTRAL's Adaptive System, which will be able to orchestrate the activation and interaction of SemAU components for personalisation purposes. Further, the basic framework for the Modelling System is being finalised. Concrete applications in this context reach from simple query refinement for the retrieval system, through context-dependent visualisations of meeting data, until the integration and adaptation of meeting data for external systems. Also as future work, evaluation results of the MISTRAL system are expected in order to proof and enhance its applicability in real-life situations.

**Acknowledgements**

The project results presented in this paper were developed within the MISTRAL research project. MISTRAL is financed by the Austrian Research Promotion Agency (http://www.ffg.at) within the strategic objective FIT-IT (project contract number 809264/9338). The support of following IICM members is gratefully acknowledged: Helmut Mader and Martin Ruhmer. We also acknowledge the M4 research project (http://www.m4project.org) for the permission to use its meeting recordings (http://mmm.idiap.ch) as test data for MISTRAL's Semantic Applications Unit.

## References

[Antunes et al. 2003] Antunes, P., Costa, C.: "Perceived Value: A Low-Cost Approach to Evaluate Meetingware"; Lecture Notes in Computer Science, 9th International Workshop CRIWG 2003, p. 109-125, 2003.

[Gütl and García-Barrios 2005] Gütl, C., García-Barrios, V. M.: "Semantic Meeting Information Application: A Contribution for Enhanced Knowledge Transfer and Learning in Companies"; in Proceedings of ICL 2005, Villach, Austria, 2005.

[Gütl and García-Barrios 2006] Gütl, C., García-Barrios, V. M.: "Smart Multimedia Meeting Information Retrieval for Teaching and Learning Activities"; in Proceedings of SITE 2006, Orlando, USA, 2006.

[Lalanne et al. 2005] Lalanne, D., Lisowska, A., and others: "The IM2 Multimodal Meeting Browser Family"; Interactive Multimodal Information Management (IM2) project report, 2005.

[Martinez 2003] Martinez, J. M.: “MPEG-7 Overview”; ISO/IEC JTC1/SC29/W11 N5525, Pattaya, March 2003.

[Mistral 2006] MISTRAL project; official Web site; http://www.mistral-project.at

[Romano et al. 2001] Romano, N.C., Nunamaker, J.F.: “Meeting Analysis: Findings from Research and Practice”; in Proceedings of HICSS 2001, Hawaii, USA, 2001.

[Whiteside et al. 1988] Whiteside, J., Wixon, D.: “Contextualism as a world view for the reformation of meetings”; in Proceedings of the 1988 ACM conference on Computer-Supported Cooperative Work, 1988.

[xFIND 2006] xFIND; official Web site; http://xfind.iicm.edu