The e-Sentencias prototype: A procedural ontology for legal multimedia applications in the Spanish Civil Courts

Pompeu Casanovas [a], Xavier Binefa [b], Ciro Gracia [b], Emma Teodoro [a], Núria Galera [a]
Mercedes Blázquez [d], Marta Poblet [e] Jordi Carrabina [c], Marius Monton [c],
Carlos Montero [c], Javier Serrano [c], and José Manuel López-Cobo [d],

[a] *Institute of Law and Tecnology*
UAB, Spain
{Pompeu.Casanovas, Emma.Teodoro, Nuria.Casellas, Nuria.Galera}@uab.cat

[b] *Digital Video Understanding Group*
ETSE, UAB, Spain
{Ciro.Gracia, Xavier.Binefa}@uab.cat

[c] *Laboratory for HW/SW Prototypes and Solutions (CEPHIS)*
ETSE, UAB, Spain
{Jordi.Carrabina, Marius.Monton, Carlos.Montero, Javier.Serrano}@uab.cat

[d] *intelligent Software Components (iSOCO)*
Madrid, Spain
{ozelin, mercedes}@isoco.com

[e] *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. We introduce a software-hardware system for the global management of the multimedia contents produced by Spanish Civil Courts. 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. This paper describes our knowledge acquisition process, sets a typology of Spanish Civil hearings as performed in practice, and a preliminary procedural ontology at its actual stage of development (e-Sentencias ontology). A discussion on procedural, contextual and multimedia ontologies is also provided.

Introduction

Legal professionals are used to consume an important part of their time searching, retrieving, and managing legal information. However, the recent explosion of multimedia legal contents has resulted in rising costs and requires more management capacities than ever before. Improving the functionalities for search, retrieval, and management of multimedia legal documents is paramount to fully unlock the potential of those contents for legal practice and to develop specific management solutions for different profiles of legal users: lawyers, judges, prosecutors, court clerks, etc.

In Spain, 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 multimedia files that have substituted the written transcripts and have become part of the judicial file, together with suits, indictments, injunctions, judgments and pieces of evidence. Lawyers, prosecutors and judges need to access these contents when preparing similar cases or, if necessary, when appealing to superior courts. To date, nevertheless, there is no available system within the judicial domain to automatically index, tag, or annotate those audiovisual contents.

Furthermore, the provisions made by the 1/2000 Civil Procedure Act for the video recording of civil proceedings in Spain do not include a protocol establishing how to obtain audiovisual records. Rather, and since an ever growing number of Autonomous Communities 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 VHS videotapes, CDs or DVDs. In addition, the procedures to store, classify, and retrieve audiovisual records may vary even from court to court, with no common database available to store the audiovisual records.

The use of ontologies applied to legal multimedia was already faced in the e-Court Project (Breuker et al. 2002). The aim of the e-Sentencias Project is to develop a software-hardware system for legal professionals to access and manage the multimedia files connected to their cases across different platforms, formats, and languages. The project includes the development of both an ontology-based meta-search engine and a specific web-based, reconfigurable hardware platform capable to accelerate and improve the quality of the document retrieval process. The objectives of this approach are: (i) to facilitate search, discovering and management of multimedia legal contents based on their meaning; (ii) to optimize search results by automatically linking video-to-text; (iii) to improve the organization of the search memory, and (iv) to save time to users. The e-Sentencias Project, in sum, combines legal and contextual ontologies with algorithmic techniques. 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 multimedia records that, coupled with textual semantics, allows the efficient navigation and retrieval of judicial documents and additional legal sources.

Section 1 of the paper offers a brief typology of civil procedures covered by e-Sentencias. In Section 2 we present an overview of the knowledge acquisition process towards the construction of a conceptual structure to classify video segments and the development of legal ontology applications. Section 3 describes the development of the e-Sentencias ontology. Section 4 sets the discussion on our research goals focusing on the state of the art of procedural, contextual and multimedia ontologies. Section 5 shows the structure and architecture of the video system prototype at the present stage of research and, finally, we conclude in section 6 by offering some conclusions and expected results.

1. Civil Procedures in Spain

The main inputs of e-Sentencias are multimedia files that contain hearings of civil cases [*vista oral*]. The 1/2000 Act establishes two basic procedures or "declarative processes": the ordinary proceedings and the verbal proceedings (see Figures 1 & 2 below). The main differences between the two procedures lie both in the value of the case-more or less than €3000, respectively-and the legal object at dispute.

The steps of civil processes also vary depending on the specific procedure. On the one hand, ordinary proceedings [*juicio ordinario*] start with a separate, independent motion called preliminary hearing [*audiencia previa*] to resolve pre-judiciary issues (documents, evidences to be accepted, etc.), while in the verbal proceedings [*juicio verbal*] all steps take place at the same judicial event. On the other hand, the claim of the plaintiff is contested in written terms in the ordinary proceeding, while in the verbal proceeding is replied orally.
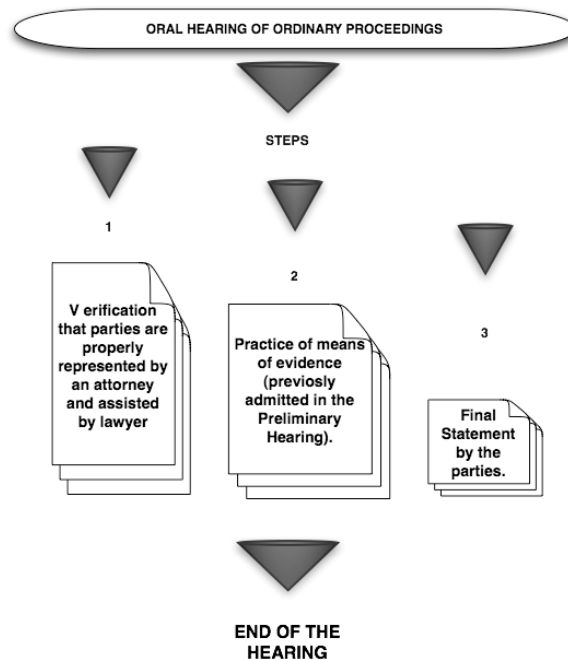

