FROM VIDEO SEGMENTATION TO SEMANTIC INDEXING:

THE PRESTOSPACE APPROACH

Basili Roberto[(1)], Marco Cammisa[(1)], Laurent Boch[(2)], Alberto Messina[(2)], Giorgio Dimino[(2)], Valentin Tablan[(3)], Borislav Popov[(4)], Werner Bailer[(5)], Walter Allasia[(6)], Michele Vigilante[(6)]

[(1)]*University of Rome, Tor Vergata (Italy), E-mail: {basili,cammisa}@info.uniroma2.it*
[(2)] *RAI Centre for Research and Technological Innovation, (Italy), E-mail: {a.messina,l.boch,g.dimino}@rai.it,*
[(3)] *University of Sheffield (United Kingdom), E-mail: v.tablan@sheffield.ac.uk,*
[(4)] *Ontotext, (Bulgaria), E-mail: borislav@sirma.bg,*
[(5)] *JOANNEUM RESEARCH, (Austria), E-mail: {werner.bailer}@joanneum.at,*
[(6)] *Eurix s.r.l., Multimedia Department (Italy), E-mail: {allasia,vigilante}@eurix.it*

ABSTRACT

This paper will present the contribution of the European PrestoSpace project to the study and development of a *Metadata Access and Delivery* (*MAD*) platform for multimedia and television broadcast archives. The *MAD system* aims at generating, validating and delivering to archive users metadata created by automatic and semi-automatic information extraction processes. The MAD publication platform employs audiovisual content analysis, speech recognition (ASR) and semantic analysis tools. It then provides intelligent facilities to access the imported and newly produced metadata. The possibilities opened by the PrestoSpace framework to intelligent indexing and retrieval of multimedia objects within large scale archives apply as well to more general scenarios where semantic information is needed to cope with the complexity of the search process.

1. Introduction

In the field of image/video processing, the "semantic gap" between high-level semantics needed for indexing audiovisual (AV) material and the low-level features available by automated analysis is often emphasized. There is a need to add and merge semantics from the analysis of available associated modalities, like speech and text. In this context, modern broadcasters have been rediscovering the value of their audiovisual archives and approaches meant to the recovery and availability of archived materials may produce consistent cost savings in the overall programme production processes [1].
Metadata, traditionally defined as "data about data", play a central role here. In the view of the broadcast archives scenario, this entails finding what information and schemes are needed to make archive users retrieve audiovisual items with effective levels of accuracy [2], [3].
In this domain, four basic retrieval patterns can be identified:

- *Retrieving audiovisual items by information*. Starting from the specification of metadata constraints the material for which the stated constraints are valid is to be retrieved. This is the traditional use of the information as "metadata".
- *Retrieving information by audiovisual item*. The access to the archive information relies on the audiovisual material as the carrier of the pieces of information the users are interested in.
- *Retrieving information by information*. In this scenario, information is reached through the use of other pieces of information that act as "metadata" with respect to the target information.
- *Retrieving audiovisual item by audiovisual item*. Audiovisual material is sought and retrieved by means of similarity searches based exclusively on the audiovisual content, i.e. regardless of the expressed meaning and content.

The PrestoSpace MAD partners have undertaken a thorough analysis of these aspects coming to the conclusion that the required information for a typical audiovisual archive exploitation process can be divided in the following fundamental classes:

- *Identification information*, e.g. titles, credits, programme publication information.
- *Editorial parts information,* i.e. information about the relevant editorial sub-items of a programme (e.g. news items).
- *Content-related information,* e.g. text of speech transcript, topics, descriptions, aural and visual low level descriptive features.
- *Enrichment information,* i.e. information coming from external sources generically or topically related to the programme content

1.1. MAD: The general architecture

The MAD Platform adopts a modular, extensible architecture to fit the above requirements. As shown in Figure 1, the MAD Documentation Platform receives as input the digitised media (video and audio files) and produces various materials, such as key frames, camera motions and semantic metadata. These materials are indexed and published on a Web server hosting the MAD Publication Platform (see Section 4).

The MAD Documentation Platform is made up of a core component (the *Core Platform*) and a set of pluggable software processors named *GAMPs* (acronym of "Generic Activity MAD Processor"). It offers the following main services:

- the *Workflow Management* service, which is responsible for starting processes in the correct order and for resolving dependencies between GAMPs;
- the *Essence and Metadata Storage* (*EMS*) system, which stores the audiovisual material sources and the associated metadata;
- the *Concurrent Versioning System*, which tracks every change to the metadata that takes place during the execution of the various GAMPs and is built on a standard CVS engine;
- the *Delivery systems* providing access to the enriched metadata and related materials created within the Documentation Platform.

