Journal of Universal Computer Science, vol. 14, no. 10 (2008), 1625-1653
submitted: 12/9/07, accepted: 22/4/08, appeared: 28/5/08 © J.UCS

Enhancements of Meeting Information Management and Application for Knowledge Access and Learning Activities

Christian Gütl
(Institute for Information Systems and Computer Media (IICM)
Graz University of Technology, Austria
cguetl@iicm.edu)

Abstract: Communication processes have become increasingly important in modern working life. Organizations invest a surprisingly high amount of financial resources and employee work time in both face-to-face and virtual meetings, yet this investment often produces poor results. To overcome this problem, research on technology-based support over a meeting's life-cycle has been increasingly conducted in recent decades. As a result of this research, particular interest has emerged in meeting information systems, which may include technology-enhanced meeting rooms as well as tools for multi-modal meeting recording, automatic meeting information extraction and annotation, in-meeting support, meeting information archiving, indexing, retrieval and visualization. Despite this great interest in and research activity on meeting information systems, insufficient focus has been paid into flexible architectures, interchangeability of meeting information as well as the integration into business processes and applications. This situation has motivated our research consortium to direct the research activity within the MISTRAL project towards a flexible and extendable system that can be easily integrated into daily working environments for knowledge access and learning activities. In this paper, we give an overview about electronic meeting systems, introduce related work on meeting information systems, outline the MISTRAL concept and its implementation, and based on that we discuss findings and problems with our research.

Keywords: meeting information system, electron meeting system, multimodal information extraction, meeting information retrieval and visualization
Categories: H.3.3, H.5.1, H.4.0

1 Introduction

In order to be competitive, modern working life in industry, governmental and research sectors is strongly influenced by intensive communication and collaboration processes; see for example [Nunamaker et al., 2001] and [Steinfield et al., 2002]. In this context, organizations invest a surprisingly high amount of financial resources and employee work time in both face-to-face and virtual meetings. Organizations, however, are frequently faced with poor meeting results far from intended expectation. This situation may be illustrated by some important findings of [Romano et al., 2001]: (1) Managers and knowledge workers spend between 25% and 80% of their working time in meetings. (2) The number of meeting participants is nine on average. (3) Most organizations spend between 7% and 15% of their personnel costs on meetings. (4) Despite the high investment of resources, meeting results often are unexpectedly poor in terms of efficiency and effectiveness.

In general, a meeting can be defined as *"a focused interaction of cognitive attention, planned or chance, where people agree to come together for a common purpose, whether at the same time and the same place, or at different times in different places."* [Romano et al., 2001] Literature studies have shown various meeting purposes, such as reaching a common understanding and group decisions, reconciling conflicts, solving problems, facilitating staff communication, exploring new ideas and concepts, exchanging knowledge, learning and training; see for example [Whiteside et al., 1988], [Nunamaker et al., 1991] and [Geyer et al., 2005]. In the light of this, interesting and important knowledge is discussed, transferred and even generated in meetings. It is obvious and supported by literature that such knowledge is worth capturing and archiving to make it accessible for further usage as well as to transfer to and integrate it with other organizational processes. [Costa et al., 2001] [Nunamaker et al., 2001] [Richter, 2005] Literature review has also shown that ineffective and inefficient meetings may result from various problems in the pre-meeting, meeting and post-meeting phases of the meeting life-cycle [Ho et al., 1999] [Romano et al., 2001] [Rienks et al., 2007].

To overcome the problems and to make use of the knowledge addressed in meetings, technology-based support for face-to-face and virtual meetings has long been a subject of research and development activities. Early systems and research are documented for example in [Turoff et al., 1977] and [Nunamaker et al., 1991]. Today, every stage of a meeting (pre-meeting, in-meeting and post-meeting phase) can be supported by a variety of technology and tools [Rienks et al., 2007] However, technology-based meeting support is still an interesting and active research topic with room for improvement and challenging issues over the entire meeting life-cycle. As a concrete application scenario within the MISTRAL research project, we have focused on the automated semantic annotation of multimodal meeting recordings in order to gain diverse knowledge from meetings and to integrate this knowledge into other organizational processes, such as knowledge access and learning activities.

The MISTRAL (Measurable Intelligent and Reliable Semantic Extraction and Retrieval of Multimedia Data) project was initiated as a result of the escalating amount of multimedia and multimodal data created by our technology-centered society. Such data are characterized by a rich and complex structure in terms of intra- and inter-relationships. Broad usage and reuse of such multimodal data is limited unless sufficient methods for semantic annotation, indexing and retrieval can be developed. In order to contribute in this context, MISTRAL innovative objectives are (1) the extraction of a large variety of relevant annotations for single modalities, (2) the merging of specifically extracted data from diverse modalities (cross-modal merging), (3) the semantic enrichment and contradiction detection based on background knowledge, (4) the feedback from cross-modal merging and semantic enrichment results to uni-modal extraction processes in order to enhance the quality of automated semantic annotations, and (5) the management, retrieval and access of extracted information [Gütl et al., 2005] [Sabol et al., 2005] [MISTRAL].

The aim of this paper is to present the solution approach and results achieved by the research project MISTRAL in the context of the meeting application scenario. In the remainder, we first give in Chapter 2 an overview about electronic meeting systems as well as introduce related work on meeting information systems and most important findings. Chapter 3 outlines the Mistral Core System, the MISTRAL semantic application for knowledge transfer and learning activities as well as discusses findings and problems. Finally, Chapter 4 addresses concluding remarks and our intended further work.

2 Electronic Meeting Systems – Background and Related Work

2.1 Technology-based Support over the Meeting Life-cycle

Over the last years, technology-enhanced group support and collaboration has gained increasing interest. Consequently, research results as well as commercial tools and applications have emerged, which can support a great variety of activities over the meeting life-cycle. Technology-based meeting support is available for meeting participants at the same location (face-to-face meetings) and at different places (virtual meetings). Based on an intensive literature survey, a brief overview about activities and a selection of supportive types of applications over the meeting life-cycle is given in the following paragraphs. Findings are based on [Nunamaker et al., 1997], [Hayne 1999], [Ho et al., 1999], [Munkvold et al., 2001] [Jain et al., 2003] [Pallotta et al., 2004], [Geyer et al., 2005], [Richter, 2005], [Sabol et al., 2005], [Rienks et al., 2007].

The pre-meeting phase aims at preparing the meeting and includes activities such as reviewing previous meetings and overall objectives, specifying the meeting goal, preparing necessary documents (agenda, presentation slides, and hand-outs), selecting appropriate meeting members, managing schedules and meeting resources, and distributing necessary information. Supportive applications include document management, meeting information archive, group calendar, agenda planning, room and resource management.

The aim of the meeting phase is to ensure an efficient and effective performance of meeting objectives. It includes guidance and moderation activities (such as chair discussion, solving conflicts, and summarizing meeting results), monitoring activities (such as keeping track of actions, participants and their level of participation, and relations between participants), and recording activities (such as transcripts, snapshots and summary). Applications include assistance in idea generation, voting, consensus mediation, note-taking and meeting recording.

The post-meeting phase aims to analyze the meeting, infer next steps and other activities, and ensure that relevant information (such as open issues, decisions, conclusions, next steps and responsibilities) is delivered and accessible to meeting participants, meeting absentees and other stakeholders. Additionally, it aims to integrate the meeting results into other business processes or processes on a higher level, such as in project management processes and knowledge management processes. Applications include semi-automatic and automatic meeting analyses, meeting transcripts, extraction of meeting minutes, meeting information indexing and retrieval.

In a complementary viewpoint, the great variety of existing applications can be classified into passive applications (users must take the initiative), interactive applications (dialog-based support) and pro-active applications (systems take the initiative). Contemporary research activities in pro-active applications experiment even with virtual meeting members coaching the meeting or portraying meeting participants. Detailed information can be found for example in [Ho et al., 1999], [Rienks et al., 2007] and [CHIL 2005a].

2.2 Meeting Information Systems

In order to narrow down to research projects and applications comparable to MISTRAL, we want to focus on specific systems, which we term Meeting Information Systems (MIS). Such systems may include technology-enhanced meeting rooms, multi-modal meeting recording tools, automatic meeting information extraction and annotation tools, in-meeting support tools, meeting recording and extracted information archiving as well as meeting indexing, retrieval and visualization. A MIS can comprise the entire set of subsystems or parts of them.

One example of an early research work in the context of MIS is the Project NICK. This project emerged in the middle of the 1980s and focused on the augmentation and analysis of face-to-face meeting. The meeting facility consisted of personal computers for each meeting participant connected in a local area network, a group work surface, software for displaying agenda and a common information space, software for enabling communication in the group and between meeting members, a video recording system, a channel for registration of the emotional state of the group, and meeting statistics capture and processing units. The system has been logically organized into three components, the capture component (collects public and private information in the meeting), the analysis component (processes the collected information after the meeting), and the presentation component (provides access to meeting information). [Cook et al., 1987]

Over the last years, meeting information systems have gained increasing interest in diverse research domains, such as automated video and audio signal processing as well as human-human and human-computer interaction. This great interest in the meeting application domain may be reflected by several interesting and partly innovative research projects in Europe and USA.

The M4 (Multimodal Meeting Manager) research project, a European Commission (EC) funded project, was started in 2002. Its aim was to develop a system that enables structuring, browsing and querying an archive of automatically analyzed meetings. The meeting room was equipped with multi-modal sensors, such as multi-channel audio and video data streams, information about presentation slides, discussion papers and whiteboard interaction. The audio-based feature extraction includes speech-to-text transformation, speaker detection, segmentation and tracking. The video-based feature extraction includes face-detection, person tracking and person action recognition. It is important to mention that the M4 research project also applies multi-modal recognition approaches for person localization and tracking. In addition, the project even focuses on group action recognition and segmentation. This great variety of the automatic creation of annotations enables a fine-grained index for synchronized multi-modal data stream to browse, search, retrieve and review meetings. [M4] [M4 2005]

The AMI (Augmented Multi-party Interaction) research project, an EC-funded project, was started in 2004 and has focused on technology-enhanced meeting rooms, which enable the collection, annotation, structuring, and browsing of multimodal meeting recordings. Modalities include audio (multiple microphones), video (multiple video cameras), slides, and textual information (such as notes and whiteboard text).

