

Improving the Effectiveness of Multimedia Summarization of Judicial Debates through Ontological Query Expansion

E. Fersini and F. Sartori

DISCo, Università degli Studi di Milano-Bicocca,
Viale Sarca, 336 - 20126 Milano, Italy
{fersini,sartori}@disco.unimib.it

Abstract. The growing amount of multimedia data acquired during courtroom debates makes information and knowledge management in judicial domain a real challenge. In this paper we tackle the problem of summarizing this large amount of multimedia data in order to support fast navigation of the streams, efficient access to the information and effective representation of relevant contents needed during the judicial process. In particular, we propose an ontology enhanced multimedia summarization environment able to derive a synthetic representation of audio/video contents by a limited loss of meaningful information while overcoming the information overload problem.

1 Introduction and Motivation

Recent technological advances in multimedia production, storage and distribution have moved to digitize the traditional analogical archives creating large digital libraries. This increases the need of tools for content analysis, information extraction and retrieval of multimedia objects in their native form in order to make easier both the access and the management of multimedia archives. These issues are strongly emphasized into the judicial domain: digital videos represent a fundamental informative source of events that occur during courtroom proceedings that should be stored, organized and retrieved in short time and with low cost. To this purpose multimedia summarization techniques, which analyze several informative sources comprises into a multimedia document with the aim of extracting a meaningful abstract, can support end users for browsing the stream, efficiently accessing to important information and effectively representing relevant contents.

By analyzing the state of the art of multimedia summarization, the most valuable contributions are mainly concerned with three research directions: (1) internal techniques, which exploit low level features of audio, video and text; (2) external techniques, which refer to the information typically associated with a viewing activity and interaction with the user; (3) hybrid techniques, which combine internal and external information. These techniques can be either domain specific, i.e. focused on typical characteristics of a given domain, or non-domain

specific, that is focused on generic features associated with a generic context (no information a priori is known).

The state of the art on multimedia summarization comprises several valuable contributions belonging to the above mentioned categories. As far as is concerned with the internal techniques the main goal is to analyze low-level features derived from text, images and audio contents within a multimedia document. Interesting example can be found in [1,2,3]. In [1] the summarization of broadcast news is addressed by automatically transcribing the utterance of the speaker. In order to identify the most relevant phrases to be included into the summary, the prior and the generative probability of each sentence is combined for deriving its relevance. In [2] the scenes containing text in football videos are recognized using OCR techniques, for then a subsequent identification of key events through audio and video features. In [3] the identification of important video events is conceived through saliency models for the audio, video and textual information. In particular, audio saliency is assessed by quantifying waveform modulations, video saliency is measured through intensity, color and motion, while the text saliency is derived from part-of-speech tagging. At the end the saliency information are integrated in a single indicator for then deriving, through a bottom-up summarization algorithm, the final video skim.

In order to reduce the semantic gap between low level features and semantic concepts, research is moving towards the inclusion of external information that usually comprise knowledge about user-based information and the context in which a multimedia document evolves. The most recent investigations are focused on domain-specific [4] and non-domain specific contexts [5]. In particular, in [4] Fendri et al. present a video summarization approach able to generate a soccer video summary responding to the user profile. The summarization approach is based on semantic classified segments and textual information related to soccer video annotations. On the other side, in [5] Manzato et al. present an architecture that supports metadata extraction by exploring interaction mechanisms among users and contents. Relevant frames can be therefore distinguished by exploiting the annotation provided by users.

Although external summarisation techniques benefit by using high-level concepts during the summary generation, the minimization of the human effort is still a pending issue. An attempt that tries to combine the peculiarities of the previous approaches is represented by the hybrid techniques. Hybrid summarisation approaches combine the advantages provided by internal and external techniques by analyzing a combination of internal and external information. Some recent examples of hybrid technique can be found in [6] and [7]. In [6] Eldib et al. present a soccer video summarization engine, which is based on the detection of replay shots as interesting events. Video shots are identified through dominant color and histogram intersection methods. Replay shots are recognized by inducing a rule-based classifier able to exploit mid-level descriptors, as goal-mouth and score board. In [7] video and textual information, through image and text processing techniques, are used for creating a classification model able to detect the four entities “who, what, where and when” occurring in a digital

video. All the shots, classified with respect to the four entities, are then grouped into scenes according to a similarity measure.

