A Social Network for Video Annotation and Discovery Based on Semantic Profiling

Marco Bertini, Alberto Del Bimbo, Andrea Ferracani, Daniele Pezzatini
Università degli Studi di Firenze - MICC
Firenze, Italy
bertini,delbimbo@dsi.unifi.it, andrea.ferracani,daniele.pezzatini@unifi.it

ABSTRACT

This paper presents a system for the social annotation and discovery of videos based on social networks and social knowledge. The system, developed as a web application, allows users to comment and annotate, manually and automatically, video frames and scenes enriching their content with tags, references to Facebook users and pages and Wikipedia resources. These annotations are used to semantically model the interests and the folksonomy of each user and resource in the network, and to suggest to users new resources, Facebook friends and videos whose content is related to their interests. A screencast showing an example of these functionalities is publicly available at:
http://vimeo.com/miccunifi/facetube

Categories and Subject Descriptors

H.3.5 [**Information Storage and Retrieval**]: Online Information Services; H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Algorithms, Design, Experimentation

Keywords

Social video tagging, internet videos, social video retrieval

1. INTRODUCTION

The proliferation of multimedia contents occurred with the emergence of Web 2.0 has required effective systems for annotation in order to enable users to search and browse huge collections of data. Tagging of multimedia content has become a common facility of many sites that offer functionalities for multimedia sharing. Flickr and Facebook for images, YouTube and Vimeo for videos, have popularized tagging practices among their users. These user-generated tags are used to retrieve multimedia content and to ease browsing and exploration of media collections also exploiting social mechanisms, e.g. reminding to some user that he was tagged by a friend. However, not all media are equally tagged by users: currently in almost all social web applications is easy to tag a single photo, and even tagging a part of a photo, like a face, has become a common practice in sites like Flickr and Facebook; instead, tagging a video sequence is a more complicated and time consuming task, so that users just tag the overall content of a video. Also tags are not equal: a common practice used in social networks like Twitter is to differentiate tags that refer to some person, e.g. using the @ sign, or tags that refer to some topic or conversation, e.g. the "hashtags" marked with the # sign, that are used to highlight and group documents or conversations. Moreover, the accuracy of tag annotations is completely dependent on users, that often use incomplete or even wrong terms because of lack of knowledge of the domain of the resource to be tagged, or because the process is completely manual and not enough assisted. In recent years the development of the semantic web and, in particular, the so called social semantic web has tried to overcome this issue, giving emphasis to the formal correctness of the annotations. In this demo we suggest a method for semantic video annotation that exploits the dynamics of social networks and of user generated content: annotations are used as a mean that allows users to expand their knowledge as well as their social network. Web 2.0 applications have shown that user friendly, easy to use, rich internet applications coupled with strategies like the use of games and competition [11] or systems based on reputation and community membership [5], like Wikipedia or StackOverflow, stimulate participation to tasks that are human-centric. This may allow to enrich unstructured media content [14] exploiting social networks and the Web of Data.

Recent advances in the computer vision and multimedia scientific communities have greatly improved the performance of methods for automatic annotation of visual content. The TRECVid benchmark has shown an increasing improvement in the performance of appropriately trained classifiers [4, 13]. Several methods have been proposed and they can be divided mainly in: *i)* supervised methods, where a set of classifiers is trained to detect scenes, objects and events, typically using methods based on the Bag-of-Visual-Words (BoVW) approach [12,15]; *ii)* unsupervised methods that exploit the plethora of user generated annotations of multimedia content to annotate images [7, 10] and more recently also to suggest and localize tags in video shots [1,2,6]. However, the performance of these systems allows to deploy them in a semiautomatic context, along with tools for manual annotation so that professional users can create semantic annotations based on ontologies [3]; even so, tools made to be deployed to general users have to reduce the cost of learning the use of ontologies such as LSCOM [8], designed for use in professional or scientific contexts, to be substituted

WWW 2012 Companion, April 16–20, 2012, Lyon, France.
ACM 978-1-4503-1230-1/12/04.

by folksonomies created without need of explicitly specifying the relations between the concepts. Moreover, there is no need to force users to produce thorough annotations of the video content. As noted in [9] analysis of the queries made to web-scale image search services are related to "unique searches" often composed by named entities, indicating a high level of specificity in image searches, particularly in the entertainment domain; also the majority of the searches were classified as being related to the "Entertainment" category, comprising movies and music. Attention to this domain and to celebrities has also been reported in a usability study made at Yahoo! Search[1]. These studies show that end users are less interested in annotations of scenes and objects that are commonly addressed by the current automatic annotation systems.

