WISMA 2010

Workshop on Interoperable Social Multimedia Applications

Proceedings of the 11th International Workshop of the Multimedia Metadata Community

Barcelona, Spain, May 19-20, 2010

General Chair:
Jaime Delgado (Universitat Politècnica de Catalunya, Spain)
Industrial Co-Chair:
Eva Rodríguez (Universitat Politècnica de Catalunya, Spain)

International Programme Committee

Anna Carreras (Universitat Politècnica de Catalunya, Spain)
Ansgar Scherp (University of Koblenz-Landau, Germany)
Bill Grosky (University of Michigan, USA)
Chris Poppe (Ghent University - IBBT, Belgium)
Christian Timmerer (Alpen-Adria-University Klagenfurt, Austria)
Dominik Renzel (RWTH Aachen University, Germany)
Frédéric Dufaux (EPFL, Switzerland)
Harald Kosch (University of Passau, Germany)
Herve Bourlard (Idiap)
Jaime Delgado (Universitat Politècnica de Catalunya, Spain)
Laszlo Böszörmenyi (Klagenfurt University, Austria)
Marc Spaniol (MPI, Saarbrücken, Germany)
Markus Strohmaier (Know Center Graz, Austria)
Mathias Lux (Klagenfurt University, Austria)
Michael Granitzer (Know Center Graz, Austria)
Oge Marques (Florida Atlantic University, USA)
Ralf Klamma (RWTH Aachen University, Germany)
Richard Chbeir (Bourgogne University, France)
Romulus Grigoras (ENSEEIHT, France)
Ruben Tous (Universitat Politècnica de Catalunya, Spain)
Savvas Chatzichristofis (Democritus University of Thrace, Greece)
Vincent Charvillat (ENSEEIHT, France)
Vincent Oria (NJIT, USA)
Werner Bailer (Joanneum Research Graz, Austria)
Yu Cao (California State University at Fresno, USA)

Local Committee at the Universitat Politècnica de Catalunya

Anna Carreras
Jaime Delgado (Chair)
Xavier Maroñas
Víctor Rodríguez

Foreword

In the Web 2.0, a growing amount of multimedia content is being shared on Social Networks. Due to the dynamic and ubiquitous nature of this content (and associated descriptors), new interesting challenges for indexing, access, and search and retrieval have arisen.

In addition, there is a growing concern on privacy protection, as a lot of personal data is being exchanged. Teenagers (and even younger kids), for example, require special protection applications; while adults are willing to have a higher control over the access to content.

Furthermore, the integration of mobile technologies with the Web 2.0 applications is also an interesting area of research that needs to be addressed; not only in terms of content protection, but also considering the implementation of new and enriched context-aware applications.

Finally, social multimedia is also expected to improve the performance of traditional multimedia information search and retrieval approaches by contributing to bridge the semantic gap. The integration of these aspects, however, is not trivial and has created a new interdisciplinary area of research.

In any case, there is a common issue that needs to be addressed in all the previously identified social multimedia applications: the interoperability and extensibility of their applications.

The meeting held in Barcelona in May 2010 was the 11th in a series of Multimedia Metadata Community workshops, this time paying special attention to annotations on content in Social Networks. The presented papers have been collected in this book.

The editors

Contents

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.

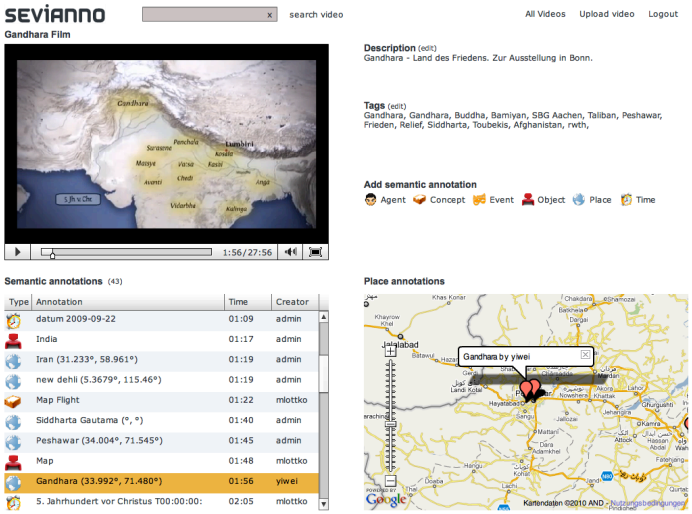


Fig. 1. The SeViAnno user interface with a video player, video information and video list, user created annotations, and Google map mashup for place annotations.

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.

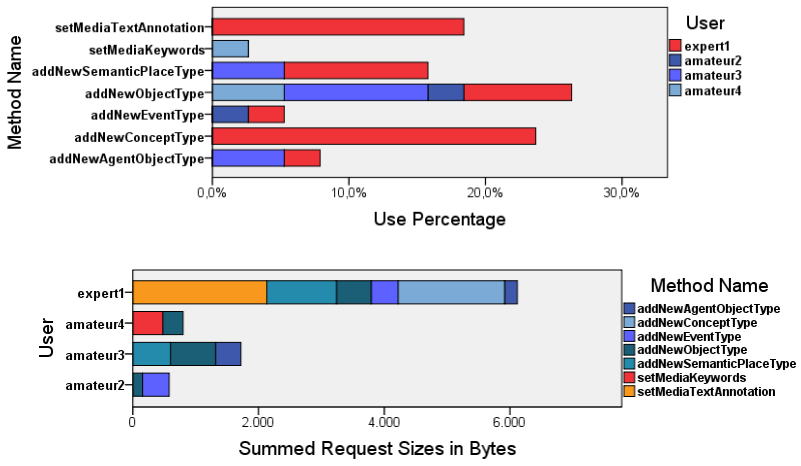


Fig. 2. MPEG-7 Service Method Invocation Statistics for Experts and Amateurs

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).

User-aware Reactive Components for Social Networks

Benoit Baccot[1,2], Romulus Grigoras[1], and Vincent Charvillat[1]

[1] University of Toulouse, IRIT-ENSEEIHT, 2 rue Camichel, 31071 Toulouse, France
[2] Sopra Group, 1 Avenue André Marie Ampère, 31772 Colomiers, France

Abstract. Social networking Web sites are extremely popular. Users spend an increasing amount of time and expect them to be a very efficient communication medium between friends. These Web sites should be as friendly as possible, and provide users with the sense of being together. Therefore Web sites should adapt to both the way users interact with their own space or socially, with the other users. In this paper we propose to extend an existing Web adaptation platform, that allows to dynamically impact on user browsing sessions on a Web site, in two ways. First, since we apply it to social networking Web sites, we propose to extend our event model and take into account social events. Second, we propose to enrich the social networking Web sites with components that react to both the *user activity level* and the *social activity level*. The user activity level describes the low-level, basic interactions, while the social activity level characterizes the social interactions between users. Experimenting with components deployed on social networks can provide researchers with a large-scale experimental testbed.

1 Introduction

Social network sites are commonly defined (e.g. in [1]) as Web sites that have three main characteristics. First, they allow individuals to construct a public or semipublic profile, in which they can put various types of information about their personnalities, their hobbies or their status. Secondly, a list of other users with whom a given user shares a connection is also available. Finaly, the main point is that a user can use these connections to view other profile pages and go through their connection lists, to make new ones. Nowadays, social network sites, such as Facebook, Myspace or Twitter, gets bigger and bigger, and are used by millions of users a day.

The particularity of social networks is that they can provide a rich context about the users, since beyond the classic information (terminal type, available bandwidth, etc.) information related to interactions between users and content from the profile pages are available.

In this paper, we deal with this rich context and try to use an adaptation platform on social networks. This adaptation platform is able to provide Web site modifications based on the context for each users. In this paper, we want to

modify a social network Web page in order to take into account the rich context, and make the site more reactive to the user.

The rest of this paper is organized in four main sections. First we present the adaptation platform, and show how it can be deployed on a social network Web site. We then define more precisely social events, that will be recorded by the platform. Two use-cases that uses the platform and the rich context available on social network Web sites are described. Some preliminary results about the implementation on Facebook and perspectives conclude this article.

2 A Dynamic Adaptation Platform for Social Networks

Generally, adaptive, dynamic multimedia applications need to take into account the context of usage in order to provide the user with a service that helps him/her accomplish a task better: find an information or a friend, communicate, play etc.

Understanding how users behave when they connect to social networking sites [2] creates opportunities for better interface design, richer studies of social interactions, improved design of content distribution systems etc. In social networks, interactions between users is paramount. These interactions raise so-called "social networking events" (here called "social events") that propagate through the network. The friends are kept posted whenever a user changes his profile, comments on another user's photo, etc. ([1]).

Interactions are mainly asynchronous: the propagation of events takes some time, and most events are generated by friends while a user is offline. Commonly the user "consumes" the events during the next session. As users are spending a large (and increasing) amount of time online, there is a need for more synchronous communication between them. Communication can be explicit (via posts, private messages etc) but also implicit. Like in real life, we may be worried if we see a friend in a sad mood. Online users may therefore be interested in knowing the state of mind of their friends at any moment. This may give them a better sense of being together ([3]).

In our previous work, we proposed a three-layer dynamic adaptation platform (figure 1) that is able to (1)observe the usage (capture various context elements), (2) analyze the usage and decide of a change to be made, and (3) apply the change. The platform naturally works in a closed-loop following the general framework of reinforcement learning: changes are evaluated with respect to some performance criteria, therefore their effectiveness can be monitored permanently. The best changes are kept, while the ones with little impact are avoided.

Also in our previous work [4] we considered the user activity level while interacting with a rich-media content. We defined the activity as the number of basic events produced by a user over a given time window. These events correspond to the basic interactions a user performs in an internet browser (e.g. mouse clicks, mouse movements, scrolls or key presses) or interactions with a multimedia component (e.g. VCR events). Considering the activity level can be useful for many applications, since adaptation actions on the Web site can be

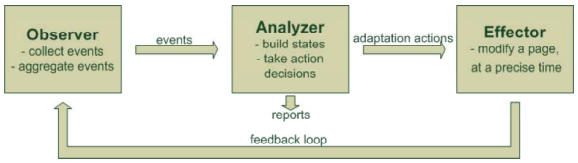


Fig. 1. The three layers of our adaptation platform.

performed upon detecting an important variation of the activity level (e.g. steep decreases or increases).

In this paper we propose to extend our dynamic platform in two ways. Firstly, since we apply it to social networking Web sites, we propose to extend our event model and take into account events at a higher semantic level: social events. Secondly, we propose to enrich the social networking Web sites with components that react to both the *user activity level* and the *social activity level*. The user activity level describes the low-level, basic interactions, while the social activity level characterizes the social interactions between users. Experimenting with our platform on a social Web site provides us with an interesting playground: once the components are deployed, and if the components enjoy a little popularity at first, the user base can increase tremendously, due to the viral nature of social networks. Indeed, this can be a very fruitful testbed on real users.

3 Social Events

Extending the event model in order to integrate social interactions is necessary for capturing user actions specific to social networks. This can be useful for both understanding (analyzing) these interaction, but also as triggers for various changes that the platform can dynamically apply to the Web site in order to adapt it to user's behaviour.

Figure 2 gives an overview of the events captured by the platform.

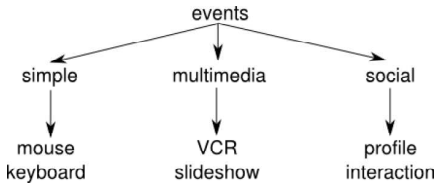


Fig. 2. The different types of events captured by our platform.

The platform is able to record simple events (events produced by mouse clicks, movements, etc.) and multimedia event (e.g. play, pause or stop on a multimedia

mponent). The specifities of social networks enable us to get new events, by using the interaction logic of such Web sites. As a result, social events can be categorized in two main groups: the ones related to publishing content on his content page (e.g. publishing a picture or writing a status) and the ones concerning interactions with other users (e.g. visiting a friend's profile or commenting a friend's publication).

Social events are more complex than simple or multimedia events, since they can have multiple targets, e.g. when a new content is published on a user profile page. The cardinality of the source or destination of these events may vary.

4 Components for social Web sites

In this section, we describe two use-cases for our platform on a social network. The first one dynamically adds a component on a social network Web page, that uses the behaviour of the current user to determine its content. The second component is much more entertaining and deals with overactive users, who are proposed a slap game in order to calm down.

4.1 User-aware Contact Bar

Using the available rich context, a dynamic contact bar can be added to a social network page. This bar recommends people that have the highest interactivity level (the most numerous interactions) with the current user. By clicking on a contact name, the user can display additional information. Interactivity level between users is defined using the social activity level: in fact, when a user interacts with a contact on a social network, it generates social events that can be recorded by our platform. The user activity level is also used since this bar is more visible when the activity of the user decreases, in order to stimulate him, at a time when, for example, he could leave the Web site.

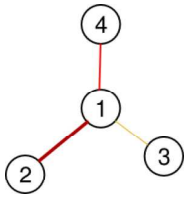


Fig. 3. The social activity of user 1. The size and the color of the branches determine the level of activity.

Social events are used to create a graph that represents the interactions between a given user and his contacts. Such a graph is presented in figure 3.

Users are the node of this graph. A link between two users means that social interactions between the users have been recorded. Links are weighted by the type and the number of times events have been captured. As a result, the graph is updated as often as social events are detected. Moreover, in order to take into account the temporal dimension, weights on link are frequently decreased. This enables to give a higher priority to recent social events, and allows the component to be more reactive to the user behaviour. Contacts to appear in the bar are then the ones with the highest interactivity with respect to the current user.

4.2 Slap Game

This use case presents a funny way to calm down overactive users, by using both the user and social activity of the user. It is really interesting to have such a funny component since it will help to disseminate the applications through through the social networks users. It also illustrates the platform ability to perform in real-time.

Generally, on a social network Web site, one can share information with other users on a profile page. Most of the time, a photo of the user is present in this page. When overactivity (in the sense of simple or multimedia events) is detected, a user is encouraged to use a slap game on a contact profile, that consists in giving slap to a contact (determined by the social activity), by moving the mouse inside and outside his contact profile picture. A score is computed, based on the velocity and the number of mouse movements. An animation is then played on the picture in order to see the effects of the slap. High scores are stored and displayed on corresponding profile pages.

5 Preliminary Results

In this section, we present the implementation of the use-cases on Facebook, using our adaptation platform. We also give some elements about the potential number of users of these components.

5.1 Deployment on Facebook

We choose to use Facebook [5] in order to implement the use cases. In fact, Facebook is unquestionably the biggest social network, with 400 millions active users who get connected in average 55 minutes a day [6]. On Facebook, the contacts of a user are called friends.

We adapt our platform (figure 1) in order to take into account social network particularities and to implement the detection of the previously called social events. We simply find rules that allow us to infer social events from simple events. Social events are of different types: chatting with a contact, adding a contact, posting a status etc.

In order to integrate it to Facebook, we had two choices:

- directly deploy the platform in the Facebook infrastructure. Obviously, as Mark Zukerberg, the Facebook creator, is not one of our acquintance, this was not a realistic solution ;).
- use a plugin for a Web browser. The plugin will be charged to dynamically add our events listeners in the Facebook Web pages, and also to add the contact bar and the slap game in the DOM tree of the Facebook pages. Then, the two components directly communicate with the platform, installed on one of our servers.

We developed a plugin for the Firefox Web browser, since installing a new plugin is quite easy (only 3 mouse clicks are needed), and the browser is used by many people. As a result, the plugin could be used by a significative number of people (see next subsection to get more details about that).

The contact bar was written in Javascript and uses the Facebook API in order to get information on the presented friends. Figure 4 presents the contact bar integrated to Facebook. The friends at the top of the bar are the one with whom the current user has the more interacted in past few minutes. It appears when a huge user activity is detected.

Fig. 4. The contact bar with two contacts.

The slap game was also written in Javascript. Figure 5 presents the slap effects on a "friend" picture profile. The slap game was proposed due to the user overactivity. As for the "friend", he was chosen using the social activity.

5.2 Dissemination of such components

Experimenting with our platform on a social Web site can provide an interesting playground. Once the components are deployed on a social network, and if the components enjoy a little popularity at first, the user base can increase tremendously, due to the viral nature of social networks. Obviously, "normal"

Fig. 5. A slapped friend and the high scores.

users needs to be interested by the components, but this is probably what may happen with the slap game. Indeed, this can be a very fruitful testbed on real users.

We present here (figure 6) the evolution of the number of users of FBOSF [7], an other Firefox plugin, that adds an "I don't care" button next to the "I like" one on Facebook friends messages, displayed on a user profile page. We have already worked with the creator of this plugin in order to deploy large scale experiments. It can be seen as a simpler component but all the same similar to the slap game, and it can enable us to think that we could have a similar dissemination for our plugin.

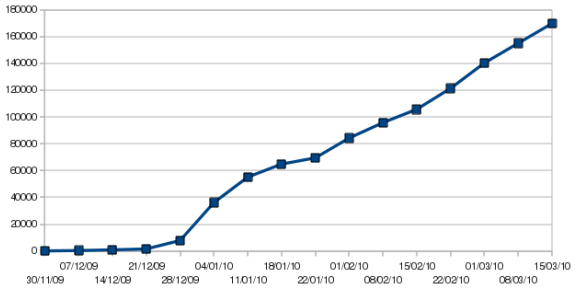


Fig. 6. Evolution of the number of users for FBOSF, 11/2009 - 03/2010.

What is really interesting is the increasing number of user (8k users after a month, 200k two months later). This encourages us to make our plugin available, in order to get as many users as possible to collect real user data.

6 Conclusion and Future Work

In this paper we presented some extensions to our original adaptation platform applied to social Web sites. In particular, we proposed novel components that can take into account both user's interactions with his own space, and also social interactions with other users.

As for the interaction graph used by our contact bar, its evolution in time (and the relation with the social graph) needs to be studied more carefully [8].

New use cases can be imagined for our platform. Examples include recommender components that suggests links according to user's behaviour, or a system that displays personalized advertisement. Statistical models such as bandit models, explored in our previous work [9], can help us optimize the performance of such components using reinforcement learning techniques that very often require a large user base. Thus, making use of the viral property of social network can enable us to test our statistical models.

7 Acknowledgments

The authors would like to thank ENSEEIHT students Anthony Luce, Guillaume Rondan, Imad El Ghalbzouri, Naoufal El Jaouhari and Simon Gay, for their participation in the development of the experimental Facebook components presented in this paper. Special thanks go to Julien Noleau, the creator of FBOSF, for his precious help.

References

1. Boyd, D., Ellison, N.B.: Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication **13**(1) (2007)
2. Benevenuto, F., Rodrigues, T., Cha, M., Almeida, V.: Characterizing user behavior in online social networks. In: IMC '09: Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference, ACM (2009) 49–62
3. Ellison, N.B., Steinfield, C., Lampe, C.: The benefits of facebook friends: social capital and college students' use of online social network sites. Journal of Computer-Mediated Communication **12**(4) (July 2007) 1143–1168
4. Baccot, B., Choudary, O., Grigoras, R., Charvillat, V.: On the impact of sequence and time in rich media advertising. In: Proceedings of the 17th ACM Conference on Multimedia. (2009)
5. Facebook: Social network. http://www.facebook.com
6. Facebook: Statistics page. http://www.facebook.com/press/info.php?statistics
7. Noleau, J.: Fbosf. http://fbosf.nolofone.com
8. Viswanath, B., Mislove, A., Cha, M., Gummadi, K.P.: On the evolution of user interaction in facebook. In: WOSN '09: Proceedings of the 2nd ACM workshop on Online social networks, New York, NY, USA, ACM (2009) 37–42
9. Baccot, B., Charvillat, V., Grigoras, R.: A bandit's perpective on website adaptation. In: Proceedings of the 10th Workshop on Multimedia Metadata (SeMuDaTe'09). (2009)

Image search based on a broker approach

Pere Torán, Jaime Delgado

Universitat Politècnica de Catalunya (UPC)
{ptoran, jaime.delgado}@ac.upc.edu

Abstract. This paper proposes a new way of searching resources, specifically images, through the web, following a broker approach. After studying some of the most important image repositories, Panoramio, Picasa and Flickr have been chosen as the resource databases where the application looks for images. Apart from the central application, which implements the search engine, a web portal has been developed in order to provide a user interface front-end where users can perform some searches in a friendly way.