As far as is concerned with the judicial debates, there is a very limited evidence about the multimedia summarization approaches [8] and in particular focused on semantic aspects. One of the main requirements of judicial users relates to the possibility of retrieving and viewing a meaningful representation of one (or more) debate(s) according to relevant events occurred during the celebration of the proceedings. This highlights two main requirements: (1) users entail to explicitly specify their needs through easy queries; (2) users want a system able to understand and augment their needs for then retrieving and presenting a limited set of relevant browseable contents. To this purpose we present a summarization environment, grounded both on machine learning and ontological query expansion techniques, for deriving a semantic-based storyboard of a courtroom proceedings recording.

The main outline of this paper is the following. In section 2 the key elements for deriving a semantic-based multimedia summary are presented. In section 3 an overview of the summarization process is given. In section 4 the core components, i.e. the ontological query expansion and the hierarchical clustering algorithm, are described. In section 5 a case study is presented, while in section 6 conclusions are briefly derived.

2 Multimedia Summarization: The Key Elements

In order to tackle the problem of generating a short and meaningful representation of a debate that is celebrated within a law courtroom, we propose a semantics-based multimedia summarization environment. The main objective is to generate a storyboard of a trial, by taking into account the semantics both beyond the user query and embedded into the courtroom recordings.

The main information sources exploited for creating a multimedia summary are represented by:

- *automatic speech transcriptions* that correspond to what is uttered by the actors involved into hearings/proceedings. Currently the speech transcriptions represent the main information source that can be semantically enriched for then supporting the multimedia summarization process. The automatic transcriptions are provided by the Automatic Speech Recognition (ASR) systems [9,10] trained on real judicial data coming from courtrooms. Since it is impossible to derive a deterministic formula able to create a link between the acoustic signal of an utterance and the related sequence of associated words, the ASR systems exploit a probabilistic formulations based on Hidden Markov Models [11]. In particular, a combination of two probabilistic models is used: an acoustic model able to represent phonetics, pronounce variability, time dynamics (co-utterance), and a language model able to represent the knowledge about the word sequences;
- *automatic audio annotations* coming from emotional states recognition. The affective state of the speakers represents a bit of knowledge embedded into

courtroom media streams that may be used to enrich the content available in judicial multimedia libraries. The emotional state annotations are derived through a framework based on a Multi-layer Support Vector Machine approach [12]. Given a set of sentences uttered by different speakers, a features extraction step is firstly performed in order to map the vocal signals into descriptive attributes (prosodic features, formant frequencies, energy, Mel Frequency Cepstral Coefficients, etc...). Then, these features are used to create a classification model able to infer emotional states of unlabelled speakers. In this work we modelled several affective states, among which *fear*, *neutral*, *anger*, *happy* and *boredom*, for introducing a first element of semantics associated to the “flat” automatic transcriptions;

- *automatic video annotations* that correspond to what happens during a debate. A further fundamental information source, for a semantic digital library into the trial management context, is concerned with the video stream. Recognizing relevant events that characterize judicial debates has great impact as well as emotional state identification. Relevant events happening during debates are “triggered” by meaningful gestures, which emphasize and anchor the words of witnesses, highlighting that a relevant concept has been explained. In order to achieve a reliable localization and tracking of people involved into the proceedings for then understanding “what is happening”, the automatic video annotations are derived through a combination of video processing algorithms. To analyze the motions taking place in a video, and to track gestures or head movements of given subjects (typically the witnesses), the optical flow is extracted as the moving points. Then active pixels are separated from the static ones using a kurtosis-based method and finally through a wavelet based approach extracting relevant features. At this stage the link between low level features and a given set of relevant actions is performed through the induction of Bayesian learner. The semantics extracted through the approaches described in [13] and [14] consist of changes of witness posture, new witnesses’ introduction, behavior of the actors, hand gestures and head movements;
- *user generated annotations* that correspond to free tags that judicial users assign to specific portions of the multimedia recordings of a given trial. Judicial users, i.e. judges, prosecutors and lawyers, usually tag manually some papers for highlighting and then remembering significant portions of the debate. This offers the possibility of personalizing contents according to the user preferences or working routines, providing then a better usability of multimedia contents. The possibility of digitally annotating relevant arguments discussed during a debate is particularly useful for future consultations and reasoning processes. For this reason, the user-generated annotations have been enclosed into the semantic-based summarization environment in order to organize and shortly represents the trials, according to the personal preferences.

All these information represent hidden knowledge that may be used to enrich the content available in multimedia digital libraries. The possibility for the end

user to consult the transcriptions, also by considering the associated semantics, represents an important achievement that allow them to retrieve an enriched written sentence instead of a flat one. This achievement radically changes the consultation process: sentences can assume different meanings according to the events that are occurring at that time.

