Collaborative Video Annotation for Multimedia Sharing between Experts and Amateurs

Dominik Renzel, Yiwei Cao, Michael Lottko, Ralf Klamma

Information Systems & Databases, RWTH Aachen University,
Ahornstr. 55, 52056 Aachen, Germany
{renzel,cao,lottko,klamma}@dbis.rwth-aachen.de

Abstract. Practices in communities can be supported through a wide use of standard multimedia and web technologies. Lately, several research domains, especially cultural heritage management, have discovered the power of collaboration with amateurs in multimedia documentation work. Amateurs often carry knowledge they would be willing to contribute with small effort. Thus, there is a need for intuitive multimedia annotation tools supporting efficient collaboration among different user communities. The MPEG-7 metadata standard has been well applied to describe rich multimedia semantics. However, the complexity of MPEG-7 should rather be hidden from the user. In this paper we present the prototype of the community-aware semantic video annotation service SeViAnno based on a combination of metadata standards and Web 2.0 technologies in the cultural heritage management domain. An evaluation was carried out with amateurs and experts to explore the influence between both communities of different expertise levels.

Keywords: MPEG-7, RIA, mashup, multimedia, metadata, Web 2.0

1 Introduction

Nowadays, it is possible for amateurs to help researchers in various fields. Cultural heritage management is one of those fields, where people from certain cultural areas can help researchers in reconstructing history. The city of Düren in Germany is currently pursuing such an approach in the context of establishing a city museum. Historic images were published on the web, and with the help of citizens being contemporary witnesses they are now able to access and reconstruct historical details related to those images[1].

The following questions are raised in the domain expert communities such as cultural heritage management: How can we design complex metadata annotation workflows for technology-inexperienced domain experts with Web 2.0 technologies? And how can we make good use of the knowledge and support from a large range of amateur communities.

In our research work, we deal with these previously addressed problems with three main aspects. First, the annotation activities of domain experts are paid more

[1] http://www.stadtmuseumdueren.de/bildersuche.html

attention to. This special support of domain experts is not well covered in common Web 2.0 multimedia platforms. On the other hand, collective intelligence of the wide "amateur" communities has been drawing attention, as how successfully those Web 2.0 sites like Twitter and Facebook nowadays work. Thus, the awareness of specialized user communities, e.g. in cultural heritage management has become one of our main research focuses. Second, multimedia metadata standards such as MPEG-7 enhance data, service and platform interoperability. Multimedia metadata standards in combination with user generated information on the Web 2.0 are used to enhance quality and quantity of multimedia annotations. For example, expert knowledge can be collected from existing systems using domain-specific metadata standards. It is not yet possible in YouTube to handle existing metadata in customized cultural heritage management platforms. Third, simplicity and intuitiveness are achievable with Web 2.0 and Rich Internet Applications (RIA) [1]. The Internet has been the pioneer platform and makes it possible now to create sophisticated and desktop-like user experiences in the web browser with high interactivity. As a proof of concept, community-aware video annotation activities among domain experts and amateur communities are traced in the Web platform prototype *SeViAnno*[2], a Flex- based approach to collaborative video annotation using the MPEG-7 standard .

The rest of this paper is structured as follows. Section 2 includes a brief state-of-the-art analysis of existing metadata standards and Web 2.0 technologies. Section 3 describes the development process of the SeViAnno prototype as a proof-of-concept. In Section 4 we describe the evaluation process and results of our prototype. Finally we conclude with an outlook to further work.

2 Related Work

Multimedia sharing between expert and amateur communities is interesting to observe and explore. Our prior research results show that tags used by experts are more concise than "amateurs" as the expert level increases [2]. Among a large number of Web 2.0 platforms, Flickr, YouTube, and Last.fm all provide the possibility to add and edit tags to multimedia content. When the tagging processes are observed, there are three different tagging concepts behind them. YouTube only allows media owners to tag their own videos. No other users are able to edit or add tags afterwards. Last.fm differentiates personal tags stored to each user individually and professional tags labeled to each piece of music. Flickr enables everybody to add tags to photos. Flickr even has a sub space called *Flickr Common* to involve users to annotate public image archives. This enables the knowledge transfer from amateurs to experts. A channel for both directions is still missing. There have been many communities for cultural heritage management, e.g. the Bamiyan Development Community[3], which support a lot of activities instead of some special and professional task support such as video tagging.