Keywords: Metadata interoperability, broker, image search.

1 Introduction

Nowadays, there are many web sites that provide a huge amount of resources. Each resource has some metadata associated, which changes from one server to another. This causes some interoperability problems between each metadata model that makes difficult performing searches against different servers.

Our goal is to implement an image searcher that can handle, in the best possible way, the negative effects of this lack of interoperability, following a broker approach.

This paper is organized as follows: Section 2 is an overview of the state-of-the-art in the use of some standard metadata models and languages to request multimedia content, and in the use of broker approaches. Section 3 is an overview of which modules integrate the whole system we have developed and how they interact with each other.

The conclusions and future work are shown in section 4.

2 State-of-the-Art

Because this is not a new problem, a lot of research work has been already done on trying to solve it. This implies, from designing a broker architectural approach or metasearches [1][4][5] to specifying standard metadata models and languages [2][6][8], to accept and respond to requests for multimedia content without the concern about where they come from [3].

2.1 Designing a Broker Architectural Approach

There are many approaches to try to solve the interoperability problem [1]. One of them is based on a broker, also called query rewriting because its job is to receive search queries in a determined metadata format and to rewrite them once for every metadata format that is supported.

After the queries have been generated, they are sent to every external server and the resulting data is processed again and mapped back to the original metadata format.

This means that it is necessary to use a metadata model in order to establish some translations, or mappings, from this model to the rest.

2.2 Working on Standardizing a Multimedia Metadata Model

There is not a metadata model considered to be the standard to use. There are several models each one of them focusing on different needs and kinds of resources.

For example, the MPEG-7 standard is a complex model that aims to describe audiovisual material such as video, sound or images [6].

Another example is the Dublin Core model, which is oriented to multimedia and text content and whose core is composed by only 15 metadata elements, but has been highly extended by some working groups [8].

The Dublin Core model is widely used in many environments, such as the eXtensible Metadata Platform [9], which is a standard for processing and storing standardized and proprietary information relating to the content of a file. This is used by the Media Annotations Working Group [7], who wants to provide an ontology and an API designed to facilitate cross-community data integration of information related to media objects in the Web, such as video, audio and images.

An interesting model that has been recently standardized is the JPSearch Core, which is the metadata core of the JPSearch framework, which wants to provide a way to decouple the in general tightly coupled systems and to provide a better interoperability during image search [2].

JPSearch goal is to provide a standard for interoperability for image search and retrieval, which matches our objective.

2.3 Defining a Language to Request for Multimedia Content

The MPEG working group has defined a language, which became an ISO/IEC standard in December 2008, to accept and respond to requests for multimedia contents. It is called MPQF (MPEG Query Format) [3].

MPQF is an XML-based query language that defines the format of queries and responses to be interchanged between clients and servers in a distributed multimedia information search-and-retrieval context. It provides interoperability between parties, e.g. clients and content providers, and platform independency.

A subset of MPQF is also used in the JPSearch part 3, which is the part that provides a standardized message protocol for image retrieval; it is called JPQF (JPSearch Query Format).

3 The Developed System

The software we have developed provides a centralized place for searching images from different images servers. In our current version we are using Panoramio, Picasa and Flickr.

Our system is compliant with the current version of the JPSearch standards. From an architectural point of view, the system follows a broker architectural approach (see section 2.1), which means that there is a subsystem that receives JPQF queries in a metadata format and rewrites them once for every metadata format that is supported (Panoramio, Picasa and Flickr). This is sketched in Figure 1.

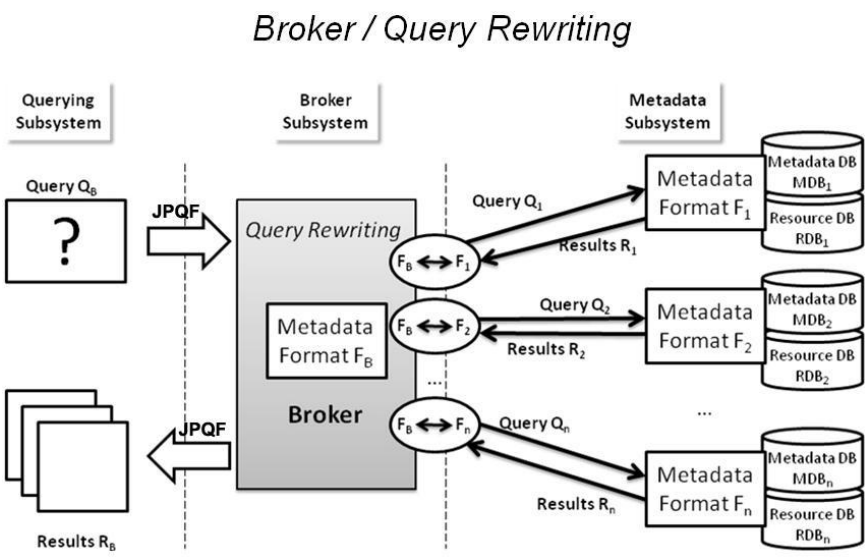


Fig 1. A broker architectural approach

The system is split into two modules: a central application that implements a broker-based metadata aggregator and a web portal as the user front-end. Modules communicate each other by using JPQF queries.

To understand how they all work together we first need to understand what each module does.

3.1 Central Application

This module takes as the input a JPQF query referring to metadata elements compliant with the JPSearch Core Schema (defined in ISO/IEC 24800-2) and returns a JPQF response.

3.1.1. Metadata Model

As mentioned before, the subsystem receives a query in a metadata format and rewrites it once for every metadata format that is supported. So, one metadata model must be used in order to establish the appropriated translations between that model and the rest.

The JPSearch Metadata Core has been chosen to be the default metadata model and has been extended with some image specific fields and with some EXIF fields, like the camera make and model.

The example in Figure 2 illustrates the model used by the application through a real search result. All the image dependent metadata fields like width, height or EXIF are included in the Photo element. Whenever it is possible, the other fields are grouped by meaning, for example, the Name and Nick fields, which are related with the Author, are included in the Author element.

```
<Metadata>
  <Title>Prenent un Bany. Taking a Bath.</Title>
  <Description>Bigui. beagle.</Description>
  <Keywords>landscape,yellow,orange,groc,taronja,Catalunya</Keywords>
  <URI>http://www.flickr.com/photos/arturdebat/1382400520/</URI>
  <Source>Flickr</Source>
  <License>1</License>
  <Author>
    <Name>Artur Debat</Name>
    <Nick>¡arturii!</Nick>
  </Author>
  <Date>
    <Created>2007-09-10 16:21:34</Created>
    <Published>2007-09-14 20:45:50</Published>
  </Date>
  <Location>
    <Latitude>41.5475158691406</Latitude>
    <Longitude>2.39347410202026</Longitude>
  </Location>
  <Photo>
    <URI>
      <Source>http://farm3.static.flickr.com/2/1892.jpg</Source>
      <Thumbnail>http://farm3.static.flickr.com/2/1892t.jpg</Thumbnail>
    </URI>
    <Width>500</Width>
    <Height>333</Height>
    <Orientation>1</Orientation>
    <EXIF>
      <Make>Canon</Make>
      <Model>Canon EOS 400D DIGITAL</Model>
    </EXIF>
  </Photo>
</Metadata>
```

Fig 2. Example of the metadata model in XML format.

3.1.2. The Broker Interface

In order to search on each server, an interface per-server must be provided. Each interface, depending on what kind of searches it receives, must know if it is possible to search on its associated server and, if so, how to launch the search and what to do with the results later.

There is a simple way to decide whenever it is possible or not to search on a server just knowing which metadata fields have their results and which of them it accepts to perform a search.

Knowing how to launch a search against a server is not only limited by using its API but, in order to expand the kind of searches that its server accepts, to properly modify the original query, for example by removing some conditions before launching it and then filtering the results by the conditions previously excluded.

For example, imagine that somebody wants to search images from Barcelona higher than 200 pixels in a server that does not allow searching by height. This search constraint will not be considered at searching time but after getting the results from Barcelona they will be reprocessed by height greater than 200 pixels.

3.1.3. Reprocessing Results

The reprocess can hence be thought as a filtering process that is done when there are search constraints that are not supported by a server.

This way of looking for images increases the search options that servers offer by default, but, on the other hand, from a user point of view, some searches can look slow because, internally, the application repeatedly launches several searches and reprocesses the results until all constraints are resolved, and that can take a while depending on how restrictive the given constraints are.

The time that these searches take can be minimized by scattering them across threads and launching them in parallel.

In order to be able to reprocess content obtained as the result of a search, they must be temporary kept in some way, for example inserted in a database. Since there is no need to keep them after being returned to the user, an in-memory database engine has been used in order to improve the efficiency.

3.1.4. Implementing a JPQF Interpreter

Our Web portal and the central application communicate each other by using JPQF queries but, internally, the application, which is programmed with Java, works with a set of Java classes that represent the JPQF query, not with JPQF expressed on XML.

This means that some kind of translation must be done from a JPQF query to the internal set of classes and vice versa. So, an interpreter that parses and transforms each incoming JPQF query must be implemented.

Since there is nothing to parse when the opposite transformation is realized, it is easier than the previous one. The interpreter only has to be aware of the JPQF query's OutputDescription, which is a part of the JPQF query, and to specify how the results

will be returned, in order to properly construct the JPQF response. For example, it is possible to specify which fields will be returned with the results, in which order they will be shown or how many of them will be returned. This is done by iterating each result's metadata field set and filling them into the JPQF.

3.2 Web Portal

The web portal is form-based and allows users to search images at different places without being concerned of how they are implemented as well as specifies some display options like sorting or the maximum number of results by page. When a user submits the form, a JPQF query representing the specified search and the display constraints is created and is sent to the application.

3.2.1 Another JPQF Interpreter

As well as we did before, a JPQF interpreter must be implemented on the client side in order to translate from HTML to JPQF and vice versa.

The HTML to JPQF translation can be done in a similar way the metadata model to JPQF translation does. Every form field becomes almost directly a JPQF condition that, concatenated with the rest using a properly Boolean operator (AND, OR), conforms the whole JPQF condition set.

Because JPQF is XML-based, the opposite translation, JPQF to HTML, can be accomplished just applying an XSLT on the JPQF response.

3.2.1 Web Design

The web is form-based, which means there are some input fields, which users can fill in with text, each one of them represents a different search field.

There is some javascript code that adds some dynamic behaviour to the search form. For example, because the form is too big, it is divided by sections (e.g. date, location, etc.) and each one of them can be collapsed resulting in a smaller form.

Also, it can be duplicated, as it is shown in figure 3, allowing launching multi-queries. In fact, they will be all assembled with an OR operand and sent to the application using a single JPQF query, but the application's JPQF interpreter will split it into several searches, one by each form, and the broker will launch them in parallel.

The search results are presented to the user as a table of thumbnails, as figure 4 shows. In order to see the information associated to each result, users must click on the appropriated thumbnail. A pop-up similar to which is shown in figure 5 will appear.

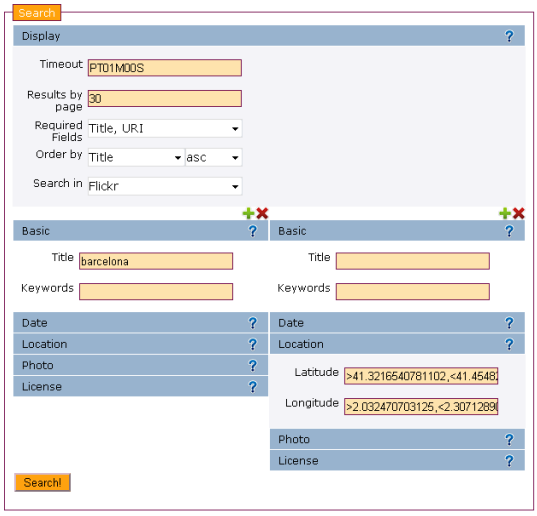


Fig 3. Example web form

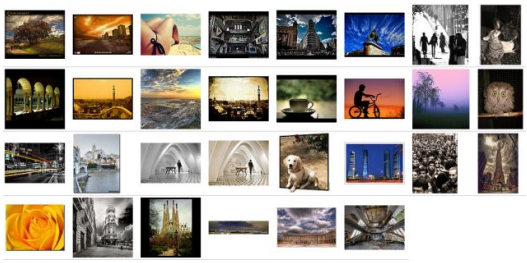


Fig 4. Result set shown in graphic mode

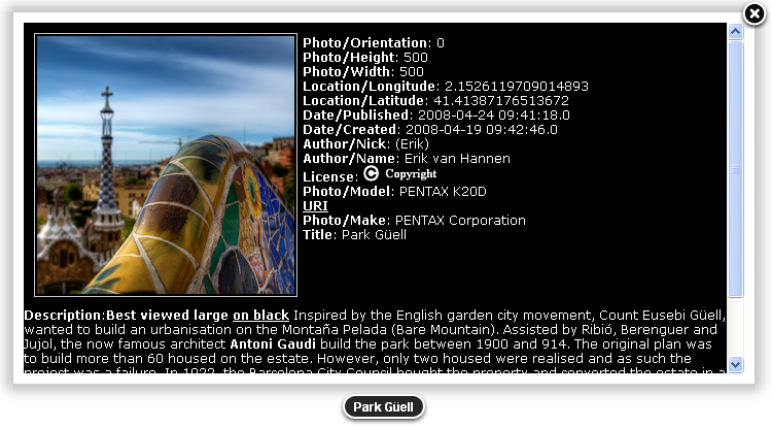


Fig 5. Result item in graphic mode

4 Conclusions and Future Work

This paper has described the design and the implementation of an image searcher that follows a broker architectural approach using some JPSearch standard parts like its metadata core and its language to retrieve images.

The main objective has been to be able to launch queries over several public image servers independently of the metadata elements used by them.

This system has been presented at the last ISO/IEC JTC1 SC29/WG1 [10] (JPEG) meeting in March 2010 and has been selected as an application to promote the use of JPEG standards. Nevertheless, our approach may work with different metadata and querying standards. For example, we have also available an implementation using the query standard from MPEG: the MPQF.

For the near future we will extend the application in order to support more content servers and more content types like audio or video.

5 Acknowledgments

This work has been partly supported by the Spanish government through the project MCM-LC (TEC 2008-06692-C02-01).

6 References

1. Toni Sala, Jaime Delgado, Rubén Tous: “Possible approaches for metadata interoperability in audiovisual content search”. DMAG Internal Report (2010). Available at http://dmag.ac.upc.edu/
2. JPSearch Metadata Core, http://www.jpeg.org/
3. MPEG Query Format standard, http://www.mpqf.org/
4. Rubén Tous, Jaime Delgado: “Advanced Meta-search of News in the Web”. Proceedings of the 6th International ICCC/IFIP International Conference on Electronic Publishing held in Karlovy Vary, Czech Republic, 6–8 November 2002.
5. Rubén Tous, Jaime Delgado, “Interoperability adaptors for distributed information search on the web”. Proceedings of the 7th ICCC/IFIP International Conference on Electronic Publishing held at the Universidade do Minho, Portugal, 25-28 June 2003.
6. MPEG-7 standard, http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm
7. Media Annotations Working Group, http://www.w3.org/2008/WebVideo/Annotations/
8. The Dublin Core Metadata Initiative, http://dublincore.org/
9. eXtensible Metadata Platform, http://www.adobe.com/products/xmp/
10. Jaime Delgado, Ruben Tous, Pere Toran, Toni Sala: “Demonstration of two JPSearch (ISO/IEC 24800) Real Applications”, ISO/IEC JTC1 SC29/WG1 document wg1n5316, March 2010. Available at http://dmag.ac.upc.edu/

A Metadata Model for Peer-to-Peer Media Distribution

Christian Timmerer[1], Michael Eberhard[1], Michael Grafl[1], Keith Mitchell[2], Sam Dutton[3], and Hermann Hellwagner[1]

[1] Klagenfurt University, Multimedia Communication Group
{*firstname.lastname*}@itec.uni-klu.ac.at
[2] University of Lancaster, Computing Department, InfoLab21
k.mitchell@lancaster.ac.uk
[3] BBC R&D Prototyping
Sam.Dutton@bbc.co.uk

Abstract. In this paper we describe a metadata solution for a Peer-to-Peer (P2P) content distribution system termed NextShare. We outline the key motivating factors for our approach, detail the overall generic architecture we have developed and present the workflow for delivering metadata through Peer-to-Peer based content distribution. The paper also presents the metadata model we have developed and we describe in detail how all the content can be packetized and distributed using NextShare. Finally, a description of the core and optional metadata attributes which may be utilized within the system is provided.

Keywords: Metadata, peer-to-peer streaming, social networks, payment and cashing, advertisements, MPEG-21 Digital Item, API.

1 Introduction

The Internet is increasingly being used to distribute both real-time and on-demand high bandwidth multimedia content to large audiences due, in part, to the increase in bandwidth available within the last mile. The server and bandwidth costs for provisioning adequate resources to facilitate high Quality of Experience (QoE) for streaming services to the end user are rapidly increasing as High-Definition (HD) content becomes increasingly dominant. One alternative to the traditional client-server or Content Distribution Network (CDN) approach is provided by Peer-to-Peer (P2P) distribution systems. In a decentralized P2P system the notion of a server does not really exist and instead all peers/nodes within a network are capable of distributing content to other peers while simultaneously consuming content. Thus, the distribution cost is shared amongst the peers themselves, potentially significantly reducing the distribution costs to content providers.

Although the number of available P2P systems is already very large and new P2P systems are frequently developed, the metadata utilized within P2P systems is usually proprietary and has not yet been standardized. In this paper, we propose a metadata model that builds upon existing standards and extends them to support P2P-specific requirements. The requirements are motivated by our work in the EU funded P2P-Next project [1] which seeks to develop an open source, standards-based P2P content

distribution platform, herein referred to as *NextShare*. The metadata model provides a solution to describe the content in P2P systems as well as a solution for structuring and packetizing the metadata and the actual audiovisual (A/V) content.

The remainder of this paper is organized as follows. Section 2 describes the general architecture and metadata workflow before we present the metadata model with its core and extensions in Section 3. Section 4 provides details on the metadata specification we have developed which is followed by an overview of the API for creation and access of metadata in Section 5. Some concluding remarks are presented in Section 6 along with some areas of future work.

2 Architecture and Workflow

This section describes a generalized architecture and workflow enabling P2P media distribution as depicted in Figure 1. The media to be distributed via P2P consists of (A/V) content and metadata. The metadata is divided into core metadata and additional (optional) metadata (i.e., extensions of the core metadata; cf. Section 3 for more details). Finally, the relationship between the A/V content and the metadata is described by structural metadata that represents the declaration of the P2P media. The combination of all these assets is referred to as the P2P-Next Item depicted in Figure 1 and is defined as follows: A/V content + metadata (core + extensions) + structure. The torrent file is generated based on these assets from the P2P-Next Item and is used for distribution via the NextShare platform [1]. Additionally, the metadata of the P2P-Next Item is used within the Atom/RSS store for the presentation/interactivity layer at the content consumption side.

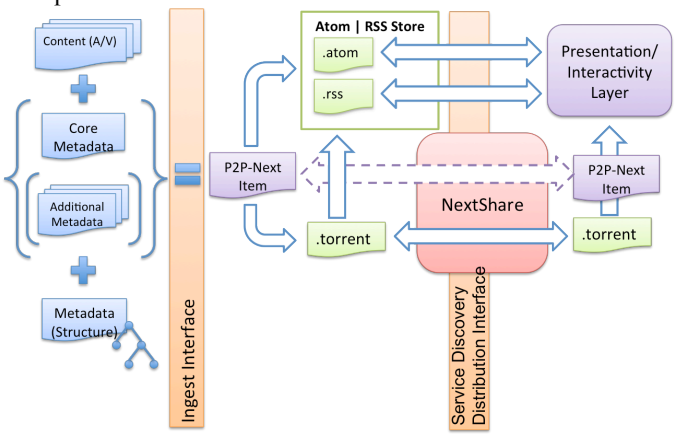


Figure 1. Architecture and Workflow for P2P Media Distribution.