These modalities are recorded time-synchronized. Feature extraction and automatic annotations from raw modalities are performed by applying state-of-the-art processing technologies. Automatic annotations include speech-to-text transcripts, face and gesture detection. The resulting multimodal recordings enriched with annotations enable structuring, indexing, browsing and querying the meeting recordings. Various meeting browsers are developed to support the access and review of meetings. [AMI] [AMI 2006]

The recently started EC-funded research project AMIDA builds on the results of the AMI project. This new project tries to move towards live meeting processing and towards the support of meeting participants, remote participants and absentees. It aims to develop advanced videoconferencing systems which include (1) filtering, searching and browsing; (2) remote monitoring; (3) interactive accelerated playback; (4) meeting support; and (5) shared context and presence. [AMIDA]

The IM2 (Interactive Multimodal Information Management) research project was initiated by Swiss National Science Foundation (SNSF) as one of the 20 Swiss National Centres of Competence in Research (NCCR) in 2002, and it brought together university institutions and a range of commercial companies from Switzerland. The project is aimed at the advancement of research and the development of prototypes in the field of multimodal information management and man-machine interaction. The consortium has also decided to focus on smart meeting management as the application domain. Input modes include speech, video, pen, touch, hand gestures, head and body movements, and even physiological sensors. Specific objectives for automatic annotation processes and for meeting information retrieval is comparable to the M4 and the AMI research projects. [IM2 2005] [IM2 2007]

There are other interesting projects worth mentioning in the context of meeting information systems. The ISL Meeting Room project, a research project from Interactive Systems Laboratories located at University of Karlsruhe in Germany and Carnegie Mellon University in USA, focuses on audio and video modes for automatic annotations on meeting recordings. [ISL] The ICSI research focuses on speech processing for meetings. It aims to develop algorithms and systems for the recognition of speech from meetings and methods for information retrieval. [ICSI 2006] ICSI research has been linked with the AMI and IM2 research project. Between 2001 and 2005, it was also part of the Mapping Meetings project, where research on speech recording was directed to speaker separation, multi-speaker language models, detecting important regions (such as 'hot spots, agreement and disagreement), segmentation, topic detection, classification, and summarization. [MMP 2003] The NIST Meeting Recognition Project aims to support the development of audio and video recognition technologies by periodic technology evaluations and workshops as well as by collecting its own multi-media meeting corpora. [NIST 2006]

Most of the above outlined projects claim to address face-to-face meetings and remote meetings as well, but our closer review of project results and publications has shown that the projects focus mainly on face-to-face meeting situations. An early approach to combine face-to-face and video-mediated meetings was the video conferencing system HERMES (Harmonious Environment by Round MEeting Space). Remote meeting participants were represented by video screens placed between physically present participants in the face-to-face meeting setup. [Inoue et al. 1997] In another approach, the Experimental Meeting System (EMS) aims to support local participants and remote participants as well. The interesting ideas, presented in [Jain et al. 2003], include a meeting room environment for *"providing immersive telepresence in meetings to enable participants from disparate locations participate as if they are in the meeting."* However, most of the ambitious ideas addressing the meeting room environment and other parts of the system have not been followed up further. [Kim 2007]

Unlike systems and research projects stated so far, some interesting projects have emerged which proactively try to assist and support meetings. The EC-funded research project CHIL (Computers in the Human Interaction Loop) started in 2004 and ran for three years. The aim of the project was to assist humans by proactive and intelligent services in situations where people interact face-to-face with people, exchange information, and collaborate or learn. To reach this goal, the system monitors and processes information from various communication channels, such as speech, gestures, body posture, other sensor data, and slides. [CHIL 2005a] [CHIL a] [CHIL b] A similarly ambitious research project, CALO (Cognitive Assistant that Learns and Organizes), was initiated by the US Defense Advanced Research Projects Agency (DARPA). The overall objective was to research and develop a cognitive agent-based system that can reason, learn, and respond to assist in military situations. [CALO 2006a] [CALO 2006b] In the context of meeting support, the Task Discussion component focuses on meeting understanding by monitoring users' collaborative multimodal dialogue during meetings. To reach this objective, the component intends to process speech, pen input, manual gestures, facial expressions, and body movements. Based on this, the planned assistance includes meetings summarization, active suggestion of relevant documents and support of collaborative creation of documents in meetings. Additionally, it also enables browsing and accessing meeting information on different detailed levels. [CHCC] A further interesting and innovative conceptual approach was introduced in 2002. The main idea is based on the support over the meeting life-cycle through three different types of virtual meeting participants based on proactive intelligent agents. One of the virtual participants takes care of organizational aspects, one focuses on social aspects and the final is responsible for providing the group with relevant information. [Rienks et al., 2007] [Ellis et al. 2003] Some of these interesting ideas including the first implementations of virtual meeting members are developed in a prototype application. [Barthelmess et al. 2005]

2.3 Findings on existing Meeting Information Systems

The aim of this section is to give a brief summary of the state-of-the-art of meeting information systems and challenges from various viewpoints, based on an evaluation of the systems stated in the previous sections.

By starting with architectural and system model aspects, most of the research projects' results are disconnected prototype tools handling specific modalities and focusing on specific functions or parts within the meeting life-cycle. Even if the projects aim towards integration, they result in inflexible approaches with less space for further enhancements such as integration of external processing tools or improvements of automatic feature extraction for a specific modality by taking into account extracted features from other modalities. In this context, a more flexible architecture, which enables easy integration of new components and external services as well as the combination and exchange of information, may be of great additional value.

From the meeting room point of view, technology enhanced equipment includes (1) presentation tools and collaborative tools, such as presenter PC and video projector, electronic blackboard, collaborative tools for information exchange and document processing, (2) supporting applications for meeting activities, such as a voting system, tools for supporting the creativity and conflict solutions, and (3) tools and applications for recording and capturing the activities in a meeting. Modalities for the recording and capturing process include audio, video, and interaction with presentation devices, collaborative and meeting-supporting applications and personal note taking. One important yet challenging aspect arises in this context and is not paid enough attention by related research projects: data streams need to be recorded and captured in a spatial and temporal fine-grained manner by accessing various systems and application without violating privacy and security policies. Another important issue is the synchronization of diverse modalities over the time period of the meeting.

By focusing on the audio modality, information extraction includes speaker identification, spatial localization and tracking, speech activity detection, and other features such as a person's emotional state. Additionally, the detection of typical sounds in meetings, such as ringing mobile phones, laughing and clapping hands are of interest. Some of the challenging problems in meeting settings are ambient noise and echo effects. Further details can also be found in [CHIL 2005b], [Fiscus et al. 2005], [AMIDA 2007a] and [AMIDA 2007b]. Speech-to-text transformation and transcripts are of particular interest for information retrieval, automatic generated meeting minutes and summaries. Ambient noise, hall effects and untrained speakers, however, are still challenging issues and interesting research topics in this context; see also [AMIDA 2007b] and [Fiscus et al. 2005].

By specifically focusing on visual modality, automatic information extraction includes face and body recognition, meeting participant identification, location and tracking, active speaker detection, and other enhanced features such as head movements, gesture and facial expression, and focus of interest. Such information extraction tasks cause a number of problems, such as video camera positions and sufficient resolution, as well as light conditions. This extraction also presents challenging tasks from the video signal processing viewpoint. This information may be backed for example by [AMIDA 2007a], [AMIDA 2007c] and [CHIL 2005b].

Additionally, other modalities from sensors in the technology-enhanced meeting room provide valuable information for processing, annotating and enriching audio-visual recording. Of particular interest is information about the interaction with the presenter PC, such as temporal information about the access to presentation slides and other related documents.

From the viewpoint of multi-modal and cross-modal information extraction, the research projects mainly focus either on multiple data streams of one modality (such as multiple audio channels or video channels) or on combining audio and video signals in order to increase the reliability of the automatic extraction of information, such as speaker identification, location and tracking by audio and visual sources. One of the interesting and challenging problems is to find related events from diverse modalities, which are fuzzy in spatial and temporal occurrence but nevertheless representative of the same information from different sources. Further information can be found for example in [AMIDA 2007a], [AMIDA 2007c] and [CHIL 2005b].

Based on the extraction of low-level features and the detection of basic events, the reasoning of further events and information on a higher semantic level is an active topic in many of the MIS research projects. Information of interest comprises the detection of meeting roles (such as meeting facilitator and meeting members), meeting actions (such as monolog, dialog, note taking, and nodding) and meeting activities (such as opening, presentation, voting, and closing). [Jovanovic 2003] [Pallotta et al., 2004] [Antunes et al. 2005]

By focusing on meeting recording and meeting information archiving, retrieval and access, relevant aspects, problems and issues can be found in [Jain et al., 2003], [Pallotta et al., 2004], [Lalanne et al. 2005], [Antunes et al. 2005] and [Gruenstein et al. 2005]. Archives are built on the file system or based on database management systems. They are designed to deliver audio and video modalities as synchronized data streams and to provide temporal time marks for other modalities (such as interaction events with presenter PC, agenda, and speech-to-text transcript) for synchronization purposes and time-based access. From the interoperability viewpoint, metadata formats used in the research projects do not sufficiently address standards, such as MPEG-7 and MPEG-21; see for example [Burnett et al. 2003], [Martinez 2002] and [Martinez et al. 2002]. This seems to be a major disadvantage in the sense of applying pre-existing indexing and retrieval systems. In this context, meeting research projects mainly focus on meeting browser approaches for searching, skimming, accessing and reviewing meetings. Requirements and ideas for accessing meetings from different viewpoints (such as employee roles, meeting participants and absentees) are addressed by some of the research projects, however project results and actual prototype implementations do not cover users' demands appropriately.