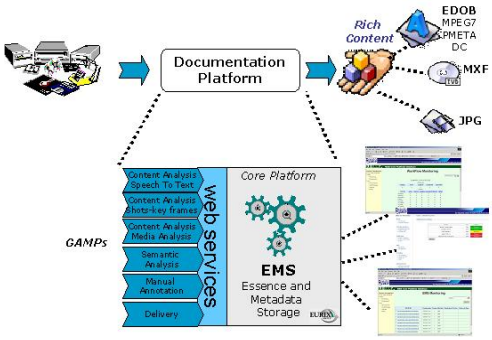

Figure 1 - Architecture of the MAD Platform

The *EMS* stores the materials on the file system, and tracks their location by means of a relational database. It is possible to have many instances of the same material, even located on various machines and accessible through different protocols (file, HTTP, FTP, SMB, ...).
The GAMPs are the software units that extract the metadata from the digitised materials. The Core Platform maintains a queue in the workflow for every GAMP, which will poll it in order to become aware of any activity to be done. In order to do their jobs, the GAMPs ask the Core Platform for the materials and the associated metadata produced up to the request time.
The current experimental instance of the MAD Platform makes use of three different categories of GAMPs: *Content Analysis*, *Semantic Analysis* and *Manual Annotation*. However any kind of new GAMP can be easily added in the future.
The overall services offered by the Core Platform are available through the following (SOAP based) Web Service interfaces: the Workflow, the EMS and the Administration. Using these services, every GAMP can poll the Core Platform asking for a job and related resources. Once the job is completed, it submits the produced data and notifies the completion of the job to the Workflow Manager. The use of web services allows the GAMPs to be written in different programming languages and deployed on different platforms and operating systems.
The overall architecture offers several benefits as follows:

- *Modularity*: GAMPs can be totally different in functionalities and implementation details, and they still interoperate with the Core Platform;
- *Scalability*: in order to add a new GAMP, it is sufficient to add a new process queue to the Workflow engine of the Core Platform;
- *Platform independency*: the GAMPs can be written in any language, provided that this supports SOAP and web services protocols.
- *Multi-tier distribution*: every GAMP can be installed on a different physical system, provided that a network link to the Core Platform exists.
- Furthermore, the Core Platform components (e.g. the EMS, the WF engine and the DBMS) can be installed on different servers as well.

The complexity of the adopted processes required a carefully design of the reference data model. This was achieved through creating a single XML-based document format, taking the best from each of two metadata standards, natively orientated to the description of audiovisual objects, MPEG-7 [4] and P_META [5].

2. Content Processing

Automatic Content Analysis methods for AV content are applied in PrestoSpace to automatically extract metadata from the multimedia material and enrich the description of the content. The automatically extracted metadata is used to help manual annotation: the discovered content structure provides the input to semantic analysis and indexing of the AV objects. A survey of the state of the art of tools for visual, audio and joint audiovisual content analysis is reported in [7].
A set of audiovisual content analysis tools have been selected for use in the documentation process and integrated in the MAD architecture described above. Due to the modular and extensible architecture of the platform, the computationally expensive jobs can be distributed among several clients.
Low-level visual feature extraction. The low-level feature extractor describes key frames or shots in terms of their color, texture and motion features. The tool extracts some of the descriptors specified in the MPEG-7 visual part ([4], part 3), namely *ColorLayout*, *ColorStructure*, *DominantColor*, *EdgeHistogram* and *MotionActivity*. The descriptors serve as a compact and efficient representation of the visual content of a shot and are used to determine visual similarity between shots.
Shot boundary detection. The shot boundary detection tool segments a video in its primary building blocks, i.e. its shots, and is capable of detecting both abrupt (cuts) and gradual transitions (such as dissolves, fades, wipes,

etc.). Shot boundaries are a prerequisite for other visual content analysis algorithms, content structuring and indexing and serve as a navigation support in the manual documentation tool. The approach used for shot boundary detection is an improvement of the one described in [9].

Key frame and stripe image extraction. The key frame detector extracts a number of key frames per shot, depending on the amount of visual change. The key frames serve as representations for the shots and are used as input for low-level feature extraction. Stripe images are spatiotemporal representations of the visual essence, created from the content of a fixed or moving column of the visual essence over time. They serve as a help for quick content overview and navigation, especially in the manual documentation tool.

Camera motion detection. The camera motion detector analytically describes four basic types of camera motion in the content (pan, tilt, zoom, roll), a rough quantisation of the amount of motion, and the length of the segments in which they appear. The algorithm is based on feature tracking. The details of the approach and evaluation results can be found in [8]. Camera motion information is an important search criteria when reusing archive material in new productions and to infer higher level information.

