Developing ontologies for legal multimedia applications

Xavier Binefa[1], Ciro Gracia[1], Marius Monton[2], Jordi Carrabina[2], Carlos Montero[2], Javier Serrano[2], Mercedes Blázquez[3], Richard Benjamins[3], Emma Teodoro[3], Marta Poblet[4], Pompeu Casanovas[3]
[1] *Digital Video Understanding Group, UAB*
[2] *Laboratory for HW/SW Prototypes and Solutions (CEPHIS) UAB*
[3] *Institute of Law and Technology, Law Dpt., UAB*
[4] *ICREA Researcher at the Institute of Law and Technology, Law Dpt., UAB*

Keywords: Semantic Search, Ontology, HW/SW Acceleration Platforms, Reconfigurable Devices, Speaker Diarization, Video Segmentation.

1. Introduction

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 domains as parliaments, courts, ministries, or security and military forces are producing enormous masses of video, audio and text files, the requirement of a specific content management solution have arisen naturally.

The aim of E-Sentencias is to develop a software-hardware system for the global management of the multimedia contents produced by Spanish civil courts. The Civil Procedure Act of January 7th, 2000 (1/2000) introduces the video recording of oral hearings. As a result, Spanish civil courts are currently producing a massive number of DVDs which have become part of the judicial file, together with suits, indictments, injunctions, judgments and pieces of evidence. This audiovisual material is used by lawyers, prosecutors and judges to prepare, if necessary, appeals to superior courts. Nevertheless, there is no available system at present to automatically annotate audiovisual contents within the judicial domain. E-Sentencias proposes a meta-search engine to manage text (legislation, jurisprudence, procedural documents, etc.), images, graph materials, and audiovisual contents in a dynamic way that combines algorithmic techniques with legal ontologies. Both automatic and semiautomatic processes facilitate the exploitation of the stored information by the users' website. In this regard, e-Sentencias involves 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 the efficient navigation and retrieval of judicial documents and additional legal sources.

Section 2 below describes the current situation concerning the audiovisual recording of civil cases in Spain. In Section 3 we offer an overview of the steps followed towards the construction of a conceptual structure to classify video segments and the development of legal ontology applications. Sections 4 and 5 depict respectively the structure and architecture of the video system prototype at the present stage of research and, finally, we conclude by offering some expected results and conclusions in sections 5 and 6.

2. Video Recording of Civil Procedures in Spain

The provisions made by the 1/2000 Civil Procedure Act for the video recording of civil proceedings in Spain do not include a homogeneous protocol establishing how to obtain audiovisual records. Rather, and since an ever growing number of Autonomous Governments in Spain hold competencies 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. And, finally, the procedures to store, classify, and retrieve audiovisual records may vary even from court to court.

As regards the basic typology of civil proceedings, the 1/2000 Act sets two declarative processes: the ordinary proceeding and the verbal proceeding. The main differences between the two lie in the value of the case – more or less than Ä3000, respectively – and the legal object at dispute.

The steps of the process also vary depending on the specific proceeding. On the one hand, the ordinary proceeding starts with a separate, independent oral hearing called "*audiencia previa*" to resolve pre-judiciary issues (documents, evidences to be accepted, etc.), while verbal proceedings take place in the same judicial 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.

