Legal Knowledge Acquisition and Multimedia Applications

Ciro Gracia[1], Pompeu Casanovas[2], Jordi Carrabina[3], Xavier Binefa[1], Emma Teodoro[2], Màrius Monton[3], Núria Casellas[2], Carlos Montero[3], Núria Galera[2], Javier Serrano[3], and Marta Poblet[2,4]

[1] Digital Video Understanding Group, ETSE, UAB, Spain
{Ciro.Gracia, Xavier.Binefa}@uab.cat
[2] Institute of Law and Tecnology, UAB, Spain
{Pompeu.Casanovas, Emma.Teodoro, Nuria.Casellas, Nuria.Galera}@uab.cat
[3] Laboratory for HW/SW Prototypes and Solutions (CEPHIS), ETSE, UAB, Spain
{Jordi.Carrabina, Marius.Monton, Carlos.Montero, Javier.Serrano}@uab.cat
[4] ICREA Researcher at the Institute of Law and Tecnology, UAB, Spain
{marta.poblet}@uab.cat

Abstract. Search, retrieval, and management of multimedia contents are challenging tasks for users and researchers alike. The aim of E-Sentencias Project is to develop a software-hardware system for the global management of the multimedia contents produced by the Spanish Civil Courts. We apply technologies such as the Semantic Web, ontologies, NLP techniques, audio-video segmentation and IR. The ultimate goal is to obtain an automatic classification of images and segments of the audiovisual records that, coupled with textual semantics, allows an efficient navigation and retrieval of judicial documents and additional legal sources

1 Introduction

The search, retrieval, and management of multimedia contents are challenging tasks for users and researchers alike. The development of efficient systems to navigate through content has recently become an important research topic. Since parliaments, courts, ministries, or security and military forces are producing enormous masses of video, audio and text files, the requirement of specific content management solutions for the legal domain has arisen naturally.

The aim of E-Sentencias is to develop a software-hardware system for the global management of the multimedia contents produced by the Spanish Civil Courts. The Civil Procedure Act of January 7th, 2000 (1/2000) introduces the compulsory video recording of oral hearings. As a result, Spanish Civil Courts are currently producing a massive number of audiovisual records which have become not only part of the judicial file, together with suits, indictments, injunctions, judgments and pieces of evidence, but also the official record of the oral hearing itself. This audiovisual material is used by lawyers, prosecutors and judges to prepare, if necessary, appeals to superior courts. Nevertheless, at present, there

are no available applications to exploit all these audiovisual materials and offer management solutions to law firms and judicial units. E-Sentencias proposes an application to manage textual case materials (legislation, jurisprudence, procedural documents, etc.), together with these judicial audiovisual contents in a dynamic way. E-Sentencias is therefore focused towards automatic annotation, segmentation and classification of these audiovisual materials, through the use of different technologies for audio and video analysis and knowledge management (i.e. ontologies). The ultimate goal is to obtain an automatic classification of the audiovisual records that coupled with textual semantics would allow, in the future, the efficient navigation, semantic search and retrieval of judicial multimedia materials.

Section 2 below describes the current situation concerning the audiovisual recording of civil cases in Spain and a short description of the basic typology of the oral hearings of the main civil procedures, which establishes the basis for the knowledge acquisition process. In Section 3 we offer an overview of this knowledge acquisition process, towards the construction of a conceptual structure for the classification of video segments of the oral hearings, keyword spotting and the development of ontologies. Section 4 depicts the structure and architecture of the video system prototype and the speech recognition aplicability at the present stage of research. We conclude with some discussions and future work in Section 5.

2 Video Recording of Civil Procedures in Spain

The 1/2000 Civil Procedure Act sets different civil proceedings. However, the ordinary and the verbal proceedings constitute the most common typologies of civil proceedings in Spain. The main differences between the two lie in the value of the case -more or less than 3000, respectively- and the legal object at dispute (see figures 1 and 2). Also, the steps followed by these two procesess are different. On the one hand, the ordinary proceeding starts with a separate, independent oral hearing called preliminary hearing [*audiencia previa*] to resolve pre-judicial issues (documents, evidences to be accepted, etc.), while the verbal proceeding takes place alltogether in the same oral event. On the other hand, in the ordinary proceeding the claim of the plaintiff is contested in written terms, while in the verbal proceeding is replied orally in the same act.