Metadata is used to describe content of multimedia files and can be classified as descriptive metadata, technical metadata or user dependent metadata. MPEG-7 is one of the most comprehensive multimedia metadata standards. It can be easily integrated

[2] http://tosini.informatik.rwth-aachen.de/media/SeViAnno.html
[3] http://www.bamiyan-development.org/

into existing systems, but has a very inclusive description scheme, and is thus very complex [3].

Rich Internet Applications aim to provide usable, complex and platform independent applications, which can be accessed from anywhere. Macromedia introduced the term in 2001 and described it as an appealing, interactive, slim and flexible web application. In 2004 Macromedia introduced Flex, which was not very successful due to its high price and the lack of an IDE. With the introduction of Flex 2 in 2006, RIA became much more popular. Recently, Adobe released Flex 4. Other technologies to create RIAs are Ajax or Microsoft Silverlight. Moreover, a mashup is a website or an application which includes data and functionality of several services to create a new service. For example, Google Maps has been widely used in travel, logistics, and customer relationship management platforms to present location-related information on a map.

There exist many *video annotation tools* for domain experts. *VideoAnt* is a web application developed at the University of Minnesota [4]. Users can annotate videos uploaded from their file system or provide a YouTube URL. *M-OntoMat Annotizer* is a desktop application developed at different universities (including University Koblenz and University Karlsruhe) [5]. Due to its complex user interface it is not suitable for non-computer experts and amateurs who only want to add small pieces of information. Nevertheless, it offers a high precision of semantization and is based on the MPEG-7 standard. Other projects dealing with multimedia content description are Boemie [6] and K-Space [7]. Boemie includes the video and image annotation tool VIA. In our previous research, the *Virtual Entrepreneurship Lab (VEL)* was developed as an interactive learning environment for entrepreneurial education [8]. An early version of MPEG-7 was used to manage metadata for different multimedia types in a consistent way. Its successor *MECCA* was designed as a multimedia screening environment to foster collaboration for movie scientists in a distributed setting [9].

3 The SeViAnno Prototype Development Process

In this section, we briefly describe the SeViAnno prototype development process. Requirement analysis has been conducted based on feedback of amateur and expert communities in the domain of cultural heritage management. The design and implementation of our community-aware semantic video annotation tool meets the three main aspects of (a) knowledge transfer between domain experts and amateurs, (b) utilization of metadata standards for interoperability, and (c) user interface simplicity and intuitiveness.

3.1 Requirements Analysis with Paper Prototyping

Based on the combination of requirements analysis and the analysis of existing Web 2.0 applications we created a paper prototype and evaluated it with different subject groups of cultural heritage management amateurs and professionals. Paper prototyping was thereby found to be a suitable method to improve user interfaces and to identify missing functionality [10]. Several interesting observations were made, e.g. that inexperienced users prefer natural data visualization such as places shown in a map or events on a timeline. All observations were then included in the

implementation of the first software prototype. It was developed using RIA technology with the dedicated goal to design an intuitive user interface hiding the complexity of MPEG-7.

3.2 Community-aware Semantic Video Annotation

A video annotation process between domain experts and amateurs comprises complex workflows. How can knowledge be well shared between amateur and experts on the level of information workflow? And how can activities on media undertaken in both communities be well influenced or interacted on the level of process workflow?

On the information workflow, valuable metadata representing professional knowledge of domain experts needs to be transferred to the amateur. Amateur users can select professional annotations from existing tag clouds to annotate a certain time point or an interval of video clips. Video tags from the amateur communities can be refereed by domain experts. At the same time, experts are able to get pre-processed video clips by large user communities and further elaborate on annotations. Several video annotation information classes are involved: annotations from complicated expert video annotation systems; selected annotations by amateurs and applied on video clips, and further added annotations by professionals on video clips. Moreover, the elaborated annotations by both communities can be used to generate more concise annotations for video clips.