Despite the broad interest in technology-enhanced meeting support and the significant number of diverse research projects in this area, concrete project results lack the appropriate integration with other business processes and information flow. To overcome this deficiency, a close integration into other business processes is a key for the actual business value of meeting information systems. Interesting and important examples of such business processes are knowledge transfer, decision making and learning activities. This may also be backed by [Costa et al., 2001], [Ingirige et al. 2002], [Gütl et al., 2005], [Page et al. 2005] and [Richter, 2005].

3 The MISTRAL System

The findings stated above motivated us to focus our MISTRAL sample application on the development of a flexible and extensible meeting information system. Based on automatic semantic annotations and enrichment of diverse modalities, it is designed to enable context-dependent access to information and easy integration into business processes such as decision support as well as learning and training activities. The MISTRAL system is logically divided into two parts: (1) the MISTRAL core system processes the multimodal input data streams and persistently stores meeting recordings and semantic annotations for further usage. (2) The MISTRAL semantic applications build on top of (1) and handle information retrieval, visualization and delivery based on concrete application scenarios.

3.1 The MISTRAL Core System

This section is based on the research work from [Gütl et al., 2005], [Gütl et al., 2005b] and [Sabol et al., 2005]. The aim within this paper is to give an overview about the concept and the architecture of the MISTRAL core system.

The MISTRAL meeting system was designed to deal with one or more video sources, multiple audio channels, and input from sensor data as well as to take into account diverse textual sources. Data sources for the first meeting setup include input stream from one high resolution video camera, a microphone array built by four far-field microphones and information about the interaction with a presenter PC (such as mouse and key events, active application and document). Furthermore, relevant meeting documents are also taken into account, such as agenda, project descriptions and the like.

The overall architecture of the MISTRAL core system is depicted in Figure 1. It is built on four unimodal extraction units for audio, video, sensor data and text sources. Detected events and extracted semantic information are temporally and spatially merged, semantically enriched and finally persistently stored in a meeting data repository for further usage.

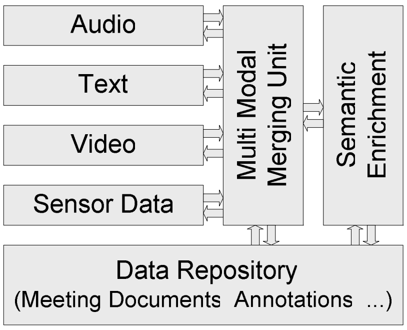

Figure 1: Units of the MISTRAL core system

The communication flow between units is bidirectional and it enables feedback loops in order to take advantage of information extracted from other units; i.e. features from other modalities can either be used to improve the extraction quality or to train or fine-tune methods for information extraction. To give one example, a robust face recognition method from the video unit can provide the audio unit with information to train an audio-based speaker identification model.

In order to gain high flexibility, extensibility, exchangeability and scalability, we have decided to follow the service-oriented approach (SOA), see for example in [Panda 2005]. It allows a loose coupling between units and components within the units by only defining contracts or interfaces. These definitions are platform-, language-, and operating-system-independent, which enables the most appropriate software environment selection for any task within the MISTRAL core system. Moreover, SOA easily supports the provision of new or improved functions by implementing and publishing a new service and integrating it within the information process workflow.

An overview over the concept and the current implementation status for each of the core units is outlined on the level of extracted semantic information in the remainder of this section. Technical details and methods for basic feature extraction are outside the scope of this paper; they are addressed in other publications by members of the MISTRAL research team and can be found elsewhere in [MISTRAL].

The Video unit is designed to detect and extract temporal and spatial information about meeting participants from one or more video streams, such as meeting participant identification, localization and tracking, facial expression and gesture, meeting participant's movements and interaction. Within the scope of interest are also the detection of relevant meeting physical objects and meeting participants' interaction with these objects. The first implementation of the video units includes identification of participants and the localization and tracking of participants.

The Audio unit processes multiple audio input channels in order to detect and extract temporal and spatial information about meeting participants, such as speaker activities, speaker identification and localization, gender estimation, emotional states, and speech-to-text transcripts. Furthermore, meeting-relevant or disturbing sound events are also of interest for the MISTRAL system, such as clapping hands, laughing, whispering, mobile phone ring tones and the like. The current implementation focuses on robust voice activity detection, speaker location estimation and speaker indexing, as well as automatic gender estimation. Furthermore, speech-to-text extractions are performed by a tool from [SAILLABS]. This method enables a speaker-independent processing on domain knowledge specific phoneme and dictionary databases.

The Sensor data unit within the MISTRAL meeting application is designed to handle a great variety of input data, such as interaction with an electronic whiteboard, presentation PC, personal devices for note taking and other sensors for capturing physical state of the meeting environment and the emotional state of the meeting participants. In our first implementation, the sensor data unit exploits the data stream from meeting participants' interactions with a presenter PC. This includes information of each event with keyboard and mouse, active application and interaction with objects and documents. For example, this module provides information about the active slide of a PowerPoint presentation or a Web site presented in the meeting. This information can be further exploited by applying text-processing tasks as described in the next paragraph.

The Text unit deals with statistical methods and natural language processing approaches to provide semantic information from text sources including speech-to-text transcripts, documents presentation slides presented in the meeting, and other meeting-relevant documents. The current implementation includes extraction of named entities, supervised concepts and unsupervised concepts such as participants' names, company names, project relevant keywords, topics and associated terms.

The Multimodal Merging unit aims at the temporal and spatial fusion of extracted meeting information and events on various semantic levels. It can merge information from different unimodal units with similar semantic meaning in order to boost or weaken the confidence level. Furthermore, the unit can also derive new information by combining extracted information and events from different units. The current implementation focuses on the fusion of extracted events and semantic information from audio and video features. The Semantic Enrichment unit handles the dissolving of identified conflicts caused by contradictory information extracted from different units. Additionally, it infers further actions on higher semantic levels such as discussion, presentation, and the like. In our first implementation, the semantic enrichment unit focuses on the detection of simple contradictory information such as how the same person cannot both give a presentation in front of a whiteboard while simultaneously sitting behind a table in a chair. Both the multimodal merging unit and semantic enrichment unit deliver information for further usage and access in a MISTRAL-specific XML data format or in the the MPEG-7 metadata standard format.

The Data Repository for MISTRAL's meeting application is designed for the persistent storage, management and access for further usage by taking into account access rights on user and group level. This includes the storage of the audio and video recordings and its delivery as streaming media files in the real media format. It also comprises the management of relevant meeting documents and its metadata as well as information about the meeting participants. In addition to these features, the data repository also handles the semantic information about the meeting extracted from unimodal and mulitimodal units, most of which are enriched with temporal information and references to the meeting recordings. In our current implementation, all data are stored in the file system and made accessible by two different media servers by using their internal rights management. The audio-visual meeting recordings are stored as real media files and delivered as streaming media by the Helix DNA Server, see also [HELIX]. Extracted semantic meeting information are stored in a MISTRAL-specific XML data format and are accessible together with relevant meeting documents in MISTRAL-predefined structure from an Apache HTTP Server, see also [Apache].

3.2 MISTRAL Semantic Application

The MISTRAL semantic application is motivated by the lack of sufficient multimodal meeting information system integration into other business processes in organizations such as knowledge management and learning and training activities. The aim of this section is to outline the design concept and our first implementation results which are based on [García-Barrios, 2007], [Gütl 2007], [Mader 2007], [Sabol et al., 2007], [García-Barrios et al., 2006], [Gütl et al., 2006], [Gütl et al., 2005] and [Gütl et al., 2005b].

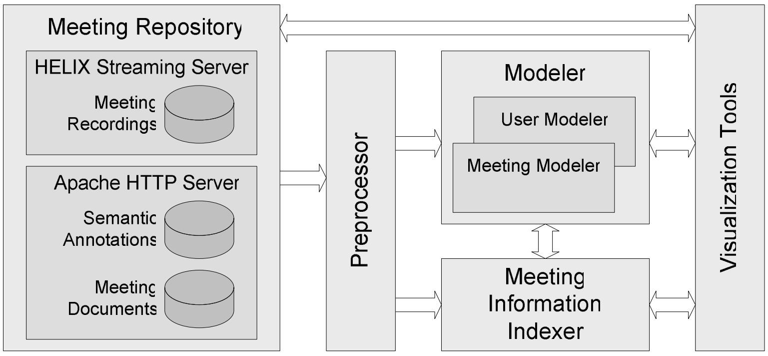

Figure 2: MISTRAL Meeting Information Retrieval and Visualization

Figure 2 illustrates the concept of the MISTRAL semantic application for meeting information retrieval and visualization. It is designed to flexibly deal with diverse meeting information from the Meeting Repository of the MISTRAL core system, which includes metadata from audio-visual meeting recordings, meeting-relevant documents and extracted semantic meeting information. The Preprocessor builds the interface between data access from the Meeting Repository at the back-end and the appropriate data delivery for the Meeting Information Indexer and the the Modeler units. Instances of the Modeler are the User Modeler, which holds information about meeting participants and the Meeting Modeler, which manages the key information about meetings. The aforementioned Meeting Information Indexer builds appropriate data structures about meetings from diverse viewpoints in order to efficiently process search requests. The Indexer's task is also to receive search requests from the front-end, process them, collect and post-process retrieved results and deliver them back to the front-end. Visualization tools handle the user interaction with the system and present the meeting information for the integrated usage in business processes.

Based on the concept of the MISTRAL Semantic Application stated above, the remainder of this section gives an overview for the current version of our implementation. The Meeting Repository is built on meeting recordings and data from the AMI Meeting Corpus and the M4 project (see also Acknowledgement section).