3 Multimedia Summarization: Overview of The Process

In this section we present an overview of the summarization process by outlining the main steps that need to be performed for extracting a semantic storyboard of recorded judicial debates (see figure 1 for a high-level workflow). In order to perform the multimedia summarization activity, all the recorded proceedings need to be represented and stored into a multimedia database that can be consequently searched through a retrieval engine. To this purpose, we represent a trial by a set of *semantic clips* that include transcriptions, audio annotations, video annotations and user generated tags. Each clip is a consecutive portion of a debate (transcription) in which there is one active speaker, i.e. there exists a sequence of words uttered by the same speaker without breaking due to others, enriched with the annotations extracted by audio and video processing and/or manually provided by the end users. Indeed, a clip comprises a textual transcription for each speaker period with the corresponding semantic tags.

Having a multimedia database that comprises all the recorded trials by their semantic clip representation, the summarization process can start by presenting to the end user a query form for specifying a set of key-words to be used to retrieve some relevant portions of a given debate (see figure 2). The user query

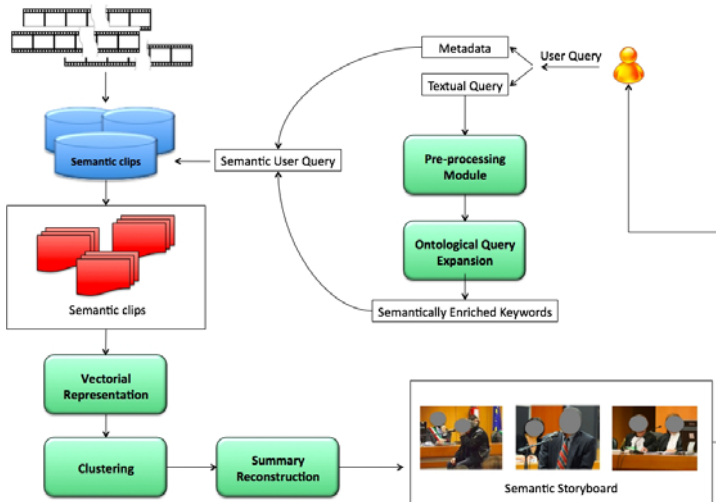
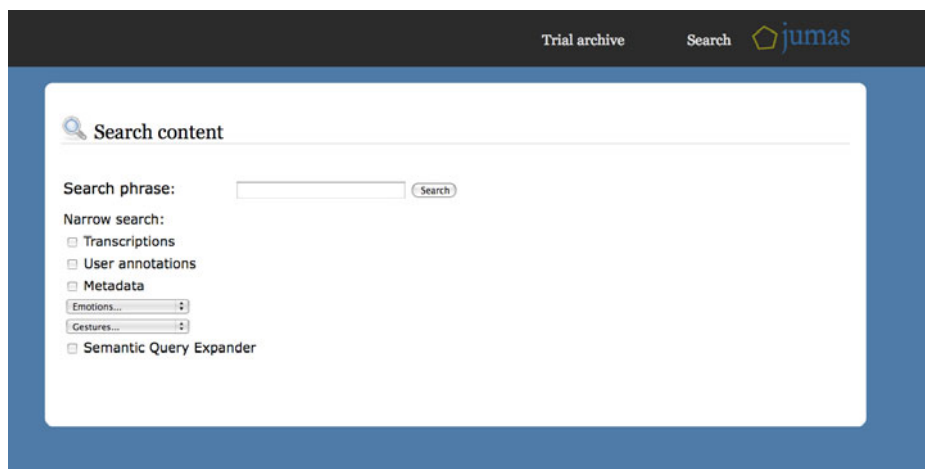


Fig. 1. Overview of the summarization process



The screenshot shows the JUMAS search interface. At the top, there is a dark blue header with the text "Trial archive" and "Search" next to the JUMAS logo. Below the header is a white search box with a magnifying glass icon and the text "Search content". Inside the search box, there is a "Search phrase:" label followed by a text input field and a "Search" button. Below the search phrase field, there is a "Narrow search:" section with several options: "Transcriptions", "User annotations", "Metadata", "Emotions...", "Gestures...", and "Semantic Query Expander". Each option has a checkbox next to it. The "Emotions..." and "Gestures..." options have dropdown arrows next to them.

Fig. 2. Query Form

can include both textual keywords and semantic annotations. In particular, the query may be composed by: (1) a *textual query* that will be used to match the transcription contents and user generated annotations, and (2) *metadata* that will be exploited to match the semantic tags provided by the automatic audio and video annotations.