As stated in Section 1, the 1/2000 Civil Procedure Act introduced the video recording of these oral hearings. However, he provisions of the Act did not include an homogeneous protocol establishing how (format, annotations, etc.) to deliver these audiovisual records. Rather, and since an ever growing number of Autonomous Governments in Spain hold competences on the organization of the judicial system there is a plurality of standards, formats, and methods to produce audiovisual records. As a result, analogical and digital standards coexist with different recording formats. The support in which copies are provided to legal professionals (i.e. to prepare an appeal) may also consist of either VHS videotapes or CDs (although the use of digital support is increasing). And, fi-

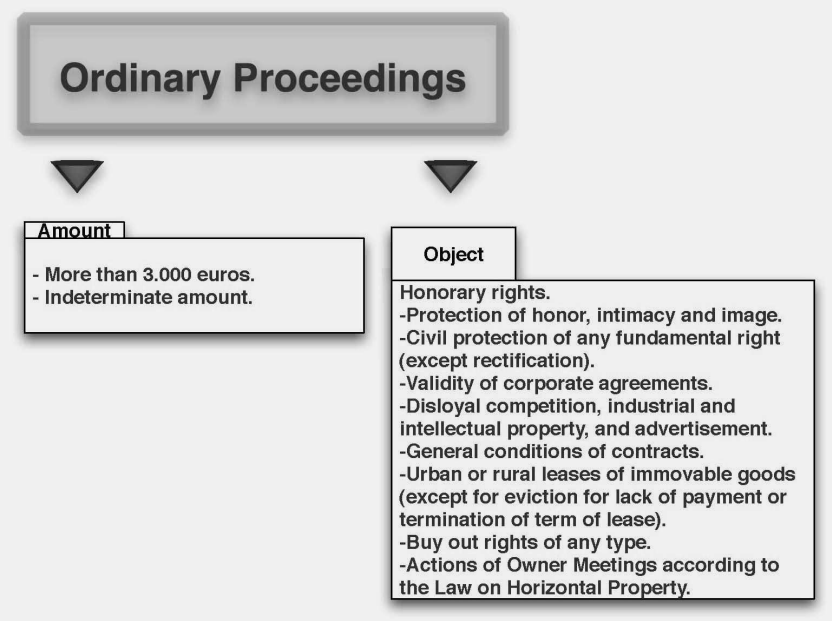

Fig. 1. The ordinary proceeding: amount and content

nally, the procedures to store, classify, and retrieve audiovisual records may vary even from court to court.

Basically, these audiovisual materials are used by legal professionals to prepare appeals to superior courts. Nevertheless, this material is just a video file and to locate specific moments or speech acts in the video is a time-consuming task. These video files do not take advantage of the specific structures of each of the types of oral hearing established by the Procedural Act or the judicial practice (practical rules and legal professional knowledge). The addition of this legal epistemological information on the video file may offer to legal professionals the possibility to improve their management capabilities and support automatic classification of video segments.

Due to the nature of the multimedia material itself and of its means of production, the knowledge acquisition process could not benefit from the start from the use of automatic information extraction tools and had to begin with the identification of the types of proceedings, their oral hearings and their structural parts and differences in order to establish an accurate workflow of the oral hearings that are contained in the video recordings. The acquisition of structural content has been performed from the analysis of legislation (Civil Procedural Act), together with the analysis of several oral hearings in order to detect peculiaritites of the hearings in the practice of the courts. Currently, we have identified the steps of the oral hearing for verbal and ordinary proceedings [figures 4 and 5].

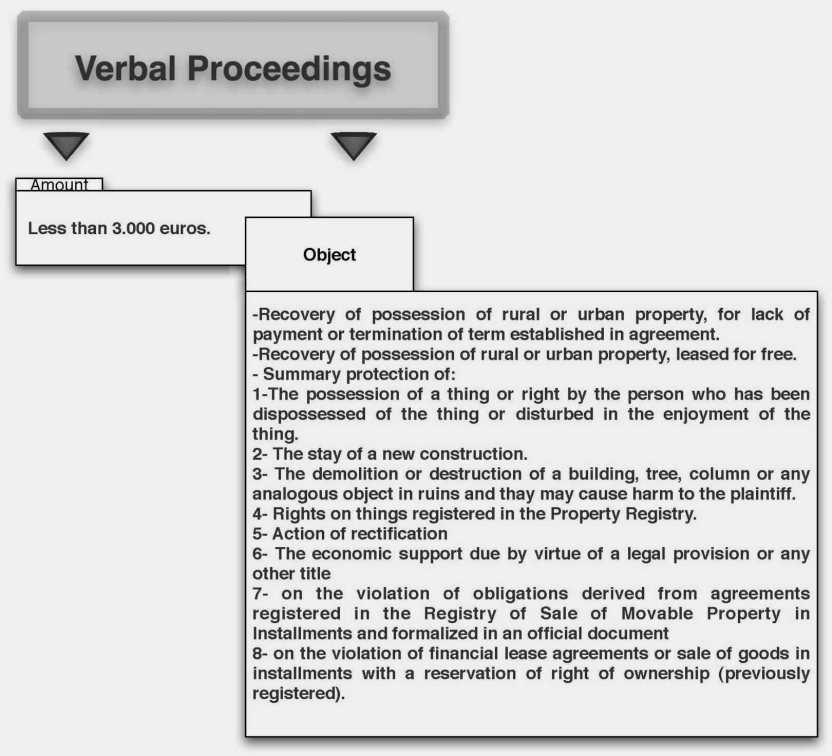

Fig. 2. The verbal proceeding: amount and content

3 Conceptual Structure and Multimedia Applications

Once this structural knowledge has been extracted, we now work to further extend the knowledge about the oral hearing to be able to detect certain important features to enable automatic classification of video segments and offer future search functionalities: how can the type of proceeding (verbal or ordinary) contained in the recorded oral hearing can be automatically detected? How can it be determined automatically when a structural step of the hearing starts or finishes?