On the annotation activity workflow, each video can be annotated with an unlimited number of tags. Each time point of a video can be annotated with rich semantic information including agents, objects, events, times, places and concepts. Community awareness is realized through a collaborative tagging process between experts and amateurs. Professionals and amateurs collaborate on video annotations to manage and share domain specific knowledge. The information about who has annotated which video segment is visible to the community. Whenever a tag is clicked, the video segment is played directly.

3.3 MPEG-7 Multimedia Annotation

Semantic annotations are realized as MPEG-7 Semantic Basetypes including *Agent*, *Concept*, *Event*, *Object*, *Place* and *Time*. Concepts, event and objects can be added by simply specifying a name. Time is specified by additionally adding a date. Places include longitude and latitude values. Existing documented or annotated videos within cultural heritage management communities might be represented in certain cultural heritage metadata standards such as CIDOC CRM. These metadata can be easily mapped to MPEG-7 and be used as initial expert knowledge for the whole video.

The management of all MPEG-7 data is implemented as LAS Web services [11]. The LAS multimedia and user management services enable Web clients to create, search, and retrieve MPEG-7 metadata. An MPEG-7 Semantic Basetype service is used for semantic annotations. Access to video segments is realized through an MPEG-7 multimedia content service. Every video can be separated into several audio visual segments, where each one has its one time point and duration. Semantic annotations are assigned to multimedia content descriptions as semantic references.

3.4 Simplicity, Intuitiveness and Interactivity

Users can annotate video segments easily by clicking the relevant semantic annotation tabs. Especially for the place annotation, a Google maps frame is applied. All place annotations are shown as markers on the map. If a user wants to get to the corresponding position in the video, he just needs to click on the marker and the video is automatically started at this position. Video segment annotations for places are possible through clicking on the map. While watching the video, the map zooms automatically into a place occurring in the respective segment.

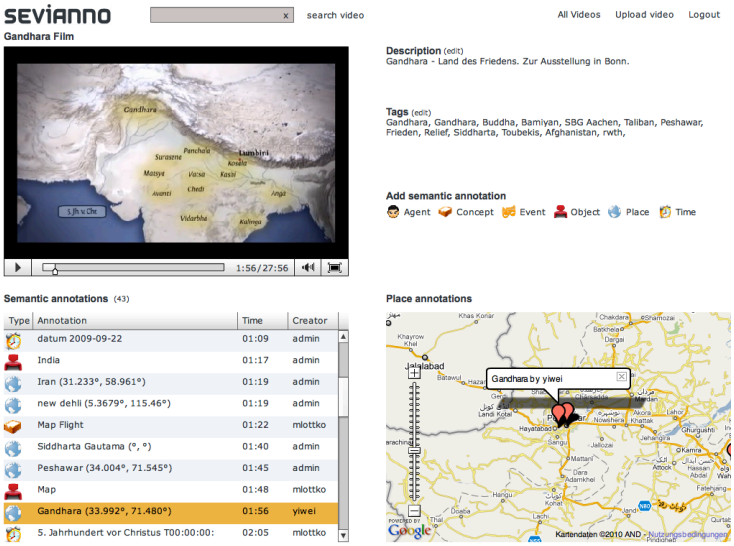

Fig. 1. The SeViAnno user interface with a video player, video information and video list, user created annotations, and Google map mashup for place annotations.

All semantic annotations are listed below the video player with an intuitive icon standing for one of the six supported types. By clicking each individual tag, users can access the related video segment. One of the main user interface improvements is automatic annotation highlighting while a video is played. In order to realize awareness for community annotation activities on a particular video, the list shows nicknames of the respective annotators. Additionally, SeViAnno supports usual plain keyword tags and text annotations stored in the MPEG-7 metadata. A screenshot of the SeViAnno user interface is depicted in Fig. 1. All user experiences including video upload, video browsing, video tagging, and video segment annotation are realized within one Web page.

4 Evaluation

In order to compare collaborative multimedia annotations of experts and amateurs, we conducted a small-scale experiment in the domain of Afghan cultural heritage with six subjects, one of them a cultural heritage expert with profound domain knowledge, the others amateurs. All subjects were asked to complete the same task of creating and/or assigning semantic tags to two videos in SeViAnno. One video showed a documentation of a 3D laser scan of a small Buddha niche in the Bamiyan valley, the other a 3D reconstruction of cities and monasteries in Gandhara. Both videos did not include any audio track or subtitle. Subjects were asked to watch each video exactly once and to add their annotations. The time for completing this task was not limited. All subject activities in SeViAnno were automatically monitored and recorded using MobSOS [12]. Monitoring log data together with generated MPEG-7 multimedia metadata were later on used for analysis.

