Cogn Comput (2012) 4:497–514
DOI 10.1007/s12559-012-9172-1

A Collaborative Video Annotation System Based on Semantic Web Technologies

Marco Grassi · Christian Morbidoni · Michele Nucci

Received: 16 August 2011 / Accepted: 21 July 2012 / Published online: 24 August 2012

Abstract In recent years, videos have become more and more a familiar multimedia format for common users. In particular, the advent of Web 2.0 and the spreading of video-sharing services over the Web have led to an explosion of online video content. The capability to provide broader support in accessing and exploring video content, and in general other kind of multimedia formats as images and documents, is becoming more and more important. In this context, the value of semantically structured data and metadata is recognized as a key factor both to improve search efficiency and to guarantee data interoperability. This latter aspect is critical to connect different, heterogeneous content coming from a variety of data sources. On the other hand, the annotation of video resources has been increasingly understood as a medium factor to enable deep analysis of contents and collaborative study of online digital objects. However, as existing annotation tools provide poor support for semantically structured content or in some cases express the semantics in proprietary and non-interoperable formats, such knowledge that users build by carefully annotating contents hardly crosses the boundaries of a single system and often cannot be reused by different communities (e.g., to classify content or to discover new relations among resources). In this paper, a novel Semantic Web-based annotation system is presented that enables user annotations to form semantically structured knowledge at different levels of granularity and complexity. Annotation can be reused by external applications and mixed with Web of Data sources to enable "serendipity," the reuse of data produced for a specific task (annotation) by different people and in different contexts from the one data originated from. The main ideas behind the approach are to build on ontologies and support linking, at data level, to precise thesauri and vocabularies, as well as to the Linked Open Data cloud. By describing the software model, developed in the context of SemLib EU project, and by providing an implementation of an online video annotation tool, the main aim of this paper is to demonstrate how such technologies can enable a scenario where users annotations are created while browsing the Web, naturally shared among users, stored in machine readable format and then possibly recombined with external data and ontologies to enhance end-user experience.

Keywords Semantic Web · Video annotation · Ontologies · Knowledge representation · Media fragment · Information sharing

M. Grassi (✉) · C. Morbidoni · M. Nucci
Semedia (Semantic Web and Multimedia),
Polytechnic University of Marche, Via Brecce Bianche I,
60131 Ancona, Italy
e-mail: m.grassi@univpm.it
URL: http://semedia.dii.univpm.it/

C. Morbidoni
e-mail: christian.morbidoni@gmail.com

M. Nucci
e-mail: m.nucci@univpm.it

Introduction

The advent of Web 2.0 and the success of online video-sharing services, like YouTube, have made of videos a popular content on the Web. The amount of videos uploaded and shared every day is constantly growing, and in a near future, videos are expected to constitute the main Web content. Such dizzying explosion of video contents on the Web poses some open challenges. Indeed, as it is happening for other kinds of digital resources, with the exponential growth of videos on the Web, traditional text-based algorithms for information retrieval are losing their efficiency and it becomes more and more difficult to find the required content among a mass of irrelevant or mildly irrelevant contents. This is all the more true dealing with video, considering that, although the relevant advances in content-based video indexing, their categorization and retrieval on the Web are still mainly based on high-level video information [1]. These are often limited to brief textual comments and tags about the whole video, as it is the case of videos uploaded on the Web by means of video-sharing services.

In these scenarios, collaboratively created resource annotations have proved to be a valuable source of metadata capable of enriching the original resource information and to be used to improve ranking, classification and retrieval of Web content and the overall user experience. The majority of the existing Web annotation systems are nowadays based on tagging. This approach is relatively easy to implement and very intuitive for users. However, it suffers from several issues related with the ambiguity of natural language since the searches are conducted at syntactic level (i.e., comparing strings), without relying on the underlying semantic (i.e., the meaning of the tag) [2]. This limits the accuracy and the efficiency of resource classification and retrieval, since:

- Tags can have ambiguous interpretations (polysemy). For example, the tag orange could be used to describe a fruit as well as a color.
- Different tags that identify the same concept (synonymy) or a subclass of it (specificity gap) would be considered completely different if analyzed only at syntactic levels.

Nevertheless, video resources play an important role also in several scientific researches, ranging from artificial intelligence to cognitive science. The availability of video dataset is in fact fundamental in many scenarios, for example, to train and test faces and shapes detection algorithms or to support studies of human cognition and emotions. In most of these cases, high-level video annotation is not sufficient. Videos need to be carefully manually labeled and annotated at fine granularity level to encode all the relevant information. Also, specific domain vocabularies have to be employed in the annotations to refer to precise agreed concepts [3]. This activity can require an elevated level of expertise and results expensive and highly time-consuming. Several software applications and annotation schemas have been developed in recent years to provide support to the annotation task [4], supplying advanced annotations capabilities and specific playback functionalities. These tools are still limited in the capability to support collaborative annotations and to create a shared structured knowledge that can be then used for research purposes: browsing, querying and reusing such a knowledge in other contexts and applications. Existing annotation tools provide in fact support for vocabularies and data schema, but the data model is not flexible enough to merge and exchange annotations. Annotations are usually encoded in ad hoc XML schema or maintained in proprietary data representations. Moreover, video integration on the Web is still limited and before the advent of HTML5 videos were mostly handled through the use of plug-ins, providing limited capabilities in particular to video fragment management [5]. As a result, most of the existing tools are desktop applications. Therefore, they do not integrate well with the place where data, information and social communities reside: the Web.

Over the years, the possibility to improve the interoperability and productivity of user-created metadata providing annotations with a clear structure and semantic has attracted a growing research interest. In particular, novel approaches and systems have started to be developed and experimented that leverage on Semantic Web (SW) technologies [6]. These allow the users in fact to:

- univocally identify Web resources and their fragments (video fragment included) founding on standard like URI and Media Fragment URI [5];
- effortlessly integrate data coming from heterogeneous sources relying on the flexible and unambiguous data model supplied by RDF [7];
- provide unambiguous machine interpretable semantics to information by using computational ontologies knowledge representation;
- reuse the giant and interconnected knowledge base provided by the Linked Open Data (LOD) cloud[1] to easily augment the original information, for example, using LOD named entities in the annotations;
- reveal the implicit information contained in data and to use it to enhance the efficiency of information and resource retrieval.

Overall, the SW technologies allow the users therefore not only to improve search efficiency but also to guarantee data interoperability. This latter aspect is indeed critical to connect different, heterogeneous contents coming from a variety of data sources. It is not a case that Europeana initiative[2] has recently adopted an RDF-based data model to aggregate independently provided digital contents.

The main idea under this work is that when properly structured and linked to LOD and ontologies, annotations can connect documents and media to the Web of Data

1 http://linkeddata.org/.

2 http://www.europeanaconnect.eu/.

allowing to reveal and explore new connections among media objects. This enables scenarios, as the one discussed in "A Simple Use Case", in which data produced for a specific task can be reused by different people and in different contexts from the one data originated from. Based on previous researches and developments in Semantic Web-oriented collaborative annotation systems (e.g., SWickyNotes[3]) and on the results of a survey about multimodal video annotation [8], this paper introduces a novel Web application for collaborative semantic video annotation. The proposed system permits annotations at different levels of granularity (allowing the annotation of temporal as well as spatial video fragments) and of complexity, supporting different types of annotations, ranging from textual comments to fully structured RDF statements. Moreover, a specific data and social model for annotation exchange are defined to foster the sharing of created annotations over the Web and among different user communities. A prototype implementation of the system is also presented, which is not intended to advance the state of the art with respect to video handling functions, but rather to demonstrate how the use of semantically structured data brings advantages in organizing, sharing and reusing collectively created knowledge.

This paper is organized as follows: "Related Works" provides a brief overview of main video annotation tools; "Background: The SemLib Project" introduces SemLib project within which this research has been conducted; "A Data Model for Semantic Annotations" and "Collaborative Annotation," respectively, describe the data and social model used for annotations encoding and sharing; "System Architecture and Implementation" discusses the proposed system presenting a fully working demonstrative prototype; "A Simple Use Case" provides a use case; finally, "Comparison with Other Systems" provides an evaluation of the system based on a comparison with main existing video annotation tools and on a user survey.

Related Works

Video annotation constitutes the starting point in many scientific researches, for example, in the study of emotion and of multimodality of human communication. In this section, we provide a brief overview of existing desktop and Web annotation tools and discuss the result of a survey that has been conducted about multimodal video annotation to analyze the applicability of Semantic Web technologies in this scenario [8].

[3] http://www.swickynotes.org.

Video Annotation Tools

Several specialized desktop applications have been developed in recent years to provide support to the annotation task in particular in the field of multimodal video annotation [9]. Anvil [10] software gives support for hierarchical multilayered annotations, visualization of waveform and pitch contour and offers an intuitive annotation board that shows color-coded elements on multiple tracks in time-alignment. ELAN [11] is a free linguistic annotation tool for the creation of text annotations for audio and video files. The annotations can be grouped on multiple layers or tiers and organized hierarchically. Each level of description is represented in ELAN on a tier. The EXMARaLDA [12] system consists of a data model, a set of corresponding XML formats and a number of software tools for the creation, management and analysis of spoken language corpora. The TASX Annotator enables an XML-based annotation of multimodal data on multiple tiers, which facilitates a parallel annotation as well as an immediate comparison across the different modalities of interest. MacVisSTA is a software program, for Mac OS, developed to code different aspects of behavior (speech, gaze, gesture, etc.).