To cope with these issues and to ease the creation of semantic tags, we propose a system for social-based video annotation that allows manual annotation of resources and people from Wikipedia and Facebook, providing also automatic extraction of entities and topics from user comments and linking them to Wikipedia resources. On the one hand this enables the discovery and browsing of videos according to the interests of users, on the other hand it generates new forms of social interaction that promote further tagging of multimedia content, with the goal of improving and encouraging semantic annotations generated by users. The discovery of new materials that may be of interest for each user is obtained either directly, through the analysis and categorization of his manual annotations (added in the comments using the Facebook and DBPedia APIs), or implicitly, suggesting resources and videos whose annotations, either created by other users or extracted automatically from their comments, belong to the same categories. Thus the annotations define the user's own interests: the more he annotates the easier it is for him to discover new things and related videos, increasing also his social influence by spreading his interests and videos in the social network. In addition the application automatically generates semantic profiles of people and resources, based on all the annotations, so that the knowledge base is expanded by the activity of the social network. Profiles and resources pages are represented in RDFa format. An overview of the system workflow is shown in Fig. 1.

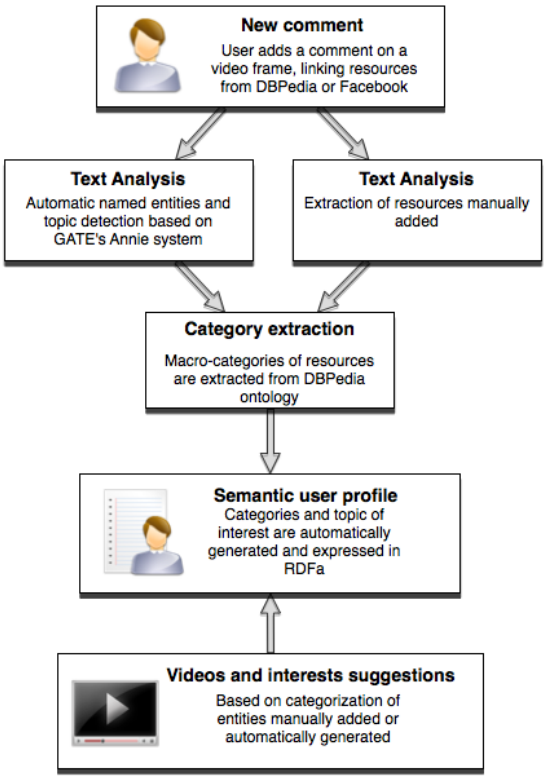

Figure 1: System workflow.

2. THE SYSTEM

The user interface has been developed in PHP and Javascript, using the jQuery framework, except for the video player that has been developed in Actionscript using the Flex framework. The backend has been developed in PHP, using the Codeigniter framework and the Facebook APIs and authentication services. Information regarding users, their social graph and their "likes" is obtained using the Facebook Query Language (FQL) and the Open Graph API. Users can upload videos (currently limited to 50MB) in the following supported formats: MPEG (up to H.264), Flash Video, Microsoft AVI and WMV, Google WebM and Apple Quicktime. Videos are then transcoded to Flash Video Format on the server side using FFMpeg. The PHP-based XoomStream server has been used to serve videos, without requiring the deployment of a full video streaming server. Video thumbnails are created in correspondence with user annotations using FFMpeg. Annotations are stored in two MySQL databases mapping the RDF triples using ARC2 RDF library for PHP[2].

Users can tag resources within comments from both Facebook and Wikipedia using the so-called status tagging feature: typing the @ character in the comment input field they can obtain the list of their friends in the social graph, whilst entering # users can retrieve, using the DBPedia API, a list of Wikipedia pages whose name matches the typed characters. Fig. 2 shows an example of annotation using a DBPedia resource.

To reduce the tagging effort required to a user and enrich the semantics, the system performs text analysis to identify potentially interesting tags. Named entities detection is based on the GATE/Annie system[3] and recognizes persons, organizations, places and dates. User annotations are also processed with LDA to identify topics. All these keywords are used to query DBPedia to provide the annotations with links to Wikipedia pages and categories.

Visualization and frame-accurate annotation of videos are facilitated by a timeline jQuery widget which allows intra video navigation by dragging horizontally the bar on which are shown the individual frames containing annotated comments. The timeline has two levels of precision to scroll through the video respectively every second or every five seconds. Each frame presents two icons of different color that indicate the number of annotations from Facebook and Wikipedia retrieved in the thread of comments. An example

[1]http://www.ysearchblog.com/2010/10/25/insights-into-multimedia-search-user-behavior-intent-and-consumption/

[2]http://arc.semsol.org/

[3]http://gate.ac.uk/ie/annie.html

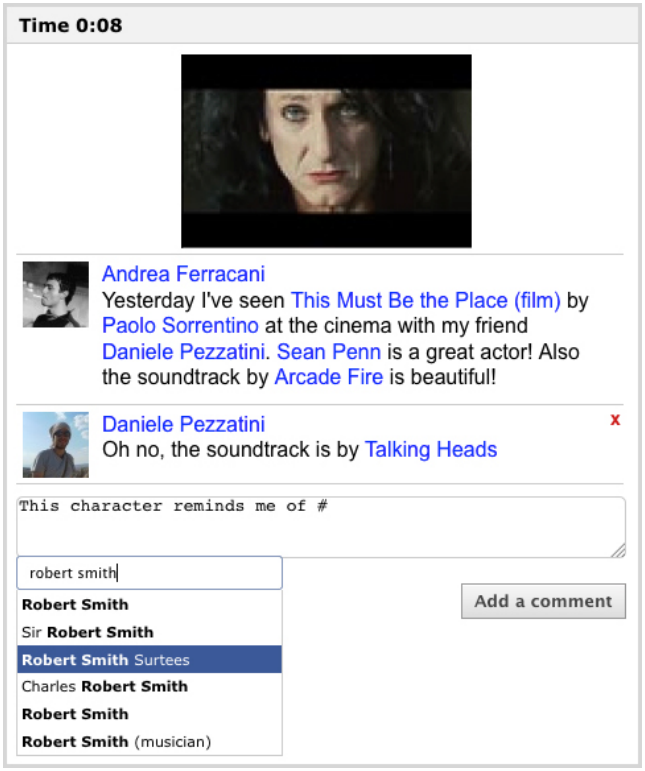

Figure 2: Example of the user interaction and annotation using a DBPedia resource. The terms highlighted in blues are Wikipedia resources and Facebook users manually annotated.

of the video player interface is shown in Fig. 3.

The recommendation system of videos is based on the analysis of four types of annotations: *i)* information extracted directly from the user Facebook profile through the Open Graph API, considering in particular the interests and "likes" of pages (this type of information is particularly helpful for suggesting interesting videos during the first access to the system); *ii)* links to Wikipedia or Facebook pages inserted manually by users in their comments; *iii)* manual annotations of other users, that belong to the same categories of interest; *iv)* named entities and topic keywords extracted automatically from text analysis.

The manual and automatic annotations are analyzed and categorized by the system using DBPedia ontology structure, in order to propose to users topics and video of interest. Fig. 4 shows an example of semantic user profile automatically generated by the system. Annotations and categories are saved as RDF triples. Profile interests and other informations about network resources can be accessed using SPARQL queries. This allows users to identify unexpected or unknown associations between topics and videos, and gives the possibility to interact with other people who share the same interests. Furthermore users can choose to keep or remove from their semantic profile interests, topics and videos proposed automatically by the system.

Finally, the application provides notification systems typical of modern social networks: the user is notified when he is tagged or any of his friends is tagged in a video, or if anyone has tagged one of his videos or videos relating to

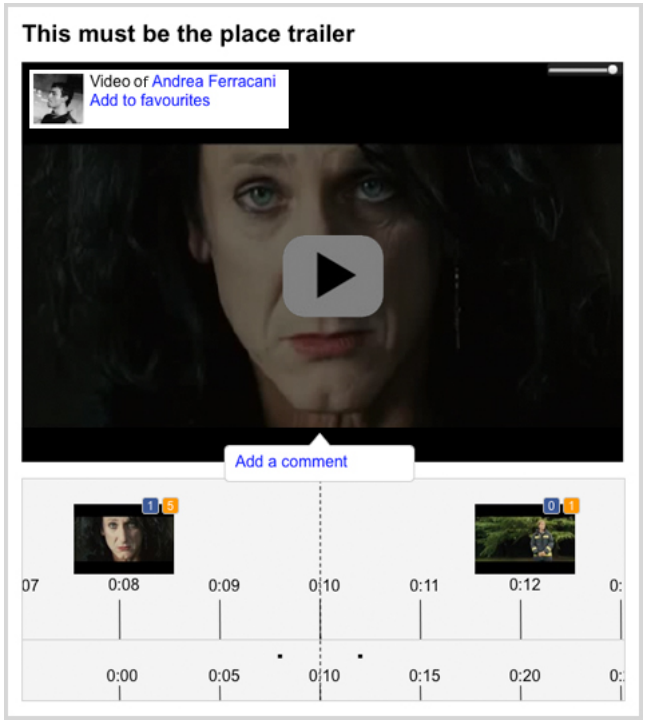

Figure 3: The video player with the timeline for intra video navigation.

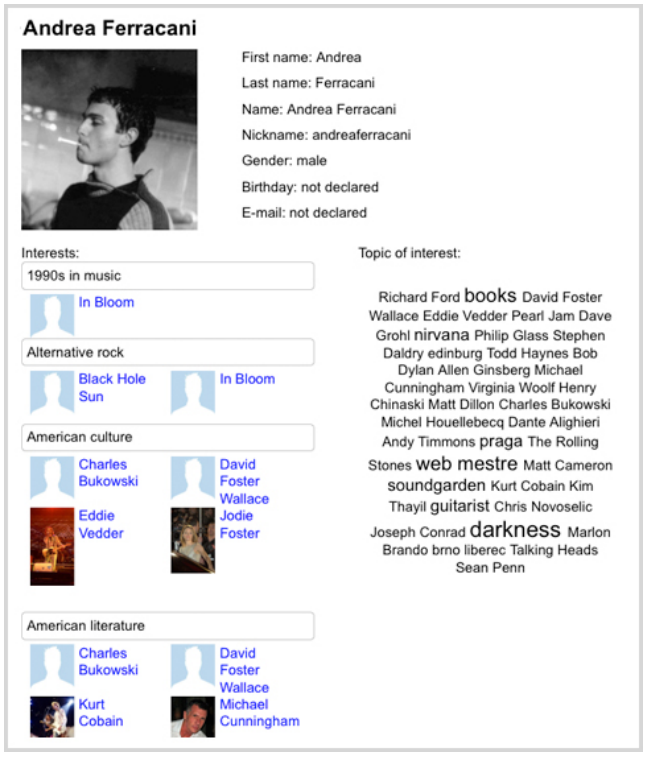

Figure 4: The automatically generated semantic profile of a user, showing his interests and the cloud of topics automatically extracted.

his interests. Each notification has a visual reference to the frame in the video where the annotation was added. Such actions are also automatically shared on the user's Facebook Wall.

3. THE DEMONSTRATION

We demonstrate the annotation modalities of the system (Fig. 2), along with its functions to browse and discover new content through the annotations and the implicit suggestions generated by users activity in the network. The main tool for manual annotations is the video annotation widget that provides a simple way to annotate videos (Fig. 3) by users that do not have any knowledge of ontologies, taxonomies or controlled vocabularies, but nevertheless exploits the structured knowledge available in DBPedia or in the Facebook graph through the real-time population of autosuggest input fields, obtained either asynchronously querying the DBPedia endpoint with SPARQL or making REST calls to the Facebook Open Graph API. We also show how entities and topics are extracted automatically when users enter a comment on a frame or respond to a comment in a thread by sending the text to a Java servlet that performs text analysis. This extraction allows the creation of several folksonomies and creates a personalized semantic profile for each user, based on his own interests: this semantic interests profile provides the possibility to watch new videos, access Wikipedia pages or contact Facebook users (Fig. 4). The social aspect of the system stemming from the network of Facebook friends of each user is also considered: it provides suggestions to check new videos either from the annotations of concepts that are part of the interests of each user or by explicit references in the annotations.

