Semantic annotation of video fragments as learning objects: a case study with YouTube videos and the Gene Ontology

Miguel Angel Sicilia¹, Salvador Sánchez-Alonso¹ and Miltiadis Lytras²

¹Information Engineering research unit, University of Alcala
Edificio Politecnico, Ctra. de Barcelona, Km.33,6 – Alcala de Henares (Spain)
{msicilia, salvador.sanchez}@uah.es

²American College of Greece
Gravias 6 St., Aghia Paraskevi (Greece)
mlytras@acgmail.gr

Abstract

Web 2.0 technologies can be considered a loosely defined set of Web application styles that foster a kind of media consumer more engaged, and usually active in creating and maintaining Internet contents. Thus, Web 2.0 applications have resulted in increased user participation and massive user-generated (or user-published) open multimedia content, some of which is potentially useful for education. In this context, the problem from the educator’s perspective is selecting and annotating existing content so that software applications can filter fragments previously marked as useful for particular learning needs. This paper discusses a solution for such problem that is non-intrusive to existing applications. This solution fits in the philosophy of multiple metadata profiles, allows for expressing fine-grained learning needs, and leverages the growing mass of contents by reusing well-established domain ontologies. A description on the technical aspects required for the infrastructure supporting such solution is first provided. Then, the solution is validated with a case study using a knowledge model of gene products —the Gene Ontology— to semantically annotate videos in YouTube that might be used in teaching biology and bioinformatics.

Keywords
Web 2.0., video clips, multimedia, learning objects, ontologies, semantic annotation, YouTube, Gene Ontology.

1. Introduction

Web 2.0 is a term which refers to several things, including behaviors, technologies and ideals. It considers services and activities that foster a new kind of media consumer who is more engaged and active in creating and adding value to the content which is the basis for using the Internet (Allen, 2008). Thus, the key of Web 2.0 applications from the viewpoint of digital resource production is that they enable mass participation in social activities structured around the contents. User-generated content (UGC) consists mostly
of micro-contents which can be considered fine-grained learning objects, provided they have the minimal accuracy and reliability requirements to be appropriate for education in participatory pedagogies (Collis and Moonen, 2008). In some cases, users do not create the contents themselves, but play the role of publishers, annotators and reviewers of contents already available. Therefore, Web 2.0 applications become massive repositories in which users share the results of their use of Web resources. A good example of this is the thousands of video clips about Deoxyribonucleic acid (DNA) in YouTube which can be annotated, commented or rated.

It is extremely attractive to reuse the growing mass of UGC, selecting content subsets that are known to be accurate and useful for particular educational needs. These resources are publicly available right now, at negligible development costs. Considering that designing and implementing quality materials is one of the most expensive parts of e-learning courses’ budget, teachers would love it if they did not have to develop new materials—a very costly choice when we speak of multimedia resources—but only to adapt pre-existing resources to each specific learning context or situation. In the following pages we will detail a form of reusing user-generated videos through the provision of metadata information in the form of semantically meaningful annotations.

There are many approaches to the annotation of digital video in the literature. Feng, Mannatha & Lavrenko (2004) introduced an automatic image and video annotation technique for retrieval based on textual queries, where the images forming the sequence are partitioned into regions. Their motivation was to improve the conventional approach (keyword-based manual annotation of each image carried out by librarians) as they felt it often did not capture the semantics of the resource. Besides their considerations, we must add that the mentioned model is too expensive in terms of cost and effort. Another interesting research direction was pointed out by Bargeron et al. (1999), who introduced MRAS, a prototype of a collaborative video annotation system. This system was intended for personal use—annotations were not intended to serve for future processes of information management—but revealed the possibilities of annotating video contents. Miyamori and Isaku (2000) proposed an automatic annotation method of sports video for content-based retrieval. Even though their motivation was "to develop integrated reasoning modules for richer expressiveness of events and robust recognition", they did not exploit the potential of Semantic Web technologies. A similar effort worthy of comment is the one by Volkmer, Smith & Natsev (2005) who “describe and evaluate a web-based system for collaborative annotation of large collections of images or temporally pre-segmented videos”. Just to mention a few other significant contributions to the annotation of video, those about speech (Brown et al., 2001) and text (Lienhart, 1996) deserve recognition.

Our main motivation is to describe a way of adapting existing video sequences, not specifically created for learning, transforming them into ready-to-use learning resources. But to achieve the goal of reusing UGC we first need to study to what extent UGC is reliable and accurate enough to be used for learning. It is reasonable to assume that part of the UGC in Web 2.0 applications is about information potentially useful for educational purposes. However, the information available is often diverse, problematic or even contradictory:
According to a recent study which examined a total of 146 unique YouTube\(^1\) video clips about human papillomavirus (HPV) vaccination, only 74% portrayed HPV vaccination in a positive manner (Ache and Wallace, 2008).

For some topics, such as contraception, most scientific claims in YouTube have been considered as unsubstantiated (Luttrell, Zite and Wallace, 2008).

For some other topics, such as surgical procedures, the information in Web 2.0 sites is mostly accurate. Devgan et al. (2007) identified that about 86% out of 35 procedures in the Wikipedia were deemed to be appropriate “for patients”.

Regarding the information categorization typical of Web 2.0 systems, there is evidence that ambiguity and polysemy in the structure of the folksonomy tags are indeed important problems according to Spiteri (2007). Scattered analyses by this author portray a very diverse landscape of quality and accuracy in the contents and organization of Web 2.0 applications.

Articles on the use of Web 2.0 technologies in education are increasingly frequent in the literature, even though there is a need to discriminate sound research from hype (Rollet et al., 2007). In any case, the role of these technologies in participatory pedagogies has been profusely discussed (Collis and Moonen, 2008; Eijkman, 2008). Web 2.0 sites like YouTube have been found to have a considerable potential in education (Milliken et al., 2008), so that they have been used as the point of departure for novel trends in educational innovation (Jenkins, 2007). In this paper, we focus only on the potential of UGC as a large base of contents from which educationally meaningful resources can be selected. However, the educational use of UGC available is not straightforward:

- Web 2.0 applications are by definition not quality-controlled, except by the social assessment of the contents. Although this social assessment includes regulations, these are not related to quality but to law enforcement.

- As learning needs are diverse, the same resource may be useful for different contexts and under different usage conditions. This idea of multiple uses is central to the concept of metadata profile (Downes, 2004) and compatible with the philosophy of many educational repositories which only store the metadata but not the resources themselves, so different repositories can hold different metadata descriptions for the same resource.

Therefore, there is a need to build filters for information in Web 2.0 systems which enable the general public to select the appropriate resources for a particular use. In the case of education, instructors would benefit of filters set by others for explicit educational purposes. This kind of peer-review mechanism is currently practiced in learning object repositories that only store metadata (not the contents themselves) such as Merlot (Cafolla, 2006). The idea is to link metadata annotations to existing resources in a non-intrusive way, i.e. external systems managing the resources. This filtering requires the provision of metadata storage facilities connected to classification systems offering an unambiguous way of describing learning needs. Existing formal ontologies provide such descriptive framework, e.g. ontology-based approaches to storing learning object metadata like the one reported by Gaševic, Jovanovic and Devedžić (2007). In the case of continuous media such as video clips, it is also a need to formally annotate resources (e.g. a fragment of a video clip) and link them to existing multimedia synchronization languages to combine them, which can be achieved using multimedia metadata ontologies.

---

\(^1\) http://www.youtube.com
This way, ontologies can be employed as a filtering mechanism for the increasing amount of UGC available in Web 2.0 applications, whereas educational considerations can be used in the filtering process by means of formally representing simple pedagogical methods.

This paper reports on an approach to ontology-based annotation of media fragments. This approach allows not only multiple descriptions but also links from those descriptions to any arbitrary domain ontology, including a knowledge basis about educational considerations. We focus on video clips as can be found in popular Web 2.0 applications like YouTube or MySpace, which can be potentially reused in pedagogies considering multimedia (Mayer, 2005). Our approach builds on existing efforts in semantic learning object metadata repositories and the possibilities offered by service-oriented architectures (Monceaux et al., 2007). The case of video clips on molecular genetics is used as an illustration of the main elements of the approach.

The rest of this paper is structured as follows. Section 2 provides some background information on multimedia and ontologies in learning technology. Section 3 describes the general approach and technical elements, whereas section 4 describes the case study. Finally, conclusions and future work are provided.

2. Background and related work

The use of video clips for the design of learning objects has been described in the literature. For instance, Cochrane (2007) reports on the use of action research for devising pedagogically sound interactive learning objects using QuickTime technology. However, the challenge of ontology-based annotation of media fragments has not yet been explicitly addressed.

There are several languages and metadata schemas specific for multimedia objects. For example, the Synchronized Multimedia Integration Language (SMIL)\(^2\), is an HTML-style language which enables the authoring of multimedia presentations. Even though such markup languages enable loose composition of media objects, nowadays they are rarely used in Web 2.0 community sites like YouTube. In contrast, these sites use specific video formats such as FLV, optimized for quick rendering. SMIL provides some media control through the `clipBegin` and `clipEnd` attributes, along with the mechanisms to pass parameters to media players. Then, existing clips annotated for educational purposes can be integrated in synchronized SMIL presentations if there is information available about both timing and educational properties.

The discussion below assumes concepts on learning objects and their relation to continuous media, as well as some basic knowledge on formal ontologies. These elements are briefly surveyed in the rest of this section.

2.1. Learning objects, granularity and media objects

The concept of learning object (LO) has been subject to considerable debate. In consequence, several definitions have been proposed (McGreal, 2004), some of them

\(^2\) http://www.w3.org/AudioVideo/
referring the need of metadata records associated to the resources. In any case, the application of Semantic Web to learning objects requires them to be annotated with rich metadata expressed with a formal, logics-based language and referring to shared domain ontologies.

However, there is a lack of tools that allow instructors or cataloguers to formally describe fragments of media assets. The growth of the Internet and particularly of some Web 2.0 tools that empower users to store, share and tag videos, picture or audio files have resulted in a huge base of potentially useful learning resources of fine granularity. A new relevant category of interest to include such capabilities is continuous media, i.e. videos or audio files.

Current learning object specifications like SCORM (Advanced Distributed Learning Initiative, n.d.) allow the composition of units of instruction. However, single files are dealt with as atomic assets, which make it impossible to reference parts of such assets for the purpose of sequencing or selecting particular parts. A possible solution could be splitting the continuous stream in smaller fragments, but this would entail redundancy because different pedagogical usages may lead to different fragmentations. Also, the philosophy of reuse in learning object fosters the reuse of resources for learning that originally did not have an educational intention. Consequently, there is a need to specify idioms that complement SCORM and other standards to deal with multimedia fragments.

2.2. Ontologies and the semantic annotation of learning contents

The “Semantic Web” vision described by Berners-Lee et al. (2001) has resulted in a considerable amount of research and development initiatives to extend the current Web technology with machine-understandable metadata. Formal ontologies (Gruber, 1993) play an essential role in the Semantic Web paradigm, providing the shared conceptualizations expressed in logics-based form that can be used by software agents to act on behalf of humans in search processes or distributed activities. In other words, ontologies provide metadata with shared semantics so that interoperable intelligent agents can act, reason and make decisions according to the information in the metadata.

Although simpler mechanisms exist such as mappings to RDF (Palmer, Nilsson and Brasse, 2003), several initiatives are working to represent the IEEE standard for learning object metadata, IEEE LOM\(^3\), in ontological form. Some of them go further than just mapping the original IEEE LOM to an ontology language like OWL (McGuinness, & van Harmelen, 2004) or WSML (de Bruijn et al., 2005). In fact, these efforts implement different mechanisms of referencing domain ontology elements inside metadata elements (Sánchez-Alonso, Sicilia & Pareja, 2007; Dodero et al., 2005). Several proposed ontological schemas for learning object metadata allow us to describe learning objects in terms of any available ontology (Sicilia et al., 2006; Gaševic, Jovanovic & Devedžic, 2007) in such a way that specialized software can be used to exploit the relationships, rules and axioms in the ontologies for (i) navigating repositories, (ii) creating tentative learning object compositions or (iii) searching for learning resources. Known applications related to the research presented here include ontology-based composition to build exercises (Fischer, 2001) and compositions tailored to personalized learner needs (Jovanovic, Gaševic & Devedžic, 2006).

\(^3\) http://ltsc.ieee.org/wg12/
Although there are works to annotate the temporal behavior of multimedia audio or video objects with ontologies (Hausmanns et al., 2003) these were used basically to index high granularity resources and not specific fragments of pre-existing media. Gahegan et al. (2007) reported on a system to exploit the richness of relationships inside ontologies integrating several facets, namely, technology, tasks, interactions, learning approaches, techniques and learning outcomes. Multimedia metadata ontologies provide a description of entire media files, even though the practicalities of annotating fragments are specified only at a high level. An example of this is the AudioVisualSegment concept in the MPEG-7 ontology defined by Hunter (2006), a concept that could be used for defining a fragment of a video clip. Other ontologies on multimedia include structural features like shape recognition (Simou et al., 2005) which are not relevant to our research, although they might be useful for describing recognized elements inside media.

Regarding the existence of similar applications to the one we present in this paper, Bagdanov et al. (2007) reported on the implementation of a software tool for semantic annotation of video in digital format using multimedia ontologies. This tool allowed “to perform higher-level annotation of the clips, to generate complex queries that comprise actions and their temporal evolutions and relations and to create extended text commentaries of video sequences”. VIA is a similar tool created for the BOEMIE (http://www.boemie.org) project. Its user friendly graphical interface allows users to "manually annotate both image and video files, using descriptors that are formalized as ontologies". However, these tools are not aimed at carrying out annotations for educational purpose, nor provide the possibility of giving the user the possibility of creating “new” virtual video sequences from the temporal delimitation and metadata annotation of existing resources.

Given that the representation of segments in existing media ontologies is not present in learning object ontologies, we identified the need for combining educational resources with the media descriptions in a single formal framework. The semantic clip annotation described in this work complements existing ontology-based annotations of learning resources by enabling the reuse of pre-existing media in Web 2.0 applications. As it integrates this kind of media in learning object ontologies in a non-intrusive way, it does not affect applications relying on them, e.g. the semantic web services offered by advanced learning object repositories such as ont-space⁴ or SLOR⁵. In this paper, we will not use any of the mentioned IEEE LOM mappings in a strict sense, but will instead provide a few ideas on how to seamlessly take profit of the educational and the media descriptions in a unique knowledge representation model.

3. Semantic annotation of existing continuous media

An approach to semantically annotate resources as generated in the current Web 2.0 social environment should meet a number of requirements specific of the nature of Web 2.0 content (Kolbitsch & Hermann, 2006). Concretely, the following requirements are important in this environment:

R1. Each uniquely identified resource (e.g. a video clip, a picture or a blog post) could be described several times, with different interpretations, following the

⁴ http://sourceforge.net/projects/ont-space/
⁵ http://slor.sourceforge.net/
idea of resource profiles (Downes, 2004). This extends the plurality of uses of micro-contents to metadata, and allows us to define and evaluate educational contexts, which is pivotal to evaluate reusability (Sicilia & García-Barriocanal, 2003).

R2. Metadata must be stored externally to the resources themselves, making the annotation non-intrusive to the original applications.

R3. Annotations could use different domain ontologies, including existing ontologies specific to pedagogical models. A single metadata record could combine links to several domain ontologies.

Let us finally discuss about the requirements. On the content application side, some form of permanent, unique identifier for each resource should exist. Fortunately, this requirement is somewhat included in most of current Web 2.0 applications. On the ontology-based annotations side, a reasonably mature and formally correct ontology must exist. Finally, the separation of metadata and the resources they refer to, e.g. would allow mashups to combine applications and metadata browsers able to dynamically combine both information spaces.

3.1. Representing learning objects as ontology instances

The main aspect of the annotation approach presented herein is to represent learning object metadata as ontology instances. This approach is similar to the TagOntology (Gruber, 2007), which enables reasoning about the learning objects. The IEEE LOM standard can be translated to ontology description languages by mapping or repurposing elements. The main elements of the adaptation of IEEE LOM to an ontology language used are described elsewhere (Sánchez-Alonso, Sicilia & Pareja, 2007). When using such ontology translations, it is not only possible to add rules and logical axioms to the metadata elements, but also to link metadata elements to instances in domain ontologies. Regarding persistency, the storage of metadata in ontological form can be carried out by using ont-space, a purposefully built open source framework used to include any number of domain ontologies in OWL. This framework was developed as part of the EU project LUISA6 and has been successfully tested in different educational environments.

Such external metadata repository infrastructure fulfils requirement R2 above, since any number of metadata repository instances can be setup. This can be used as a semantic-based extension of the approach of distributed annotation servers chosen in Annotea annotation servers (Kahan et al., 2002). In formal terms, we should speak about resource profiles, as a resource such as a video clip identified by a URI is represented by a number of potentially distributed metadata records.

3.2. Mapping continuous media fragments

The annotation process starts by creating a LearningObject instance in the ontology for a permanent identifier (if not yet existing). Subclassing allows for combining by subsumption the concepts AudioVisual—in MPEG-7 ontologies (Hunter, 2006)—and LearningObject into a new concept labeled AVLearningObject. Instances of

6 http://www.luisa-project.eu
this concept are any video sequences in digital format which can be (re)used in different educational context, and which include metadata descriptions in a standardized form.

Our first problem is that the just mentioned mapping assumes that a metadata record describes an entire media resource: the one identified by URI in the Identification section of IEEE LOM. This is in many cases inappropriate, especially in the case of continuous media, where segments of the full content can be considered learning resources. MPEG-7 provides the Segment concept to model fragments of multimedia content, and the decomposition property relates segments to media contents. However, it is not possible to specify the starting and ending time points of a segment. Consequently, a necessary and sufficient condition has to be used to automatically classify both segments and learning objects describing segments:

\[
\text{ContinuousSegment} = \text{MultimediaContent and exactly partOf.MultimediaContent and exactly starts.float and exactly ends.float}
\]

The partOf property has the same meaning that the homonymous property in the IEEE LOM relationship category. It refers to strict semantics based on physical containment as related to continuous entities. In this way, the identity of a part is simply determined by the extent, i.e., equality in start and end. This simple representation is sufficient for video clips or audio files. Figure 1 illustrates the ontology classes mentioned in this section and their relations.

Figure 1. A partial view of the classes in the audiovisual learning object ontology
As Figure 1 shows, instances of \texttt{AVLearningObject} are indirectly subsumed by \texttt{MultimediaContent} so that such classification criterion serves also for metadata profiles of learning resources.

![Identifier: http://www.youtube.com/watch?v=teV62zrm2P0](image)

Figure 2. Annotation cases for a given resource.

Figure 2 shows a possible configuration of annotations with different options. Profile 1 represents a conventional annotation of the entire asset, so it is not a \texttt{ContinuousSegment}. Profiles 2 and 3 are continuous segments and represent two different educational perspectives on the same resource. For example, it may be that the first one is annotated for the context of secondary education, while the second one is annotated for bioinformatics courses, so that the educational uses and properties are widely diverging. Using an abstract syntax to define it, profile 2 would look like:

\begin{verbatim}
ContinuousSegment(?profile2) ∧
partOf(?profile1) ∧
educationalContext(?secondary-education) ∧
starts(?13:01) ∧
ends(?25:34)
\end{verbatim}

Profile 4 is an additional perspective with a subsegment overlapping some other profiles. Each of the profiles in this example would be represented by an instance of the concept \texttt{AVLearningObject}. At the same time, profiles 2, 3 and 4 would be related to profile 1 with a \texttt{partOf} predicate.

### 3.3. Linking to domain ontologies
In our case, the main purpose of reusing domain ontologies is to express learning objectives mainly through the use of the IEEE LOM element 9.Classification. Although other forms of linking learning objects to ontology elements could have been devised – e.g. defining a property for linking video fragments to concepts of domain ontologies instead of using IEEE LOM elements—, we chose this alternative because unlike other options, this decision has the advantage of producing IEEE LOM conformant metadata records. Besides, IEEE LOM Category 9 is a multi-purpose metadata element aimed at describing the contents of the learning resource, so we can use any ontology translation of IEEE LOM for this purpose. To proceed with this form of annotation, a two step process will be followed:

- **Specify the purpose:** This requires the compilation of a number of purposes of the classification that combine the content of the resource with the type of interaction, as will be described in what follows.
- **Select the TaxonPath:** To be set depending on the domain and the purpose of the classification.

As an example, let us consider the purpose of visualizing processes. This could be represented as a specific purpose from which a logical rule can infer that the associated taxon path is a process. SWRL\(^7\), a high-level abstract syntax for rules in OWL, can be used to represent such rule:

\[
\text{LearningObject(?x)} \land \\
\text{classification(?x, ?c)} \land \\
\text{purpose(?c, visualizing-process)} \land \\
\text{taxonpath(?c, ?tp)} \\
\rightarrow \text{Process(?tp)}
\]

This serves two purposes. First, if the associated taxon path for the classification was something incompatible with a process, an inconsistency would be detected. Second, if the inference is successful, then it will be possible to use knowledge about it if such knowledge is formally represented in the ontology. For instance, if there is a formal description of processes and their constituents, it is possible to link them by using some sequencing technique as in elaboration theories (Reigeluth, 1999). Assuming that, formal definitions of processes (such as the processes of replication of DNA, or the photosynthesis process) can also be used to select related resources.

The notion of Process as “sequence of steps” is recurring in different upper ontologies such as OpenCyc (Lenat, 1995), DOLCE (Gangemi et al., 2002) or SUMO \(^1\). Analyzing some upper concepts can lead to a number of stereotypes in element purpose that can be defined to enable concrete kinds of reusable reasoning. For example, stereotypes regarding visualization of processes entail a very concrete learner cognitive activity which is considered in theories of multimedia learning. Another similar reusable purpose is visualizing-composition, which in this case refers to any kind of formal mereology, as may be the parts of the cell, or the parts of the digestive system of humans.

### 3.4. Exploiting ontologies for instructional purposes

There is a considerable amount of research that points to the appropriateness of animations and videos as a resource in education. Some of them highlight the importance

\(^7\) [http://www.w3.org/Submission/SWRL/](http://www.w3.org/Submission/SWRL/)
of having adequately short videos, because it is less likely that learners will watch the complete clip if this is too long (Korakakis et al., *in press*). In addition, there is a need to express the role of multimedia elements in the delivery of learning resources. Current theories of multimedia learning (Mayer, 2005) provide a number of empirically validated guidelines on how to combine visualizations with verbal representations, emphasizing at the same time the use of two modes of representation rather than one when explaining a concept (Mayer & Moreno, 1997). In the case of devising full instructional sequences, the role that multimedia presentations can play is supporting concrete, fine-grained concepts that are embedded in the activity sequence. A checking mechanism would be that of marking those learning objects aimed at teaching something about a process, which do not include any audiovisual representation.

(1)  
\[
\text{LearningObject}(?lo) \land \\
\text{classification}(?lo, ?c) \land \\
\text{purpose}(?c, \text{competency}) \land \\
\text{taxonPath}(?c, ?p) \land \\
\text{Process}(?p) \land \\
\text{hasPart}(?lo, ?lo2) \land \\
\text{AVLearningObject}(?lo2) \land \\
\text{classification}(?lo2, ?c2) \land \\
\text{purpose}(?c, \text{visualizing-process}) 
\rightarrow \text{AVProcessLO}(?lo)
\]

(2)  
\[
\text{AVLearningObject}(?lo) \land \\
\text{classification}(?lo, ?c) \land \\
\text{purpose}(?c, \text{visualizing-process}) \land \\
\text{taxonPath}(?c, ?p) 
\rightarrow \text{AVProcessLO}(?lo)
\]

(3)  
\[
(\text{not } \text{AVProcessLO})(?x) \rightarrow \text{LOMissingVisualization}(?x)
\]

The above rules examine whether the learning object is about a process according to an upper ontology generically denoted with the namespace upper. If this is the case, then the rules check whether it is itself a visualization or contains a visualization. The resulting classification can be used to trigger a search for a visualization of the process.

In our formal representation, this can be translated into constraints on the IMS Learning Design (Koper, Olivier & Anderson, 2003) structure. For example, the requirement “Provide the learner control of the sequence when lengthy instructional sequences must be completed by the student in no specific order” (Stemler, 1997) can be implemented as a constraint on a learning activity. Using the `lessThan` SWRL built-in, which satisfies iff the first argument `a1` is less than the second argument `a2` according to a given ordering –temporal sequence in this case–, and the IMS-LD ontology by Amorim et al. (2006), the constraint would look like:

\[
\text{Learning-Activity}(?a1) \land \\
\text{Learning-Activity}(?a2) \land \\
\text{Activity-Structure}(?as1) 
\]
The above rules check if a given learning activity is based on audiovisual material (which can be defined similarly to AVLearningObject) and whether it is sequenced or not. If both questions are answered in the affirmative, the rule recommends removing the sequence. This kind of checks and constraints based on pedagogical guidelines can be used to develop reusable libraries. In that case, instructional designers have two options. On the one hand, they can check that their design complies with a subset of the library. On the other hand, they can select some of the formally represented guidelines so that an intelligent authoring environment could help them in the process of composition of units of learning.

The presented rules are just examples aimed to illustrate the ideas introduced in this section. As such, they do not form part of any available set of rules we have developed. However, they show the path for the development of a larger, organized set of rules. This would of course imply a wider effort, which would include full development and evaluation over selected video fragments.

4. **Case study**

The case study reported here is an application of the semantic annotation concepts described in the preceding sections to the YouTube Web site. As widely known, YouTube is one of the largest and most successful Web 2.0 applications, where users upload, watch and share video clips. Among the millions of videos available, YouTube contains many clips on the dynamics molecular biology processes which can be used as resources in teaching biology and bioinformatics for very diverse learner profiles. Even though this is a particular domain, it serves as a proof of concept on how the mass of resources shared in current community sites could be reused in a non-intrusive way.

In this section, some background on the Gene Ontology (GO) is provided. The main elements of the design of the annotation tool are described along with a concrete example of pedagogical annotation.

4.1. **The Gene Ontology and biological processes**

The Gene Ontology (GO)\(^8\) project is a collaborative effort to address the need for consistent descriptions of gene products in different databases (GOC, 2004). Each entry in the GO has a unique numerical identifier of the form \textit{GO:nnnnnnn}, and a term name.

\(^8\) http://www.geneontology.org/
The GO includes a detailed representation of biological processes, a series of events accomplished by one or more ordered assemblies of molecular functions. Examples of broad biological process terms are cellular physiological process or signal transduction. However, a biological process is not equivalent to a pathway. At present, the GO does not try to represent the dynamics or dependencies that would be required to fully describe a pathway.

The statement of relationships between processes in the GO is specified by a straightforward rule on the degree of overlap, which avoids ambiguous interpretations. To determine whether a process term should be an is_a or part_of child of its parent we need to check if an instance of the child process is also an instance of the entire parent process, i.e. if the whole process takes place from start to finish. If answered in the affirmative, the child is is_a; otherwise, if it is only a portion of the parent process, the child is part_of.

4.2. Example annotation

As a case of annotation, we will deal with the video clip “From RNA to Protein Synthesis”. This video visualizes an abstract view of protein synthesis, which can be identified with the GO biological process coded GO:0006412 and labeled “translation”. This process has the following parts in the GO:

- Translational initiation (GO:0006413), the process preceding formation of the peptide bond between the first two amino acids of a protein.
- Translational elongation (GO:0006414), the successive addition of amino acid residues to a nascent polypeptide chain during protein biosynthesis.
- Translational termination (GO:0006415), the process resulting in the release of a polypeptide chain from the ribosome, usually in response to a termination codon (UAA, UAG, or UGA in the universal genetic code).

Translational initiation starts at second 62 in this clip and lasts until second 102. From that point to second 134 some examples of elongation are provided and the rest of the video illustrates the termination phase. This would result in four learning object metadata records, one for the full clip and the other three declared as partOf of that one. The annotations for the purpose would be set to process-visualization, while the taxonpath would match the corresponding GO term identifier. The GO does not contain information on the sequence of sub-processes in its current version. It however requires every process definition to have a starting and end points, so that in this case it is possible to derive the sequence. The video clip does also contain the visualization of a fragment of another process in the beginning, concretely corresponding to the process of mRNA transcription (GO:0009299). In this way, different parts of the video can be used for different learning objectives.

An interesting feature of the multiple annotation approach is that different aspects can be considered when describing learning objects. For example, the video “DNA translation animation”10 starts with a description of aminoacids preceding the description of the

---

9 identified in YouTube with the URI: http://www.youtube.com/watch?v=NJxobgkPEAo
10 identified in YouTube with the URI: http://www.youtube.com/watch?v=ncjl1D51hQk
translation process. Furthermore, it is two-dimensional and introduces additional concepts such as *codons*, *anti-codons*, and *peptide bonds*, including details of its chemical structure, providing a more detailed level. These elements can also be described as annotations in a simple way, in this case, using classification *purpose* equal to *describing-concept* and specifying the corresponding ontology concepts.

Regarding relationships, it is important to point out that the *partOf* GO relation can be used in several ways in the context of learning resource search. These include the following:

- Expanding a query using a term, to its constituent parts (which are in turn terms in the ontology). In this manner, each part of a process might be visualized by a fragment of a different video clip. This process can be further exploited through the several levels of traversal of the *partOf* hierarchy.
- Aggregating pre-requisites. In the case of intermediate or final parts of whole processes, knowledge on preceding sub-processes could be used either to check the pre-requisites on the learner side or to find additional background material for understanding the complete process.

There are other interesting relations in the GO that could be exploited for this case. Existing *isA* relations can be used to implement “examples first” tactics in sequencing instruction, which would typically go bottom-up from the leaves of the *isA* hierarchy. Also, domain-specific relationships are a source of interesting knowledge for instruction. Following the preceding example about the video *From RNA to Protein Synthesis*, learning resources annotated with positive regulations of translation (GO:0045727) could be used for instruction, having molecular-level regulation of processes as an objective. These regulations are represented in the GO explicitly with relations like *positively_regulates* or *negatively_regulates*.

In addition to relating the resources to the elements in the ontology, a different source of knowledge that annotators could easily provide are relations between different resources covering the same content. Following the example of video clips visualizing translation, the first one stays at a less detailed level than the second (which even provides chemical formulae for some elements). It is difficult to represent absolute levels of detail in IEEE LOM, since the interpretations of some metadata elements such as *DifficultyLevel* are tied to subjective interpretation. An alternative approach is to state a relationship like *(A moreDetailedThan B)* in the ontology representation. This relationship between learning objects is useful for loosely specifying partial orders of difficulty when searching resources for a given educational level.

### 4.3 Designing the annotation tool

Devising an annotation tool to capture different usages of media fragments requires capturing starting and ending offsets in existing video clips, along with a flexible way for creating metadata records connected to the ontologies of interest.
The prototype shown in Figure 2 is an Adobe Flex 3 graphical interface for YouTube videos. As users paste the identifiers of clips on the left hand side of the tool, they are automatically played. On the right hand side, the user can create a new learning object metadata record for the video in the left part of the tool. Clicking on the left buttons the user can set the exact start and end points of the segment being annotated. It is also possible to select previously created metadata records, and create contained records (that will be implicitly related by partOf). Then, the rest of the metadata can be edited using an ontology-enabled metadata editor such as SHAME\textsuperscript{11}. As long as editors like SHAME can be set up to work with existing ontologies by means of annotation profiles, the annotations related to the GO can be integrated in the ont-space by using the OWL version of the ontology.

5. Conclusions and future work

The exponential growth of user generated content available through Web 2.0 applications represents an opportunity for educators seeking for quality resources on the Web. However, neither sites collecting such resources apply quality control procedures nor educational value is one of their objectives. In consequence, educators would need to select and annotate those resources deemed appropriate for particular learning needs. This can be performed by creating metadata pointing to the original contents, an approach which results easy to use and almost cost-free. Considering economic reasons as the central target of our effort, we aim at making new educational resources available with negligible development costs. Adapting pre-existing materials through semantic annotation provides the ability of solving this issue in a fairly quick and effective manner.

The use of existing mature domain ontologies for creating that metadata together with repository technology supporting their storage enables reuse and plays the role of a filter. Furthermore, such technology enables the composition of multimedia presentations combining them, and the semi-automated composition of resources based on instructional

\textsuperscript{11} http://kmr.nada.kth.se/shame/
and domain knowledge represented in ontologies. This paper has described an approach for such purpose that can be used for any kind of Web 2.0 application that (i) provides permanent identifiers, e.g. permanent URIs, and (ii) allows multiple descriptions of the same or overlapping fragments and their combination. A case study in the domain of molecular genetics has been provided as a proof of concept of the possibilities of the technology developed.

Future work will investigate the connection between user reviews, comments and the structure of internal bookmarking in applications like YouTube with the educational quality of the resources. At first glance, there are many motives that lead users to positively rate or comment on a piece of content. Even if these reasons are almost not related to its potential educational usage, these ratings may serve as a first filter for finding potentially useful user generated content. Another interesting research direction for future work is to examine how the semantic coherence of social nets around content (Paolillo, 2008) could be exploited for educational purposes.

**Acknowledgements**

This work has been partially supported by project PERSONAL (Personalizing the learning process through Adaptive Paths based on Learning Objects and Ontologies), funded by the Spanish Ministry of Education - Project code TIN2006-15107-C02-01 and by project LUISA (Learning Content Management System Using Innovative Semantic Web Services Architecture), code FP6–2004–IST–4 027149.

**References**


General comments:

It is not clear which (if any) of the presented artifacts (ontological representation of the IEEE LOM standard, rules for defining constraints over the instructional process, the tool for annotations of YouTube videos) has actually been developed. The authors should make that more explicit. For example, even though the authors have mentioned that mappings of the IEEE LOM standard to ontology languages such as OWL, have already been defined and are available, it is not stated whether the authors are using some of those (preexisting) mappings or they have developed their own. If they are using one of the existing mappings, they should state which one; if not, they should explain why they have decided to introduce yet another mapping and what its advantages over existing ones are.

Similarly, regarding the rules presented in Section "Exploiting ontologies for instructional purposes", it is not clear whether the presented examples are just to illustrate the idea or they are part of a larger set of rules that have been developed and maybe even tested.

Added more information on this in sections 2 and 4.

Regarding the tool presented in the "Case Study" section: the end users acceptance of the suggested tool and the suggested form of interaction is highly questionable - dealing with an ontology editor is difficult for teachers and learning content authors as it is for any person without Computer Science (Artificial Intelligence, Knowledge Engineering) background. The suggested ontology-enabled metadata editor SHAME consists of huge forms that overwhelms users and they very rarely fill them up. The tool also assumes familiarity with the Gene Ontology, since without being familiar with this ontology an end user would have to spend a lot of time searching for appropriate concepts. All in all, I would not expect broad acceptance of this tool.

Broad acceptance of the tool has never been one of our objectives. We agree with the reviewer that SHAME-based annotation tools are not ready for the general public to use it. The ideas behind it, however, can be the germ for future efforts on the implementation of helpful interfaces, plugins that automatically suggest the right ontology or the right terms in an ontology according to the users previous inputs or history, etc. This is exactly what we aim for, opening a debate and providing fresh air on this field through the introduction of ontology annotations as a powerful and under-utilized way of improving media annotations.

Related work regarding semantic annotation of multimedia content is missing. Even though the field of multimedia annotation is not among my primary research interests, I'm aware of the research work been done in that field and the research efforts aimed at annotating media fragments. Attempts have also been made at automatic annotation of multimedia resources. For example, have a look at the following resources:

- the tool for video annotation developed in the scope of the BOEMIE (Bootstrapping Ontology Evolution with Multimedia Information Extraction) project (http://www.boemie.org/node/368)
- Also, Yahoo has recently launched its VideoTagGame (http://sandbox.yahoo.com/VideoTagGame/).
So, the authors should present relevant research in this field and compare their work to the existing solutions.

*We did so with the first 2 references provided. We tried yahoo TagGame but we can not link our research to a game where users receive more points for a tag if another player has entered the same tag close in time in a parallel annotation race. The relationship between what we report on and that game is very distant.*

The flow of narration should be improved, especially in section "Background and related work" where the authors frequently jump from one topic to the other without making connections between them and without sufficiently explaining some of the introduced topics.

*We did our best to clarify the text shortening a few long sentences and clarifying others. We hope the reviewer (and everyone else) find it easier to read now.*

**More specific comments:**

Page 2: Try to reformulate the text starting from the sentence: "In this paper, we focus only on the potential of UGC as a large base of contents..." till the end of that paragraph - the given statements are not clear. Specifically, it is not clear what are "preconditions to reuse of content" and the things mentioned in the points 1 and 2 are "two main reason" for what?

*Reviewer was right, it was unclear. Reformulated.*

Page 6-7: It would be useful to add a figure to section "Mapping continuous media fragments" in order to illustrate the ontology classes mentioned in this section and their relations. Likewise, I would suggest providing an ontological representation of one of the profiles from the example presented in Figure 1; that can be done in a manner similar to the one used for presenting rules in the subsequent sections (i.e. using some form of pseudo code), it is not necessary to use ontology language primitives.

*Done.*

Page 8, paragraph 1: 
"...ContinuousFragment. Then, profiles 2 and 3 are continuous fragments and represent..." Is it continuous fragment or continuous segment? Previously in that section the ContinuousSegment class was introduced and explained while the term fragment was not mentioned at all, so I suppose this is a mistake.

*Fixed. Thanks.*

Page 8-9: in the presented rules, lr namespace prefix was used - to which namespace it refers to?

*All prefixes removed for clarity. The example looks better now, as prefixes were not essential to the discussion. We did the same for other examples further down in the paper.*

Page 10, example of a rule:
"COMP_showsBefore(?a1, ?a2)
AVLearningActivity(?a1)
AVLearningActivity(?a1)..."

I suppose that in the third row "?a2" should be used instead of "?a1".

*You were right. Fixed, thanks.*
--------- Comments by reviewer 2 ---------

a) Technical comments:
a.1. The paper is difficult to read. There are many very long and ambiguous sentences. We did our best to clarify the text shortening some long sentences and clarifying others.

a.2. English is not good enough. It demands significant improvements. Again, we did our best here to improve it. We hope is good enough now.

a.3. There are some formatting issues to be solved. For example, it is difficult to distinguish between sections and subsections since they are not numerated and the font size is the same.
Sections now have numbers.

A.4.- The bullet lists use different bullet shapes (dashes, circles, squares), and so on. Fixed, now all bullets are circles.

b) Major comments:
b.1. The idea seems interesting and relevant to be investigated. However, deeper discussion on the motivation and how proposed solution will enhance learning process is necessary.
We further explained our motivation. In addition, the conclusions detail a little bit more of the advantages of our approach.

- Related work is focused on the annotation of learning objects while the annotation of media (video) content is almost not discussed. There are many approaches to the annotation of video clips on different levels: whole clip, segments, frames or even objects/figures in the frame. I agree with the author that most of them do not use ontology-based annotation schemes, as the one proposed in this paper, but discussion on them would enable comparative analysis and emphasize potential advantages of the approach.
Done, a few references and a paragraph on related work to video annotation has been introduced in section 1.

- The main contribution of this work, in my opinion, is the creation of the AVLearningObject concept that unifies the AudioVisual and LearningObject concepts by using the sub-classing mechanism. However, there is no schema (or ontology) definition of the AVLearningObject provided. Also, I think that once you have introduced this new concept, you could define some new properties of the concept as well. For example, you can define a property for linking video fragments to concepts of domain ontologies, instead of using some LOM elements, which are not primarily created for this purpose.
Using specific properties as suggested could be interesting from the point of view of flexibility and richness. Of course the definition of those properties would give us more possibilities in a semantic environment, but we didn’t want to forget that our annotations will not always be processed by semantic-engines so we need to provide compatibility with LOMv1.0. A mixed schema where AVLearningObject instances were annotated both through specific properties and through LOM-compliant elements was also devised, but it is still not developed. In fact the benefits of the former are still unclear. This choice is now included and justified in the text (section 3.3)
- The weakest part of the paper is a lack of the evaluation. The section you entitled as "case study" is the description of an annotation example, rather than the case study. Case studies provide a systematic way of looking at tasks, collecting information, analyzing information, and reporting the results. The case study should be conducted to show how your approach enhances the learning process (e.g., discoverability and reusability of learning content and so on).

_We agree on that. There is no way of conducting a proper case study under the current circumstances though. The lack of time and available users make very difficult to conduct it, to process the data and finally to write down the report to be included in our paper. However, we thought that our initiative was interesting enough as to be presented as is, even if it is not yet completely polished, and that the ILE readers would appreciate it. We hope the editors to share this vision with us._

- It is not clear what is the status of the prototype development. Is there the operable version of the tool? If so, it would be useful to provide the link where readers can try the tool?

_It is currently on development. Most features have been codified and tested, while some others, those of low priority, are still being polished and tested. We hoped to have it ready at the same time the paper was delivered, but we couldn’t finally get to the deadlines. It is currently planned as part of the development effort of the Organic.Edunet project (http://www.organic-edunet.eu/organic/index.html) so it will be released at the end of this year at the latest._

**Some comments/questions/suggestions for each section:**

1. **Introduction**

- The first paragraph contains several arbitrary interpretations of Web 2.0, which are not connected coherently. For example, Web 2.0 referees to behaviors, technologies, ideas, services, activities and so on. The last sentence of the paragraph is not justifiable since "the day of this writing" is unknown information for readers.

_Thanks, a good suggestion._

- In the second paragraph you say: "We feel it is extremely attractive to reuse the growing mess of UGC", Growing mess can not be reused, UGC is something what can be reused.

_Rewritten

- You defined the term UGC, and you should continue using it in the rest of the paper, instead of introducing one more term with the same meaning such as "user-contributed content".

_Nice suggestion, thanks._

- The bullet list items should contain/describe a number of characteristics for a given classification/categorization. The bullet list in the third paragraph does not contain information to be structured as a list of items. You should transform it into paragraph.

_Although you have some reason here, we prefer to keep it like that for the sake of legibility._

- The first sentence of the fourth paragraph: "considerable number" is unclear in this context.
Rewritten

- The last paragraph gives the outline of the paper, which is difficult to match with the sections of the paper. The exact titles of the sections do not appear in the outline.  

Rewritten according to new numbering

2. Background and Related Work

- Tagging media fragments with ontologies is rarely used term; more common are ontological annotation or ontology-based annotation.

Nice suggestion, thanks.

- There is missing reference for the SCORM model.

Included

- ALOCOM model is another LO model which could be relevant for the discussion.

We don’t think that speaking of Alocom would bring any additional information. It seems to be a very restricted and somewhat local model.

- What is WSML? Do you mean Web Service Modeling Language? If so, can you also reflect on its relations with the standard ontology language OWL, and why you have chosen WSML.

It is only mentioned in the text as an example of ontology language. Further explaining on WSML wouldn’t bring anything interesting in our opinion.

3. Semantic Annotation of Existing Continuous Media

- The requirements for semantic annotation (e.g., you listed three of them) of Web resources that are important for your approach should be deeply discussed. At some point you say that they will be discussed, but what follows are just two sentences.

Fixed

- ont-space project that you refer to is empty! Please, check the link!

Fixed

- There should be provided and discussed a formal definition (ontology or schema) of the AVLearningObject.

Provided

- The discussion of ontologies for instructional design seems a little bit inappropriate here; it could be moved to related work. Actually, I couldn't see anything new what your approach brings to that.

You are right; all the discussion on LD ontologies has been removed

4. Case Study

- Take a look on the comment about the evaluation, which I put in the major comments.

Done.