Some of these tools also make use of Semantic Web technologies for video annotation. OntoELAN [13] inherits all ELAN's features and extends the tool with an ontology-based annotation approach. It can open and display ontology specified in OWL (Ontology Web Language) to create language profile and ontological tier. M-OntoMat-Annotizer (M stands for Multimedia) [14] is a user-friendly tool that supplies a graphical interface to load and process visual content (images and videos), to extract visual features and to associate these with domain ontology concepts. VIA (Video Image Annotation Tool) [15] allows users to import its descriptors from predefined OWL ontology files and to perform the annotation of specific video regions and enable the captivation of movement trajectories. SVAT (Semantic Video Annotation Tool) [16] enables film analysts to efficiently annotate video footage. SVAT also support a plug-in for object recognition and search tool based on Difference-of-Gaussian local key points and computed SIFT descriptors as local low-level features for every frame in the video.

Most of these software tools also allow the users to perform complex searches based on temporal and/or structural relations in the single generated annotation files. Some of them also support simple text searches across multiple files. However, all these software tools, being conceived as desktop applications, perform search queries only locally in the user machine. In addition, no handling for different users permissions is provided, and everyone can modify, read or create new annotations.

In recent years, as a result of the great spreading of videos over the Web, video content management has been increasingly supported by Web application and also tools for general purpose video annotation have started to be developed. YouTube itself [17] enables video uploaders to create textual annotations in the form of text bubbles or notes, also highlighting part of the screen and to make these annotations visible to all YouTube users when the video is played. VideoAnt [18] is a Web application that also uses YouTube as videos source, which allows the users to insert markers in video timeline and to associate textual annotation. The created annotations are sent by e-mail to be accessed also by other users. Project Pad [19] is a project to build a web-based system for media annotation and collaboration for teaching, learning and scholarly applications. Project Pad provides an open source Web application, distributed under GPL license, developed in Java and Flash available both as a standalone application and as part of Sakai. The application allows selecting video segments and creating textual annotations. It also provides a timeline visualization of the annotations. Kaltura Advanced Editor [20] provides several functionalities for online video editing, supporting timeline-based editing and video and audio layers. It allows the users to add soundtracks and transitions to import videos, images and audio while editing, and to add effects and textual annotation. EuropeanaConnect Media Annotation Suite (ECMAS) [21] is an online media annotation suite based on Annotea that allows users to extend existing bibliographic information about digital items like images, audio and videos. It includes a client application for video annotation. It allows users to select video segments adding a marker and to add textual annotation. EuropanaConnect video annotator also includes Semantic Web capabilities that enable users to augment existing online videos with related resources on the Web, provided by Linked Data Cloud. This augmentation happens on-the-fly while the users are writing their annotations, the application proposes in fact related resource derived from DBpedia [22], a semantic compliant version of Wikipedia. User has to verify the semantic validity of a link or to disambiguate between eventual homonyms before they become part of the annotation. Such linked resources can be then exploited in the underlying search and retrieval infrastructure. Inside the EU IM3I[4] project, an integrated web system for video search and annotation has been developed [23], which consists in three main components: an ontology-based search engine (Orione), a search interface (Sirio) and web-based video annotation tool (Pan). The video annotation tool allows users to browse videos, insert or modify annotations, selecting concepts from an ontology structure. Ontologies can be imported and exported using the MPEG-7 or OWL ontology or created from scratch by users. Annotations are allowed only for the entire video, and it is not possible to select video fragments.

Currently, between the existing Web video annotation tools, only ECMAS and Pan exploit some of the possibilities offered by Semantic Web technologies. In particular, ECMAS only allows the management of video fragments but without following the Media Fragment URI standard, which limits the interoperability and accessibility of the annotated fragments. In addition, apart from Kaltura editor, existing applications for video annotation provide quite poor interfaces and limited performances in comparison with existing desktop tools, not supporting, for example, drag and drop functionality.

A Survey About Multimodal Video Annotation

In a previous work [8], we conducted a survey about multimodal video annotation, a particularly interesting use case of video annotation due to the wide number and complexity of created annotation, between members of COST 2102,[5] a research action about the Cross-Modal Analysis of Verbal and Non-Verbal Communication. Sixteen members with experience in multimodal annotation applied to the survey. Rather than comparing in detail the performances of the existing tools and annotation schemas, the main purpose of the survey was to inquire about the actual state of the art in multimodal video annotation and its possible enhancements, particularly through the application of SW techniques.

According to the survey results, freely available on the Web, [6] used annotation tools' functionalities (mostly ELAN and Anvil) are satisfactory in annotations creation while it could benefit from some improvement in annotation searches to enhance their accuracy. Also users authentication and management are not supported by currently used tools, which limit the possibility of collaborative annotations. In fact, survey result has also revealed that there is no specific support for annotation and video sharing. Annotation is stored in local files and is exchanged mostly using mass storage or e-mail exchange. Neither centralized databases nor video-sharing services are exploited to improve the accessibility of video and annotation leveraging on the Web infrastructure.

Survey participants, most of whom had no knowledge about SW initiative, were also inquired about the applicability of SW technologies to support the process of multimodal video annotation creation and sharing. In particular, they were asked to evaluate a simple demonstrative Web

[4] http://www.im3i.eu.

[5] COST 2102 Cross-Modal Analysis of Verbal and Non-Verbal Communication http://cost2102.cs.stir.ac.uk/.

[6] Multimodal Video Annotation Survey http://www.semedia.dii.univpm.it/survey/videoannotation.

site,[7] based on the Exhibit[8] framework to implement the faceted browsing paradigm for videos and annotations visualization and navigation from a RDF file. All of them found the proposed demo Web site an interesting way to display and manage the annotated video segments and agreed that the use of SW techniques could be beneficial in the field of multimodal video annotation.

Background: The SemLib Project

The proposed video annotation system has been developed inside the SemLib project,[9] a 24-month project founded by European commission that aims to improve the current state of the art in Digital Libraries (DLs) and Web resources annotation through the application of SW technologies for data representation and management.

One of the main expected outputs of SemLib project is the design and implementation of a general purpose annotation system able to enrich and interconnect digital objects published on the Web, specifically targeting DLs and multimedia archives available through the Web. Also, SemLib aims to provide tools and models capable of leveraging the process of collaborative and community-driven annotation of DLs items, from the simple social engagement factor to a source for creating/generating added value, as envisioned by the crowdsourcing paradigm [24].

With such purposes, in SemLib an on-purpose annotation model and an annotation-sharing model have been defined as described in "A Data Model for Semantic Annotations" and "Collaborative Annotation," which has also been adopted in structuring and organizing the annotations created with the proposed video annotation system. Also, the SemLib annotation server, described in "Annotation Server," has been used for annotation storage.

A Data Model for Semantic Annotations

Annotations are pieces of information that a user attaches to a resource by adding some knowledge that is not explicit in the resource itself. Such added knowledge can later be exploited by the user herself or by others. A tag, for example, can be used for retrieving a resource based on some criteria (e.g., classification, related topic). A comment posted by a user that links to a related video might be useful for another user to discover new interesting contents. Also, contextual information assumes a particular significance in dealing with annotations. Attributes that specify the creation date, the author and other metadata are again useful for users to organize, share or reject annotations. Allowing computers to better understand annotation content and context represents a key factor to improve browsing, searching and consuming content. It is an understood fact that, in order to enable this, annotations should be semantically structured. Once relations among annotated object (as well as annotation context) are properly represented, computers can take advantage of them and applications can behave in smarter way. As users are accustomed to interact with data and with each other in a Web environment, having a web-based annotation system and making annotations available on the Web is of primary importance. Linked Data and SW are used in the presented system to provide such an interoperability layer, which, basing on W3C standards like RDF [7] and SPARQL [25], effectively allows completely decoupled applications to reuse annotations and their content. In other words, to enable collaborative annotation in online communities, the system must produce machine readable annotations with explicit semantics.

In this chapter, the data model used in the proposed system for representing annotations, semantically encoding both their context and their informative content, is described.

Annotation Metadata (Context)

Open Annotation Collaboration (OAC) [26] and other annotation ontologies, like Annotea [27] and Annotation Ontology [28], provide an understood model for representing annotation context and to link annotations to the resources they refer to. Figure 1 provides an example of how an annotation is represented using the OAC model.

Among the contextual information included in the annotation context, relevant ones are those specified by the `aoc:hasTarget` and `oac:hasBody`. The first one specifies what resources are involved in the annotation, so that a software can retrieve all the relevant annotations given one or more resources. In some cases, such targets are not entire media objects, but rather fragments (e.g., a temporal segment of a video). In this case, according to the OAC specifications, the system bases on Media Fragments [5]. This specification by W3C provides a standard format for univocally identifying and accessing fragment of images, audio and video resources on the Web through a URI introducing three main parameters for resource fragment identification: time, space and track.