Towards the acquisition of more knowledge regarding the oral hearings, we have started from scratch by transcribing a small set of oral hearings. These textual transcriptions offer, first, an insight to the structure of the hearing, to the legal concepts (themes) being discussed (i.e. judgment, injunction, cause of necessity, deed, etc.) and to the legal expressions used (i.e. *con la venia*: "with the permission of your Honor"). All these different keywords or phrases can be used to detect certain parts of the oral hearing or even certain actors of this hearing (judge, plaintiff, etc.).

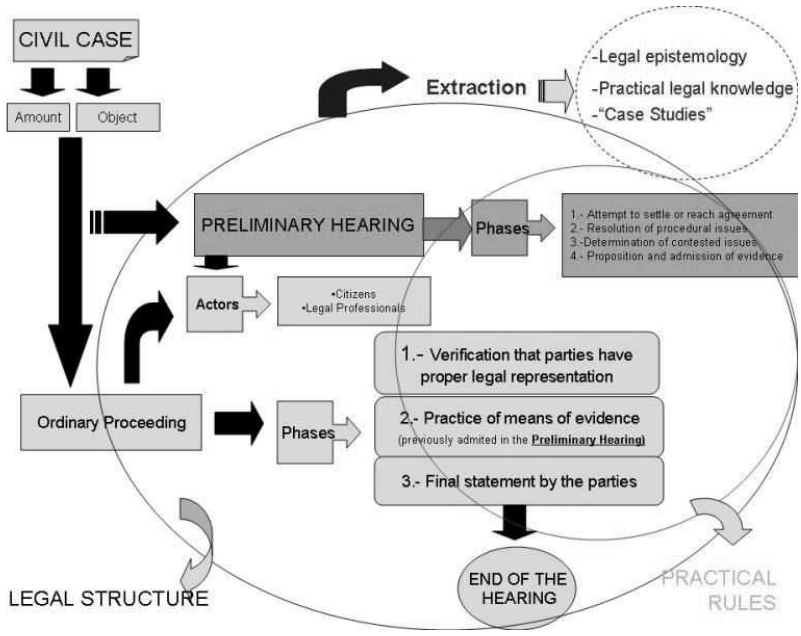

Fig. 3. Knowledge acquisition possibilities from the Court video files

In addition, they facilitate the coding of practical rules regarding judicial procedures that are implicit in the video sequences, and that can be used also to detect struture and content within the multimedia files. See, for example, the following piece of transcription:

<actor name="judge" tc="00.01.30">
Let us see mr. *** DEFENDANT STANDS UP AND APPROACHES TO THE MICROPHONE come to the microphone [PROCEDURAL FORMULA, EXCLUSIVE USE BY THE JUDGE]
</actor>
<actor name="defendant" tc="00.01.31">
yes
</actor>
<actor name="judge" tc="00.01.38">
and answer the questions that both attorneys are going to formulate, starting by the attorney of the plaintiff [GENERAL RULE: IF BOTH PARTIES HAVE REQUESTED EXAMINATION, THE PLAINTIFF'S ATTORNEY ALWAYS COMES FIRST IN EXAMINATING THE DEFENDANT, AND THEN CONTINUES THE DEFENDANT'S ATTORNEY].
</actor>