Audio structuring and segmentation. This analysis consists in classifying segments of audio in four principal categories (silence, music, speech, noise). This information is mainly considered as a support for manual annotation.

Editorial parts segmentation. *Editorial parts* are considered by any modern television archivist as the basic indispensable entities for the documentation of an archived programme. They can be defined as the constituent parts of the programme from the editorial point of view, i.e. that of the creators of the programme (e.g. news items in a newscast programme). Several techniques have been investigated to solve the hard problem of identifying editorial parts from the low-level analysis of raw content [7], though none is solving the problem generally. Due to this, the PrestoSpace MAD unit limits the use of automatic editorial segmentation in the news domain, choosing a multi-layer approach that merges visual and aural information for the detection of news items in the mainstream newscast editions.

Reference video clip detection. Reference video clip detection is the task consisting in detecting replica of a reference clip into a visual content. This simple activity is particularly useful when applied in broadcast archived material, where jingles, color bars and other effects are used as visual separators between the parts of a programme. The produced information constitutes one of the inputs to the more complex editorial segmentation task.

3. Dealing with Semantic Information

The MAD platform aims at exploiting human language technologies for Information Extraction (IE) from the AV data made available by large archives. The nature and complexity of management, search and reuse of archive materials require complex storage and retrieval functionalities. These activities ask for:

- Recognition and indexing of *suitable generalizations* of relevant archive concepts as people names, organizations and locations
- Effective retrieval functions that improve indexing at the simple textual level and support *conceptual rather than string retrieval*
- Interoperability at the levels of abstractions required by the AV contents. For example, AV data should be published, queried and exchanged in a distributed fashion. The development of Web publication should support distributed querying and semantic service-based instantiation and invocation. The semantic data descriptions are critical in these activities and interoperable models (ontologies) are needed.

Semantic Analysis is applied in MAD to fit such high-quality requirements from the available multimedia properties (e.g. audio) to suitable generalizations and ontological representation. In the Semantic Web area, the processes going from raw and textual data to ontological annotations are typically called Information Extraction processes.

Thus, the starting point for semantic analysis is the Automatic Speech Recognition (ASR) from the audio data content**.** Extracting text from spoken content of audiovisual material is a fundamental step allowing for several documentation tasks, as well as representing an important core of searchable data in the publication system. In the current set-up of the documentation platform an automatic speech-to-text engine is used, developed by ITC-IRST [6], capable of extracting text from English and Italian.

The redundancy that AV objects guarantee at the data level needs to be explored in order to govern the retrieval complexity at the proper quality. The problems due to noisy nature of the extracted data (e.g. errors in the ASR that produce mistakes in the grammatical recognition) should be properly limited. The aim is to make as much information as possible available to the overall extraction and retrieval components of MAD. In this perspective larger data sets should be taken into account than just the source AV data. The input textual material should be processed and enriched by the following relevant evidences as semantic metadata:

- *Terminological* and *lexical information* local to the AV input data (via ASR)
- Recognition of citations to *Named Entities* (e.g. people or organization) from local data as well as from reachable external sources

- *Automatic computation of useful hyperlinks* between the archived AV data (e.g. the individual segments in broadcasted TV journals) and the distributed sources (e.g. Web-based newspaper portals and pages). These sources include assessed textual descriptions of topics related to the AV segment contents and are trusted.
- *Ontological information* contained in all the above sources, as representation of classes (e.g. geographical locations, organizations or persons), individuals (e.g. *John Coltrane* or *USA/United States*) and topical classes (e.g. *Education* vs. *Sport*, *Foreign Politics* vs. *Economics*).

The extraction of this rich variety of information, required by MAD, is the target of specialized GAMPs called *Semantic Analysis GAMPs*. GAMPS falling in this class are language specific, so that two SA_GAMPs have been designed for Italian and English information extraction respectively. In the following the Italian SA_GAMP will be used as the reference example for the discussion, while technical details of the English ones are found in [17].

3.1. Semantic Analysis In MAD

In MAD, a cascade of processes are used to enrich the multimedia data with semantic metadata. All these processes are organized (or synchronized) by a specific module called *"Workflow Manager"*. This module calls the processors according to their dependencies as shown in Fig. 2. The invoked modules are:

- an *Intaker*, that normalize the input AV segment text obtained by ASR
- the *News Categorizer*, a topical categorization module that assign a specific category to incoming segments
- the *Natural Language Parser*, [10] that recognize lexical units in the ASR transcripts and provides grammatical disambiguation (POS tagging)
- a *Named Entity (NE) Recognizer* that extracts citations of people, organizations, locations and other interesting entities (e.g. dates)
- an *Ontology-based NE recognizer* that links the discovered NE to known individuals and entities of the reference ontology (see Section 3.2)
- a *Web Aligner* that search the Web for pages related (or equivalent) to the source individual AV segments

