

# The IM2 Multimodal Meeting Browser Family



Denis Lalanne<sup>4</sup>, Agnes Lisowska<sup>5</sup>  
Eric Bruno<sup>6</sup>, Mike Flynn<sup>3</sup>, Maria Georgescu<sup>5</sup>, Maël Guillemot<sup>3</sup>,  
Bruno Janvier<sup>6</sup>, Stéphane Marchand-Maillet<sup>6</sup>, Mirek Melichar<sup>2</sup>,  
Nicolas Moenne-Loccoz<sup>6</sup>, Andrei Popescu-Belis<sup>5</sup>, Martin Rajman<sup>2</sup>,  
Maurizio Rigamonti<sup>4</sup>, Didier von Rotz<sup>1</sup>, Pierre Wellner<sup>3</sup>

<sup>1</sup>EIF - Fribourg

<sup>2</sup>EPFL – LIA/CGC - Lausanne

<sup>3</sup>IDIAP Research Institute - Martigny

<sup>4</sup>University of Fribourg – DIVA

<sup>5</sup>University of Geneva – ISSCO/TIM/ETI

<sup>6</sup>University of Geneva – Viper/CVML

March 2005

## Abstract

The aim of this report is to describe the browsers that have been developed by various groups within the IM2<sup>1</sup> project, highlighting goals, design methodologies, key functionalities and evaluation methods used by each. The paper concludes with a tabular overview of the media, input and output modalities and special functionalities handled by each browser, as well as providing specific contact points and references.

## 1. Introduction

Within a project that treats as diverse a group of research topics as does IM2 and the multimodal meeting domain, it is not unreasonable to expect different research groups to develop means to access and manipulate the data that is of most relevance to them. In many cases, the interfaces that are developed serve as a view on a particular data collection, where the requirements of the interface are those of the researchers/developers themselves. We will call these research-driven browsers. In other cases, the goal of the interface is to facilitate access to data for external parties such as the general population who might be using a meeting browsing/searching application. We will call these user-driven browsers. These two types of browsers serve different purposes and are based on often divergent sets of user requirements, but in virtually all cases, their existence can serve complementary purposes.

Research-driven interfaces are important not only to facilitate the work of the researchers, but also to explore in a more constrained environment the various functionalities that can be associated with the manipulation of a particular data type. These functionalities, once in a stable state, can then be considered as a mechanism to port to user-driven interfaces, where while the underlying functionality may remain the same, the manner in which it is presented to the user might change. Furthermore, the existence of data-type specific browsers allows for the investigation of different modalities and their role in interaction with a specific data type. This work can later be used to strengthen the interaction in more general, user-driven browsers, making them a priori easier to use and more 'intelligent'.

User-driven interfaces on the other hand focus on the needs of everyday end-users. They should be designed to meet the needs of a more diverse audience and focus on finding an appropriate equilibrium between usability and user-friendliness of the system as a whole and the incorporation and representation of available technologies.

## 2. The Browsers

Each of the following sections describes one of the browsers, including the purpose of the browser, the main media and modalities used, key functionalities, testing and evaluation methodologies used, main labs involved, and a graphic of the interface.

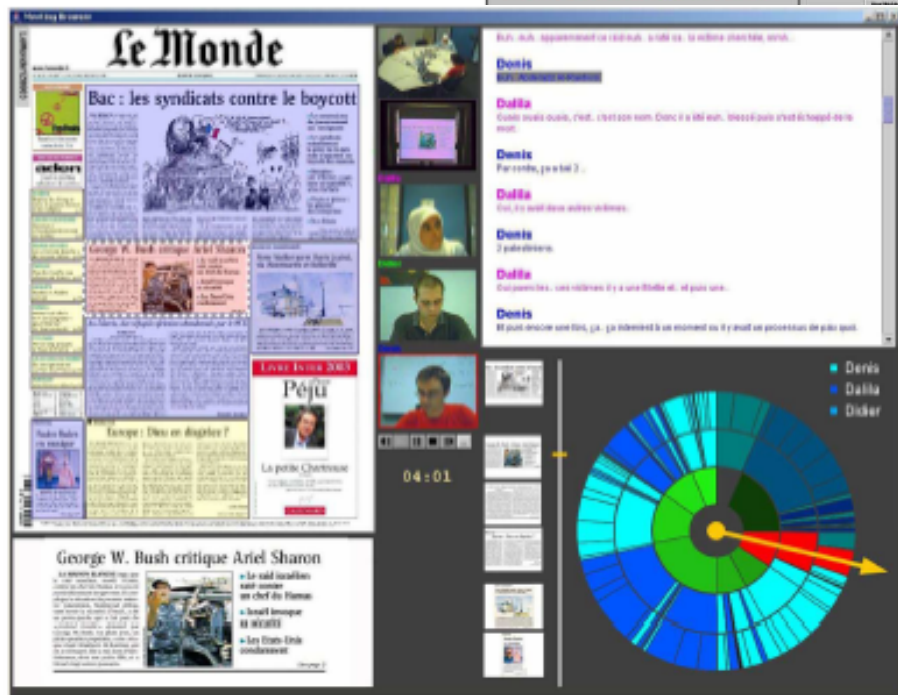
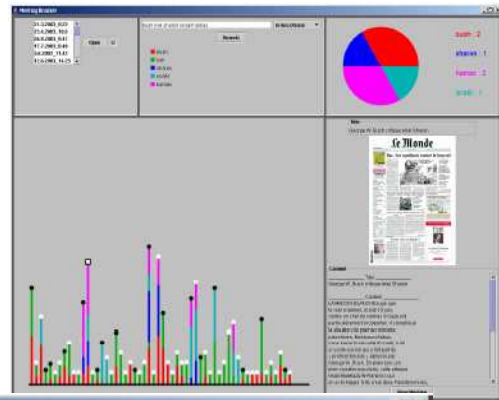
---