A key requirement for the P2P-Next project is to ensure backwards compatibility with other BitTorrent clients. Thus, a torrent file compatible with the BitTorrent protocol [2] needs to be provided as top-level information. However, torrent files contain only a small portion of the metadata (i.e., the core metadata) needed to represent rich media content. The high-level structure of the P2P-Next Item is shown in

Figure 2. The torrent file contains the references and hash values for the media resources, i.e., the MPEG-2 Transport Stream (TS) containing the video and the audio content, and possibly the Scalable Video Coding (SVC) enhancement layers in case the video content is scalable. In particular, an MPEG-21 Digital Item Declaration (DID) [3] is included in the torrent file which might be encoded as binary XML or just provided as plain XML. The DID included in the torrent file contains the core metadata from the P2P-Next Rich Metadata specification [4] and references to other, optional, metadata and resources packaged separately. One way to package the optional content would be to packetize it into an MPEG-21 file (.m21) [5], including an additional DID that describes the individual optional content items. Alternatively, the additional DID could be provided on its own (i.e., as plain XML) and reference the optional content items, which could be, e.g., distributed through the NextShare system. The main reason for storing only the core metadata directly in the torrent file is to keep the size of the torrent file as small as possible (by referencing the other data) and still provide sufficient data to enable search on the content of the torrent file. A more detailed description of the structure of the DID is provided in the next section.

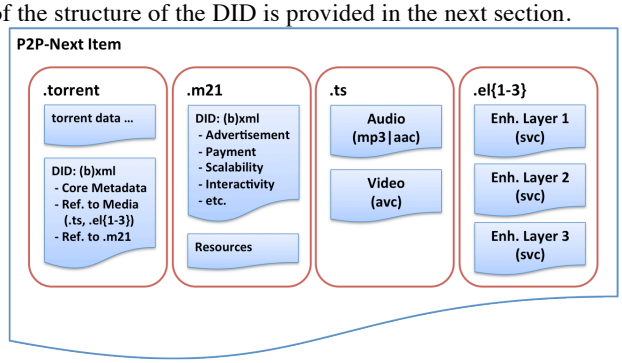


Figure 2. High-Level Structure of P2P-Next Item.

The DID for the optional content (packetized into an .m21 file in Figure 2) provides access to optional metadata related to advertising, media rating/review, content provider, payment, scalability, or interactivity. Furthermore, the metadata required for interactivity references additional resources (text, images, small audio/video clips, etc.) which could be included within an MPEG-21 file. Note that the MPEG-21 file format is based on the ISO base media file format that also provides the foundation for the well-known MP4 file format.

The actual A/V content is multiplexed within an MPEG-2 TS and is encoded with Advanced Video Coding (AVC) and MPEG-1 audio layer 3 (MP3) or Advanced Audio Coding (AAC) respectively. Additionally, enhancement layers for the video content are provided as separate bit-streams which are encoded with SVC having the base layer as part of the MPEG-2 TS.

3 Metadata Model

The P2P-Next Item contains all media resources, metadata, and possible additional data related to a single digital object. The P2P-Next Item is described by a DID which structures content and metadata. The DID contains relevant metadata and provides references to media resources and distributed metadata. The core metadata from the P2P-Next Rich Metadata (RM) specification are directly included in the top levels of the DID. All optional RM, such as for payment, advertising, media rating/review, and scalability, are referenced via XML Inclusions (XInclude) [6], as these metadata are only required for specific services and might be provided separately, e.g., on a secure server for payment. Additionally, data for interactivity is also referenced by means of XInclude. Within the P2P-Next project, the interactivity layer is termed LIMO (Lightweight Interactive Media Objects). LIMO data uses features of HTML 5 [7] and may include HTML, JavaScript code, style sheets, and/or additional media resources. As it is difficult to include media resources into XML files (i.e., base64 encoding is not a feasible solution for large media resources), the LIMO content can be packetized into the MPEG-21 file format, which contains another DID that references all the LIMO content within the MPEG-21 file. In this way, the part of the DID that is stored directly in the torrent file is kept as small as possible while retaining a complete DID structure and conveying the RM core metadata inside the torrent file for increased search performance. Thus, there are two (or more) DIDs within the NextShare system representing the P2P-Next Item.

The main (or master) DID conveys the overall structure of the P2P-Next Item and core metadata. It is stored inside the torrent file. Any additional DID can be stored in the XML box of an MPEG-21 file and can contain the additional metadata as well as data for LIMO. The torrent file references that MPEG-21 file. The main document references parts of the additional document as described above. However, the conceptual DID model for the P2P-Next Item should be seen as one entity, only its physical representation is split into two (or more) documents. Please note that the usage of an MPEG-21 file is not mandatory. The second DID document does not necessarily need to be stored within an MPEG-21 file. If it is more advantageous to distribute the LIMO content and the additional metadata in separate files, the DID just needs to reference these files and they can be provided separately, e.g., through the NextShare system or on traditional servers. This approach is especially useful when only parts of the LIMO content might be required for playback for the user. The MPEG-21 file just provides one way to store all the additional data together, if such a packaging mechanism is desired.

Figure 3 outlines the metadata model based on MPEG-21 DID for a P2P-Next Item and, thus, for P2P media distribution. The building blocks of a DID are shown as part of the legend at the bottom of the figure. Shapes with a dotted outline indicate data that are not included directly within the main DID, but are included by reference, in order to ensure the small file size of the torrent file.

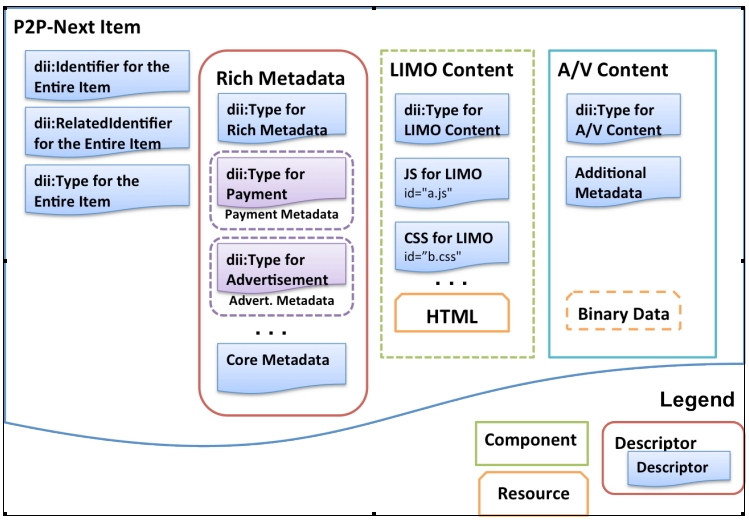


Figure 3. Metadata Model for P2P Media Distribution.

An `Item` represents the P2P-Next Item. The `dii:Identifier` is an MPEG-21 Digital Item Identifier (DII) [3] for the entire item, e.g., the Uniform Resource Name (URN) "urn:p2p-next:item:bbc-bbcone-b00n9p5x". The `dii:RelatedIdentifier` enables identification of the underlying work described by a Digital Item (DI). In this case, the `RelatedIdentifier` defines an "isAbstractionOf" relation to the underlying media content, e.g., identified by "urn:bbc:bbcone-b00n9p5x". This underlying media content is independent of the P2P-Next system. The `dii:Type` identifies the structure of this DID. It is set statically to the URN "urn:p2p-next:type:item:2009", thus determining the position within the DID and which building blocks are allowed. The structure of this DID is defined as part of [8].

The RM is represented following the Rich Metadata specification [4]. The actual RM is contained within a `Descriptor` and is structured as follows. The core metadata are contained within a `Statement` which is typically the last element in this `Descriptor`. Some nested `Descriptors` precede that `Statement`. The first one contains the `dii:Type` for the core metadata. This `Type` is set statically to "urn:p2p-next:type:rm:core:2009" and identifies the structure of the RM core metadata within a P2P-Next Item.

All additional RM assets are contained in further nested `Descriptors` which are referenced via XInclude. For each Descriptor, an `xi:include` element points to a `Descriptor` in the additional DID document. Each of these `Descriptors` has a `dii:Type` to identify its structure and consequently its purpose. Each `Descriptor` contains a `Statement` which conveys the RM representation of the appropriate RM part. More details on the URNs defined and utilized within NextShare are provided in Annex A of [8].

There are two `Components` in the DID for the P2P-Next Item. The first `Component` contains all data for LIMO. It is stored in the additional DID document inside a separate MPEG-21 file. Furthermore, it is included into the main document by means

of an `xi:include` element. The `dii:Type` in its first `Descriptor` is set statically to the URN "urn:p2p-next:type:limo:2009". Further `Descriptor`s contain resources required by the actual LIMO resource (i.e., the HTML content). These resources may be JavaScript files, Cascading Style Sheets (CSS) as well as JPEG or PNG images. Each `Descriptor` contains an `id` attribute, uniquely identifying that resource within the P2P-Next Item. It is proposed to use the original file name of a resource in all lower-case characters for the `id` value if applicable. The resources with text content are contained directly in a `CDATA` section of the `Resource`. On the other hand, resources with binary content, such as JPEG images, are bundled in the MPEG-21 file. The `Resource` of the `Component` contains the HTML page representing the LIMO content. As the required resources are not in actual files but rather in one MPEG-21 wrapper file, all references to the original files have to be replaced in the HTML document by the corresponding `id`s. For example, the reference to the file "script.js" would be replaced by "#script.js", a reference to the `Descriptor` with the `id` "script.js". These replacements could be accomplished by means of an XSL Transformation (XSLT). The HTML document for LIMO is contained in the `CDATA` section of the Resource.

The second `Component` in this `Item` represents the actual media content. The media content shall typically be packed into an MPEG-2 TS. The `Component` contains a `Descriptor` with a `dii:Type` which is set to "urn:p2p-next:type:content:2009". Furthermore, a second `Descriptor` may be present, conveying technical metadata about the TS (such as bitrate, size, etc.). The actual binary data of the TS is referenced through a `Resource` within the `Component`.

For examples of such DIDs the interested reader is referred to [8].

4 Metadata Specification

This section provides an overview of the core and optional metadata utilized within the NextShare system. More information on the presented metadata can be found in [4].

4.1 Core Metadata

The core metadata contains the essential content-related information that is required to search for a specific P2P-Next Item. All of the core metadata attributes need to be static, as these attributes are included within the torrent file and cannot be changed during the lifecycle of the torrent's swarm. The core metadata attributes include among others the title, the description, the genre, the creator, and also technical attributes like the resolution and the frame rate of the A/V content.

4.2 Optional Metadata

The optional metadata contain information that is related to the content and required for specific services, but might not always be needed for a certain P2P-Next Item. The individual optional metadata types are briefly described below.

– *Advertising* describes which types of advertisement are suitable and may be included in the P2P-Next Item. The attributes include the formats and advertisement types that may be used within the P2P-Next Item as well as information about the target group for advertising.

- *Payment* describes if the content is available for free or has to be paid for. The attributes include the price, the payment options and recipient, and information about possible donations.
- *Scalability* is only required if the video content is scalable and describes the properties of the scalability layers.
- *Media Review* reflects the perception of content by the audience such as user ratings. RM supports two kinds of media reviews. Individual media reviews performed by individual users and aggregated media reviews which summarize sets of individual media reviews. A media review comprises annotation of content by means of keyword tags and textual reviews, rating of content, rating of identity (e.g., to mark inappropriate content or spam), and rating of quality. The individual media review is always linked to its reviewer in order to prevent abuse.
- *User Profile* consists of core information (such as name and contact information) and additional information. The additional information ranges from usage preferences over usage history to social metadata. Social metadata comprises various kinds of information that a user might want to share with a specific subset of users, e.g., personal interests, education, employment, and information on social relations. User profile information is not included within the P2P-Next Item but is conveyed through other channels (e.g., a secure server). However, the P2P-Next Item contains a reference to its content provider.

5 Application Programming Interface

In order to utilize the above metadata model we have defined an Application Programming Interface (API) that serves two major purposes: Firstly, during the creation of the torrent file, the API is utilized to create the MPEG-21 DID describing the content package. Secondly, during the distribution of the content in the NextShare system, the API is utilized to access the different parts of the DID such as the core metadata. The API is implemented in C++ using the CubeWerx BXML library [9] and provides Python bindings for usage with the content ingestion tools implemented in Python (see Section 3.2.3 of [8]).

The API for NextShare provides two major interfaces. The *DIDCreator* enables creation of MPEG-21 DIDs containing the core metadata, references to optional metadata and to media resources, while the *DIDParser* allows extraction of the content from DIDs.

Due to space restrictions only an excerpt of the DIDParser API is provided below; for details the interested reader is referred to [8].

Method	Description
`const CW_XML_NODE* getContentByDiiType (CW_XML_SCAN* xmlScan, const char* diiType)`	Extracts the content from the DID document based on the specified `dii:Type`.
`const CW_XML_NODE* getContentByID (CW_XML_SCAN* xmlScan, const char* id)`	Extracts the content from the DID based on the `Descriptor`'s `id`. The `Descriptor id` is only utilized for LIMO content.

Interestingly, as the `getContentByDiiType` method returns parts of the DID identified by a standardized URN (i.e., based on the specified `dii:Type`), these parts can be accessed (i.e., parsed) by a software module conforming to the API that corresponds to the given `dii:Type`. For example, the parts of the DID representing the core rich metadata (i.e., identified by "urn:p2p-next:type:rm:core:2009") are handed over to the core RM API implementation and LIMO (i.e., identified by "urn:p2p-next:type:limo:2009") is forwarded to the LIMO implementation. In this way, interoperability is guaranteed through the usage of standardized (meta-)data formats and well defined APIs and URNs.

6 Conclusions and Future Work

In this paper we have presented the architecture and metadata model utilized within the NextShare system. In particular, we have shown how the A/V content, the core and optional metadata as well as the metadata structure can be packetized into a P2P-Next Item and distributed through the NextShare system. The major advantage of our approach is *interoperability* – thanks to the usage of existing, standardized representation formats for both media and metadata – and *backwards compatibility* to the well-known BitTorrent protocol.

NextShare including our metadata model and the core metadata (cf. Section 4.1) has been successfully demonstrated at various events (e.g., IBC'09, NEM-Summit'09; cf. [1] for details) and is currently being evaluated within the P2P-Next Living Lab [10]. Future work includes the complete definition, design, implementation, and validation of the optional metadata as outlined in Section 4.2 and full support of LIMO content within our metadata framework.

Acknowledgments. This work was supported in part by the European Commission in the context of the P2P-Next project (FP7-ICT-216217).

References

1. The P2P-Next Project, http://www.p2p-next.org
2. B. Cohen, "The BitTorrent Protocol Specification", http://www.bittorrent.org
3. Burnett, I.S., Davis, S.J., Drury, G.M.: MPEG-21 Digital Item Declaration and Identification - Principles and Compression. IEEE Transactions on Multimedia, vol. 7. no. 3, pp. 400--407 (2005)
4. P2P-Next D5.3.1, "Tools for Rich Metadata and Signposts", 2009, http://www.p2p-next.org/download.php?id=C48E37B2671AE3F3B8459942B2C7E484
5. Singer, D., Visharam, M.Z.: MPEG-21 File Format White Paper, ISO/IEC JTC 1/SC 29/WG 11N7925, Montreux (2006)
6. W3C Recommendation, "XML Inclusions (XInclude) Version 1.0", http://www.w3.org/TR/xinclude/
7. W3C Working Draft, "HTML5", 25 February 2010, http://dev.w3.org/html5/spec/Overview.html
8. P2P-Next D5.2.1, "Content Packaging, Ingestion, and Adaptation", 2009, http://p2p-next.org/download.php?id=98079FB406C2CBD78766E20BBF531F4E
9. CubeWerx BXML library, http://www.cubewerx.com/bxml
10. P2P-Next Living Lab, http://www.livinglab.eu/

Enabling Interoperability between Multimedia Resources: An Ontology Matching Perspective

Nicolas James, Konstantin Todorov, and Céline Hudelot
{nicolas.james,konstantin.todorov,celine.hudelot}@ecp.fr

MAS Laboratory, École Centrale Paris, F-92 295 Châtenay-Malabry, France

Abstract. The semantic annotation of images can benefit from representations of useful concepts and the links between them as ontologies. Recently, several multimedia ontologies have been proposed in the literature as suitable knowledge models to bridge the well known semantic gap between low level features of image content and its high level conceptual meaning. Nevertheless, these multimedia ontologies are often dedicated to (or initially built for) particular needs or a particular application. Ontology matching, defined as the process of relating different heterogeneous models, we will argue, is a suitable approach to solve interoperability issues in semantic image annotation and retrieval. We propose a generic instance-based ontology matching approach, applied to an important semantic image retrieval issue: the bridging of the semantic gap by matching a multimedia ontology against a common-sense knowledge resource.

1 Introduction

The fast growth of shared digital image and video collections together with the intensive use of visual information for decision making in many domains (medicine, geosciences, etc) require new effective methods for search and retrieval in these collections. In order to enable and improve the communication and the interface between humans and computers, it is necessary to understand the semantic content of images and to built linguistic descriptions of their content in an automatic way. Following decades of research on Content Based Image Retrieval (CBIR), automatic image annotation is nowadays an active research topic which aims at bridging the semantic and the perceptual levels of abstraction, known as the *Semantic gap* problem [11]. In most of the image annotation approaches, the computed linguistic description is often only related to perceptual manifestations of semantics. Nevertheless, as explained in [5], the image semantics cannot be considered as being included explicitly in the image itself. It rather depends on prior knowledge and on the context of use of the visual information. In consequence, explicit semantics, represented by ontologies, has been intensely used in the field of image retrieval recently.

With the growth of the application of ontology-based solutions in the multimedia domain, a lot of interoperability issues have arisen: (a) *At the semantic level* – between different representations of the same domain knowledge; (b) *At*

the visual level – between different multimedia ontologies; (c) *Between the visual level and the semantic level*, i.e. the semantic gap problem. Ontology matching, widely used for semantic web applications and rarely in the context of image sharing and retrieval, that we defined as the process of relating heterogeneous knowledge models, can be used to solve these kinds of interoperability issues. This paper proposes a generic approach to address the question of filling the semantic gap by matching an ontology at the semantic level (Wordnet[1] associated to the image database LabelMe[10]) with an ontology at the visual level (LSCOM [12]).

Next section is a short review of existing multimedia ontologies and related approaches. Section 3 describes the ontology matching framework which forms the methodological background of our approach, presented in turn in Section 4. Results of our preliminary experiments are discussed in Section 5; Section 6 concludes.

2 Related Work

In the past few years, *concept-based multimedia retrieval* has been a very active research field with a major effort in the automatic detection of semantic concepts from low level features with machine learning approaches. Despite these efforts, the semantic gap problem is still an issue for the semantic understanding of multimedia documents. Recently, many knowledge models have been proposed to improve multimedia retrieval and interpretation by the explicit modeling of the different relationships between semantic concepts. Indeed, many generic large scale multimedia ontologies or multimedia concept lexicons together with image collections have been proposed to improve multimedia search and retrieval by providing an effective representation and interpretation of multimedia concepts [13,12,1]. We propose to classify these ontologies in four major groups: (1) semantic web multimedia ontologies often based on MPEG-7, reviewed in [1] (2) visual concept hierarchies (or networks) inferred from inter-concept visual similarity contexts (among which VCNet based on Flickr Distance [15] and the Topic Network of Fan [3]), (3) specific multimedia lexicons often composed of a hierarchy of semantic concepts with associated visual concept detectors used to describe and to detect automatically the semantic concepts of multimedia documents (LSCOM [12], multimedia thesauri [13]) and (4) generic ontologies based on existing semantic concept hierarchies such as WordNet populated with annotated images or multimedia documents (ImageNet [2], LabelMe [10]). These ontologies have proved to be very useful mainly in the context of semantic concept detection and automatic multimedia annotation but many problems still remain unsolved among which enabling the interoperability between visual concepts and high level concepts. Although there exist attempts to solve these problems by manual concept mappings [13], little effort has been directed towards performing them in an automatic manner. Moreover, these ontologies are often dedicated to (or built for) particular needs or a particular application and are

[1] `http://wordnet.princeton.edu/`

complementary knowledge sources. While studies have been done to analyze the different inter-concept similarities in different multimedia ontologies [8], to the best of our knowledge, there are no studies which propose a cross analysis and a joint use of these different and complementary ontologies.