The Information Extraction chain first applies the *"Intaker"* module. It collects and normalizes the incoming broadcasted news items as they are transcribed and segmented by the speech recognition GAMP. Then, the *News Categorizer* is invoked to assign the suitable topical categories (and an associated confidence scores) according to the target classification scheme. In the Italian semantic analysis GAMP the RAI internal classification scheme has been adopted. Concurrently, these news items can be parsed (via the *"Parser"*) to detect Named Entities (via the *"Named Entity Recognizer"*), these provide a set of significant metadata that can be used by the *Aligner* module to search for candidate news items that are similar to the pages downloaded by a *Web Spider*. The retrieved Web pages are also parsed[1] and indexed according to traditional IR techniques. For each news item, the *Alignment* process selects the suitable Web pages from the set of the retrieved candidates and sets direct hyperlinks to them. These links are used to include further (external) metadata, auxiliary to the internal ones, to improve the overall accuracy that can be affected by wrong or irrelevant information[2]. Finally, a module using an Ontology to annotate the news item is applied (in PrestoSpace, the KIM platform [13] has been used). More details on the Italian SA GAMP can be found in [11], while the English SA GAMP is discussed in detail in [17]. KIM will be further discussed in section 3.2.

3.2. The role of Ontological Information

The ontological component in PrestoSpace is managed by the KIM platform [13] which supports information extraction based on an ontology and a massive knowledge base. The KIM platform implements semantic annotation as an innovative model for automatic semantic content enrichment, it enables new information access methods, and extends the existing ones. In this way, KIM supports applications such as highlighting, indexing and retrieval, categorization, generation of advanced metadata, smooth traversal between unstructured text and the available relevant knowledge. The information extraction approach employed in KIM has roots in the conception that certain entities, a content refers to, are of significant importance for the meaning of the content they appear in. To clarify why named entities constitute an important part of the semantics of the documents, consider the sentence "the first president of the United States". Understanding the meaning of the constituent words is not enough to understand the meaning of the sentence.. Unlike words, named entities denote an (often concrete) individual and not a class or any member of the class. When describing the meaning of words, lexical semantics and/or common sense would suffice; to understand the meaning of named entities more specific knowledge about the world is required.

Semantic annotation is a generation of specific metadata. It is the process of assigning to the named

1 The parsing process is different in the two cases as automatic transcriptions are error prone so that specific grammars are required.

2 Mistakes made by the speech recogniser over ASR transcriptions can be improved by the better quality of the external, i.e. Web originated, metadata.

entities in the text links to their semantic descriptions.

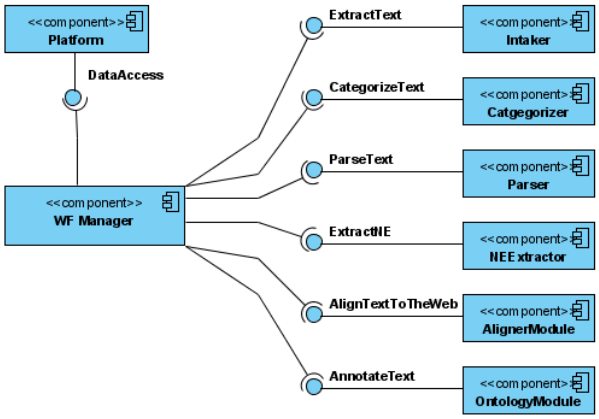

Figure 2: The Ita SA GAMP structure

The semantic annotation process in KIM is based on a simple model of real-world entity classes (i.e. an ontology) and a massive knowledge base. The semantic annotation (metadata) has certain prerequisites for its representation: (1) an ontology (or at least, a taxonomy), which defines the entity classes; (2) entity identifiers, which allow entities to be distinguished and linked to their semantic descriptions; (3) a knowledge base with entity descriptions. KIM relies on two types of ontologies: *upper-level* (PROTON[3]) (roughly domain-independent) and *domain-specific*.

The PROTON upper level ontology encodes the most common aspects of any considerable description no matter of the specificity of the domain (weather forecast, popular science documentary, etc.), regardless of the specificity of the task in view (for example - classification of movies, access to news emissions, description of the themes of the documentaries). PROTON was designed to address the requirement of being suitable for open-domain general purpo semantic annotations as well as to allow easy extensions according to specific needs. It currently contains about 300 classes and 100 properties.

