SEMANTICS AND MACHINE LEARNING FOR BUILDING THE NEXT GENERATION OF JUDICIAL COURT MANAGEMENT SYSTEMS

E. Fersini, E. Messina
University of Milano-Bicocca, Viale Sarca 336, Milan, Italy

D. Toscani, F. Archetti, M. Cislaghi
Consorzio Milano Ricerche, Via Cozzi 53, Milan, Italy

Keywords: e-Justice, Digital trial folder, Information extraction and retrieval, Speech processing, Video analysis.

Abstract: Information and Communication Technologies play a fundamental role in e-justice: the traditional judicial folder is being transformed into an integrated multimedia folder, where documents, audio and video recordings can be accessed and searched via web-based judicial content management platforms. Usability of the electronic judicial folders is still hampered by traditional support toolset, allowing search only in textual information, rather than directly in audio and video recordings. Transcription of audio recordings and template filling are still largely manual activities. Thus a significant part of the information available in the trial folder is usable only through a time consuming manual search especially for audio and video recordings that describe not only what was said in the courtroom, but also the way and the specific trial context in which it was said. In this paper we present the JUMAS system, stemming from the JUMAS project started on February 2008, that takes up the challenge of using semantics towards a better usability of the multimedia judicial folders. The main aim of this paper is to show how JUMAS has provided the judicial users with a powerful toolset able to fully exploit the knowledge embedded into multimedia judicial folders.

1 INTRODUCTION

The use of Information and Communication Technologies (ICT) is considered one of the key elements for making judicial folders more usable and accessible to the interested parties, reducing the length of judicial proceedings and improving justice. The progressive deployment of ICT technologies in the courtroom (audio and video recording, document scanning, courtroom management systems), jointly with the requirement for paperless judicial folders pushed by e-justice plans (Council of the European Union, 2009), are quickly transforming the traditional judicial folder into an integrated multimedia folder, where documents, audio recordings and video recordings can be accessed usually via a web-based platform (Velicogna, M, 2008). This trend is leading to a continuous increase in the number and the volume of case-related digital judicial libraries, where the full content of each single hearing is available for online consultation. A typical trial folder contains:

- audio hearing recordings
- audio/video hearing recordings
- transcriptions of hearing recordings
- hearing reports
- attached documents (scanned text documents, photos, evidences, etc.)

The ICT container is typically a dedicated judicial content management system (court management system), usually physically separated and independent from the case management system used in the investigative phase, but interacting with it. Most of the present ICT deployment has been focused on the deployment of case management systems and ICT equipment in the courtrooms, with content management systems at different organisational levels (court or district). ICT deployment in the judiciary has reached different levels in the various EU countries, but the trend toward a full e-justice is clearly in progress. Accessibility of the judicial information,

Fersini E., Messina E., Toscani D., Archetti F. and Cislaghi M..
SEMANTICS AND MACHINE LEARNING FOR BUILDING THE NEXT GENERATION OF JUDICIAL COURT MANAGEMENT SYSTEMS.
DOI: 10.5220/0003099300510060
In *Proceedings of the International Conference on Knowledge Management and Information Sharing* (KMIS-2010), pages 51-60
ISBN: 978-989-8425-30-0

both of case registries, more widely deployed, and of case e-folders, has been strongly enhanced by state-of-the-art ICT technologies. Usability of the electronic judicial folders is still affected by a traditional support toolset, being information search limited to text search, transcription of audio recordings (indispensable for text search) is still a slow and fully manual process, template filling is a manual activity, etc. Part of the information available in the trial folder is not yet directly usable, but requires a time consuming manual search. Information embedded in audio and video recordings, describing not only what was said in the courtroom, but also the way and the specific trial context in which it was said, still needs to be exploited. While the information is there, information extraction and semantically empowered judicial information retrieval still waits for proper exploitation tools. The growing amount of digital judicial information calls for the development of novel knowledge management techniques and their integration into case and court management systems. Several EU research projects have proposed actionable models of legal knowledge (ESTRELLA, METALEX), and investigated interoperability of legal documents and the potential of text-based semantic analysis. Research about multimedia judicial trial folder and courtroom technologies in criminal trials has been addressed by e-COURT project and SecurE-Justice projects in FP5 and FP6, with the main objective of the digital trial folder and its secure accessibility. JUMAS project (JUdicial MAnagement by digital libraries Semantics), started of February 2008 and under validation in the Court of Wroclaw (Poland) and in the Court of Naples (Italy) with the support of the Polish and Italian Ministries of Justice, faces the issue of a better usability of the multimedia judicial folders, including transcriptions, of information extraction and semantic search, to provide to users a powerful toolset capable to fully address the knowledge embedded in the multimedia judicial folder. The JUMAS project has several scientific objectives:

- Knowledge Models and Spaces: Search directly in the audio and video sources without a verbatim transcription of the proceedings.
- Knowledge and Content Management: Exploit hidden semantics in audiovisual digital libraries in order to facilitate search and retrieval, intelligent processing and effective presentation of multimedia information. Research addresses also multiple cameras and audio sources.
- Multimedia Integration: Information fusion from multimodal sources in order to improve accuracy in automatic transcription and annotation phases.
- Effective Information Management: Streamline and Optimize the document workflow allowing the analysis of (un)structured information for document search and evidence base assessment.
- ICT Infrastructure: Service Oriented Architecture supporting a large scale, scalable, and interoperable audio/video retrieval system.