The demo will also point out how the automatically generated page of each resource in the network (people, videos, named entities, topics) is expressed in RDFa syntax using microformats to bind data to structured vocabularies recognized on the Web like DBpedia, the Open Graph Protocol, FOAF[4] and Dublin Core[5].

The focus of the demo of the system is to annotate videos related to entertainment, music and arts but in principle it can be used in other categories such as sports, cars and races. A screencast showing an example of these functionalities is publicly available at: http://vimeo.com/miccunifi/facetube

4. CONCLUSIONS AND FUTURE WORK

In this demo we have presented a system that allows social network users to discover new videos whose content matches their interest profile. These profiles are automatically created through the semantic analysis of the annotations created by users themselves. Our future work will deal with improved text analysis of user comments, the use of part of speech analysis in order to expand the annotation based on DBPedia, and with exploitation of unsupervised automatic video annotation techniques based on user generated annotations.

5. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union's Seventh Framework Programme managed by REA-Research Executive Agency ([FP7/2007-2013] | [FP7/2007-2011]) under grant agreement n. 262428 (www.eutvweb.eu). The authors thank Giannantonio D'Avico for his contribution to the development of the system.

[4]http://www.foaf-project.org/
[5]http://dublincore.org/

6. REFERENCES

[1] L. Ballan, M. Bertini, A. Del Bimbo, M. Meoni, and G. Serra. Tag suggestion and localization in user-generated videos based on social knowledge. In *Proc. of ACM Workshop on Social Media (WSM)*, 2010.

[2] L. Ballan, M. Bertini, A. Del Bimbo, and G. Serra. Enriching and localizing semantic tags in internet videos. In *Proc. of ACM Multimedia*, Nov 2011.

[3] M. Bertini, A. Del Bimbo, A. Ferracani, L. Landucci, and D. Pezzatini. Interactive multi-user video retrieval systems. *Multimedia Tools and Applications*, 2012.

[4] A. G. Hauptmann, M. G. Christel, and R. Yan. Video retrieval based on semantic concepts. *Proceedings of the IEEE*, 96(4):602–622, 2008.

[5] S. Kuznetsov. Motivations of contributors to wikipedia. *SIGCAS Comput. Soc.*, 36, June 2006.

[6] G. Li, M. Wang, Y.-T. Zheng, and T.-S. Chua. ShotTagger: Tag location for internet videos. In *Proc. of ACM ICMR*, 2011.

[7] X. Li, C. G. M. Snoek, and M. Worring. Unsupervised multi-feature tag relevance learning for social image retrieval. In *Proc. of ACM CIVR*, 2010.

[8] M. Naphade, J. Smith, J. Tesic, S.-F. Chang, L. Kennedy, A. Hauptmann, and J. Curtis. Large-scale concept ontology for multimedia. *IEEE Multimedia*, 13(3):86–91, July-Sept. 2006.

[9] H.-T. Pu. A comparative analysis of web image and textual queries. *Online Information Review*, 29(5):457 – 467, 2005.

[10] B. Sigurbjörnsson and R. van Zwol. Flickr tag recommendation based on collective knowledge. In *Proc. of WWW*, 2008.

[11] K. Siorpaes and M. Hepp. Games with a purpose for the semantic web. *IEEE Intelligent Systems*, 23:50–60, May 2008.

[12] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *Proc. of ICCV*, 2003.

[13] C. G. M. Snoek, M. Worring, J. C. van Gemert, J.-M. Geusebroek, and A. W. M. Smeulders. The challenge problem for automated detection of 101 semantic concepts in multimedia. In *Proc. of ACM MM*, 2006.

[14] T. Steiner. DC proposal: Enriching unstructured media content about events to enable semi-automated summaries, compilations, and improved search by leveraging social networks. In *Proc. of ISWC*. 2011.

[15] J. Zhang, M. Marszałek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. *International Journal of Computer Vision*, 73(2):213–238, 2007.