For the monitoring log data analysis, we considered descriptive statistics on measures such as total session duration, method invocation frequency, number of bytes sent, etc. as proxies for the comparison of annotation activity. We first analyzed the total duration of each subject's SeViAnno evaluation session. While the average session duration for amateurs was at 28.91 min (min: 12.8 min; max: 51.5 min), the professional spent 136.7 min, i.e. more than four times longer than the amateur average. Compared to the total video duration of 6.25 min, subjects spent the major part of their session time on browsing, revisiting and annotating the multimedia material. However, session time alone was not sufficiently expressive to allow statements about annotation activity. Therefore, we analyzed LAS method invocation frequency as the next proxy measure for activity. 44.79% of all LAS method invocations were made by the professional, resulting in an average of 11.04% for each of the amateurs.

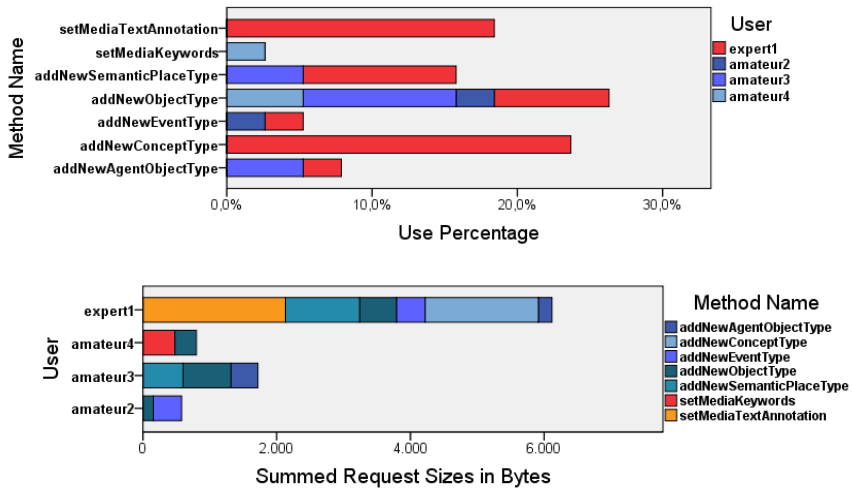

Fig. 2. MPEG-7 Service Method Invocation Statistics for Experts and Amateurs

More detailed statistics on MPEG-7 service method invocation are shown in Figure 2. Since our analysis concentrated on annotation activity, we only included those methods in our analysis which actively contributed to multimedia annotations. Two of the amateurs did not contribute any annotations, probably because they considered previously made annotations as sufficient. The upper diagram clearly shows that the majority of annotation method invocations was executed by the domain expert. While amateurs annotated using plain keyword tagging and at most two distinct semantic base types per subject, the expert used the full range of all types supported in SeViAnno and additionally specified full text descriptions. Furthermore, we analyzed the number of bytes sent as parameters with all MPEG-7 annotation method invocations as proxy for annotation length. The lower diagram in Figure 2 paints a clear picture that again the expert contributed more than all amateurs together. However, the above quantitative statistics do not provide any information on the actual quality of annotations. Therefore, further analysis of the actual annotation quality was conducted based on the generated MPEG-7 multimedia metadata descriptions and observations during the evaluation session. One interesting observation was made regarding the annotation of places with Google Maps. Amateurs just typed in the name of a location and blindly trusted the partially wrong coordinates returned, introducing annotation imprecision. The professional always checked the proposed location for correctness resulting in higher precision and annotation quality. Amateurs often reused already existing semantic base types for tagging instead of creating new elaborate ones. Some amateurs abused annotation functionality for asking questions, e.g. by creating an object with title "What are these green areas?" which can serve as request for more precise annotation by a professional, but again introduce a certain decrease of annotation quality. However, altogether, we found that annotations of amateurs and professional complemented each other. While amateurs provided their annotations quickly, but not too profound, professionals spent a lot of time to provide annotations as detailed as possible.