In this paper we present how the results of JUMAS can help to meet the challenges in analyzing audio and video recordings and outline the impact on ICT infrastructure in the court.

2 THE JUMAS SYSTEM

2.1 The JUMAS Concept

In order to explain the relevance of the JUMAS objectives we report some volume data related to the judicial domain context. Consider for instance the Italian context, where there are 167 courts, grouped in 29 districts, with about 1400 courtrooms. In a law court of medium size (10 court rooms), during a single legal year about 150 hearings per court held with an average duration of 4 hours. Considering that approximately in 40% of them only audio is recorded, in 20% both audio and video while the remaining 40% has no recording, the multimedia recording volume we are talking about is 2400 hours of audio and 1200 hours of audio/video per year. The dimensioning related to the audio and audio/video documentation starts from the hypothesis that multimedia sources must be acquired at high quality in order to obtain good performances in audio transcription and video annotation, which will in turn affect the performance connected to the retrieval functionalities. Following these requirements one can figure out a storage space of about 8.7 MB/min for audio and 39 MB/min for audio/video. The total amount of data to process in one year in a court is summarized by the following table:

Table 1: Dimension of the problem.

Hypothesis of Data Amount per Year /court		
	Hearing Duration (hrs)	Required Space (TB)
Audio	2400	1,2
Audio/Video	1200	2,8

During the definition of the space dimension required on a single site, the estimation will also take into account that a trial includes some additional data: (1) textual source as for example minutes in .doc and .pdf

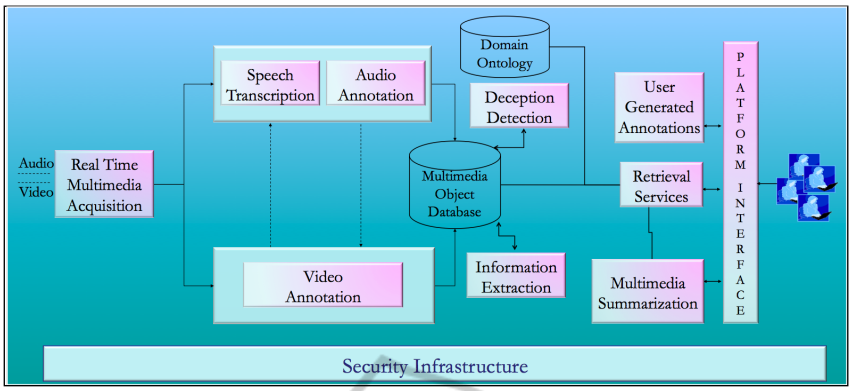


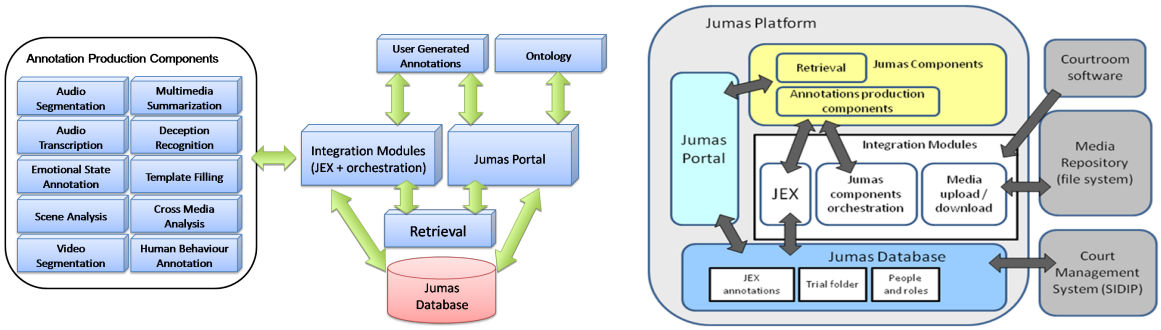
Figure 1: Information flow in JUMAS.

