Social and Automatic Annotation of Videos for Semantic Profiling and Content Discovery

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

ABSTRACT
This demo presents a system based on social relationships, social knowledge and automatic video and textual content analysis for the discovery of videos in social networks. The system, developed as a web application, allows users to annotate, manually and automatically, and comment 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 profile of each user extracting and expanding his interests and folksonomy, as well as resources of interest in his social graph. The automatically generated profile page is used to suggest to users new resources, Facebook friends and videos whose content is related to their interests and allows profile curation.

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
In recent years tagging of multimedia content has become a common facility of many sites that offer functionalities for multimedia sharing. People sharing images on Flickr and Facebook, or sharing videos on YouTube and Vimeo, are now accustomed to tagging practices. These user-generated tags are used to retrieve multimedia content and to ease browsing and exploration of media collections; often tags are used to exploit social mechanisms, e.g. reminding to some user that he was tagged in a photo by a friend. However, not all media are equally tagged by users: currently the practice of tagging a single photo, or even just 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 is to differentiate tags that refer to some person, e.g. using the @ sign, while tags that refer to some topic or conversation are marked to highlight and group documents or conversations, e.g. using the the # sign ("hashtag"). Finally, 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 simply because the process is completely manual and cumbersome. 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 propose a method for manual and automatic 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, rich internet applications coupled with strategies like the use of games and competition [4] or systems based on reputation and community membership [1], stimulate participation to tasks that are human-centric. This may allow to enrich unstructured media content [5] 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. 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 enrich semantic annotations based on ontologies; on the other hand tools made to be deployed to general users have to deal with folksonomies created without need of explicitly specifying the relations between the concepts. Studies like those made by Yahoo! Search1 show that end users are less interested in annotations of scenes and objects that are commonly addressed by the current automatic annotation systems. As noted in [3] 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

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

in image searches, particularly in the entertainment domain, that is also the domain that contains the majority of the searches.

To deal with these issues and to ease the creation of semantic tags, we propose a system\(^2\) for social-based video annotation that allows manual annotation of resources and people from Wikipedia and Facebook, providing also automatic selection and suggestion of “interesting” keyframes based on their visual saliency \(^2\) and automatic extraction and semantic expansion of entities and topics from user comments, that are then linked to Wikipedia resources. This enables the discovery and browsing of videos according to the interests of each user, and it generates new forms of social interaction that increase tagging of multimedia content in order to improve and encourage user-generated semantic annotations. 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.

2. THE SYSTEM

The frontend 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 that are then transcoded to Flash Video Format and processed to select the most salient keyframes. Videos are served to users through the PHP-based XoomStream server, without requiring the deployment of a full video streaming server. Video thumbnails are created in correspondence with salient keyframes and user annotations using FFmpeg. Annotations are stored in two MySQL databases that map the RDF triples using the ARC2 RDF library for PHP.

Users can tag resources within their 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, through the DBpedia API, a list of Wikipedia pages whose name matches the typed characters.

The research leading to these results has received funding from the European Union’s 7th Framework Programme managed by REA-Research Executive Agency ([FP7/2007-2013][FP7/2007-2011]) under grant agreement n. 262428 (www.euvtweb.eu). The authors thank Giannantoni D’Avico for his contribution to the development of the system.

To reduce the user’s effort required for tagging and to enrich the semantics of annotations, the system performs text analysis of users’ comments to identify potentially interesting tags. Named entities detection is based on the GATE/Annie system 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 automatically 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.

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 useful 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.

Annotations and categories are saved as RDF triples; profile interests and other informations about network resources are accessed by the system using SPARQL queries. The manual and automatic annotations are analyzed and categorized by the system using DBPedia ontology structure, in order to present to users videos and topics of interest. 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 curate their personal page and 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 other videos that are related to 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 Facebook Wall of the users.

3. REFERENCES


