Conference knowledge modelling for conference-video-recordings querying and visualization

Maria Sokhn[1], Elena Mugellini[2], Omar Abou Khaled[3], Ahmed Serhrouchni[4]
[1, 2, 3] Multimedia Information System Group
Dept. of Information & Communication Technologies
University of Applied Sciences of Western Switzerland
Boulevard de Pérolles 80, 1700 Fribourg
Switzerland
{maria.sokhn, elena.mugellini, omar.aboukhaled}@hefr.ch

[4]Multimedia information system group
Dept. of Network & Security
Telecom ParisTech
46 rue Barrault 75634 Paris cedex 13
France
ahmed.serhrouchni@telecom-paristech.fr

ABSTRACT: *The evolution of the web in the last decades has created the need for new requirements towards intelligent information retrieval capabilities and advanced user interfaces. Nowadays, effective retrieval and usage of multimedia resources have to deal with the issues of creating efficient indexes, developing retrieval tools and improving user oriented visualization interfaces. To that end we put forward an integrated framework named CALIMERA. The framework is based on a High-level modEL for cOnference (HELO) and aims at enhancing the information management, retrieval and visualization of recorded talks of scientific conferences. This paper presents the conference model and its uses within the framework: performing high level annotation of scientific talk recordings, offering granular search facilities and complex queries, and enhancing the knowledge visualization of the recordings. As a proof-of-concept we present the prototypes that have been implemented.*

Keywords: Knowledge modelling, Ontologies, Knowledge-based information management, Visualization, Virtual querying

Received: 12 March 2010, Revised 11 April 2010, Accepted 19 April 2010

1. Introduction

While recording technology is becoming easier to use and more affordable due to technological developments, an increasing number of conferences and scientific events are now being recorded. However the resulting multimedia data, such as the recordings of the talks, is accessible today only if considerable effort is made to analyze, extract and create appropriate indexes. In fact, video data presents challenges to information retrieval due to the dynamic nature of the medium and due to the lack of advanced and precise methods that handle non-textual features (1). This multimedia data lacks semantic content annotation, giving place to the so-called semantic gap. This gap was defined by Smeulders (2) as “the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data has for a user in a given situation”. Other papers presented similar definitions (3; 4; 5). In recent years, ontologies being abstract models of some domains have been adopted to enhance content based annotation and, accordingly, retrieval of multimedia resources (6; 1; 7). Indeed, as argued in Hare (8), the use of ontologies improves both automatic annotation and retrieval process. Using ontologies to describe multimedia resources provides methods to define well structured and related concepts facilitating the burden of annotation and retrieval (9). Ontologies are also very useful for visual representation of resources (10; 11), and may considerably enhance browsing, navigation and data accessibility. This paper presents the ontology model for conference video-recordings, HELO and its uses within the conference framework CALIMERA. HELO stands for High-level modEL for cOnference. It models the information and knowledge conveyed within a conference life cycle. By using HELO we aim at bridging the semantic gap by providing possibilities to perform high level and granular annotations to scientific talk recordings, to enhance information retrieval and visualization of the recordings, and to allow granular search facilities and complex queries based on semantic criteria such as: Find a talk sequence recording of the workshop held in Utah, in 2007, in relation with the event "ACM MEDES" where the speaker talked about "CALIMERA" project and presented a demo. The speaker is Tom Smith's colleague and works in the "MIT" research group and wrote the paper "A survey of multimedia retrieval". The remaining part of the paper is structured as follows: In section 2, we review related works. In section 3, we introduce the framework CALIMERA. In section 4, we describe the conference knowledge model HELO. In section 5, we present the uses of HELO through three model-based processes involved along an information retrieval: Section 5.1, model based annotation of the scientific talk recordings, section 5.2, model based querying for the scientific talk recordings followed by the description of the virtual query engine section 5.3 and section 5.4, model based visualization interface. Finally, we conclude with the section 6 and we present future work.

2. Background

Many publications and projects similar to the ones described in (12; 13; 14; 15; 4; 16; 9) have been designed to handle multimedia information retrieval. Some of them, such as the PHAROS project (13), are dedicated to generic multimedia information retrieval. Others are designed to handle specific domains, such as COALA (16) for TV news information retrieval, ConKMeL (15) for e-learning information retrieval, the work conducted with Bertini et al (9) for soccer video annotation and retrieval. In our work we focus on conference and scientific events domain. We aim to enhance and facilitate the retrieval, the navigation through and the replay of scientific talk recordings, relying on conference and scientific events model. Much effort has been dedicated to modelling scientific events. To name only few we mention: The AKT Reference Ontology (17), the Conference ontology by Jen Golbeck (18), the eBiquity Conference Ontology (19), the ESWC2006 and ISWC2006 ontology (20; 7), and the Semantic Web Conference Ontology (SWC) (25). According to a detailed analysis within the ESWC and ISWC metadata project, existing ontologies, except the SWC, lack the expressiveness required to assert the entire knowledge and content of a conference (7). Therefore more expressive ontologies, the ESCW2006 and ISWC2006, have been developed. Lately SWC ontology was developed. It draws heavily on ontologies developed for ESWC2006 and ISWC2006. It is mainly a convention of how to use classes and properties from other ontologies, most prominently FOAF (for people) and SWRC (their BibTeX elements, for the papers). These ontologies lack expressiveness required to model the multimedia data generated from scientific events such as the videos of the recorded talks. To address this issue in addition to several restructuring within designed concepts, we propose a new conference model named HELO detailed in section 4. HELO, as mentioned in the introduction models the knowledge conveyed within a scientific conference life cycle and has the advantage of being integrated in a framework designed to handle conference information management and retrieval. This framework named CALIMERA is introduced in the following section.

3. CALIMERA Framework

CALIMERA is an integrated framework aiming to facilitate the retrieval of videos of recorded talks within a conference. CALIMERA framework provides a solution for two main tasks; the knowledge and information management of a conference video recording as well as their retrieval. We argue that the retrieval process is enhanced if the content and the context of a conference during its life cycle are taken into account for describing the resources.

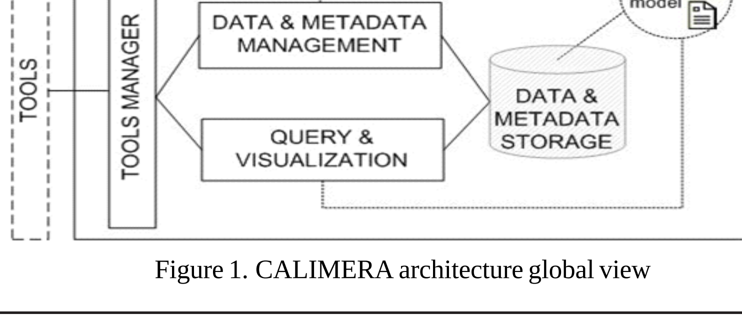

Figure 1. CALIMERA architecture global view

Figure 1 outlines the global view of the framework which is composed of the following modules:

- **Tools manager:** CALIMERA is a tool independent framework. The tool manager allows any user to integrate a tool that may be used for data metadata management, query and visualization or both (Figure 2). For a proof of concept we integrated four principal tools. SMAC is a tool we developed, to record conference talk and automatically segments the recording of this talk based on slide change detections. INDICO is an integrated tool developed by the CERN (21) that manages the administrative part of a conference (such as the conference planning, logistics, etc.). INVENIO and CALISEMA (presented in section 5.1.1) are two different tools for video semi-automatic and manual annotation, guided by the conference model, HELO. INVENIO has also a set of modules for automatic features extraction of multimedia data and information indexing.

- **Data and metadata management module** (Figure 2) consists of handling the conference high-level information, such as recording talks, segmenting video recordings, annotating video segments, managing the context information of these talks, etc.

- **Data and metadata storage** (Figure 2) integrates existing data and metadata formats such as MPEG-7 (22) which is one of the most widely used standard for multimedia description, RDF (23) and OWL (24), which are a more semantically oriented standards for multimedia description that integrates high level semantic description.

- **Query and visualization module** (Figure 2) queries the data and metadata storage in order to return the video or the set of video sequences of recorded talks the users are seeking for. In order to handle the heterogeneity of annotation standards and formats we designed a format independent interface aiming at querying the heterogeneous data storage. This query engine is based on different concepts such as query analysis, query reformulation, result reasoner, etc. and is presented in section 5.3.

- **Conference model:** HELO (Figure 1 and 2), is a conference model we designed to model the conference high-level information conveyed within a conference life cycle. As cited in the introduction HELO is based on existing ontologies related to conference domain.