<sup>1</sup> The Swiss National Center of Competence in Research (NCCR) on Interactive Multimodal Information Management

## 2.1 FriDoc

### Introduction

Our assumption is that in a large proportion of multimedia applications (e.g. lectures, meetings, news, etc.), classical printable documents play a central role in the thematic structure of discussions [5]. The **FriDoc** browser, a document-enabled multimedia browsing system, proposes to consider static documents as structured and thematic vectors towards multimedia archives. The main purpose of the **FriDoc** browser is to assess the usefulness of document alignments, and more generally to evaluate the role of cross-media analysis, for multimedia browsing. Our hypothesis is that finding links between documents and multimodal annotations of the audio and video streams will permit the design of user-interfaces that improve retrieval tasks. Further, we believe that meeting browsers will be improved if they can visualize not only the different kinds of media that can be captured during a meeting, and their related annotations, but also visualize cross-media alignments.



### Description and key functionalities of the browser

Users can first search at a cross-meeting level, by typing a set of keywords, in order to retrieve all the relevant documents. Clicking on a document or an article allows users to view its content and its related annotations, e.g. period discussed or projected, keywords, authors, etc. Further, users can view the related multimedia data attached to

this document element in the intra-meeting navigator, in which all the components (documents, audio/video, transcription, annotations) are synchronized through the meeting time, thanks to the document alignments; clicking on one of them causes all the components to visualize their content at the same time. For instance, clicking on a journal article positions audio/video clips at the time when it was discussed, positions the speech transcription at the same time, and displays the document that was projected.

The sunBurst visualization at the bottom-right of Figure 1 represents the complete meeting's duration. It is a visual overview of the overall meeting and can serve as a control bar. Each layer of the disk stands for a different temporal annotation: speaker turns, utterances, document blocks and slides projected. Other annotations could be displayed depending on the meeting type (topics, silences, dialog acts, pen-strokes for handwritten notes, gesture, etc.). Those temporal annotations are currently stored in the form of XML files, which hold timestamps for each state change (i.e. new speaker, new topic, slide change, etc.) and spatial information for documents. For example, the speech transcript contains speaker turns, divided in speech utterances, with their corresponding start and end times. We believe that the sunBurst or other similar visualizations can reveal some potential relationships between sets of annotations, synergies or conflicts, and can bring to light new methods in order improve the automatic generation of annotations.

### **Testing and Evaluation**

22 meetings recorded in the Fribourg meeting room, of roughly 15 minutes each, and 1 meeting recorded in the IDIAP meeting room, have been integrated into our meeting browser, both at the cross-meeting and intra-meeting levels [4]. Based on those data, a preliminary user evaluation of this document-centric browser has been performed on 8 users. The goal was to measure the usefulness of document alignments for browsing and searching through a multimedia meeting archive. Users' performance in answering questions, both mono-modal and multimodal (e.g. "*Which articles from the New York Times have been discussed by Denis?*"), were measured on both qualitative and quantitative basis (e.g. task duration, number of clicks, satisfaction, etc.). Users browsing meetings using document alignments solved 76% of the questions and users browsing meetings without the document alignments solved 66% of the questions. The performance difference became particularly significant for multi-modal questions, i.e. requiring information both from the speech transcript and from document discussed or projected. In this case, around 70% of the questions were solved when users were benefiting from the alignments and only half of the questions were solved without the alignments.

### **Who, what and where**

The FriDoc interface is being developed at the University of Fribourg. New browsing interfaces and applications, e.g. for conference archives (SMAC) and more generally for digital libraries, are currently under study in collaboration with EIF, CERN and EPFL.

## **2.2 Archivus**

### **Introduction**

The Archivus system was conceived as a multimodal interface that allows users to access and interact with stored meeting data as naturally, flexibly and efficiently as possible. One of the primary goals is to determine which modalities in which combinations are most suitable for the five general tasks envisioned for this domain – a manager tracking employee performance, a manager tracking project progress, a new employee wanting to learn about a project, an employee who needs to catch up on a meeting they have missed, and an employee who needs to do some fact checking. Thus, the system was designed with more of a commercial application view in mind, rather than as a tool for researchers to continue their work.

The development of the Archivus system follows the traditional software design lifecycle, with a particular emphasis on user-centered design. The design is based on an analysis of user requirements elicited from real users [7] and will be revisited and if necessary modified as a result of testing and evaluation of the system. The cycle will be repeated until the system has reached a sufficiently high level of user satisfaction.

### **Media and modalities used**

The Archivus browser is intended to access all the possible media formats available within the IM2 project - video, audio and document representation (paper artifacts, slides, handwritten notes and visualization of whiteboard activity). Input modalities include simultaneous use of keyboard, mouse, touchscreen and voice, with possible extension to pen input in later iterations. Output modalities are graphics, text, audio (both for system prompts and data replay) and video (data replay only).

### **Key elements and functionalities**

The Archivus system allows user to both browse the meetings, or pinpoint specific elements of a meeting through interactive constraint satisfaction. To facilitate comprehension of the interface and the data that is accessible, we have couched the system in the metaphor of a person interacting in a library or an archive. The library represents the entire database of stored meetings, and each meeting is represented by a single book in that library. Some of the key functionalities of the interface are:

- Constant and customizable view of entire search space and its relation to current search results
- Search possible across whole database of meetings, not just a single meeting
- Dialogue based guidance for searches, with both auditory and visual system prompts
- A visual, chronological, interactive view of all constraints imposed on the system
- Oral and written natural language input is displayed back to the user and alterable in case of errors
- Searches on highly customized search spaces are possible via the Archivus notebook, which a user can also use to take personal notes or copy parts of meetings as they work for later reference
- Criteria refinement buttons guide users in their search definition and help them better understand the range of criteria that they can set for the system - date, place, participant, topic, dialogue element and event

- Each meeting is represented as an interactive book where the overall arrangement of a meet is represented in the standard structure of a book, the transcript of the meeting is the body of the book, and the multimedia elements of the meeting such as sound, video and accompanying documents are accessible through graphical icons logically placed in the margins of the book.