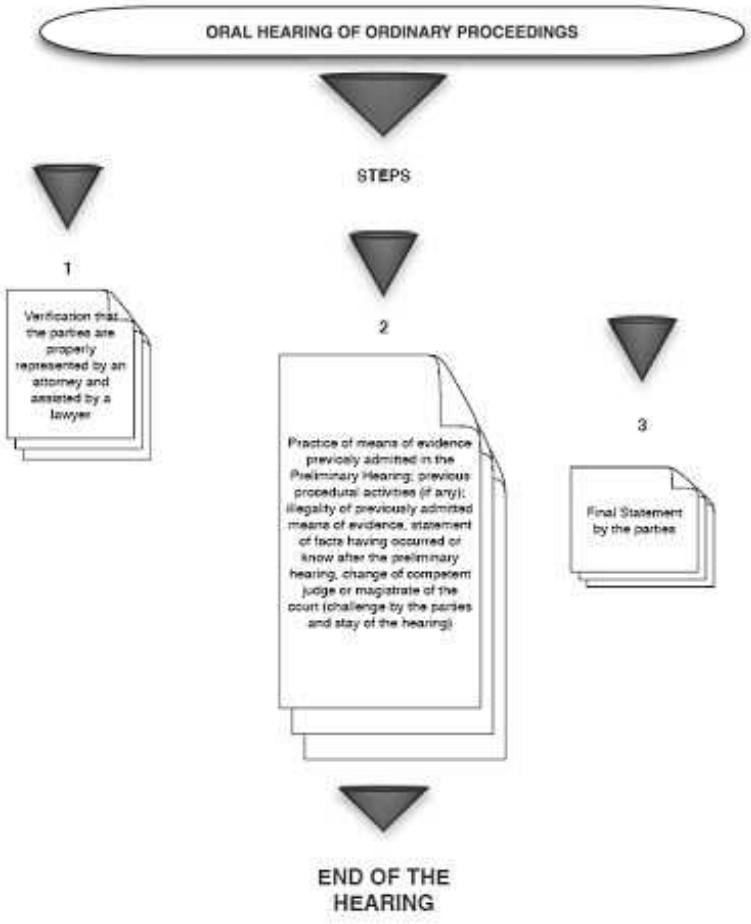

Fig. 4. Steps of the process in ordinary proceedings

This is only a first level of textual and visual annotation of judicial hearings, but it is also the basis to create specific annotation templates at different levels (concepts, legal formulae, practical rules of interaction, etc.) that facilitate the construction of different types of ontologies. On one hand, legal domain ontologies based on the structure and content of the oral hearing and, on the other, multimedia ontologies that describe the input at hand. Figure 6 below shows the annotation tool (used for the transcription process) we built up to capture and point out the pragmatic, cognitive (emotional) and linguistic features of legal discourses in Court interactions.

In practice the use of ontologies for different tasks and purposes requires to consider the particular task as context for the ontology. The reason is that ontologies are often not really designed independent of the task at hand [1].

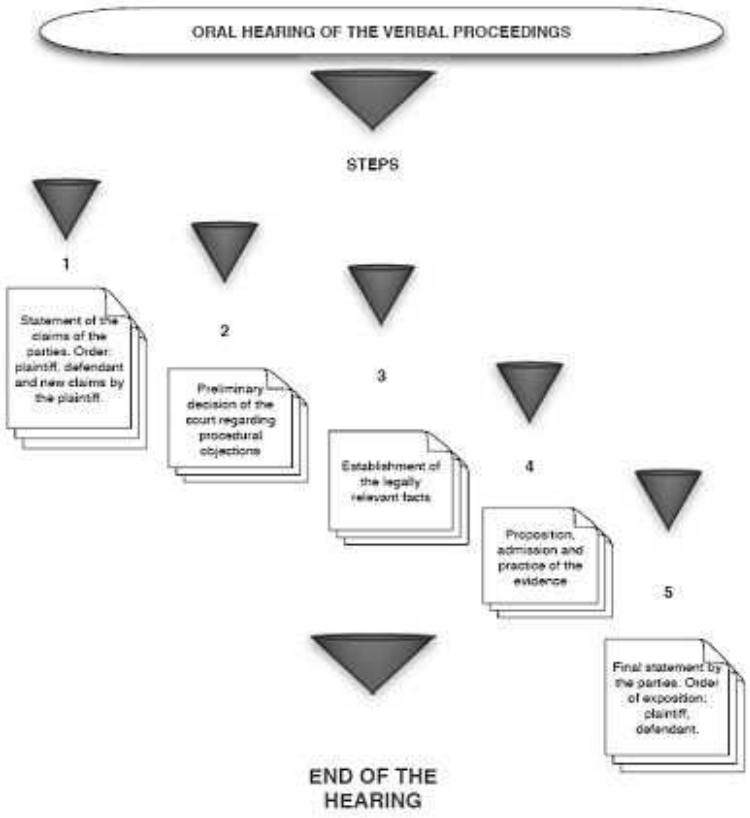

Fig. 5. Steps of the process in verbal proceedings

From a legal multimedia user-centered perspective there are two problems that have to be addressed (i) the definition of context in merging and aligning legal domain and multi-media ontologies; (ii) the specific exophoric nature of the legal videorecording.[5]

Researchers on contextual ontologies use to define "context" as local (not shared with other ontologies) and opposed to content ontologies themselves (shared models of a domain) [1] [2]. For example, to cope with the directionality of the information flow, local domains and the context mapping, which cannot be represented with the current syntax and semantics of OWL, C-OWL (*Context* OWL) is being developed [3].

[5] *Exophoricity* means that references, co-references and context of the legal discourses being developed in court (original facts, quoted precedents ....) are to be found outside of the present legal setting. E.g the narratives of the story told by witnesses took place in a different time and space; past precedents quoted by lawyers refer to past decisions on different cases etc.