Figure 1. Steps of the process in ordinary proceedings [*juicio ordinario*]
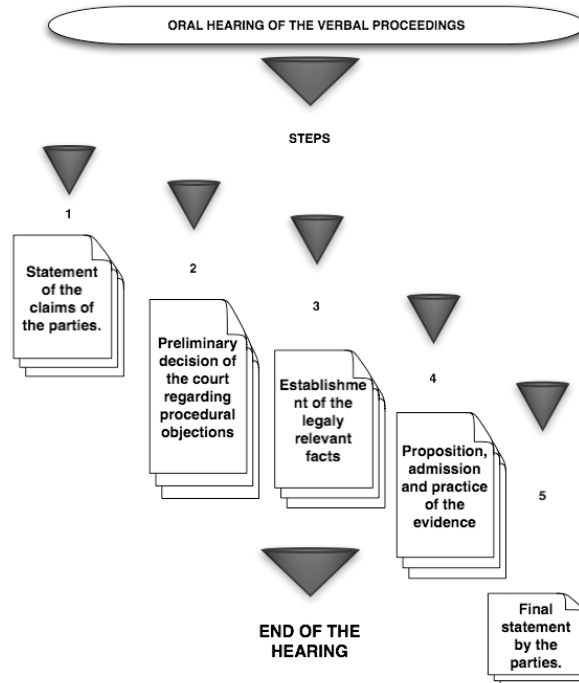
Figure 2. Steps of the process in verbal proceedings [*juicio verbal*]

2. Knowledge Acquisition Methodology

The figures above represent the steps of civil processes as formally set by the 1/2000 Act. However, the routines and constrains of daily practice in court result in a far more complex typology, revealing interesting differences between the formal provisions of the law and the actual development of the hearings. Thus, we identified 14 different types of proceedings populating the ordinary/verbal division. Therefore, a new typology of hearings has to be superimposed over the original division. This fact adds some complexity to the conceptualization of the hearings, because each particular judgement has to be classified as (i) preliminary hearing, ordinary proceeding, verbal proceeding, or precautionary/provisional measure and (ii) within one type of the proposed typology.

Preliminary hearing *Audiencia previa*	Preliminary hearing *Audiencia previa*
Ordinary proceedings *Juicio ordinario*	Ordinary proceedings *Juicio ordinario*
Verbal proceedings *Juicio verbal*	Verbal proceedings *Juicio verbal*
	Proceedings regarding the capacity of persons (special proceedings) *Procesos sobre capacidad de las personas (procesos especiales)*
	Matrimonial proceedings (special proceedings) *Procesos matrimoniales y menores (procesos especiales)*
	Judicial division of states *División judicial de patrimonios*
	Exchange proceedings (special proceedings) *Procedimiento cambiario (procesos especiales)*
	Summon proceedings (special proceedings) *Procedimiento monitorio (procesos especiales)*
	Mortgage proceedings *Procedimientos hipotecarios*
	Extraordinary appeal for procedural infringements and cassation proceedings *Recurso extraordinario por infracción procesal y casación*
	Opposition to judicial determination of fees for undue fees *Impugnación tasación de costas por honorarios indebidos*
	Opposition to enforcement *Oposición a la ejecución*
Precautionary and provisional measures *Medidas cautelares y provisionales*	Precautionary measures *Medidas cautelares*
	Provisional measures *Medidas provisionales*

Figure 3. Typology of civil hearings as emerged in daily practice

To proceed with the knowledge acquisition process, we represented the new typology as a set of workflows able to contain (and interact with) other structures of practical knowledge (i.e. sub-typologies, concepts, keywords, content groups, etc). The basic units of the *workflows* are 'procedural stages'. Stages are the unities in which 1/2000 Act segments civil procedures and they are also the basic video sequences of the e-Sentencias prototype. In workflows, procedural stages include: (i) actions (rhombus) which imply to move forward or backward to another stage; (ii) procedural legal concepts, (iii) and content of legal concepts (boxes). An example of such a workflow is shown in Figure 4.
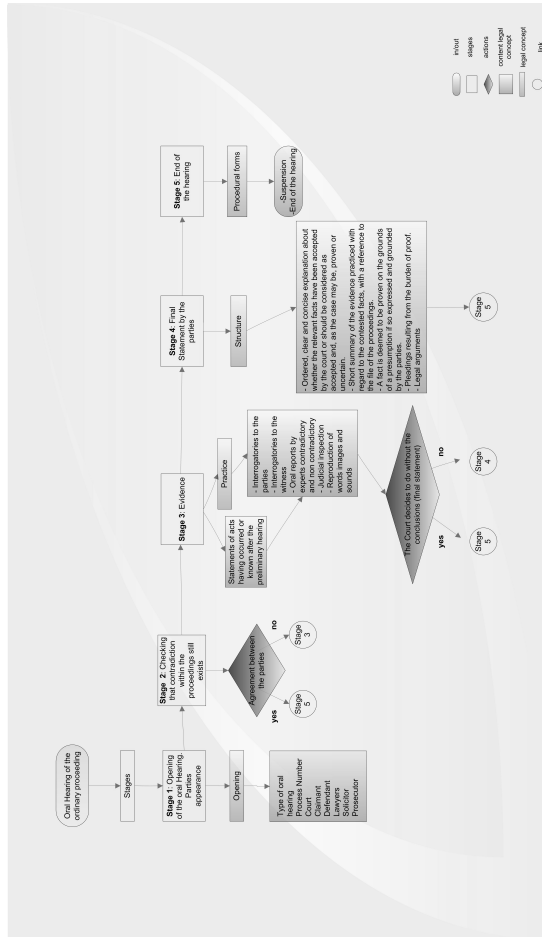
Figure 4. General structure of the hearing in ordinary proceedings [*juicio ordinario*]