More details on these and other functionalities, including dialogue management can be found in [1, 6].

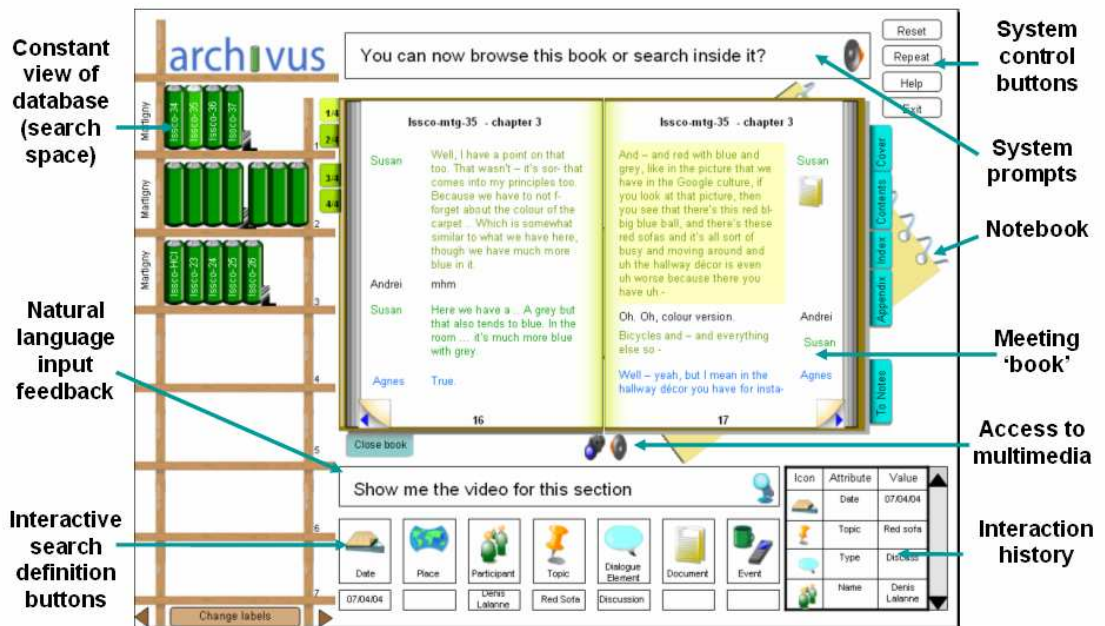


Figure 2: The Archivus system interface

### Testing and Evaluation

Due to high overhead in developing a fully functional user-centered multimodal system, the Archivus system will be tested iteratively using the Wizard of Oz methodology [2]. In Wizard of Oz experiments, users interact with a system, unaware that the system is not in fact fully operational. Instead, a Wizard sits in another room, watching the user and their interactions, and supplementing the functionalities that are missing in the system in real time. Thus, the user is given the impression of a fully and smoothly functioning system while developers are not bound to various constraints, and can explore the system needs/requirements and flaws more fully.

### Who, what and where

The Archivus interface is being developed by the IM2.MDM group in a joint effort between various individuals at ISSCO/TIM/ETI at the University of Geneva and LIA/CGC at the EPFL. Additionally, we are collaborating with IM2.DI to integrate novel features of interaction with documents.

## 2.3 Ferret

### Introduction

Understanding what users are interested in finding within recording meetings motivates the features we choose to implement, and help us to evaluate browser performance. The precise information that people find interesting obviously depends on the individuals, and on the specific meeting in question. Ferret is a framework for experimenting with alternative browser components. It is expected to be flexible in terms of interface customization and versatile to be able to cope with various input sources and data types. Originally dedicated for researchers in computational perception fields such as group activity modeling in meetings, we continue the development of Ferret in a user-driven methodology.

### Description and functionalities of the browser

Ferret works over the Internet, using a normal web browser as the client application. The main meeting corpus also resides on the Internet <http://mmm.idiap.ch>, along with much processed data – such as transcripts and recognition results. Here are two illustrations of Ferret. This first GUI (Figure 3) contains a media player for synchronized playback of audio and video, along with various temporal streams: speaker segments, transcript, slides. VCR controls, active areas and zoom in/out facility allows to jump quickly into moments of interest.

