MYSTORYPLAYER: SEMANTIC AUDIO VISUAL ANNOTATION AND NAVIGATION TOOL

Pierfrancesco Bellini, Paolo Nesi, Marco Serena
Distributed Systems and Internet Technology Lab, Department of Systems and Informatics
University of Florence, Via S.Marta 3, Florence, 50139 Italy, tel +39-055-4796523
nesi@dsi.unifi.it , http://www.disit.dsi.unifi.it

ABSTRACT

Most of the solutions for accessing semantically annotated media provide clients for interpreting simplified information and not to directly executing the semantic annotations. Thus, the interactivity in the navigation among the semantic concepts is closed on the server side, leaving at the client/player tool a passive role. In this paper, a uniform solution (called MyStoryPlayer) that allows modeling and playing non-linear stories by following their semantic including temporal and logical relationships, is proposed. In MyStoryPlayer, any media segment can be an annotation for another media element, plus any semantic description. The user may navigate in the audiovisual annotations and media creating its own non-linear experience/path. The resulting solution includes a uniform semantic model, a corresponding semantic database for the knowledge, a distribution server for semantic knowledge and media, and the MyStoryPlayer to be used in web applications

Index Terms— audio visual annotation, semantic model, semantic player.

1. INTRODUCTION

The content models for entertainment and edutainment are suffering a range of transformations in the recent years. Multistream TV programs such as granprix, bigbrother, etc., are proposing synchronized multiples views/streams of the same event, on which the user may decide to switch. Despite of this effort, the TV is still less attractive than Internet since the possibilities and interactivity are limited (even adopting innovative streaming solution MPEG-21 DIS, digital item streaming [Burnet et al., 2005], or by the many projects on multimedia and multicontext streaming such as LIVE [http://www.ist-live.org]). The most famous TV serials such as Lost, Flashforward, and also the more dated Odissey 5, Doctor Who, etc., may find a more attractive modality of fruition on more interactive channels for fruition. Internet fruition is mainly non-linear. To this end, new models have been proposed in the context of IPTV and/or WebTV. They may take advantage of the possibility of providing content and stream on demand and by the possible direct connection with portals of content distributors, via the so-called hypervideo and/or multimedia links from videos.

The internet capability of rapid context change and the virtually unlimited number of combinations/paths are probably the most relevant motivations for the fast penetration of Web solutions for entertainment, despite the present limitations in network bandwidth. In Internet, on web browsers or via dedicated applications, a large range of entertainment and edutainment solutions are growing in which the user may see at the same time: video and slides, video and chat, etc. Moreover, the adoption of semantic web technologies on internet are transforming the information search and access. Some of the most accessed web portals are starting using semantic web technologies; by indexing content on the basis of their semantic description, and allowing users to provide annotations and classifying content and contributions. See for example users' annotations on images of Flick, and the annotations on videos of YouTube. In the latter case, an annotation may allow to select the successive scene, passing to different videos, modeling the paradigm of Hypervideo (see MIT Hypersoap [Dakss et al., 1998], etc.). Simple paths may be created starting a video from an overlapped link placed on another. Thus, these solutions give at the user the possibility of creating a unique personal experience in the content fruition. In most cases, queries are still traditionally based on keywords, and in some cases full text indexing via some fuzzy support. Thus, the insertion of hyperlinks on videos as annotations does not correspond to enabling semantic queries. That means for example, that it is not possible to perform semantic queries such as requesting *"all videos where two men are talking each other in a car"*. To this end, semantic descriptors of the scene have to be indexed to enable semantics queries, for example in SPARQL. Moreover, for huge collections the computational complexities for semantic extraction and about the inferential processing on semantic indexing related to queries may be huge. Complexity may also depend on the semantic model and functionalities offered to the final users. Therefore, a large effort for content service improvement has been focused to the extraction and association of semantic descriptors with content.