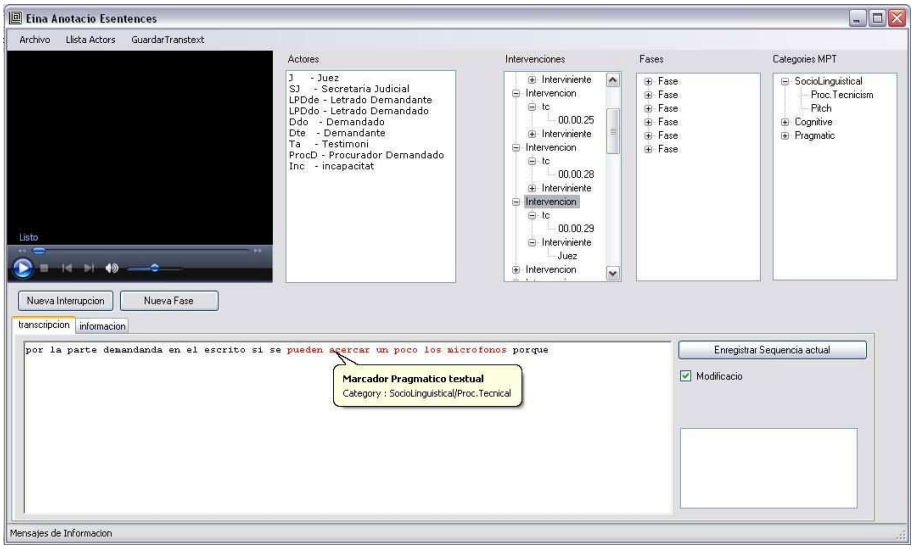

Fig. 6. Annotation tool

From the multimedia researchers' point of view, context is defined currently as "the set of interrelated conditions in which visual entities (e.g. objects, scenes) exist" [4] [5]. This grounds the strategy of the direct vs. indirect exploitation of the knowledge base to annotate the content of the videos, using *visual* and *content* descriptors alike [6]. But, most important, this definition of context entails a theoretical approach in which "actions and events in time and space convey stories, so, a video program (raw video data) must be viewed as a document, not a non-structured sequence of frames" [7] [8]. In such an approach, visual low level features, object recognition and audio speaker diarization (process of partitioning the audio stream in homogenous segments and clustered according to the speaker identity) are crucial to analyze e.g. a sport or movies' sequences.

However, the audiovisual documents that are recorded in Spanish courtrooms do not convey actions, but *legal narratives*. Motion and color are generally uniform, since they are not considered the relevant aspect of those documents. Thus, Court recordings are technically very poor, filmed using a one-shot perspective (the camera is situated above and behind the judge, who never appears on the screen). Rather than *telling a story*, the video structures a single framework in which a story is referred, conveyed and constructed by the procedural actors (judge, counsels, testimonies, secretary, and court clerks).

Here lies the *layered exophoricity* of the legal discourse. Actions, events and stories are referred into a contextually embedded discourse, procedurally-driven, and hierarchically conducted by the judge (judge-centered). Therefore, a strong *décalage* is produced between audio and video as sources of information. The information provided by a judicial video record depends greatly on the audio,

because we may only infer procedural (but not substantial) items from motion. What is important is what is *said* in Court, not really what is *done*. Visual images are only ancillary related to the audio stream. This is an important feature of the records, which has to be taken into account in the tasks of extracting, merging and aligning ontologies, because what the different users require (judges, lawyers, citizens) is the combination of different functionalities focused on the legal information content (legislation quoted, previous cases and judgments -precedent-, personal professional records, and so on). This is the reason for a *hybrid user-centered approach* that is the kernel of our theoretical approach.

Currently, the team of legal experts is extracting keywords (legal themes) that could provide de basis to classify the different proceedings at hand and is extracting legal concepts and expressions for word spotting that are important towards the automatic detection of structure, and is improving the knowledge on the structure with more information regarding the actors and acts that participate and occur during the hearing. Also, the construction requirements of a domain ontology of the oral hearing is being outlined. Moreover, multimedia ontologies are being studied towards its use together with the legal domain ontology (i.e. ontology alignment and merging techniques) to facilitate automated annotation and classification. Finally, research towards the acceleration of the semantic search (through ontologies) is also performed to offer an integrated software-hardware accelaration platform.[6]

[6] The legal field constitutes a privileged domain for the application of Semantic Web technologies and several legal ontologies are being used to construct tools and prototypes to support the management, organization, search and retrieval of documents stored in legal databases [11]. Therefore, the developments regarding ontology-based semantic search offer encouraging qualitative results in the legal domain, as they reduce significantly the amount of information retrieved (compared to indexing) and, at the same time, they improve the quality of the document retrieval process (it filters the non-semantically related documents) [12]. However, quantitatively, semantic searches are not yet sufficiently efficient. The process to cover an ontology based on 10.000 nodes (with different connectivity degrees) might take from 9 hours up to 4 days. The improvement of the computational time would result in a more efficient and cheaper application of semantic technologies. Our purpose is the development of a system capable of prototyping the implementation for a specific problem, semantic search, with reconfigurable devices. A system like this is configured in compilation time from description written in a high level language, partitioned into processes to be executed different resources ranging from processors to application specific hardware resources. This process of describing both hardware and software is based on a set of design methodologies known as "hardware-software co-design". Our acceleration platform will be based on a PCI expansion card for a standard PC computer. This PCI card will contain a FPGA as main computational device and the amount of memory required for complex problems. This expansion board will accelerate the process of finding next suitable vertex and maintaining the edges set updated.