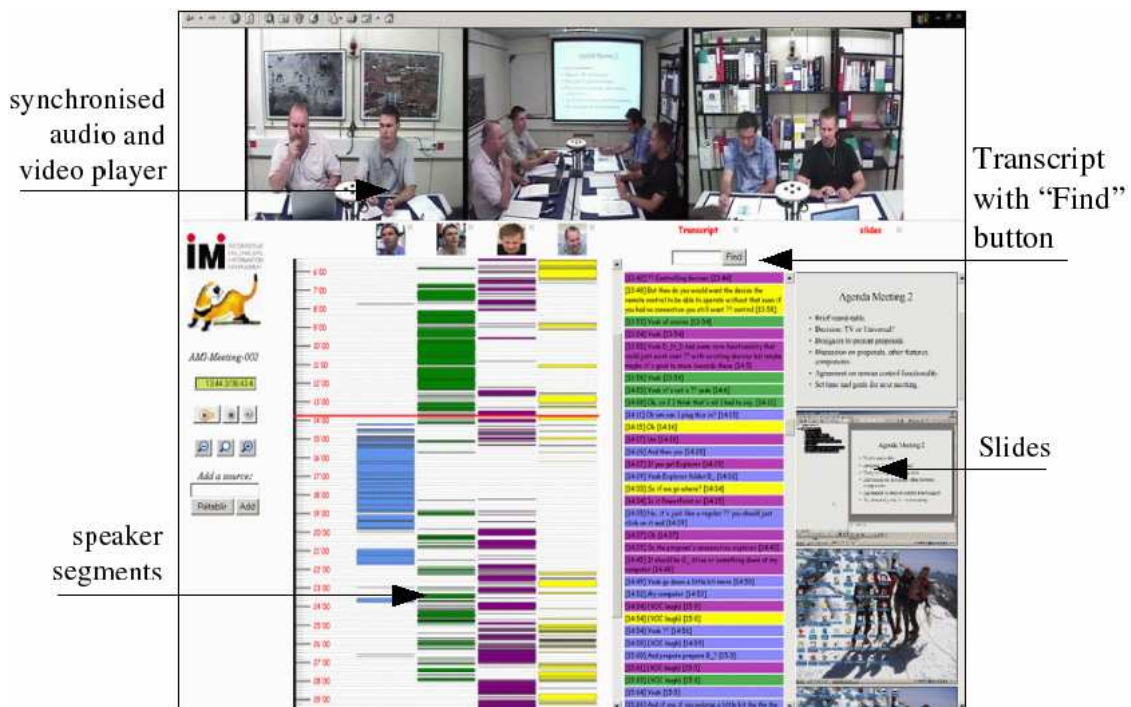


Figure 3 - The Ferret browser.

Another example (Figure 4) with audio conference call recordings where slides are shared by multiple parties remotely located.

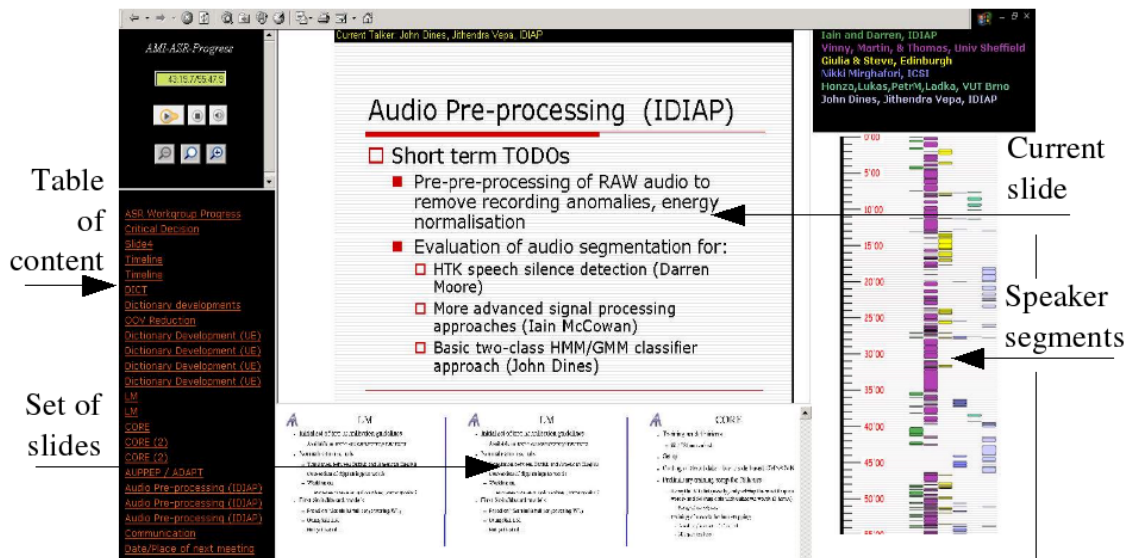


Figure 4 - Browsing audio conference calls with Ferret.

Key functionalities of Ferret (see [15] for more details) are flexibility in terms of interface customization and versatility to be able to cope with various input sources and data types. In addition to the data currently displayed by Ferret - audio streaming, video streaming, transcripts, slides, speaker turns, whiteboard strokes --, we are in the process of adding additional streams such as pen strokes, and representations of continuous data, such as energy levels or head positions. A new version written all in Java is being developed.

### Testing, evaluation and results

To evaluate Ferret (and other meeting browsers), we are developing a test that aims to be a) an objective measure of browser effectiveness based on user performance rather than judgement; b) independent of experimenter perception of the browsing task; c) able to produce directly comparable numeric scores, automatically; and d) replicable, through a publicly accessible web site. The browser evaluation test is detailed in [16], and depends on the collection of "observations of interest" made by people who view the meetings. Initial results show that speed and accuracy ratio are higher [0.6; 67.7%] using Ferret browser than in conditions where subjects have only access to a media player with basic CVR controls [0.52; 63.5%].

### Who, what and where

Ferret Meeting Browser is being developed in IDIAP Research Institute.

## **2.4 TQB: Transcript-based Query and Browsing Interface**

### **Introduction**

TQB is a transcript-based interface to meeting databases. The main idea behind TQB is that the content of the meeting (the meaningful pieces of information that constitute a meeting) is conveyed mainly by the meeting transcript. Therefore, search and browsing over an enriched dialogue transcript aligned with other media should be an efficient way to access a meeting.

### **Purpose**

TQB is developed to give straightforward access to a database of processed meetings. The generic goal is to help people who did not attend a meeting, or want to review it, to search for a particular piece of information. In its present state, TQB is mainly used for accessing the various dialogue annotations and for exploring the utility for browsing of different presentation styles of the enriched transcript and other media. TQB aims at showing the relevance of transcript-based interfaces for query and browsing.

TQB is constrained by the annotations that can be extracted from meetings. TQB is thus a compromise between utility, assessed by user studies [7, 8], and intrinsic interest assessed by its creators.

### **Media and modalities**

While many media (if not all) can be captured in a Smart Meeting Room, not all the conveyed modalities are equally relevant to meeting understanding. As meetings are interactions between rational agents, the highest informational content is conveyed by *language*, used in conversations (speech: audio or transcribed) and in meeting documents (reports, slides). Other relevant modalities are mimics (from video), positions of participants (from video), and emotions (from audio and video). Therefore, it makes sense to base the meeting rendering on speech, and transcripts are easier to grasp.