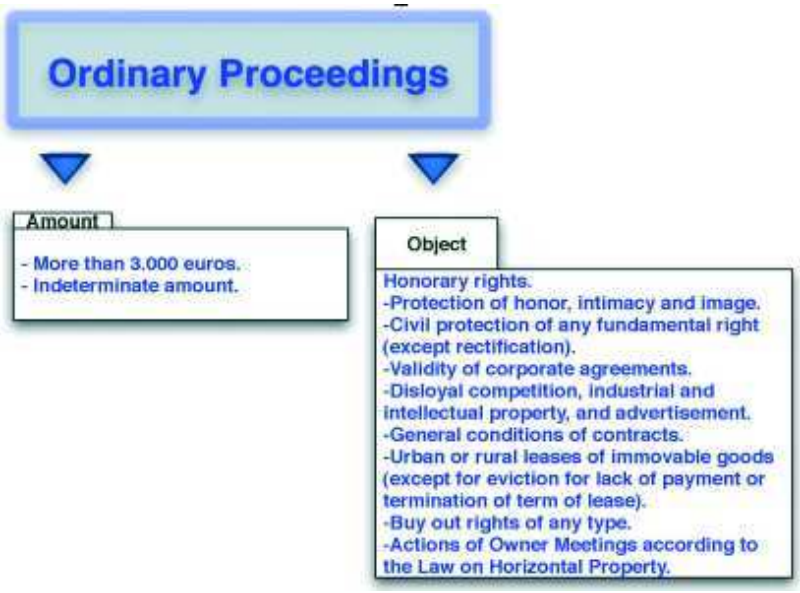

Figure 1. The ordinary proceeding: amount and content

3. Conceptual structure and ontology legal applications

One of the core objectives of e-Sentencias is to develop automatic classification strategies to classify video segments. To do so, we have started from scratch by transcribing a small set of oral hearings (corresponding to fifteen civil cases). Textual transcriptions also mark the different steps of the oral hearing and include a manual coding of legal concepts (i.e. judgment, injunction, cause of necessity, deed, etc.) and legal expressions (i.e. "with the permission of your Honor". In addition, they facilitate the coding of practical rules of procedure that are implicit in the video sequences, such as the following piece of transcription shows:

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.

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 (Haase et al. 2006). In general, the context of use has an impact on the way concepts are interpreted to support certain functionalities. As some aspects of a domain are important in one con-

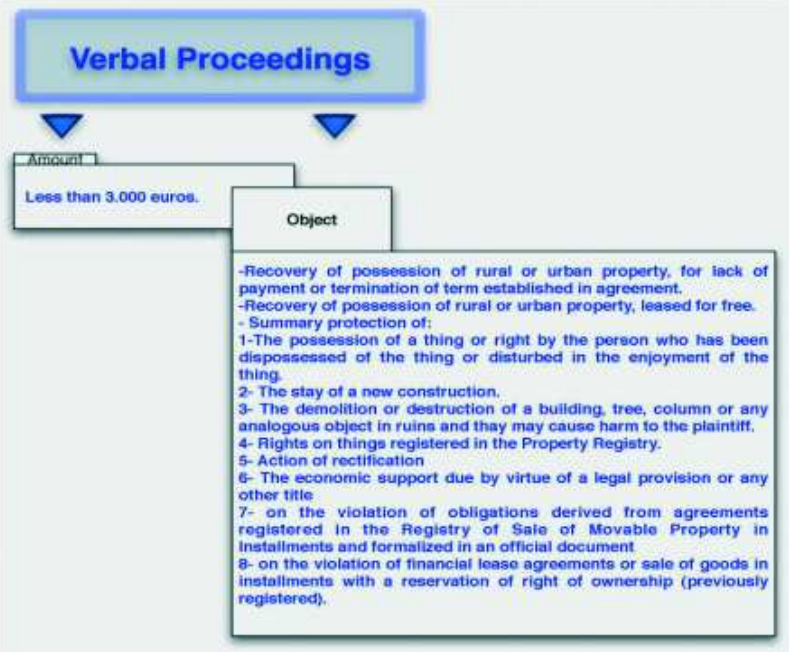

Figure 2. The verbal proceeding: amount and content

text but do not matter in another one, an uncontextualized ontology does not necessarily represent the features needed for a particular use. In order to solve this problem, we have to find ways to enable the representation of different viewpoints that better reflect the actual needs of the application at hand.
When talking about viewpoints, we can distinguish two basic use cases: In the first case, the aim is to provide means for maintaining and integrating different existing viewpoints. In the second use case, one may want to extract a certain viewpoint from an existing model that best fits the requirements of an application.