With the exception of the Modeler unit and specialized visualisation tools, the meeting information retrieval unit is based on the open source xFIND search system, which is built on three services: (1) the Gatherer requests, filters, retrieves and pre-processes documents in order to provide a configurable set of metadata for further processing, (2) the Indexer caches the metadata, builds on top of that, and indexes and post-processes search results for presentation, and (3) the Broker provides the interface to the user and prepares the search results for information delivery. Further information can be found for example in [Gütl 2002].

The Preprocessor has been implemented by an enhanced xFIND Gatherer that is specialized to retrieve data from the meeting repository and pre-process them for the further data processing chain. In the current version, the following data sources are processed: (1) metadata about the meeting such as meeting title, place and time, meeting duration, participants and related project. (2) Extracted semantic information such as speech-to-text excerpts, active speakers, visible meeting participants and presenter PC interaction. (3) Relevant meeting documents such as agenda, presentation slides and other related back-ground documents, for example project description, design documents and the like. (4) Video data from meeting recordings in order to create scene snapshots for information presentation.

The Meeting Information Indexer has been realized by the xFIND Indexer component configured for the specific needs of our meeting application scenario. The xFIND Indexer supports necessary data types (text, number, data, enumeration) out of the box. The pre-processed data provided by the adapted Gatherer component is cached by the Indexer for (1) building indexes on top of it for efficient search tasks (e.g. full text index from speech-to-text transcripts), and (2) preparing information for the search results (e.g. text snippets around the search terms from speech-to-text transcripts). Both data fields to be indexed and to be included for search results can be configured for the Indexer component, and therefore they can easily be adapted for specific application scenarios. Unlike in other systems and in order to cover a wide range of information needs, we have decided to build indices on and enable retrieval requests for different granularities of a meeting: (1) the entire meeting, (2) segmentation based on speaker activity, (3) segmentation based on topics according to the agenda of the meeting, and (4) segmentation based on the granularity of presentation slides. The system enables the search for these granularities for one or more media. To give an example, one can search for 'design decision' in speech-to-text transcripts and in the content of the presentation slide presented within the selected meeting granularity. Unlike other systems, which present separate search results for matching speech-to-text sections and documents, we present the results in a unified way based on the selected meeting granularity (see also Figure 4).

Within the MISTRAL project, we have decided to adopt two different approaches for meeting information presentation and interaction with users. The first one provides a "search-engine-like" user interface. This was motivated by the fact that such interfaces are commonly used and well-known by most users, and therefore the learning curve for effective usage is short. This interface was implemented by using the xFIND Broker component. Search forms and search results presentations have been tailored by means of the xFIND Broker's template engine. Different search forms support the users to describe their information needs. Power users get the freedom to construct complex search queries by applying Boolean expressions, specifying specific indexes and granularities in an expert search form (see Figure 3). A simple search form hides this complexity and enables users to search over any indexed information on the granularity of entire meetings by typing only one or several keywords. Specific forms support the users to search for meeting information on the granularity of speaker segments or agenda topics. The search result presentation depends on the selected granularity in the search process. Figure 4 depicts the search results for selected granularity of speaker segments. Speaker segments matching the search query are grouped by meetings. For each meeting, an informative overview is given including meeting title, date and location. For each speaker segment, the meeting segment title and relevant text snipped from the speech-to-text transcripts are given. Additionally, information is rendered about the meeting participants present in the segment and their activities within the scene. It also renders thumbnails from the meeting scene temporally linked to the speaker segment and visualizes the timeframe of the segment within the meeting. Furthermore, information is rendered about the segment-specific meeting document actively used by the presenter PC. Hyperlinks provided within the search result presentation enable users to request more detailed information such as the speech-to-text transcript of the specific sequence or information about the meeting participants. These hyperlinks also enable users to stream audio-visual media for a specific scene (only selected streams or the entire multimodal recording).

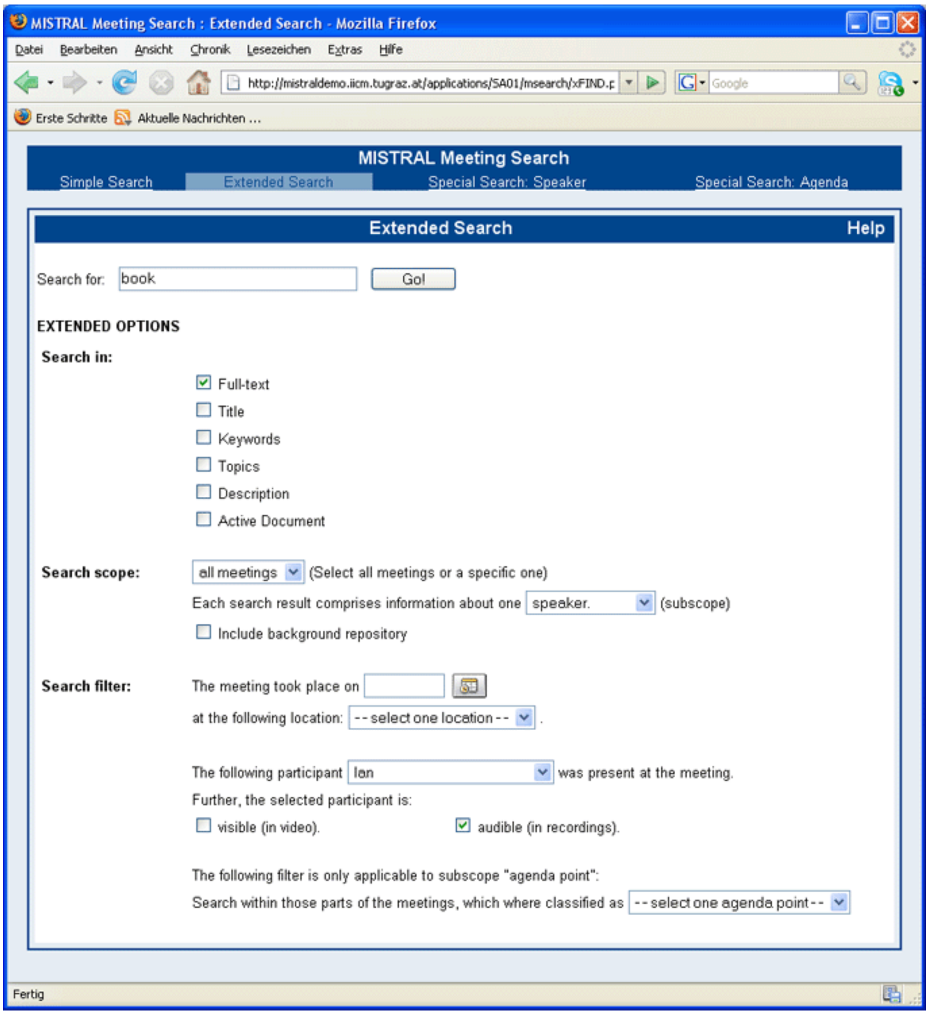

Figure 3: Extended Meeting Information Search Form

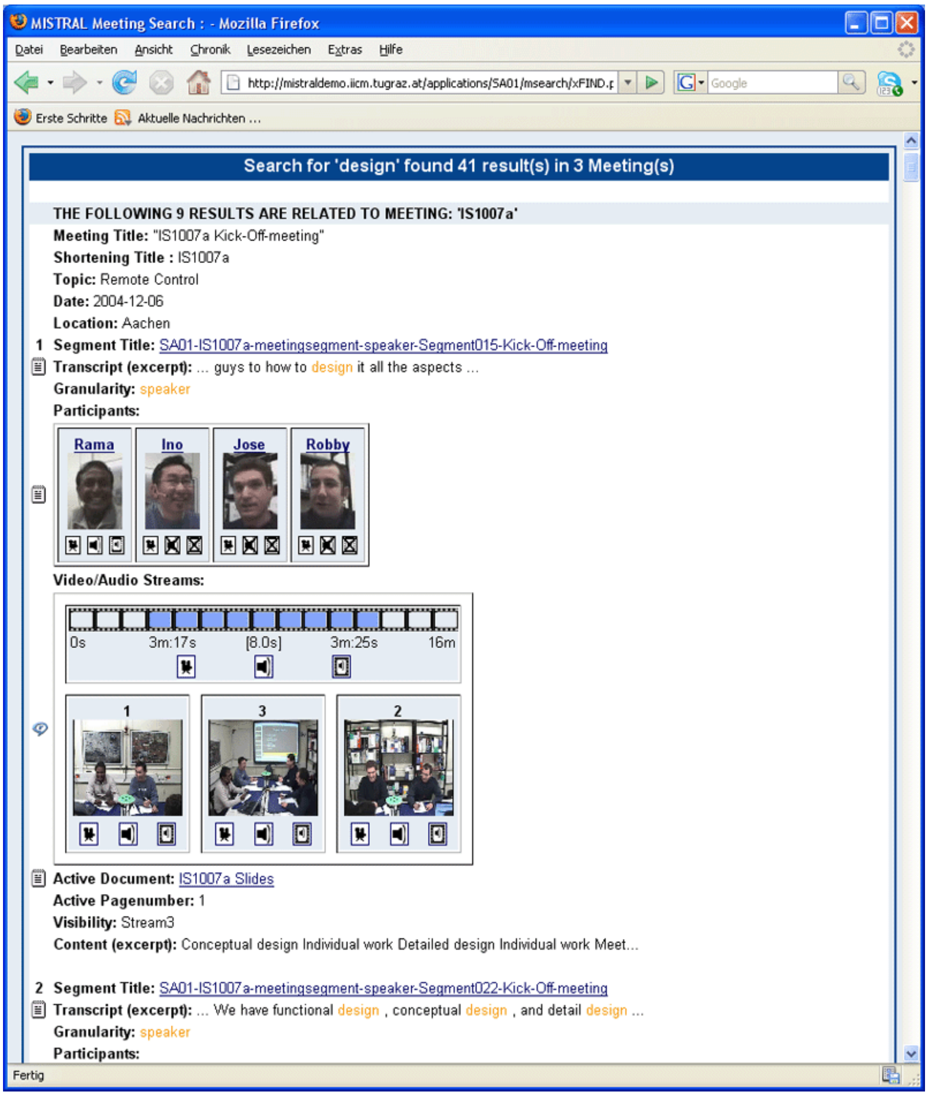

Figure 4: Information result presentation for a search request on the granularity of speaker segments