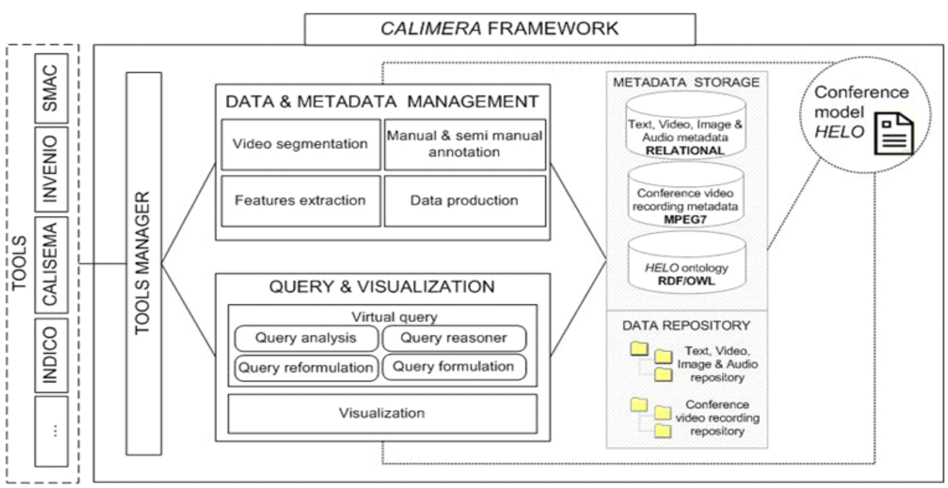

Figure 2. CALIMERA architecture detailed view

4. HELO: Conference knowledge model

As stated above, the model HELO has been designed to enhance multimedia information retrieval and more precisely the scientific conference recordings. We decided to describe this model using ontologies. In fact, the use of ontologies makes possible the extension of existing vocabularies and concepts. Moreover data and object properties such as symmetry, transitivity, inverse functional, etc. can largely simplify and enhance the knowledge modelling. The knowledge conveyed along a scientific conference life cycle, mainly concerns three entities.

- **Event:** events may consist of single talks such as paper presentations or multiple talks within a session, or industrial track or special events such as an inauguration or a software release.

- **People:** people in a conference may be speakers, colleagues of a speaker, attendees, organizers, members of a community, etc.

- **Artefact:** artefacts are about multimedia data such as full papers or posters, conference proceedings, speaker's slide set, and most importantly and of great value -according to our work-, the videos of recorded talks.

HELO is an ontological model that describes and structures the information conveyed within a conference life cycle. Wherever possible we thought to reuse established concepts. Indeed HELO is based on the effort made in ESWC2006, ISWC2006 and SWC (25). It integrates several other concepts extracted from other established ontologies related to conference and other broad domains. For instance, HELO makes use of the event concept from the SWC ontology, the Person concept from the FOAF ontology (26), the ResearchTopic concept from the SWRC (27) ontology. In addition to reused concepts/classes HELO developed additional classes related to the following issues: describing the video resources of recorded talk, increasing the integration of existing concepts, and taking advantage of all the possibilities the ontology may offer such as rules usage. As shown in the Figure 3, the HELO model can be graphically presented as a two-layer set-of-concepts: the Annotation oriented layer and the User oriented query and visualization layer, all of which are extensively sub-classed to provide a high-level degree of expressiveness. By that, the knowledge conveyed within a conference is on the one hand described and annotated according to granular concepts (Annotation oriented layer) and on the other hand it is searched using a user oriented and simplified process (User oriented query and visualization layer).

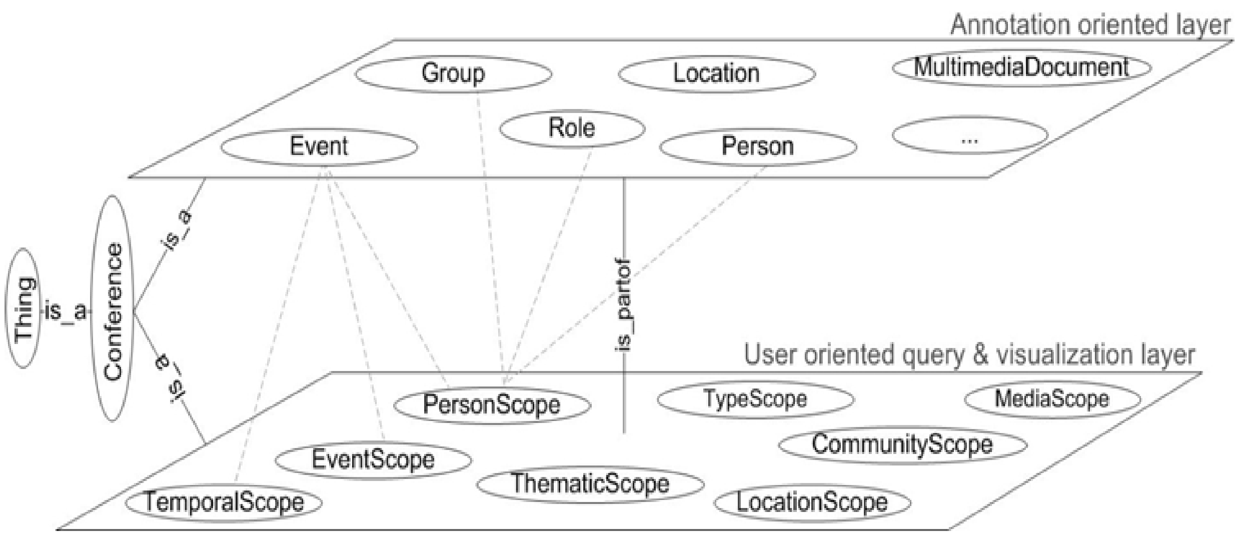

Figure 3. HELO model representation

- **The Annotation oriented layer:** the Annotation oriented layer offers granular concepts allowing users to annotate conference information, and mainly scientific talk recordings, with a high-level degree of expressiveness. This layer is composed of the following set-of-concepts: Group, Person, OrganisedEvent, Location, MultimediaDocument and Topic. Other important concepts have been defined through sub-classes or properties such as the concepts of Role, Expertise, VideoRecordings, etc. The major improvements led into HELO compared to the above listed ontologies are the definition of the MultimediaDocument concept which describes all the media conveyed within a conference. The VideoRecording (a sub-class of the MultimediaDocument concept) which describes the video recordings of scientific conference talks. The Expertise concept which describes the expertise level per domain. Finally, the Topic concept which describes the thematic occurring in scientific conferences, Topics concept is based on the ACM taxonomy (28) therefore guiding the users in their annotations.

- **The User oriented query and visualization layer:** In order to provide a common retrieval framework for different users we proposed to model a scientific conference through different views, referred in our work as Scopes. These scopes, detailed in the following list, correspond to a set of interpretations of the audiovisual content based on users' retrieval needs. In order to identify these different scopes, we studied a set of queries (such as the ones listed in the introduction) put forward by different users. Based on that we have identified 8 Scopes, which correspond to the most common retrieval activities of users in a scientific conference environment. Each scope is a concept that may be linked to one or more concepts of the Annotation oriented layer.

1. **PersonScope** includes information about people involved in a conference e.g. names, roles, affiliation. It allows users to make queries such as: find the video-recording of the talk where a colleague of the chairman made a presentation.

2. **LocationScope** contains information about the conference location e.g. continent, city, building, room. It allows users to make queries such as: find the video-recording of the talk that took place in a city in France.

3. **TemporalScope** concerns conference planning e.g. starting time, parallel sessions, breaks. It allows users to make queries such as: find the video-recording of the talk that took place in the afternoon in parallel with talk B.

4. **TypeScope** lists several categories of conferences e.g. workshop, lecture. It allows users to make queries such as: find a talk given in the academic lecture Y.

5. **MediaScope** gathers all the media information linked to a talk e.g. video-recording of the talk, slide set document, papers, and books. It allows users to make queries such as: find the video-recording sequence of the talk related to this paper.

6. **ThematicScope** affiliates a conference to a domain, topic, related events e.g. video-recording indexing, biology. It allows users to make queries such as: find the video-recording part of the talk related to knowledge management.

7. **CommunityScope** defines communities such as laboratories, research groups, conferences group. It allows users to make queries such as: find the video-recording of the talk where a professor from France University in the MEDES program committee made a presentation.