Several modalities can be used to query the meeting database. We chose for TQB a form-based approach, which provides simpler, though less flexible and less natural interaction. Moreover, the form-based approach is not *a priori* less efficient than a language-based one. Human-computer dialogue seems useful for sequences of query/answer pairs that refine an initial query, but TQB does not implement such a dialogue.

### **Meeting database used by TQB**

Design was constrained by the dialogue annotations targeted by IM2.MDM, within the shallow dialogue analysis model [12]. The annotations are done offline by automatic methods (nearly completed) and stored as XML annotations on transcripts (manual). Stylesheets are used to convert the XML data in order to feed the database and to produce an HTML enriched transcript (as “columns” or “lines”).

### **Actual functionalities of TQB**

TQB is a lightweight search and browsing interface, matching the MDM data model. TQB allows the user to search for a specific utterance, than view the surrounding meeting context. The user first selects a meeting based on its name only (number 1 in Figure 5), and the interface displays a form with the input fields that can be used, and their allowed values for the selected meeting (number 2). At present, these fields are: episode (pick one from list), speaker (same), dialog act (same), time interval (enter

start/end time) and words (search for a string). Once the form is filled, the user hits the Submit button (3).

The results of the query (4) are the utterances from the database that fulfill all the criteria entered by the user. These are hyperlinked to their respective positions in the dialogue transcript, so when the user selects one utterance by clicking on it, the surrounding dialogue context is displayed in the central frame (5), and the user can read it. Reading is facilitated as the transcript is speaker-segmented using colors, hyperlinked to the audio (play/pause by clicking on an utterance), and to the meeting documents – at present based only on referring expressions. Clicking on such an expression highlights the corresponding document element in the rightmost frame (6).

The form-based interface is implemented using JSP as a Web service client. The allowed arguments of the queries are computed online from the Postgres database. Some of the frames simply display HTML files (transcript, documents). Hyperlinks to the audio use Javascript calls to an embedded RealPlayer which could also display video.

### Plans for evaluation

TQB was tested and validated informally by its authors, and was demonstrated at several events (LREC '04, MLMI '04, UniGe Open Day). The performances depend greatly on the quality of the data. Therefore, we use manually annotated data (100% correct). To evaluate TQB on such data, we foresee user-based evaluation such as in BET [16]. Since TQB is web-based, it is only necessary to enter the new data and annotations (if available) into the database to make TQB ready for BET.

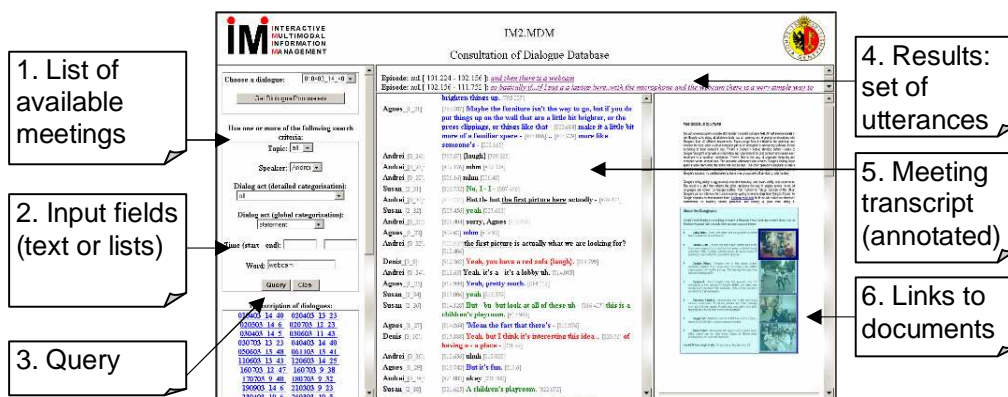


Figure 5: The TQB interface

### Who, what and where

TQB was developed within IM2.MDM at ISSCO/UniGe.

## 2.5 ViCoDE: Video Content Description and Exploration

### Introduction

The aim of ViCoDE is to propose a transversal approach to video document retrieval and management. As such, it proposes a collection-level management of the data, associated with content-based retrieval capabilities and document-level browsing features. ViCoDE is a general purpose video browser and is also well-suited for meeting data, as it is flexible in terms of the features used for document representation and also in terms of the level of understanding of each document. It is further able to handle parallel heterogeneous synchronized temporal streams (e.g. audio, video and text).

### Purpose

The primary goal of ViCoDE is to provide a query-by-example (QBE) retrieval engine on a large video document database. Search is performed by selecting video segments and marking them as positive or negative examples. The interface makes available a number of data feature spaces with respect to which similarity is to be defined. Data features used are typically based on color, motion, speech (audio signal and text transcript) and associated metadata. Based on the highlighted samples, a non-linear learning machine is then trained to define an adaptive multimodal similarity measure. This measure is used to sort the complete set of segments (collection) by similarity to the positive sample set. From there, segments are taken as initial points of a collection exploration. Further, the organization created under the interactive similarity measure definition allows for handing an ordered collection and to perform an organized annotation of the data.

The development purpose of ViCoDE is threefold. First, as a scientific prototype, it is a framework within which multimodal indexing and retrieval procedures may be tested interactively with relevant collection sizes. Second, it aims at demonstrating the feasibility of online large-scale content-based video information management. For that purpose, a specific storage backend has been developed to allow for full random access within video files [10]. Last but not least, ViCoDE is designed to accommodate several video collection navigation modes at various levels.