In many application domains (such as law) it is acknowledged that the creation of a single universal ontology is neither possible nor beneficial, because different tasks and viewpoints require different, often incompatible conceptual choices. As a result, we need to support situations where different parties commit to different viewpoints that cannot be integrated by imposing a global ontology. This situation demands for a weak notion of integration, in order to be able to exchange information between the viewpoints (Stuckenschmidt, 2006). Stuckenschmidt describes one of such examples from oncology. Oncology is a complex domain where several specialties, e.g. chemotherapy, surgery,

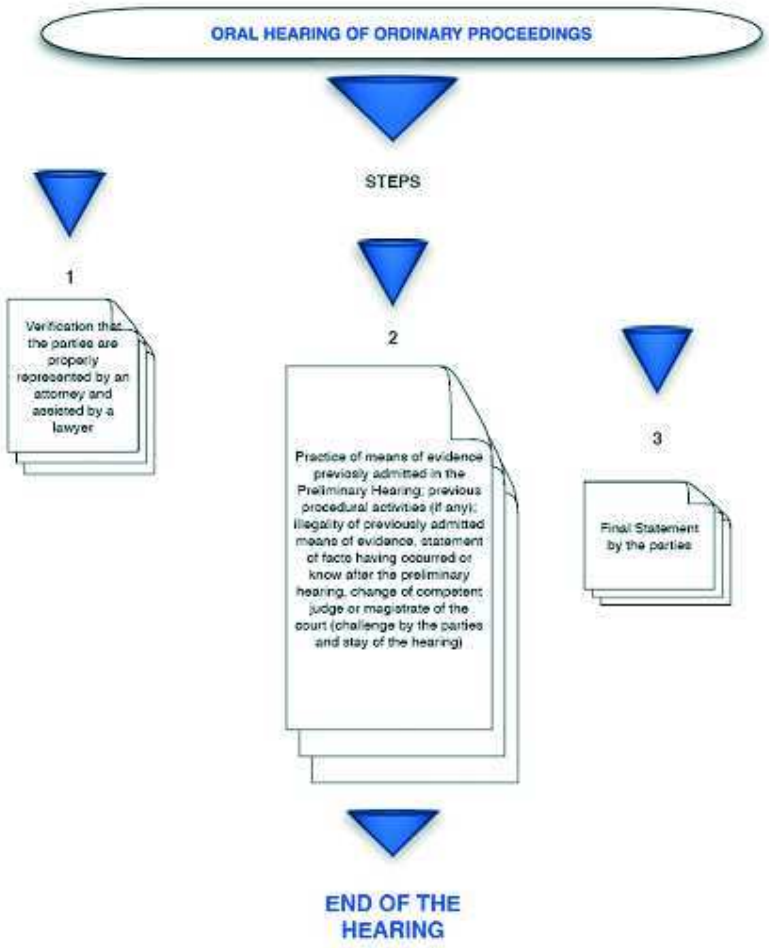

Figure 3. Steps of the process in ordinary proceedings.

and radiotherapy are involved in a sequence of treatment phases, each representing a particular viewpoint.

Law is also a complex domain, where several roles are involved (judge, prosecutor, defendant ...). They must be represented from different points of view, thinking of the possible use of the images of the hearings for multiple (and adversarial) purposes.

We find in the recent literature several approaches to this perspective problem and the so-called 'semantic gap': (i) multi-context ontologies vs. mono-context ontologies (Bensliman et al. 2006 ; Arara and Laurini, 2005 ; Dong and Li, 2006); (ii) low-level descriptors [pixel color, motion vectors, spatio-temporal relationships] vs. semantic descriptors [person,