8. **EventScope** describes the events related to a conference event. It allows users to make queries such as: find the video-recording of the talk related to the LHC inauguration event.

Alongside with the establishment of the HELO model we conceive applications highlighting the enhancement due to the use of ontology in multimedia information retrieval, more precisely the retrieval of the recorded talks of a conference. The following section describes the three model-based processes used in the retrieval of the recorded talks: the model based annotation of scientific talk recordings, the model based querying for scientific talk recordings and the model based visualization interface.

5. HELO based retrieval

Multimedia information retrieval, more precisely the conference recordings retrieval, is enhanced by using a conference ontology model throughout the three main activities of an information retrieval process which are: the annotation process (section 5.1), the querying process (section 5.2) and the results rendering/visualization process (section 5.3). To that end we build the framework CALIMERA, based on the model HELO. The following sections describe the HELO based annotation, the HELO based querying and the HELO based visualization, respectively.

5.1 HELO based annotation

Annotations are an important key-success in the emergence of the semantic web. The need to semantically annotate resources using ontologies has been clearly identified (32). Since manual annotation is a laborious task, it is important and required to develop annotation tools that facilitate this task. In the last decades several interesting projects have been conducted. In our work we developed in collaboration with the University of Athena, CALISEMA a tool for manual and semi-manual video annotations (section 5.1.1) and in collaboration with a Swiss institute, an automatic annotation tool based on a speech to text plug-in (section 5.1.2).

5.1.1 Semi-manual annotation prototype

In order to facilitate the annotation of videos and more precisely the videos of the recorded talks we developed CALISEMA. What characterizes CALISEMA from other existing tools is that it is fully integrated and adapted to conferences. In fact CALISEMA is integrated in the information management process of the framework CALIMERA. This enables CALISEMA to automatically collect information (e.g. name of the speaker, duration of the video, etc.) from other tools used in our framework. Moreover, CALISEMA integrates several video segmentation algorithms allowing users to choose the one that most meets their requirements. One of these algorithms offers users the possibility to segments the video of the recorded talk according to slide change detection. This tool also allows annotations according to the conference ontology HELO. Figure 4 shows a screen shot of the CALISEMA prototype.

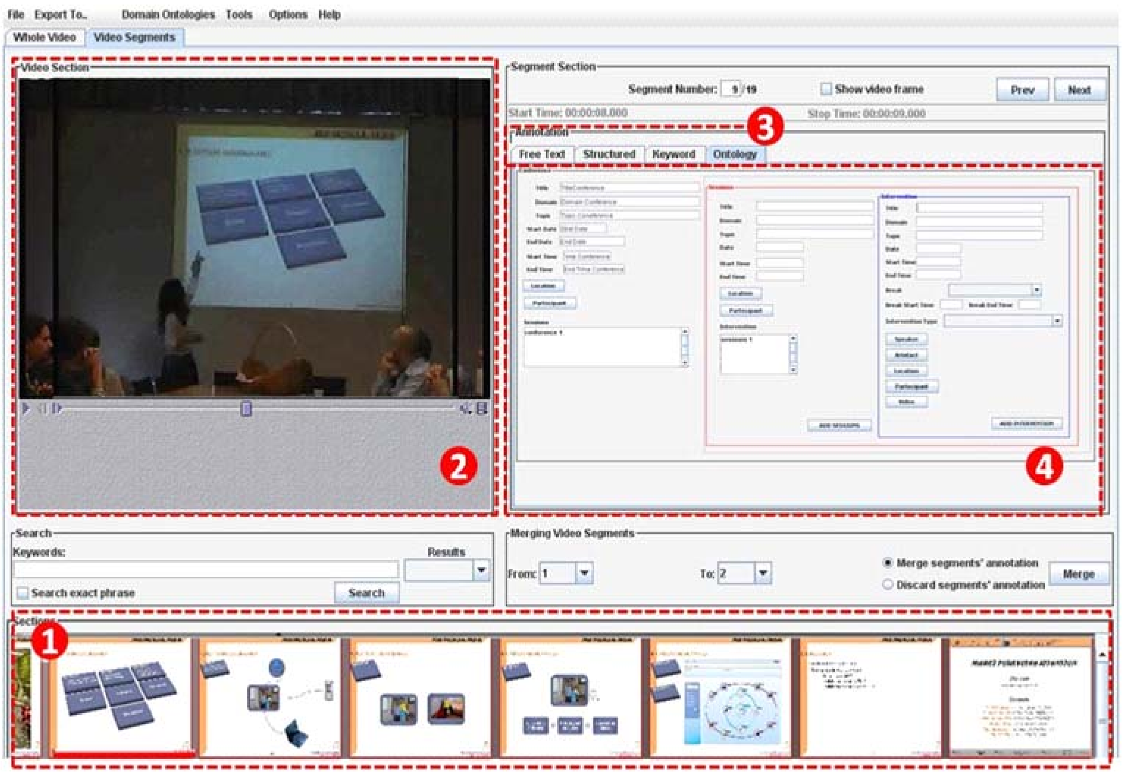

Figure 4. HELO based annotation prototype

The bottom part -1- shows the key-frames of the video segments of the recorded talk. Every key-frame corresponds to a slide within the talk presentation slide- set. In the left part -2-, we visualize the video sequence corresponding to the chosen key-frame. In the top-right part -3-, we identify the three different ways of annotation offered to the users: free text, structured, keywords and ontology. The middle-right part -4-, is dedicated for annotating; Figure 4 is an example of an HELO ontology based annotation. CALISEMA provides users the possibility to handle (deleting, adding, or merging) video segments. The annotation is exported in either MPEG-7 format, OWL format or both.

5.1.2 Automatic annotation prototype

The integration of a speech-to-text plug-in within CALIMERA framework is essential as it allows associating automatically additional metadata to the recorded talks. In order to analyze the audio content of the video, a first step is to transcript the spoken words. The voice recognition technology uses methods from the signal processing and artificial intelligence domains. A recorded and digitized sentence is pushed to the voice recognition software in order to be transcript. There are several voice recognition software solutions, among the best we can list Crescendo (29), Dragon Naturally Speaking (30) and Sphinx4 (31). Sphinx4 has been developed by the University Carnegie Mellon, Sun Microsystems laboratory and Hewlett Packard. More over it is a flexible and modular framework. Adding to it offers (as open source) the framework and several implementations examples. Based on these advantages we developed our plug-in using the Sphinx4 technology. In order to store the audio time-based metadata we used the MPEG-7 technology. In fact MPEG-7 is standardized which allows us to offer interoperability with other systems. In addition to that, it offers possibilities of future extension. Improving the efficiency of the speech-to-text algorithm is an important point. With Sphinx4 technology, we got 78% of correct words. The experiment was based on an internal test bed composed of English speaking conferences with different accents, audio quality and length). In order to improve this percentage we decided to apply some optimizations. Two optimizations technique were combined in order to enhance the result quality: on the one hand by removing the rare words, and on the other hand by modifying the segments durations of the audio track. The result is shown in figure 5.

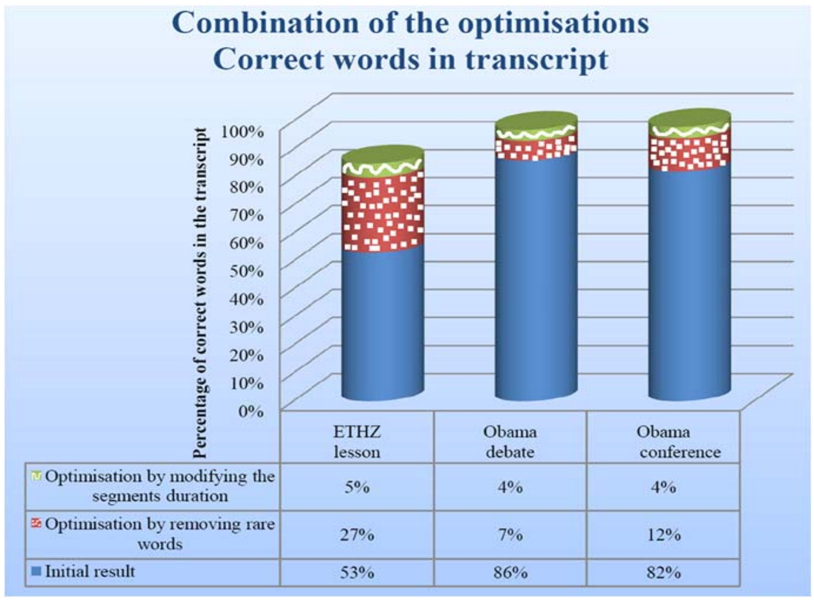

Figure 5. Speech-to-text results

Algorithm concept

The relevance value defines the importance of a word. This relevance can be compared to the page rank in Google for instance. When searching a word in Google, a list of websites is displayed. This list starts with the results which are very relevant. Defining relevance in speech-to-text plug-in is not a trivial task. In fact, we only know the words which were spoken and the time they were evoked within the video. Based on this information, a relevance value needs to be computed. To get a correct relevance value it is important to define some criteria. We based our analysis on three criteria:

• The relevance of the word in the entirety of the audio track. To determine this relevance we take into consideration the number of occurrences of each word in the entirety of the audio track. If a word occurs frequently, it is then considered as a relevant word within the video.