Semantic descriptors are typically modeled as media or multimedia annotations formalized in MPEG-7 and/or RDF [MPEG-7], annotations can be also in MPEG-21 [Bellini 2011]. **Vannotea** solution has been proposed for collaborative annotation of videos [Kosovic et al. 2004].

Vannotea allows the collaborative discussion on video content in real-time. While, the execution model is linear along the video time line. The annotations database has been developed by exploiting the Annotea model and solution W3C [Koivunen et al., 2003].
In this context, critical points are: (i) the detailed modeling of the relationships among media especially along the temporal line, and thus of the ontological model adopted, (ii) the interpretation of the annotations in terms of ontological model with time, and (iii) the effort performed by the server in searching and providing this information to the rendering and exploiting client tools. In most cases, there exists a semantic gap; the server is capable to store and keep knowledge model in terms of ontologies and relationships, while the players receive simple knowledge about the semantic context, which is limited to information strictly needed to the rendering of connected media. In all cases, the capabilities of navigating into the knowledge, and thus the complexity are located on the server side. Thus, the interactivity in the navigation among the semantic concepts is confined on the server side, leaving at the client/player tool a passive role.

In this paper, a uniform solution (called MyStoryPlayer) that allows to model and play annotations and thus non-linear stories by following their internal temporal and logical relationships is presented. The idea is to model audiovisual annotations in terms of RDF including classical descriptors and temporal relationships, for media. The MyStoryPlayer may continuously obtain the successive descriptors and relationships when the user changes the context, for example, selecting media annotations. The final effect consists in allowing the user to create their own experience in navigation by executing the semantic relationships among media and annotations. The solution has been studied and developed as a generalization of the models that may be used to describe and annotate non-linear stories such as Lost, Flashforward, Odissey 5, and classical learning content in histories and humanities in which several different time relationships may be defined. With MyStoryPlayer and tools, the final user may start by making a semantic query to the MyStory Server to obtain as a result a set of possible entry points. During this experience, the user may decide to navigate interacting with them, thus, changing the context in terms of time and related annotations. This allows the user to build a personal experience on the related story.
The paper is organized as follows. In Section 2, the conceptual semantic model of MyStory is presented. Section 3 reports the general architecture, giving details about the server and clients. In Section 4, the internal details of the MyStoryPlayer client tools are presented. An example, which can be accessed via web, is presented on Section 5. Conclusions are drawn in section 6.

2. CONCEPTUAL AND SEMANTIC MODEL

The semantics of relationships among media essences and annotations can be regarded as simple descriptors or effective temporal associations to synchronize them. The annotations may be capable to create/formalize relationships from them and the annotated resources. Moreover, the annotations may consist in semantic descriptors of the content. For example a description of the scene, the description of the image (e.g., the scene is in the forest, or the forest is represented in the scene, Al and Jack are talking each other, Jack has a gun). Thus, each single element of the scene may have its corresponding descriptors as annotations.
The main idea of MyStoryPlayer is to put in the hands of the final users a tool for navigating in the annotations including temporal relationships and with a full access to the single media segments, which may refer to independent media segments of: images, video, and audio essences with associated descriptors. Each of them may be executed according to their semantics and nature. *Thus the execution of a certain scenarios with all its related annotations, leads to the execution of multiple synchronized media.*
In MyStoryPlayer navigation, several annotations and thus audiovisual elements may be executed at the same time. These audio/visual annotations can provide additional and more accurate information, may:

- explode a single time instant with a more complex scenario,
- bring back the story to reminded past events, or to possible futures,
- show a different point of view, or what happen at the same time in another place,
- provide the comment of the director,
- show a video with the scene without visual effects,
- show the historical scenarios,
- show the advertising details of the car which the present in the scene, etc.