This paper proposes to situate these problems in an O[ntology] M[atching] framework. The OM-approach presented in next section is much in line with the tradition of *extensional* matching. This comprises a set of techniques which base the similarity of concepts on characteristics of the instances that these concepts contain [6].

3 An Ontology Matching Approach

An ontology is based on a set of *concepts* and *relations* defined on these concepts, which altogether describe the knowledge in a given domain of interest. Due to the fact that different communities, independently from one another, tend to conceptualize differently the same domain of interest, a growing number of *heterogeneous* ontologies, describing similar or overlapping parts of the world are created. An OM procedure aims at reducing this heterogeneity by linking the correspondent elements of two ontologies in an automatic or semi-automatic manner.

Formally, a **populated ontology** will be defined by $O = \{C, \texttt{is\_a}, R, I, g\}$, where C is a set whose elements are called concepts, `is_a` is a partial order on C, R is a set of other (binary) relations holding between the concepts from the set C, I is a set whose elements are called instances and $g : C \rightarrow 2^I$ is an injection from the set of concepts to the set of subsets of I.

We note that the sets C and I are compulsorily non-empty, in contrast to R. Thus, the definition above describes an ontology which, although not limited to subsumptional relations, necessarily contains a hierarchical backbone, defined by the partial order. The set I may contain text documents, images or other (real world data) entities. By assumption, every instance can be represented as an n-dimensional real-valued vector, defined by n input *variables* of some kind which are the same for all instances in I.

In the context of semantic image annotation, WordNet together with the LabelMe database [10] and LSCOM [12] together with the TRECVID 2005 database are two examples of such populated ontologies. Concepts are the nodes of the WordNet hierarchy in ImageNet or the LSCOM categories, while instances are the images in the associated databases, which are labeled by these concepts. It is important to note that the set R is empty for the LSCOM ontology. In the case of WordNet, R contains several useful relations like `is_a_member_of`, `is_a_part_of`, `opposes`, etc.

Often the outcome of an OM-procedure is a set of cross-ontology concept alignments, issued from a **measure of concept similarity**. The measures used in the current study are based on *variable selection* and we will describe them in more detail.

Variable selection techniques (reviewed in [4]) serve to rank the input variables of a given problem (e.g. classification) by their importance for the output (the class affiliation of an instance), according to certain evaluation criteria. A real valued *score* which accounts for this importance is attached to every variable. In our case, this can be of help for uncovering latent input-output dependencies. Assuming that instances are represented as real-valued vectors, the computed scores would indicate which of the vector dimensions are most important for the separation of the instances (within a single ontology) into those that belong to a given concept and those that do not and thus best characterize this concept.

We define a binary classification training set S_O^c for each concept c from an ontology O by taking I, the entire set of instances assigned to O and labeling all instances from the set $g(c)$ as *positive* and all the rest $(I \backslash g(c))$ as *negative*. By the help of a variable selection procedure performed on S_O^c, we obtain a representation of the concept c as a list

$$L(c) = (s_1^c, s_2^c, ..., s_n^c), \tag{1}$$

where s_i^c is the score associated to the ith variable. To compute a score per variable and per concept, we apply the S[upport] V[ector] M[achine]-based variable selection technique introduced in [14]. A series of SVMs is learned on the training set S_O^c by subsequently removing a variable at a time. The ability of each variable to discriminate c from the other concepts in O is evaluated by measuring the sensitivity of the VC-dimension, an important SVM parameter, with respect to the variable in question.

By following the described procedure, given two source ontologies O_1 and O_2, a representation as the one in (1) is made available for every concept of each of these ontologies. The similarity of two concepts, $A \in O_1$ and $B \in O_2$ is then assessed in terms of their corresponding representations $L(A)$ and $L(B)$. Several choices of a similarity measure based on these representations are proposed and compared in [14]. In the experimental work contained in this paper, we have used Pearson's, Spearman's and Kendall's measures of correlation calculated on the variable scores or ranks (integers corresponding to the scores) given by

$$sim_{Pearson} = \frac{\sum_{i=1}^{n}(s_i^A - s_{mean}^A)(s_i^B - s_{mean}^B)}{\sqrt{\sum_{i=1}^{n}(s_i^A - s_{mean}^A)^2}\sqrt{\sum_{i=1}^{n}(s_i^B - s_{mean}^B)^2}}, \tag{2}$$

$$sim_{Spearman} = 1 - 6\frac{\sum_i d_i^2}{n(n^2-1)}, \quad sim_{Kendall} = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}. \tag{3}$$

In the formulae above, s_{mean}^A and s_{mean}^B are the means of the scores over all input variables, d_i is the difference of the ranks calculated for the ith variable w.r.t. the two concepts, and n_c and n_d are the numbers of concordant and discordant pairs among the lists of scores $L(A)$ and $L(B)$.

4 Filling the Semantic Gap with Mapped Concepts

As noted in the introduction, many challenging issues in the field of image retrieval stem from the semantic gap problem. Two examples are the construction of robust high level concept detectors and the creation of user oriented annotations with high level semantics. In this section, we propose an attempt to fill the semantic gap by matching two complementary resources: a *visual* and a *semantic* thesaurus. Contrary to [13], our approach is *automatic*, *generic* (ontology independent) and makes use of the *visual knowledge* shared by the source ontologies.

On one hand, we chose LSCOM [12], an ontology dedicated to multimedia annotation. It was initially built in the framework of TRECVID[2] with the criteria of concept usefulness, concept observability and feasibility of concept automatic detection. LSCOM is populated by the development set of TRECVID 2005 videos (news broadcasting). On the other hand, we used WordNet [9] populated with the LabelMe dataset [10]. Many interoperability issues can be addressed for these two ontologies among which semantic interoperability and semantic gap interoperability. Aligning these resources allows for the semantic enrichment of concepts belonging to a multimedia ontology with high level linguistic concepts from a general and common sense knowledge base and the evaluation of the quality of the baseline concept detectors by studying the link between concepts whose semantics is related to their perceptual manifestations and concepts whose semantics is related to common sense.

In our setting, the instances that extensionally define a concept are images whose annotations contain the name associated to this concept. An image is represented as a vector of descriptors. We use a codebook built on a bag-of-features model and histograms of codewords which is, nowadays, the best approach in the state-of-the-art [7]. In that, the variables which describe the instances are the bins of these histograms. The generic variable selection approach described in Section 3 is applied directly on our data. In result, we obtain a concept representation as the one introduced in eq. (1) for every concept of our two source ontologies. As stated above (Section 3), there exist several plausible choices of a measure of similarity for two concepts represented in this manner. In our experiments, we have tested the three measures of correlation given in (2) and (3). Regardless of the particular choice, the similarity is always based on *visual criteria*, since the underlying concept representations are obtained by using visual characteristics of the instances (in the particular case of LSCOM and WordNet these are the sets of images of either TRECVID or LabelMe).

Aligning LSCOM to WordNet allows to infer knowledge about the LSCOM concepts (dedicated to the multimedia document annotation) with regard to the concepts of WordNet and the alignment could be used to build a linguistic description of the concepts of LSCOM, or, in other words, to answer the question *"What is an LSCOM concept in WordNet?"* in an automatic manner. This improves the retrieval process in several ways: (1) through query expansion and

[2] http://www-nlpir.nist.gov/projects/tv2005/

reformulation, i.e. retrieving documents annotated with concepts from an ontology O_1 using a query composed of concepts of an ontology O_2, (2) through a better description of the documents in the indexing process. However, note that this relation is not symmetric: alignments in the other sense are prompt to fail to be of any help, since WordNet concepts are rather atomic (such as "car") as compared to the more complex LSCOM concepts (e.g. "Natural Disaster Scene").

5 Experimental Results

We use a part of the LSCOM ontology, LSCOM_Annotation_v1.0[3], which is a subset of 449 concepts from the initial LSCOM ontology, and is used for annotating 61,517 images from the TRECVID2005 development set. Since this set contains images from broadcast news videos, the choosen LSCOM subpart is particularly adapted to annotate this kind of content, thus contains abstract and specific concepts (e.g. `196_Science_Technology`, `330_Interview_On_Location`). To the contrary, our sub-ontology defined from WordNet populated with LabelMe (3676 concepts) is very general considering the nature of LabelMe, which is composed of photographs from the daily life.

In this way, to provide a preliminary evaluation of the suggested approach, we chose three concepts from the LSCOM ontology and five concepts from the WordNet ontology. The choice of the selected concepts was made on several criteria: (1) the number of associated instances, (2) for every selected concepts there is no semantic ambiguity in our dataset, (3) for WordNet only: a high confidence (arbitrarily decided) in the discrimination of the concept using only perceptual information.

Table 1. LSCOM/TRECVID2005 against WordNet/LabelMe from left to right rowwise: Variable selection-based concept similarities with Pearson's, Spearman's and Kendall's coefficients and a manual annotation of the TRECVID2005 images.

Concept Names	Man	Car	Boat	TV	House	Man	Car	Boat	TV	House
Natural Disasters	**0.28**	0.12	-0.08	-0.02	**0.17**	**0.37**	0.15	-0.33	0.12	**0.44**
US Flags	0.17	0.01	0.15	-0.15	0.063	**0.21**	0.09	0.01	0.05	0.05
Single Family Homes	**0.21**	0.13	-0.22	0.05	**0.23**	**0.20**	0.18	-0.36	0.13	**0.41**
Natural Disasters	**0.36**	0.19	-0.15	0.016	**0.37**	103	51	4	0	73
US Flags	0.15	0.13	0.09	0.012	0.11	434	16	0	2	28
Single Family Homes	**0.23**	0.19	-0.18	0.016	**0.35**	205	73	1	0	184

[3] http://www.ee.columbia.edu/ln/dvmm/lscom/

Fig. 1. An LSCOM image annotated by LSCOM:Single_Family_Homes, and four sample images from WordNet/LabelMe. The LSCOM image can be effectively annotated and retrieved using the WordNet concepts Man and House.

To construct image features, we use a bag-of-features model with a visual codebook, built classically using the well known SIFT descriptor and a K-Means algorithm. The quantification of the extracted SIFT features was investigated in two ways: (1) over all the instances associated to the selected concepts (LSCOM and WordNet), (2) only over the LabelMe images and quantification per concept. The two experimentations gave very similar results, and the results of the experiment based on the first codebook are resumed in Table 1.

The values in the first three matrices are correlations indicating high similarity for positive values (low for non-positive). As we can see, the concept WordNet:TV is weakly correlated to the chosen LSCOM concepts, and the concept WordNet:House is highly correlated with LSCOM:Natural_Disasters and LSCOM:Single_Familly_Homes but not with LSCOM:US_Flags. This is coherent with the TRECVID2005 data considering that the images annotated with LSCOM:US_Flags are mostly images from speeches of politicians during presidential elections. An example of an LSCOM image annotation that could be extended to WordNet by the help the concept mapping is given in Fig. 1.

6 Conclusion

The paper proposes an ontology matching technique to solve interoperability issues in the area of semantic image annotation and retrieval. In particular, we have addressed the problem of bridging the semantic gap by the help of a generic instance-based ontology matching approach which aims at automatically producing concept-based annotations enriched with a lexical description of the concepts. In preliminary experiments, we have tested a concept similarity measure on two small sets of concepts taken from the LSCOM ontology and WordNet/LabelMe. Our results are in good agreement with the nature of the

instances associated to the selected LSCOM concepts. However, the efficiency of the approach has to be tested on larger sets of concepts (currently in progress). A large-scale application would also allow us to benefit from all the semantic relations in WordNet, like hypernymy, meronymy, antonymy. In the future, we plan to investigate the qualities of our automatic approach in terms of retrieval efficiency as compared to approaches that solely rely on manual mappings.

Acknowledgments. This work is funded by the French National Research Agency (ANR) through the COSINUS program (project COLLAVIZ ANR-08-COSI-003) and by the region Île de France through the SEBASTIAN2 project (Cap Digital cluster).

References

1. S. Dasiopoulou, V. Tzouvaras, I. Kompatsiaris, and M.G. Strintzis. Enquiring MPEG-7 based multimedia ontologies. *Multimedia Tools and Applications*, pages 1–40, 2010.
2. J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, and L. Fei-Fei. ImageNet: a large-scale hierarchical image database. In *CVPR*, pages 710–719, 2009.
3. J. Fan, H. Luo, Y. Shen, and C. Yang. Integrating visual and semantic contexts for topic network generation and word sense disambiguation. *ACM CIVR'09*, pages 1–8, 2009.
4. I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *JMLR*, 3(1):1157–1182, 2003.
5. C. Hudelot, N. Maillot, and M. Thonnat. Symbol grounding for semantic image interpretation: from image data to semantics. In *SKCV-Workshop, ICCV*, 2005.
6. A. Isaac, L. van der Meij, S. Schlobach, and S. Wang. An empirical study of instance-based ontology matching. *The Semantic Web*, pages 253–266, 2008.
7. Y.G. Jiang, J. Yang, C.W. Ngo, and A.G. Hauptmann. Representations of keypoint-based semantic concept detection: A comprehensive study. *IEEE Trans. on Multimedia, in press*, 2010.
8. M. Koskela and A. Smeaton. An empirical study of inter-concept similarities in multimedia ontologies. In *CIVR'07*, pages 464–471. ACM, 2007.
9. G.A. Miller. WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41, 1995.
10. B.C. Russell, A. Torralba, K.P. Murphy, and W.T. Freeman. LabelMe: a database and web-based tool for image annotation. *IJCV*, 77(1):157–173, 2008.
11. A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. Patt. An. Mach. Intell.*, pages 1349–1380, 2000.
12. J.R. Smith and S.F. Chang. Large-scale concept ontology for multimedia. *IEEE Multimedia*, 13(3):86–91, 2006.
13. C.G.M. Snoek, B. Huurnink, L. Hollink, M. De Rijke, G. Schreiber, and M. Worring. Adding semantics to detectors for video retrieval. *IEEE Trans. on Mult.*, 9(5):975–986, 2007.
14. K. Todorov, P. Geibel, and K.-U. Kühnberger. Extensional ontology matching with variable selection for support vector machines. In *CISIS*, pages 962–968. IEEE Computer Society Press, 2010.
15. Lei Wu, Xian-Sheng Hua, Nenghai Yu, Wei-Ying Ma, and Shipeng Li. Flickr distance. In *MM'08*, pages 31–40. ACM, 2008.

Access control issues in Social Networks

Anna Carreras, Eva Rodríguez, Jaime Delgado, Xavier Maroñas

Distributed Multimedia Applications Group (DMAG)
Universitat Politècnica de Catalunya (UPC), Jordi Girona 1-3, E-08034 Barcelona, Spain
{annac, evar, jaime.delgado, xmaronas}@ac.upc.edu

Abstract. Social Networks, as the main axis of Web 2.0, are creating a number of interesting challenges to the research and standardisation communities. In this paper, we analyse the current and future use of access control policies in Social Networks. Subsequently, two main issues are addressed: the interoperability amongst systems using different policy languages and the lack of elements in the existing policy languages when trying to express Social Networks' access control. In particular, our approach is based on the use of the XACML standard.

Keywords: Privacy, social networks, access control policies, XACML.

1 Introduction

In the last few years, social networks have been actually *the* Internet phenomenon, and the main axis of the so-called Web 2.0, while creating a number of new interesting challenges to the research community. *Online social networks* are communities in the Internet, usually around one website, which connect users voluntarily sharing information. In this context, mainly due to the growing amount of (personal) data being shared nowadays through internet, users' concern about privacy has risen. From our previous work [1]-[4], we have identified two important issues that still need to be fully solved by the standardization and research communities. First, the existing standardized access control policy languages (i.e. Extensible Access Control Markup Language (XACML) [5]) are missing some elements when trying to express Social Networks current and future privacy policies. And second, the interoperability between different policy languages still needs to be solved. Thus, in the next section, we will first go into details of the aforementioned open issues, and then, in Section 3, we will present our initial approach to solve them, as well as some preliminary work done in this direction. Finally, Section 4 will conclude the paper.

2 Open issues on access control policies languages for Social Networks

As introduced in Section 1, Social Networks present new interesting challenges when trying to address the protection and/or governance of the shared data. In few words, they have created a highly dynamic environment in which users have a producer-consumer role and their actions are based on the idea of "trust". Furthermore, new types of "resources" need to be protected (such as "relationships" or "events"), and a high degree of expressiveness is demanded by users in order to define their own access control (privacy) preferences. Policy expressions mainly depend on the context of the access (apart from the nature of the "resource" that needs to be protected, and the "user's" characteristics). Although XACML has been proved to be flexible enough to describe any type of access control policy, it presents some limitations when applied to privacy protection for Social Networks. For example, as identified during the last W3C Workshop on Access Control Application Scenarios [6], it does not support neither credential based access control, which would allow a more user-friendly privacy protection for Social Networks, nor the use of sticky policies, suitable when addressing data handling. Furthermore, it lacks of attributes to guarantee semantic interoperability among different Social Networks. Indeed, this lack of semantic interoperability, apart from being an issue to be solved at an application-specific level, should be also addressed in a more generic way. Nowadays, different applications and services using different access control policy languages may also need to be interoperable. This time, the incompatibility is not only between different Social Networks but also between a number of heterogeneous services/applications, and thus the interoperability between the different access control policies languages may be even harder to achieve.

Users voluntarily share information, but not only content, also actions and personal information. In addition, service providers are collecting even more information on users' behaviour. However, not only this "voluntarily" provided shared information must be protected. There is an increasing amount of "third parties" which have seen a business opportunity in Social Networks, and are offering all kind of applications to these communities of users. It is important that users had means to decide the access control policies applying to "friends", but also to these "third parties". The implementation of an access control model based on a symmetric level of trust would be recommendable, for example, including the possibility of negotiating policies.

3 Our approach and preliminary work

In this section we address the two main issues previously identified: the lack of interoperability between different Access Control Policy Languages, and the lack of privacy protection when dealing with third-party applications within Social Networks. Thus, we propose a possible architecture which allows (Social Networks') users to control the access to their content without the need of giving it to the Social Network Provider and through the use of policies based on XACML. Furthermore, the use of

some translators (detailed in [3]) guarantees the interoperability between RELs without losing information. The proposed architecture is shown is Fig. 1.

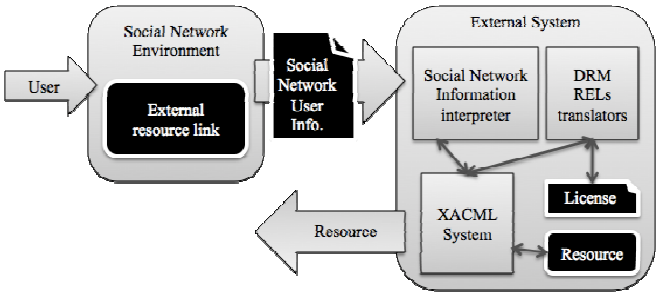


Fig.1 Proposed architecture for access control in Social Networks application scenario

A user would be able to publish an external resource link in her user profile in order to share some of her pictures stored in an External System (external from the Social Network, for example, it could be her private server) with some access control. Then, when another user would check that link, she would be redirected to the external system. The later would extract the necessary context from the Social Network and process the request. Finally, if the object license is not in the XACML language, the RELs translators would generate the appropriate policy and the result would be passed to the XACML system. This module is in charge of authorising the access, and is also detailed in [3]. If the authorisation were positive, the system would access the content, and would show it to the user. If not, it would just show the user a message telling him that she has no rights to do that.