• The relevance of the word in the context. To determine this relevance we take into consideration the number of occurrences of each word within an interval of time. If a word occurs several times it is considered a relevant word.

This relevance can define whether the word is relevant in the context of the respective current subject in the audio track or not.

• Relevance value in the proximity. This parameter is determined based on the time that separates two occurrences of a same word. For instance, a word occurring twice in a short period of time is considered as a relevant word. Each word has a relevance value which describes the relevance of it along the entire video, its relevance throughout the context and its relevance according to the proximity words.

5.2 HELO based querying

The HELO model is based on the study of user needs and requirements when searching for scientific recorded talks. We set up a graphical search interface based on users needs and requirements and fully adapted for scientific events domain (Figure 6).

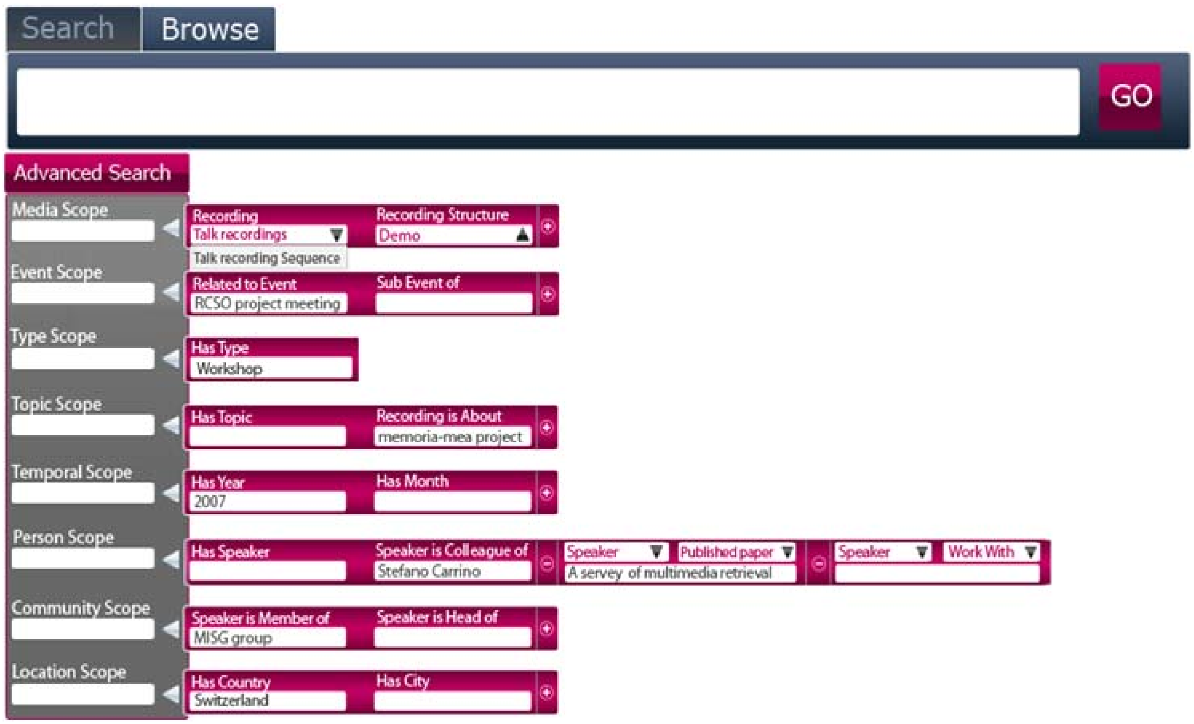

Figure 6. HELO based query interface

This interface offers the users the possibility to make, on the one hand, “traditional” keyword based searches by using the simple search bar, and on the other hand, to perform complex and granular queries such as the one listed in the introduction by using the “Advanced Search” option. The “Advanced Search” is composed of 8 fields that correspond to the 8 defined scopes (MediaScope, EventScope, TypeScope, ThematicScope, TemporalScope, PersonScope, CommunityScope, LocationScope). Each field contains one or more sub-categories making granular queries possible (Figure 6). The query formulated by the users through this graphical user interface is translated to a SPARQL query language with a user transparent way. This translation is ensured by a knowledge base engine. This engine uses Jena's framework to handle ontologies (definition of data), data (instances of the ontologies) and inferences (additional deductions based on rules).

The resulting information takes the form of a Uniform Resource Identifier (Figure 7). These identifiers are then interpreted by the graphical user interface which part of it is presented in section 5.4.

In order to enrich the query results integrating data not only from ontological resources we need to address the issue of handling queries over heterogeneous and distributed data. In fact the CALIMERA framework allows the integration of data from multiple heterogeneous sources such as manual annotation with CALISEMA, automatic annotation from the speech to text plug-in, etc. All these sources (may) have different data storage models and techniques. This leads to the coexistence of different data models and scheme and therefore different query languages. The structural and semantic heterogeneity of data makes the development of custom solutions for querying these data time-consuming and complex. This issue can be divided into three main parts: dispatching important search information depending on the data sources, creating database-specific queries and merging results from several sources. To handle this issue we designed a query system named Virtual-Q described in the following section.

5.3 Virtual query engine

Virtual-Q aims to provide users transparent and easy access to retrieve data from heterogeneous sources. Virtual-Q system is a novel approach based on innovative virtual query engine architecture. This new architecture integrates concepts such as

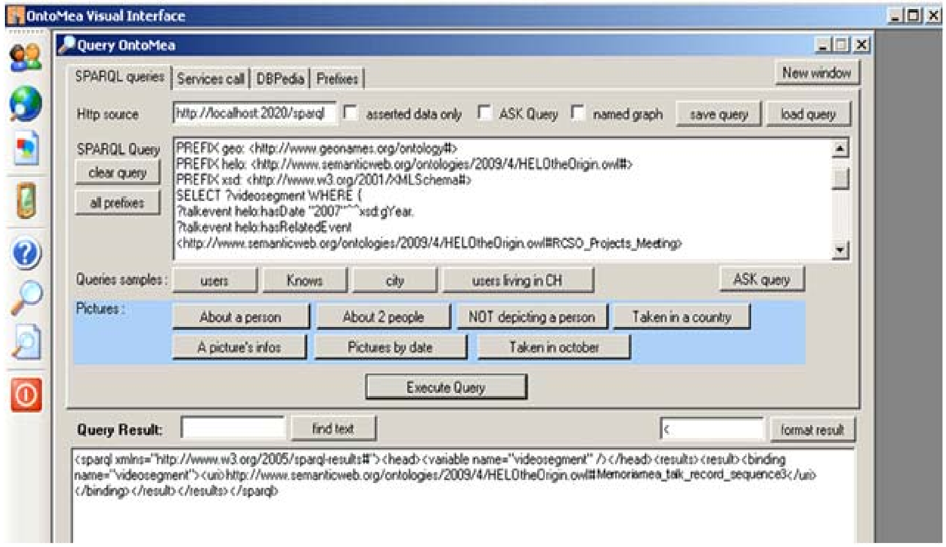

Figure 7. Knowledge based engine

query analysis or sub-queries formulation, which facilitate transparent access to heterogeneous data. Existing architectures are often based on existing elements that simplify sub-queries reformulation, such as global and local ontology for each data source using only Xquery and working on XML data with metadata describing sources' structure, or an administrator which determines similarities and conflicts between sources and defines rules for inter-schema correspondences. Virtual-Q aims to provide mainly an easy-to-use system by avoiding the complex administration work. Therefore we designed an autonomous query engine able to query heterogeneous data sources even when ignoring the data structure. The user takes advantage of this approach indeed he easily adds or removes data sources at any time. The global system (Figure 8) is divided into four main parts: -1- the user interface, -2- the Virtual Query Engine (VQE), -3- the external data sources and -4- the external reasoning models.