In MyStoryPlayer solution, the main aim is on defining a general model to cope with a set of annotations which can be used to navigate among audiovisual scenes according to different possible paths – e.g., the aim of director, the linear time, the activities of a given actor, the movements of an object. The MyStoryPlayer may be used to create new content experiences and models for accessing media in which one may access to a story in a given point and from that may navigate to an indefinite network of annotations.
In MyStoryPlayer, media are synchronized via annotations and segments, and may be executed by the user at the same time, giving the opportunity to the user to jump on a different context by following those annotations, that in this manner become the main stream. The user may pass from one video/essence to another and to return back in the stack

of events. The single resources may not be simply connected as temporally aligned videos such as the views of the different cameras in granprix, multiview football match, big brother multiview, etc. This means that the switch from one essence to another may change the context and thus some of the other essences in execution jumping to a different execution time and annotation domain.

The proposed MyStoryPlayer model and tool exploits a semantic model which allows a full RDF navigation among temporal models and audiovisual descriptors and reduces the gap between annotated media essences and the annotations, since in MyStoryPlayer: both of them belong to the same pool of audiovisual.

The main elements of the ontological model of MyStoryPlayer knowledge are reported in Figure 1. The figure does not include the relationships with basic elements such as: URLs, strings, date, time, integer, floats, etc. According to Figure 1, the MyStoryPlayer model is mainly focused on the concept of Annotation. Each Annotation includes a set of descriptions and a media references.

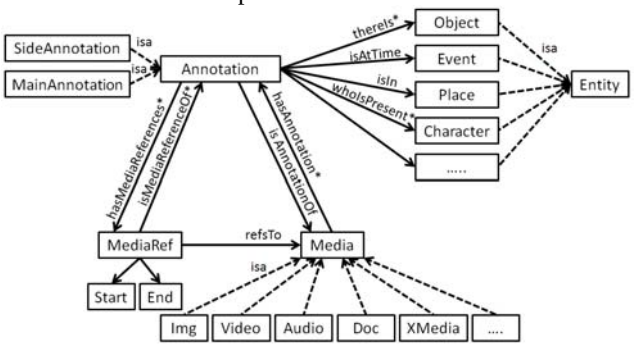

Figure 1. Simplified MyStoryPlayer semantic model.

The scene Annotation descriptor may include multiple details about Characters, Objects in the scene. These kinds of descriptions are useful to perform semantic queries of the scenes. Moreover, Places and Event are singular features. The scene representation model can be augmented with additional ontologies; thanks to the RDF flexibility other features can be added to the model in order be more accurate, complete, suitable for any specific domain and contexts. Any media essence may have multiple annotations; each of them with its own description and reference to other media elements. Among the description also a textual annotation, a starting and ending date/time instant.

The MyStoryPlayer annotations have a starting time, and an ending time. Annotations can be SideAnnotation or MainAnnotation. The two kinds of semantic annotations may be composed to create more complex cases. And the same media segment may be used as SideAnnotation and/or MainAnnotation in different contexts.

The **SideAnnotations** are associated with the main audiovisual by means of a starting time instant, and from that time instant they are played synchronously with the annotated audiovisual (but aside). They start at a given time instant and have a time duration, for that duration the two audiovisual are synchronously played: the main audiovisual in the main area and the annotation aside (the annotation may have to show some associated descriptions as text or additional information). The SideAnnotations are typically adopted for showing audiovisual and/or document/text synchronously such as slides aside the teacher video, different views of the same theatrical scene or sport event.

The **MainAnnotations** are associated with the main audiovisual at a given starting time. From that time instant the annotation is “explosive”. This means that, the audiovisual annotation go forward putting in pause the main audiovisual for the duration of the annotation itself. After that, the context returns on the annotated main audiovisual. The return back is automatic if the MainAnnotation is executed up to its conclusion, while the user may decide to change the context during the execution of the MainAnnotation as well. MainAnnotations may be adopted to provide explanations, to attract the interest of the user on different scene, etc.