As a next step, it would be very interesting (in the Social Networks application scenario) to give the opportunity to users and service providers to negotiate the access control policies. This is mainly due to the dynamicity of the application scenario being addressed in this paper, and in order to give the maximum control to users over the protection of their contents. For this purpose, a message expressing an "offer" instead of imposing a policy may be required. XACML, as well as RELs, can be used to express offers in which users of a system may propose to other users of the system usage rules for their content according to the rights and conditions that they negotiate. MPEG-21 REL [7] defines the "obtain" right for this purpose, which can be conceptualised as an advertisement to share or sale the associated grant. Within this grant, the rights and conditions initially stated by the offer maker will be defined. Then, in XACML, a similar mechanism can be used to provide this capability. Nevertheless, in [6], held in November 2009, the question on how to establish a user feedback channel in XACML stayed unsolved.

4 Conclusions

In this paper, some novel issues on the access control in Social Networks application scenario have been analysed. In particular, two main issues have been addressed. On the one hand, the interoperability among Social Networks which are using different policy languages and, on the other hand, the lack of elements of the current existing standards trying to express access control policies in Social Networks. In our approach, we have shown how the desired (syntactic) interoperability could be achieved by using policies and REL (Rights Expression Languages) translators in a distributed access control architecture based on XACML. And finally, we have presented how XACML could be used in the negotiation of access control policies. In the future, we will try to contribute on solving the lack of semantic interoperability among Social Networks. In this line, we are already studying some standardisation initiatives such as the Delivery Context Ontology (DCO) [8], the Friend Of A Friend (FOAF) project [9], and the work developed by the Policy Language Interest Group (PLING) [10].

5 Acknowledgments

This work has been partially supported by the Spanish government through the projects MCM-LC (TEC 2008-06692-C02-01) and Segur@ (Centre for the Development of Industrial Technology (CDTI), CENIT-2007 2004, under a subcontract with Safelayer Secure Communications).

6. References

1. E. Rodríguez, V. Rodríguez, A. Carreras, J.Delgado, "A Digital Rights Management approach to privacy in online social networks", in Proc. of the 1st Workshop on Privacy and Protection in Web-based Social Networks (within ICAIL '09), Barcelona, Spain, June 2009. IDT Series, vol. 6, ISSN 2013-5017.
2. V. Rodríguez, A.Carreras, E. Rodríguez, J. Delgado, "Applications to improve privacy on online social networks", in Proc. of the First Workshop on Law and Web 2.0, Antoni Roig (ed.), September 2009.
3. X. Maroñas, E. Rodríguez, J. Delgado, "An architecture for the interoperability between rights expression languages based on XACML", in Proc. of the 5th International ODRL Workshop (within Virtual Goods' 09), Nancy, France, September 2009, ISBN: 978-2-905267-69-6.
4. A. Carreras, J. Delgado, E. Rodríguez, R. Tous, "The Impact of Contextual Information on User Privacy in Social Networks", in Proc. of the 1st Workshop on Privacy and Protection in Web-based Social Networks (within ICAIL '09), Barcelona, Spain, June 2009. IDT Series, vol. 6, ISSN 2013-5017.
5. T. Moses (Ed.): eXtensible Access Control Markup Language (XACML) Version 2.0, Feb. 2005. [Online]. Available: http://docs.oasis-open.org/xacml/2.0/access control- xacml-2.0-core-spec-os.pdf.

6. W3C Workshop on Access Control Application Scenarios, 17-18 November 2009, Luxembourg. [Online] Available: http://www.w3.org/2009/policy-ws/.
7. International Standards Organisation. Information technology – Multimedia Framework (MPEG-21) – Part 5: Rights Expression Language. ISO/IEC 21000-5:2004.
8. Delivery Context Ontology (DCO). W3C Working Draft 16 June 2009. [Online]. Available: http://www.w3.org/TR/2009/WD-dcontology-20090616/.
9. The Friend of a Friend (FOAF) Project. [Online]. Available: http://www.foaf-project.org/.
10. W3C Policy Language Interest Group (PLING). [Online]. Available: http://www.w3.org/Policy/pling/wiki/Main_Page.

Live-Ticker Supported Sports Video Annotation

T. R. Mayer[2], D. Coquil[1], M. Döller[1], L. Limam[1], F. Stegmaier[1], and H. Kosch[1]

[1] Chair of Distributed Information Systems
University of Passau, Germany
[2] Laboratoire d'InfoRmatique en Image et Systèmes d'information (LIRIS)
INSA de Lyon, France

Abstract. Automatic semantic annotation of videos remains an open research problem. For soccer matches, live tickers, are very interesting sources of information. Indeed, being written by humans, they may contain pertinent semantic information that can be used to annotate the corresponding videos. This paper presents a prototype that processes live ticker texts and transforms them into machine-readable annotations. The analyzer produces semantic annotated actions as RDF-graphs. This annotation may be used for other purposes, such as tactical analysis orsemantic multimedia query processing.

Key words: Semantic video annotation, ontology, text analysis

1 Introduction

The analysis of sports videos is increasingly gaining in importance for professional and private consumption. Thanks to online video portals (e.g. YouTube), users can catch up with missed highlights of their favorite sport. To enable the retrieval of such data, annotation of this content is required. Several approaches have been proposed for automatic, semi-automatic or manual video annotation. In view of the amount of video material, automatic or at least semi-automatic techniques are preferable. As this annotator automatically processes live ticker texts written by human operators, it is a semi-automatic approach that benefits from human interpretation capabilities while still being able to cope with the mass of videos. Moreover, live tickers are available for free on the internet. They are thus a very good source of semantic annotations for sport videos. This annotation can be used for queries or also as support for video annotation systems. The remainder of the paper is structured as follows. After an overview of related work, the prototype architecture is introduced. This is followed by a summary of the challenges that have been faced. Finally, a conclusion is given.

2 Related Work

Soccer video analysis is an emerging research field in the domain of multimedia annotation. Early works [1] focused on the recognition of the edges of the soccer

field (e.g. penalty box). Later approaches dealt with for instance shot boundary detections with respect to events such as goals [7]. Besides, techniques have been introduced for player tracking [5] or player identification [2]. The trajectory of players and the ball has been analyzed in [3] to generate tactic annotations (e.g. "attack from left"). A first work targeting live ticker texts has been proposed in [6]. There, the individual entries in the text are assigned with the comments and timestamps to the video source.

3 Prototype Architecture

3.1 Current Prototype

The prototype consists of four modules: *Preprocessor*, *Text Analysis*, *Postprocessor* and *Annotation Engine* (see figure 1). A *Domain-Specific Web Ontology Language (OWL)* is used by the Annotation Engine. The text processing components, which have already been implemented, are described here in more detail. The Annotation Engine, which has not yet been realized is briefly outlined in section 3.2. The first three modules form the ticker text processing components, which create an informational structure for use by the Annotation Engine. The goal is to transform the raw ticker text into machine understandable data, and to enrich it by lexical and semantic information.

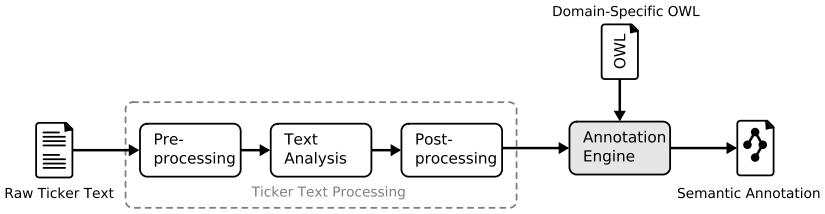


Fig. 1. Overview of the system architecture

Beginning with the *Preprocessor*, the raw ticker text, provided as a text file, is read. This character stream is then split into events. Each event represents a message submitted by the ticker author; it may consist of multiple sentences. The events are separated using timestamps provided in the original text, resulting in a list of tuples <*timestamp,event-text*>, which we call *event list*. In addition to In-Game (IG) events, the ticker text also contains Out-of-Game (OoG) events such as announcements. OoG events can be distinguished from IG event due to differences in the timestamp format; they are thus identified and removed using regular expressions. The result is an event list in which each entity is assigned to a phase of the match (first half-time, extra time of second half-time etc.) based on the timestamp and game rules. The last action of the preprocessor is to correct wrongly separated words, which usually occur for "-"-composed words with

special character. This most likely occurs because the ticker text is in German, whereas we used text analysis libraries dedicated to English.
The *Text Analysis* is the second part of the text preprocessing. It enriches the ticker events with lexical and semantic information. This is done in two steps. First, the Stanford Parser[3] performs lexical analysis and Part-of-Speech-Tagging (PoS-Tagging). Basing on statistical models, this parser tries to identify the lexical category (PoS-Tag) of a word in a sentence (e.g. verb, adjective). The library provides a model for German language. In the second step, the Stanford Named Entity Recognizer[3] or NE-Tagger identifies names in the event texts. As no German model is provided, training data have been used to let the NE-Tagger learn such a model. To this end, the ticker text has been annotated manually using domain-specific tags: PERSON, TEAM, LOC (locations, e.g. places or cities) and OTHER (for a word that is not a name). Using this as training data, a general model for the German language has been produced. Both, the PoS- and NE-Tagger use Sequence Labeling algorithms (see e.g. [4]) and calculate labels that are added to each word of an event text.

In the Text Analysis step, the PoS- and NE-Tagger split the event text into tokens and add their calculated labels. Due to different splitting rules, they may generate a different count of tokens for an event. The *Postprocessor* corrects this by merging multi-token entries, e.g. "Tim" and "Borowski" to "Tim Borowski". Furthermore, names are often written only as fore-/surname rather than completely. The Postprocessor corrects this by extending all names to their full version. This is done by comparing them with a lexicon, which has been developed by parsing a soccer website[4].

3.2 Future Work

An ontology has been created manually for the soccer domain to be used by the *Annotation Engine* (see figure 1), which has not yet been developed. This component will process text in two steps. The first step is similar to the present work: labels, which are references to OWL entities, will be assigned to tokens of the event text. A relation to a semantic context will then be assigned to each token. As a second step, the Annotation Engine will create an RDF graph using the domain-specific OWL and annotated event text tokens of the previous step. This graph can then be used for additional processing such as a tactical analysis.

4 Challenges

We had to face several challenges when developing our live ticker analysis prototype. As the ticker text is not machine understandable, it must first be transformed to a normalized format. This is a hard task due to different character encodings, different timestamp formats or incorrect punctuation. Writing styles

[3] http://nlp.stanford.edu/software
[4] http://www.fussball-wm-statistik.de

(e.g. jargon language) must be adapted. In the end, such a normalization engine is hard-linked to one specific ticker text provider. Another problem was that the analysis tools sometimes have trouble with special cases of word/sentence splitting. Finally, differences in timestamp formats complicated the assignment of an event to a phase of the game. As a summary, it seems that a tool with standardized output format would avoid many problems and enhance the ticker text production process.

5 Conclusion

This paper presents the current state of a live ticker analyzer for automatic generation of semantic annotations. A difficult problem is the lack of a standardized format for the ticker text. For now, texts must be normalized to a processable format for each provider, which is a hard task only valid until structural changes. Other issues such as unreliable punctuation complicate the semantic analysis. Once the text has been normalized and preprocessed, the annotation generation itself can take place. The result can be used for semantic multimedia queries and supporting additional annotation. In the future, the prototype will output an RDF graph including timestamps and representing game events. This will enable additional benefits such as the analysis of game tactics.

References

1. Gong, Y., Sin, L. T., Chuan, C. H., Zhang H., Sakauchi, M.: Automatic Parsing of TV Soccer Programs. Proc. of the International Conference on Multimedia Computing and Systems, 167–174 (1995)
2. Bertini, M., Del Bimbo, A., Nunziati, W.: Player identification in soccer videos. Proc. of the 7th ACM International workshop on Multimedia information retrieval. 25–32 (2005)
3. Zhu, G., Huang, Q., Xu, C., Rui, Y., Jiang, S., Gao, W., Yao, H.: Trajectory based event tactics analysis in broadcast sports video. Proc. of the 15th international conference on Multimedia, 58–67 (2007)
4. Nguyen, N., Guo, Y.: Comparisons of sequence labeling algorithms and extensions. Proc. of the 24th international conference on Machine learning (ICML '07), 681–688 (2007)
5. Miura, J., Kubo, H.: Tracking players in highly complex scenes in broadcast soccer video using a constraint satisfaction approach. Proc. of the International conference on Content-based image and video retrieval, 505–514 (2008)
6. Xu, C., Wang, J., Lu, H., Zhang, Y.: A Novel Framework for Semantic Annotation and Personalized Retrieval of Sports Video. IEEE Transactions on Multimedia, Vol. 10, Issue 3, 421–436 (2008)
7. Saric, M., Dujmic, H., Baricevic, D.: Shot Boundary Detection in Soccer Video using Twin-comparison Algorithm and Dominant Color Region. Journal of Information and Organizational Sciences, Vol. 32, Issue 1, 67–73 (2008)

Are Sensory Effects Ready for the World Wide Web?

Christian Timmerer, Markus Waltl, and Hermann Hellwagner

Klagenfurt University, Multimedia Communication Group, Universitätsstrasse 65-67,
9020 Klagenfurt, Austria
{*firstname.lastname*}@itec.uni-klu.ac.at

Abstract. The World Wide Web (WWW) is one of the main entry points to access and consume Internet content in various forms. In particular, the Web browser is used to access different types of media (i.e., text, image, audio, and video) and on some platforms is the only way to access the vast amount of information on the Web. Recently, it has been proposed to stimulate also other senses than vision or audition while consuming multimedia content through so-called sensory effects, with the aim to increase the user's Quality of Experience (QoE). The effects are represented as Sensory Effects Metadata (SEM) which is associated to traditional multimedia content and is rendered (synchronized with the media) on sensory devices like fans, vibration chairs, lamps, etc. In this paper we provide a principal investigation of whether the sensory effects are ready for the WWW and, in anticipation of the result, we propose how to embed sensory effect metadata within Web content and the synchronized rendering thereof.

Keywords: MPEG-V, Sensory Effects, Quality of Experience, World Wide Web.

1 Introduction

In recent publications, it was proposed [1][2] to annotate multimedia content (i.e., videos) with a so-called Sensory Effect Metadata (SEM) description that can be used to stimulate also other senses than vision or audition, e.g., olfaction, mechanoreception, or thermoception which shall lead to an enhanced, unique user experience. That is, the actual video content is rendered on traditional devices (e.g., TV screens) and the sensory effects (i.e., described via SEM and synchronized with the video) are rendered on sensory devices like fans, vibration chairs, lamps, perfumer, etc. The concept of this approach is depicted in Figure 1 with the aim to provide an enhanced, unique, and worthwhile user experience. Note that the SEM is currently being standardized within Part 3 of MPEG-V Media Context and Control [3].

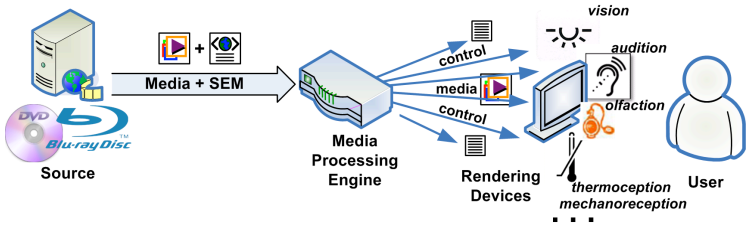


Figure 1. The Concept of Sensory Effects [1].

In [4] we present promising results regarding the usage of sensory effects in conjunction with video content. However, current implementations primarily focus on video content. In this paper we argue that this could be easily extended to Web content, of course including videos also. Therefore, in Section 2 we propose an approach on how to embed SEM descriptions into Web content (i.e., (X)HTML) and the processing thereof as a browser plug-in based on an MPEG Extensible Middleware (MXM) [5] engine capable of parsing the SEM description. To the best of our knowledge, this is the first attempt trying to bring sensory effects to the World Wide Web. The conclusions and future work are outlined in Section 3.

2 Proposed Approach for Integrating Sensory Effects within Web Content

The preferred way to link resources such as a SEM description to a Web document – herein after referred to as document – is the `link` element. An example usage of the `link` element to reference a SEM description is shown in Listing 1.

Listing 1. Example usage of the `link` element to reference a SEM description.

```
<link rel="alternate" type="text/xml" title="SEM description"
      href="http://server.org/sem.xml" media="light, wind, vibration"/>
```

The attributes `type`, `title`, and `href` are used as specified by HTML. The `rel` attribute indicates that the SEM description provides an *alternate representation of the current document* which, in our view, conforms with the concept of sensory effects as depicted above. In particular, if used in conjunction with the `media` attribute, the `ref` attribute indicates that the referenced SEM description is intended for use with the media specified. Please note that the media attribute must support some version of the media query language [6]. Currently, the following media types are defined: 'aural', 'braille', 'handheld', 'print', 'projection', 'screen', 'tty', 'tv'. As none of them match with the concept of sensory effects we propose adding new media types corresponding to the sensory effects as defined in [3], e.g., 'light, wind, vibration' as shown in Listing 1.

Based on the information in the `link` element, the SEM description could be parsed either directly via JavaScript or within a browser plug-in utilizing an MXM engine as proposed in [5]. The SEM MXM Engine – part of the MPEG-V reference

software – is used to access and parse the data within the SEM description and used to control the actual rendering devices (e.g., fans, vibration chairs, lamps). Therefore, the corresponding application programming interfaces (APIs) of the rendering devices must be accessed from within the Web application/browser. In case of the JavaScript solution, this might be an issue but for the browser plug-in the rendering devices could be accessed directly from within the plug-in.

However, the main issue is the synchronization with possible multimedia resources included within the document such as video or audio content. Therefore, it is required to access the current playback position of the actual multimedia resources which is provided thanks to HTML5's `video` and `audio` elements and corresponding DOM attributes [7]. In particular, the `currentTime` DOM attribute provides the current playback position in seconds that can be used for synchronization with the sensory effects. Additionally, the `timeupdate` event indicates that the current playback position changed as part of normal playback which also serves the synchronization purpose. Listing 2 shows an excerpt of how to use the `timeupdate` event and `currentTime` attribute for the synchronization of the video content with the sensory effects.

Listing 2. Usage of `timeupdate` and `currentTime` for synchronization of video content with sensory effects.

```
<script>
...
function time_update() {
  var vid = document.querySelector('video');
  var now = vid.currentTime;
  // note: sem holds sensory effects information previously parsed
  for (var i=0; i<sem.length; i++ {
    if (now >= sem[i].start && now <= sem[i].end) {
       update_sensory_devices(sem[i]);
       break;
} } } // end function time_update()
...
</script>
...
<video src="http://server.org/video.mp4" ontimeupdate="time_update()"
       autobuffer controls>
```

Finally, in [1] we presented means for the automatic extraction of color information from the video frames in order to control additional light sources, e.g., placed behind or next to the TV screen. In particular, we evaluated a choice of algorithms within different color spaces (average color: RGB; dominant color: RGB, HSV, HMMD). However, in all cases we need to access the content of the frame which is not possible within HTML5 as something like `currentFrame` does not exist and is not foreseen. On the other hand, the `canvas` element allows for manipulating the display's content (examples can be found in [8]) and can be used for this purpose. In particular, the `canvas` element is used to extract the pixel information from the display and, consequently, this information is used for extracting the color information. For the actual color extraction and its usage the interested reader is referred to [1].

3 Conclusions and Future Work

In this paper we proposed an approach that brings sensory effects to the World Wide Web. In particular, it provides a conceptual answer to the question raised in the title of this paper. That is, yes, sensory effects are ready for the World Wide Web but this approach requires some implementation work to be done as part of our future work. Furthermore, currently it is not clear how these sensory effects are perceived by the users of the Web content which calls for formal subjective quality assessments. Finally, we would like to study how to apply sensory effects for the whole Web site/content instead of only focusing on audio and video parts.

Acknowledgments. This work was supported in part by the European Commission in the context of the NoE INTERMEDIA (NoE 038419), the P2P-Next project (FP7-ICT-216217), and the ALICANTE project (FP7-ICT-248652).