The second approach for meeting information visualization is designed to get a more comprehensive overview and to enable information discovery by various visualization metaphors. This meeting information visualization tool also acts as an xFIND broker, i.e. it provides the interface to the users and communicates with the MISTRAL Meeting Information Indexer unit. The handling and inspection of meeting data is supported by three generic panels: Filter and Control Panel, Linear List Panel (lists in a table-like view selected data fields of selected search results) and Info Panel (provides information about all data fields of a selected search result). For information visualization and information discovery, two diverse metaphors have currently been implemented: (1) The Scatter Plot view (see left side of Figure 5) uses ordinate, abscissa, icons or pictures and their size to visualize four diverse dimensions of the multi-dimensional data in a two-dimensional plane. Users are free to choose any dimension from the data vector and assign them to one of the four visualization dimensions. (2) The VisIsland view (see right side of Figure 5) is based on the visualization principle to place graphical representations of data vectors in accordance to their similarity (for example by using the Cosine similarity coefficient) in a two-dimensional plane. Consequently, similar data vectors are spatially grouped together and form islands of similar objects in the graphical view. Both visualizations enable users to gain insights about aspects of one or several meetings such as activities of meeting participants, frequency of topics addressed in meetings, and the like. Further details about this visualization tool can be found in [Mader 2007].

The two Modeler units are in an experimental stage and the full integration within the system is not yet finished. Currently, a manually-started script exploits data from the xFIND Index's cache, generates statistic data and feeds the two Modeler instances. The User Modeler holds basic information about users and their presence or absence in meetings, as well as their roles and activities within the meeting. The Meeting Modeler keeps summarized information about the meetings such as meeting duration, location, meeting topics and the like. A Web interface built on top of the servelet container Apache Tomcat (see for example in [TOMCAT]) allows users to browse and search for meeting and user information.

Figure 5: Meeting information visualization based on the scatter plot metaphor (left side) and the VisIsland methaphor (right side).

3.3 Concept-based Access for Knowledge Transfer and Learning Activities

The MISTRAL meeting information retrieval and visualization application enables users to search, render and make specific information accessible in employees' daily working processes. However, the solution proposed so far is not fully integrated in these daily business processes, and employees need to access an extra application. In order to overcome this, we have also designed a solution for the easy integration of meeting information into employees' working environments. The solution is aimed at simplifying knowledge access and stimulating additional learning activities based on "real life business" information from meetings.

Figure 6 outlines our proposed solution based on a context-dependent concept-based access to meeting information. This solution enables users to retrieve contextualized meeting information according to their working tasks without leaving their familiar daily working environment. Thus, users can interact with information systems (such as knowledge management systems, workflow systems and learning management systems) and retrieve requested information in the usual way based on specific business cases as well as user and group roles. Additionally, our solution offers context-based and personalized related meeting information on the logical level of concepts to the users. Each of the concepts represents specific information from the meeting archive and each is linked with a proper search query for the Meeting Information Retrieval and Visualization system. By selecting one of these concepts, a request is sent to the MISTRAL system and a dynamically compiled, up-to-date set of search results is delivered to the user. To illustrate this, consider a user who has enrolled in a course at the corporate learning management system. In order to contextualize learning content within the learning process, not only course content is delivered, but information about practical experiences from corporate memory as well. By applying our approach, personalized and related concepts are offered in addition to the learning content for retrieving information about meetings. One example is given in Figure 7: a user has requested information about the topic “speech recognition” from the learning management system. Additionally, related concepts for explorative learning activities on meeting information are provided as background knowledge in a tree-view-like sidebar. This tree-view representation enables users to request information about meetings where the user was present or absent for the given topic and projects what the user was involved or not involved in for the given topic. It also provides information about the meeting participants who actively spoke about the specific topic addressed by the concept. Moreover, users can request meeting-relevant documents dealing with the given topic and information about experts who are often involved in the specific topic.

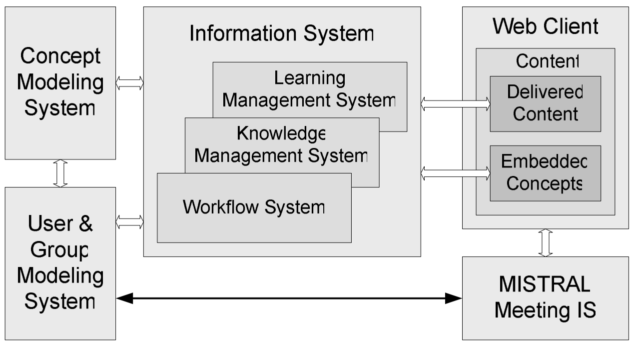

Figure 6: Concept-based access to the MISTRAL meeting information system from users’ daily working environments

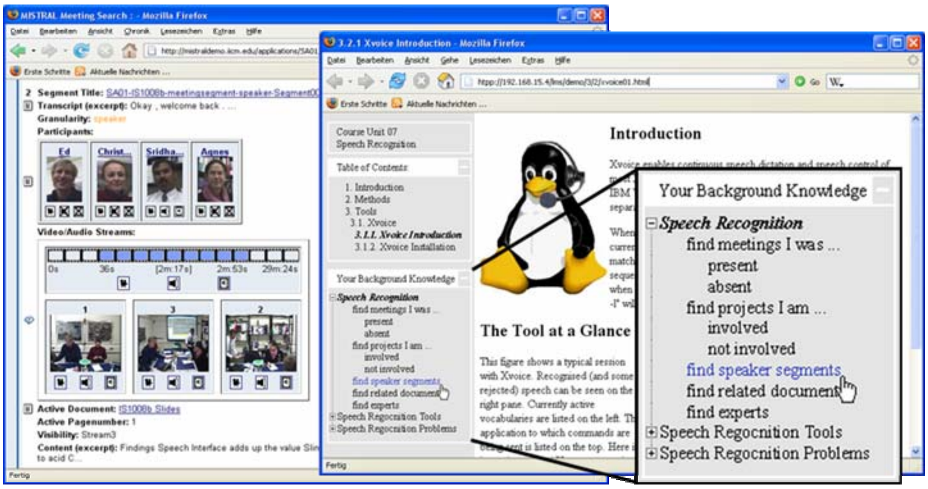

Figure 7: Integration of "real business life" information in a corporate learning environment

Focusing on the technical design of the proposed solution, Figure 6 outlines the architectural overview. At the outer-left side, the core units are responsible for providing context-dependent and personalized concepts. The User & Group Modeling System holds most information about users, such as roles, memberships to user groups and projects. Meeting-relevant user information is fed by the MISTRAL Meeting IS, such as presence or absence in meetings and the like. The Concept Modeling System enables users to manage and deliver context-dependent personalized concepts. For each context, a number of context items can be assigned and concepts can be linked to one or more concept items. Each of the concepts is linked with one or more queries in order to request background information from information systems, such as the MISTRAL Meeting IS. To illustrate this, let us assume that one workflow managed by a workflow system is considered to be one context and each of the tasks in the workflow is represented by one context item. Based on this, a number of relevant concepts for background knowledge can be defined for each of the tasks. To simplify the creation of queries, generic query templates with placeholders can be defined. For each new concept, the pre-defined query templates are instantiated with information from the concrete concept description (name of concept or synonyms). In order to personalize the queries, the same mechanism is used and placeholders of the query templates are replaced by user and group information (delivered by the User & Group Modeling System). This approach supports personalized information requests, such as "show me meetings about topic x I was present". The information flow for the users is as follows: they interact with information systems by a Web client, and according to their requests, information is delivered to and rendered by the Web browser. Additionally, personalized and context-dependent concepts are delivered to and rendered by the client. There are two possible approaches. First, the Concept Modeler gets the information about the context and the sub-context (context item) as well as user information by the information system and the relevant concepts are delivered via the information system to the user client. Second, the user client implements a XUL application (XML User Interface Language, see also [XUL]) and sends the aforementioned information directly to the Concept modeller and renders the information within a separate window. Finally, users can request background information from the MISTRAL Meeting IS by selecting concepts offered by the system. In our first experimental implementation, a simple learning management application has been prototyped which sends context and user information to the Concept Modeler and integrates the retrieved information within the content pages for the users, see also Figure 7. The Context-Sensitive Concept Modeler, developed at the IICM (see for example [Gütl et al., 2006]) and MISTRAL User Modeler have been used for managing and delivering personalized concepts.

4 Lessons Learned

The aim of this section is to outline the most important findings and areas for improvement from our informal qualitative evaluation. In this process, members of the research team, other domain experts, and students were involved. Moreover, within the exploitation work package of the MISTRAL project, feedback has been collected from application providers and vendors as well as from their customers in the domain of knowledge management and meeting support systems.

4.1 MISTRAL Core System

From the architectural point of view, the service-oriented architecture (SOA) is very promising in terms of flexibility, extensibility, exchangeability and scalability. In particular, SOA supports modern business strategies and enables the integration of the MISTRAL system or selected MISTRAL units into companies' existing IT infrastructure. However, servicing a completely distributed system in terms of distributed data processing and information extraction is restricted because of the high network traffic caused by moving large datasets of meeting recordings. To overcome that, research for an agent-based approach by applying mobile agents to reduce network traffic and enable flexibility is of great interest.