From the temporal point of view, the time instants of the annotation (starting and ending) are distinct concepts with respect to the Event concept associated with the Annotation. An important feature of scene description is the property *isAtTime* that links the Annotation class with Event, which gives (when possible) at the scene described a temporal collocation, and enables a semantic separation between the real temporal line. Indeed, there are many cases in which the moviemaker decides an order to represent the movie, which does not necessarily follow the occurrence in the real world. Therefore, with this feature, we can separate two parallel temporal lines, the one referred to video, and the other referring to the event ordering into the narrative, to allow the user make temporal queries in which he can ask: “all the scenes before/after event X”. The Event concept is related to the narrative semantics of the scene, and represents a symbolic or a time event in the story (a symbolic event is represent by a nickname, for example the “the second goal”; a time event is something which may be defined as DD:MM:YYYY, HH:MM:SS). This means that a clear distinction from time into the story and the time relationships among the digital essences and annotations has been made. For example, it is possible to have set of media and annotations representing a possible experience related to a football game; some videos may include a linear view of the game while others may represent parallel point of views, comments of experts, jump back or forward to other events, detailed facts occurred during the advertising, etc. So that, a combination of Side and Main Annotation is needed to model all kinds of annotations. And, they can be mounted on a fewer number of video files, even in a single file.

Moreover, MyStoryPlayer annotations may be nested, and may be used to organize/create non-linearly and recursive

complex paths. An additional flexibility to the model is offered by the fact that each single media file may be referred by any number of annotations, overlapped each other or not.

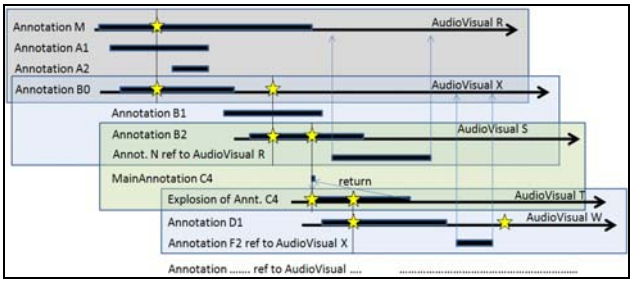

Figure 2. An example described in the next steps:

1. user started with the context of AudioVisual R (for example Video R.flv) on Annotation M (a video segment with association a set of scene descriptors and a text). During the execution of AudioVisual R also Annotation A1, A2 and B0 are executed synchronously according to the time schedule.
2. at the first yellow star (time instant marked with a star), the user decided to change the context by following the content played by Annotation B0. Thus, AudioVisual X was loaded together with its annotations: B1, B2 and Annotation N on AudioVisual R.
3. user completed Annotation B0 and continued on AudioVisual X for a while. Then, the user decided to change context passing at Annotation B2, with the corresponding change to AudioVisual S. AudioVisual S has MainAnnotation C4 (an explosive annotation) aligned at the second star. Thus, the execution passes on Exploding Annotation C4, changing the context on AudioVisual T, to returning back to AudioVisual S. The context of AudioVisual T has annotations D1 and F2. Please note that also Annotation F2 of AudioVisual T may lead again to move the context on AudioVisual X.
4. user, in the middle of executing Annotation C4, decided to pass at Annotation D1 changing again the content to go on AudioVisual W.

The example in Figure 2 represents a case in which the user starts in a certain context and changes the context (audiovisual execution) by following a number of annotations. Thus, these annotations have to be loaded from the knowledge repository by sending a Semantic Navigation Query. The advantages of MyStoryPlayer reside into the semantic model and player tools which include: the management of multiple annotations, the execution of multiple synchronous annotations with time relationships, the possibility of returning back on the stack of context changes including annotations and execution time, in the recording of the experience performed, and in the possibility of having and cumulating a portion of the knowledge on the client side to perform internal navigation without reloading in the RDF model. The possibility of receiving the RDF model of the annotations allows at the MyStoryPlayer to implement strategies on the loading and play of the successive audiovisual/media streams. In fact, it is possible to load in advance the next possible RDF descriptors without waiting for the change of context. This may accelerate the change of context and to perform a partial semantic reasoning on the client side.