Once the query has been specified by the end user, the textual part is submitted to the pre-processing module, which is aimed at reducing the size of vocabulary through stop words removal and stemming. Then, the reduced textual query is submitted to the ontological query expansion module for deriving a set of semantically enriched keywords. The augmented textual part, joint with the required semantic audio and video annotations, are provided to the search engine for retrieving those trials that match the required information.

Given the retrieved semantic clip, the next step is aimed at determining an appropriate representation of contents to be submitted to the clustering component that generates the browseable storyboard. To this purpose, we combine transcriptions, audio annotations, video annotations and user generated tags by using a vectorial representation of each retrieved semantic clip. A first set of coordinates are dedicated to the textual information, i.e. transcriptions and user tags, while a second set are appointed to the metadata, that is audio and video annotations. In particular the textual information are represented through TF-ISF coordinates, i.e. each term - belonging either to the speech transcription or to the user tags of a retrieved semantic clip - is represented through its TF-ISF weight [15]. On the other side, the metadata are represented through binary coordinates that denote the presence or absence of a specific audio/video annotation associated with the corresponding transcription or user tags.

As far as is concerned the TF-ISF representation, each term is weighted by considering its frequency into the semantic clip (TF) tuned by its inverse semantic clip frequency (ISF). Before computing the TF-ISF score of each term, the

vectorial representation is refined. The original query terms specified by the user, joint with the expanded terms derived through the ontological query expansion, are used to bring the semantic gap that is still present into the vectorial representation. The relationships between the textual query terms and the corresponding ontological augmentation, i.e. synonyms and related terms, are exploited for tuning the TF scoring. The extracted synonyms and related terms - which have been originated starting from the user query and at the same time occur along the textual information of a given semantic clip - are removed from the vectorial representation, while their TF are added to the corresponding terms.

Now, having a refined vectorial representation of the semantic clip, the multimedia summarization module may start the extraction of the storyboard. In particular the refined vectorial representation is submitted to a clustering component, based on the Induced Bisecting K-means [16], for creating a hierarchical organization of the semantic clips according to their similarity. The last step relates to the summarization reconstruction, where the final storyboard is derived from the dichotomic tree structure produced by the Induced Bisecting K-means algorithm. Given the dichotomic tree, a pruning step is performed in order to choose only those clusters of semantic clips that satisfy a given intra-cluster similarity requirements [17]. Finally, only the first frames related to the representative semantic clip of each cluster are presented to the end user as pictures that could be clicked to start the corresponding audio-video portion.

4 Multimedia Summarization: The Core Components

4.1 The Semantic Component: Ontological Query Expansion

In order to enhance the precision and recall of user query, as briefly introduced in the previous section, a Query Expansion platform has been enclosed in our multimedia summarization environment.

There are two main approaches to accomplish the Query Expansion task [18]: probabilistic and ontological. The first is based on statistic data that indicates the frequency of terms in a collection of supposed relevant documents, or the frequency of co-occurrence of terms. The second is based on knowledge models, in particular ontologies, in which the method search for the terms that have to be added to the query.

Although probabilistic query expansion is the dominant approach, it has shown limits in different fields and tasks: Carpineto et al. [19] highlight how this approach weights the terms on the basis of their relevance within the text rather than the real benefits for the user; moreover, it has been pointed out [20] [21] that probabilistic methods are very influenced by the corpus of documents and their relevance to the query.

On the contrary, ontological approaches are less developed and studied, but they virtually have a lot of undiscovered possibilities to semantically improve queries: being *corpus-independent*, they are more precise than probabilistic methods in the text disambiguation within a given domain. Moreover, they are partic-

ularly suitable to treat short queries. Anyway, as shown in [22] and [23], ontological methods have some important drawbacks: ontologies are typically difficult to create and maintain in order to guarantee the necessary level of precision to avoid the decrease of performance. The so called *query drift* (i.e. the choice of an expansion direction that is out of user scope) phenomenon is more probable with ontological approaches than probabilistic ones.

The query expansion module is a web service, realized with the JAX-WS 2 (Java API for XML Web Services) technology, which is part of the Java EE 5 platform. The chosen semantic framework is Jena, an open source Java framework for building Semantic Web applications grown out of work with the HP Semantic Web.

Currently, the adopted ontology is a *thesaurus* of legal terms, written both in *Italian* and in *Polish*: thesauri are structured vocabularies where relationships among terms are specified. They can be considered as a subclass of ontologies, since they represent only a part of the knowledge involved and their power of expression is limited [24]. Anyway, their main features are compatible with most of the ontological methods for QE available at the moment.