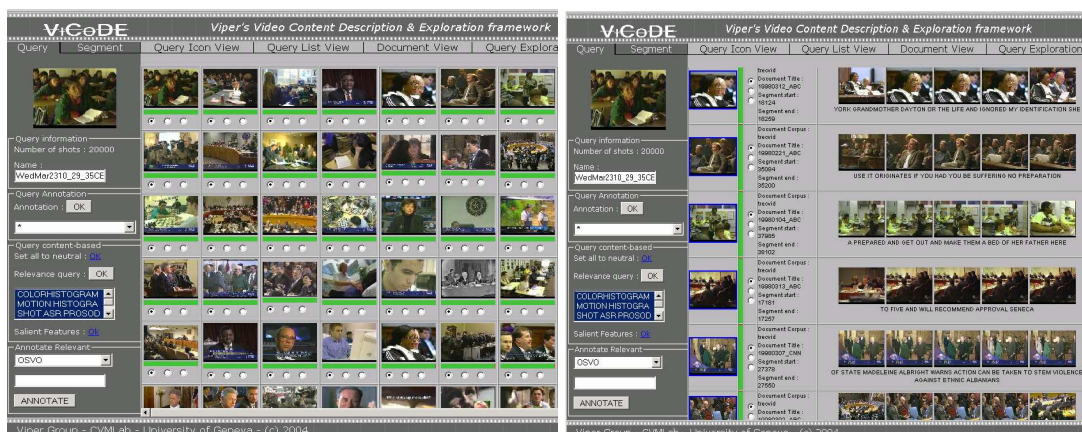


Figure 6: ViCoDE interfaces: (left) Query - (right) Local temporal browsing

## Targeted functionalities

ViCoDE is both a content-based query engine and a video browsing tool. Video documents are divided into relevant individual segments (e.g. shots or stories [3]). The query interface is proposing independent video segments (Figure 6 - left), whose features are the base for defining the similarity perception of the user from its input (over selected modalities). After retrieval, each individual matching segment becomes the starting point for a document exploration. This may first be done locally. The browser then displays row-wise keyframes and data associated to each individual segment, along with the last keyframe of the previous segment and the first keyframe of the next segment (Figure 6 - right). Browsing may also be done document-wise. The complete document is represented by a series of keyframes at the chosen scale. The active segment is located within the document by highlighting its keyframe using a fisheye effect (Figure 7).

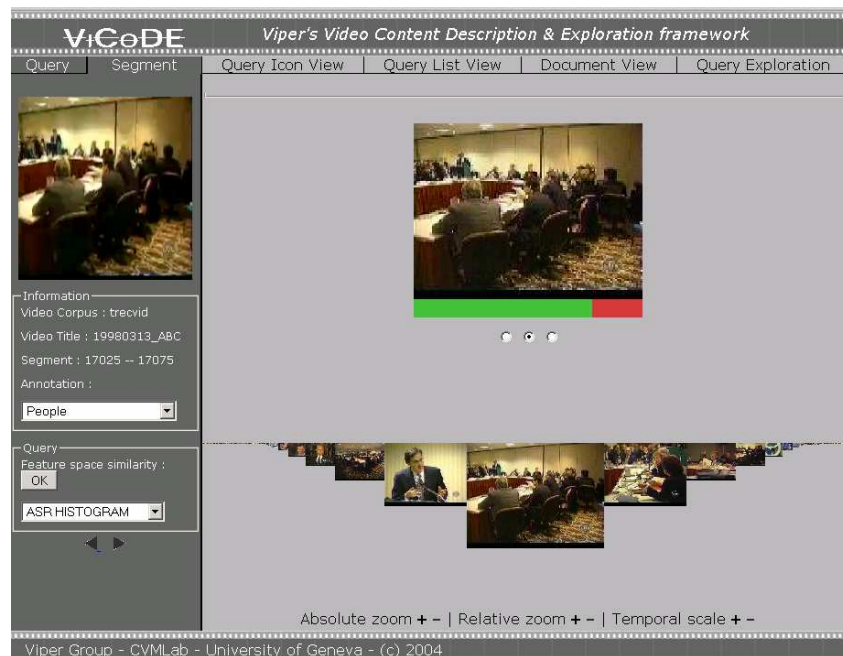


Figure 7: Document browsing interface using a fisheye effect on a series of keyframes

## Evaluation

ViCoDE's retrieval capabilities (*i.e.* collection browsing features) are evaluated against predefined ground truth obtained thru our active participation in the TRECVID Video retrieval and exploration contest. TRECVID also provides a thorough framework for the unified evaluation of exploration and retrieval video management systems. Locally, the evaluation of the document browsing capabilities of ViCoDE are under investigation in collaboration with the development of an online video service in UniGe's library. Document fast browsing interfaces were tested against actual naïve users. New tests will be initiated (both locally and in the TRECVID context) when the interface will be fully powered with our novel Collection Guiding capabilities [9].

## Who, what and where

ViCoDE is developed within IM2.IIR at *Viper/CVML/UniGe*.

### 3. Conclusions

As the descriptions above indicate, the interfaces being developed serve quite diverse purposes and have arisen from different research needs. They do however have an important element in common at a more abstract level. Notably, four of the five browsers (FriDoc, Archivus, TQB and Ferret) fall under the category of discourse browsers following the Tucker and Whittaker classification [14]. Discourse browsers use derived data streams [5] such as transcripts and documents, possibly in combination with raw data, to drive the browsing mechanism, which implies a higher degree of complexity at the interface level.

In order to have useful derived data streams for retrieval by the interfaces, processing on the raw data must be segmented into homogenous regions, and annotated, where the annotations serve as a relation between the semantic meaning of the annotation itself and the time-stamp reference to the raw data in the meeting. The levels of annotation utilized by the various browsers, as well as a more general comparative summary of the key aspects of each interface are outlined in the table below.