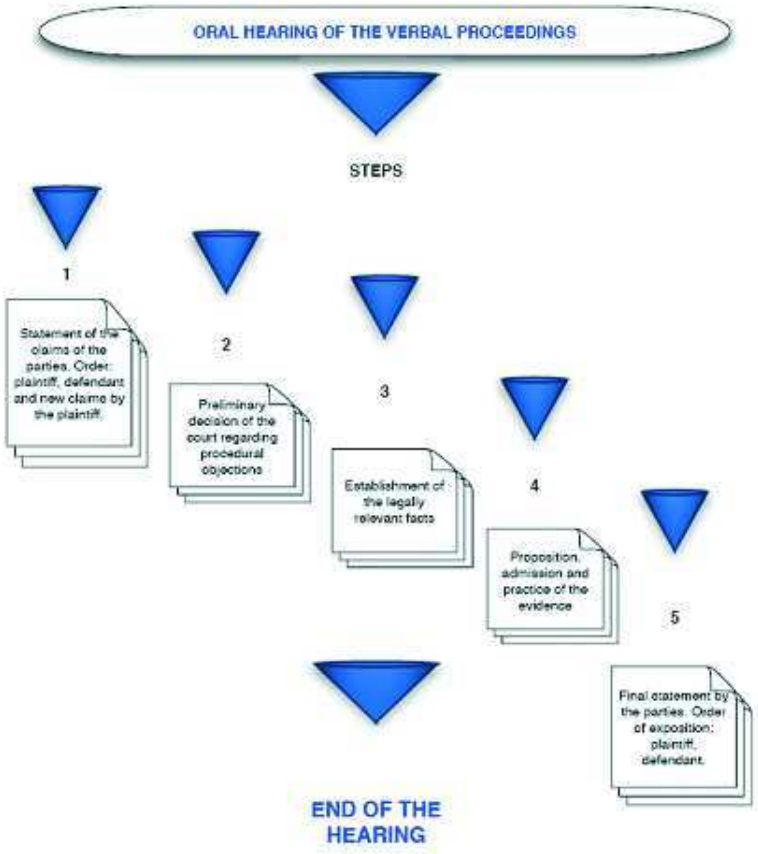

Figure 4. Steps of the process in verbal proceedings.

vehicle...] (Petrides et al. 2005, Athanasiadis et al. 2005, Boehorn et al 2005) ; modal keywords of perceptual concepts [aural, visual, olfactory tactile, taste] vs. content topics (Jaimes et al. 2003a; Jaimes et al. 2003b); (iii) cross-media annotation (Deschachts and Moens 2007).
From a legal multimedia user-centered perspective there are two problems related to these proposals that have to be addressed (i) the definition of context in merging and aligning legal and multi-media ontologies; (ii) the specific exophoric nature of the legal videorecording.
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) (Bouquet et a. 2004; Haase et

<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>
<actor name="plaintiff's attorney" tc="00.01.38">
With the permission of your honor [PROCEDURAL FORMULA, EXCLUSIVE USE BY THE ATTORNEYS] eh do you know whether mrs. **** is being living with her grandmother mrs ** since january 2001
</actor>

Figure 4.

al. 2006).[1] Therefore, to cope with the directionality of information flow, the local domains and the context mapping, which cannot be represented with the current syntax and semantics of OWL, C-OWL is being developed.[2]

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' (Jaimes et al. 2003a,b). 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 (Bloedhorn et al. 2005).[3] But, most important, this definition of context entails a theoretical approach in which 'actions and events in

[1] 'It can be argued that the strengths of ontologies are the weakness of contexts and vice-versa' (Bouquet et al., Haase et al. ibid.).

[2] *Directionality of information flow*: keeping track of the source and the target ontology a specific piece of information; *local domains*: giving up the hypothesis that all legal ontologies are interpreted in a single global domain; *context mapping*: stating that two elements (concepts, roles, individuals) of two ontologies, though extensionally different, are contextually related, e.g., because they both refer to the same object in the word (Bouquet at al. 2004).

[3] 'The main idea of our approach lies in a way to associate concepts with instances that are deemed to be prototypical by their annotators with regard to their visual characteristics' (ibid. 2005: 593).

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' (Song et al. 2005, 2006). 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 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 colour are generally uniform, since they are not considered the relevant aspect of those documents. Thus, court records are technically very poor (see fig. 5), 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. A legal court video record would be completely useless without the audio, because we may only infer procedural (but not substantial) items from the motion. What is important is what is *said* in court, not 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 judgements –precedent-, personal professional records, and so on). This is the reason for a hybrid user-centred approach that is the kernel of our theoretical approach.