By focusing on the sound processing unit, MISTRAL's setup using an array of far field microphones has resulted in the advantages that meeting participants can move and act dynamically within the meeting environment. However, processing and information extraction tasks from signals of far-field microphones are faced with echo effects and higher intensity of ambient noise. Near-field or body-mounted microphones can improve the situation and this setup will also satisfy most business applications. Another interesting research approach is to post-process the signals in order to build a virtual directional microphone focusing on the spatial area of interest, e.g. to the current active speaker. The performance of speech-to-text transcripts based on speaker-independent methods applied on noisy signals was very poor. This is contrasted by business needs, which are of particular interest in this part of the MISTRAL application. Further research of interest in this area includes the application of user-dependent methods and meeting-specific (topic) dictionaries.

From the video-processing viewpoint, an improved and robust method has been developed for the MISTRAL project, which can not only detect faces but also track these detected faces. This feature enables further training data collection for the improvement of the face recognition process, i.e. it provides additional data for improving underlying model for recognition process. Main problems are caused by poor light conditions and low pixel resolution (e.g. participant is far away and face covers only a small area of the video). Further research of interest for "electronically" zooming into the most interesting parts of meeting activities in real-time is motivated by concrete business application needs in order to support remote meeting participants focusing on main activities.

By focusing on the sensor data processing unit, information about meeting participants' interaction with the presenter PC provides helpful information about documents used in meetings (e.g. presentation slides, Web pages and documents from corporate knowledge space). Additionally, topics and main concepts currently talked about can be inferred by means of the text processing unit. Personal note taking and documents used within the meeting by meeting participants' notebooks are indeed of interest, but they also raise great security and privacy aspect concerns.

The text-processing unit is mainly applied to extract named entities and key concepts from textual information, such as speech-to-text transcripts, presentation slides and meeting-relevant documents. From the business application viewpoint, such extracted information is of great interest in order to enrich meeting sequences with metadata which will subsequently be indexed. Furthermore, extracted information from presenter PC and the meeting domain knowledge are of interest for improving the extraction methods and therefore speech-to-text transcripts.

From the viewpoint of the merging unit, it is particularly challenging to merge fuzzy temporal and spatial co-occurrence of information from diverse modalities. However, the fusion of comparable semantic information from different modalities can be used to boost or weaken the confidence level of semantic information extracted by the unimodal units. Based on former extracted information, the semantic enrichment unit focuses on the inference of higher semantic actions and concepts as well as on detection of contradictions. From this perspective, it is very complex to model such actions and concepts, and it requires the ability to compromise settlements. For example, it is not clear how to model the difference between a short interruption in a presentation and a discourse. Both units deliver their results either in a MISTRAL-specific XML format or it can also be transformed into the MPEG-7 data format. Specifying the data structure has taken a lot of effort. After evaluating some alternatives, we have decided to follow an event-based notation. Data structure mapping into the MPEG-7 format was again very complex. We have encountered interoperability problems because of the openness of definitions in the standard. To overcome this, a mapping was defined, which builds on the MPEG-7 Detailed Audiovisual Profile, see also [Bailer et al. 2006].

Focusing on the data repository, not only access rights are important for this unit, but also privacy aspects and legal problems may become an issue in practical applications. According to business application requirements, the simple centralised management of meeting recordings and meeting-related documents in file hierarchies need to be improved in order to support multi-media databases and existing document repository infrastructure.

4.2 MISTRAL Semantic Application

From the architectural viewpoint, our approach for meeting information retrieval and visualization has been shown to be very flexible in terms of integrating different data sources for the indexing process and for accessing these different sources by the visualization tools. Moreover, meeting information can be integrated into other systems at different levels, from pre-processed data structures to meeting information retrieval results to meeting information visualization. This flexibility meets in particular business requirements for diverse application scenarios identified from the MISTRAL exploitation work package.

By focusing on the Preprocessor unit, it enables a flexible support (1) to handle diverse document and file formats, (2) to access different data repositories, and (3) to prepare a selective number of metadata (metadata fields as well as type and extent of content) on different granularities of the meeting (e.g. meeting segments and meeting agenda). Our approach builds on the xFIND Gatherer, which supports the retrieval of documents by different protocols such as http, ftp and file access. Different document types are processed by appropriate filters in order to extract metadata and content. For the MISTRAL project, specialized filters have been implemented to deal with the MISTRAL-specific file formats. The current solution has two drawbacks: first, document filters need to be implemented as an xFIND Gatherer component. Second, all data files from repositories need to be retrieved and pre-processed to extracting data of interest. This can cause heavy network loads in a distributed system, for example if video data from meeting recordings are used to produce thumbnails of meeting scenes. To overcome these drawbacks, other approaches need to be researched, such as applying mobile agents and the like.

Evaluation results in the context of the Meeting Information Indexer unit has shown that the indexing of and searching for meeting information at different granularities (entire meeting as well as segments of speaker activities, agenda topics and presentation slides) meet users information needs given by identified business application scenarios. However, specifically for speaker segment granularity, we have also encountered the problem that some segments contain minimal valuable content. To illustrate this, let us consider meeting participant A, who starts a sentence and will be interrupted by participant B, who finally takes over the discussion. Short statements that agree or disagree to previous statements are another type of problem. Such segments are useless without the context of the entire discussion. Within the scope of the MISTRAL retrieval and visualization application, this is still an open issue and it provides the possibility for further interesting research. The powerful search functions are also positively evaluated, such as searching in specific metadata fields (e.g. search only in speech-to-text field), Boolean search of specific metadata fields from different modalities (e.g. search for speaker segments having term ‘x’ in speech-to-text field and having term ‘y’ in presentation slide content), and filter functions for the search results (e.g. search for a term in speech-to-text and filter by one or more meeting participants). However, this great variety of retrieval functionality has also evaluated as too complex for simple users or employees. This has resulted in simplified user interfaces for the MISTRAL visualization tools and in the approach for concept-based access to meeting information, which is also discussed in the remainder of this section.

By focusing on our experiences of the MISTRAL Visualization Tools, we have started with the "search-engine" like approach in order to take the above evaluation results into account. Test users have agreed that the search forms keep the cognitive load for query formulation low. In particular, specialized search forms for meeting information, speaker segments, speaker information and agenda information are positively noted. Subjects have also very positively evaluated search results presentation providing most important information based on the meeting granularity level of interest. However, they also emphasized the problem that for some information needs, parts of the information rendered for a search result are not that important, and they suggested to improve the UI by simply enabling configuring the extent of the search results. Furthermore, we have also learned that some application scenarios require larger sets of meeting data, such as the need to find out how often meeting participants talk about a specific topic in all meetings for one project. This has motivated us to develop a meeting information visualization to provide comprehensive overviews and to enable users to discover information from different visualization metaphors. This part of the MISTRAL project has been recently finished but not yet evaluated.

The two instances of the Modeler unit, namely the User Modeler and the Meeting Modeler, have only been prototyped in a very experimental stage. First experiences have shown that comprehensive information about meetings and about meeting participants can be usefully integrated in meeting applications. This fits well also into identified business application needs. In order to take further advantage of the modeling systems within the MISTRAL application, further research work needs to be conducted to discover how to integrate user and group-specific meeting information into business-wide user and group profiles.

4.3 Concept-based Access for Knowledge Transfer and Learning Activities

The MISTRAL Meeting Retrieval and Visualization Application has been evaluated overall as quite good for enabling retrieval and access of relevant meeting information and for supporting the exchange of knowledge. It could even be used to access learning-relevant "real-life" business information, such as discussions of specific problems about a topic and decisions reached with how to handle similar problems and the like. However, a drawback has been identified in that the MISTRAL application is still an isolated application that needs to be actively accessed by employees. This finding is also supported in literature as outlined in 2.3. To overcome this problem, we have developed an approach for concept-based access to the MISTRAL Meeting Information Application from employees' daily working environments. Our experimental prototype implementation has shown the successful integration of meeting information as learning background information (additional learning content) within a learning environment. Furthermore, we have learned that the solution approach is also applicable for other information systems. The context-dependent and personalized offer of relevant additional content on the logical level of concepts has been seen as the major strength of the application. It supports explorative learning activities and enables employees or learners to contextualize main learning content by business-centric additional information. Additionally, hiding complex search queries for specific information needs by simply selecting corresponding concepts has been evaluated positively. Despite the simple way to create and manage concepts, the human effort has been seen as the major drawback of the approach. Based on that, research needs to be conducted to find out how to further support users or even automatically compile relevant concepts for such applications.

4.4 Further Remarks

As one of the concluding remarks, we want to stress the main advantages of the MISTRAL system compared to other meeting information systems: (1) flexible architecture based on the service-oriented approach, (2) feedback of extracted information from one unit to improve extraction performance form other units, (3) usage of content information from presenter PC to gain information about the agenda topic and to improve speech-to-text transcripts, (4) MPEG-7 support, (5) indexing and retrieval granularities based on the entire meetings, speaker segments, agenda topics and presentation slides, (6) simple “search-engine” like user interface and search results presentation providing most important information based on the meeting granularity level, (7) integration of the meeting application within daily business working environment, and (8) support for learning activities.

Within the exploitation work package, we have identified together with knowledge management vendors the following promising business application: (1) Completely Virtual meetings become increasingly important in modern business life and tools basically support these communication forms. However, there is still a lack in terms of ‘intelligently’ archiving the meeting information and to fully integrate this information in project management tools or in knowledge management systems. (2) Nevertheless, face-to-face meetings are still important because they are necessary to foster and preserve trust. However, it is very unlikely to bring together a large group of people at the same place because of today’s high business travel activities and distributed organizational structures. Therefore, a combination of face-to-face meetings with remote meeting participants often becomes what we call a “hybrid meeting”. A modern meeting information system can not only help to archive meeting information, but also support remote participants focusing on the main activities by simply transmitting the most important section of the video picture. This could be particularly helpful in case of using mobile devices and low bandwidth connections.

Finally, we have also learned within the MISTRAL project that such systems could raise problems concerning privacy aspects. We therefore recommend the implementation of policies for recording, archiving, accessing and analysing such information.

5 Conclusions and Further Work

In this paper we have focused on technology-enhanced meetings and meeting support. Based on the motivation for such systems, we have outlined a variety of tools and technologies supporting meetings through their every stage, namely pre-meeting, in-meeting, and post-meeting.