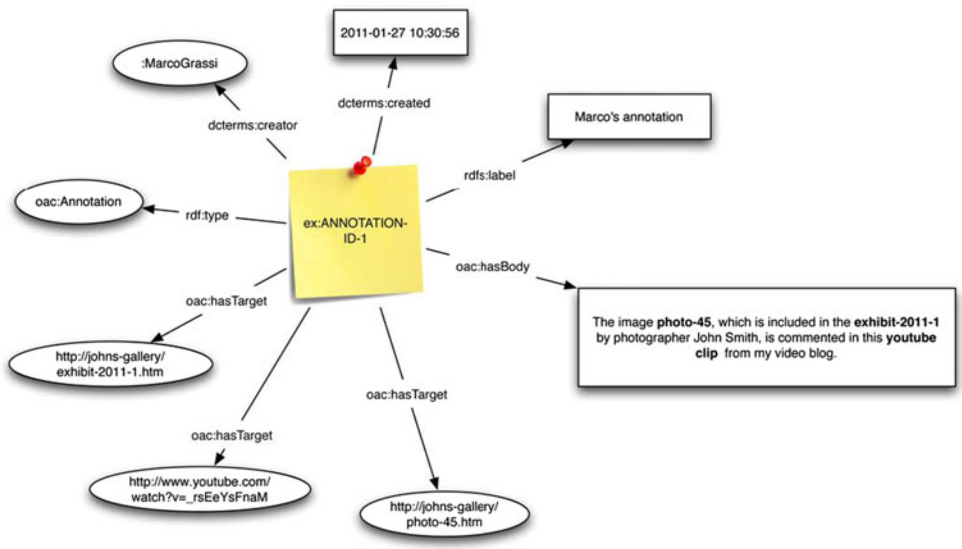
In Fig. 2, the representation of a video segment is shown. The `dc:isPartOf` property specifies the corresponding entire video from which the segment has been taken. This allows the system for retrieving all the annotations that have as target a segment of a specified video.

[7] Video Exhibit http://videoemotion.altervista.org/videoemotion.html.

[8] Simile Exhibit: http://www.simile-widgets.org/exhibit/.

[9] SemLib project: Semantic Web Tools for Digital Libraries.

Fig. 1 A simple annotation represented using the OAC model



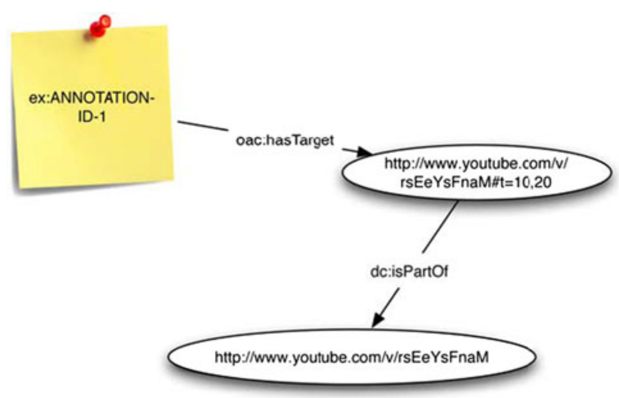


Fig. 2 The representation of a video segment

The `oac:hasBody` property is used to link the actual content of the annotation, which is where semantic information expressed by users resides. While the cases where such a body is a web document or a textual comment are extensively documented in the AOC specifications, there is still discussion going on about how representing bodies that contain semantic graphs themselves. As, one can see in Fig. 1, the body of an annotation might contain semantic relations among the annotated resources, but without Natural Language Processing or entity extraction algorithms, it is impossible for a machine to meaningfully interpret such semantics and effectively use it.

Informative Semantic Content

The main idea behind the proposed annotation system is that of representing the annotation body as a RDF graph itself, where semantic relations among the resources are made explicit and encoded in a machine readable way. At user interface level, it will be shown later in this paper how users are enabled and supported in the process of creating semantic annotations by grouping and relating different kind of resources. From a data representation viewpoint, to model this situation the system is based on named graphs. A named graph is a sub-graph of an RDF graph that can be identified by a URI. This allows the graph to contain triples (statements) about the named graph itself. Figure 3 illustrates the use of named graphs to embed semantic statements expressed by users. In this example, two bodies are associated with the annotation. The first one is a textual comment, as in the previous examples, and the second is a named graph and contains RDF triples whose semantics is structured. Using named graphs, which are fully supported by the SPARQL query language, allows users to perform queries on a single annotation content or, more interestingly, to group annotations with specified characteristics and query the sum of their semantic content. For example, one could query for all the annotations whose target is a specific video and whose author is one (or more) specific user and then extract all the resources that "comments" the video according to the selected annotations.

While the OAC ontology is used to represent contextual information, the semantic content cannot be represented based on a fixed ontology. Different users communities operating in specific domains need specific shared vocabularies (ontologies) of terms and relations that they can use in annotations. Interoperability of semantic content within a given community depends on the agreement on common entities and relations identification. At RDF data storage level, the system is agnostic with respect to the domain ontologies used in structuring annotation informative semantic content, and specific configuration at application level can be used to build an ad hoc vocabulary for each

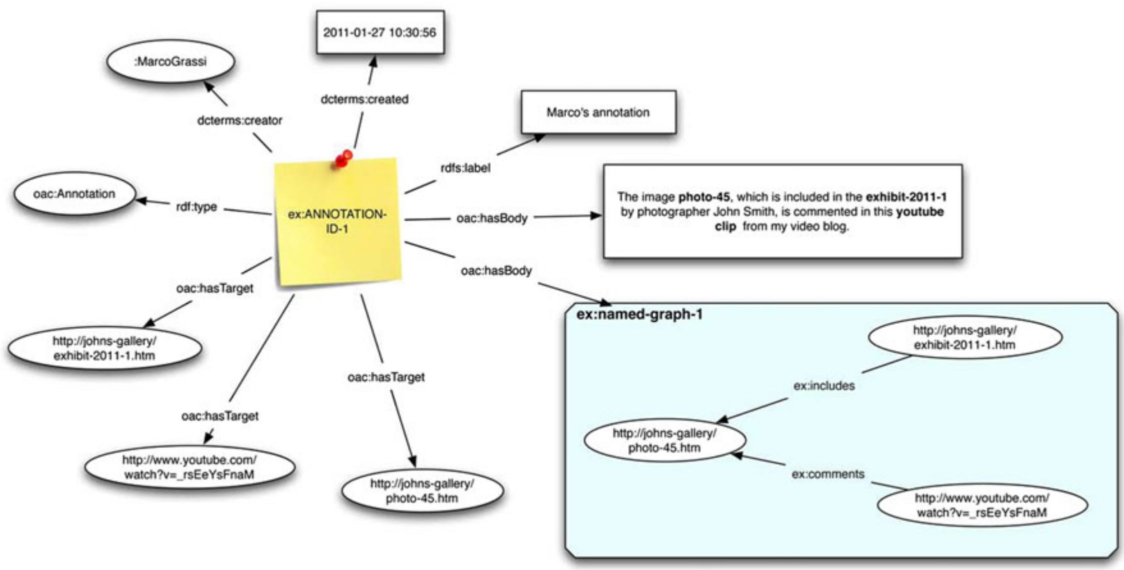


Fig. 3 The representation of an annotation using a named graph

community addressed. In "SemTube: Semantic YouTube Video Annotation" a specific domain, the annotation of human emotions, is considered as a case study for the prototype implementation. It is worth to notice, however, that different annotations using different domain ontologies are still interoperable as they base on the RDF data model and can be merged. In this case, however, if the different ontologies are not aligned, performing meaningful queries could be problematic. For example, if the same entity is identified by different URLs in different ontologies, then the system will treat the two entities as distinct ones.

Notebooks

As a convenient way of collecting annotation, the system uses the concept of notebooks. Notebooks are resources that put together a set of annotations so that they can be retrieved and queried as previously discussed. By default, each user has a proprietary notebook where all her annotations are collected. Notebooks can have read/write privileges and can be used for giving users control over her annotations, for example, sharing the notebook with others (e.g., assigning read privileges to a specific user). Notebooks can also be open, allowing read/write to all users. This is the case, for example, of the prototype discussed later in this paper.

In Fig. 4, the RDF representation of a notebook including two distinct annotations is illustrated. This picture does not show all the data and semantic relations among entities, but rather focuses on the graph path that a query engine would follow to dynamically collect all annotations in a given notebook and, more interestingly, retrieve and merge all the RDF statements asserted by users in such annotations (their semantic content).

RDF data retrieval can be performed through SPARQL query language. The example query shown in Listing 1 returns all the named graphs whose semantic content is stored given a notebook URL:

Listing 1 An example query returning all the named graph given the notebook URL

```
SELECT DISTINCT ?graph
WHERE {
  <http://swickynotes.org/notebook/resource/d3369cd2> <http://purl.org/
   swickynotes/ao/v1#includes> ?annotation .
 ?annotation <http://www.openannotation.org/ns/hasBody> ?graph .
```

Results from this query are then used in other subsequent queries that retrieve and merge the actual RDF data, as shown in Listing 2.

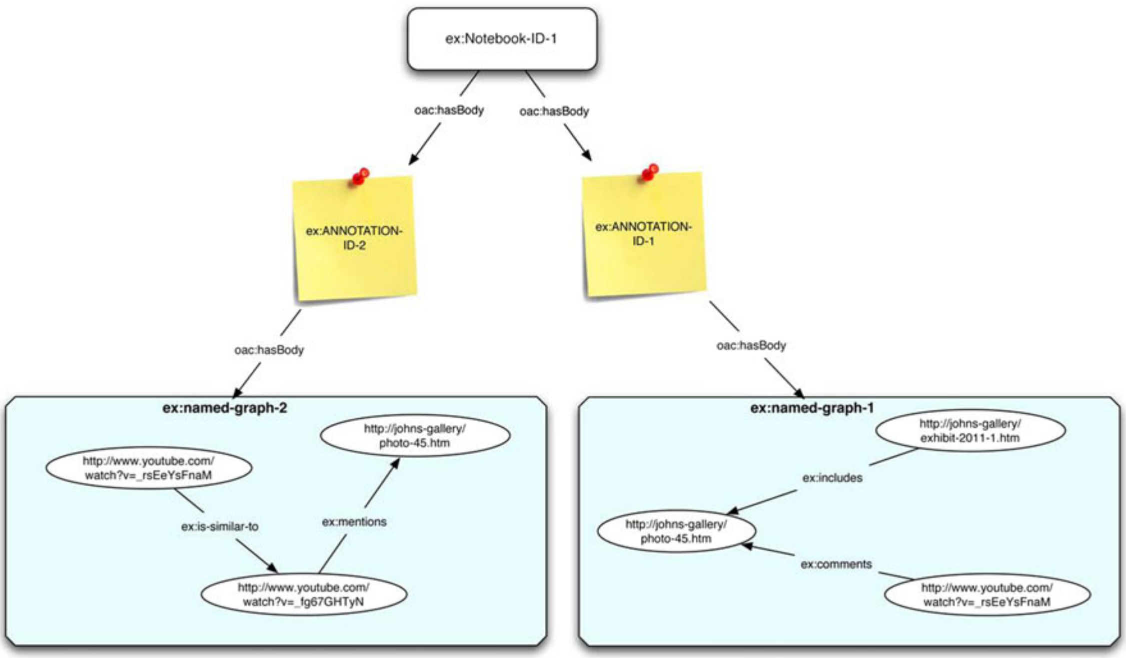


Fig. 4 The RDF representation of a notebook including two distinct annotations

Listing 2 An example query for returning results in the form of an RDF graph

```
CONSTRUCT { ?s ?p ?o }
FROM <http://swickynotes.org/notebook/resource/graph-56af4f30>
FROM <http://swickynotes.org/notebook/resource/graph-a76db5a9>
WHERE { ?s ?p ?o }
```

where `http://swickynotes.org/notebook/resource/graph-56af4f30` and `http://swickynotes.org/notebook/resource/graph-a76db5a9` are the named graphs returned by the first query. The CONSTRUCT command allows users to get the result in the form of a RDF graph encoded in RDF/XML syntax.

Collaborative Annotation

In SemLib, notebooks have a central role in collaborative annotation. In particular sharing a notebook with other people is similar to sharing a web calendar or a collaborative document. The invited users will see annotations from friends or colleagues while browsing annotated videos. As sharing a notebook is as easy as sharing a URL to issue a web service call, existing social media can also be used to share or open notebooks to the public.

Providing support for collaborative annotations has proved to be an enabling feature for scholars to actually benefit from the digital world in their everyday work. Experiments conducted within the Discovery[10] European project have shown that building structured information by annotating Web documents can be a valuable mean of representing aspects of the study process, for example, in e-learning or classroom activities. In [24], authors make a distinction between social engagement, where users annotate contents for their own purposes (e.g., to better organize study resources), and crowdsourcing, where social engagement is used within groups of users (communities) to achieve a shared goal by working collaboratively together as a group. In Digital Libraries, metadata crowdsourcing refers to the ability of leveraging users annotations to improve the digital archives, for example, by building personalized recommender systems (as the SemLib project aims to do) or simply collecting and selecting structured metadata that can drive browsing and searching capabilities. Semantic Web Digital Libraries (DLs) have already been developed, and the support for RDF metadata is growing even in main stream channels, as the BBC[11]. By providing RESTful

[10] ECP 2005 CULT 038206 project, EC eContentplus programme.

[11] http://www.w3.org/2001/sw/sweo/public/UseCases/BBC/.

APIs and standard SPARQL endpoints to retrieve and query annotations in the form of RDF graphs, the system enables this scenario by giving applications (e.g., DLs) the ability of easily importing metadata produced by user communities. As an example, annotations from known (or trusted) users that involve a set of known web media objects can be harvested periodically with a simple script.

For simplicity, the client prototype application later discussed in this paper, whose primary goal is to demonstrate how annotations can be created from web pages, makes use of base configuration of the SemLib annotation sharing system, where all users own a public notebook that is by default visible to all users, which are logged into the system.

System Architecture and Implementation

The presented annotation system has a client–server architecture, as shown in Fig. 5, where the SemLib annotation server provides RESTful APIs for creating, managing and retrieving data and the SemTube prototype client integrates with users web browser and handles semantic video annotation authoring. In this chapter the main characteristics of the two components are discussed. Successively, the prototype is demonstrated in a specific annotation domain.

Annotation Client

The client-side component comprises a set of sub-modules developed in Javascript using Dojo[12] framework to facilitate cross-browser support, to implement graphic effects and widgets and to ease the implementation of client–server AJAX communication. The client-side module implements the graphical user interfaces to create and browse annotations as well as modules dedicated to the communication with the server-side components.

Player The annotation functionalities of the proposed system require the use of a player with additional capabilities other than the common playback, in particular making possible the selection and visualization of temporal video fragments. With such purpose a Javascript custom player has been developed that provides an abstract interface for all the required functionalities and implements a special YouTube player. YouTube has been chosen as a video source considering that it is currently the most popular video-sharing service on the Web, with more than 3 billions video watched daily and more than 48 h video uploaded every minutes.[13] Also, YouTube provides the JavaScript Chromeless Player APIs[14] that gives full control over YouTube videos. Using such APIs a custom player has been developed, which in addition to the common playback functionality allows users to select frame and video for annotation. In particular, a custom video progress bar has been created in Javascript, allowing to place markers for video fragment selection.

Fragment Handler This component is responsible for the selection of both temporal and spatial video fragments. For spatial fragment selection, an on-purpose Javascript library has been developed called *CanvasSelector* that allows users to select regions of web pages and pages elements, as images and videos, and can be integrated in any application. The CanvasSelector allows superimposing a transparent Canvas, a new element introduced by HTML 5 for supporting drawing and animation, to selected DOM elements of the Web page. Such Canvas is used as a drawing mask for region selection. Different selection shapes can be chosen as rectangle, ellipses and polygons, using different fill colors and alpha levels.

The *Fragment Handler* returns the Media Fragment URI of the selected video fragment. Currently, Media Fragment URI only allows the identification of rectangular regions. Therefore, the standard has been extended introducing an additional parameter *shape* to identify the kind of geometric region. In the case of polygonal shape, the parameter vertex has been introduced to identify the pair of (x,y) coordinates of the polygon vertex.

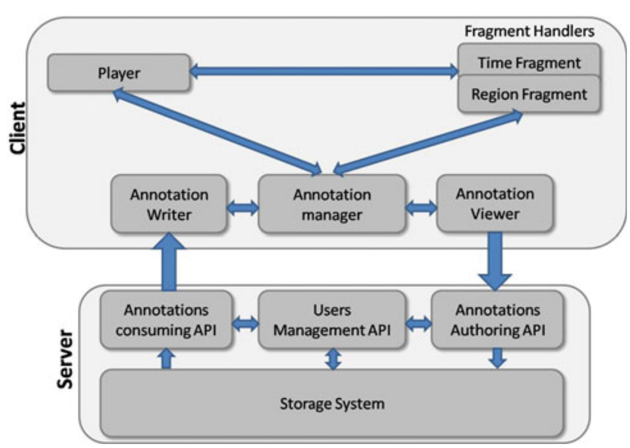


Fig. 5 Simplified architecture of SemTube video annotation system

[12] http://dojotoolkit.org/.

[13] http://www.youtube.com/t/press_statistics.

[14] http://code.google.com/apis/youtube/getting_started.html#player_apis.

Listing 3 An example of an extended Media Fragment URI representing an elliptic region inscribed inside a rectangle with upper corner in (10,10) having width and height respectively equal to 30 and 15 belonging to a video fragment, starting at 10 s and ending at 20 s of a youtube video, whose id is *ORqru0hOa5k*

```
http://www.youtube.com/v/ORqru0hOa5k\#t=10,20\&shape=ellipse\&xywh
  =10,10,30,15
```

Named Entity Selector This component is responsible for retrieving named entities using Freebase Reconciliation Service[15], which takes a query and attempts to return records that best match it using string distances and probabilistic matching.

Vocabulary Selector This component is responsible for the management of the vocabularies used for the annotation. Dojo Tree Widget has been used to create a tree-based visualization of the taxonomical organization of concept inside a vocabulary.

Item Manager The Item Manager is responsible for the local storage and the management of the selected items (selected video fragment, Freebase and vocabulary items, etc). In particular, it is responsible for retrieving and visualizing the created items that are relative to the current video or page to allow user to easily reuse these without needing to selecting them again. Items can also be filtered according to their type: as video and video fragment, relations and entities. Also, this component offers management for recent and favorite items, respectively, memorizing the last created items and allowing the user to save favorite items. These items are stored locally using HTML5 localStorage (when supported by user browser) or in alternative using cookies.

Annotation Manager The annotation manager is responsible for annotation creation and editing. An on-purpose graphic interface has been developed that allows user to create RDF triples by drag&dropping resources (video and video fragment, predefined property, ontology and vocabulary items and literals) in three boxes representative of the subject, the predicate and the object of the statement. Range and domain of the properties and resource type are controlled to check acceptance of dragged items in order to ensure consistency of the created statements.

Annotation Writer The annotation writer is responsible for sending the created annotations to the server Annotation Authoring API using AJAX-based communication and a specially conceived JSON format for data encoding. This is similar to the one used in the TALIS project[16] and allows users to map RDF triples in JSON allowing faster processing in JavaScript.

Annotation Visualizer This component is responsible for showing the created annotations, by communicating with Annotation Consuming API of the Annotation Server. Different types of visualization and filtering are allowed. In particular, a dedicated panel allows users to explore the annotations relative to the current video that have been created by the current user as well as public annotations created by other users. In addition, a visualization based on the faceted browsing paradigm is used for showing information contained in a user notebooks, based on Exhibit framework[17]. The faceted classification allows the assignment of multiple categories to an object, enabling the classifications to be ordered in multiple ways, rather than in a single, predetermined, taxonomic order. This makes possible to perform searches combining the textual approach with the navigational approach, enabling users to navigate a multidimensional information space by concurrently writing queries in a text box and progressively narrowing choices in each facet dimension.

Annotation Server

The Annotation Server is a modular web service designed and developed using the REST paradigm. It provides annotations storage, users authentication and management in addition to APIs for annotations authoring, consuming and sharing.

The storage module defines a completely generic interface, designed to support different kind of storage systems ranging from traditional relational databases to NoSQL databases (e.g., RDF triplestores). In the prototype version, the storage is implemented using the Sesame triplestore[18], but it can be easily replaced with other data storage systems (e.g., relation databases) developing new dedicated plug-ins.

The annotation server RESTful API allows users to create new notebooks and annotations supporting different data formats (e.g., RDF, JSON, etc.) and to browse notebooks and related annotations, providing advanced searching and query facilities based on Semantic Web technologies. Each notebook is provided with a dedicated SPARQL endpoint. The annotation server queries the data model, described in Listing 4 to perform the queries only to those named graphs that contain semantic data enclosed in annotations included in the notebook. Methods are

[15] http://data.labs.freebase.com/recon/query.

[16] Talis Platform http://www.talis.com/platform/.

[17] http://www.simile-widgets.org/exhibit/.

[18] openRDF.org http://www.openrdf.org/.

provided for searching annotations. As an example, an API call for all the annotations mentioning a specific video (and all its fragments) is given in Listing:

natural language processing and artificial intelligence techniques to related video comments [33, 34].

Listing 4 API call for all the annotations mentioning a specific video (and all its fragments)

```
/notebooks/$notebook-id/search?query={''resources'': [''http://youtube.com/
   v/15tyUJ98'']}
```

As one can see, JSON (properly encoded) is used as a syntax for passing parameters, in this case specifying the video by its YouTube URL.

The annotation system has been designed to work in a distributed cross-domain context. Scripts running in random web pages should be able to communicate via AJAX with remote components. In order to overcome the limitation of the same-origin policy that prevents cross-domain AJAX request, CORS standard [29] has been implemented using Tomcat's CORS filter[19].

SemTube: Semantic YouTube Video Annotation

A demonstrative prototype of the video annotation system has been implemented[20] that allows users to semantically annotate YouTube videos. Video from YouTube can be loaded in SemTube page both copying their URL in an textbox input and using a special bookmarklet that can be dragged on the browser favorites. Clicking this bookmark while watching a video on YouTube redirects to SemTube page automatically loading the video.

In order to prove the applicability of the proposed tool for affective information encoding and multimodal video annotation in general, HEO ontology [30] is provided between the default domain ontologies. HEO provides a description model for human emotions that grants at the same time enough flexibility to describe all the main features of an emotion, by allowing the use of a wide and extensible set of descriptors, and interoperability, by allowing to map concepts and properties belonging to different emotion representation models. HEO offers a wide set of classes and properties for describing emotions both in terms of category, dimension and other affective-related concepts. Together with standardized ontologies for Web resources and people description as Dublin Core Ontology [31] and FOAF (Friend of Friend Ontology) [32], HEO provides a complete framework for encoding affective information of generic Web resources. For this reason, HEO has already been used to encode affective information automatically extracted from YouTube videos by applying

SemTube player GUI allows users to select entire videos, temporal video fragments (by adding a temporal marker) and video fragment regions (by drawing shapes over the video fragment). Annotations can be created using two different GUIs. Comments and semantic tags can be added using the Comment/tags panel. This interface allows the user to type a comment and to automatically extract tags from it using Dbpedia Spotlight[21] service. User can remove suggested tags that are not considered relevant or add others using Dbpedia Lookup[22] service. In addition, the "Advanced Annotation Editor" bar allows users to edit annotations using drag&drop to populate statements in the form of subject-object-predicate with previously selected resources. Different types of resources can be used. The entire video, temporal video fragments and video fragment regions can be selected using the player interface. Named entities can be retrieved from Freebase. Entities and relation can be chosen from vocabularies and ontologies, which can be loaded at run time. Also literal can be edited. An on-purpose tab is provided by the GUI for each type of resource. Recently selected items are stored using HTML5 localStorage, when supported by the browser, or in alternative using cookies. Also selected resources can be added to favorites and stored to the server to be displayed also in different pages for the one in which they have been selected. This allows the creation of cross video annotation, for example, to state that fragments from different video are similar.

In Fig. 6a screenshot shows semantic annotation creation and visualization in SemTube. In this example, a video fragment region has been selected and some annotations added. A textual comment has been typed in the Comment/Tags panel, and some tags have been extracted from it using Dbpedia Spotlight. The selected fragment has also been annotated using one of the relations offered by the HEO ontology, `heo:showsEmotion`, and one of the accepted values of such property, `heo:ArchetypelAnger`, has been selected from the controlled vocabulary (based again on the HEO ontology). The resulting statements say that the video fragment shows a well-defined emotion type, heo:ArchetypelAnger, which belongs to a specific emotion sets (`heo:ArchetypalEmotion`) included in the ontology, and is in relation

[19] CORS Filter: http://software.dzhuvinov.com/cors-filter.html.

[20] SemTube: Semantic YouTube Video Annotation: http://semedia.dii.univpm.it/semtube/.

[21] DBPedia spotlight, http://spotlight.dbpedia.org/demo/index.html.

[22] DBPedia lookup, http://dbpedia.org/lookup.

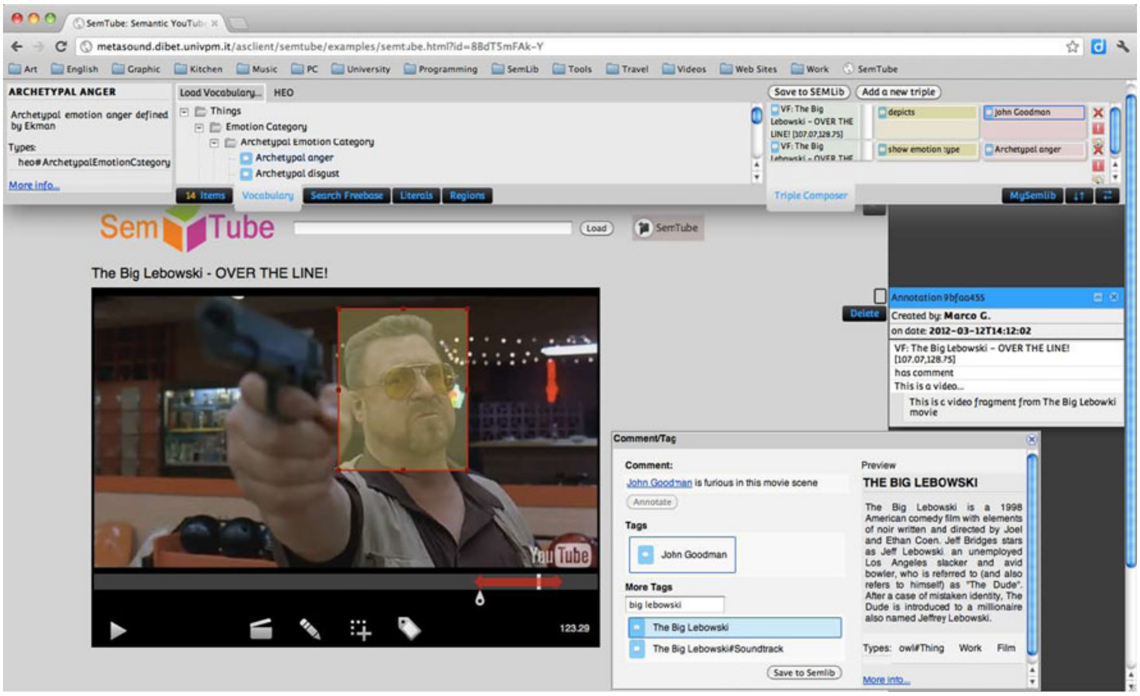


Fig. 6 A screenshot of SemTube demo application

to other emotion categories defined in the ontology. Similarly, the user has provided other information about the person that is highlighted in the video region. In this case, the property `semlib:depicts` has been selected, and the value (the actor John Goodman) has been searched into Freebase via the dedicated tab. The result is a statement that can be browsed to reach the Freebase representation of the entity. For humans this is a semantic link that can help in browsing and understanding the content. For machine such a link can be automatically followed to retrieve RDF data from Freebase and possibly process it to learn additional facts (e.g., John Goodman's date of birth or movies with John Goodman). Such additional information can be used to discover new relations between user-annotated contents, as described in "A Simple Use Case."

The "Annotation Visualizer" bar allows to explore all the annotations associated with the current video. In the current demo implementation, all the created annotations are stored in a public notebook and can be visualized by every user that is logged into the system. Annotations are refreshed every time that a new one is created and become immediately available to other users. In this way, the system makes annotation sharing completely transparent to the users and allows them to collaboratively annotate videos.

In addition, a demonstrative prototype has been set up, using SIMILE Exhibit APIs, that allows to explore videos and their annotations combining faceted browsing paradigm with text-based search,[23] as shown in Fig. 7. Videos and their annotations can in fact be filtered both using the search box, to perform keyword-based queries, and by adding or removing constraints on the facet properties. In this way it becomes very easy and intuitive to accomplish non-trivial queries, such as "find all the video related to John Goodman that show an emotion of type anger." In addition, since RDF is used to store information, which allows to automatically merge information coming from different sources about the same entity univocally identified by its URI, user-created annotations can be effortless augmented with external data. In this way both user-added properties and new properties extracted on-the-fly from DBPedia are presented in a uniform way and can be used to filter results and browse graph. As a proof of concept, in the demo, videos can be filtered not only according to their tags but also according to tag categories provided by Dbpedia. This allows users, for example, to search not only "all the video fragment that show a certain type of emotion" (as explicitly stated by users) but which have been "interpreted by someone who is a tv actor" (as discovered in Dbpedia).

A Simple Use Case

In this section, a simple use case is presented as proof of concept of how the proposed system, leveraging on accessibility of the web architecture, relying on the use of standardized domain vocabularies to provide semantics to annotations

[23] http://metasound.dii.univpm.it/asclient/semtube/examples/files/exhibit/videoexhibit.html.

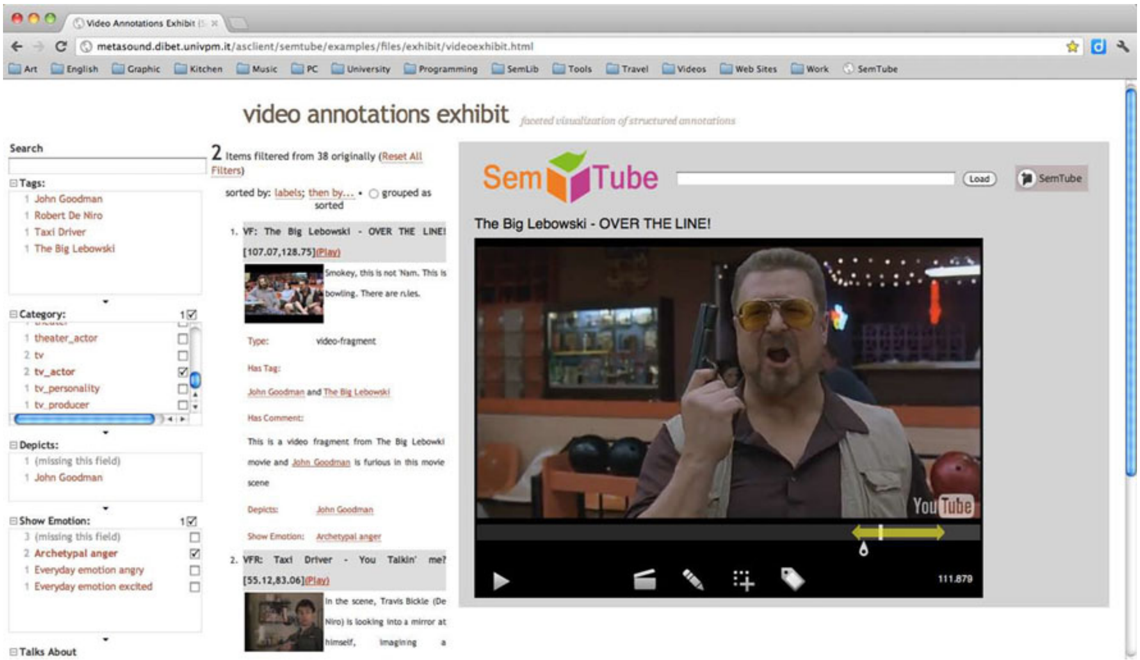


Fig. 7 Exploring annotated videos using faceted browsing

and exploiting the knowledge provided by the LOD can be beneficial in fostering annotation sharing and reuse.

Alice is a psychologist interested in human emotions. She is watching videos on YouTube and finds an interesting one to show it in a conference presentation as an example of the emotion ”rage.“ She decides to annotate it with Semtube and uses the Comment/Tag panel to add a comment about the video that includes the name of playing actor (John Goodman) and the movie from which the scene has been taken (The Big Lebowski). When she clicks on the annotate button, such entities are recognized and added as tags. She also loads the emotions vocabulary that allows her to express a precise relation (`heo:showsEmotion`) between the annotated video fragment and one entity (`heo:ParrotSecondaryEmotionRage`) from the vocabulary.

the domain ontologies used in the annotations, which provides semantic glue to connect different annotations.

A second user, Bob, is looking for samples to train an algorithm of speech emotion recognition. His algorithm aims to discriminate between the six basic emotions defined by Ekman. He searches in the Web site where Alice annotation is also contained to find annotated video fragment that he can use. Since the ontology used in annotation (and contained in the repository) defines `heo:ParrotSecondaryEmotionRage` as a narrower category of `heo:ArchetypalAnger`, which has “anger” as `rdfs:label`, searching with the keyword “anger,” Bob can also find the video fragments annotated by Alice. He also finds some entire videos that have been entirely tagged with emotion terms and the playing actor. He uses the annotation system

Listing 5 An excerpt of Alice annotation

```
:v-fragment-1 semlib:hasTag <http://dbpedia.org/resource/John_Goodman >,
  <http://dbpedia.org/resource/The_Big_Lebowski>
:v-fragment-1 heo:showsEmotion heo:ParrotSecondaryEmotionRage
heo:ParrotSecondaryEmotionRage rdfs:label ‘‘rage’’
```

Alices video annotations are collected and merged with other video annotations, created for different purposes by a virtual community of users, into a Web site exposing a RDF-based faceted browser that allows users to navigate and search the videos. The annotation repository also contains all

to refine the previously created annotations by selecting and annotating the exact fragment of video in which the emotion has a manifestation. Between those, he annotates a video fragment from the Taxi Driver movie in which Robert De Niro interprets anger.

Listing 6 An excerpt of Bob annotation

```
:v-fragment-2 semlib:hasTag <http://dbpedia.org/resource/Robert_De_Niro
  >, <http://dbpedia.org/resource/Taxi_Driver>
:v-fragment-1 heo:showsEmotion heo:ArchetypalAnger
heo:ArchetypalAnger rdfs:label ‘‘anger’’
```

Charlotte is conducting a research about emotion expression in different cultures. In particular, she wants to find evidences that confirm the universality of the six basic emotions by Ekman and to find examples of emotions that are expressed differently in different cultures. The repository in which Alice and Bob annotations are contained automatically loads the respective Dbpedia semantic data each time an annotation mentions a Dbpedia resource. In our case the system has imported the RDF representations of both Robert DeNiro and John Goodman. Charlotte can now explore a new relation between the videos annotated by Bob and Alice, as the system knows that both the mentioned persons belong to the category `yago:AmericanFilmActor` and have as `dbpedia:birthPlace` a US city, as shown in Fig. 8. So, Charlotte can filter the video fragments according to both the emotions represented and the nationality of the actor playing. In this way, she can, for example, compare video fragments in which the same emotion is played by American actors and Japanese ones.

Evaluation

Comparison with Other Systems

Table 1 proposes a comparison between SemTube and four other existing tools, widely used in research domains as human emotions annotations. Different features have been taken into account, as video playback and video fragments selection, annotation expressivity and complexity and annotations search and visualization capability. Also, particular attention has been payed to aspects related to data interoperability and annotation sharing.

Compared with existing desktop video annotation tools that count on several years of experience and development, SemTube provides more basic capabilities in terms of video playback, video segment selection and visualization and at the current state of development does not provide support for multiple video annotation tracks. However, compared with the existing Web tools, SemTube provides more advanced management of video fragment region,

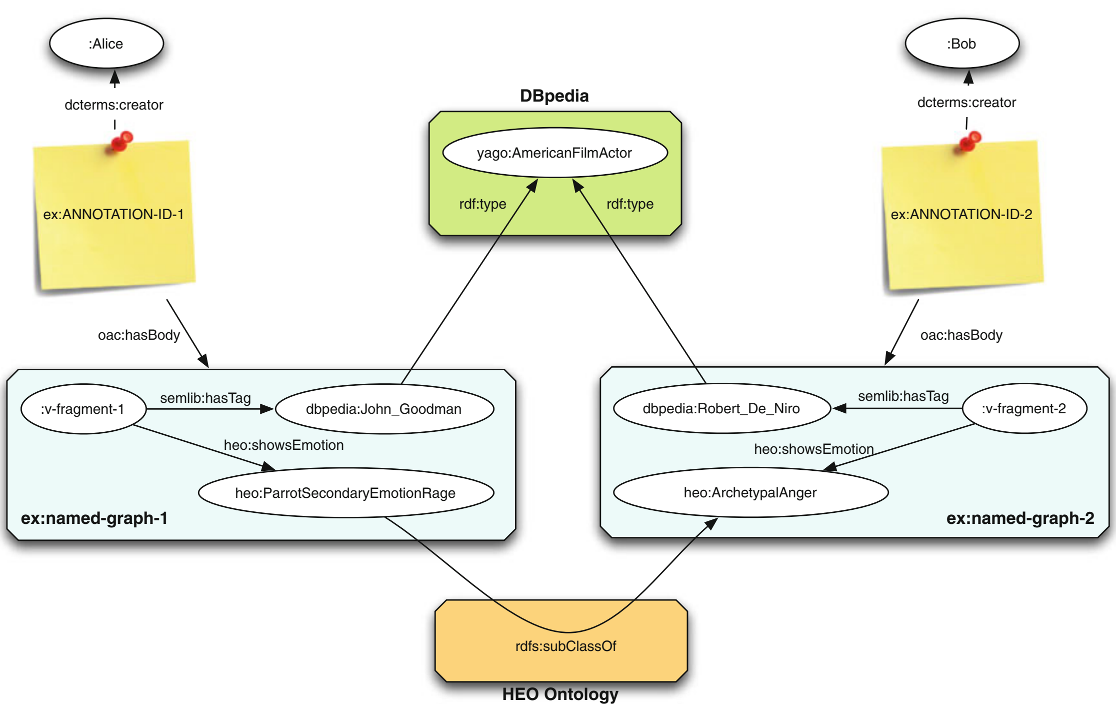


Fig. 8 Using Linked Data Cloud and HEO ontology to discover new data relations

Table 1 Feature comparison of video annotation tools

	SemTube	Anvil	ELAN	ECMAS	Sirio-Orione-Pan
Representation of annotation metadata (context)	RDF/OWL, OAC data model	XML	XML	RDF	Information not available
Representation of knowledge expressed by annotations	RDF/OWL ontologies	XML schema vocabulary, plain text	XML schema vocabulary, plain text	Plain text, semantic tags, geo tagging	OWL ontologies, MPEG-7, geo tagging, plain text
Semantic expressivity/complexity of annotations	Free-text, (semantic) tags, attributes, relations, ontology based	Free-text comment, attributes, relation	Free-text comment, attributes, relation	Free-text, semantic tags	Information not available
Support for controlled annotation vocabularies	Custom vocabularies, taxonomies, relations and constraints on relations	XML schemas vocabulary	XML schemas vocabulary	No	OWL ontologies
Vocabulary/schema configurability	JSON configuration file	Load controlled vocabulary from file or URL/possibility to edit the vocabulary	Load XML schema vocabulary from file or URL/possibility to edit the vocabulary	No	Possibility to import and edit ontologies in OWL and MPEG-7
Data storage	Dedicated annotation server based on Sesame TripleStore (inference)	Local files	Local files	Dedicated annotation server based on Mongo DB	Annotation server
Annotation grouping and collections	User personal notebooks, flexible selection via API/SPARQL	No	No	No	No
Data import/export and annotation reuse in other applications	Linked Data and W3C (RDF, SPARQL), RESTful APIs	Import/export in different XML-based format	Import/export in different XML-based format	RESTful APIs. RDF exports	Possibility to import/export annotations in MPEG-7 and OWL
Support for temporal fragment selection	Basic support for time fragment selection and multiple fragment visualization	Advanced support for time fragment selection and multiple visualization	Advanced support for time fragment selection and multiple visualization	Basic support for time fragment selection	No
Support for video region selection	Support for different shapes	Under construction	No	No	No
Annotation visualization	Contextual visualization of the annotation for each video, facet browser visualization	Time aligned visualization of the annotation grouped in different layers	Time aligned visualization of the annotation grouped in different layers	Contextual visualization of the annotations for each video	Contextual visualization of the annotations for each video
Annotation search	Keyword-based and faceted search (enriched by Linked Data content). Semantic SPARQL-based searches	Keyword-based search on single and multiple files	Keyword-based search on single and multiple files	Keyword-based search (enriched by Linked Data content)	Keyword-based and queries with Boolean and temporal operators
Annotation sharing	Collections of annotations can be made public or shared by a dereferenceable URL	No	No	No	No
Integration with the Web	Web-based	No (Desktop)	No (Desktop)	Web-based	Web-based

relying also on Media Fragment URI standard to identify selected region. Based on a semantic graph data model, SemTube supports the higher level of expressivity in annotation content. This makes the user interaction more complex to handle (e.g., one needs to build statements and not simply type a tag), but the complexity seems acceptable in scenarios where researchers are used to complex annotations and needs to encode non-trivial data. For less experienced users, a more intuitive comment/tags GUI has been provided.

The main focus on SemLib project is to enhance the creation of annotations and to foster their reuse by external applications, providing REST APIs and SPARQL endpoints to consume created data and metadata. In this vision, content and annotation search is mostly demanded to external applications. However, as proof of concept of the potential of exposing structured data with a well-defined semantic, a demonstrative web application has been developed that provides advanced search capabilities implementing faceted browsing paradigm. In addition, de facto standard technologies like Apache Solr[24] could be easily applied to index content for the RDF graph. Experimental technologies as Siren,[25] which are designed to build semantic aware high-performance indexes over RDF repository, are planned to be investigated to provide such functionalities. In addition, leveraging on the notebook organization of the annotation, it is possible to query only restricted subsets of the created annotations. At the current state, only Sirio-Orione-Pan provides more advanced search functionalities relying on a dedicated search engine. Search granularity in this case is limited to entire video and not to video fragment like in SemTube.

SemTube is also the only tool that has been conceived to provide specific support for annotation sharing, relying on the mechanism of notebooks to aggregate relevant information and make these available to both other users and third-party applications by means of a dereferenceable URI, to be easily consumed leveraging on the uniform model provided by RDF and on the clear defined semantic provided by the use of pluggable ontologies.

User Evaluation

In order to evaluate the performances and the usability of the prototype and plan future developments, a survey[26] has been conducted over a set of 12 people: four with expertise in multimodal video annotation, four with expertise in SW technologies and Web development and four non-technically skilled users. Participants have been asked to watch a demonstrative video[27] and to perform some basic operations, that is, to select a video fragment and a video fragment region and to create some annotations using both Comment/Tag panel to add a textual comment and some tags and the "Advanced Annotation Editor" to create full subject-predicate-object statements.

Most of the users have found video fragment (both temporal and spatial) selection easy and intuitive. Over 75 % of them have indeed performed selection in less than 20 s and consider the time required operation fast enough. Comment/Tag panel has been conceived as a simplified annotation GUI that could also be used by users with no particular experience in annotation or in Semantic Web technology. Survey results seem to confirm that the purpose has been achieved: More than 83 % of the users consider it intuitive and easy to use and the average time required to create an annotation containing a short textual comment and two tag retrieved using Dbpedia Lookup service of 32 s. The time required for the annotation creation has been evaluated as fast enough by 91 % of the participants. The same is not true for the Advanced Annotation Editor. In this case, user's answers differ remarkably and mostly depend on their experience. In fact, if users that are familiar with Semantic Web or with the use of complex annotation tools (as the ones used for multimodal video annotation) evaluate positively tool usability, 16 % of the users consider the triple editor very difficult to use. The average time required to edit a single statement is of 31 s. Even if the complexity appears to be acceptable for experienced users, these results also seem to suggest that ad hoc interfaces should be developed for single use cases to fasten complex annotations editing and hide the complexity of triple model to non-technically skilled users.