For the purposes of semantic annotation, indexing, and retrieval of documents, KIM also uses a seed *knowledge base* (KB). The knowledge base (KB), in this context, is a body of formal knowledge about entities, a means for the representation of non-ontological formal knowledge. It consists of instance data – descriptions of entities and their interrelations, i.e. for each entity, the KB contains information about the entity's type, aliases (including a main alias, i.e. official or well-known name), attributes, and relations. The KIM KB provides coverage of popular real-world entities of common interest, which are considered well-known and thus not explicitly introduced in the documents. Most important and used entities in the KIM KB are *geographic names and organizations*. The entities representing geographical features are imported from *GNS* (*GEOnet Names Server*) and other sources. They are organized so as to represent instances of *Location* (and its subclasses) having the property *subRegionOf* as it is applied between *Continents*, *GlobalRegions*, *Countries*, and other subclasses of *Location*. Some subtypes of *Location* which are contained in KIM KB are *Country*, *Province*, *County*, *CountryCapital*, *City*, *Ocean*, *Sea*, etc. The locations are given together with several among their aliases, including English and French aliases, as well as with their geographic coordinates (*Long/Lat*), the designator (*DSG*) and Unique Feature Index (*UFI*), according to *GNS*. All this provides a useful basis for cross-linguistic querying and retrieval. The entities in the KB are derived or collected from various sources as geographical and business intelligence gazetteers.

One of the roles of KIM in MAD is to provide a language independent representation for Named Entities as a specific metadata common to the two languages. As an example consider that the *"White House"* is translated in other languages (e.g. in Italian the correct translation is *"Casa Bianca"*). The ontology representation for this entity is via an unique id (i.e. an Uniform Resource Identifier *"URI"*), that is for its nature language independent. This realizes a systematic and consistent approach to multilingual indexing and searching.

4. Information Retrieval IN MAD

The rich variety of information extracted by the GAMPs poses several requirements to the Information Retrieval functionalities in the publication phase. First, the user interface should model access methods according to different (and integrated) capabilities:

- *Full text search* as usually applied by mostly popular search engines
- *Natural Languages Questions*
- *Semantic browsing* as navigation through concepts, relations and instances of the ontology

All the above functionalities are to be intended as language neutral: full texts should be searchable in different languages, while ontological information as well as NEs should be properly represented so that language ambiguity and variability are taken into account. Second, all the above search modalities should be offered in a language independent fashion. The discussion of technological solutions to support the above processes is reported below in this document, as they have a relevant impact on the accuracy reachable by the PrestoSpace solutions to CLIR issues.

The viable solutions to the above problem concern:

- The adoption of language neutral representation (via the KIM ontology)

[3] PROTON: see http://proton.semanticweb.org/

- Query processing (expansion and translation) for dealing with multilingual information during search

4.1. Ontology-driven retrieval

In MAD, the KIM platform is in charge of making extensive ontological knowledge about the news domain available, and supporting indexing and navigation functionalities. It provides a novel Knowledge and Information Management infrastructure and services for automatic semantic annotation, indexing, and retrieval of unstructured and semi-structured content/documents. It differs from other systems and approaches in that by providing semantic annotations it supports also IR services based on the results.
Different KIM front-end user interfaces are possible given the KIM API, which provides the functionality and infrastructure for the semantic annotation, indexing and retrieval, as well as document management, and KB navigation. The KIM web user interface (KIM Web UI, Fig. 3) allows traditional access methods (key word search) and semantic ones (entity search, pattern search), too. Via semantic search the user is allowed to express querying about specific entities restricted by formal constraints over properties. This can be done by navigating the ontology or filling special purpose templates. The interface can return either a set of entities that satisfy the query or the set of documents that refer to these entities. The user can access the document content enriched with the associated metadata on the document level (such as title, author, the target entities, ...)
A *plug-in* for Internet Explorer browser has also been created. The KIM plug-in provides lightweight semantic annotations of the Web pages displayed in the browser. On a specific tab (on the left), the plug-in displays the entity type hierarchy (a branch of the KIM ontology). Each entity type has an associated color used to highlight the annotations of this type. Check boxes for each entity, allow the user to select or hide the different entity types (and colors). This way the user can directly navigate from the annotations to the instances that they are linked to in the KB. Via this explorer, the KB could be further explored by choosing one of the related entities, or the entity class.
More details about the KIM and PROTON technology for Ontology-driven IR can be found in [14, 13].

4.2. Cross-language Retrieval

Cross-language Information Retrieval (CLIR) is supported in the MAD publication platform by a specific server called *CLIR Server*.
As described in Figure 4, the CLIR Server includes several components:

- The *NL Parser*, to extract *Named Entities* and other nouns from the query *q*, in the source language *L*;
- *Pseudo Context Generator*, to generate for each target lexical item *t* in *q*, the most relevant terms that are topically related to *t*;
- *Sense Disambiguator*, to disambiguate all common nouns in the source language *L*;
- *Translator*, to translate the disambiguated common nouns from the source language *L* to the target language *L2*;
- *Kim Server*, to annotate the ontological entries as they are found in a query *q*;
- *Text Categorizer* that classifies the query *q*.