References

1. Waltl, M., Timmerer, C., Hellwagner, H.: A Test-Bed for Quality of Multimedia Experience Evaluation of Sensory Effects, In: 1st Int'l. Workshop on Quality of Multimedia Experience (QoMEX'09), San Diego, USA (2009)
2. Choi, B. S., Joo, S., Lee, H.: Sensory Effect Metadata for SMMD Media Service, In: 4th Int'l Conf. on Internet and Web Applications and Services (ICIW'09), pp. 649-654, Venice/Mestre, Italy (2009)
3. ISO/IEC 25005-3 FCD: Information technology – Media context and control – Sensory Information, ISO/IEC JTC 1/SC 29/WG 11/N10987, Xi'an, China (2009)
4. Waltl, M., Timmerer, C., Hellwagner, H.: Increasing the User Experience of Multimedia Presentations with Sensory Effects, In: 11th Int'l. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'10), Desenzano del Garda, Italy (2010)
5. Timmerer, C., Chiariglione, F., Preda, M., Rodriguez-Doncel, V.: Accelerating Media Business Developments with the MPEG Extensible Middleware, In: FIA Valencia Book 2010, IOS Press (2010)
6. Lie, H., Çelik, T., Glazman, D., van Kesteren, A. (eds): Media Queries, Candidate Recommendation, W3C (2009), http://www.w3.org/TR/2009/CR-css3-mediaqueries-20090915/
7. Hickson, I., Hyatt, D.: HTML5, Working Draft, W3C (2009), http://www.w3.org/TR/2009/WD-html5-20090825/
8. Manipulating video using canvas, https://developer.mozilla.org/En/Manipulating_video_using_canvas

A Cloud Multimedia Platform

Dejan Kovachev and Ralf Klamma

Information Systems & Database Technologies, RWTH Aachen University
Ahornstr. 55, D-52056, Aachen, Germany
{kovachev|klamma}@dbis.rwth-aachen.de

Abstract. Social networking web applications such as Facebook and Flickr present new challenges for storing and processing user generated content, i.e. multimedia. Handling massive amounts of data requires special systems that need upfront investment, which may hinder the realization of new innovative ideas. Instead, cloud computing as a new emerging operations model promises to deliver elastic on-demand unlimited computing resources as a utility. In this position paper we propose architecture for a Cloud Multimedia Platform that does the heavy-lifting for massive amounts of multimedia storage and processing in the spirit of the cloud computing paradigm.

1 Introduction

Online social communities are an integral and important part of the Web. The Web has become a powerful and ubiquitous delivery channel for mass social interaction and collaboration applications. As a result, an enormous quantities of data are produced/consumed every day [1]. This data comes in forms of content (multimedia), structure (links) and usage (logs). Storing, managing, searching and delivering such data volumes introduce new challenges; thus motivating development of novel data systems.

Cloud computing envisions the notion of delivering software services and customizable hardware configurations to public access, similar how public utilities (electricity, water, etc.) are available to the common man [2]. The concept of cloud computing can include different computer technologies, such as networking infrastructure, Web 2.0, virtualization, SOA and other technologies. The cloud abstracts infrastructure complexities of servers, applications, data, and heterogeneous platforms, enabling users to plug-in at anytime from anywhere and utilize storage and computing services as needed at the moment.

Handling user generated content imposes problems more related to the data volumes, privacy, and delivery latency. Buying upfront the entire infrastructure storage and processing user generated content for a new social networking site can be problematic due to resource scarcity. Furthermore, setting up the computing environment can take additional resources, such as technical personal and time. However, it happens that sites do not go off. As an alternative, the cloud computing model fits naturally in such scenarios [3]. The cloud basically delivers the needed services to the users no matter the load, where charging happens

only by the used resources (CPU, storage, network bandwidth consumption), e.g. at Amazon's Elastic Compute Cloud. In the fast changing Web, delivery of new ideas is often crucial. The developers don't have the time, resources, or expertise to implement scalable infrastructure. Therefore, we believe that providing multimedia processing and storage in the cloud can leverage the dynamics of social networking sites, as there is an economic benefit of doing so.

2 A Cloud Computing Model Need

Consider an example of a social networking site. When the number of users on some social networking site starts to grow, then grow the amounts of user generated data.

An example for a multimedia processing leveraged in the cloud is a production of multiple versions of the same multimedia artifact. These versions could be different image sizes or image quality (thumbnails or mobile phone version). This task normally does not represent a difficult problem. However, in the case of social network site the size really matters (Facebook had 20 billion images as of 2009). That means processing the multimedia in reasonable time can be delivered by using the "unlimited" on-demand computing power of the cloud. Another example is image retrieval from large datasets. Finding the most similar image from a given set of known images requires feature extraction, such as color, texture, histogram, etc. Furthermore, it requires online real time feature comparison between the search item and the dataset.

We have assessed the needs of multimedia data management in cloud computing environment settings. As a result, we have extracted a set of principal requirements for the cloud multimedia platform:

R1: *Scalability.* Handling more data should be done by simply adding more machine instances; thus more power to the system.

R2: *Elasticity.* The infrastructure must be able to elastically adjust assigned resource (scale up as well scale down) according the usage rates.

R3: *Abstraction.* Developers are insulated from the details of provisioning servers, replicating data, recovering from failure, adding servers to support more load, securing data.

R4: *Simplicity.* There are simple interfaces for a development of new services and resource provisioning.

R5: *Interoperability.* The interoperability enables exporting from and importing in the cloud.

R6: *Distributed Data Management.* Large scale data is stored and managed in a distributed manner. This includes also processing and managing user generated content from a social network with millions of users spread around geographically.

R7: *Responsive Services.* The services respond timely and effectively regardless to the amount of data.

R8: *Service Orchestration.* Site administrators can orchestrate multiplex services to compose uniform flow to the user.

If we suppose that our social networking site was built on the cloud computing platform, and the site becomes popular, managing and operating the user generated content is handled with less problems.

3 Architecture for Cloud Multimedia Platform

We build a set of scalable, highly-available multimedia storage and processing services, that can be used across different applications. It is easy to build and rapidly evolve a Web-scale applications on top of the offered horizontal services. Developers focus on the applications logic, and the work around scaling and high availability is done in the cloud layer, instead in the application layer.

The architecture for the Cloud Multimedia Platform is shown on Fig.1. Since storage and computing on massive amounts of data are the key technologies for a cloud computing infrastructure [4], using virtual machines in combination with suitable middleware provide convenient solution for scalable computing systems. Our physical infrastructure is based on *Sun Solaris Containers* virtualization technology[1], which allows the user to allocate a system's various resources, such as memory, CPUs, and devices, into logical groupings and create multiple, discrete systems, each with their own operating system, resources, and identity within a single computer system. The virtualization layer based on the Solaris Containers technology enables provisioning of resources as needed. (Requirements R1, R2 and R3)

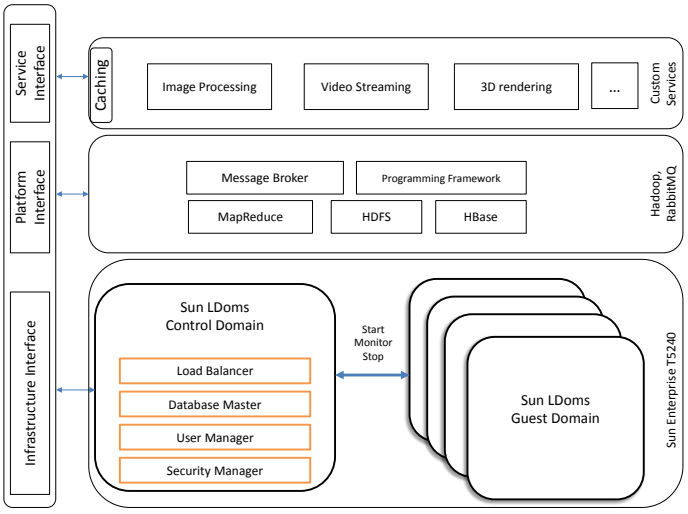


Fig. 1. Cloud computing architecture supporting multimedia services

[1] http://www.sun.com/software/solaris/virtualization.jsp

To ease the development of large scale data processing and storage applications we use the distributed computing framework *Hadoop*[2], which enables a cloud service similar to Platform as a Service (PaaS). The core of Hadoop are MapReduce, Hadoop Distributed File System (HDFS) and HBase. *MapReduce* [5] is one of the most popular programming paradigms for convenient large-scale computing on commodity hardware. *HBase* is a scalable, distributed database that supports structured data storage for large tables. (Requirements R3, R4 and R6)

Since we use many virtual machines running different disparate services, we use a messaging tier that helps tie the services work together. For example, we use *RabbitMQ*[3], which is an open source message broker based on the Advanced Message Queuing Protocol(AMQP).

For example, users produce a lot of multimedia content that is stored in HDFS. When some processing over the content is needed, such as feature extraction, transcoding, or resizing, Hadoop MapReduce can do it in a batch processing. Rankings, tags or other multimedia related metadata is stored within HBase.

4 Future Work

This work is only the first step of many. We plan to investigate which multimedia services are reasonable to be delivered through the cloud computing model. We need to expose simple interfaces to horizontal services, which are used among different vertical applications. The horizontal services in our cloud provide platforms to store, process and effectively deliver data to users. (Requirement R8) The interoperability can be enhanced by using standards for the multimedia metadata. (Requirement R5) We want to evaluate the effectiveness of the cloud computing model; we can easily test within our infrastructure the trade-off between putting more power on a single machine (scale-up) and putting multiple virtual machine instances (scale-out).

References

1. Baeza-Yates, R., Ramakrishnan, R.: Data challenges at Yahoo! In: EDBT '08, New York, NY, USA, ACM (2008) 652–655
2. Expert Group Report: The Future of Cloud Computing. Opportunities for European Cloud Computing Beyond 2010 (2010)
3. Armbrust, M., Fox, A., et al.: Above the Clouds: A Berkeley View of Cloud Computing. Technical report, EECS Department, University of California, Berkeley (February 2009)
4. Peng, B., Cui, B., Li, X.: Implementation Issues of A Cloud Computing Platform. IEEE Data Eng. Bull. **32**(1) (2009) 59–66
5. Dean, J., Ghemawat, S.: Mapreduce: simplified data processing on large clusters. In: OSDI'04, Berkeley, CA, USA, USENIX Association (2004) 10–10

[2] http://hadoop.apache.org
[3] http://www.rabbitmq.com

Towards Detecting Faked Images

Khaled A. N. Rashed and Ralf Klamma

Databases & Information Systems
RWTH Aachen University, Ahornstr. 55, D-52056 Aachen, Germany
rashed,klamma@dbis.rwth-aachen.de

Abstract. An important facet of traditional retrieval models is that they retrieve images and videos and consider their content and context reliable. Nevertheless, this consideration is no longer valid since they can be faked for many reasons and at different degrees thanks to powerful multimedia manipulation software. Our goal is to investigate new ways detecting possible fake in social network platforms. In this paper, we propose an approach that assets identification faked images by combining standard content-based image retrieval (CBIR) techniques and community as alternative solution for supporting such semantic multimedia tasks in an interoperable way using ontologies.

1 Introduction

Nowadays, in the Web 2.0 platforms, there has been an explosive growth of multimedia content. One of the results of the dynamic nature of multimedia content in social network sites is the fake of images. By image fake we mean the changing of an imageś context or its content or both of them for whatever reason. Identification of such fake is considered an important issue for a number of applications such as digital cultural heritage, eTourism, eLearning and multimedia management in general. Identification of faked images in large scale social network sites such as Flickr[1] remains an elusive goal. Over the last years, a number of approaches have tackled the underlying task such as digital watermarking approaches [7] that for a long time have been proposed to provide image authenticity. These approaches rely on embedding additional information within the image content. Their limitation is the requirement of embedding information in images before making them public, which excludes images that are already in public platforms. In addition, watermarking modifies the image content that might create a serious problem where the quality is highly required. Digital forensics tools [6] aim to determine whether digital photos have been manipulated. These tools can measure statistical inconsistencies in the underlying image pixels, detecting traces of resampling, improbable lighting and shadow, physically distortion, and other artifacts. Disadvantage of such tools is that their use in public domains is computationally expensive. Content based approaches such as replica detection [5], near-duplicate detection [3] aim at detecting all

[1] http://flickr.com

images reproduced from the original image. They are based on the image similarity. Their respond speed and efficiency of such detection scheme is largely affected by the size of the original/reference image dataset. The main advantage of such approaches stems from the fact that no additional information should be embedded within the image, as they consider the images itself is the watermark. Content-based image retrieval is the main element of these systems. In addition, it has the scalability property that makes it a suitable alternative to be used to tackle the fake detection problem in social network platforms. The amount of images uploaded to the social network sites to be analyzed by other communities growths exponentially. Therefore, developing new techniques, that can help to make a statement about the searched or retrieved image is strongly needed. Our goal is to combine semantic multimedia metadata, communities and Web 2.0 with content based image retrieval to solve the fake images problem.

CBIR Techniques for Faked Image Detection

The success of utilizing content-based image retrieval (CBIR) techniques [2] in widely application areas motivates us to use them to manage low-level image features. In this regard, images are represented in the feature vector space by means of MPEG-7 visual descriptors [4]. MPEG-7 standard facilitates the quality access to the image content, and support the interoperability issue. Furthermore, MPEG-7 captures the different aspects of color, texture and shapes of the images. They are also robust against the common image transformations. Similarity between two images is measured by calculating the distances between two feature vectors using standard functions.

Web 2.0, Community and Collaborative Activities

Web 2.0 refers to the infrastructures that support humans working together by means of tools and technologies. Community approach in the context of fake detection in such platforms focuses on detecting semantic inconsistencies in images. Rating plays an important role in influencing decision making and peoples choices. Therefore, we will use the collaborative rating to images against the suspected fake. Collaborative rating in Web 2.0 allows classification of experts in dynamic manner. One of most important problems in this regard is the trust of individual users and the quality of ratings. Users rate images and create the associations between the users, rates and the images. Collaborative filtering is used to automatically identify high quality rates for users. Collaborative rating network can be modeled as social network. Such network presents the graph with ternary edges, where each edge represents as the fact that a particular user rated certain image with certain rate. Trust can be calculated among network users based on their dynamic relationships.

2 Objectives and Approach

We intent to develop a system that enables to make sense about the images fake. Such system needs to be able to manage the image content as well as semantic users' interactions. We believe that content-based image retrieval (CBIR) tech-

niques and harnessing collective intelligence[2] is the appropriate solution. In our work, we consider the image as faked if some details are deleted from or inserted to it and photomontage. Another types of image transformation (i.e. rotation, smoothing, luminance change, compression and additive noise) are accounted as versions of the original image. Since a faked image is generated based on an original image, the requirements of the image fakery is that it retains similar visual content to the original image. Therefore, the faked image shares lots of information with the original enough to be distinguished from any other images. Based on this fact, detecting faked images requires robust and efficient visual features to be managed. To fulfill the above requirements and to achieve the interoperability, well established MPEG-7 standards and tools [8] will be used, MPEG-7 is able to express image content covering the most aspects including low level technical information and high level content semantics. Among many useful MPEG-7 visual descriptors, we will consider: *Color Layout*, *Scalable Color*, *Homogeneous Texture*, *Edge Histogram* and *Color Structure* to create the image signature. The low level image features do not provide a comprehensive infor-

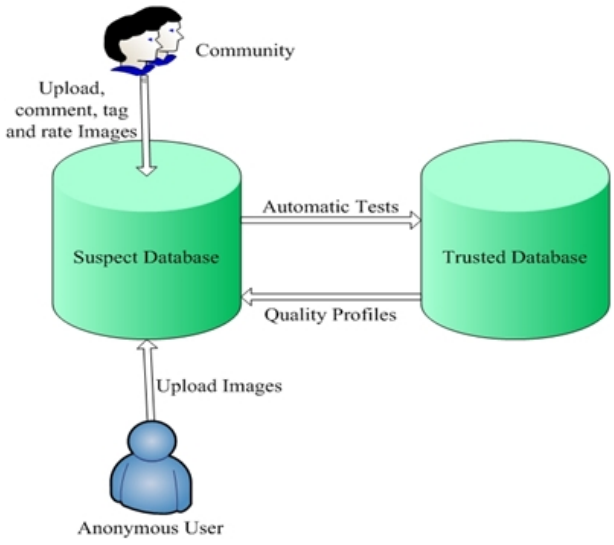


Fig. 1. Overview of the System

mation about the image fake, but they can be support the community decision. On the other hand, we also intent to deploy community and Web 2.0 aspects in our fake detection approach. The key idea is behind the fact that most of discovered faked images were identified by some one who knows the original and then

[2] http://www.oreilly.de/artikel/web20.html

those images have been analyzed against the fake. The overview of the proposed approach is visually described in the fig.(1). Normal users and community can upload images to the system from social network sites such Flickr. These images can be peer rated and commented in term of the fake (degree of fake). For each uploaded image extracted the image signature, and will be stored together with corresponded rates and possible tags in suspected database that should be publicly accessed. Trusted database stores all information about original images, including MPEG-7 visual descriptors, tags, ratings and comments. In case of existing reference trusted image, low level features will be compared, the results will be visualized and then the user judgment will be taken by rate and tag of suspected image. This procedure will be happened reciprocally. Evidence from the community knowledge can be complemented by the evidence from the low level features and vice versa. Information from the trusted database used to improve quality of rating in the case of existing the reference image. When no reference image existed we rely only on the community knowledge. Users (community members) look at image and try to find out the semantic errors that can be exist in the image, annotate and rate the image.

3 Conclusions

In this paper, we presented an overview of our proposed approach to detect faked images. We believe that with combining CBIR with emerging Web 2.0 concepts such as collective intelligence and collaborative rating the fake detection problem in the large scale community image sharing platforms can be solved. Prototype to prove the concepts described above will be implemented. Ontologies can serve as the basis to combine two mention above approaches. In the future work we plan also to investigate on more types of fake issues. Considering image signature tools [1]. We also intend to test some combinations of different feature sets, not only to detect the manipulated images but also to detect the similar fake.

References

1. *"ISO/IEC 15938-3:2002/Amd.3 Image Signature Tools "*, Mar 2009.
2. R. Datta, D. Joshi, J. .li, and J. Z. .Wang. Image retrieval: Ideas, influnces, and trends of new age. *ACM Computing Surveys (CSUR)*, 40:60, 2008.
3. Y. Ke, R. Sukthankar, L. Huston, Y. Ke, and R. Sukthankar. Efficient near-duplicate detection and sub-image retrieval. In *In ACM Multimedia*, pages 869–876, 2004.
4. MPEG-7. Multimedia content description interface part 3: Visual. *ISO/IEC JTC1/SC29/WG11,Doc.N4062*, Mar 2001.
5. S. Nikolopoulos, S. Zafeiriou, and N. Nikolaidis. Image replica detection system utilizing r-trees and linear discriminant analysis. *PR*, 43(3):636–649, March 2010.
6. A. C. Popescu and H. Farid. Exposing digital forgeries by detecting traces of resampling. *IEEE Transaction on signal processing 53*, 53(758-767), 2005.
7. C. Rey and J.-L. Dugelay. A survey of watermarking algorithms for image authentication. *EURASIP J. Appl. Signal Process.*, 2002(1):613–621, 2002.
8. G. Shih-Fu Chang, A. B. Chan, and P. J. Moreno. Overview of the mpeg-7 standard. *IEEE Trans. Circuits and Systems for Video Technology.*, 11(0):688 – 695, 2001.