At this point, we followed two different knowledge acquisition strategies. First, we set up a focus group composed of professional attorneys, judges and solicitors. We used standard eliciting techniques (e.g. interview, discussion, and brainstorming) to extract from their professional knowledge the conditions of use and estimated frequencies of expressions and legal concepts as they appear in recorded hearings. As a result of that, we elicited a list of roughly 900 terms and legal practical expressions, and we linked them to the general typology of 14 oral hearings and to concrete subjects contained in each type. Figure 5 below shows this kind of work for one of the legal processes (subtypes).

Terms	Less frequent				Most frequent
Action for non-consented sub-lease and infringement of agreement/	1	2	3	4	5
Contract					x
Sub-lease			x		
Renting					x
Owner					x
Lease					x
Consent			x		
Non consent			x		
Summons				x	
Renting					x
Gratuitous	x				
Dwelling					x
House/ flat/ apartment				x	

Figure 5. Estimated frequency of terms associated with hearings

The second strategy focuses on the transcripts of the hearings. We transcribed a small set of fifteen hearings marking their different steps and coding manually legal concepts (i.e. 'injunction', 'cause of necessity ', 'deed', etc.), legal expressions (i.e. 'with the permission of your Honor'), and practical procedural rules that are implicit in the video sequences, such the following piece shows:

```
<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>
```

Figure 6. Transcription of a Civil Hearing

This is only a first level of textual and visual annotation, 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. To facilitate both the transcription and annotation processes we set up an annotation tool interface (Fig. 7). A second list of terms was then extracted, and a third list of concepts was built by convention according to the following parameters: (i) speech markers, (ii) pragmatic markers, (iii) cognitive markers.

Speech markers are of sociolinguistic nature (pitch on word, pitch on phrases, short pause, emphatic stress, overlapping speech, rising shift in intonation ...). *Pragmatic markers* are of discursive and rhetoric nature (opening speech, ending expression, ending pause, competitive talk, spacing out ...). *Cognitive markers* are of communicative (link of subjects to reference and co-reference phenomena: metaphor, anaphora, cathafora) and emotional nature (stress, plea, anxiety, surprise ...).

This methodology comes from the fields of linguistic, discourse and cognitive analysis, reconstructing the main specific frameworks in which situated meaning is produced (Clancey, 1994; Nerlich and Clarke, 2001). We may notice here that there is no need to use all the results obtained through this kind of annotation work to build up the final ontology. Using discourse, pragmatic and cognitive markers is only a way to clarify and organize the experts' knowledge on these particular legal frameworks. However, this is crucial in order to identify the most common patterns of meaning used in each type of civil proceeding, avoiding misinterpretations.

This expert linguistic knowledge allows grasping the main features needed to build the procedural cognitive contexts in which ontologies may add semantics to the further diarization of the multimedia content. *Diarization* refers to the process of audio stream segmentation. We will be come back to it later on.

It may be noticed by now that MPEG-7 descriptors which are designed primarily to describe low-level audio or visual features such as color, texture, notion, audio energy, and so forth, or attributes of AV content such as location, time, quality and so forth, have
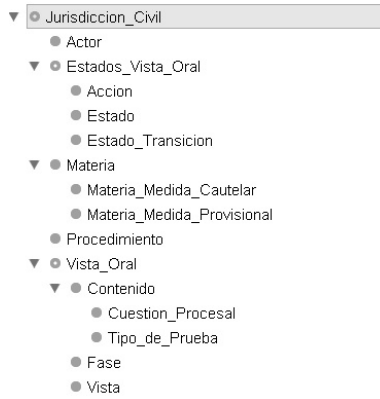
been loosely followed to *interpret* the main features of the speech (Martínez, 2004). On the contrary, higherlevel AV features such regions, segments, objects and events have been substituted for the cognitive and emotional features of the actors link to the content of the hearing.

Figure 7. Annotation tool interface

3. The e-Sentencias ontology

We have proceeded to model the e-Sentencias ontology on the basis of the procedural knowledge practically used in the Court hearings. This procedural knowledge is what we have been labeling as "professional legal knowledge" (PLK) (Casanovas et al. 2004, 2005, 2006; Casellas et al. 2006). PLK is the kind of knowledge which is produced through the daily practice and behaviour of professional people. In this case, as a result of the knowledge acquisition process, we obtained several lists of concepts, expressions and terms ordinarily used and shared by judges, lawyers, court secretaries and clerks in the civil procedures. We organized them following the discourse and cognitive guidelines described so far, but bearing in mind the needs and requirements of the final users of the system.

To date, the e-Sentencias ontology consists of 16 classes, 22 attributes and 169 instances. The ontology provides a model of the following aspects:

- The hearings as found in the civil jurisdiction practice and their different phases, stages and actions.
- The agents who take part in each one of the phases of the hearings (lawyer, plaintiff, judge, prosecutor, etc.)
- The contents dealt with in each one of the modelled hearings

Figure 8. Hierarchy of classes

We may identify the different phases the hearings in different proceedings consist of, bearing in mind that within each phase we have identified an order of precedence among the phases. This order of precedence does not take place within a hearing, but rather, and depending on the different actions, the hearing jumps into another phase. For this purpose, we have created the classes Phase and Hearing in the ontology, where a hearing is characterized by a set of phases and agents involved and by the fact that it is related to a certain type of civil-jurisdiction proceedings. An example of an instance within the hearing is the one shown in Figure 9.

Figure 9. Instances of the class Hearing

Furthermore, the type of claim is also modeled (or else the issues dealt with) within the class hearing, which make sense within the course of a hearing, as can be seen in the right part of Fig. 9. As mentioned above, there is a relationship of precedence between the different phases a hearing consists of. For instance, Fig. 4 shows that the oral hearing of the ordinary proceedings has five phases: (i) opening of the hearing, (ii) verification that the action is pending, (iii) evidence, (iv) final conclusions, (v) and end of the hearing. The hearing consists of a set of phases. We define the order of precedence within the transition stages. Fig. 10 below offers an example of instance of phase. In such class we determine to which hearings a certain phase relates to (in the example, the phase of evidence takes place in all hearings) and we determine also what sub phases the phase is divided into (the phase of evidence consists of three sub phases: carrying out the evidence, admitting the evidence and putting forward the evidence).
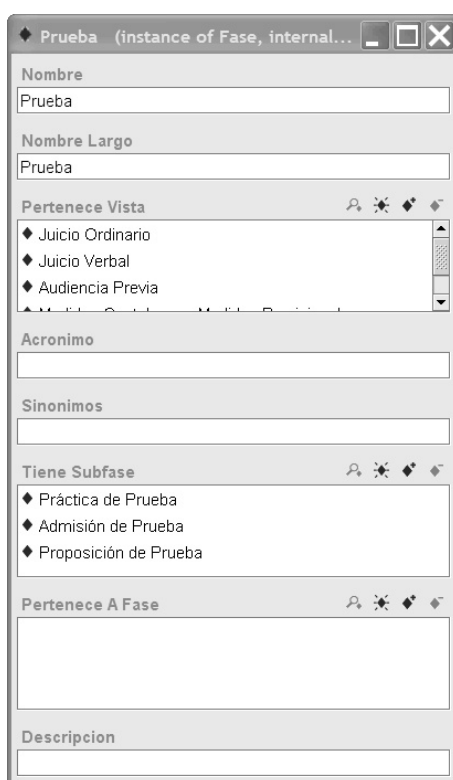
Figure 10. Instances of the class Hearing

Another aspect which has been taken into account when modeling the ontology are the specificities identified during the hearings, as, although there is an order of prece-dence of the phases within the hearing, there are certain points where we change from one phase to the other, that is to say, there is a diagram of stages. A clear example can be seen in the prior graphic (Fig. 4), where, for example, if there is coincidence between the phases we pass directly from phase 2 to checking the subsistence of proceedings in phase 5 or end phase of the hearing. These transitions between phases have been modeled in the class StageTransition. Another example can be found in Fig. 11 below.
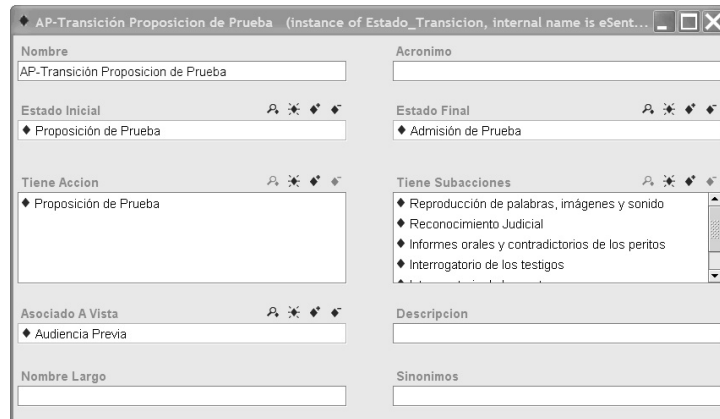
Figure 11. Instances of the class stage of transition

In the stages of transition we may identify the initial stage, the final stage, the hearing where the transition takes place, as well as the related action. Furthermore, we have also identified a series of sub actions which may or not occur during the transition. Thus, in the period for submitting the evidences [*proposición de prueba*], different sub actions may occur, such as the oral and contradictory reports of the experts, the interrogatories made to the witnesses, etc., which may or not be carried out within this period.

Finally, this ontology is used in combination with a set of thesaurus of synonyms where different expressions related to one instance are specified, for example, for the instance of the phase 'Statements of the Claim', the following synonyms have been defined:

- Statement of the claim of the party
- This party ratifies the statement of claim
- This party ratifies the written statement of the claim
- I reaffirm again
- I ratify
- I confirm the written statement of defence
- I ratify the written defence
- I require these proceedings
- I petition these proceedings to be admitted to evidence
- Pleas provided as grounds for the petition
- Challenging of the document
- Statement regarding a plea of the other party
- Put forward an exception

These synonyms have a common trait: they are all expressions which are formulae used by the agents involved in the proceedings and will thus allow us to carry out the semantic searches in the audiovisual context by processing the audio.

4. Discussion: procedural, contextual and multimedia 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 context 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 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, it may be required 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 among the viewpoints (Stuckenschmidt, 2006). Stuckenschmidt describes one of such examples from oncology. Oncology is a complex domain where several specialties, e.g. chemotherapy, surgery, and radiotherapy are involved in a sequence of treatment phases, each representing a particular viewpoint.

Procedural law -Court proceedings- 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.

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). 'It can be argued that the strengths of ontologies are the weakness of contexts and vice-versa' (Bouquet et al., 2004; also, Haase et al. 2006). In this specific sense, contextual ontologies have to deal at least with three problems: the directionality of information flow, local domains and context mapping.

Directionality of information flow means keeping track of the source and the target ontology as specific piece of information; *local domains* means giving up the hypothesis that all ontologies are interpreted in a single global domain; *context mapping* means 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 world (Bouquet at al. 2004). The e-Sentencias ontology faces these three problems, and it is being modeled within the workflow of each type of hearing. Thus, it is contextsensitive. We modeled the directionality of the informational flow as transitional states of the different steps of the oral hearings. We treated the specific domain of modeling an ordinary proceeding (*juicio ordinario*) as a local domain.

Regarding the context mapping, we are dealing with two separate ontologies with a different scope. The aim of the first one is to model judicial professional knowledge -the

OPJK ontology we have been developing since the European Project SEKT (Davies et al. 2006). The second one models the procedural steps and development of the Spanish verbal and ordinary hearings, including all their phases and some common elements contextually related (e.g, the agents involved in the hearings). Figure 12 shows the contextual relations of transitional stages for the opening of the hearings [*apertura*].
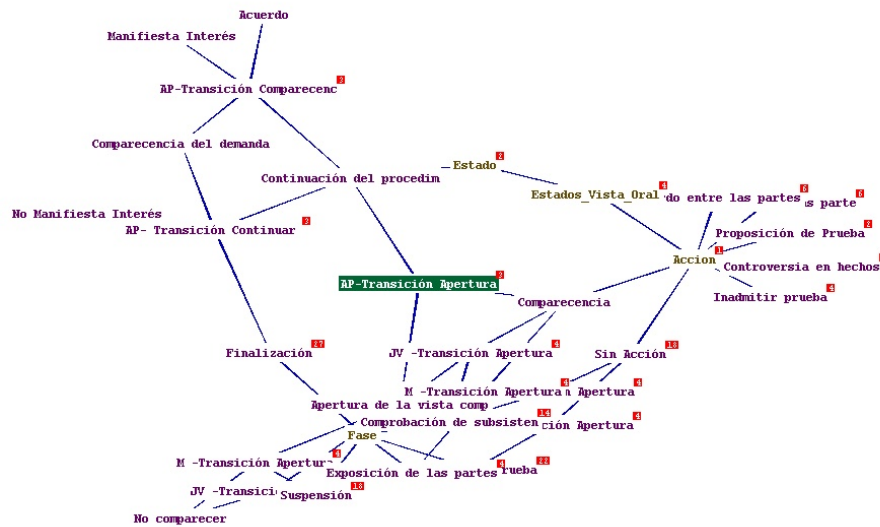

Figure 12. Transitional stage graph (apertura)

Strictly speaking, multimedia ontologies take a different point of view onto the context. Context is defined currently in a more cognitive way 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. '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' (Bloedhorn et al. 2005).

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' (Song et al. 2005, 2006). In such an approach, visual low level features, object recognition and audio speaker *diarization* -the 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. The distance between low-level features and their possible interpretation is known as 'the semantic gap'.

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

scriptors [person, vehicle ...] (Petrides et al. 2005, Athanasiadis et al. 2005, Boehorn et al 2005) ; (iii) modal keywords of perceptual concepts [aural, visual, olfactory tactile, taste] vs. content topics (Jaimes et al. 2003a; Jaimes et al. 2003b); (iv) cross-media annotation (Deschachts and Moens 2007).

In the e-Sentencias ontology we have not properly taken a pure multimedia perspective because 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 records are technically very poor (see fig. 13), 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. *Exophoricity* means that what is said in court refers to other scenarios and ambiences situated outside the court. However, actions, events and stories are referred as well to 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 task of building an ontology, 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.

The e-Sentencias ontology is focused on civil proceedings as they are really performed into the Spanish Courts, it takes into account the different contexts embedded into the workflows of each type of hearing, and it is addressed to multimedia content. But it is not a properly contextual or multimedia ontology, but a *procedural* one.

Figure 13. Image quality

5. Structure and architecture of the e-Sentencias Prototype

The parallel development of an intuitive user interface, adapted to different profiles of legal professionals, 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 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 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.
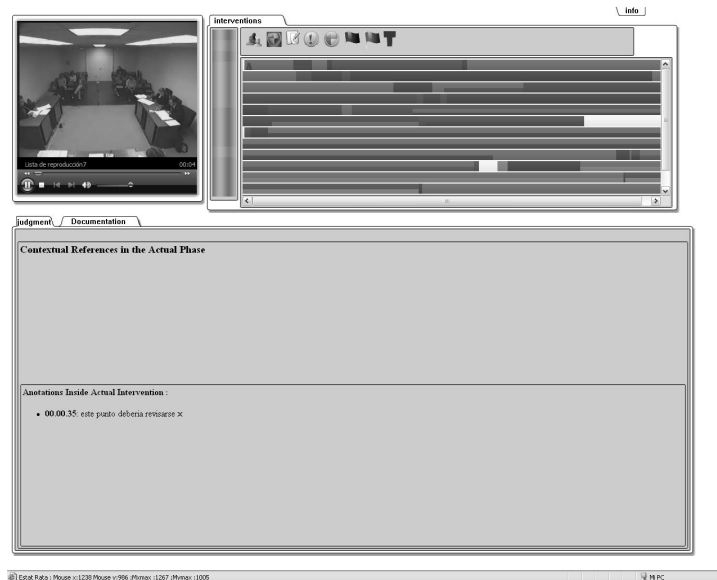
Figure 14. User interface

As Figure 14 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.
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.
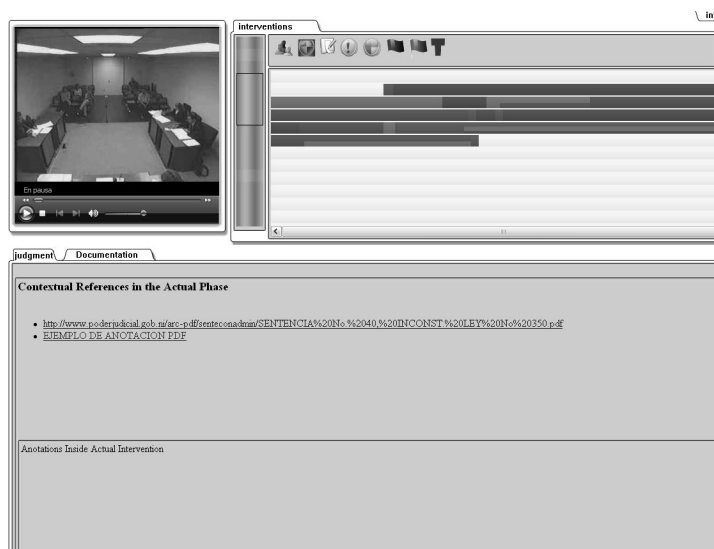

Figure 15. Interventions of one procedural phase and related information

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.

The architecture of the system, finally, 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.
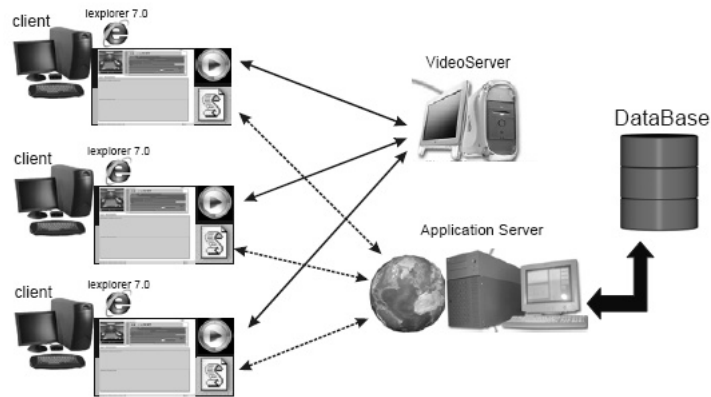
Figure 16. Architecture components and interactions between them

6. Conclusions and expected results

The bottom-up methods employed so far deserve a further discussion, since no clustering or semi-automatically ontology extraction tools have been used. And, yet, it is our contention that the e-Sentencias application can reach its maximum efficiency by automatically segmenting and annotating the video contents. However, this is a research goal that requires, first, a clear understanding of the conceptual content of the hearings. Therefore, we developed a research strategy in which keywords, ontologies and the user interface prototype would be worked out at the same time.

A second reason of this bottom-up approach lies on the very nature of the institutional conditions we are dealing with. Video records of Spanish courtrooms are neither officially transcribed nor digitally stored in large databases. Yet, most multimedia ontology extraction methods are applied to vast collections of digital images or video sets stored in libraries or large databases (Yahoo news, broadcasting TV, sports such as soccerĚ). The knowledge acquisition and ontology building processes are therefore addressed to fill up the semantic gap between metadata, low-level (physical features) and high level descriptors (interpretation, human interaction) in order to perform a more effective classification, information retrieval or search and browsing within databases or websites.

Several techniques have been proposed: (i) *Automatic Image Annotation (AIA) and Confidence Clustering*(CC) (Stauder et al. 2004); (ii) *fusing* MPEG-7 visual descriptors (Sirou et al. 2005); (iii)*pictorially and multimedia enriched ontologies* (Bertini et al. 2006, Bimbo and Bertini 2007); (iv) *salience* and *visualness* calculation of multimedia entities(Moens, 2007; Deschacht and Moens, 2007; Deschacht et al. 2007); (v) use of *probabilistic grammar, Bayesian networks* combined with semantic constraints (Town, 2006); (vi) *inter and itra-contextual segments for video segment retrieval* (Pattanasri et al. 2005; (vii) interoperability of OWL with the MPEG-7 MDS (MOREL ontology) (Tsinaracki et al. 2004)

In our case, however, lawyers and judges store the multimedia files obtained from the court office on individual basis (typically archiving them together with the case records).

No written transcription of the hearing is provided, so that professionals have to exclusively rely on multimedia contents. Our main goal, then, is not focused on image, shot, event, scene, frame detection or recognition patterns, but offering a semi-automatic way to easily search each step or phase of the hearing in every single record.

These are the reasons why we adopted a bottom-up strategy, carefully transcribing a sample of each type of hearing, building from scratch the conceptual structure, and hand making a preliminary segmentation of the hearings. Some empirical studies show that 'retrieval based on transcription of the speech in video data adds more to the average precision of the result than content-based retrieval' (Hollink et al. 2004). However, this leads to a kind of paradox, since the amount of videos to manage grows so large that it becomes impossible to process them manually. This is also the main management problem that lawyers, judges and court offices are facing with their multimedia files.

Therefore, once done the manual annotation process and built a preliminary procedural ontology, we propose to handle the problem by using speech recognition, which is as an already mature technology in other domains (i.e. information services). Speech recognition 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* (Huang, 2001).

The required metadata needed to draw the *oral hearing axis* are temporal marks at which each procedural phase of the judicial process starts or switches. Each process has its own steps, and each phase has its own actors. During these phases, all the actors start and end up their interventions with the same utterances, i.e. in Spanish, *con la venia* [approx. 'thank you your Honor'], *'no hay más preguntas Señoría'* ['no more questions, your Honor']. We state that by spotting these and similar utterances, we will be able to settle the starting and ending point at any step of the hearing.

This approach is simple, but relies on the effectiveness of the recognition of the keywords. In order to improve this effectiveness, we are training the acoustics models of the speech recognition system, nowadays based in a commercial motor (Loquendo), with data stemming from the videotaped hearings. We expect to recognize close to a 100system will be trained with the whole data set.

On the other side, required metadata for the oral hearing line are the moments when actors start and finish. These moments are found by means of speaker identification techniques. This part of the research is still at an early stage and we do not have significant results yet. We should have in mind that speech recognition system cannot be completely trusted to obtain a perfect speech to text transcription yet (Junqua, 1996).

In the e-Sentencias and other related projects we expect to obtain two different types of results. On the one hand, a fully annotated legal corpus of multimedia oral hearings classified in 14 procedural classes, as regulated by the 1/2000 Act. On the other hand, an operational system with a humancomputer interface, adapted to different user profiles, 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 will also be evaluated in cross-oral hearings retrieval based on hardware accelerated and specifically implemented multimedia ontologies.

Acknowledgements

References

[1] Arara, A.; Benslimane, D: Towards Formal Ontologies Requirements with Multiple Ontologies. H. Christiansen et al. (Eds.), FQAS, LNAI 3055, Berlin, Heidelberg: Springer-Verlag (2004), 150-160.

[2] Arara, A.A., Laurini, R.: Formal Contextual Ontologies for Intelligent Information Systems. Enformatika 5 (2005), 303-306.

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

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

[5] Benslimane, D.; Roussay, C.; Vangenot, C.; Arara, A.: Towards a Contextual Content of Ontologies. N. Zhong et al. (Eds.) ISMIS-03, LNAI 2871, Berlin Heidelberg, Springer Verlag (2003), 339-343.

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

[7] Bertini, M.; D'Amico, G.; Bimbo del, A., Torniai, C.: Using knowledge Representation Languages for Video Annotation and Retrieval. H. Legind Larsen et al. (Eds.), FQAS-06, LNAI 4027, Berlin, Heidelberg: Springer Verlag (2006), 634-646.

[8] Bimbo del A.; Betini, M.: Multimedia Ontology Based Computational Framework for Video Annotation and Retrieval. N. Sebe, Y. Liu, Y. Zhuang, MCAM-07, LNCS 4577, Berlin Heidelberg: Springer Verlag (2007), 18-23.

[9] 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, Lecture Notes in Computer Science 3532, Belrin, Heidelberg,: Springer Verlag (2005), 592-607.

[10] Bouquet, P.; Giunchiglia, F., van Harmelen, F., Serafín, L., Stuckenschmidt, H. C-OWL: Contextualizing Ontologies. In D. Fensel et al. ISWC, Lecture Notes in Computer Science 2870, Berlin, Heidelberg: Springer Verlag (2003), 164-179.

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

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

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

[14] Casanovas, P., Casellas, N., Vallbé, J.-J., Poblet, M., Ramos, R., Gorroñogoitia, J., Contreras, J., Blázquez, M. and Benjamins, V.R.: 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, 2005.

[15] Casanovas, P., Casellas, N., Vallbé, J.-J., Poblet, M., Benjamins, V.R. Blázquez, M., Peña-Ortiz, Rl and Contreras, J. :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, Chichester (2006), 259-280.

[16] Casellas, N., Jakulin, A., Vallbé, J.-J. and Casanovas, P.: Acquiring an ontology from the text. In M. Ali and R. Dapoigny, editors, Advances in Applied Artificial Intelligence, 19th 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 (2006), 1000-1013.

[17] Clancey, W.: Situated Cognition: How representations are created and given meaning, in Lessons from Learning, R. Lewis and P. Mendelsohn (Eds.), Amsterdam: North-Holland (1994), 231-42.

[18] Davies, J.; Studer, R.; Warren, P. (Eds.): Semantic Web Technologies. Trends and Research in Ontology-based Systems. John Wiley & Sons, Ltd, Chichester UK, 2006.