format; (2) images; (3) other digital material. Under these hypotheses, the overall size generated by all the courts justice system only for Italy in one year is about 800 terabyte; this shows how the justice sector is a major contributor to the data deluge (The Economist, 2010). The typical deployment of JUMAS is at district level (about 5 courts), in which it is performed the coordination among courts. Installations in specific courts can be justified by the needs of particular trials, which involve huge quantities of very sensitive data (e.g. class actions, organized crime). The issue of scalability has to be considered for a proper dimensioning of the hardware infrastructure in which to run the system: the most computationally intensive part of JUMAS is the automatic audio transcription; the transcription engine based on an ensemble of Hidden Markov Models requires high performance computing. Also video processing has high computational needs. In order to manage such quantity of complex data, JUMAS aims to:

- Optimize the workflow of information through search, consultation and archiving procedures.
- Introduce a higher degree of knowledge through the aggregation of different heterogeneous sources.
- Speed up and improve decision processes discovering and exploiting knowledge embedded into multimedia documents, in order to consequently reduce unnecessary costs.
- Model audio-video proceedings in order to compare different instances.
- Allow traceability of the proceedings during their evolution.

2.2 System Architecture

The architecture of JUMAS is based on a set of key components: a central database, a user interface on a web portal and the integration and orchestration modules which allows composing several media analysis components, as described in the following. Figure 1 presents the information flows in JUMAS: the media stream recorded in the courtroom includes both audio and video which are analyzed to extract semantic information used to populate the multimedia object database. The outputs of these processes are Annotations, which are the common name in JUMAS to identify tags that are attached to media streams and stored in the database (Oracle 11g). The links between audio and video analysis components in Figure 2(a) show that the algorithms apply exchange information, using the annotations produced by one component as input to improve the performance of the other (e.g. through face recognition performed on video is possible to identify changes in speaker, which is used to cut audio streams to separate transcriptions). Figure 2(b) displays a higher level overview of the JUMAS components and the software architecture, which shows also the links with the external courtroom infrastructure. The integration among modules is performed through a workflow engine and a module called JEX (JUMAS Exchange library). The workflow engine is a service application that manages all the modules for audio and video analysis (described in Sect. 3). It is written in Java; it defines an entity model for annotations with a corresponding XML schema for web services and a database schema for Hibernate persistence. JEX provides a set of services to upload and retrieve Annotations to and from the JUMAS database; these services are hosted in a J2EE application server, with SOAP support for direct usage. The integration of the JUMAS components is supported by a client library that allows manipulating objects that are stored in the DB through web services or direct database access, without having to cope with web services, SQL or XML. A client console multi-platform application exposes all the functionality of the JEX services through XML



(a) JUMAS components and data flows

(b) Software architecture of JUMAS

Figure 2: The JUMAS system.

files. This integration schema allows clear and simple data exchange, making JUMAS flexible enough to be deployed in different configurations, including all or part of its components, and is also open to future improvements with the inclusion of new modules. In JUMAS it has been defined a wide set of annotations categories and events, to deal with the potentially wide set of annotations produced by the media analysis modules. The JEX client library (and the applications) can use two communication protocols (web services and database) because each one has pros and cons: web services are easier to be accessed by remote modules; moreover, this approach can be used directly in different languages/platforms. On the other side, the direct database access does not require the presence of the JEX service (i.e. of the J2EE application server). The users interact with JUMAS through the JUMAS Portal, a web application whose server runs where JUMAS is deployed. Clients can access the data through secure authentication. JUMAS actually integrates a court management system (namely SIDIP - Sistema Informativo DIbattimento Penale - (as shown in Figure 2(a)) for the creation of the trial folder, the management of lawyers and judges registries and the attachment of textual documents. The same results can be obtained by including alternative Court Management Systems (CMS). All the annotations produced by the JUMAS components and stored in the Database serve as basis for the Hyper Proceedings View described in Sect. 5.1.

3 KNOWLEDGE EXTRACTION

3.1 Automatic Transcription

A first fundamental information source, for a digital library related to the courtroom debate context, is represented by the audio recordings of actors involved into hearings/proceedings. The automatic transcription is provided by an Automatic Speech Recognition (ASR) system (Falavigna et al., 2009) trained on real judicial data coming from courtrooms. Currently two languages, Italian and Polish, have been considered for inducing the models able infer the transcription given the utterance. Since it is impossible to derive a deterministic formula able to create a link between the acoustic signal of an utterance and the related sequence of associated words, the ASR system exploits a statistical-probabilistic formulation based on Hidden Markov Models (Rabiner and Juang, 1993) In particular, a combination of two probabilistic models is used: an acoustic model, which is able to represent phonetics, pronounce variability and time dynamics (co-utterance), and a language model able to represent the knowledge about word sequences. The audio acquisition chain in the courtroom has been designed specifically to improve the Word Error Rate (WER), using lossless compression such as FLAC and cross-channels analysis. This allows a good trade-off between the conflicting needs of a manageable dimension of the audio file and good quality recording. The ASR system developed was facing with several domain constraints/limitations:

- Noisy audio streams: the audio stream, recorded during a judicial trial, can be affected by different types of background environmental noises and/or by noises caused by the recording equipment.
- Spontaneous Speech: the sentences uttered by a speaker during a trial are characterized by breaks, hesitations, and false starts. Spontaneous speech - with respect to read utterance - plays a fundamental role in ASR systems, by originating higher values of WER.
- Pronounce, language and lexicon heterogeneity: the actors interacting during a trial may be different for language, lexicon and pronounce types. In particular, the judicial debates could contain many

words (e.g. person names, names of Institutions or Organizations, etc.) that are not included in the dictionary of the ASR, thus increasing the number of out-of-vocabulary words and, consequently, the resulting WER.

- Variability of a vocal signal: the word sequences provided by the ASR can be influenced by different circumstance such as posture of the speaker, emotional state, conversation tones, and different microphone frequency responses. These elements introduce perturbations that are difficult to be taken into account during automatic speech transcription activities.
- Non-native speakers: the actors involved in a proceeding can be characterized by linguistic difficulties or can be non-native speaker. These linguistic distortion negatively affect the accuracy of the produced transcriptions.

Currently the ASR modules in the JUMAS system, offer a 61% accuracy over the generated automatic transcriptions and represent the first contribution for populating the digital libraries behind the judicial trials. In fact, the produced transcriptions are the main information source that can be enriched by other modules and then can be consulted by the end users through the information retrieval system.

3.2 Emotion Recognition

Emotional states represent a bit of knowledge embedded into courtroom media streams. This kind of information represents hidden knowledge that may be used to enrich the content available in multimedia digital libraries. The possibility for the end user to consult the transcriptions, also by considering the associated semantics, represents an important achievement that allow them to retrieve an enriched written sentence instead of a flat one. This achievement radically changes the consultation process: sentences can assume different meanings according to the affective state of the speaker. In order to address the problem of identifying emotional states embedded into courtroom events, an emotion recognition system is comprised into the JUMAS system. A set of real-world human emotions obtained from courtroom audio recordings has been gathered for training the underlying supervised learning model. In particular, this corpus encloses a set of 175 sentences uttered by different actors involved into the considered debates, i.e. judges (46 samples), witnesses (67 samples), lawyers (29 samples) and prosecutors (33 samples). The corpus contains emotional speech signals coming from 95 males and 80 females, with a duration ranging from 2 to 25 seconds. The dataset contains the following emotional states: anger, neutral, sadness and happiness. Given the emotional corpus, a features extraction step is performed in order to map the vocal signals into descriptive attributes (prosodic features, formant frequencies, energy, Mel Frequency Cepstral Coefficients, etc...). Given this representation, a supervised emotion recognition model can be trained. Into the emotion recognition component, a Multi-layer Support Vector Machines (SVMs) have been defined (Fersini et al., 2009). At the first layer a Gender Recognizer model is trained to determine the gender of the speaker, for distinguishing the "male" speakers from the "female" ones. In order to avoid overlapping with other emotional states, at the second layer gender-dependent models are trained. In particular, Male Emotion Detector and Female Emotion Detector are induced to produce a binary classification that discriminates the excited emotional states by the not excited ones (i.e. the neutral emotion). The last layer of the hierarchical classification process is aimed at recognizing different emotional state using Male Emotion Recognizer and Female Emotion Recognizer models, where only sentences uttered as excited are used to train the models for discriminating the remaining emotional states. Since SVMs are a linear learning machine able to find the optimal hyperplane separating two classes of examples (binary classification) and in our final layer we have a multi-class problem, we adopted the pairwise classification approach (Hastie and Tibshirani, 1998). In this case, one binary SVMs for each pair of classes is learned to estimate the posterior probability to assign an instance to a given class label. A given instance is finally associated to the class with the highest posterior.

3.3 Human behaviour Recognition

A further fundamental information source, for a semantic digital library into to the trial management context, is concerned with the video stream. Recognizing relevant events that characterize judicial debates have great impact as well as emotional state identification. Relevant events happening during debates trigger meaningful gestures, which emphasize and anchor the words of witnesses, highlighting that a relevant concept has been explained. For this reason, human behaviour recognition modules have been included into the JUMAS system. The human behaviour recognition modules capture relevant events that occur during the celebration of a trial in order to create semantic annotations that can be retrieved by the end users. The annotations are mainly concerned with the events related to the witness: change

of posture, change of witness, hand gestures, fighting. The Human Behaviour Recognition modules are based on motion analysis able to combine localization and tracking of significant features with supervised classification approaches. In order to analyze the motions taking place in a video, the optical flow is extracted as moving points. Then active pixels are separated from the static ones for finally extracting relevant features and for recognizing relevant judicial events (Briassouli et al., 2009), (Kovács et al., 2009). The set of annotations produced by the human behaviour recognition modules provide useful information for the information retrieval process and for the creation of a meaningful summary of the debates (see section 5.2). The human behaviour recognition modules developed was constrained by several domain limitations:

- Quality of the input video stream: the video capturing equipment used into courtroom is mostly low-cost. This resulting in low quality input video stream has a great impact into the recognition of relevant events.
- Stationary camera hampers shot detection: cameras are usually installed in fixed positions with a background that remains nearly stationary during the whole process. Indeed, it is difficult to cut the video-stream into shots for then understanding eventual relevant events.
- Long distance shots: video is usually shot from long distance. This implies that all involved people belong to the shot and, consequently, features of peoples faces are most of the time not distinguishable. Therefore, video analysis tasks, as for instance face recognition or expression analysis, become extremely difficult or even impossible.

In order to address the mentioned challenges, the acquisition chain related to the video source has been opportunely tuned. In order to allow the video analysis procedures to use a high quality stream albeit limiting the growth of the dimension of the video into the judicial folder, a double chain has been developed. The high quality video is analyzed by the human behaviour recognition components, while the low quality video is stored into the judicial folder and synchronized with the extracted semantic annotations.

3.4 Deception Detection

The discrimination between truthful and deceptive assertion is one of the most important activity performed by judges, lawyers and prosecutors. In order to support their reasoning activities, aimed at corroborating/contradicting declarations (lawyers and prosecutors) and judging the accused (judges), a deception recognition module has been developed as a knowledge extraction component. The deception detection module stands at the end of the data processing chain, as it fuses the output of the ASR, Video Analysis, and Emotion Recognition modules. This module is based on the idea initially presented by Mihalcea and Strapparava (Mihalcea and Strapparava, 2009), where a variety of machine learning algorithms have been investigated to distinguish between truth and falsehood. The deception detection module developed for the JUMAS system is based on Nave Bayes classifier (Ganter and Strube, 2009). To study the distinction between true and deceptive statements, we required a corpus with explicit labelling of the truth value associated with each statement. In order to train the model, a manual annotation of the output of the ASR module - with the help of the minutes of the transcribed sessions has been performed. The knowledge extracted for then training classification models is concerned with lies, contradictory statements, quotations and expressions of vagueness. To date, 8 sessions ($\sim$70.000 words) have been annotated, yielding 88 lies, 109 contradictory statements, and 239 expressions of vagueness. Given the training data, the Nave Bayes classifier can be induced for providing indications about future trials. In fact, the deception indications are provided only to the judges by highlighting into the text transcription, through an interactive interface, those relevant statements derived from verbal expression of witnesses, lawyers and prosecutors. In this way the identified statements may support the reasoning activities of the judicial actors involved in a trial by triggering relevant portion of the debate representing cues of vagueness and contradiction.

3.5 Information Extraction

The current amount of unstructured textual data available into the judicial domain, especially related to transcriptions of debates, highlights the necessity to automatically extract structured data from the unstructured ones for an efficient consultation processes. In order to address the problem of structuring data coming from the automatic speech transcription system, we defined an environment that combines regular expression, probabilistic models and relational information. The key element of the information extraction functionality is represented by the Automatic Template Filling component, which is based on a probabilistic model for labelling and segmenting sequential data by handling the correlation among features. In particular, a probabilistic framework based on Conditional Random Fields (Lafferty et al., 2001)

for labelling a set of trial transcriptions coming from an ASR system, has been developed for JUMAS. The core elements of the information extraction module can be distinguished in:

- Data: training data represented by annotated transcriptions and domain knowledge information available into (national) relational databases. The main information sources used for producing a structured view of unstructured texts are represented by:
  - Automatic speech transcriptions that correspond to what is uttered by the actors involved into hearings/proceeding. The ASR output related to 20 sessions has been annotated (165.000 words), yielding about 3.500 semantic annotations. The semantic items stated for annotating include: name of the lawyer, name of the defendant, name of the victim, name of the witness, name of cited subjects, cited date and date of the verdict.
  - Domain knowledge information, which correspond to databases containing information about National Lawyers, National Judges, Common Weapons, etc
- Model: Conditional Random Fields, which are discriminative graphical models trained with both transcriptions as training examples and domain knowledge as additional information. The traditional model has been extended in order to train from ASR features and from domain knowlegde databases.

The Information Extraction functionalities are provided to judges, lawyers, prosecutors and court clerks. In particular, the structured information is exploited in two different ways: (1) to provide additional information, to the Information Retrieval component, for an efficient storage and retrieval of proceedings; (2) to provide, through an interactive user interface, an immediate overview of trial contents for consequently speeding up the consultation process.

4 KNOWLEDGE MANAGEMENT

4.1 Information Retrieval

Currently the retrieval process of audio/video materials acquired during a trial needs the manual consultation of the entire multimedia tracks. The identification of a particular position on multimedia stream, with the aim at looking/listening at/to specific declarations, participations and testimonies, is possible either by remembering the time stamp in which the events were occurred or by watching the whole recording. In order to address this problem, an Information Retrieval system should address the following challenges for retrieving relevant multimedia objects:

1. Users tend to specify the queries by using only few keywords.
2. Search terms might be ambiguous or simply not the right ones.
3. The information retrieval system might not be able to automatically extract the information necessary, e.g., in case the search term relates to a high-level concept that cannot be understood at the machine level.

The conjunction of automatic transcriptions, semantic annotations and ontology representations (outlined in section 4.2), allow us to build a flexible retrieval environment based not only on simple textual queries, but on wide and complex concepts. In order to define an integrated platform for cross-modal access to audio, video recordings and their automatic transcriptions, a retrieval model able to perform semantic multimedia indexing and retrieval has been developed (Darczy et al., 2009). In particular, a linear combination of the following information has been developed:

- Similarity of representative frames of shots.
- Face detector output for topics involving people.
- High level feature considered relevant by text based similarity.
- Motion information extracted from videos.
- Text similarity based on ASR lattices.

The main goal of the information retrieval module is to provide the users with a flexible search system on judicial documents through the realization of the following type of services:

- Basic search: specification of single terms, list of terms, pairs of adjacent terms, prefixes or suffixes.
- Advanced search: specification of linguistic query weights associated with single terms, specification of linguistic quantifiers (most, all, at least n) to aggregate the terms, searching in specified XML sections of documents ordered by importance scores, and query translation based on bilingual dictionaries.
- Semantic search: the user may not only define simple text keyword queries, but also rely on wide and complex concepts based on multimedia content or ontology usage.

In this way, all the relevant information of a trial may be retrieved in terms of multimedia objects (audio,

video and text with the associated embedded semantics) by using low level textual queries and high level semantic concepts.

4.2 Ontology as Support to Information Retrieval

An ontology is a formal representation of the knowledge, which characterizes a given domain, through a set of concepts and a set of relationships that hold among them. Into the judicial domain, ontologies represent a key element that support the retrieval process performed by the end users. Textual-based retrieval functionalities are not sufficient for finding and consulting transcriptions (and other documents) related to a given trial. A first contribution of the ontology component developed in the JUMAS system is concerned with its query expansion functionality. Query expansion aims at extending the original query specified by the end users with additional related terms for then automatically submitting the whole set of keywords to the retrieval engine. The main objective is to narrow the focus (AND query) or to increase recall (OR query). When expanding the query, new terms are enhanced with a confidence weight used by the scoring function of the retrieval component. Let's introduce an example in order to understand the aim of the query expansion functionality in the whole searching process. Suppose that a judge needs to retrieve the transcription about a weapon crime: a knife. By specifying knife in the search form the system finds only few documents containing this word. Then the retrieval module invokes the query expansion web service in order to obtain the related terms. The web service finds the knife instance in the ontology and explores the ontological relationship in order to find related terms. The query expansion module is a web service based on Jena (Carroll et al., 2003) that provides a programmatic environment for OWL ontologies and includes a rule-based inference engine. A further functionality offered to the end user is related to the possibility of knowledge acquisition. In fact, the ontology component offers to the judicial users the possibility of acquiring specific domain knowledge, i.e. they have the opportunity of specifying semantic relationships among concepts available into the trial transcriptions. This knowledge management component provides not only the possibility of more advanced and accurate search, but also the opportunity to contribute to the construction of sophisticated domain ontology without any background knowledge about the ICT and ontological modelling aspects.

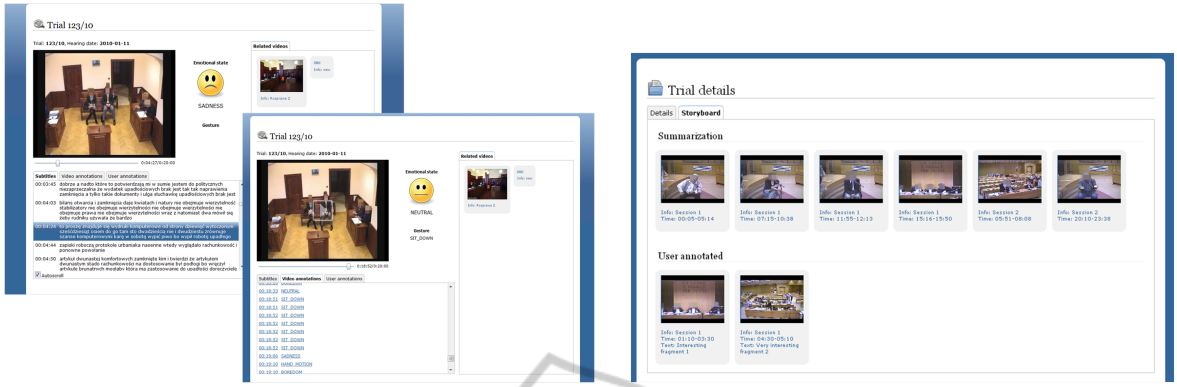
4.3 User Generated Semantic Annotations

Judicial users usually tag manually some papers for highlighting (and then remembering) significant portion of the debate. An important functionality offered by the JUMAS system relates to the possibility of digitally annotating relevant arguments discussed during a debate. In this context, the user-generated annotations may help judicial users for future retrieval and reasoning processes. The user-generated annotations module enclosed into the JUMAS system allows the end-users to assign free tags to multimedia content in order to organize the trials according to their personal preferences. It also enables judges, prosecutors, lawyers and court clerks to work collaboratively on a trial, e.g. a prosecutor who is taking over a trial can build on the notes of his predecessor. The tags are analysed to suggest related tags to the user for search and to automatically find related documents that contain related terms. The user-generated semantic annotation module exposes all annotations to the common JUMAS JEX annotation infrastructure, which can be searched in by the retrieval models. To allow users of JUMAS to browse through the various available documents in a focused manner, upon viewing a specific document, a dedicated interface shows several tags, which can be used to browse the documents. The tags recommended by the module for a specific document are found by several techniques, which have been combined to a meta-recommender as proposed in (Jäschke et al., 2009). In particular, Collaborative Filtering, Tag co-occurrence-based and occurrence-based recommendations have been enclosed into the meta-recommender module. As outcome, this module offers the possibility of personalizing contents according to the user preferences or working routines, providing then a better usability of multimedia contents.

5 KNOWLEDGE VISUALIZATION

5.1 Hyper Proceeding Views

As introduced in Section 2.2, the user interface of JUMAS is a web portal, in which the contents of the Database are presented in different views, to support the operations of clerks, judges and all the people involved in the trial. The basic view is the browsing of the trial archive, like in a typical court management system, to present general information (dates of hearings, name of people involved) and documents attached to each trial in the archive. JUMAS has also



(a) Hyper proceedings view (b) Multimedia Summarization Visualization

Figure 3: Knowledge visualization.

distinguishing features among which the automatic creation of a summary of the trial, the presentation of user generated annotations and, primarily, the Hyper Proceeding View (see Figure 3(a)), i.e. an advanced presentation of media contents and annotations that allows to perform queries on contents, jumping directly to relevant parts of media files. The Annotations are shown in the portal while media files are played and it is also possible to browse media by clicking on annotations. The contents can be browsed over several dimensions: audio, video, text annotations; the user can switch among them. This approach, typical of web applications, provides a new experience for the user that, in existing systems, can only passively watch the contents or perform time-consuming manual search.

5.2 Multimedia Summarization

Digital videos represent a fundamental informative source of those events that occur during a trial: they can be stored, organized and retrieved in short time and with low cost. However, considering the dimension that a video source can assume during a trial recording, several requirements have been pointed out by judicial actors: fast navigation of the stream, efficient access to data inside and effective representation of relevant contents. One of the possible solutions to these requirements is represented by multimedia summarization aimed at deriving a synthetic representation of audio/video contents, characterized by a limited loss of meaningful information. In order to address the problem of defining a short and meaningful representation of a debate, a multimedia summarization environment based on an unsupervised learning approach has been developed (Fersini et al., 2010). The main goal is to create a storyboard of either a hearing or an entire proceeding, by taking into account the semantic information embedded into a courtroom recording. The storyboard construction creates a unified representation of contents. In particular, the summarization module exploits two matrices: one matrix associated with speech transcription and one matrix associated to the audio/video annotations. The first matrix represents textual transcription scoring, obtained through the TFIDF weighting technique (Salton and Buckley, 1998). The second matrix, defined as binary, represents the presence or absence of a specific audio/video annotation associated to a given transcription segment. Starting from these two matrices, the multimedia summarization module may start the summary generation. The core component is based on a clustering algorithm named Induced Bisecting K-means (Archetti et al., 2006). The algorithm creates a hierarchical organization of (audio, video and textual as well automatic annotations) clips, by grouping in several clusters hearings or sub-parts of them according to a given similarity metric. At the moment the relative length of the summary is set by the system administrator. In the next release each user will be enabled to control the length of his/her summary. The resulting storyboard is presented to judge, prosecutor, lawyer and court clerk as shown in Figure 3(b).