3. THE GENERAL ARCHITECTURE

According to the above-presented scenarios, the user starts accessing to the MyStory knowledge by performing a semantic query on a web page. As a result, he/she receives back a list of possible annotations representing scene descriptors from which the navigation may start. Once an annotation/scene is selected, the user experience starts; for example, from video Lost-Season-3-Episode-12 at the time instant 23:34:12. The chosen annotation is put in execution with its corresponding digital media essence, thus the MyStoryPlayer automatically shows a set of synchronized annotations and thus related essences along the execution time line. Annotations are executed aside the main essence according to the time line, and the user may decide to start navigating among them. Thus creating his experience among the possible paths which are modeled by the annotations. The set of annotations are stored into the ontological knowledge database, managed by SESAME, see Figure 3. The MyStory WebPortal front-end is only a web server providing a set of web pages for collecting information, collecting queries, providing support for the client MyStoryPlayer into the internet Browser. The MyStoryPlayer is a player specifically designed and developed for executing the semantic model above presented. It has been developed in Adobe ActionScript of flash, so that it is automatically provided and loaded during the web page loading as flash player tool.

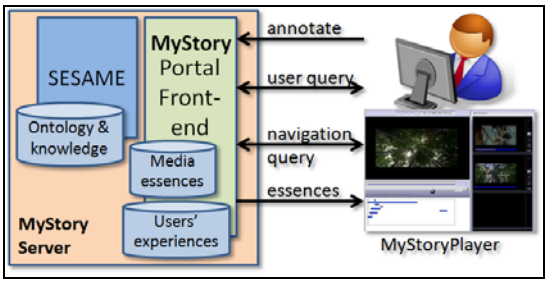

Figure. 3. General architecture of MyStoryPlayer.

The solution proposed works as a client/server application. The solution may be accessed via web browser and may provide more complex experiences including new annotations provided by other users, Http://www.mystoryplayer.org . On the other hand, the video streaming or progressive download is quite expensive for the network and thus having a number of them is presently possible only exploiting low resolutions video or very powerful connections.

The main functionalities are reported in the architecture of Figure 3. Among them the:

- **addition of annotations** to the knowledge database. The annotations are added providing a set of information such as: ID of the digital resources already in or provided file, starting point, ending point, model of play (sync or explosive), classification metadata, text description, etc.
- **SPARQL queries** on the knowledge database to get a list of possible entry point scenarios.
- **Semantic Navigation Queries** in SPARQL on the knowledge database, to acquire updated information every time the context is changed. For example, when the user click on a video/image/audio on the right side, the MyStoryPlayer poses a query to the server to get all the triples connected to the next context, to explode it in terms of related audio/visual, and thus to start the progressive download and player of the related resources.
- **Saving users' experiences on the server.** When a users takes a decision on recording the experience in navigating. This implies that for action the player has to communicate to the server the time instant at which the action has been performed on the timeline of the current media resources. This allows keeping trace of the whole user experience, and may be used to share this experience with other users. The sharing of experience is of interest for educational and entertainment purposes.

Potentially the user may navigate in any direction among the annotations contained into the knowledge database, following the links and audio/video/images which are presented along the experience and navigation. The recorded users' experiences may be re-proposed in a second moment to the same user or to other users that may be interested to make the same experience.

4. INSIDE MYSTORYPLAYER

One of the most interesting tools of the proposed solutions is the MyStory client tool, namely the MyStoryPlayer. In order to satisfy the above-described main features, the MyStoryPlayer has been designed to perform:

- interpretation of RDF information and query results coming from the MyStory Server,
- understanding and recognizing the different kinds of annotations,
- interpreting media references and thus relationships among elements,
- supporting Main and Side Annotation semantics,
- execution/interpretation of multiple annotations and thus of multiple audiovisual streams related to main media essences and Annotations. The streams are received as progressive download flash video streams;
- change of context when the user click on an Annotation (video, audio, image), implies to dynamically perform the access to the media of the next Annotation on WebPortal of MyStory.
- visualization of information and text associated with the current main audiovisual and of the Annotations; Textual descriptors to the single annotation and resources can be used to make in evidence specific aspects and to start discussions.