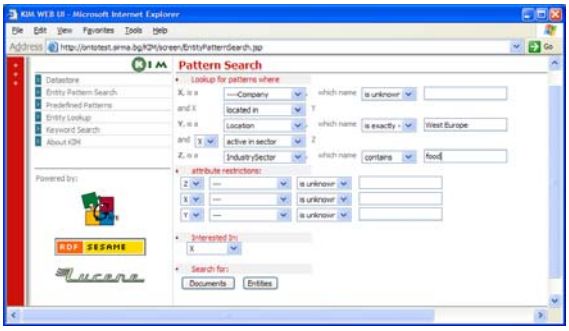

Figure 3: Ontological queries to the KIM platform

The CLIR Server communicates with these components and manages the internal workflow. The NL parser, Text Categorization and Kim annotation processes are the same used for the Semantic Analysis GAMP.
A distinctive feature of the CLIR server is the adopted technique for Sense Disambiguation and Translation. Translation of all common nouns is required as they are very language specific and must be consistently combined to the language-independent representation of Named Entities.
The sense disambiguation algorithm adopted has been presented in [12]. The aim of the method is to automatically extend the information about a query via text mining techniques, disambiguate nouns through Wordnet senses and use them to select suitable translations in the target language.
In particular, a query expansion process is first applied through a *Latent Semantic Analysis* (LSA, [15]) approach. The initial query *q* is mapped into an LSA space (previously obtained from news corpora in both languages): this allows to associate to all nouns in *q* the closest terms, i.e. a lexicon *dom(q)* associated to the *q*'s topical domain (*Pseudo Context Generation*). Within this lexicon a sense disambiguation process is applied: an *n*-ary similarity metric (see [16]) is used here to rank Wordnet senses of individual nouns in *q* given *dom(q)* (*Sense Disambiguation*). As sense ambiguity is much lower within a domain, the sense disambiguation in *dom(q)* is very effective. Preferred senses are finally used to generate translations (*Translation*). The interlingual interfaces of Wordnet, in fact, link *synsets* in different languages. The best senses (*synsets*) for nouns in *q* are then used to derive their best translations

in the target language. The resulting query includes named entities, query category and all the synonyms of original nouns in the target language. The method runs in a fully automatic way as LSA can be applied without human intervention. The sense disambiguation algorithm is much more effective when combined with LSA as discussed in [12].
As an example a natural language query and the results of the CLIR are described in the following Table, where individual translations are showed.

Input Query	*Blair calls on NATO member to contribute more troops to Afghan force.*		
Chaos NEs	Blair [person] NATO [organisation]		
KIM NEs	NATO [Organization] Blair [Person]		
Nouns, Translations	**Noun**	**Input language senses**	**Target language senses**
	NATO	**North_Atlantic_Treaty_ Organization, NATO**	**n.a.t.o., organizzazione_del_trattato_nordatlantico**
	member	Member	componente, membro
		penis, phallus, member	asta, fallo, membro, membro_virile, pene, verga
		Member	appartenente, componente, iscritto, membro
		extremity, appendage, member	arto, estremita', membro
		Member	**membro**
	troops	**military_personnel, soldiery, troops**	**truppa**
	force	Force	forza
		military_unit, military_force, military_group, force	**arma**
		violence, force	forza, violenza
		effect, force	effetto, forza
		force, persone	forza, personale
		force, forcefulness, strength	corpo, energia, forza, lena
Output Query	Person:Blair & Organization:Nato & (n.a.t.o | "organizzazione del trattato nordatlantico") & membro & truppa & arma		

An example of translation from Italian to English is shown in the following Table:

Input Query	*Berlusconi al parlamento sulla missione di guerra in Iraq.*		
Chaos NEs	Berlusconi [person] Iraq [paese]		
KIM NEs	Iraq [Location]		
Nouns, Translations	**Noun**	**Input language senses**	**Target language senses**
	parlamento	**parlamento**	**parliament**
	missione	**delegazione, deputazione, missione, rappresentanza**	**deputation, commission, delegation, delegacy, mission**
		missione	mission, military_mission
	guerra	guerra	war, warfare
		battaglia, combattimento, conflitto, guerra, lotta, scontro	battle, conflict, fight, engagement
		discordia, disunione, guerra, zizzania	discord, strife
		guerra	**strife**
Output Query	Person:Berlusconi & Location:Iraq & parliament & (deputation | commission | delegation | delegacy | mission) & strife		

4.3. User Browsing In Publication

The MAD Publication Platform provides retrieval and browsing functionalities. It deals with instances of documents conforming to the MAD metadata format and makes them available in a web-based representation. It also gives access to the materials exported from the Core Platform.
The Publication Platform architecture is based on a Web application as user interface, a DBMS storing the available information related to programmes, and the KIM indexing and search engine. The Publication Platform offers two main features:

a. *data import* for submitting (and index) materials
b. data search and browsing

The search interface supports the various retrieval approaches described in the previous sections, and the user can choose the target of his/her search (e.g. a programme or a news item), which can be filtered by title, broadcast date and service, contributions (e.g. authors, journalists, directors), classification (topics, categories), text of description.

