SEMANTIC VIDEO ANNOTATION IN E-LEARNING FRAMEWORK

Antonella Carbonaro, Rodolfo Ferrini
Department of Computer Science
University of Bologna
Mura Anteo Zamboni 7, I-40127 Bologna, Italy
Tel.: +39 0547 338830 Fax: +39 0547 338990
e_mail: carbonar@csr.unibo.it, ferrini@csr.unibo.it

Abstract. The aim of this paper is twofold. In the first part we will consider the new technologies proposed as solution for the Semantic Web and then we try to outline possible applications in different fields of e-learning giving example of actual works. In the second part of the paper our proposal of improvement of an e-learning with a semantic video annotation module is presented.

1. Introduction

Nowadays Web is maybe the larger available repository of resources. But when one has to look for within an enormous set of information, the research process i.e. learning material, may look as too expensive without the help of machines. The problem is that the Web was designed (and currently is) for human usage and its resources are not machine-understandable. It’s necessary adding some machine understandable meta-data to resources in order to provide and build services, agents or other kind of applications that help students in their tasks.

In last years we have had an improvement in the kind of resources used as learning material. A tangible example of such trends are video resources. Automatic systems for video segmentation and annotation are the requirements of all digital video management systems. The goal is to find automatic and general procedures to segment videos into scenes and to annotate them with textual data or with metric information. Annotations could be useful for further indexing, retrieval, recommendation and so on, performed both by human users and by automated applications.

The Semantic Web [Berners-Lee et al., 2001] is the proposed solution in semantic resource annotation perspective.

The Semantic Web can be defined as an extension of the current Web in which meaning is added to resources so that machines are allowed to understand them better. This new architecture is based on the annotation of web documents with additional semantic data. In these last years a number of new languages have been proposed in order to carry out this task.

In [Bighini and Carbonaro, 2004] is introduced the InLinx (Intelligent Links) system, a Web application that provides an on-line bookmarking service. The overall system it has been firstly improved in [Carbonaro and Ferrini, 2005], introducing concepts for classification, recommendation and document sharing to provide a better personalized semantic-based resource management. Most recently we have introduced the video annotation module Scout-v [Carbonaro et al., 2006]. Scout-V performs automatic shot detection and supports user during
annotation phase in a collaborative framework proposing suggestion on the basis of the effective user needs and modifiable user behaviour and interests. The paper is organized as follows. Section 2 provides an overview of semantic web technology. Some examples of current research projects are given in section 3. In section 4 an inLinx architectural overview is described while a description of Scout-V module is given in section 5.

2. Semantic web technologies

In the above section we have introduced the development of languages for resource annotation. But in the semantic web architecture there is another important layer to take into consideration.

![Semantic Web Layers](image)

XML allows the definition of personalized tags while RDF allows the metadata realization. But if you want to univocally describe a resource, you need to refer your RDF statement to an ontology. Such a term is defined as a shared specification of a conceptualization [Gruber, 1993] and can be used as a knowledge representation of a particular domain by allowing applications to communicate each other and to “understand” the meaning of the annotated resources.

In the rest of this chapter we will briefly describe some of the main technologies developed for the Semantic Web. Most of them have been proposed as standard by the WWW Consortium, the leftover ones are alternative solutions.

- **XML** [XML, 2004]: is a well known Mark Up Language recommended by W3C. This is a technology that allows to go deeper into the page structure provided by HTML.

- **RDF** [RDF, 2004]: is the W3C recommendation for the creation of metadata about resources. With RDF one can make statements about a resource in the form of a subject-predicate-object expression. The described resource is the subject of the statement, the predicate is a specified relation that links the subject and the object that is the value assigned to the subject through the predicate.
- **OWL** [OWL, 2004]: is the W3C recommendation for the creation of new ontology optimized for the web. The Web Ontology Language OWL is a semantic markup language for publishing and sharing ontologies on the World Wide Web. OWL is developed as a vocabulary extension of RDF and it is derived from the DAML+OIL Web Ontology Language. For these reasons it provides a greater machine interpretability of Web content than that one supported by its predecessors. Essentially, with OWL one can describe a specific domain in terms of class, properties and individuals. It has three increasingly-expressive sublanguages: OWL Lite, OWL DL and OWL Full.

- **UML** [UML, 1999]: is a specification language originally designed for software engineering. UML provides a set of graphical notations to describe an abstract model of a system. A lot of efforts have be recently done in order to map UML notation with Semantic Web technology like OWL.

- **RuleML**\(^*\): is a language defined by the Rule Markup Initiative for the definition of XML rules. The main objective of this project is to provide a basis for an integrated rule-markup approach that will be useful for many interconnected purposes.