The MyStoryPlayer client tool has been designed by using object-oriented technology, see Figure 4. It has been designed to be loaded into a large range of Internet Browsers and thus in a range of web based solutions and applications. According to the design, the player is capable to receive annotations in RDF, where they are internally interpreted according to their semantic: audio, video, image, etc.

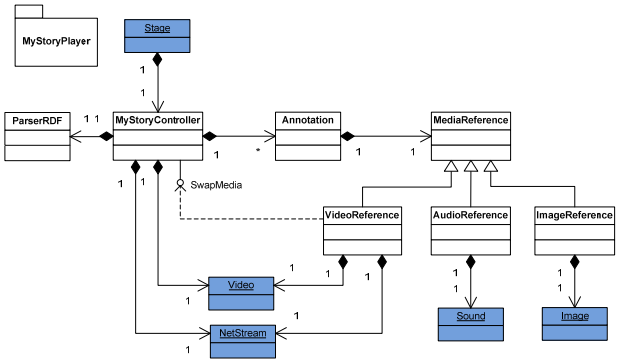

Figure 4. The main classes of MyStoryPlayer client tool and their relationships

5. USING MYSTORYPLAYER

According to the semantic model adopted, the user may perform semantic queries in SPARQL. For example, we have coded and annotated a subset of Lost TV serial to show the mechanism, and the user may perform the following SPARQL queries to get all scenes (the authors are requesting the rights to publishing the example proposed by using Lost serial, while other examples can be used as well): in which John and Jack are in a Forest; in which is present a Knife and are on the Beach. More complex semantic models could be adopted supporting also queries as: in which Kate has a Gun after the Second Airplane Disaster (as Event); on the Beach or in the Forrest, etc. For example, an example of semantic query in SPARQL to get "All the scenes where is present Jack and there is a gun", is reported in the sequel:

```
SELECT DISTINCT ?Video ?Annotation WHERE {
        ?Video a msp:Video. ?Video msp:hasAnnotation
        ?Annotation.
        ?Annotation msp:whoIsPresent ?character.
        ?character rdfs:label "Jack".
        ?Annotation msp:thereIs ?object. ?object
        rdfs:label "Gun".
}
```

After some seconds of the main video execution, other 4 video annotations start and are executed at the same time on the right side of the MyStoryPlayer in web page. The videos on the right side may contain explanations, simple flashback, contemporaneous events, as well as related issue. Each of them presented a different starting point and duration. In this context, the user may decide to interact with the presented media objects and controls.