5 Conclusions & Outlook

We have realized the concept to support the knowledge sharing and community awareness across expert communities and amateur communities on deploying the prototype SeViAnno. The MPEG-7 metadata standard is employed to enhance interoperability and rich semantic annotation. The interface based on Web 2.0 and RIA technologies was designed to hide the complexity of MPEG-7. We furthermore identified that information and activity workflows between professional domain experts and amateur communities are complex, however complementary processes. There are still a list of open questions and tasks for further research. Browsing and annotation processes can be further explored and compared between different communities. How can comment, ranking, question-answer sessions be employed to improve the communication between experts and amateurs? Other future scenarios can be illustrated. SeViAnno can be applied for other communities. Such an "amateur-amateur" scenario for fun can be imagined. Hollywood fan communities can annotate film videos with clips related to some popular travel destinations, which could be well shared by travelers. In addition, it could be interesting to explore the

influence of annotation activities between experts and amateur communities. Questions could be addressed whether amateurs feel self-confident to annotation videos in such platforms, in contrast to a Web 2.0 video sharing site. The level of community awareness can be further extended by introducing more direct interaction between community members, e.g. with an included chat functionality, member presence information, etc. The effects of such extensions on annotation quality can then be further explored and exploited.

References

1. Allaire J.: Macromedia Flash MX – A next-generation rich internet client. (2002)
2. Klamma, R., Cao, Y., and Jarke, M.: Storytelling on the Web 2.0 as a New Means of Creating Arts, Borko Furht (Eds.): Handbook of Multimedia for Digital Entertainment and Arts, pp. 623-650. Springer (2009)
3. Manjunath, B.S., Salembier, P., and Sikora, T.: Introduction to MPEG-7: Multimedia Content Description Interface. Wiley & Sons (2002)
4. Hosack, B., Miller, C., and Ernst, D.: VideoANT: Extending Video beyond Content Delivery through Annotation. T. Bastiaens et al. (Eds.), In: Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2009, pp. 1654-1658, Chesapeake, VA: AACE (2009)
5. Bloehdorn, S., Petridis, K., Saathoff, C., Simou, N., Tzouvaras, V., et al.: Semantic Annotation of Images and Videos for Multimedia Analysis, ESWC 2005, 2nd European Semantic Web Conference, Heraklion, Greece (2005)
6. Paliouras G: Final Activity Report of the EU FP6 Project BOEMIE. (2009)
7. Izquierdo, E., Chandramouli, K., Grzegorzek, M., and Piatrik, T.: K-Space Content Management and Retrieval System. In: Proceedings of the 14th international Conference of Image Analysis and Processing - Workshops (September 10 - 13, 2007), pp. 131-136. IEEE Computer Society, Washington DC (2007)
8. Spaniol, M., Klamma, R., and Jarke, M.: Semantic processing of multimedia data with MPEG-7 for comprehensive knowledge management. B. Grosky (Ed.): Proceedings of SOFSEM 2002 Workshop on Multimedia Semantics, Milovy, Czech Republic, November 27-28, pp. 56-65. (2002)
9. Klamma, R., Spaniol, M., and Jarke, M.: MECCA: Hypermedia Capturing of Collaborative Scientific Discourses about Movies. Informing Science. The International Journal of an Emerging Discipline, Volume 8, N. Sharda (Ed.): Special Series on Issues in Informing Clients using Multimedia Communications, 3-38 (2005)
10. Snyder, C.: Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces (Interactive Technologies). Morgan Kaufmann, San Francisco (2003)
11. Spaniol, M., Klamma, R., Janßen, H., and Renzel, D.: LAS: A Lightweight Application Server for MPEG-7. In K. Tochtermann, H. Maurer (Eds.): Proceedings of I-KNOW'06, 6th International Conference on Knowledge Management, Graz, Austria, September 2006, J.UCS (Journal of Universal Computer Science) Proceedings, pp. 592-599. Springer (2006)
12. Renzel, D., Klamma, R., and Spaniol, M.: MobSOS - A Testbed for Mobile Multimedia Community Services. 9th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '08). Klagenfurt, Austria (2009).