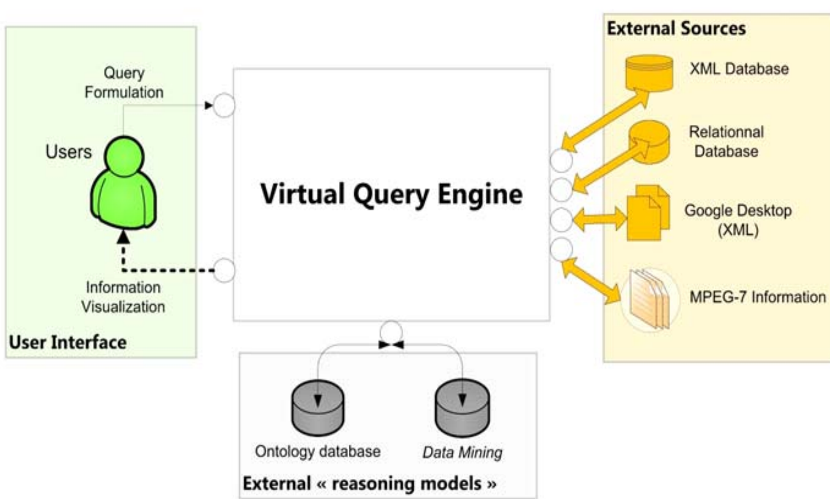

Figure 8. Virtual-Q system overview

The user interface displays the information for the user and proposes an interface to formulate the queries. The external data sources are the heterogeneous sources the engine will query. The external reasoning models are sources that could help the engine when it reasons on the results. The Virtual Query Engine is the core of the global system which is (Figure 9).

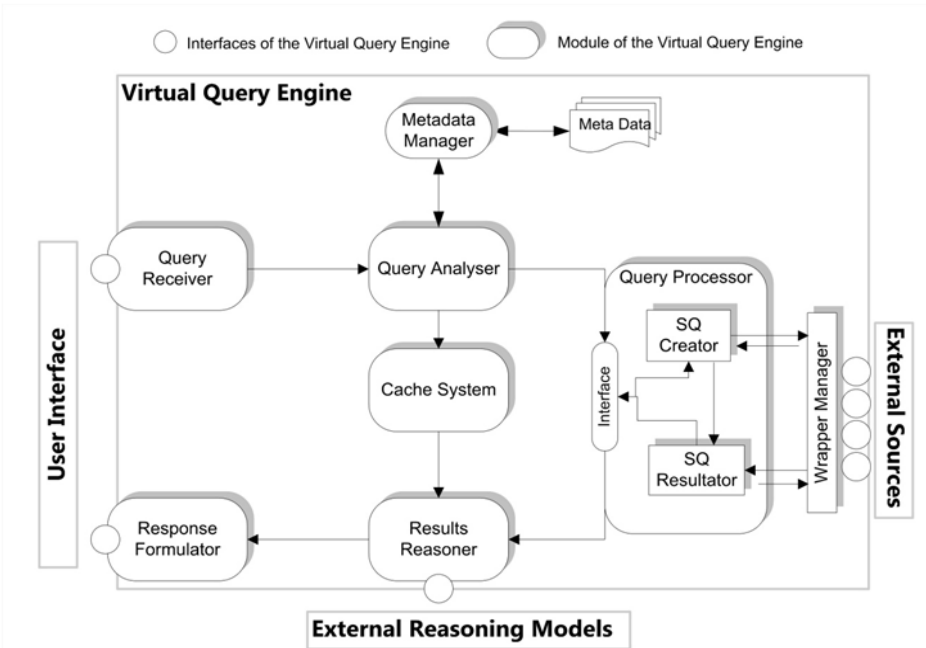

Figure 9. Virtual Query Engine architecture

5.3.1 Virtual query engine architecture

The Virtual Query Engine is composed of several modules. The modules that process the query are linked together and form a chain (Figure 10). Once the query is transmitted by the user interface to the query engine, it is transformed into a format allowing the addition of extra data. This query passes through the chain of modules. Each module is free to use these data or to add additional. In principle, the data are added during the analysis phase and used during the next phases like sub-queries reformulation or results reasoning. The query process acts as follows: -1- The user enters the query and transmits it to the engine. -2- The query is analyzed, using the best data source schema if necessary. -3- The query is reformulated in source-specific queries, which are executed on the corresponding sources. The query is first run on the best source in order to find more helpful data for sub-queries reformulation and then run on the other sources. -4- The engine merges the results and reasons on them to remove useless or organize them for example. -5- The results are formatted in the correct output format. -6- Finally, the results are displayed to the user. This modularity allows to clearly defining the task(s) of each module. We also can easily change a module or add a new one. In addition, several modules could be chained to perform specific tasks. Two modules are responsible for managing the internal "configuration" of the engine and the wrappers (especially forwarding the query to the right wrappers, thus to the right data sources). The different modules composing the Virtual Query Engine and their behaviours are described below.

Query process module

• **Query Receiver:** receives the original user query and transforms it in the format used inside the Virtual Query Engine. We refer to this as the User Query.

• **Query Analyzer**: analyses the free query to find pertinent data that could be used later in the process. Therefore, it retrieves first the best schema trough the Metadata Manager and the Wrapper Manager. Then, with different methods analyses the query and stores those data in the query. We refer to this as the Virtual Query. The query analysis allows finding pertinent information for users' queries. These data are then be used along the query process. In order to improve the query analysis, we suggest the following approach: -1- Data sources schemes: Each data stored into a source is compliant to a schema or model, even the most basic one. According to the kind of source, this schema will be more or less complex and

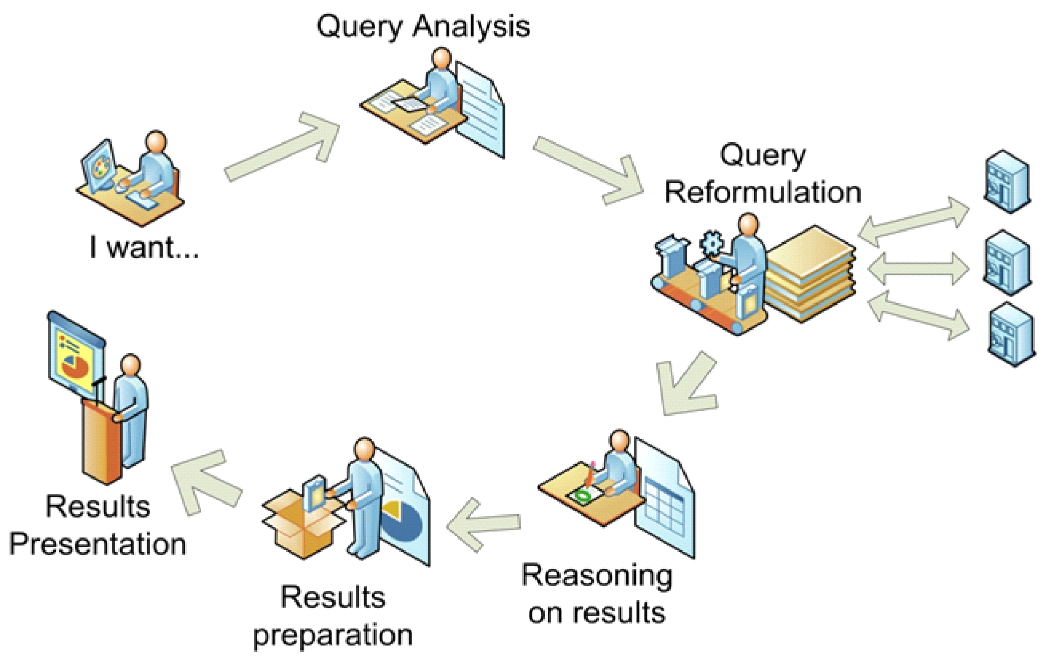

Figure 10. Query process approach

complete. The idea is to request this schema to the data source. Normally, the most data sources allow retrieving their schemes or models. Sometimes, it can also be known differently. -2- Ranking of the schemes: One knows as well that some schemes are stronger than others. For example ontology provides more structured and rich information than then Google Desktop XML Schema. So, the second idea is to rank the available schemes for each theme and to store it in the metadata. Like this, the VQE knows the best schema for a given theme. -3- Using the schema: Once we have the best schema or model, we can use it to analyze the queries. This could be done exploiting the fields' names or types for example. Moreover, since we know the best schema, we can refine the creation of sub-queries using the results of the query on the best data source. Secondly, it is important to know which information could be "produced" from the query and how. Some possibilities are: relevant data sources (using metadata and source ranking), Keywords (fields' names of data sources, lists, ontologies, etc.), stop words (Black list, database, etc.), summary (who, what, where, etc.), Boolean functions (predefine keywords or symbols: OR, AND, &&, etc.), relations between terms (ontologies, etc.). In most cases users only write few keywords for their query, consequently some of the methods presented above are no more adequate. Sub-queries formulation presents a key-solution for this issue.

• **Query Processor:** contains two internal modules. The first internal module creates a first query run on the best data source, possibly analyses the results for additional helpful data and transmits the Virtual Query to the second module. That one forwards the query to all the concerned wrappers through the Wrapper Manager and lastly merges the results.

• **Result Reasoner**: reasons on the results, guided by external sources if necessary to improve the results (suppression of useless results, better organization of the results, etc.

• Response Formulator: reformulates the response into a standard format before transmitting it to the user. The query reformulation is one of the main challenges of Virtual-Q. Indeed the integration of heterogeneous data sources and the fact that the user asks free queries add the constraint of transforming the original query into source-specific queries. This means that for the query Greece, the correct SQL query is Select X From Y Where Z for a relational database or XQuery Full-text search query for XML documents must be generated. Moreover, the approach chosen by Virtual-Q (no global schema and free query) increases the difficultly of the query reformulation. The analysis module gets round the free queries creating structured queries. Then the sub-queries are composed thanks to this structured information. Concerning the global schema, Virtual-Q does not have an established one, we consider that the module is self-learned based and therefore it keeps track of links between keywords, local information and between local schemes. The complexity in developing a wrapper depends on the kind of source that is being handled. Wrappers could be Full-text Search, API-based or fields-based.

Resource management:

• **Metadata Manager:** manages the internal metadata of the Virtual Query Engine (connected data sources, installed wrappers, themes with source ranking and reasoning sources).
• Wrapper Manager: is responsible for forwarding the queries to the concerned wrapper(s).

5.3.2 Virtual query engine prototype
In order to validate the new concept proposed in this paper, a preliminary prototype has been developed in Java 1.6. Currently, bases of the query process have been implemented as well as the Wrapper Manager and the Metadata Manager. Two wrappers have also been created: one for Google Search Desktop through an API and another for searching inside MPEG-7 files (an XML based language for multimedia contents description). The graphical interface is composed of two windows. The user window allows running a query and the administration window (Figure 11) to manage the wrappers and data sources. It also displays the log of the engine.

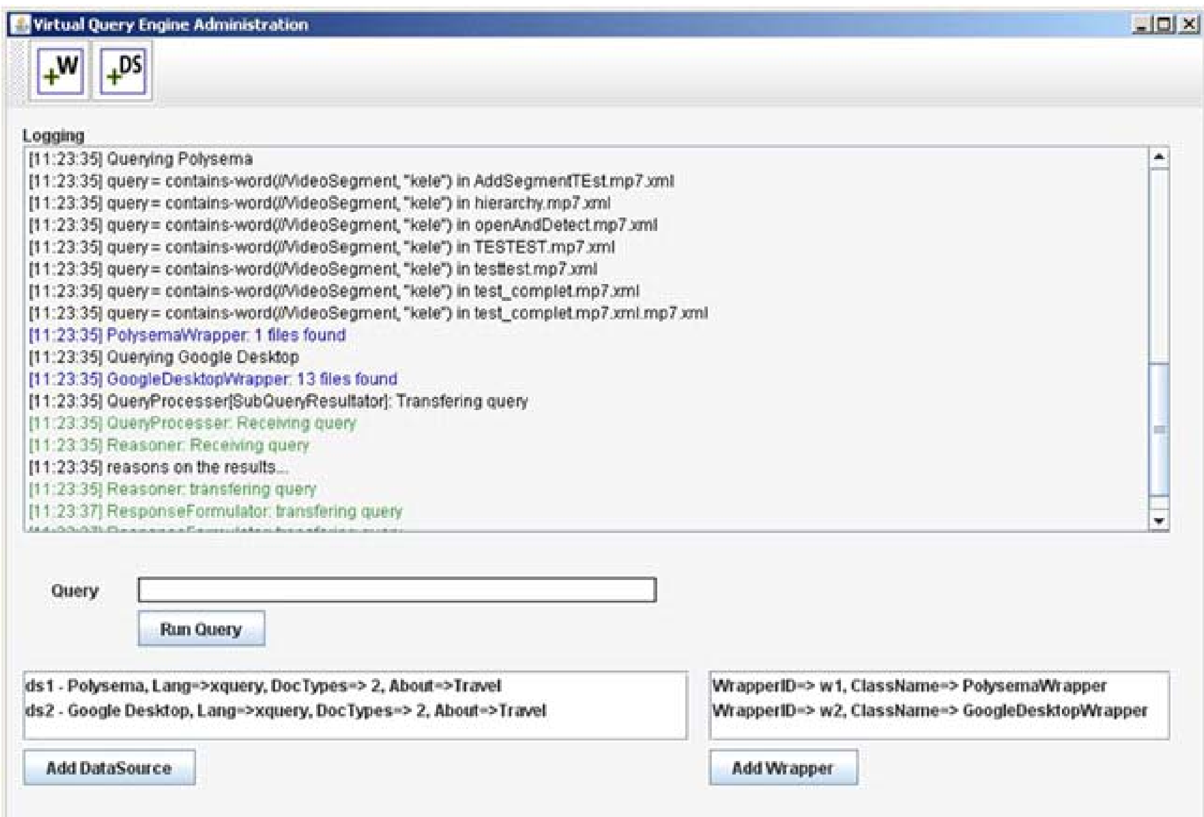

Figure 11. Virtual-Q prototype: administration window

The other technical choices explained in the following relate to the data format for queries, the communication between the different modules and the metadata management.

• **Query:** The principal format used all along the process is XML. First, XML is an easy-readable format for which many tool already exist. Secondly, XML is the base for OWL and RDF. So, it is possible to easy translate data from on format into another when needed. Thirdly, XML allows envisaging external treatments about the query because the transmission of data is easy and even Web Services technologies could be used. In the engine, the XML data are encapsulated in Java Object. The “XML query” is compliant to an XML Schema but no validation is done because only the software is expected to modify these data.

• *Inter-modules communication:* The second problem is the communication between the modules. The only data that is exchanged between them should be the query in the corresponding format. The Source-Listener solution has been chosen for the implementation. Each module is aware of which one wants to receive the query after having processed it. When thec hain of modules is created, a module can add itself as a listener of another. Then the source of the event calls a specific method of each listener. For this reason, each listener must be conforming to a common specification for all listeners.

• **Metadata:** The Metadata are stored in an XML file. The access to the data is made using the JAXP API. However, like the Metadata Manager does not directly work on the raw data, it is still possible to change the storage type without affecting the whole engine. The data are compliant to an XML Schema but no validation is done because only the software is expected to modify the file.

5.4 HELO based visualization

As mentioned earlier, using ontology is a powerful key in multimedia information retrieval. Still, to be effective, the rendering of the query' result has to be easy to read and simple to navigate through. To address this issue and further enhance the retrieval of scientific recordings of the talks, we designed a set of graphical renderings of information based on the 8 different Scopes of the HELO conference model. Using HELO for visualization provides the users with an interface where the descriptions are grouped based on their use, offering them the ability to explore the information content in an interactive way. In fact, multimedia content description based on a model which structures the characteristics and the relationships of a set of scopes helps the user apply a simple and structured retrieval method. This will help users improve their query formulation by expressing their requirements in a more precise way. Indeed these features become essential in multimedia retrieval which has a complex content hardly searched by "traditional" keyword-based searching. The following section presents a proof-of-concept prototype, NAVIR that focuses on the PersonScope concept. Once the media is retrieved, the users have the possibility either to play the video within the search interface (Youtube-like visualization) or to get redirected to a new window where the video of the recorded talk is played within a conference based visualization interface using

SMAC (Figure 12). The visualization interface presents three distinct blocks: the recorded talk block on the right, the slide set block on the left and the information block on the bottom. The recorded talk is synchronized with the slides extracted from a presentation slide set. The navigation banner at the bottom of the slide set block allows the user to select any slide from the presentation in order to get the corresponding talk video sequence played. Using SMAC to replay the recorded talks provide users simple and easy navigation through the different sequences of the talk, in addition to the display of annotations, provided by HELO, for each video sequence. At any time the users can go back to the search interface to perform other queries.