4 Structure and Architecture of the Video Prototype & Speech Recognition Techniques

At the current stage of research, we outline the initial structure and architecture of the video prototype, with special regard to the user interface, and the utilities provided by video analysis and speech recognition. The prototype will be adapted and improved according to the knowledge acquired from the transcriptions and the analysis of video and text being currently performed.

4.1 Structure and Architecture of the Video Prototype

The development of an intuitive user interface constitutes a central requirement of the system. While preserving the simplicity of use, the application allows: a) access to the legally significant contents of the video file; b) integration of all procedural documents related to the oral hearing; c) management of sequential observations, and d) semantic queries on the contextual procedural aspects.

The structure of the application is based on two intuitive and semantically powerful metaphors: the *oral hearing line* and the *oral hearing axis*. The *oral hearing line* presents a timeline divided into segments. Each segment represents a different speech, produced by one of the participants in the process: judge, secretary, attorneys, witnesses, etc. Participants are represented by a different color to obtain an identification at first glance of their interventions. Therefore, it is possible to visualize specific contents of the video by merely clicking on a particular colored sequence. Moreover, it is possible to add textual information at any moment of the intervention.

The *oral hearing axis* consists of a column representing the different phases of the event as defined by procedural legislation. Different phases (as opening statements, presentation of evidences, concluding statements, etc) are represented by different colors, allowing a quick access. It is also possible to access to legal documents related to each phase (i. e. pieces of evidence such as contracts, invoices, etc.) as well as to jurisprudence quoted in the oral hearing and detected through phonetic analysis. This legal information is also structured in directories and folders.

As Figure 7 shows, the user interface is divided into two main parts: the upper part contains the video player, the oral hearing axis and the oral hearing line. The lower part is devoted to external information layers (i.e. references to articles, documents annexed, manual annotations, links to jurisprudence, etc.). This part is divided into two tabs. The first one contains important information of the selected phase, allowing the addition of the different documents presented during the phase. The second tab contains historical information of the process and all the related information available in advance.

The main functionalities offered in the upper part of the user interface are:

1. The information tab: this is a scrollable tab containing the most relevant data of the process.

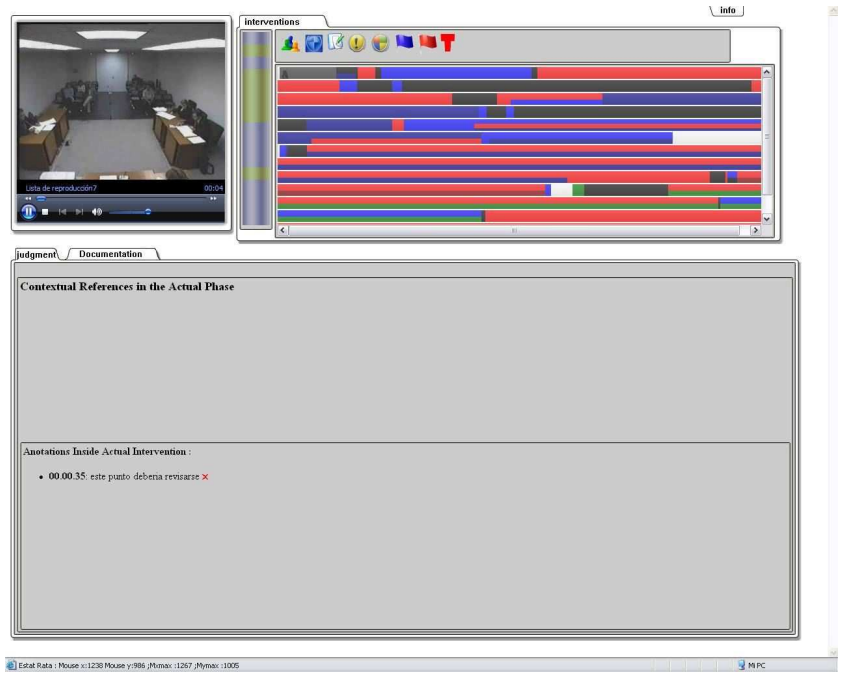

Fig. 7. User interface

2. The oral hearing line: the timeline of sequences and interventions assigned to the different actors of the process. One single sequence of the video may contain interventions of different actors. Therefore, sequences may be either mono-colored (intervention of one single part) or multi-colored (more than one part intervening in the same sequence). The horizontal length of each segment of the timeline is proportional to its length in seconds. The application includes two modes of playing video, apart of the usual one. It is possible to select either the visualization of all the interventions by a single participant or, in turn, all the interventions on a given phase.
3. The list of intervening parties: Each actor intervening in the process is represented by an icon. As in the case of the oral hearing line, we may choose to visualize only those sequences appearing one specific participant (i.e. the judge or de defense attorney).
4. The oral hearing axis: this is the vertical line representing the procedural phases of the process. The judicial process is therefore divided in procedural phases which can, as well, be subdivided in interventions. The vertical axis has the advantage of providing quick access to interventions belonging to a given phase.