4. Structure 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*

Figure 5. Image quality.

axe. The *oral hearing line* presents a timeline divided into segments. Each segment represents a different speech, produced by one of participants in the process: judge, secretary, attorneys, witnesses, etc. Each participant is 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 to any instant of the intervention.
The *oral hearing axe* 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 6 shows, the user interface is divided into two main parts: the upper part contains the video player, the *oral hearing axe* 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

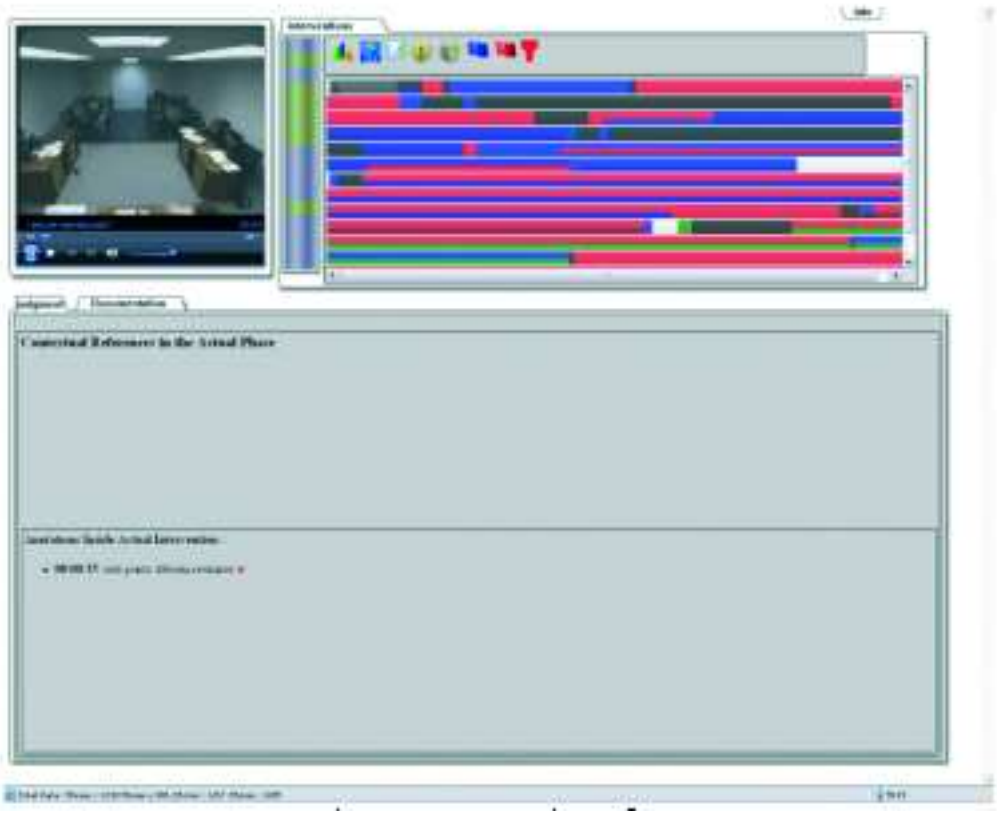

Figure 6. User interface.

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.

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 axe*: 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 axe has the advantage of providing quick access to interventions belonging to a given phase.