6 CONCLUSIONS

JUMAS project is demonstrating that it is possible to enrich the court management system with an advanced set of tools for extracting and using the knowledge embedded into the multimedia judicial folder. Automatic template filling, semantic enrichment of the judicial folder through audio and video processing, enhanced transcription process, help judges, prosecutors not only to save time, but in a special

way to enhance the quality of their judicial decision and actions. These improvements are mainly due to the possibility to search not only text, but also events that occurred in the courtroom. The first outcome in JUMAS indicates that transcription and audio analysis with an acceptable word error rate and video analysis can provide additional information in an affordable way. The demonstration and validation in progress are also providing indications about the requirements for the next generation of ICT-empowered courtrooms. Recording systems and ICT infrastructure in the courtroom that are actually under deployment or to be deployed in the near future can be designed in order to support audio and video processing capabilities, while information retrieval relies on state-of-the art ICT infrastructure. JUMAS is providing not only tools for a better usability of the trial folder, being the audio-related tools the closest to a first deployment, but also inputs for a more future oriented specification of e-justice systems in different countries.

ACKNOWLEDGEMENTS

This work has been supported by the European Community FP-7 under the JUMAS Project (ref.: 214306).

REFERENCES

Archetti, F., Campanelli, P., Fersini, E., and Messina, E. (2006). A hierarchical document clustering environment based on the induced bisecting k-means. In *Proc. of the 7th Int. Conf. on Flexible Query Answering Systems*, pages 257–269.

Briassouli, A., Tsiminaki, V., and Kompatsiaris, I. (2009). Human motion analysis via statistical motion processing and sequential change detection. *EURASIP Journal on Image and Video Processing*.

Carroll, J., Dickinson, I., Dollin, C., Reynolds, C., Seaborne, A., and Wilkinson, K. (2003). Implementing the semantic web recommendations. Technical Report HPL-2003-146, Hewlett Packard.

Council of the European Union (2009). Multi-annual european e-justice action plan 2009?2013. In the Official Journal of the European Union (OJEU) 31.3.2009 C 75/1.

Darczy, B., Nemeskey, D., Petrs, I., Benczr, A. A., and Kiss, T. (2009). Sztaki @ trecvid 2009.

Falavigna, D., Giuliani, D., Gretter, R., Loof, J., Gollan, C., Schlueter, R., and Ney, H. (2009). Automatic transcription of courtroom recordings in the jumas project. In *ICT Solutions for Justice*.

Fersini, E., Messina, E., and Archetti, F. (2010). Multimedia summarization in law courts: A clustering-based environment for browsing and consulting judicial folders. In *Proc. of 10th Industrial Conference on Data Mining*.

Fersini, E., Messina, E., Arosio, G., and Archetti, F. (2009). Audio-based emotion recognition in judicial domain: A multilayer support vector machines approach. In *Proc. of the 6th International Conference on Machine Learning and Data Mining in Pattern Recognition*, pages 594–602.

Ganter, V. and Strube, M. (2009). Finding hedges by chasing weasels: hedge detection using wikipedia tags and shallow linguistic features. In *Proc. of the ACL-IJCNLP 2009 Conference*, pages 173–176.

Hastie, T. and Tibshirani, R. (1998). Classification by pairwise coupling. In *Proc. of the 1997 conference on Advances in neural information processing systems*, pages 507–513.

Jäschke, R., Eisterlehner, F., Hotho, A., and Stumme, G. (2009). Testing and evaluating tag recommenders in a live system. In *Proc. of the 3rd ACM conference on Recommender systems*, pages 369–372.

Kovács, L., Utasi, A., and Szirányi, T. (2009). Visret - a content based annotation, retrieval and visualization toolchain. In *Advanced Concepts for Intelligent Vision Systems*.

Lafferty, J. D., McCallum, A., and Pereira, F. C. N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. of the 18th International Conference on Machine Learning*, pages 282–289.

Mihalcea, R. and Strapparava, C. (2009). The lie detector: explorations in the automatic recognition of deceptive language. In *Proc.of the ACL-IJCNLP 2009 Conference*, pages 309–312.

Rabiner, L. and Juang, B. H. (1993). *Fundamentals of Speech Recognition*. Prentice-Hall Inc.

Salton, G. and Buckley, C. (1998). Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 24(5):513–523.

The Economist (2010). Data, data everywhere. Feb 25th 2010. The Economist.

Velicogna, M (2008). Use of information and communication technologies (ict) in european judicial systems. In European Commiission for the Efficiency of Justice Reports, Council of Europe.