In addition to these functionalities, it is possible make a manual annotation of the sequence. Double-clicking with the right bottom of the mouse over a

sequence running on the video screen opens a pop-up with a manual annotation tool.

As regards the lower part of the user interface, this area contains all the relevant information and documents of the process, but also enables the user to add and organize the information appearing during the different phases. This part is divided into two different sections:

1. An area enabling the visualization of all the references related to each phase of the process. References consist of data (i.e. Civil Code articles, judgments, Internet links, etc.) automatically introduced through semantic annotation.
2. An area including all manual annotations of the sequences made by the user.

The architecture of the system is based on a web system including the following components:

1. Video server WMS: a server based on Windows 2003 Enterprise server with a streaming Windows Media services which allows video broadcast of audiovisual content of the judicial processes under demand. Application server TOMCAT: the application serves web contents and provides the required interaction with the database by means of Java Server Pages.
2. Mysql Database: the Mysql database contains the information related to all processes and their respective annotations
3. Client browser IE 7.0: It allows the management of the user interface and the management of the user interaction with the embedded Windows Media Player 11 that streams the video.

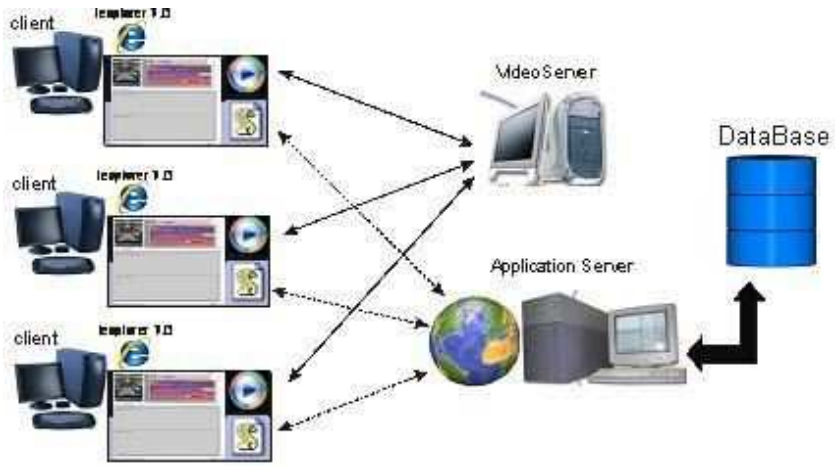

Fig. 8. Web Distribution System

4.2 Automatic segmentation from speech recognition technologies

The system described until now can reach its maximum efficiency if we are able to automatically segment and annotate the video content. The amount of videos to manage is so large that it would be impossible to process them manually. We have already emphasized the crucial importance of the audio string in Court hearings. We propose to use speech recognition to manage this problem, as an already mature technology in other domains (i.e. information services). These technologies are being used in two different manners: (1) keyword spotting for the oral hearing axis and (2) speaker recognition for the oral hearing line [9].

Required metadata to draw the oral hearing axis are temporal marks pointing at the beginning of each procedural phase in Court. Every process contains several phases, and every phase contains its own actors. During these phases, all the actors start and finish their interventions with the same type of ritualized utterances, i.e. in Spanish, "con la venia", "no hay más preguntas" etc. So, by spotting these utterances, we are able to set the switching and starting point of each procedural phase.

This approach is simple, but relies on the effectiveness of keyword recognition. In order to improve this effectiveness, we train the acoustics models of the speech recognition system, nowadays based in a commercial motor (Loquendo ASR 7.4), with data from the courtroom videos. We expect to recognize close to 100% of keywords when our system will be trained with the whole data set, even if the speech recognition system cannot be completely trusted to obtain a perfect speech to text transcription (because of audio quality, environment variability, noise) [10].

5 Discussion and Future Work

As mentioned in Section 3, the team of legal experts is currently extracting keywords (legal subjects/themes) that could differenciate the oral hearings at hand and extracting legal concepts and expressions for word spotting for automatic detection of structure and is improving the knowledge on the structure (actors and acts). They will have to be evaluated and validated. These keywords, legal concepts and expressions are also to be used as the input data to model an ontology for the oral hearing process and the process of ontology construction will be started when sufficient knowledge has been extracted and the ontological requirements are established. Also, decisions towards the integration or merging of multimedia and contextual ontologies will be taken. Finally, research towards the acceleration of the semantic search (through ontologies) is also performed to offer a software-hardware accelaration platform.