Technological meeting support has been an active research for decades, however, particular interest in meeting information systems has emerged within the last years. Such systems may include technology-enhanced meeting rooms as well as tools for multi-modal meeting recording, automatic meeting information extraction and annotation, in-meeting support, meeting information archiving, indexing, retrieval and visualization. Various research projects in Europe and in USA have focused on multimodal meeting recordings and on their semantic information extraction and retrieval. Despite these great interest and research activities, insufficient attention has been paid to flexible architectures, interchangeability of meeting information, and the integration of this meeting information into business processes and applications.

Research and development activities within the MISTRAL project have been directed to overcome the problems stated above. The two-year project resulted in an approach towards a flexible and extendable system that can be easily integrated into daily working environments. It enables users to search, retrieve and access relevant information on different granularities of meetings. First experiences have shown that the system can be applied to support knowledge transfer for various business processes, and that it can also be part of the corporate learning environment. Overall, first experiences are promising, however, a great deal of further research can still be conducted.

For future work, we intend to evaluate two aspects in detail: first, the application of meeting information for corporate learning activities in a real-life setup, and second, the usability of our implemented meeting information visualization metaphors. Furthermore, we want to research enhanced methods for speech-to-text transcripts by taking into account speaker characteristics, knowledge domain of the meeting and information from the presenter PC. We are also interested in applying the system for processing recorded lectures (e.g. MIT open courseware repository) to make them searchable and accessible on a semantic level. Finally, together with a big software developing company on knowledge management system, we are currently working in the early planning stages of a concept being designed to integrate parts of our research results with the improved archiving and retrieval features of the company's virtual meeting tool.

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 author likes to acknowledge the valuable contributions of all Institutions involved and the members of the MISTRAL team. In particular, the support of following IICM members is gratefully acknowledged: Victor Garcia, Helmut Mader, Martin Ruhmer, and Christian Safran.

We also grateful acknowledge the following resources for meeting data provision: (1) The AMI Meeting Corpus, which have been published under the Creative Commons Attribution NonCommercial ShareAlike 2.5 License. (2) I. McCowan, S. Bengio, D. Gatica-Perez, G. Lathoud, F. Monay, D. Moore, P. Wellner, and H. Bourlard, "Modeling Human Interaction in Meetings", in Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Hong Kong, April 2003. For further information see also http://www.m4project.org and http://mmm.idiap.ch.

References

[AMI] AMI: “AMI Overview”, Official Website, AMI Consortium, last retrieved 2007-03-26 from http://www.amiproject.org/ami-scientific-portal

[AMI 2006] AMI: “AMI overview and prospects for future research”; State of the Art White Paper, AMI Consortium, January 2006, last retrieved 2007-03-26 from http://www.amiproject.org/ami-scientific-portal/documentation/annual-reports/pdf/AMI-overview-and-prospects-for-future-research-Jan2006.pdf

[AMIDA] AMIDA: “AMIDA Overview”, Official Website, AMI Consortium, last retrieved 2007-03-26 from http://www.amiproject.org/amida-scientific-portal/overview

[AMIDA 2007a] AMIDA: “State-of-the-art overview: Localization and Tracking of Multiple Interlocutors with Multiple Sensors”, Technical Paper, AMIDA Consortium, January 2007, last retrieved 2007-04-03 from http://www.amiproject.org/ami-scientific-portal/documentation/annual-reports/pdf/SOTA-Localization-and-Tracking-Jan2007.pdf

[AMIDA 2007b] AMIDA: “State-of-the-art overview: Conversational multi-party speech recognition using remote microphones”; Technical Paper, AMIDA Consortium, January 2007, last retrieved 2007-04-03 from http://www.amiproject.org/ami-scientific-portal/documentation/annual-reports/pdf/SOTA-Conversational-multiparty-ASR-using-remote-mics-Jan2007.pdf

[AMIDA 2007c] AMIDA: “State-of-the-art overview: Recognition of Attentional Cues in Meetings”; Technical Paper, AMIDA Consortium, January 2007, last retrieved 2007-04-03 from http://www.amiproject.org/ami-scientific-portal/documentation/annual-reports/pdf/SOTA-Focus-of-Attention-Jan2007.pdf

[Antunes et al. 2005] Antunes, P., Carriço, L.: “Modeling the Information Structures of Meetingware” Sistemas de Informação, ournal of the Portuguese Association of Information Systems, vol. 17, pp. 21-37, 2005.

[APACHE] APACHE: Apache HTTP Server. The Apache Software Foundation, last retrieved 2007-08-01 from http://projects.apache.org/projects/http_server.html

[Bailer et al. 2006] W. Bailer, P. Schallauer: The Detailed Audiovisual Profile: Enabling Interoperability between MPEG-7 Based Systems. In: 12th International Multi-Media Modelling Conference Proceedings, 2006.

[Barthelmess et al. 2005] Barthelmess, P., Ellis, C.A.: “The Neem Platform: An Evolvable Framework for Perceptual Collaborative Applications”; Journal of Intelligent Information Systems, Springer Netherlands, Volume 25, Number 2 / September 2005, pp. 207-240.

[Burnett et al. 2003] Burnett, I., V. d. Walle, R., Hill, K., Bormans, J., Pereira, F.: “MPEG-21: Goals and Achievements”; In IEEE Multimedia, Vol. 10, Oct-Dec 2003, pp. 60-70.

[CALO 2006a] CALO: “About CALO”; SRI International, last modified 2006, last retrieved 2007-03-31 from http://caloproject.sri.com/about/

[CALO 2006b] CALO: “CALO: Cognitive Agent that Learns and Organizes”; SRI International, last modified 2006, last retrieved 2007-03-31 from http://caloproject.sri.com/

[CHCC] CHCC: “CALO”; Center for Human-Computer Communication, OGI School of Science & Engineering, Beaverton, USA, last retrieved 2007-04-01 from http://www.cse.ogi.edu/CHCC/Projects/CALO/main.html

[CHIL 2005a] CHIL: “Fact Sheet Software Architecture”; CHIL Project Consortium, last modified May 2005, last retrieved 2007-03-31 from http://chil.server.de/servlet/is/5270/software_architecture.pdf?command=downloadContent&filename=software_architecture.pdf

[CHIL 2005b] CHIL: “Deliverable D7.4 Evaluation Packages for the First CHIL Evaluation Campaign”; Technical Paper, CHIL Consortium, March 2005, last retrieved 2007-04-03 from http://chil.server.de/servlet/is/2712/CHIL-D7.4-EvalPackages-V2.1-2005-02-28-PU.pdf?command=downloadContent&filename=CHIL-D7.4-EvalPackages-V2.1-2005-02-28-PU.pdf

[CHIL a] CHIL: “Project Description”; Original Project Website, CHIL consortium, last retrieved 2007-03-31 from http://chil.server.de/servlet/is/104/

[CHIL b] CHIL: “Scenarios”; Original Project Website, CHIL consortium, last retrieved 2007-03-31 from http://chil.server.de/servlet/is/105/

[Cook et al., 1987] P. Cook, C. Ellis, M. Graf, G. Rein, T. Smith: “Project Nick: meetings augmentation and analysis”; ACM Transactions on Office Information Systems, 5, 2, 1987, 132-146.

[Costa et al., 2001] C.J. Costa, P.A. Antunes, J.F. Dias, A Model for Organisational Integration of Meeting Outcomes. In Maung K. Sein, and others (Eds.), Contemporary Trends in Systems Development. Kluwer Plenum, 2001. (Papers from the Ninth International Conference on Information Systems Development, ISD 2000. ISBN: 0-306-46608-2).

[Ellis et al. 2003] Ellis, C., Barthelmess, P.: "The neem dream”; Proceedings of the 2003 conference on Diversity in computing (TAPIA '03), Atlanta, Georgia, USA, 2003, pp. 23-29.

[Fiscus et al. 2005] Fiscus, J.G., Radde, N., Garofolo, J.S., Le, A., Ajot, J., Laprun. C.: “The Rich Transcription 2005 Spring Meeting Recognition Evaluation”; Rich Transcription 2005 Spring Meeting Recognition Evaluation Workshop, Edinburgh, UK, July 2005, last retrieved 2007-04-03 from http://www.nist.gov/speech/publications/papersrc/rt05sresults.pdf

[García-Barrios et al., 2006] García-Barrios VM, Gütl C: Semantic Applications on MPEG-7 Descriptions of Multi-modal Meeting Corpora: First Results. Special Issue "Multimedia Metadata Community Workshop Results 2005 Dissemination", JUKM, Vo. 1, No. 1, pp. 45-53.

[García-Barrios, 2007] García-Barrios VM: Personalisation in Adaptive E-Learning Systems. A Service-Oriented Solution Approach for Multi-Purpose User Modelling Systems. Dissertation, Graz University of Technology, May 2007.

[Geyer et al., 2005] W. Geyer, H. Richter, G.D. Abowd, Towards a Smarter Meeting Record—Capture and Access of Meetings Revisited. Springer Science + Business Media, Multimedia Tools and Applications, 27, 393–410, 2005.

[Gruenstein et al. 2005] Gruenstein, A., Niekrasz, J., Purver, M.: “Meeting structure annotation: data and tools”; In Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue, Lisbon, Portugal, September 2005.

[Gütl 2002] Gütl C: Approaches in modern knowledge discovery in the Internet. Dissertation, Graz University of Technology, September 2002.

[Gütl et al., 2005] C. Gütl, V.M. García-Barrios, Semantic Meeting Information Application: A Contribution for Enhanced Knowledge Transfer and Learning in Companies. In Proceedings of 8th International Conference on Interactive Computer Aided Learning (ICL 2005), Villach, Austria, 2005

[Gütl et al., 2005b] C. Gütl, V.M. García-Barrios, MPEG-7 Requirements and Application for a Multi-Modal Meeting Information System: Insights and Prospects within the MISTRAL Research Project. Invited Talk at MPEG and Multimedia Metadata Community - 2nd Workshop, 5th International Conference on Knowledge Management I-KNOW'05, Graz, Austria, 2005

[Gütl et al., 2006] Gütl C, Safran C: Personalized Access to Meeting Recordings for Knowledge Transfer and Learning Purposes in Companies. In Proceedigns of m-ICTE 2006.