3. **Semantic web and e-learning**

How can semantic web improve e-learning systems? This new web perspective promises more reusability of resources as well as much more flexibility in systems architecture and in interoperability between them. The research on e-learning and Web-based educational systems traditionally combines research interests and efforts from various fields, in order to tailor the growing amount of information to the needs, goals, and tasks of the specific individual users. Semantic Web technologies may achieve improved adaptation and flexibility of e-learning systems and new methods and types of courseware which will be compliant with the Semantic Web vision. In the following sections we will describe some examples of existing projects thanks to which we will be able to outline what the current research on these fields offers.

3.1. **IMS Learning Design**

IMS LD\(^†\) is defined as a language for modelling units of study. Its specification language is founded on the Educational Model Language\(^‡\) [Koper, 2001]. IMS LD extends EML by covering not only the learning process modelling but also integrating the aspects of other specifications like IMS content packaging and IMS LOM metadata. The starting point in IMS LD development is the separation between learning activities and their learning resources. LD doesn’t completely capture the aim of the semantic web but the conceptual model expressed in UML and the constructs specification expressed in XML

\(^*\) [RuleML](http://www.ruleml.org/)
\(^†\) [IMS Global](http://www.imsglobal.org/learningdesign/index.cfm)
\(^‡\) [EML](http://eml.ou.nl/introduction/docs/ped-metamodel.pdf)
can provide a semantic notation that e-learning systems could extend with their preferred semantic web technologies.

### 3.2. Edutella

Edutella§ is defined as a multi-staged effort to scope, specify, architect and implement an RDF-based metadata infrastructure for P2P-networks for exchanging information about learning objects. Edutella P2P architecture is essentially based on JXTA and RDF. JXTA** is an open source technology that provides a set of XML based protocols supporting different kind of P2P applications.

According to [Nejdl et al., 2002] three types of services, which a peer can offers, are defined in an Edutella network:
- Edutella Query Service: the basic service in the framework. It provide a common, RDF-based query interface (the Query Exchange Language – RDF-QEL) for metadata providing and consuming through the Edutella network;
- Edutella Replication: it provide replication of data to additional peers to ensure data persistence;
- Edutella Mapping, Mediation, Clustering: this kind of services manages metadata allowing semantic functionality of the global infrastructure;

An important point to underline is that Edutella doesn’t share resource content but only metadata.

### 3.3. Smart Space for Learning

Smart Spaces for Learning is the result of the Elena project†† work [Simon et al., 2003a] and according to [Simon et al., 2003b] a Smart Space for Learning can be defined as a set of service mediators which support the personalized consumption of heterogeneous educational services provided by different management systems. But what is a learning service? Learning services are entities designed to satisfy a specific purpose (e.g. the delivery of a course). They may use resources as learning objects (e.g. exercises and exams) and Web services‡‡ to interface the formers with learners.

The system architecture of Smart Learning Spaces is essentially composed by this two building blocks:
- an Edutella network;
- a set of ontologies;

In a Smart Space for Learning, providers of learning services are connected to a learning management system which is based on Edutella. Ontologies has to describe the learning domains using concepts and relations that may be referred to in the annotations of the learning services. A brief description of an ontology language (the Web Ontology Language OWL) is given in section 2.

### 3.4. HyCo

---

** [http://www.jxta.org/](http://www.jxta.org/)
†† [http://www.elena-project.org/](http://www.elena-project.org/)
‡‡ W3C’s overview on Web services at [http://www.w3.org/2002/ws/](http://www.w3.org/2002/ws/)
HyCo [García et al., 2004] stands for Hypermedia Composer and it is a multiplatform tool that supports the creation of learning materials. HyCo is the result of the development of an authoring tool created with the intent in order to define ALDs.

According to [Berlanga and García, 2005] ALDs there are some units of learning that contains personalized behaviour in order to provide each student with a learning flow which is to be adequate to her/him characteristics. ALDs are semantically structured according to IMS LD (described above) in order to allow reusability.

The last version of HyCo also manage a kind of resources named SLOs. A SLO is a Learning Object compliant with IMS Metadata§§. Every resource created with HyCO is turned into a SLOs. Whenever the conversion process is finished an XML file is generated for the new SLO and stored in a repository.

4. **inLinx**

In this section we are going to briefly introduce the InLinx (Intelligent Links) system [Bighini and Carbonaro, 2004], a Web application providing an on-line bookmarking service. InLinx is the result of three filtering components integration corresponding to the following functionalities:

- bookmark classification (content-based filtering);
- bookmark sharing (collaborative filtering);
- paper recommendation (content-based recommendation);

Over the years we have introduced several extensions to the original architecture. Generally, recommender systems, like inLinx, uses keywords to represent both users and resources. Another way to handle such data is using hierarchical concept categories. This issue will enable users and the system to search, handle or read only those concepts of interest in a more general manner. For example, synonymy and hyponymy can reveal hidden similarities, by potentially leading to better classify and recommend. The advantages of a concept-based document and user representation are: (i) ambiguous terms inside a resource are disambiguated, allowing their correct interpretation, consequently, a better precision about the user model construction (e.g., if a student is interested in computer science resources, a document containing the word ‘bank’ in the financial context will not be relevant); (ii) synonymous words belonging to the same synset can contribute to the user model definition (for example, both ‘mouse’ and ‘display’ brings evidences for computer science documents, improving the coverage of the document retrieval); (iii) finally, classification, recommendation and sharing phases take advantage from the word senses in order to classify, retrieve and suggest documents with high semantic relevance with respect to the user and resource models.

5. **The Scout-V Module**

The Scout-V module assists authors in the task of annotating video sequences. Multimedia annotation systems need standard output, compliant with other tools

§§ [http://www.imsglobal.org/metadata/index.cfm](http://www.imsglobal.org/metadata/index.cfm)
for browsing or indexing. MPEG-7 [ISO/IEC, 2002] standard was defined to this purpose. It represents an elaborate standard in which a number of fields ranging from low level encoding scheme descriptors to high level content descriptors are merged, useful for describing a video and part of it.

Each shot belonging to the video sequence can be annotated on the base of underling ontology. These descriptions are labeled for each shot and are stored as MPEG-7 descriptions in the output XML file. Scout-V can also save, open, and retrieve MPEG-7 files in order to display the annotations for corresponding video sequences.

The Scout-V main page shows all the videos that should be elaborated performing shot detection, editing or removing, as illustrated in Figure 2.

Figure 2  System video processing: shot detection phase is selected.

Given the segmentation of video content into video shots, the second step is to define the semantic lexicon in which to label the shots. A video shot can fundamentally be described using five basic classes: agents, objects, places, times and events. These five types of lexicon define the initial vocabulary for our video content; they correspond to the SemanticBase MPEG-7 tags.

We have also defined attributes to describe class characteristics. Each attribute corresponds to a specified MPEG-7 tag in storing phase.

Using the defined vocabulary for static agents, key objects, places, times and events, the lexicon is imported into Scout-V for describing and labeling each video shot. The shots are labeled for its content with respect to the selected lexicon. Note that the lexicon definitions are database and application specific, and can be easily modified and imported into the annotation tool.

Scout-V annotation tool is divided into three graphical sections, as illustrated in Figure 3. On the upper left-hand corner of the tool is located the Scene Matching frame in which are specified the algorithms that can be used to obtain video annotation recommendations (Block Truncation Coding, edge histogram,
colour histogram). On the bottom left-hand portion of the tool is placed the Ontology Visualization frame providing interactivity to assist authors of the annotation tool. On the right, is the Video Presentation frame with key frame image display and frame characteristics. The Ontology Editor module allows to modify the ontology tree creating needed classes and instances. The two figures show how to create ontology structure introducing new classes and their hierarchy and how to populate created class with a specific instances. The aim of the instance creation phase is to effectively represent the domain knowledge, achieving better precision in the annotation task.

![Figure 3 Scout-V annotation tool.](image)

Figure 4 shows annotation procedure apply on the first scene using checkboxes and comments to better describe the selected video. Annotations are then stored and used by recommendation procedure to help users finding similar frames annotated also by different users. When all the scene have been annotated the system produces the MPEG-7 file.
6. Conclusions

The paper addresses a key of the semantic web technology usage in e-learning. We have explored some of the most important W3C’s standard as XML, RDF, OWL as well as some alternative solutions proposed in other research fields like RuleML.

We have tried to address our main purpose by giving examples of our current research project aiming to extrapolate which and how the semantic web technologies are involved in their architecture.

Subsequently a brief introduction of inLinx and of its new semantic video annotation module has been given. inLinx is an hybrid recommender system that provides an on-line bookmarking service. Scout-V performs automatic shot detection and supports user during annotation phase in a collaborative framework proposing suggestion on the basis of the effective user needs and modifiable user behaviour and interests.

7. References


[Carbonaro and Ferrini, 2005] A. Carbonaro, R. Ferrini. “Considering semantic abilities to improve a Web-Based Distance Learning System”. ACM International Workshop on Combining