All the participants agree that the proposed tool can improve significantly (50 and 25 % of the participants, respectively, answer much and very much) the productivity of their annotation process in particular making transparent the annotation sharing with other users. User has also been asked to discuss which of the feature provided by the tool was more interesting. Also in this case, not surprisingly, answers vary according to participant expertise: Users with experience in multimodal video annotation have found very useful the support for ontology in the annotation, users with knowledge of Semantic Web have found interesting annotation linkage with the Linked Data Cloud, and users with no experience in these field have found useful the possibility to tag region, a functionality that is becoming common with images but is not for video yet.

Participants have also been asked to choose between some additional functionalities to implement. The most required feature (56 %) is the handling of different user

[24] http://lucene.apache.org/solr/.

[25] http://siren.sindice.com/.

[26] http://www.3medialabs.org/survey/index.php?sid=73483&lang=en.

[27] http://www.youtube.com/watch?v=e1-DV1A%5Ftlw.

permission in accessing created annotations, which is expected to be implemented in the following tool development. Other highly required features, in particular by users with experience in multimodal video annotations, are a time-aligned visualization of the annotations (50 %) and a wider timeline to make temporal video fragment selection more accurate (33 %).

Overall, 72 % the participants have affirmed that they would use the tool once it reaches a more mature state of development.

Conclusion and Future Efforts

In this paper, a novel video annotation system has been introduced, which relies on SW technologies allowing semantic annotations of temporal and spatial video fragments with different levels of complexity and structure. The employment of such technologies together with a data and social model developed on purpose is expected to foster video resource management on the Web and to enhance collaborative video annotation in different application scenarios.

A fully working prototype implementation of the system, for the semantic annotation of YouTube videos, has also been presented discussing the experimental client-side GUI for annotation creation and the server's RESTful APIs for annotation storage, sharing and consumption. A comparison with existing tools and a user survey gave clues that the application of Semantic Web technologies and a Web 2.0 annotation tool like SemTube could bring advantages in a collaborative scenario, by producing interoperable data and fostering annotation sharing and querying. The high expressivity of semantics annotations makes so that user interaction is less intuitive than in general purpose annotation tools.

Several challenges are still open in improving annotation creation/visualization, by introducing novel GUIs and enhancing the ease of the use of existing ones, and annotation sharing, extending server capabilities, which will be tackled in future releases of the annotation system.

Acknowledgments The research leading to these results has received funding from the European Union's Seventh Framework Programme managed by REA-Research Executive Agency (http://ec.europa.eu/research/rea) ([FP7/2007-2013][FP7/2007-2011]) under grant agreement no. 262301. This work is also supported by COST 2102, EUCogII and SSPNET.

References

1. Hu W, Xie N, Li L, Zeng X, Maybank S. A survey on visual content-based video indexing and retrieval. IEEE Trans Syst Man Cybern Part C Appl Rev. 2011;41(6):797–819.
2. Andrews P, Zaihrayeu I, Pane J. A classification of semantic annotation systems. Technical Report DISI-10-064, Ingegneria e Scienza dell'Informazione. University of Trento. 2010.
3. Grassi M, Morbidoni C, Nucci M. Semantic web techniques application for video fragment annotation and management. In: Proceedings of the SSPnet-COST 2102 PINK international conference on analysis of verbal and nonverbal communication and enactment: the processing issues. 2011. p. 95–103.
4. Dasiopoulou S, Giannakidou E, Litos G, Malasioti P, Kompatsiaris Y. A survey of semantic image and video annotation tools. In: Paliouras G, Spyropoulos CD, Tsatsaronis G, editors. Knowledge-driven multimedia information extraction and ontology evolution, 6050/BOEMIE EU project. Berlin: Springer; 2011.
5. Media Fragments URI 1.0. W3C Working Draft 24 June 2010. [Online]. Available: http://www.w3.org/TR/media-frags/.
6. Morbidoni C, Grassi M, Nucci M. Introducing SemLib project: semantic web tools for digital libraries. To Appear in: International workshop on semantic digital archives—sustainable long-term curation perspectives of cultural heritage held as part of the 15th international conference on theory and practice of digital libraries (TPDL). 29.09.2011 in Berlin.
7. Manola F, Miller E (eds). RDF primer W3C recommendation 10 February 2004. Available: http://www.w3.org/TR/REC-rdf-syntax.
8. Grassi M, Morbidoni C, Piazza F. Towards semantic multimodal video annotation. Lect Notes Comput Sci. 2011;6456/2011: 305–16.
9. Rohlfing K et al. Comparison of multimodal annotation tools—workshop report. Gesprachsforschung-OnlineZeitschrift zur verbalen Interaktion 7. 2006;7(7):99–123.
10. Kipp M. Anvil—a generic annotation tool for multimodal dialogue. In: Proceedings of the 7th European conference on speech communication and technology (Eurospeech). Aalborg; 2001. p. 1367–70.
11. Wittenburg P et al. ELAN: a professional framework for multimodality research. In: Proceedings of the 5th international conference on language resources and evaluation (LREC 2006). p. 1556–9.
12. Schmidt T. Transcribing and annotating spoken language with EXMARaLDA. In: Proceedings of the LREC-workshop on XML based richly annotated corpora. Lisbon 2004. Paris: ELRA; 2004.
13. Chebotko A et al. OntoELAN: an ontology-based linguistic multimedia annotator multimedia software engineering. In: Proceedings of the IEEE sixth international symposium on 13–15 Dec. 2004. p. 329–36.
14. Petridis K et al. MOntoMat-Annotizer: image annotation. Linking ontologies and multimedia low-level features in KES 2006—10th international conference on knowledge based, intelligent information and engineering systems.
15. Video Image Annotiation (VIA) Informatics and Telematics Insitute (CERTH-ITI). Available: http://mklab.iti.gr/via/.
16. SVAT Semantic Video Annotation Tool by JOANNEUM RESEARCH. Available: http://www.joanneum.at/en/fb2.html.
17. YouTube Video Annotations. Available: http://www.youtube.com/it/annotations_about.
18. VideoANT. Available: http://ant.umn.edu/.
19. Project Pad. Available: http://dewey.at.northwestern.edu/ppad2/.
20. Kaltura Video Editing and Annotation. Available: http://corp.kaltura.com/video_platform/video_editing.
21. Haslhofer B, Momeni E, Gay M, Simon R. Augmenting Europeana content with linked data resources. In 6th international conference on semantic systems (I-Semantics), September 2010.
22. Bizer C, Lehmann J, Kobilarov G, Auer S, Becker C, Cyganiak R, Hellmann S. DBpedia a crystallization point for the web of data. J Web Semant Sci Serv Agents World Wide Web. 2009;7: 154–165.

23. Bertini M, D'Amico G, Ferracani A, Meoni M, Serra G. Sirio, orione and pan: an integrated web system for ontology-based video search and annotation. In: Proceedings of the international conference on multimedia (MM '10). New York: ACM. p. 1625–8. doi:10.1145/1873951.1874305. http://doi.acm.org/10.1145/1873951.1874305.
24. Holley Rose Crowdsourcing: How and Why Should Libraries Do It?. D-Lib Magazine, The Magazine of Digital Library Research. March/April, 2010.
25. Sparql query language for rdf, W3C Recommendation, January 2008. [Online]. Available: http://www.w3.org/TR/rdf-sparql-query/.
26. Sanderson R, Van de Sompel H (eds). Open annotation: Alpha3 data model guide 15 October 2010. Available: http://www.openannotation.org/spec/alpha3/.
27. Kahan J, Koivunen MR. Annotea: an open RDF infrastructure for shared web annotations. In: Proceedings of the 10th international conference on World Wide Web. 2001. p. 623–32.
28. Annotation Ontology http://code.google.com/p/annotation-ontology/.
29. Cross-Origin Resource Sharing. W3C Working Draft 27 July 2010. Available: http://www.w3.org/TR/cors/.
30. Grassi M. Developing HEO human emotion ontology. LNCS, vol 5707. Berlin: Springer; 2009. p. 244–51.
31. Dublin Core http://purl.org/dc/elements/1.1/.
32. FOAF Vocabulary Specification 0.98. Namespace Document 9 August 2010. Available: http://xmlns.com/foaf/spec/.
33. Grassi M, Cambria E, Hussain A, Piazza F. Sentic web: a new paradigm for managing social media affective information. Cognitive computation. Berlin: Springer. In press. doi:10.1007/s12559-011-9101-8.
34. Cambria E. In: Roelandse M, editor. Sentic computing: techniques, tools and applications. Berlin: Springer. In press.