The ontological query expansion platform has been designed and implemented to support the execution of the hierarchical clustering algorithm for the retrieval of audio/video data in the judicial domain. This algorithm is the subject of the next section: further details about the design and implementation of the QE platform are out of paper scope and can be found in [25][26].

4.2 The Computational Component: Hierarchical Clustering

The approaches proposed in the literature for hierarchical clustering are characterized by good performance in terms of accuracy, but are also affected by high computational complexity. In order to take advantage of hierarchical approaches, but at the same time to reduce their time complexity, we introduced into the summarization system a novel approach based on Induced Bisecting k-Means [16]. It starts with a single cluster of multimedia clips and works in the following way:

The distance a_{ij} is computed as an average linear combination of a cosine-based distance $1 - \cos_T(m_i, m_j)$, for the TF-ISF coordinates concerned with transcriptions and user tags, and the jaccard distance $Jdist_B(m_i, m_j)$, for the binary coordinates concerned with audio/video annotations. The parameter α tunes the importance of each component, TF-ISF vs binary, during the distance computation. The same approach has been exploited for the similarity metric: it is an average linear combination of the cosine similarity $\cos_T(m_i, m_j)$, for TF-ISF coordinates, and the jaccard similarity $Jsim_B(m_i, m_j)$, for the binary coordinates. Also in this case, the parameter α tunes the importance of each component, TF-ISF vs binary, during the similarity computation.

At the end of the summarization process, a dichotomic tree of semantic clips is generated. In order to derive the final storyboard we need to collapse the tree into

Algorithm 1. Induced Bisecting K-Means

- 1: Set the Intra Cluster Similarity (ICS) threshold parameter τ
- 2: Build a distance¹ matrix A whose elements a_{ij} represents distance between two semantic clips m_i and m_j , computed as

$$a_{ij} = \frac{\alpha \times (1 - \cos_T(m_i, m_j)) + (1 - \alpha) \times Jdist_B(m_i, m_j)}{2} \quad (1)$$

- 3: Select, as centroids, the two semantic clips i and j s.t. $a_{ij} = \max_{l,m} A_{lm}$
- 4: Find 2 sub-clusters S_1 and S_2 using the basic k-Means algorithm²
- 5: Check the ICS of S_1 and S_2 as²

$$ICS(S_k) = \frac{1}{|S_k|^2} \sum_{m_x, m_y \in S_k} \frac{\alpha \times (1 - \cos_T(m_x, m_y)) + (1 - \alpha) \times Jdist_B(m_x, m_y)}{2} \quad (2)$$

- 6: If the ICS value of a cluster is smaller than τ , then reapply the divisive process to this set, starting from step 3
 - 7: If the ICS value of a cluster is over a given threshold, then stop. 6. The entire process will finish when there are no sub-clusters to divide.
-

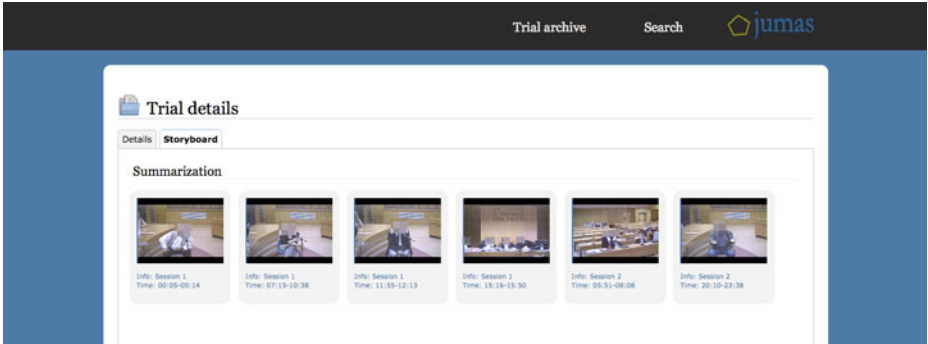


Fig. 3. Storyboard of a sample judicial debate

“flat” clusters. To this purpose the dichotomic tree is visited in order to find those nodes (clusters) that respect a given intra-cluster similarity constraints, allowing us to select the most homogeneous groups of semantic clips. Now, suppose that the pruning activity after the Induced Bisecting K-means returns $C1$, $C2$ and $C3$ as the resulting clusters containing the semantic clips $1, \dots, 9$. The presentation activity considers the most representative clip of each cluster (centroid) as the relevant element to be show into the summary, providing then a browseable storyboard as depicted in figure 3.

5 An Explanatory Judicial Case Study

5.1 Case Study

In order to understand the entire summarization process with respect to the judicial domain, we present a case study able to explain the impact that the ontological query expansion may have on real judicial multimedia summarization. As outlined in section 3, the entire process starts when the end user specifies a set of keywords for retrieving all the relevant debates related to his/her query. Now suppose that a judge is interested in viewing, through an interface that makes available the relevant browseable clips, those debates that are concerned with “white weapon”.

In a traditional summarization system, where no ontological query augmentation is available, only those debates with clips that merely contain the terms “white weapon” are retrieved for the subsequent steps. Of course, this reduces the ability of the system to include a larger set of debates that do not explicitly contains the term “white weapon”, but are stricly related to the corresponding concept. This means that the retrieval engine, and consequently the storyboard construction, will not take in consideration all the debates related to “knife”, “blade” or “stab” (see Fig. 4).

In the proposed system, the multimedia summarization process has a different outcome. When the user provides the query “white weapon” the ontological query expansion component accomplishes its enhancement task by enriching the query term with the additional related concepts among which “knife”, “blade” and “stab”. After the retrieval phase, which is driven by the semantically enriched query terms, a greater set of debates than the ordinary may be available for the subsequent steps. Now, consider the representation step aimed at creating a suitable vectorial representation of the retrieved semantic clips. At this stage we can exploit the knowledge about the relationships that exist among “white weapon”, “knife”, “blade” and “stab” for tuning the TF-ISF representation. While in a fool-

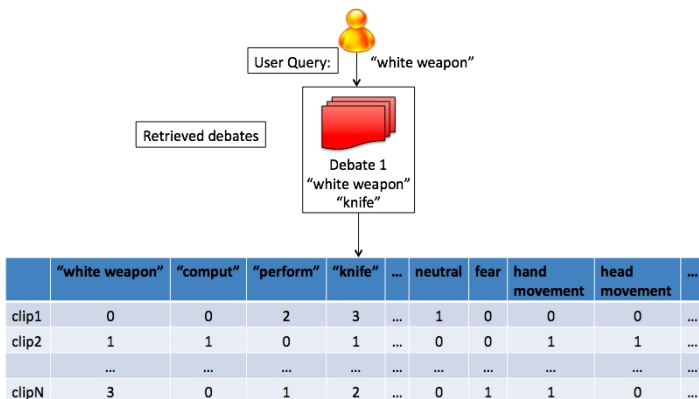


Fig. 4. Storyboard construction without QE

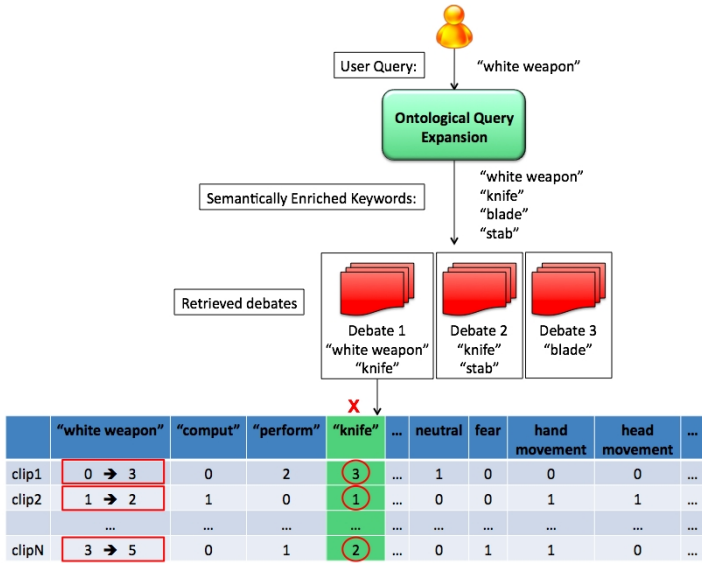


Fig. 5. Storyboard construction with QE

ish system the TF-ISF scoring is computed through a simple frequentist approach, where “white weapon”, “knife”, “blade” and “stab” belonging to a given semantic clip are considered as independent, in the proposed system their relationships strongly influence the TF-ISF representation. In particular, as highlighted in Fig. 5, the vectorial representation of the semantic clips is revised according to the knowledge about the augmented terms: the terms “knife”, “blade” and “stab” of a given semantic clip are removed and their TF contribute to improve the TF of the “white weapon” term originally stated by the user. This improvement of TF, due to the semantic relationships among terms, implies an increasing relative importance that will have impact on the subsequent clustering phase.

5.2 Experimental Investigation

The semantic-based multimedia summarization environment has been validated by using real trials recorded at the Court of Naples. In particular, the validation was focused on 4 trials:

- Trial 1, composed of one session of 15 minutes
- Trial 2, composed by one session of 35 minutes
- Trial 3, composed by two sessions: Session One, with a 59 minutes duration, and Session Two, with a 3h 33min duration.
- Trial 4, composed of two sessions: Session One with a duration of 17 minutes and Session Two with a duration of 1h 15min.

For each trial, two different types of summarization have been constructed:

- Base summary: a basic summary with no ontological query expansion.

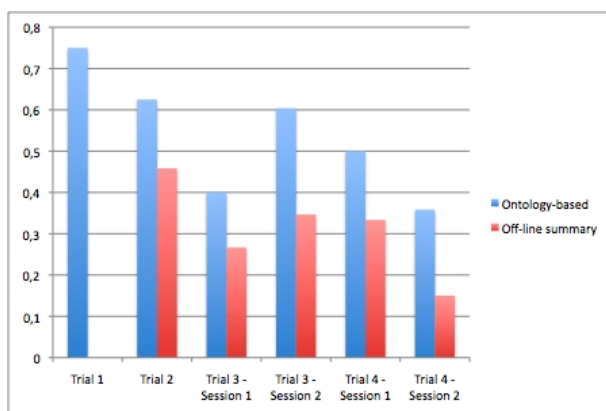


Fig. 6. Experimental Results

- Ontology summary: a summary created using, together with ASR transcriptions and audio/video annotations, a query for which the Query Expansion module has been invoked.

In particular, the following queries have been exploited:

- Trial 1: “coroner” and “defendant”
- Trial 2: “verdict” and “slaughter”
- Trial 3: “sequestration” and “preliminary investigation” for Session One; “ambush” and “gun” for Session Two.
- Trial 4: “homicide” for Session One; “homicide” and “eavesdropping” for Session Two.

Since the summarization process could be strongly affected by the number and the quality of the expanded query, only related terms with a relationship weight greater than 0.7 have been considered. A user opinion investigation has been performed on the produced summary. Each trial has been submitted to five users in order to discriminate relevant clips from the non-relevant ones. In particular, the whole collections of clips comprised in the considered trials have been evaluated and labelled with relevant and non-relevant labels. Starting from the results of the users opinions the accuracy of the derived summaries has been computed. A synthetic representation of the user perception of the generated storyboard is depicted in figure 6. The results show that the ontology-driven multimedia summarization approach outperforms the traditional approach. Interesting future comparisons could be related to Latent Semantic Analysis approaches, where the semantic gap that could exist among terms detected and managed by the Singular Value Decomposition technique.

6 Conclusions

In this paper a multimedia summarization environment based on semantic query expansion has been presented in order to allow judicial actors to browse

multimedia documents related to judicial proceedings. In particular, by exploiting ontological query expansion and the representation of semantic clips of debates, the proposed system is able to derive a compressed and meaningful representation of what happens into a law courtroom. Our work is now focused on creating a testing environment for a quality assessment of the storyboard in order to verify the concrete applicability of the approach in Italian and Polish courtrooms, that is the final aim of the project.

Acknowledgment

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

References

1. Chen, Y., Chen, B., Wang, H.: A Probabilistic Generative Framework for Extractive Broadcast News Speech Summarization. *IEEE Transactions on Audio, Speech & Language Processing* 17(1), 95–106 (2009)
2. Jiang, R.M., Sadka, A.H., Crookes, D.: Hierarchical Video Summarization in Reference Subspace. *IEEE Transactions on Consumer Electronics* 55(3), 1551–1557 (2009)
3. Evangelopoulos, G., Zlatintsi, A., Skoumas, G., Rapantzikos, K., Potamianos, A., Maragos, P., Avrithis, Y.: Video event detection and summarization using audio, visual and text saliency. In: *Proc. of the 2009 IEEE International Conference on Acoustics Speech and Signal Processing* (2009)
4. Fendri, E., Ben-Abdallah, H., Hamadou, A.B.: A Novel Approach for Soccer Video Summarization. In: *Proc. of the Second International Conference on Multimedia and Information Technology* (2010)
5. Manzato, M.G., Coimbra, D.B., Goularte, R.: Multimedia Content Personalization Based on Peer-Level Annotation. In: *Proc. of the 7th European Conference on European Interactive Television Conference* (2009)
6. Eldib, M.Y., Zaid, B., Zawbaa, H.M., El-Zahar, M., El-Saban, M.: Soccer video summarization using enhanced logo detection. In: *Proc. of the 16th IEEE International Conference on Image Processing* (2009)
7. Bo-Wei, C., Jia-Ching, W., Jhing-Fa, W.: A Novel Video Summarization Based on Mining the Story-Structure and Semantic Relations Among Concept Entities. *IEEE Transactions on Multimedia* 11(2), 295–312 (2009)
8. Fersini, E., Messina, E., Archetti, F.: Multimedia Summarization in Law Courts: A Clustering-based Environment for Browsing and Consulting Judicial Folders. In: *Proc. of the 10th Industrial Conference on Data Mining* (2010)
9. Löff, J., Gollan, C., Ney, H.: Cross-language Bootstrapping for Unsupervised Acoustic Model Training: Rapid Development of a Polish Speech Recognition System. *Interspeech*, 88–91 (2009)
10. Falavigna, D., Giuliani, D., Gretter, R., Löff, J., Gollan, C., Schlüter, R., Ney, H.: Automatic Transcription of Courtroom Recordings in the JUMAS project. In: *Proc. of the 2nd International Conference on ICT Solutions for Justice* (2009)
11. Rabiner, L.R.: A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE* 77(2), 257–286 (1989)