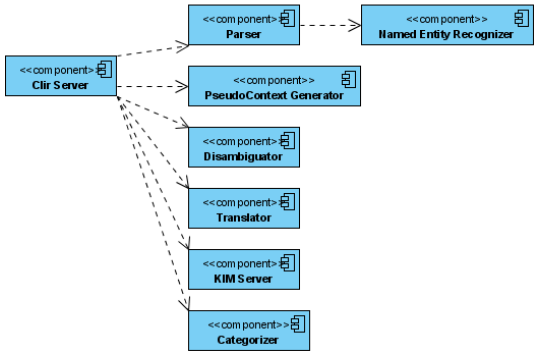

Figure 4: The CLIR Server

As the user selects an item (table row), a *browsing* window is opened which presents all the details of the specific item. The window is made up of four frames: a video preview, the editorial parts tree, the key frames, and an extensible multi-tab frame, each of which is representing a specific elaboration result. The content of all the frames is synchronised during user interaction.
The following tabs of the multi-tab frame have been implemented so far:

- *Info*. It contains the general metadata about the programme such as title, subtitle, publication dates and channels, contributors.
- *Transcriptions* (see Figure 5). This tab shows the output of the speech-to-text GAMP. The text is divided into segments representing individual news items. The interface also allows the user to select a specific text segment.
- *Semantic analysis* (see Figure 6). This tab shows a navigable tree that can be explored interactively. It shows the entities found by the semantic GAMPs.
- *Content analysis*. Here the user can view the stripe images and the related camera motion information on the timeline.

The video preview makes use of Windows Media Player and allows synchronization among all the available tabs. The user can experience different browsing approaches. For example, it is possible to navigate the items of a programme using the tree located just below the player, and at the bottom of the page key frames and visual shots segmentation are browsable. By selecting the corresponding tabs it is possible to view information about programmes and news items, to view audio transcripts aligned with the timeline, and classifications on the semantic analysed content.

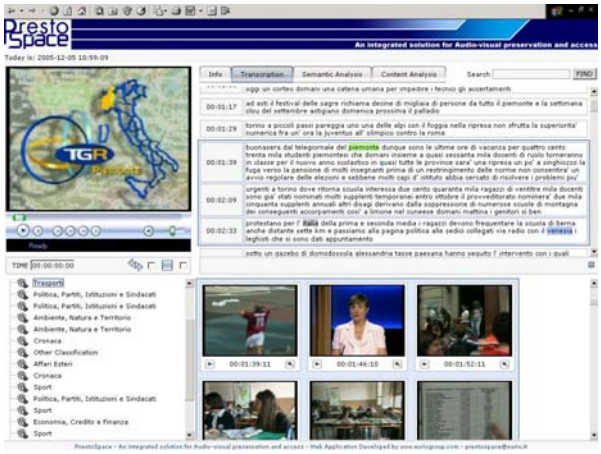

Figure 5 - Speech to text visualization

5. Conclusions

In this paper, the approach to multimedia indexing developed within the PrestoSpace project has been discussed. Benefits range from a substantial enrichment and generalization of the input raw material during the documentation stage to the enabling of advanced information retrieval, such as ontology-driven browsing of AV data as well as multilingual NL queries. Although some of the technologies have not been subject of extensive quantitative evaluation yet, the early qualitative data analysis suggests that the overall reachable accuracy is very good. The general framework proposed by the PrestoSpace technology thus opens the way for a variety of applications, including processing of AV data different from TV broadcasted news and automation of semantic extraction related to more complex information (e.g. events, as relations among individuals and concepts detected in the AV input). The availability of large archives digitised and enriched with semantic metadata will enable future extensions like cross-media mining and development of more advanced user interfaces based on social network and virtual community models.

6. Acknowledgment.

This paper has been partially funded under the IST PrestoSpace project, EU - IST - FP6-2002-IST-1.

7. REFERENCES

[1] R. Del Pero, G. Dimino, and M. Stroppiana, "Multimedia Catalogue – the RAI experience", *EBU Technical Review nr. 280*, European Broadcasting Union, Geneva, Summer 1999, pp. 1-13.