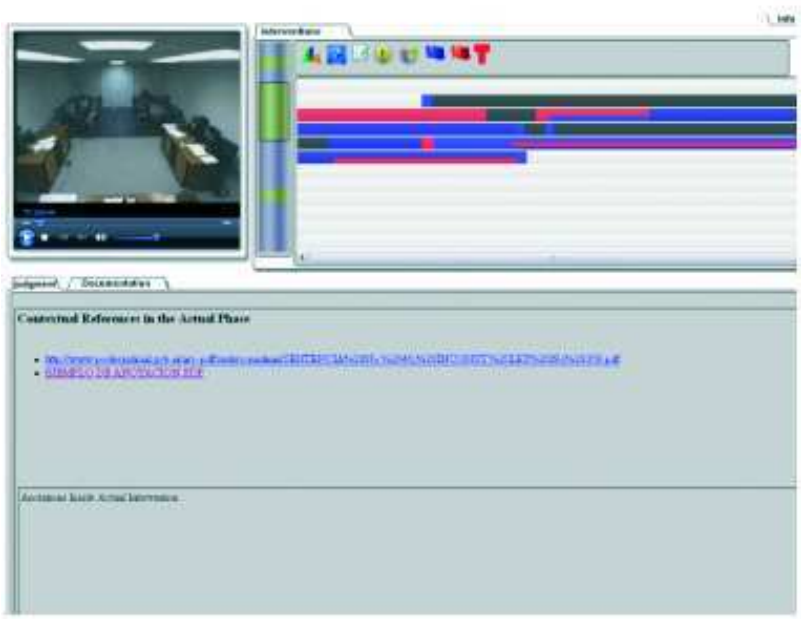

Figure 7. Interventions of one procedural phase and related information.

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.

5. Architecture of the video prototype

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.

6. Conclusions and expected results

In the E-Sentencias project we expect to obtain two different types of results. On the one hand, a fully annotated legal corpus of multimedia oral hearings classified in 15 procedural classes, as regulated by the 1/2000 Act. On the other hand, an operational system with a human-computer interface as described in this paper. Using the system prototype, the automatic capabilities of speaker interventions and phases detection will be tested against manually annotated corpus. It

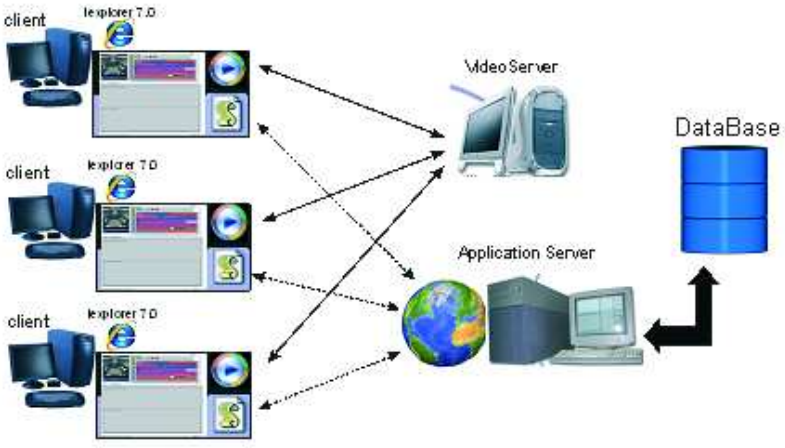

Figure 8. Architecture components and interactions between them.

will also be evaluated in cross-oral hearings retrieval based on hardware accelerated and specifically implemented multimedia ontologies.

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 España, IUAB 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

Arara, A.A., Laurini, R. Formal Contextual Ontologies for Intelligent Information Systems, 2005. *Enformatika* 5: 303-306.

Athanasiadis, T., Tzouvaras, V., Petridis, K., Precioso, F., Avrithis, Y., Kompatsiaris, Y. Using a Multimedia Ontology Infrastructure for Semantic Annotation of Multimedia Content, 2005. Proc. of 5th International Workshop on Knowledge Markup and Semantic Annotation ('05), Galway, Ireland, November 2005

Benjamins, V.R., Casanovas, P., Gangemi, A. and Breuker, J. (ed.). 2005. *Law and the Semantic Web: Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications.* Lecture Notes in Computer Science. Berlin, Springer Verlag.

Bensilimane, D., Arara, A., Falquet, G., Maamar, Z., Thiran, P., Gargouri, F. 2006. Contextual Ontologies. Motivations, Challenges and Solutions. Fourth Biennial International Conference on Advances in Information Systems ADVIS, October 18^{th}-20^{th}, Ankara.

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. A.Gómez Pérez and J. Euzenat (eds.) ESWC 2005, Lecture Notes in Computer Science 3532, 592-607.

Bouquet, P.; Giunchiglia, F., van Harmelen, F., Serafín, L., Stuckenschmidt, H. C-OWL: Contextualizing Ontologies. In D. Fensel et al. ISWC 2003, Lecture Notes in Computer Science 2870: 164-179.