12. Fersini, E., Messina, E., Arosio, G., Archetti, F.: Audio-based Emotion Recognition in Judicial Domain: A Multilayer Support Vector Machines Approach. In: Proceedings of the 6th International Conference on Machine Learning and Data Mining in Pattern Recognition, pp. 594–602 (2009)
13. Avgerinakis, K., Briassouli, A., Kompatsiaris, I.: Video processing for judicial applications. In: Proc. of the 2nd International Conference on ICT Solutions for Justice (2009)
14. Kovács, L., Utasi, A., Szirányi, T.: Visret - a content based annotation, retrieval and visualization toolchain. In: Proc. of the 11th International Conference on Advanced Concepts for Intelligent Vision Systems (2009)
15. Larocca Neto, J., Santos, A.D., Kaestner, C.A.A., Freitas, A.A.: Document clustering and text summarization. In: Proc. 4th Int. Conf. Practical Applications of Knowledge Discovery and Data Mining, p. 4155 (2000)
16. Archetti, F., Fersini, E., Campanelli, P., Messina, E.: A Hierarchical Document Clustering Environment Based on the Induced Bisecting k-Means. In: Proc. of the Int. Conf. on Flexible Query Answering Systems (2006)
17. Kashyap, V., Ramakrishnan, C., Thomas, C., Bassu, D., Rindfleisch, T.C., Sheth, A.: Taxaminer: An experiment framework for automated taxonomy bootstrapping. *International Journal of Web and Grid Services* 1(2), 240–266 (2005)
18. Tuominen, J., Kauppinen, T., Viljanen, K., Hyvnen, E.: Ontology-based query expansion widget for information retrieval. In: Proceedings of the 5th Workshop on Scripting and Development for the Semantic Web (SFSW 2009), 6th European Semantic Web Conference (ESWC 2009) (May 31–June 4) (2009)
19. Carpineto, C., de Mori, R., Romano, G., Bigi, B.: An information-theoretic approach to automatic query expansion. *ACM Trans. Inf. Syst.* 19(1), 1–27 (2001)
20. Voorhees, E.M.: Query expansion using lexical-semantic relations. In: SIGIR 1994: Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 61–69. Springer-Verlag, New York Inc. (1994)
21. Ruthven, I., Tombros, A., Jose, J.M.: A study on the use of summaries and summary-based query expansion for a question-answering task (2001)
22. Navigli, R., Velardi, P.: An analysis of ontology-based query expansion strategies. In: Proc. of Workshop on Adaptive Text Extraction and Mining (ATEM 2003), in the 14th European Conference on Machine Learning (ECML 2003), September 22–26, pp. 42–49 (2003)
23. Durão, F.A., Vanderlei, T.A., Almeida, E.S., de L. Meira, S.R.: Applying a semantic layer in a source code search tool. In: SAC 2008: Proceedings of the 2008 ACM Symposium on Applied Computing, pp. 1151–1157. ACM, New York (2008)
24. Wielinga, B.J., Schreiber, A.T., Wielemaker, J., Sandberg, J.A.C.: From thesaurus to ontology. In: K-CAP 2001: Proceedings of the 1st International Conference on Knowledge Capture, pp. 194–201. ACM, New York (2001)
25. Sartori, F.: A Comparison of Methods and Techniques for Ontological Query Expansion. *Communications in Computer and Information Science*, Vol. 46, pp. 203–214 (2009)
26. Sartori, F., Palmonari, M.: Ontological Query Expansion for the Legal Domain: a Case Study from the Jumas Project. *LNBIP*, vol. 62 pp.107–122 (2010)