Figure 12. Talk replay interface

5.4.1 Social network visualization prototype

NAVIR exploits the CALIMERA ontology model. It focuses on the scientific community member relationships and their related multimedia data (scientific publications, recorded talks, etc.). To help making the good choices for the visualization techniques, we have first explored the information visualization domain, and especially relative to the scientific domain. This area has been explored through the completion of a literature review, to carefully examine the different visualization techniques available. This study was conditioned by Daniel Keims' classification (33), which separates the domain into three main axes: data to be visualized (one / two / multi-dimensional, hierarchies, graphs, etc.), visualization techniques (standard 2D/3D display, stacked display, etc.), interaction and distortion techniques (projection, filtering, distortion, etc.). This study has helped to motivate the choices of visualization and interaction techniques based on the data to visualize. Technology and Framework choice: Among the existing libraries providing interesting navigation features, we can cite Processing (34), javascript InfoVis Toolkit (35), Raphael (36), Prefuse library (37) available in a version for Adobe Flash technology called Flare (38), etc. Due to the wide range of provided layouts, Flare library was chosen to develop our prototype. As shown in the figure 13, the resulting application is composed of two visualizations: the social graph on the left and the related multimedia content visualization on the right.

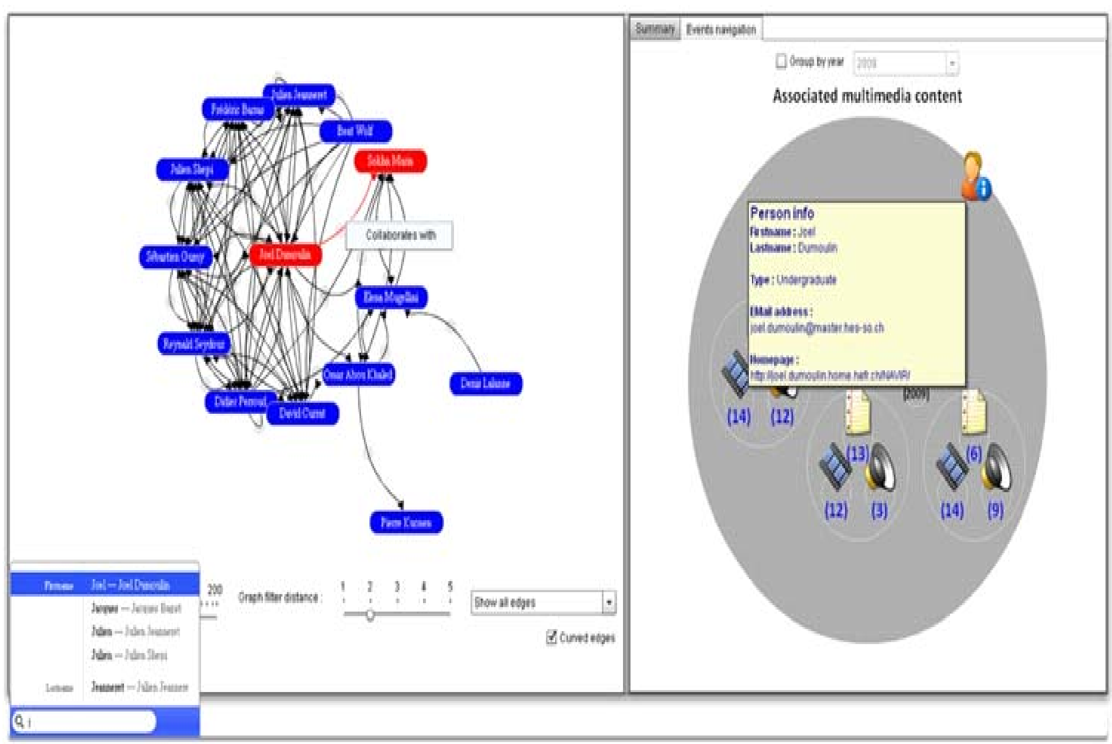

Figure 13. NAVIR interface overview: Social graph (left), related multimedia resources (right)

Social graph

Using a social graph enhances the community browsing by offering a map of the connections between the persons. The interface integrates several features facilitating the browsing and navigation. Each node represents a person and the links between the nodes represent the relationships between the persons.

• **Relationships**: by passing over an arrow linking two nodes the concerned nodes are highlighted and the type of relation is popped (Figure 14).

• **Focusing**: by double clicking on a node we set the focus on the chosen node.
• **Dragging**: the nodes can be dragged, allowing modifying the nodes position.

• **GraphDistanceFilter**: the depth of the tree can change allowing a better overview of the social network related to the focused

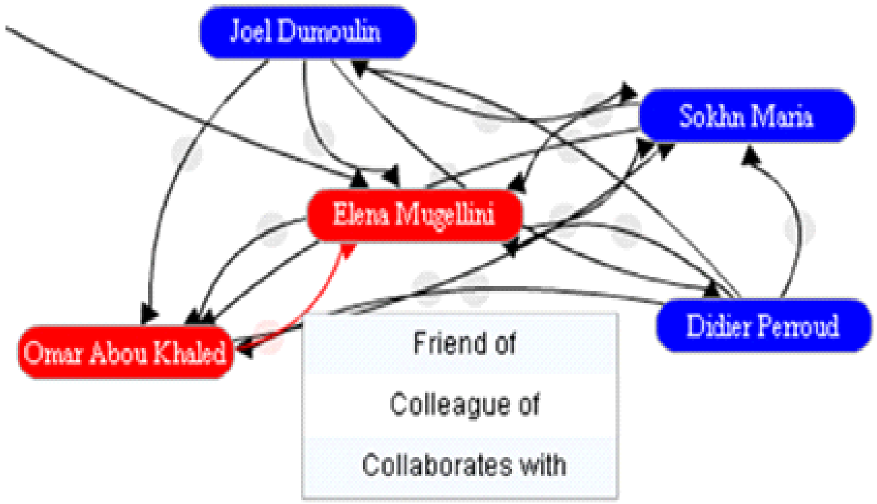

Figure 14. Social graph: Relationships between persons

node. More features such as the searching bar, the highlighting, etc. are described within the accompanying video.

Related multimedia: The related multimedia content interface offers users the possibility to search with the multimedia resources related to the person selected within the social graph visualization. The displayed resources are grouped by conference event and can be filtered by date. Several features are offered by the interface:

• **Person info:** by passing over person icon, general information relative to the selected person is displayed (Figure 13, right part).

• **Event info**: by passing over an event circle, information about the event is displayed (Figure 15).

• **Related multimedia resources**: by clicking on an event circle, we expand its content showing the related multimedia resources such as the videotaped talk which can be accessed by clicking on it.

6. Conclusion and future work

In this paper we presented the scientific conference model HELO (High-level modEL for cOnference) used to enhance the retrieval process of talk recordings (annotation, querying and visualization). HELO structures the knowledge and information conveyed along a conference life cycle. It was built on the basis of a study that analyses the user's needs and requirements when searching for videos of recorded talks. Based on this analysis, 8 different Scopes were identified: MediaScope, EventScope, TypeScope, ThematicScope, TemporalScope, PersonScope, CommunityScope, LocationScope. The model is used to perform and build efficient content video annotations, consequently allowing users to submit complex queries. It is also used to provide users with a graphical search interface based on the Scopes and therefore on their requirements. And finally HELO has been used to enhance browsing, navigation and data accessibility. Different prototypes have been presented in this paper as a proof-of concept. In the future, our goal is to enhance the HELO based visualization part from its current basic stage, integrate the virtual query engine within our framework and evaluate its global performance. This evaluation focuses mainly on three areas:

1. The evaluation of the ontological model HELO: This evaluation should validate the adaptability of the scopes and their coherence from different users' point of view. It should also measure the facility of learning and using this model.

2. The evaluation of the semantic search engine interface: This interface should be evaluated in a set of experiments to verify both its expressiveness and its effectiveness.

3. The evaluation of the multimedia rendering and browsing interfaces: This evaluation should compare the proposed interfaces versus the existing ones in different retrieval situations.