Bouquet, P., Giunchiglia, F., van Harmelen, F., Serafín, L., Stuckenschmidt, H. 2004. Contextualizing ontologies. *Journal of Web Semantics* 26): 1-19.

Breuker, J., Elhag, A., Petkov, E. and Winkels, R. 2002. Ontologies for legal information serving and knowledge management. In *Legal Knowledge and Information Systems, Jurix 2002: The Fifteenth Annual Conference.* IOS Press.

Casanovas, P., Poblet, M., Casellas, N., Vallbé, J.-J., Ramos, F., Benjamins, V.R., Blázquez, M., Rodrigo, L., Contreras, J. and Gorroñogoitia, 2004. J. *D10.2.1 Legal Case Study: Legal Scenario.* Technical Report SEKT, EU-IST Project IST-2003-506826.

Casanovas, P., Casellas, N., Vallbé, J.-J., Poblet, M., Ramos, R., Gorroñogoitia, J., Contreras, J., Blázquez, M. and Benjamins, V.R. 2005. Iuriservice II: Ontology Development and Architectural Design. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law (ICAIL 2005).* Alma Mater Studiorum-University of Bologna, CIRSFID.

Casanovas, P., Casellas, N., Vallbé, J.-J., Poblet, M., Benjamins, V.R. Blázquez, M., Peña-Ortiz, Rl and Contreras, J. 2006. Semantic Web: A Legal Case Study. In J. Davies, R. Studer and P. Warren, editors, *Semantic Web Technologies: Trends and Research in Ontology-based Systems.* John Wiley & Sons.

Casellas, N., Jakulin, A., Vallbé, J.-J. and Casanovas, P. 2006. Acquiring an ontology from the text. In M. Ali and R. Dapoigny, editors, *Advances in Applied Artificial Intelligence, 19^{th} Internatoinal Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (IEA/AIE 2006).* Annecy, France, June 27-30 2006, Lecture Notes in Computer Science 4031, Springer, 1000-1013.

Deschacht, K. and Moens, MF. 2007. Text Analysis for Automatic Image Annotation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Prague, Czech Republic, June 23rd–30th, 2007.*

Dong, A., Li, H. 2006. Multi-ontology Based Multimedia Annotation for Domain-specific Information Retrieval. Proc. Of the IEEE International Conference on Sensor Networks, Ubiquitous and Trusworthy Computing (SUTC' 06).

Haase, Peter; Hitzler, Pascal; Rudolph, Sebastian; Oi, Guilin; Grobelnik, Marko; Mozeti, Igor; Damjan, Bojad Ziev; Euzenat, Jerome; d'Aquin, Mathieu; Gangemi, Aldo; Catenacci, Carola. 2006. *D3.11 Context Languages-State of the Art.* NeOn Project EU-IST Integrated project (IP) IST-2006.

Jaimes, A., Smith, J.R. 2003a. Semi-automatic, Data-driven Construction of Multimedia Ontologies. ICME 203, IEEE.

Jaimes, A., Tseng, B., Smith, J.R. 2003b. Modal Keywords, Ontologies, and Reasoning for Video Understanding. CIVR 2003, E.M.Bakker et al. (eds.) Lecture Notes on Computer Science 2728: 248-259.

Petridis, K., Precioso, F., Athanasiadis, T., Avrithis, Y., Kompatsiaris, Y. 2005. Combined Domain Specific and Multimedia Ontologies for Image Undertanding.

28th German Conference on Artificial Intelligence, Koblenz, Germany, September 11-14.

Song, D., Liu, H.T., Cho, M., Kim, H., Kim, P. 2005. Domain Knowledge Ontology Building for Semantic Video Event Description. W.K. Leow et al. (Eds.) CIVR-05, Lecture Notes in Computer Science 3568, 267-275.

Song, D., Cho, M., Choi, C., Shin, J., Park, J., Kim, P. A. Hoffmann et al. (Eds.) PKAW -06, Lecture Notes in Artificial Intelligence 4303, 144-155.

Stuckenschmidt, Heiner. 2006. Toward Multi-viewpoint Reasoning with Owl Ontologies. In ESWC, pages 259.272.