Content Security and Privacy Preservation in Social Networks through Text Mining

José María Gómez Hidalgo[1], José Miguel Martín Abreu[1],
Pablo García Bringas[2], Igor Santos Grueiro[2]

[1] Optenet, José Echegaray 8, 28230 Las Rozas, Madrid, Spain
{jgomez, jabreu}@optenet.com
[2] S3Lab, Deustotech, Universidad de Deusto,
Avda. Universidades 24, 48007, Bilbao, Bizkaia, SPAIN
{pablo.garcia.bringas, isantos}@deusto.es

Abstract. Due to their huge popularity, Social Networks are increasingly being used as malware, spam and phishing propagation applications. Moreover, Social Networks are being widely recognized as a source of private (either corporate or personal) information leaks. Within the project Segur@, Optenet has developed a number of prototypes that deal with these problems, based on several techniques that share text mining as the underlying approach. These prototypes include a malware detection system based on Information Retrieval techniques, a compression-based spam filter, and a Data Leak Prevention system that makes use of Named Entity Recognition techniques.

Keywords: Text Mining, Information Retrieval, Named Entity Recognition, Malware, Spam, Data Leak Prevention

1 Introduction

As Social Networks (SNs) gain more and more popularity, they are becoming a major source and propagation vector for malware, spam, phishing and private information leaks. It is imperative to control these threats in order to make SNs as useful as they promise to be. The project Segur@ has helped Optenet to frame a number of research developments addressing these problems, and sharing Text Mining as the base of them. These developments, along with the prototypes developed, are described in the next sections.

2 Malware Detection using Information Retrieval with Opcodes

Malware (trojans, spyware, viruses, etc.) is plaguing SNs, mostly in the form of obfuscated variants of an original malicious software that adapts to the current user machine specifics. In order to detect these new variants and even new malware families (i.e., completely new malicious software), already known software (both

malicious and legitimate) can be represented in terms of operational codes (assembler code instructions), and new software similarity to these known executables can be computed. We have designed an approach in which:

1. Software is represented as sequences of operational codes (opcodes) of several lengths, after selecting the most relevant ones according to its Mutual Information with respect to being malware or legitimate software. Opcode n-grams are terms in a Vector Space Model, weighted using a weighted term frequency formula.
2. New software instances similarity with respect to malware or legitimate software is computed using the cosine similarity between the vectors that represent all software instances. A new instance is considered malware or legitimate depending of its similarity to both kinds of instances. This way, variants of currently known malware are detected.
3. A Machine Learning approach is used to derive a classifier able to detect new malware instances that share properties with existing code samples. Several Machine Learning approaches have been tested, including Support Vector Machines (SVMs), Decision Trees, or k Nearest Neighbour classifiers.

Overall, a Text Classification approach is used, but on assembler code instead of text. The experiments carried out for these approaches are very encouraging, as the results obtained on a test collection of over 26,000 samples of malware and legitimate software show that:

1. It is quite easy to find a discriminative threshold between malware and legitimate software in the detection of variants of known malware instances when using opcode bigrams, leading to high detection rates and no false positives.
2. For instance, SVMs and Decision Trees (Random Forests) over one and bigrams lead to detection rates over 95% with false positives below 5% by cross-validation on the test collection.

A prototype has been developed in order to demonstrate the analysis process on a Microsoft Windows Systems. The prototype allows loading a software instance, disassembling it, and comparing it to the current database of software samples.

3 Compression-based Spam Detection

SNs are being used as a prominent method for disseminating spam and phishing attacks, what represents a major threat for SNs users. A wide range of learning based systems and approaches for spam detection (Bayesian filters) have been developed in recent years. Despite of their theoretical effectiveness, most of them are very sensitive to the tokenization process, in which messages are represented as sequences of character strings intended to capture either message semantics or spam properties.

State of the art on Bayesian filters as demonstrated in the TREC Spam Filtering Tracks [1] shows that compression based learning, often applied to Text Categorization (see e.g. [3]), does not face the same problem. The reason for this is that messages are not explicitly represented in terms of strings resembling works, but in terms of bit sequences hardly hacked by spammers in order to overcome the filter.

We have developed a compression based spam filter that includes a number of text compression approaches [1], namely Dynamic Markov Chains, Prediction by Partial Matching, and Lempel-Ziv variants as GZip. The compression filter is able to classify email messages (as those received through SNs with updates and status summaries) as spam or legitimate, and it has been implemented resembling Spamassassin command line operations, in order to simplify its integration with Mail User Agents like Thunderbird or KMail over Linux.

4 Privacy Preservation through Named Entity Recognition

Perhaps the most dangerous current security threat for SN users is private information leaks. From phishing attacks to unnoticed users mistakes, leaks are getting more and more common, and Data Leak Prevention (DLP) tools must be made more and more effective in order to preserve user and corporate privacy.

Most DLP tools are very effective when protecting already known private information. However, a great amount of private information is not recognized until it has been disclosed to unknown users or competition enterprises. In order to detect unknown private information leaks, we have designed a system based on:

1. Text Classification techniques that are used to process any text disclosed to a SN in search for Named Entities like person, organization or product names. Language Independent techniques (e.g., n-gram features) are used to detect unknown patterns (entities), using the package Freeling [2] for testing our approach on Spanish and English text instances.
2. A user-learning loop similar to that of personal firewalls. The user is alerted when a new entity is detected, allowing them to classify as private or public, and to block or allow text including the detected entity to be posted to the net. Subsequent occurrences of the entity are blocked or allowed by using a white/black list. In consequence, the software gradually adapts its performance to the user/organization behaviour.

The Named Entity Recognition module of our prototype has been tested on a sample of entities and text postings collected over a year on the micro-blogging Twitter SN, covering a range of popular brand and product searches according to Google Insights statistics. Results of our experiments show accuracy over 90% in English and Spanish, and very low false negative rates. Moreover, most false negative text instances (which correspond to undetected entities which may lead to private information leaks) are detected in other instances, and in consequence, the user is alerted regarding them anyway.

The prototype configuration screen is presented in Fig. 1, showing the whole range of privacy enabling data fields (from credit card numbers to passwords, etc.). Named Entities detected are stored as user-defined fields (*Other Keys* section). The prototype is available for Microsoft Windows.

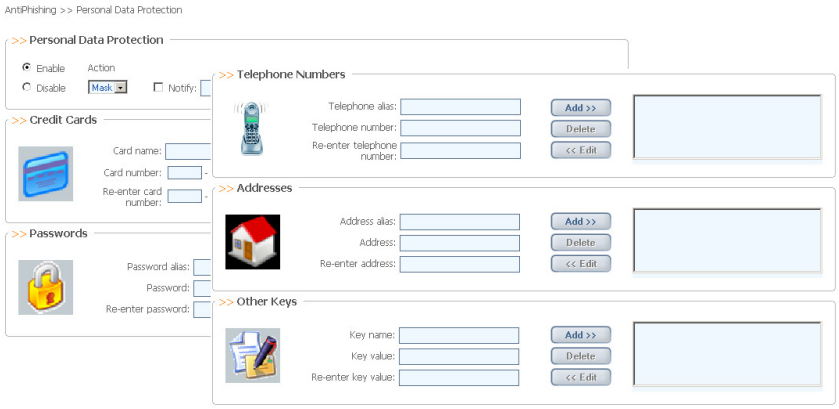


Fig. 1. Data fields used in privacy protection features. Named Entities are stored in the Other Keys section, as the users are getting alerted and they decide to supervise them.

5 Future Work

Current prototypes future work plans include improving both the text analysis techniques involved, and getting the user interface getting integrated into the Optenet Security Suite for PC, enabling current and future Optenet users to benefit from these advances.

Acknowledgments. The prototypes described in this paper have been developed within the Project Segur@ (CENIT-2007 2004).

References

1. Cormack, G.: TREC 2007 Spam Track Overview. In Voorhees E.M., Buckland, L.P. (eds): NIST Special Publication 500-274: The Sixteenth Text REtrieval Conference Proceedings (TREC 2007).
2. Freeling Home Page, http://www.lsi.upc.edu/~nlp/freeling/.
3. Teahan, W.J., Harper, D. J.: Using compression based language models for text categorization. In Callan, J., Croft, B., Lafferty, J., (eds): Workshop on Language Modelling and Information Retrieval, pages 83--88, Carnegie Mellon University, 2001.

Security 2.0: Trusted Identity Service

Vanessa Alvarez[1], Antonio Amaya[1], Miguel Ochoa[1], David Prieto[1]

[1] Telefonica I+D
{vac, amac, ochoa, dprieto}@tid.es

Abstract. Social network interactions need a recompilation of information that can be used to evaluate the trust that can be deposited on each peer. Existing solutions centered on service providers have both quantitative and qualitative limitations. We introduce a new approach based on a Trusted Identity Service located on the access provider that will provide service providers with information about a global and unchangeable trust level of the users.

1 Introduction

Internet services boom imply an increase of private and confidential information deposited by individuals and companies on the service providers. There's also a constant increase of the economic value of the online transactions.

At the same time the economic value and type distribution of cybercrime is increasing. New cybercrime types include identity theft and stealing of personal data to be used on Internet frauds –like stealing of money from bank accounts and buying with stolen credit cards, or distributing unwanted mail or publicity to the contact list of the compromised person- or real world crime –get personal information from person to blackmail them, steal on his home..- Because of all this it's necessary having mechanisms that allow the users to evaluate the trust level they can deposit on a peer on any online transaction.

While security and trust on the real world is based and relies usually on physical presence of the peers involved on any transaction on the same location – that way documents, physical aspect or signatures may be verified in situ, and thus the trust level of the peers may be established- on Internet there's no such possibility, since peers communicate remotely from locations on any part of the world, and they interact using computing devices and communication media that can be controlled by third parties.

Thus, during electronic transactions there's a recompilation of information that can be used to evaluate the trust that can be deposited on each peer, protecting the information that's being used, avoiding an illicit use of the information, on that same or a later moment.

On one hand, users accessing a service provider have different methods to evaluate the trust level of the service provider using server certificates, web filtering services, …

On the other hand, service providers also need to evaluate during electronic transactions the trust level they can deposit on any user – is he the one he says he is?

Is he using a secure device and communication channel, on which the confidential information is secure against a later illicit use? To this extent, there are also several solutions that get useful information to evaluate the trust level using user identification mechanisms based on several authentication factors, solutions like firewalls and antimalware that try to avoid information theft on the customer's device, solutions that provide a service provider some other information that can be used to evaluate the trust level like usual or close to a recent usual IP address, browser known security exploits, …

Depending on the risk assessment, a service provider can make several decisions: it can reject the user or ask him for a different level of authentication. Our solution is centered on the evaluation, by the service provider, of the trust that he can assign to the user, behaving in that way like a "trust authority".

1.1 Problems with existing solutions

Social network sites have to evaluate the trust level of the user trying to access its services and this trust evaluation depends on the quantity and quality of available data.

As a part of that trust determination process, a service provider has to answer the questions: is the user the one he says he is? Is he accessing our servers from a secure environment (location, device) on which confidentiality and security of transmitted data will be assured? To that extent they use information they have available, related with electronic communications like security of the equipment/device the user is using to access the service, user location, user's behavioral analysis or user's authentication.

But usage of electronic communications information by part of the service providers with current techniques has a quantitative limitation due to the problem of using partial information, and a qualitative limitation because this information can be manipulated by malicious attackers. Current systems also require that the user uses specific mechanisms for each service provider (authentication mechanisms, antimalware software). This might hinder user's enjoyment of the services if he usually accesses several service providers.

2 Trusted Identity Service

The solution proposed is based on a trust generation system located on the access provider that will provide service providers with information about the trust level of the users. This information will be more complete and of more quality than the information the service providers are using actually. This way the access provider will become a 'trust authority'.

This is accomplished by gathering information directly on the Internet access provider that the users are using to contact service providers. A device located on the access provider internal network gathers information about: user's identity, user's traffic that will be used to analyze his behavior, security status of the user's device,

and geographical location. All this information is analyzed and summarized on a 'trust ticket' that will be sent to service providers.

The defined system will provide service providers with the trust level they can assign to a given user more complete and with more quality than the one they're currently using, since the trust level will be based on information gathered directly on the network access provider that the user's are using to connect to the service providers. This information is more difficult to manipulate by malicious third parties than the information service providers are currently using; thus, IP address used as part of the analysis is assigned by the network access provider and cannot be manipulated as it could be if the service provider were getting the IP from the user's device.

2.1 Security as a Service Capability

Telcos have chosen a Service Oriented Architecture (SOA) as a base of their Service Delivery Platform (SDP) for applications, services and contents distribution. SDP allows to Telco's enterprises to increase incomes per user with new added-value services, reducing development, deployment and operational costs. SOA technology provides the integration of different systems on a consistent environment where the existing assets can be re-use to create a new value, where time to market is reduced from months and years to days and weeks.

So security functionalities that are deployed as network capacities can be used alone or in an aggregate form (multifactor security) by third-party applications, Telco's applications or by other elements within the SDP architecture, to enrich their functionalities, increase user loyalty or increase the security level which is developed.

2.2 Privacy by Design

The Trusted Identity Service developed as a RESTful web service, even though shown in an aggregate form, discloses user personal information. So, we use OAuth protocol for intercepting queries that come from online services that act as consumers, conducted to this resource and checking that they hold a valid authorization ticket issued by the personal information owner.

Moreover, OAuth [1] extension has been developed to introduce a privacy policy associated to the lifetime and uses of authorization tickets.

3 Conclusion

Global social network services require global and trusted identity services. Access providers have a global and unchangeable knowledge of users and services to become a 'trust authority'. This Trusted Identity Service can be exposed as a SDP Service Capacity to service providers but from the beginning conforming to privacy principles of visibility, transparency and respects for user privacy.

Acknowledgments. This work has been developed by Telefonica I+D and funded by the Centre for the Development of Industrial Technology (CDTI) of Spain, within the framework of the Segur@ project [2], reference CENIT-2007 2004 of the CENIT Program (part of the INGENIO 2010 initiative).

References

1. OAuth, http://oauth.net
2. Proyecto SEGUR@ - Seguridad y Confianza en la Sociedad de la Información", http://www.cenitsegura.com

Distributed Access Control Management in Federated Identity Systems

Manuel Gil Pérez[1], Gabriel López[1], Antonio F. Gómez Skarmeta[1], Alberto Sicre Vara-De-Rey[2] and Aljosa Pasic[2]

[1] Departamento de Ingeniería de la Información y las Comunicaciones
University of Murcia, Spain
Email: {mgilperez,gabilm,skarmeta}@um.es
[2] ATOS Origin, Albarracin 25, 28037 Madrid
Email: {alberto.sicre,aljosa.pasic}@atosresearch.eu

Abstract. Identity federation provides a powerful way for managing sensible information of its users. However, as the number of members increases, the management of the policies defined by the federation is becoming more and more complex. In this paper, we present a mechanism to manage this complexity by means of the *administrative delegation*. This allows administrators of an institution to delegate part of these policies to other users, who also will have more knowledge within the scope where these policies will be enforced. This proposal also introduces a way for shortening the added complexity that supposes the introduction of this new sort of policies to users without deep knowledge in the policy management area.

Key words: administrative delegation, delegation policies, access control, distributed authorization

1 Introduction and Motivation

As the number of members of an institution increases, new institutions join the same federation, or the policies defined in a federation change due to the highly dynamic nature of this sort of systems, the management of their policies becomes more and more complex. This is mainly due to the great amount of policies that should be managed, such as access control policies, privacy policies, or validation policies based on LoA (*Level of Assurance*), among others.

In order to reduce this complexity, the system administrator of an institution can delegate to third parties, called *delegates*, the management of a subset of the system policies. Thus, we are not only distributing the management of those policies to other people, but also they are being delegated to those who have more knowledge in the application area where they will be used. This process, where the administrator transfers the management of a subset of policies to a delegated person, is commonly known as *administrative delegation* [1, 2].

The introduction of this new sort of policies supposes a new value-added service for the current identity systems, but also presents some drawbacks that have to be treated suitably:

- The number of policies to manage increases, so that the system administrator will have to manage both the policies that already existed previously (access control policies to services and resources, privacy policies, etc.) and this new sort of policies.
- The delegates are usually users with no knowledge in the policy management, access control languages such as XACML, etc. Therefore, we will have to make easier to these people the generation and management of this new sort of policies, and make it as easy and intuitive as possible.

As we can see, even though the workload of the administrator is reduced, and considerably distributed among several delegates, the policy management (including the administrative ones) will also be more complex. Therefore, it is not enough to define policies for the administrative delegation, but also it will be necessary to define an infrastructure that can manage these policies, thereby helping, mainly to the delegates, to carry out these tasks.

2 Application Scenario

As application example of the administrative delegation in real environments, let us suppose a scenario where the administrator of an institution delegates to all heads of department the policies about granting access to those people under their supervision. These heads of department will have more knowledge about their own employees than the system administrator. In this scenario, the system administrator will be able to delegate in each head of department the definition of which employees will have access to the network, as well as the connection schedule according to their workday; that information is perfectly known by each head of department.

In this example scenario we can see the use of the administrative delegation, where we are avoiding that the system administrator has to create access control policies on a set of people that he (probably) does not know. In this case, each head of department, once the system administrator authorizes him as a delegate, will be the person in charge of controlling the access to the network of his employees by creating the needed access policies with the adequate information.

3 Delegation Policy Management

As a solution to the problems previously commented, we have included in this work a set of new components to the Segur@-DAMe identity federation [3] for managing the complete life cycle of these new policies. We have also defined a mechanism by which we allow generating a set of templates (Web forms) to help delegates to perform these administrative tasks in a simple and intuitive way. These templates are automatically generated by the infrastructure from the administrative delegation policies created by the system administrator.

To this end, our infrastructure makes use of XACML 3.0 (*eXtensible Access Control Markup Language*) [4], which includes in its specification new advanced

features for the definition of delegation policies. These changes have been made to allow any person of an institution, who owns a certain privilege, to delegate it to another person.

In this sense, XACML 3.0 defines a new element, called *PolicyIssuer*, to indicate who has issued a given policy. With this element, the system can identify and verify whether the corresponding issuer is valid to delegate the enclosed privilege just before the policy is used. A policy with no issuer element is considered as trusted and, therefore, will be managed by the PDP (*Policy Decision Point*) as a traditional policy.

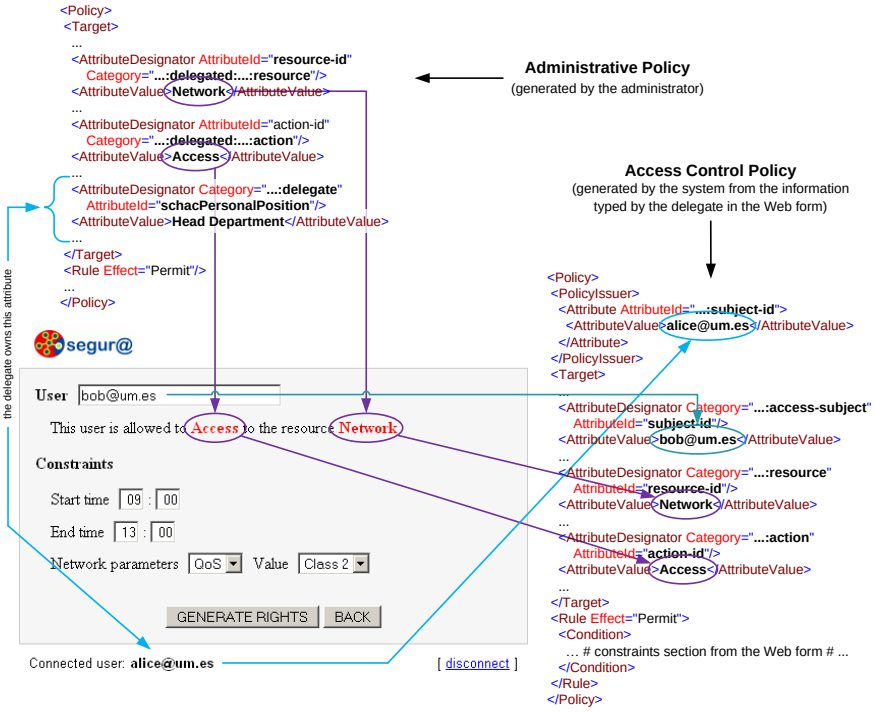


Fig. 1. Automatic generation of templates and access control policies

As we have commented before, the main goal of our infrastructure is to provide delegates with an easy and intuitive way for creating the access control policies for which they are responsible. For this purpose, we make use of a PMT (*Policy Management Tool*) which is capable of parsing administrative policies and extracting the information necessary to generate the corresponding templates in an automatic fashion. The PMT will then make use of XSL transformations [5] to carry out this task, taking as input the administrative policy

generated by the administrator. The rest of information that cannot be directly found in that policy is requested to the delegate as input fields in the template.