In the E-Sentencias project we expect to obtain several results. First, a fully annotated legal corpus of multimedia oral hearings classified in their corresponding procedural types. Second a set of domain and multimedia ontologies that represent the knowledge acquired from the oral hearings and their contextual and multimedia features. And finally, a system for automatic annotation of the

judicial recordings, which would include search and retrieval capabilities. The automatic capabilities of the system to automatically detect interventions from different actors and the various phases within the hearing will be then tested against the manually annotated corpus. The search and retrieval performance of the system will also be evaluated.

Acknowledgements

E-Sentencias (E-Sentencias. *Plataforma hardware-software de aceleración del proceso de generación y gestión de conocimiento e imágenes para la justicia*) is a Project funded by the Ministerio de Industria, Turismo y Comercio (FIT-350101-2006-26). A consortium of: Intelligent Software Components (iSOCO), Wolters Kluwer Spain, UAB Institute of Law and Technology (IDT-UAB), Centro de Prototipos y Soluciones Hardware - Software (CHEPIS - UAB) y Digital Video Semantics (Dpt. Computer Science UAB).

References

1. Haase, P., Hitzler, P., Rudolph, S., Qi, G., Grobelnik, M., Mozetic, I., Bojadziev, D., Euzenat, J., d'Aquin, M., Gangemi, A., Catenacci, C.: D3.1.1 Context Languages - State of the Art. NeOn Project EU-IST Integrated project (2006)
2. Bouquet, P., Giunchiglia, F., van Harmelen, F., Serafn, L., Stuckenschmidt, H.: Contextualizing ontologies. Journal of Web Semantics **26** (1) (2004) 325-343
3. Bouquet, P., Giunchiglia, F., van Harmelen, F., Serafini, L., Stuckendschmidt, H.: C-OWL: Contextualizing Ontologies. In Fensel, D. et al. (eds.), ISWC 2003 Second International Semantic Web Conference, Lecture Notes in Computer Science **2870** (2003) 164-179
4. Jaimes, A., Smith, J.R.: Semi-automatic, Data-driven Construction of Multimedia Ontologies. ICME 2003: Proceedings of the 2003 International Conference on Multimedia and Expo, IEEE (II) (2003) 781-784
5. Jaimes, A., Tseng, B., Smith, J.R.: Modal Keywords, Ontologies, and Reasoning for Video Understanding. In Bakker, E.M. et al. (eds.), Image and Video Retrieval. Second International Conference, CIVR 2003 Urbana-Champaign, IL, USA, July 24-25, 2003 Proceedings, Lecture Notes in Computer Science **2728** (2003) 248-259
6. Bloedhorn, S., Petridis, K., Saathoff, C., Simou, N., Tzouvaras, V., Avrithis, Y., Handschuh, S., Kompatsiaris, Y., Staab, Y., Strintzis, M.G.: Semantic Annotation of Images and Videos for Multimedia Analysis. In Gmez-Prez, A. and Euzenat, J. (eds.), The Semantic Web: Research and Applications: Proceedings of the Second European Semantic Web Conference, ESWC 2005, Heraklion, Crete, Greece, May 29-June 1 (2005), Lecture Notes in Computer Science **3532** (2005) 592-607
7. Song, D., Liu, H.T., Cho, M., Kim, H., Kim, P.: Domain Knowledge Ontology Building for Semantic Video Event Description. Leow, W.K. et al. (eds.), Image and Video Retrieval. 4th International Conference, CIVR 2005, Singapore, July 20-22 (2005), Lecture Notes in Computer Science **3568** (2005) 267-275
8. Song, D., Cho, M., Choi, C., Shin, J., Park, J., Kim, P.: Knowledge Representation for Video Assisted by Domain-Specific Ontology. in Hoffmann, S.A. et al. (eds.), Pacific Rim Knowledge Acquisition Workshop, PKAW 2006, Guilin, China, August 7-8 (2006), Lecture Notes in Artificial Intelligence **4303** (2006)144-155

9. Huang, X., Acero, A., Hon, H.: Spoken Language Processing: A Guide to Theory, Algorithm and System Development. Prentice Hall PTR (2001)
10. Junqua, J.C., Haton, J.P.: Robustness in Automatic Speech Recognition: Fundamentals and Applications (The International Series in Engineering and Computer Science). Kluwer Academic Publisher (1996)
11. Benjamins, V.R., Casanovas, P., Gangemi, A. and Breuker, J.: Law and the Semantic Web: Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications, Lecture Notes in Artificial Intelligence **3369** (2005)
12. Casanovas, P., Casellas, N., Vallbé, J., Poblet, M., Benjamins, V.R. Blázquez, M., Pea, R., Contreras, J.: Semantic Web: A Legal Case Study. In Davies, J., Studer, R. and Warren P. (eds.), Semantic Web Technologies: Trends and Research in Ontology-based Systems. John Wiley & Sons (2006) 259-280