[2] A. Messina, and D. Airola Gnota, "Automatic Archive Documentation based on Content Analysis", *IBC 2005 Conference Publication, International Broadcasting Convention*, Amsterdam, September 2005, pp. 278-286.

[3] A. Messina, "Documenting the Archive using Content Analysis Techniques", *EBU Technical Review nr. 305*, European Broadcasting Union, Geneva, January 2006.

[4] ISO/IEC 15938, *Multimedia Content Description Interface*.

[5] EBU Tech3295, European Broadcasting Union (EBU) P_META Metadata Exchange Scheme.

[6] Brugnara, F., Cettolo, M., Federico, M., and Giuliani, D. (2000). A system for the segmentation and transcription of Italian radio news. In *Proceedings of RIAO, Content-Based Multimedia Information Access*, Paris, France.

[7] W. Bailer, F. Höller, A. Messina, D. Airola, P. Schallauer, M. Hausenblas, State of the Art of Content Analysis Tools for Video, Audio and Speech, Deliverable 15.3 of the IST PrestoSpace project, March 2005.

[8] W. Bailer, P. Schallauer G. Thallinger, "Joanneum Research at TRECVID 2005 – Camera Motion Detection", Proc. of TRECVID Workshop, Gaithersburg, MD, USA, Nov. 2005.

[9] W. Bailer, H. Mayer, H. Neuschmied, W. Haas, M. Lux, W. Klieber, "Content-based video retrieval and summarization using MPEG-7", Proc. Internet Imaging V, San Jose, CA, USA, Jan. 2004, pp. 1-12.

[10] Basili R., F.M. Zanzotto, Parsing Engineering and Empirical Robustness, 8 (2/3) 97120, *Journal of Language Engineering*, Cambridge University Press, 2002

[11] Roberto Basili, Marco Cammisa, Emanuale Donati, RitroveRAI: A Web Application for Semantic Indexing and Hyperlinking of Multimedia News, in "*International Semantic Web Conference*", Y. Gil, E. Motta, V.R. Benjamins, M.A. Musen Eds., Lecture Notes in Computer Science, LN 3279, 97-111, 2005.

[12] R. Basili M. Cammisa, A. Gliozzo, Integrating Domain and Paradigmatic Similarity for Unsupervised Sense Tagging, *Proceedings of the European Conference on Artificial Intelligence*, Riva del Garda, (Italy), 2006.

[13] A. Kiryakov, B. Popov, D. Ognyanoff, D. Manov, A. Kirilov, M. Goranov, Semantic Annotation, Indexing, and Retrieval. *Elsevier's Journal of Web Sematics*, Vol. 2, Issue (1), 2005.

[14] B. Popov, A. Kiryakov, D. Ognyanoff, D. Manov, A. Kirilov, KIM - a semantic platform for information extraction and retrieval, *Journal of Natural Language Engineering*, Vol. 10, Issue 3-4, Sep 2004, pp. 375-392, Cambridge University Press.

[15] Berry, M.W., Dumais, S.T., O'brien, G.W. Using linear algebra for intelligent information retrieval, *SIAM Review*, Vol. 37, No. 4, pp. 573-595, December 1995.

[16] R. Basili, M. Cammisa, F.M. Zanzotto, A semantic similarity measure for unsupervised semantic disambiguation, *Proceedings of the Language, Resources and Evaluation LREC 2004 Conference*, Lisbon, Portugal, 2004.

[17] M. Dowman, V. Tablan, H. Cunningham and B. Popov. Web-Assisted Annotation, Semantic Indexing and Search of Television and Radio News. *14th International World Wide Web Conference*. Chiba, Japan, 2005.

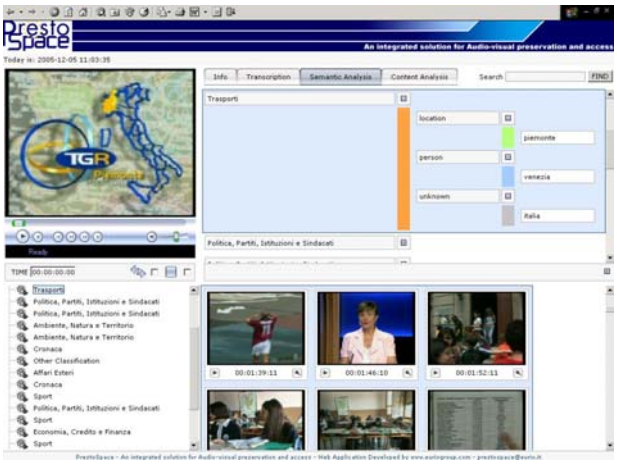

Figure 6 - Publication Platform: Semantic analysis