Fig. 1 depicts this process, in a schematic way, for: 1) generating the templates (Web forms) from the administrative policy created by the system administrator (top of the image); and 2) generating the final access control policies from some pieces of the administrative policy and the information provided by the delegate in the Web form (right-hand size of the image). Both processes are automatically carried out by making use of the XSL transformations mentioned above.

4 Conclusion

As we have seen throughout this paper, we have presented a new way for managing the administrative delegation in which system administrators or any institution, belong to the same identity federation, can delegate part of their work to third parties. With the help of this system, the policy management is being distributed to those people who have more knowledge than the system administrator on the scope where those policies will be applied later.

To this end, our infrastructure is capable of automatically generating a set of templates (or Web forms) from the administrative policies created by the administrators. These delegates, who have (probably) no idea about how to create policies, will be able to fill in these templates in an easy and intuitive way.

Acknowledgment

This work has been funded by the CENIT Segur@ (Seguridad y Confianza en la Sociedad de la Información) project. Authors would also like to thank the Funding Program for Research Groups of Excellence with code 04552/GERM/06 granted by the Fundación Séneca.

References

1. E. Rissanen and B.S. Firozabadi. "Administrative Delegation in XACML - Position Paper". Swedish Institute of Computer Science, September 2004.
2. K. Gaaloul and F. Charoy. "Task Delegation Based Access Control Models for Workflow Systems". In *I3E '09: Proceedings of the 9th IFIP Conference on e-Business, e-Services and e-Society*, pages 400–414, 2009.
3. The CENIT Segur@ Project. "Seguridad y Confianza en la Sociedad de la Información". http://www.cenitsegura.com.
4. E. Rissanen (editor). "XACML v3.0 Administration and Delegation Profile Version 1.0". Committee Draft 01, April 2009.
5. M. Kay (editor). "XSL Transformations (XSLT) Version 2.0". W3C Recommendation, January 2007.

A trustworthy and privacy-enhancing registration process for social network services

Ana Ballester[1], David Campos[1], Francisco Jordan[2], Jorge Palacio[1], Helena Pujol[2]

[1] Distributed Multimedia Applications Group, Universitat Politècnica de Catalunya
[2] Safelayer Secure Communications, S.A.
{anab, dcampos, jpalacio}@ac.upc.edu
{jordan, helena.pujol}@safelayer.com

Abstract. The increasing popularity of social network services (SNS), being used by millions of people everyday, has arisen some security and privacy issues. While users must be free to choose which amount of personal information they provide, SNS providers should get more involved to guarantee that their services are correctly exploited. This leads us to propose a more secure but still privacy-enhancing solution for registration processes, that involves the use of digital certificates —which are a reliable source of personal information—, as well as information cards —which hold privacy properties. Safelayer applications PKI Trust Center and Interidy Identity Provider implement this proposal.

Keywords: digital certificate, information card, privacy, social network service, tagging, trust

1 Introduction

Online Social Network Services (SNS) have revolutionized the way we communicate and cooperate with others. Most users seem to be willing to publish their personal data, agenda and even thoughts, but they are seldom aware of the potential risks they may be exposed to, when they make this kind of information available indiscriminately or when they get in touch with unfamiliar people. Obviously, this is critical when children and young people are involved.

Concerned about this sociological challenge, the European Commission has actively contributed to develop the Safer Social Networking Principles for the EU [1], together with 20 SNS providers and a number of NGOs, which propose good practice recommendations to enhance the safety of children and young people using these services.

Although access control and transparency, as well as education, are essential to improve security in the Information Society, it is also important to improve the trustworthiness of registration processes. For example, most SNS are targeted at users 14 and over, but there are also other SNS that only accept users aged 13 to 17, like Teen Second Life. Besides, there are SNS that are targeted at a professional audience, whereas others belong to a single organization. Therefore, some mechanisms should

be established to ensure that registered users hold some particular attributes, while they still should have the chance of registering with privacy, for instance, using a pseudonym.

1.1 Providing trustworthy attributes

SNS providers usually delegate the trustworthiness of the information collected during the registration process to the user. This is achieved by including an explicit clause in the terms of service, which the user must declare to have read and accept in order to register. However, SNS providers, authorities and even users should not rely on such a weak trust mechanism when their very privacy and security may be exposed.

Digital certificates are a trustworthy source of personal data. They involve a Certification Authority, which is a trusted third party which guarantees that the personal attributes stated by the certificate are true and correspond to the certificate's holder. The legal document that describes how a Certificate Authority manages the life-cycle of the certificates it issues is known as Certification Practice Statement, and follows the RFC 3647 [2]. Consequently, not all digital certificates are equally reliable, as each Certification Authority is ruled by its own procedures. In particular, many Certification Authorities require the physical presence of the future certificate holder at one point of the certificate issuance process, while others don't. This kind of requirements, as well as the cryptographic characteristics of the associated keys, determines the level of trust that can be put in a Certification Authority.

1.2 Ensuring privacy

Digital certificates represent a huge source of reliable data. In the particular case of Spain, more than 14 million citizen e-ID cards [3] have already been issued. However, digital certificates identify users unequivocally, and therefore they are not suitable for applications that require a higher level of privacy during the user registration process, like SNS. So, in order to get benefit from this already deployed infrastructure, we could think of a mechanism that complemented it by providing more privacy.

This may be achieved with information cards [4], which can be seen as an electronic version of conventional cards in our wallets. Information cards involve the user in the personal data disclosure process, allowing him to decide which particular personal attributes (name, age, nationality, etc.) he wants to reveal to the service provider, in this case, the SNS.

2 Signing up with trustworthiness as well as privacy

By combining the trust infrastructure of digital certificates and the privacy-enhancing properties of information cards, a more reliable and confidential authentication mechanism can be implemented.

Safelayer's PKI Trust Center compiles and displays Certification Practice Statements information in a user-friendly way, and compliments it with contributions from Certification Authorities administrators and users, in a Web 2.0 style. The technology underneath is based on ontologies and Semantic Web standards like OWL and RDF, which make information machine readable and processable. By the use of data mining techniques a trust rating is automatically inferred for each Certification Authority.

In addition, Safelayer's Interidy Identity Provider is set to issue information cards with personal attributes that are imported from digital certificates. The criterion that determines if the imported personal data is trustworthy enough is the trust rating provided by PKI Trust Center.

So, a user should fulfill two steps in order to register in a social network application with privacy. First, the user should generate an information card from one of his personal certificates using Interidy Identity Provider. This card would contain a set of verified claims. Afterwards, the user could complete the registration process by presenting this information card to the social network registration service

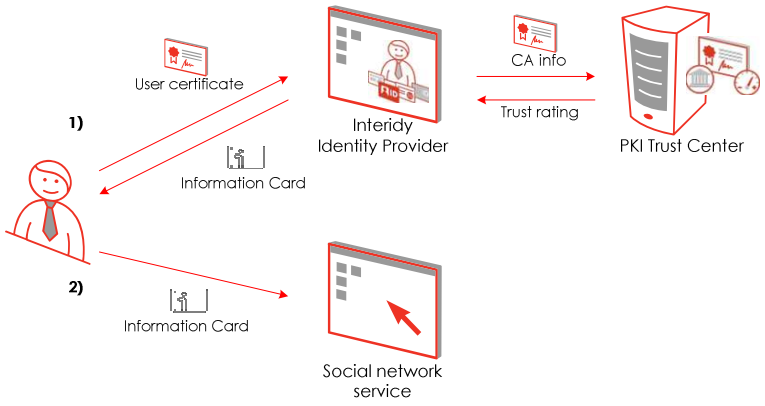


Fig. 1. First, the user generates an information card from a digital certificate, and afterward, the user completes the registration process at the social network application.

With this solution, as the user is able to reveal only a minimum subset of required personal attributes (for example, being over 14), his privacy is preserved and he is entitled to use pseudonyms all through the registration process.

3 Conclusion

Two main requirements should be ensured during the registration process of a user in a social network online service. On the one hand, a user should be able to register with enough privacy guarantees, even with a pseudonym in case he didn't want to

disclose his real name. On the other hand, the service provider should be able to validate that its users fulfill certain criteria, being one of the most critical, the fact that they all belong to a particular age range. Therefore, a third party is needed to act as a Certification Authority that will bind authentic identity attributes to end entities, but a privacy-respecting authentication mechanism is desirable so that users are required to disclose the minimum set of identity attributes.

This solution may be achieved by combining the functionalities of Interidy Identity Provider and PKI Trust Center applications, which are available at Safelayer Sandbox website [5].

Acknowledgments. This research has been supported by Safelayer Secure Communications and the Centre for the Development of Industrial Technology (CDTI) of Spain, within the framework of the Segur@ project, reference CENIT-2007 2004 of the CENIT Program (part of the INGENIO 2010 initiative) [6].

References

1. "The Safer Social Networking Principles for the EU", Europe's Information Society Thematic Portal, Safer Social Networking Activity, 2009, http://ec.europa.eu/information_society/activities/social_networking/docs/sn_principles.pdf
2. S. Chokhani, W. Ford, R. Sabett, C. Merrill, S. Wu, "RFC 3647 - Internet X.509 Public Key Infrastructure Certificate Policy and Certification Practices Framework", Internet Engineering Task Force, 2003
3. Portal oficial sobre el DNI electrónico, http://www.dnielectronico.es/
4. Information Cards, http://informationcard.net/
5. "Proyecto SEGUR@ - Seguridad y Confianza en la Sociedad de la Información", http://www.cenitsegura.com
6. Safelayer Sandbox, http://sandbox.safelayer.com

VAnalyzer: a MPEG-7 based Semantic Video Annotation Tool

Florian Stegmaier, Mario Döller, David Coquil, Vanessa El-Khoury
and Harald Kosch

Chair of Distributed Information Systems, University of Passau, Germany
{forename.surname}@uni-passau.de

1 Introduction

Much recent research deals with the automatic extraction of multimedia metadata. While automatic extraction is already well understood for low-level features (e.g., color), it remains an open issue for high-level (semantic) features [1].

In this context, an evaluation of currently available MPEG-7 annotation tools and frameworks has been performed [2]. Most tools can extract low-level features automatically, and some also provide functionalities to describe high-level features manually. This paper highlights VAnalyzer[1], an MPEG-7 based semantic video annotation tool, which is also able to extract certain semantic annotation automatically. Here, we only introduce the features of the VAnalyzer without discussing its architecture. Finally, the demonstration procedure will be highlighted.

2 Main features of VAnalyzer

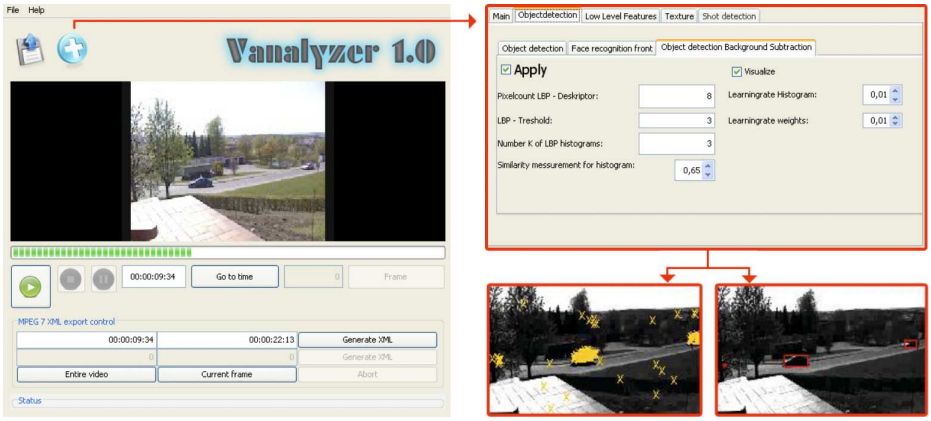
VAnalyzer is implemented in Java and based on the NetBeans Platform[2], a generic framework for Swing applications. The application offers regular video player functions (e.g., stop or play) and produces valid MPEG-7 metadata descriptions. The raw (decoded) video frames are accessed through the Java Media Framework (JMF)[3] and its plug-in FOBS4JMF[4]. This FFmpeg wrapper plug-in adds a broad support of media formats in addition to the few that JMF can natively process.

Figure 1 outlines VAnalyzer's interface: (a) shows the main menu of the application; (b) illustrates the user interface to (de-) activate or configure the metadata extraction algorithms, completed by a visualisation of the algorithm processing.

The actual content of the extracted metadata depend on the selected algorithms. The following algorithms have been integrated in the VAnalyzer:

– *MPEG-7 low-level features extraction*[5] processes for Color Layout, Color Structure, Dominant Color and Edge Histogram.

[1] `http://www.dimis.fim.uni-passau.de/iris/index.php?view=vanalyzer`
[2] `http://platform.netbeans.org`
[3] `http://java.sun.com/javase/technologies/desktop/media/jmf`
[4] `http://fobs.sourceforge.net`
[5] Adapted from the Caliph & Emir project: `http://www.semanticmetadata.net`



(a) Main menu & player functionalities (b) Algorithm configuration & visualisation

Fig. 1. User interfaces and process visualisation of the VAnalyzer

– Two approaches for *object recognition and tracking*. The first approach makes use of the OpenCV[6] project for object (e.g., faces) detection using Haar cascades. The second follows a modified approach of Heikkilä et al. [3], which detects objects using background substraction and background models.
– A *shot detection* algorithm based on a modified approach described in [4]. Our improvements concentrate on an enhanced dissolve and fade detection through a new one sided feature approach and efficient block matching technique.

3 Demonstration procedure

The demonstration will contain an (optional) talk regarding the architectural facets of the application, a computer-assisted presentation of the central features and the validation of the extracted metadata. Thereby, the used data set will consist of a video surveillance footage.

References

1. J. S. Hare, P. A. S. Sinclair, P. H. Lewis, K. Martinez, P. G. B. Enser, and C. J. Sandom, "Bridging the Semantic Gap in Multimedia Information Retrieval: Top-down and Bottom-up approaches," in *Proceedings of the 1st International Workshop on Semantic Web Annotations for Multimedia*, (Edinburgh, Scotland), 2006.
2. M. Döller and N. Lefin, "Evaluation of available MPEG-7 Annotation Tools," in *Proceedings of I-MEDIA/I-SEMANTICS '07, 1st International Conference on New Media Technology*, (Graz, Austria), September 2007.
3. M. Heikkilä and M. Pietikäinen, "A Texture-Based Method for Modeling the Background and Detecting Moving Objects," *IEEE TPAMI*, vol. 28, no. 4, pp. 657–662, 2006.
4. H. Fang, J. Jiang, and Y. Feng, "A Fuzzy Logic Approach for Detection of Video Shot Boundaries," *Pattern Recogn.*, vol. 39, no. 11, pp. 2092–2100, 2006.

[6] `http://opencv.willowgarage.com`

Callisto: Tag Recommendations by Image Content

Mathias Lux[1], Arthur Pitman[1], Oge Marques[2]

[1] Klagenfurt University, Universitätsstr. 65-67
9020 Klagenfurt, Austria
mlux@itec.uni-klu.ac.at, arthur.pitman@uni-klu.ac.at

[1] Florida Atlantic University, 777 Glades Rd.
Boca Raton, FL 33431 USA
omarques@fau.edu

Abstract. While assigning uncontrolled keywords to photos – a process called tagging – solves a lot of problems with retrieval of visual information still many photos published on the web go untagged or are tagged with non-descriptive or too few tags. In this extended abstract we demonstrate our tag recommendation prototype *Callisto* incorporating statistical tag co-ocurrence as well as image contents. We also outline qualitative experiments that have shown the difference in results between solely statistical and content based approaches.

Keywords: Tagging, recommendation, content based image retrieval

1 Introduction

User generated content is on the rise millions of people upload and publish digital photos on a daily basis. Flickr alone claims thousands of uploaded photos per minute with reports stating numbers from 3,000-6,000 uploads depending on the time of day. Many of these photos are uploaded as-is, with only a minimum amount of metadata attached. This typically includes the EXIF metadata and the name of the image automatically assigned by the camera. A considerable share of photos is tagged, which means the photo is annotated by a set of keywords, but by far not every single photo was tagged by the uploader. As retrieval of the photos heavily depends on the annotations the amount and quality of tags is critical to every retrieval scenario. Therefore we identify a need for applications supporting the users in the tagging process. Classic approaches are limited to co-occurrence analysis of tags and therefore typically suggest tags that are most frequent and not very distinctive such as beautiful, topphoto or flickrdiamond. However, our assumption is that in many scenarios tags describing the actual content of the images, like flower, sunset or wood, are needed. In this paper we present a software prototype, which is able to suggest tags based on (i) one or more initial tags and (ii) image content. This novel combination allows for a more content-related suggestion of tags and might help users to find more and more descriptive tags to annotate their uploaded images.

2 Callisto

Our tag recommendation prototype, called *Callisto*, allows for input of a photo as well as one or more start tags. As depicted in Fig. 1 Callisto downloads images based on the start tag and ranks the set of retrieved images based on image content. Photos with highest ranks are taken into account for tag analysis, which then leads to the recommendation of a number of tags (typically 5-10).

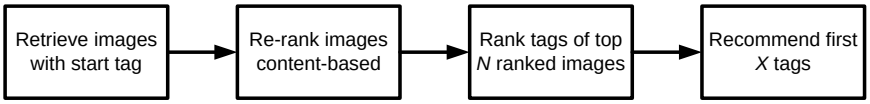


Fig. 1. Main four steps in the recommendation process. Images are retrieved based on the provides start tag, then re-ranked based on image content. From this list the top *N* images are considered for tag analysis, which leads in step 4 to the recommendation of *X* tags.

The screenshots in Fig. 2 show Callisto in use. Left screenshot shows a photo of a fire juggling act, while the right one shows a scene of a person juggling clubs. With the start tag juggling the classical statistical approach (named "Sugg. Stat" in Fig. 2) are the same for both photos: juggler, fire, juggle, balls, etc. However, taking into account content based low level features leads to different recommendations per image (named "Sugg. NCP" in Fig. 2). For the fire juggling photo the tag *fire* is ranked first, while for the club juggling photo *juggler* is the first.

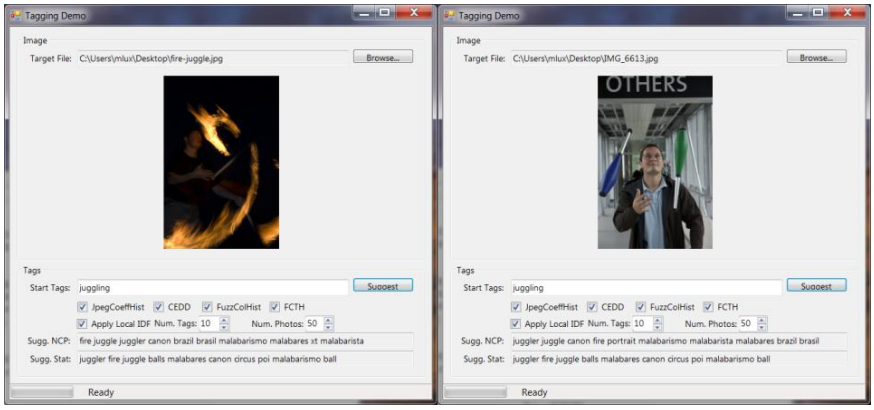


Fig. 2. Screenshot of the Callisto tag recommendation system for two different photos with the same start tag.

Callisto utilizes the data of Flickr (http://www.flickr.com) and heavily depends on the quality of annotations of the photos retrieved for tag recommendation. However in our experience the tag recommendations got better the more photos were considered in the process (step 1 in Fig. 1). With a critical mass of start tags recommendation also yields good results without a start tag and can be employed for auto-tagging.

Author Index