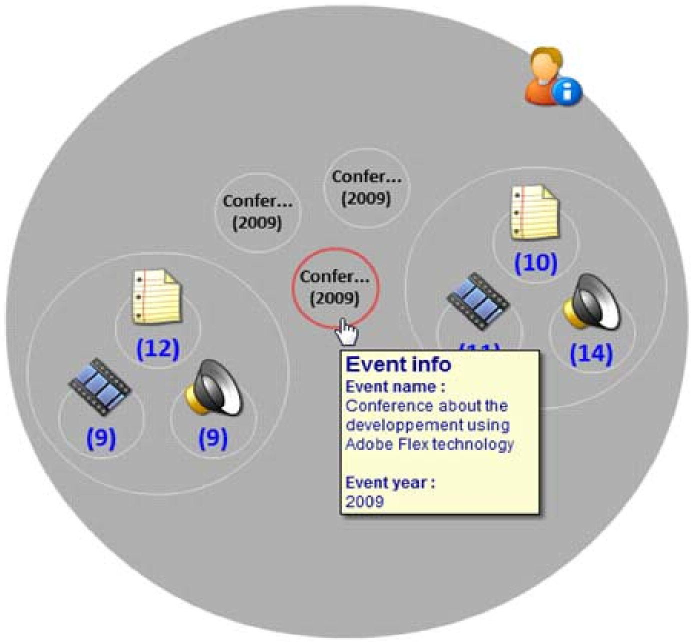

Figure 14. Social graph: Relationships between persons

References

[1] Andrew, D., Bertini, Bagdanov and Marco., Del Bimbo,Alberto., Serra,Giuseppe., Torniai, Carlo (2007). Semantic annotation and retrieval of video events using multimedia ontologies. *In:* ICSC, P. 713-720, 2007.
[2] Smeulders, A.W. M. , Worring , M., Santini , S., Gupta, A., Jain. R. (2000). Content-based image retrieval at the end of the early years, *IEEE Transactions on Pattern Analysis and Machine Intelligence.*
[3] Liu, Y., Zhang, D., Lu, G., Ma,W. Y. (2007). A survey of content-based image retrieval with high-level semantics.
[4] Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., Jain, R. (2001). Content-based image retrieval. *IEEE Trans. Pattern Anal., Machine Intell.*, 23 (9) 1349-1380.
[5] A. P. G. (2006) .PhD thesis, Queen Mary, University of London.
[6] Hollink, L., Worring, M. (2005). Building a visual ontology for video retrieval, *In:* Proceedings of the 13th annual ACM international conference on Multimedia, November 06-11, 2005, Hilton, Singapore.
[7] Knud, M., Tom, H., Siegfried, H.,John, D.: The ESWC and ISWC Metadata Projects. In ISWC2007, pp.802-815, Korea (2007)
[8] J. S. Hare, P.A.S. Sinclair, P.H. Lewis, K. Martinez,P.G.B. Enser, and C. J. Sandom. Bridging the semantic gap in multimedia information retrieval - top-down and bottom-up approaches. In Proc. 3rd European Semantic Web Conference, Budva (Montenegro), June 2006
[9] Bertini, M. Del Bimbo, A. Torniai, C. Soccer Video Annotation Using Ontologies Extended with Visual Prototypes. Content-Based Multimedia Indexing, 2007. CBMI '07. International Workshop, 25-27 June 2007
[10] Vladimir Geroimenko, Chaomei Chen. Visualizing the Semantic Web: XML-based Internet and information visualization. Birkhuser (2006)
[11] Van Harmelen, F. Broekstra, J. ,Fluit, C., ter Horst, H., Kampman, A., van der Meer, J., Sabou, M. (2001). Ontology-based information visualisation, 546-554.
[12] An advanced software framework for enabling the integrated development of semantic-based information, content, and knowledge (ick) management systems. http://www.vikef.net/
[13] Platform for searching of audiovisual resources across online spaces. http://www.pharos-audiovisualsearch.eu/
[14] Rick, K., Reem A. H., William, H., Marilyn M. (1996). Four Paradigms for Indexing Video Conferences. *IEEE MultiMedia arch.* 3. 63-73.
[15] W. H., A. M. (2006). Conkmel: A contextual knowledge management framework to support multimedia e-learning, August.
[16] Fatemi,N. (2003). PhD thesis, A semantic views model for audiovisual indexing and retrieval, Ecole polytechnique federale de Lausanne.
[17] AKT, http://www.aktors.org/publications/ontology/

[18] Conference ontology by Jen Golbeck http://www.mindswap.org/golbeck/web/www04photo.owl/
[19] eBiquity Conference Ontology, http://ebiquity.umbc.edu/ontology/conference.owl/
[20] Semantic Web Technologies ESWC2006, http://www.eswc2006.org/technologies/ontology-content/2006-09-21.html
[21] CERN, Centre Europeen de Recherche Nucleaire, http://www.cern.ch
[22] General introduction to MPEG-7, http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm
[23] RDF, Resource Description Framework, http://www.w3.org/RDF/
[24] OWL, Web Ontology Language, http://www.w3.org/TR/owl-features/
[25] Semantic Web Conference Ontology, http://data.semanticweb.org/ns/swc/swc 2009-02-27.html
[26] Friend Of A Friend Vocabulary Specification, http://xmlns.com/foaf/0.1
[27] Sure, Y. Bloehdorn, S. Haase, P. Hartmann, J., Oberle, D (2005). The SWRC ontology Semantic Web for research communities. In 12th Portuguese Conference on Artificial Intelligence (EPIA 2005), Covilha, Portugal, December 2005. Springer.
[28] Top-Level Categories for the ACM Taxonomy, www.computer.org/mc/keywords/keywords.htm
[29] Crescendo, http://www.crescendo.com/en/speech recognition.php
[30] Dragon Naturally Speaking, http://www.nuance.com/naturallyspeaking/
[31] Sphinx4, http://cmusphinx.sourceforge.net/sphinx4/
[32] Jonquet, C., Musen, M. A., Shah,N. H. (2008). Help will be provided for this task: Ontology-Based Annotator Web Service. Technical Report (2008)
[33] Keim,Daniel A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8. 18.
[34] Processing, http://www.processing.org
[35] Javascript InfoVis Toolkit, http://www.thejit.org
[36] Raphael, http://www.raphaeljs.com
[37] Prefuse library, http://www.prefuse.org
[38] Adobe ash technology, Flare, http://are.prefuse.org

Author Biographies

Maria Sokhn is a PhD student at TelecomParisTech and the University of applied sciences of western Switzerland, Fribourg, EIA-FR. She is member of the Multimedia Information System Group, MISG. She graduated from the Engineering Faculty of Saint Joseph University, ESIB, with a Computer and Communication Engineer diploma, and holds a specialized master diploma from TelecomParisTech in Multimedia production and creation. Now she is working on her PhD thesis at EIA-FR in the MISG team in collaboration with the European Organization for Nuclear Research, CERN and TelecomParisTech. Her researches interest focuses on multimedia information retrieval, knowledge management, modelling and visualization. She is currently the head of the technological centre for eGovernment at the EIA-FR.

Elena Mugellini is Professor at the Information and Communication Department of the University of applied sciences of western Switzerland, Fribourg (EIA-FR). Elena is the leader of MISG research group. She is also member of the Telematics Technology Laboratory at the University of Florence. She holds a PhD in Telematics and Information Society received from the University of Florence in 2006, and a Master in Telecommunication Engineering from the same university received in 2002. Her current research interests are on the areas of Ambient Intelligent, Multimodal Interaction, Tangible User Interface, Personal Information Management, and Document Engineering.

Omar Abou Khaled is Professor in ICT Department of the University of applied sciences of western Switzerland, Fribourg (EIA-FR). He holds a PhD in Computer Science received from the Perception and Automatic Control Group of HEUDIASYC Laboratory of "Université de Technologie de Compiègne", and a Master in Computer Science from the same university. He is Head of MRU "Advanced IT Systems Architects" at EIA-FR. Until 2007 He was leader of the MISG (Multimedia Information System Group). He is responsible of several projects in the field of Document Engineering, Multimodal Interfaces, Context Awareness, Ambient Intelligence, Blended Learning, and Content-Based Multimedia Retrieval. He is involved in the teaching of several courses related to Information Systems, Web Technologies and Multimodal Interfaces.

Ahmed Serhrouchni is an associate professor at the department of Computer Science and Networks at Telecom arisTech, CNRS LTCI-UMR 5141 laboratory.He received his PhD in Computer Science from University Pierre et Marie Curie in 1990. His main research interests include computer networks, security architectures and secure communication protocols, and software environments for their deployment.