<b>Media handled</b>			<b>FriDoc</b>	<b>Archivus</b>	<b>TQB</b>	<b>Ferret</b>	<b>ViCoDE</b>
	Documents	Pdf, ps, doc	√	√	√		
		Slides	√	√	√	√	
		Notes	√			√	
	Video	√	√		√	√	
	Audio	√	√	√	√	√	
	Whiteboard				√		
	Images	√	√		√	√	
<b>Input modalities</b>			<b>FriDoc</b>	<b>Archivus</b>	<b>TQB</b>	<b>Ferret</b>	<b>ViCoDE</b>
	Keyboard	√	√	√	√	√	
	Mouse	√	√	√	√	√	
	Touchscreen		√				
	Voice		√				
<b>Output modalities</b>			<b>FriDoc</b>	<b>Archivus</b>	<b>TQB</b>	<b>Ferret</b>	<b>ViCoDE</b>
	Graphics	√	√	√	√	√	
	Audio	System audio		√			
		Data audio	√	√	√	√	√
	Video	√	√		√	√	
	Multi-video streams	√			√		
	Text	√	√	√	√	√	
Structured documents	√	√	√		√		
<b>Annotations</b>			<b>FriDoc</b>	<b>Archivus</b>	<b>TQB</b>	<b>Ferret</b>	<b>ViCoDE</b>
	Transcript of the meeting	√	√	√	√	√	
	Dialogue Acts		√	√		√	
	Thematic segmentation	√	√	√		√	
	Speaker segmentation	√	√	√	√	√	
	Argumentative segmentation		√			√	
	Meta-data (who, when, where etc)	√	√		√	√	
	Document content reference	√	√	√		√	
	Document image reference	√	√			√	

<b>Other functionalities</b>	<b>FriDoc</b>	<b>Archivus</b>	<b>TQB</b>	<b>Ferret</b>	<b>ViCoDE</b>	
Cross-meeting	√	√	√		√	
Content-based retrieval		√			√	
Document description					√	
Interface customization				√		
Data base		√	√		√	
Multiple visualization techniques	√	√		√	√	
Web-based			√	√	√	
Portable (i.e. java based)	√	√	√		√	
Dialogue driven		√				

## 4. Doctoral work related to the browsers

There are 3 PhD theses underway that are directly related to the work on the browsers described in this report.

1. Agnes Lisowska, from the ISSCO/TIM/ETI group at the University of Geneva, is working on a thesis under the supervision of Susan Armstrong concerning the user-driven design of a multimodal interface to multimodal meeting data. This will be done by extending methods from traditional HCI to the multimodal domain, following the standard software engineering lifecycle and considering psychological and cognitive factors imposed by users. Specifically, the key research issues addressed in the work will be determining 1) which modalities and in which combination are the most appropriate for both input and output for various tasks in the envisioned interface, 2) what language processing capabilities are required for a multimodal natural language interface, 3) what is the relationship between the design of the interface and the way in which the data in the database is stored and how that relationship can be exploited to improve the interface.

2. Maurizio Rigamonti, from the University of Fribourg, has started a PhD, under the supervision of Denis Lalanne and Rolf Ingold, on the design and implementation of a framework that allows 1) browsing in a world composed of heterogeneous and multimedia documents; 2) visualizing and validating documents derived data, links and annotations; 3) creating summaries of regions of the world. The documents in this world consist of media (static documents, video, etc.), multimedia documents (e.g. slide shows), strictly related documents, i.e. clusters and events such as conference and meetings, and finally persons. Search and navigation will be based on the document derived data, in order to retrieve and present relevant information to the user. The validation aims at accepting or correcting results of the various document analysis, annotations and links between related documents (e. g. document/speech thematic alignment). This validation task will allow 1) final users to navigate in a consistent world for retrieving information and 2) researchers to correct and refine algorithms for multimodal data analysis. Finally, summary creation targets the production of new documents, annotations and links, associating existing data with new ones.

3. Mirek Melichar, from the LIA lab at the EPFL will be starting a PhD under the direction of Martin Rajman. The objective of the thesis will be to find new approaches to multimodal dialogue management. In order to achieve a good compromise between interaction robustness and interaction flexibility, the chosen focus will be on frame-based systems (the frame is used to accumulate the user requests and to determine the next dialogue turn). We suggest supplying the dialogue manager with a set of different frame

templates, allowing the definition of more sophisticated dialogue management strategies making the system able to process different types of user queries. Another aspect of the system is multimodality. In order to allow an interaction with the system as flexible as possible, we plan to provide the user with a consistent access to a fixed variety of modalities that should be completely interchangeable. Since modality fusion is a very difficult task in general we have decided to circumvent the need for complex multimodal fusion techniques by a proper design of a multimodal interface that strictly controls the possibilities provided to the user to make complex cross-modal actions. The integrated system will emphasize portability of the complete solution. In addition, the produced dialogue manager should be able to run in two modes: a fully automated mode and a simulation mode, adequate for WOz experiments.

## 5. Future work and possible collaborative points

Several points of collaboration have already been established between the various groups working on the browsers. Ferret, ViCoDE, Archivus, TQB and FriDoc are using wherever possible common raw data, transcripts and derived annotations. Work on alignment between documents and references to those documents from within transcriptions is under way, as is the incorporation of that work and the work on document and topic alignment into the TQB and Archivus interfaces (the FriDoc browser provides a much more extensive range of document-driven manipulations of meeting content).

Furthermore, collaboration between all the groups presented above has started in 2004 for setting up a common Browser Evaluation Test continuing and expanding on an earlier version of the Browser Evaluation Test described in [15].