[19] Deschacht, K. and Moens, MF.: Text Analysis for Automatic Image Annotation. In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Prague, Czech Republic, June 23rdŰ30th, 2007.

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

[21] Haase, P.; Hitzler, P.; Rudolph, S.; Oi, G.; Grobelnik, M.; Mozeti, I.; Damjan, B.Z.; Euzenat, J.; d'Aquin, M.; Gangemi, A.; Catenacci, C.: D3.11 Context Languages-State of the Art. NeOn Project EU-IST Integrated project (IP) IST-2006.

[22] Hollink, L.; Nguyen, G.P; Koelma, D.C.; Schreiben, A.T.; Worring, M. :User Strategies in Video Retrieval: a Case Study. In P. Enser et al. (Eds.), CIVR-04 , LNCS 3115 (2004), 6-14.

[23] Huang, X., Acero, A., Hon, H.H; Reddy, R.: Spoken Language Processing: A guide to theory, algorithm and system development. Prentice Hall. New Jersey, 2001.

[24] Jaimes, A., Smith, J.R. : Semi-automatic, Data-driven Construction of Multimedia Ontologies. ICME 2003, IEEE.

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

[26] Jung, J.J. ; Kim, Hong-Gee ; Jo, Geun-Sik. : Alignment-Based Preeprocessing of Personal Ontologies on Semantic Social Network. Apolloni et al. (Eds.) KES-07/WIRN-07, Part II, LNAI 4693, Berlin Heidelberg: Springer- Verlag (2007), 255-262.

[27] Junqua, J.C., Haton, J.P.: Robustness in Automatic Speech Recognition, Kluwer Academic Publishers, 1996.

[28] Leenheer, de P.; de Moor, A.; Meersman, R.: Context Dependency Management in Ontology Engineering: A Formal Approach. S. Spaccapietra et al. (Eds.), Journal on Data Semantics VIII, LNCS 4380 (2007), 26-56.

[29] Martínez, J.M. (Ed.) MPEG-7 Overview. ISO/IEC JTC1/SC29/WG11. http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm (November 6th, 2007).

[30] Moënne-Loccoz, N.; Janvier., B.; Marchand-Maillet, S.; Bruno, E.: An Integrated Framework for the Management of Video Collection. S. Bengio, H. Boulard (Eds.), MLMI-04, LNCS 3361 (2005), 101-110.

[31] Nerlich, B., Clarke, D.D. : Semantic fields and frames: historical explorations of the interface between language, action, and cognition, Journal of Pragmatics 32 (2001), 125-150.

[32] Pattanasri, N.; Chatvichienchai, S.; Tanaka, K. : Digital Libraries : Implementing strategies and sharing experiences. E.A. Fox et al. (Eds.), ICADL-05, LNCS 3815 (2005), 119-128.

[33] Pertridis, K., Precioso, F., Athanasiadis, T., Avrithis, Y., Kompatsiaris, Y. : Combined Domain Specific and Multimedia Ontologies for Image Understanding. 28th German Conference on Artificial Intelligence, Koblenz, Germany, September 11-14, 2005.

[34] PPorzel, R.; Gurevych, I.: Contextual Coherence in Natural Language Processing, P. Blackburn et al. (Eds.) CONTEXT-03, LNAI 2680, Berlin, Heidelberg: Springer-Verlag (2003), 272-285.

[35] Rifaieh, R.; Arara, A.; Benharat, A.N.: MurO: A Multi-representation Ontology as a Foundation of Enterprise Information Systems. G. Das and V.P. Gulati (Eds.): CIT-04, LNCS 3356, Berlin, Heidelberg, Springer- Verlag (2004), 292-301.

[36] 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.

[37] 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 (2006), 144-155.

[38] Spyrou, E.; Le Borgne, H.; Mailis, T. ; Cooke, E. ; Avrithis, Y. ; O'Connor, N.: Fusing MPEG-7 visual descriptors for image classification. AceMedia IST FP6-001765,2005. http://www.acemedia.org/aceMedia/files/document/wp7/2004/ewimt04- dcuThom.pdf (November 6th, 07).

[39] Stauder, J.; Sirot J.; Le Borgne, H.; Cooke, E. ; O'Connor, N.E.: Relating Visual and Semantic Image Descriptors. AceMedia IST FP6-001765, 2004. http://www.acemedia.org/aceMedia/files/document/wp7/2004/ewimt04- dcuThom.pdf (November 6th, 07).

[40] Stuckenschmidt, H.: Toward Multi-viewpoint Reasoning with Owl Ontologies. In ESWC (2006), 259-272.

[41] Town, Christopher: Ontological inference for image and video analysis. Machine Vision and Applications 17(2), (2006), 94-115.

[42] Tsinaraki, C; Panagiotis, P.;Christodoulakis, S.: Interoperability Support for Ontology-Based Video Retrieval Applications. P.Enser et al. (Eds.), CIVR-04, LNCS 3115 (2004), 582-591.

[43] Vieira,V.; Tedesco, P.; Salgado,A.C.: Towards an Ontology for Context Representation in Groupware. H. Fucks, S. Luckosch, A.C. Salgado (Eds.) CRIWG-05, LNCS 3706 (2005), 367-375.

[44] Zhu, X.; Fan, J.;Elmagarmid, A.; Wu, X.: Hierarchical video content description and sumarization using unified semantic and visual similarity. Multimedia System 9 (2003), 31-53.