[Gütl 2007] Guetl C: Context-sensitive and Personalized Concept-based Access to Knowledge for Learning and Training Purposes. In Proceeding in ED-MEDIA 2007, Vancouver, Canada.

[Hayne 1999] S. C. Hayne: “The facilitators perspective on meetings and implications for group support systems design”; SIGMIS Database, 30, 3-4, 1999, pp. 72-91.

[HELIX] HELIX: The Helix DNA™ Server. Helix Community, last accessed 2007-08-01 from https://helix-server.helixcommunity.org/

[Ho et al., 1999] T. Ho, P. Antunes, Developing a Tool to Assist Electronic Facilitation of Decision-Making Groups. Proceedings of the String Processing and Information Retrieval Symposium & International Workshop on Groupware, 1999, 243 – 253.

[ICSI 2006] ICSI: “Meeting recorder Project”; International Computer Science Institute, last modified 2006-04-03, last retrieved 2007-03-27 from http://www.icsi.berkeley.edu/Speech/mr/

[IM2 2005] IM2: “Research”; Official Website, IDIAP Research Institute, last modified 2005-11-08, last retrieved 2007-03-26 from http://www.im2.ch/research/

[IM2 2007] IM2: “Portrait”; Official Website, IDIAP Research Institute, last modified 2007-03-23, last retrieved 2007-03-27 from http://www.im2.ch/portrait/

[Ingirige et al. 2002] Ingirige, B., Sexton, M., Betts, M.: “The suitability of IT as a tool to facilitate knowledge sharing in construction alliances”; International Council for Research and Innovation in Building and Construction, CIB w78 conference 2002, June 2002, last retrieved 2007-04-04 from http://itc.scix.net/data/works/att/w78-2002-85.content.pdf

[Inoue et al. 1997] Inoue, T., Okada, K., Matsushita, Y.: “Integration of face-to-face and video-mediated meetings: HERMES”; In Proceedings of the international ACM SIGGROUP conference on Supporting group work (GROUP '97), Phoenix, Arizona, 1997, pp. 405-414.

[ISL] ISL: “Meetign Room Project”; Official Website, Interactive systems laboratories, last retrieved 2007-03-27 from http://penance.is.cs.cmu.edu/meeting_room/intro.html

[Jain et al., 2003]: R. Jain, P. Kim, Z. Li: „Experiential meeting system”; In Proceedings of the 2003 ACM SIGMM workshop on Experiential telepresence (ETP '03), Berkeley, California, 2003, pp. 1-12.

[Jovanovic 2003] Jovanovic, N. Recognition of meeting actions using information obtained from different modalities -a semantic approach-. Technical Report TR-CTIT-03-48 Centre for Telematics and Information Technology, University of Twente, Enschede, 2003, last retrieved 2007-04-03 from http://eprints.eemcs.utwente.nl/5806/01/000000de.pdf

[Kim 2007] Kim, P.: “EMS:Experiential meeting system”; Official Project Website, last modified 2007-02-04, last retrieved 2007-03-31 from http://www.pilhokim.com/index.php?title=EMS:Experiential_meeting_system

[Lalanne et al. 2005] Lalanne, D., Lisowska, A., Bruno, E., Flynn, M., Georgescul, M., Guillemot, M., and et.al.: “The IM2 Multimodal Meeting Browser Family”; Technical report, March 2005, last retrieved 2007-04-03 from http://www.issco.unige.ch/projects/im2/mdm/docs/im2browsers-report-march2005.pdf

[M4] M4: “M4 Project”, M4 Project Website, last retrieved 2007-03-26 from http://www.dcs.shef.ac.uk/spandh/projects/m4/

[M4 2005] M4: “Multimodal Meeting Manager”; Deliverable D4.3: Report on Final Demonstrator and Evaluation, M4 Consortium, February 2005, last retrieved 2007-03-26 from http://www.dcs.shef.ac.uk/spandh/projects/m4/publicDelivs/D4-3.pdf

[Mader 2007] Mader H: Visualizing Multidimensional Metadata. Development of the Visualization Framework MD2VS. Master’s thesis, Graz University of Technology, March 2007.

[Martinez et al. 2002] Martinez, J.M., Koenen, R., Pereira, F.: “MPEG-7: The Generic Multimedia Content Description Standard, Part 1”; In IEEE Multimedia, Apr. 2002, 78-87.

[Martinez 2002] Martinez, J.M.: “MPEG-7: Overview of MPEG-7 Description Tools, Part 2”; In IEEE Multimedia, Jul. 2002, 83-93.

[MISTRAL] MISTRAL, official website, last received 2007-01-10, http://mistral-project.tugraz.at/

[MMP 2003] MMP: “Mapping Meetings project”, Official Website, Columbia University, last modified 2003-08-11, last retrieved 2007-03-27 from http://labrosa.ee.columbia.edu/mapmeet/

[Munkvold et al., 2001] B.E. Munkvold, R. Anson: “Organizational adoption and diffusion of electronic meeting systems: a case study”; In Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work (GROUP '01), Boulder, Colorado, USA, 2001, pp. 279-287.

[NIST 2006] NIST: “NIST Meeting Recognition Project: Overview”; National Institute of Standards and Technology, last modified 2006-12-22, last retrieved 2007-03-31 from http://www.nist.gov/speech/test_beds/mr_proj/index.html

[Nunamaker et al., 1991] J.F Nunamaker, A.R. Dennis, J. S. Valacich, D.R. Vogel, J. F. George, Electronic meeting systems to support group work: theory and practice at Arizona. Communications of the ACM, 34, 7, 1991, 40-61.

[Nunamaker et al., 1997] J.F. Nunamaker, R.O. Briggs, D.D. Mittleman, D.R. Vogel, Lessons from a Dozen Years of Group Support Systems Research: A Discussion of Lab and Field Findings. Journal of Management Information Systems, 1997.

[Nunamaker et al., 2001] J.F. Nunamaker, N.C. Romano, R.O. Briggs, A Framework for Collaboration and Knowledge Management. Proceedings of the 34th Hawaii International Conference on System Sciences, 2001.

[Page et al. 2005] Page, K.R., Michaelides, D.T., Buckingham Shum, S.J., Chen-Buger, Y.-H., Dalton, J., Eisenstadt, M., and et al.: “Collaboration in the semantic grid: a basis for e-learning”; Applied Artificial Intelligence, 19 (9-10), 2005, pp. 881-904.

[Pallotta et al., 2004] V. Pallotta, H. Ghorbel, A. Ballim, A. Lisowska, S. Marchand-Maillet, TOWARDS MEETING INFORMATION SYSTEMS: Meeting Knowledge Management. International Conference on Enterprise Information Systems (ICEIS 2004), 2004.

[Panda 2005] D. Panda, An Introduction to Service-Oriented Architecture from a Java Developer Perspective. ONJava.com, O’Reilly, last edit Jan 26 2005, last visit Aug 1 2007, from http://www.onjava.com/pub/a/onjava/2005/01/26/soa-intro.html

[Richter, 2005] H. Richter, Designing and Evaluating Meeting Capture and Access Services. Dissertation, Georgia Institute of Technology, April 2005, last received 2007-01-09, http://www.smartech.gatech.edu/bitstream/1853/6933/1/richter_heather_a_2000505_phd.pdf

[Rienks et al., 2007] R. Rienks, A. Nijholt, P. Barthelmess, Pro-active Meeting Assistants: Attention Please! *AI & Society*, The Journal of Human-Centred Systems. Springer-Verlag London Ltd, 2007, to appear.

[Romano et al., 2001] N.C. Romano, J.F. Nunamaker, Meeting Analysis: Findings from Research and Practice. Proceedings of the 34th Annual Hawaii International Conference on System Sciences, 2001.

[Sabol et al., 2005] V. Sabol, M. Granitzer, K. Tochtermann, W. Sarka, MISTRAL – Measurable, Intelligent and Reliable Semantic Extraction and Retrieval of Multimedia Data. 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies, London, UK, 2005

[Sabol et al., 2007] Sabol V, Gütl C, Neidhart N, Juffinger A,Klieber W, Granitzer M: Visualization Metaphors for Multi-modal Meeting Data. In Proceedings of the 6th Workshop of the Multimedia Metadata Community (WMSRM07), Aachen, Germany.

[SAILLABS] SAILLABS, Speech Analytics: Maximizing the Value of Customer Communications. SAILLABS Technology, last visit 2007-07-15, http://www.sail-technology.com/index.html?solutions/html/cm_sa.html

[Steinfield et al., 2002] C. Steinfield, C. Yang, J.M. Huysman, K. David, J. Lloyd, et al., Communication and collaboration processes in global virtual teams. INTEnD report, Michigan State University, East Lansing, Michigan, 2002, last received 2007-01-08, http://cscw.msu.edu/papers/INTEnD_Summary.pdf

[TOMCAT] TOMCAT: Apache Tomcat. The Apache Software Foundation, last retrieved 2007-08-01 from http://tomcat.apache.org/

[Turoff et al., 1977] M. Turoff, S. Hiltz,. Meeting through your computer. IEEE Spectrum, May 1977, p. 58-64.

[Whiteside et al., 1988] J. Whiteside, D. Wixon, Contextualism as a world view for the reformation of meetings; In Proceedings of the 1988 ACM conference on Computer-supported cooperative work, 1988, p. 369 – 376.

[XUL] XUL: XUL (XML User Interface Language). Mozilla Developer Center, last visit 2007-08-08 from http://developer.mozilla.org/en/docs/XUL