## 6. Contact Points

Browser	Name	Institute	E-mail
FriDoc			
	Denis Lalanne	UniFri	Denis.Lalanne@unifr.ch
	Didier von Rotz	EIF	<a href="mailto:Didier.vonrotz@eif.ch">Didier.vonrotz@eif.ch</a>
	Maurizio Rigamonti	UniFri	Maurizio.Rigamonti@unifr.ch
Archivus			
	Martin Rajman	EPFL	Martin.Rajman@epfl.ch
	Agnes Lisowska	UniGe	Agnes.Lisowska@issco.unige.ch
	Mirek Melichar	EPFL	Miroslav.Melichar@epfl.ch
Ferret			
	Pierre Wellner	IDIAP	<a href="mailto:wellner@idiap.ch">wellner@idiap.ch</a>
	Mike Flynn		<a href="mailto:flynn@idiap.ch">flynn@idiap.ch</a>
	Maël Guillemot		<a href="mailto:mael.guillemot@idiap.ch">mael.guillemot@idiap.ch</a>
ViCoDE			
	Nicolas Moenne-Loccoz	UniGe	<a href="mailto:moennen@cui.unige.ch">moennen@cui.unige.ch</a>
	Eric Bruno		<a href="mailto:bruno@cui.unige.ch">bruno@cui.unige.ch</a>
	Stéphane Marchand-Maillet		<a href="mailto:marchand@cui.unige.ch">marchand@cui.unige.ch</a>
TQB			
	Andrei Popescu-Belis	UniGe	Andrei.Popescu-Belis@issco.unige.ch
	Maria Georgescu		Maria.Georgescu@eti.unige.ch

## 7. References

- [1] Bui, T.H., Rajman, M., and Melichar, M. Rapid dialogue prototyping methodology. In Petr Sojka, Ivan Kopeček, and Karel Pala, editors, Proceedings of the 7<sup>th</sup> International Conference on Text, Speech and Dialogue-TSD 2004, Lecture Notes in Artificial Intelligence LNCS/LNAI 3206, pages 579-586, Brno, Czech Republic, September 2004. Springer-Verlag.
- [2] Dahlbäck, N., Jönsson, A. and Ahrenberg L., "Wizard of Oz Studies – Why and How", in W.D. Gray, W.E. Helfley and Murray, D. (eds). Proceedings of the 1993 Workshop on Intelligent User Interfaces (pp. 193/200) Orlando, FL. New York, ACM Press.
- [3] Janvier, B., Bruno, E., Marchand-Maillet, S., and Pun, T. *Information-theoretic framework for the joint temporal partitioning and representation of video data*, in European Conference on Content-based Multimedia Indexing (CBMI03), Rennes France, September, 2003
- [4] Lalanne, D., Ingold, R., von Rotz, D., Behera, A., Mekhaldi, D., Popescu-Belis A. (2005)- "Using static documents as structured and thematic interfaces to multimedia meeting archives". In Bourlard H. & Bengio S., eds., Multimodal Interaction and Related Machine Learning Algorithms, LNCS, Springer-Verlag, Berlin, pp. 87-100.
- [5] Lalanne, D., Sire, S., Ingold, R., Behera, A., Mekhaldi, D., von Rotz, D. (2003), [Research Agenda For Assessing The Utility Of Document Annotations](#). In Multimedia Databases Of Meeting Recordings. In proceedings of the 3rd International Workshop on Multimedia Data And Document Engineering, September 8th 2003, Berlin, Germany.
- [6] Lisowska A., Rajman M., & Bui T.H. (2004) - ARCHIVUS: A System for Accessing the Content of Recorded Multimodal Meetings. *Proceedings of the Joint AMI/PASCAL/IM2/M4 Workshop on Multimodal Interaction and Related Machine Learning Algorithms*, Martigny, Switzerland, 8 p.
- [7] Lisowska, A. (2003) - [Multimodal Interface Design for the Multimodal Meeting Domain: Preliminary Indications from a Query Analysis Study](#). Project Report IM2.MDM-11, November 2003, 30 pages.
- [8] [Lisowska A., Popescu-Belis A. & Armstrong S. \(2004\)](#) - User Query Analysis for the Specification and Evaluation of a Dialogue Processing and Retrieval System. *LREC 2004 (Fourth International Conference on Language Resources and Evaluation)*, Lisbon, Portugal, vol.III, p.993-996.
- [9] Marchand-Maillet, S., and Bruno, E., Collection Guiding: A new framework for handling large multimedia collections, in First Workshop on Audio-visual Content And Information Visualization In Digital Libraries, AVIVDiLib05, Cortona, Italy, May, 2005.
- [10] Moënné-Loccoz, N., Janvier, B., Marchand-Maillet, S., and Bruno, E. *Handling Temporal Heterogeneous Data for Content-Based Management of Large Video Collections*, in Multimedia Tools and Applications, (to appear), 2005
- [11] [Popescu-Belis A., Georgescu M., Clark A. & Armstrong S. \(2004\)](#) - Building and Using a Corpus of Shallow Dialogue Annotated Meetings. *LREC 2004 (Fourth International Conference on Language Resources and Evaluation)*, Lisbon, Portugal, vol.IV, p.1451-1454.
- [12] [Popescu-Belis A., Clark A., Georgescu M., Zufferey S. & Lalanne D. \(2005\)](#) - Shallow Dialogue Processing Using Machine Learning Algorithms (or not). In Bengio

S. & Bourlard H., eds., *Machine Learning for Multimodal Interaction*, LNCS 3361, Springer-Verlag, Berlin/Heidelberg, p.277-290.

- [13] [Popescu-Belis A. & Lalanne D. \(2004\)](#) - Reference Resolution over a Restricted Domain: References to Documents. *ACL 2004 Workshop on Reference Resolution and its Applications*, Barcelona, Spain, p.71-78.
- [14] Tucker, S., Whittaker, S. (2005), Accessing Multimodal Meeting Data: Systems, Problems and Possibilities, In: Bengio, S., Bourlard, H. (Eds.) *Lecture Notes In Computer Science*, 3361, 1-11.
- [15] Wellner, P., Flynn M., & Guillemot M., (June 2004), "Browsing Recorded Meetings With Ferret", *MLMI'04 Workshop*, Martigny, Switzerland.
- [16] Wellner, P., Flynn, M., Tucker, S., Whittaker, S. (2005), A Meeting Browser Evaluation Test, Presented at the Conference on Human Factors in Computing Systems (CHI), Portland, Oregon, USA, 2nd-7th April.