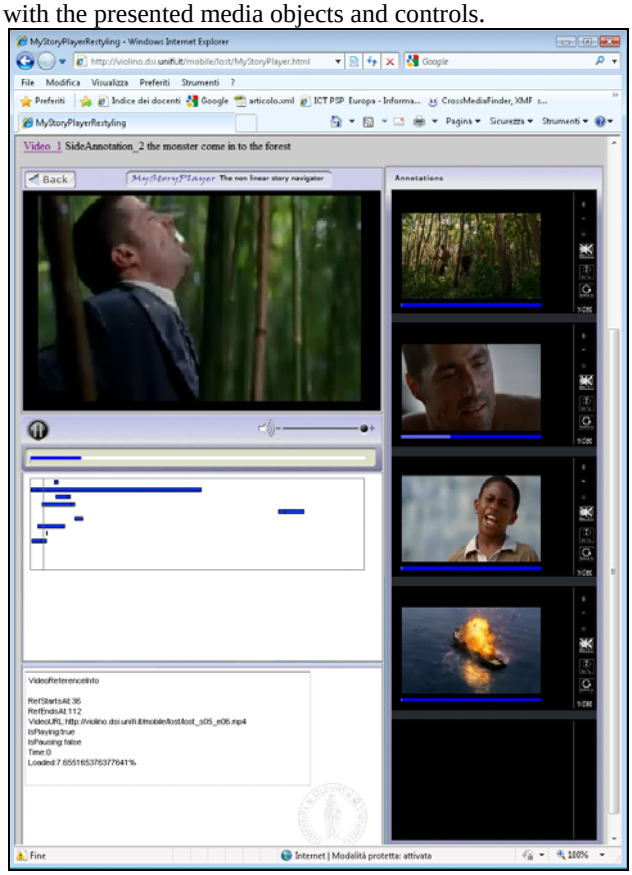

Figure 5. MyStoryPlayer at work

Figure 5 shows a case in which an annotation started with a video (the experience can be performed on http://www.mystoryplayer.org by looking for all scenes with Jack and a gun).

For example: to follow the play of the multiple streams by keeping active the audio on one or more of them, or may decide to change the context; to see the textual annotations associated with one to the video by selecting corresponding "info" button. By clicking one of the videos (audio or images as well) on the right provokes the change of context, bringing the clicked media in the first place (on left). From the point of view of the ontological model, the change of context means to contact the MyStory Server with a query to get the annotations of the selected media and to recreate the knowledge into the MyStoryPlayer. The knowledge may grow around the present network of annotations and the past annotations may be discharged after some jumps. It is evident that new annotations have been loaded while the backlog of the past annotation is still accessible. In order to guarantee a number of back jumps to rewind the experience, the stack of the activities performed have to be maintained, storing also time instants in which the changes of context have been performed.

6. CONCLUSIONS

The paper presented MyStoryPlayer model and tool to execute non-linear stories by following their internal temporal and logical relationships formalized via semantic annotations. In MyStoryPlayer, any media segment can be an annotation for another media element. The user may navigate in the audiovisual annotations creating its own experience. The resulting solution includes a uniform semantic model, a semantic database, a distribution server for semantic knowledge and media, and the MyStoryPlayer to be used in web applications. The solution has been tested on a number of scenarios related to the modeling of non-linear story telling. An example is also accessible via web for the reviewers of the paper. In the short future, the solution would be adopted for modeling educational material such as those that can be created for medical and theatrical environments and in particular for ECLAP PsP project of the European Commission. In both cases, the scenes are typically recorded from more than one point of view, additional explanation video have also to be added showing the historical aspects (on different time and spaces) or about the basic technologies/interpretations or comparison.

Acknowledgement

The authors would like to thanks to ECLAP project partners.

7. REFERENCES

Bellini, P., I. Bruno, P. Nesi, "Exploiting Intelligent content via AXMEDIS MPEG-21 for modeling and distributing news", on Int. Journal of Software Engineering and Knowledge Engineering, vol.21, n.1, 2011

Burnett, I.S., Davis, S.J., Drury, G. M., "MPEG-21 digital item declaration and Identification-principles and compression", IEEE Transactions on Multimedia, Vol.7, N.3, pp.400-407, 2005.

Dakss J., Stefan Agamanolis, Edmond Chalom, V. Michael Bove, Jr., ``Hyperlinked Video," Proc. SPIE Multimedia Systems and Applications, v. 3528, 1998.

Koivunen, M., Swick, R., and Prud'hommeaux, E., "Annotea Shared Bookmarks", 2003, In Proc. of KCAP 2003, http://www.w3.org/2001/Annotea/Papers/KCAP03/annoteabm.html

Kosovic D., Ronald Schroeter, Jane Hunter, "FilmEd: Collaborative Video Annotation, Indexing and discussion tool over Broadband networks", International Conference on

Multimedia Modeling archive Proceedings of the 10th International Multimedia Modelling, Page: 346 , 2004.

MPEG-7: http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm