Sören Auer Sebastian Dietzold
Steffen Lohmann Jürgen Ziegler (Eds.)

Proceedings of the International Workshop on

Interacting with Multimedia Content in the Social Semantic Web

Co-located with the 3rd International Conference on Semantic and Digital Media Technologies (SAMT 2008)

Koblenz, Germany, December 3, 2008

Title: Proceedings of the International Workshop on Interacting with Multimedia Content in the Social Semantic Web (IMC-SSW'08)

Editors: Sören Auer, Sebastian Dietzold, Steffen Lohmann, Jürgen Ziegler

ISSN: 1613-0073

CEUR Workshop Proceedings Series
(CEUR-WS.org)

Editors' addresses:

University of Leipzig
Agile Knowledge Engineering and Semantic Web (AKSW)
Johannisgasse 26, 04103 Leipzig, Germany

{auer | dietzold}@informatik.uni-leipzig.de

University of Duisburg-Essen
Interactive Systems and Interaction Design
Lotharstr. 65, 47057 Duisburg, Germany

{steffen.lohmann | juergen.ziegler}@uni-duisburg-essen.de

Preface

Media sharing and social networking websites have attracted many millions of users resulting in vast collections of user generated content. The contents are typically poorly structured and spread over several platforms, each supporting specific media types. With the increasing growth and diversity of these websites, new ways to access and manage the contents are required – both within and across Web platforms. For these reasons, the convergence of the Social and the Semantic Web is of great potential for the future evolution of the Web.

The goal of the International Workshop on Interacting with Multimedia Content in the Social Semantic Web (IMC-SSW'08) is to provide a forum for researchers and practitioners from the Semantic Web, Human-computer Interaction, and multimedia communities to discuss these topics and share experiences from a multidisciplinary perspective. The focus is on new approaches that follow Web 2.0 principles of simplicity and/or social navigation in combination with the representation, annotation, and linking power of the Semantic Web. The half-day workshop is co-located with the third International Conference on Semantic and Digital Media Technologies (SAMT 2008) that takes place in Koblenz on December 3 2008.

The workshop consists of two sessions each containing two full and one position paper. The first session starts with Tobias Bürger who presents a model for deploying multimedia content descriptors that he calls Intelligent Content Objects (ICOs) on the Semantic Web in order to improve reusability. The approach consists of a data model, a metadata model and a way to deploy the ICOs inline as part of HTML pages using RDFa. Avare Stewart, Ernesto Diaz-Aviles, and Wolfgang Nejdl demonstrate how user activities on different social networking websites can be combined to improve recommendation and discovery of social links. They present empirical results from the music domain showing that it can be worthwhile to consider cross-site relationships. The position paper by Tobias Bürger and Michael Hausenblas defines some general principles and requirements for interlinking multimedia content on the Web. The authors describe different methods to achieve this by using Semantic Web and especially Linked Data principles. They identify user interactions as a major source to establish links between multimedia resources.

The second session starts with Andreas Walter and Gábor Nagypál who present an approach of semantically annotating images available on the Web and show how to implement a mashup based on their application ImageNotion and popular Web platforms such as Flickr. Furthermore, they describe how the mashup can be extended to a Semantic Web Service that semi-automatically creates semantic image annotations for the Web. Philipp Heim, Jürgen Ziegler, and Steffen Lohmann propose a novel approach for browsing interrelated Web data by combining a graph-based visualization with faceted filtering. They illustrate the benefits of this approach by a prototype and a scenario. Finally, they provide some first evaluation results based on their prototype. The position paper by Wolfgang Halb and Michael Hausenblas discusses the issues of provenance, trust, and privacy when interacting with contents in social media environments. The authors present a general model and describe how Semantic Web technologies can help to overcome limitations in these areas.

We thank all authors for their submissions and the members of the IMC-SSW program committee for providing their expertise and giving elaborate feedback. All contributions that have been selected by the program committee are published in these proceedings. Last but not least, we would like to thank the organizers of the SAMT conference for their support.

November 2008

Sören Auer, Sebastian Dietzold,
Steffen Lohmann, Jürgen Ziegler

Program Committee

- Scott Bateman, University of Saskatchewan, Canada
- Simone Braun, FZI Research Center for Information Technology, Germany
- Davide Eynard, Politecnico di Milano, Italy
- Michael Hausenblas, Johanneum Research, Austria
- Andreas Heß, Lycos Europe GmbH, Germany
- Knud Möller, DERI Galway, Ireland
- Jasminko Novak, University of Zurich, Switzerland
- Alexandre Passant, Université Paris-Sorbonne, France
- Yves Raimond, Queen Mary University of London, UK
- Harald Sack, University of Potsdam, Germany
- Andreas Schmidt, FZI Research Center for Information Technology, Germany

Contents

Towards Increased Reuse: Exploiting Social and Content Related Features of Multimedia Content on the Semantic Web

Tobias Bürger

Semantic Technology Institute (STI),
University of Innsbruck, Austria,
tobias.buerger@sti2.at

Abstract. While the amount of multimedia content on the Web is continuously growing, reuse of multimedia content remains low and automated processing of content remains hard. An increased reuse of content however would result in a greater consistency, quality and lowered cost of the production of new content. The retrieval of multimedia content on the Web is a continuous challenge which is due to the lack of (formal) descriptions and generic multimedia analysis algorithms. We present a model and a set of ontologies to mark up multimedia content embedded in web pages which can be used to deploy its descriptions on the Semantic Web and which in turn can be used to reason about the contents of multimedia resources. This model is part of a method to raise the reusability potential of content which in turn is expected to lower production costs of new content.

1 Introduction

In 2001, Tim Berners-Lee et al. introduced their vision of an augmented Web in which information is meaningful for machines as well as for humans. Since the introduction of this vision, the Web more and more turned into a multimedia environment and became a place to share great amounts of professionally and user generated content.

One of the challenges brought forward by the Semantic Web is the need to enrich existing digital content in such a way that machines can determine what the content is about, how it can be used, and whether one needs to pay for it or not. This part of the vision, richly annotated and formally described content which supports its automated negotiation, is sometimes referred to as *Intelligent Content*. Not only the ever growing amount of digital content raised interest in this topic in the recent years, but also the lack of appropriate multimedia description standards for the explication of features of content [22, 16]. Relevant features include not only the semantic content of images but also structural, legal or behavioral issues as especially multimedia content has many characteristics that for different usage scenarios need to be described.

One question that we especially intend to answer is how multimedia content which is published on the Web can be described to efficiently be reused, republished or reformatted for different purposes or different target media. And more concrete: How can social and content-related features and descriptions of content on the Web be exploited to increase the reusability potential of content.

The contribution of this paper is a conceptual model to describe and represent multimedia content including its context on the Web. The model is especially designed to raise the reusability potential of content published in typical web pages (e.g. embedded in news stories), social media or professional image licensing sites. The model is supported by a set of ontologies to mark up multimedia resources inline of HTML pages using RDFa[1].

2 Motivating Scenario: Towards Web-scale Reusability of Content

Supporting the reuse of content can provide significant improvements in the way how content is created and used, including increased quality and consistency, long-term reduced time and costs for development, maintenance, or adaptation to changing needs [18]. The amount of content available on the Web grows every day and the amount of professionally produced content in local or commercial databases also stays on a high level all of which could potentially be reused. The wish for reuse of content comes inline with the need for automation of associated tasks like search & retrieval, selection, or adaptation of content. However, automated handling is mostly hindered by the fact that users search for content based on the aboutness of the contained information which is – if at all – represented by tags attached to the content on the Web. Still, high-level features which are of high importance for retrieval of content [13] are not automatically derivable by most analysis algorithms which is due to the Semantic Gap [19], which commonly refers to the large gulf between automatically extractable low-level features and high-level semantics which are typically derived based on the background of a human being. Furthermore licenses and conditions of use are mostly encoded in web pages using natural language which is understandable by humans but not by machines. With the aim to automate handling which includes selection of the right pieces of content, this fact demands for richer semantic descriptions of content. Having richer semantic descriptions in turn implies improvements in reuse and automation.

Currently content on the Web is published based on different metadata standards and with different intention in mind. End users either publish images for non-commercial aspects, i.e. for others to watch or to gain reputation, or they publish it out of a commercial reason, either because they want to sell it or to grant access in order to gain revenues.

Current Web-based content reuse is difficult because amongst others

- Presentations are mostly available as PDF files with some tags attached to them. However fine granular descriptions of images, which were used in some slides and which would be candidates for reuse, are mostly missing.
- The same is true for videos: Tags are provided which are often not covering the semantics of particular scenes but only of the whole video.
- Cross-site searches in commercial image libraries are often hindered by the fact that images are described differently across sites.
- A huge amount of sites are using images for illustrative purposes which are not explicitly described. Sometimes textual descriptions are provided with the images but these are not explicitly assigned to the image which again makes retrieval hard.

[1] `http://www.w3.org/TR/xhtml-rdfa-primer/`

We intend to overcome these difficulties using a **two-fold strategy** in order to raise the reusability rate of content: (1) Raise the findability rate of content by unlocking the reusability potential of content published on social media, non-commercial or commercial sites by formally describing their content related- and social features and (2) Increase the ad-hoc adaptability of content by allowing to select and describe parts of it instead of only the singular datastream / resource.

3 Requirements for Intelligent Content with respect to Reusability

Content is an *individual securable and targeted reproduction of implicit information done by humans*. Important aspects of content, which are important with respect to reuse, include:

- the *contextual aspect* (ie. what the information or the content is about),
- the *technical aspect* (ie. how is it technically represented,),
- the *economic aspect* (ie. what is the value of the content), and
- the *legal aspect* (ie. what are the rights to use the content)

Intelligent Content Objects (ICOs) – as we understand them – are inspired by the vision of smart content objects [1], which define a package structure including the content, knowledge about its properties and several interfaces to interact with the smart content object. This is similar to packaging standards like MPEG-21 [4] or OAI Information Packages [7] with the difference that semantic technologies are explicitly used to provide machine understandable descriptions.
Requirements for multimedia content descriptions have been researched before and investigations of the combination of multimedia descriptions with features from the Semantic Web are yet numerous which we summarize in [5]. We want to highlight this issue again with respect to reusability of content which we believe deserves special attention.

We identified 3 aspects that need to be fulfilled in order to increase the reusability potential of content:

1. **Findability**: Reuse is often hindered by the fact that people are not aware of content to be re-used because it can not be found. This is especially true for multimedia content. Thus there is a need for a metadata model especially supporting findability and reusability of content. The model has to support descriptive information but also needs to support linking and referencing of secondary information and to acknowledge the fact that on the Web2.0 different users may provide metadata by tagging, rating, or referencing.
2. **Adaptability**: Resources published on the Web are mostly atomic. They are mostly available in a single file even if the file has been assembled out of different datastreams. This makes content difficult to repurpose. The existence of a reusability-friendly format that makes structure explicit could however enable the reuse of components as well. Thus there is the need for a model that allows to access subcomponents and which enables to structure content and to identify and select its sub-parts.

3. **Cross-Community Interoperability**: Query mechanisms for content must reflect habits of people from different communities. People from the E-Learning domain are used to think in terms of learning objects and content fragments while people from the archival domain communicate in terms of Information Packages or Information Objects, etc. This demands for a basic compatibility with existing standards.

4 A Conceptual Model for Intelligent Content for the Semantic Web

In this section we present a reference model for intelligent content which takes the characteristics of the Semantic Web [5] and the characteristics of the Web 2.0 as being a paradigm for rich social interaction on the Web into account.

4.1 A Data Model for Intelligent Content

Firstly there is a need for an abstract model that offers a set of well-defined concepts and vocabularies to sketch the problem of how content can be described and how it supports *adaptability* and *findability* while remaining compatible with existing data models from the multimedia, E-Learning and archival domain to support *interoperability*.

The general aim is to lay a graph over published contents on the Web, associating descriptions in a Web page to it and to bind metadata to content and descriptive information. This approach is similar to the way how digital assets are organized in the information domain. Here digital assets aggregate multiple-streams of relevant data, descriptive metadata and secondary data into one compound object which is then managed by a single entity (cf. section 6).

In [3] Boll et al. compared multimedia document models according to advanced requirements for reusability, adaptability and usability from a technical perspective. Important aspects regarding reusability are: (1) Granularity of media elements, fragments and documents (2) Kind of reuse, i.e. structural or identical, and (3) Identification and selection. Based on an assessment of different models according to these characteristics we selected the MPEG-21 Digital Item Declaration (DID) Abstract Model [4] as a data model that fulfills the basic characteristics of an adaptable data model, i.e. granular description of fragments, media elements (resources), grouped resources (i.e. components) and identification and selection of (parts of) resources. The basic parts of an MPEG-21 Digital Item which are interesting from this aspect are depicted in Figure 1: These include most notably containers in which identifiable digital assets are included (items) and which contain (multimedia) resources. Fragments of these resources can be selected and both resources and their fragments can be described via descriptors or annotated via annotations. Our proposed data model is realized by an ontology which covers the MPEG-21 DID Abstract Model and which amongst others makes the semantic types of relations between media elements, components or fragments and their descriptors explicit. The ontology is briefly described in section 4.3. The MPEG-21 DID Abstract Model is amongst other compatible with the OAI Abstract Information Model as shown in [2] and our first investigations also indicate that it is compatible to existing learning content models.

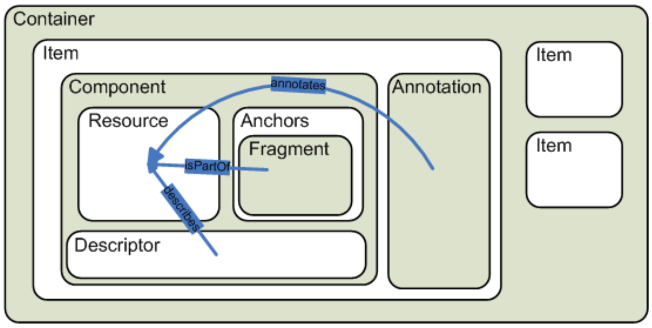


Fig. 1. MPEG-21: Main elements within the Digital Item Declaration Model

4.2 A Metadata Model to Increase the Findability Rate of Content

Multimedia retrieval is the discipline of applying information retrieval techniques to non-text- based content. The critical point for these techniques is, that users search for non-text based content based on the aboutness of the contained information [13] which is – if at all – represented by tags attached to the content or by information derived from the deployment context of the content which – when used for search – is blurred in the retrieved result sets. This is why reliable metadata is often essential to enable retrieval of multimedia content. It is commonly acknowledged that a metadata model which increases the findability rate of content also increases the reusability potential of content [18]. We follow a **two-fold strategy to increase the findability rate of content**: (1) First descriptions provided along with the content on web sites are explicitly related with the content to provide hints for search engines where to find information that can be used for indexing. This is supported by the data model introduced in section 4.1. (2) Secondly, content will be accompanied by metadata sets following a metadata model which captures relevant social and content related features and which is outlined in the subsequent section.

Social Aspects: The Role of Different Users in Content and Metadata Production We analysed the life cycle of both content and metadata and the roles of distinct user groups in order to determine components that our metadata model has to provide. Here, we especially took social aspects of content into account which is an important indicator for reuse. The content lifecycle consists of the following different dimensions (cf. [14]):

1. **The User Dimension**: Content and metadata is produced, altered and consumed by different users playing different roles: Production related users who create, process, resell, or publish content and end users who mainly consume but also increasingly produce content.
2. **The Content Dimension**: During its lifetime content is transferred between different stages. According to the canonical processes of media production [10] it is premediated, created (processed), annotated, packaged, organized and distributed.
3. **The Metadata Dimension**: Metadata is potentially being added by different users in every step of the content lifecycle. In [9] four different metadata creation roles

are introduced: The content creator who directly provides metadata, professional metadata creators who get paid for annotating content, technical metadata creators who just add basic technical metadata, and community enthusiasts which are very prominent in the Web 2.0 and tag content.

From this observation we are able to derive that metadata about content is not static and should be changeable during lifetime. Furthermore metadata is potentially being provided by different parties.

Unblurring Content Descriptions: Supporting Multiple Metadata Sets In the literature the distinction is made between authorative and non-authorative metadata. Authorative metadata is contributed by the author (creator) of the content and reflects persistent information about the content. Non-authorative metadata is provided by the consumer or a third party and provides contextual and changing aspects [17]. Non-authorative metadata is especially useful for recommendations based on collaborative filtering techniques and thus is critical in the effective discovery and reuse of content. To reflect different opinions and interpretations of content, metadata provided by different parties should be connected to its originator which is an important indication of its quality and trustworthiness and thus should be kept separate. Our model therefore explicitly supports *one authorative metadata and multiple non-authorative sets to be attached to different parts of the data model*. While authorative metadata is explicitly added, non-authorative metadata may explicitly be provided (through annotations, reviews, ratings, etc.) or implicitly be generated (through harvesting or usage analysis).

Types of Metadata As previously said, metadata is critical for the discovery of non-text based content on the Web. Metadata standards or vocabularies for multimedia are yet numerous as we summarize in [12].

Our investigations of standards and types of metadata focused on a core set which reflects the properties of content in its lifecycle. This set mostly covers the core facets which are used across a variety of domains and which we believe are important to support findability with respect to reuse:

1. **Bibliographic metadata** is traditionally concerned and related to the authorship of content and includes basic fields like identification, naming, publication or categorization.
2. **Technical metadata** typically describes physical properties of content, like format, bit-rate and what is mostly called low-level features of content.
3. **Classification metadata** might include keywords or tags but also domain specific classification information
4. **Evaluative metadata** includes ratings and qualitative assessments of the content. A collaborative evaluation model based on evaluative metadata could provide invaluable information regarding reuse of content. This could include dimensions like usefulness, presentation aesthetics, or design.
5. **Relational metadata** is one of the most important parts to be able to explicitly define relations between the content and other related information / content. Relations can include explicit ones which are given through the design of the content object

or external objects. However it might also contain implicit relations gathered by observations or the usage history.
6. **Rights metadata** are of utmost importance with respect to reusability as they declare the terms of use.
7. **Functional metadata**: Functions may be supported to alter the presentation of the content, to customize or personalize the content, or to provide access to different versions.

The selection of the different metadata types with respect to reuse was based on a set of expert interviews and is currently empirically validated in a survey. The assignment of metadata to resources is not restricted to the above mentioned types but is open to domain-specific assignments like educational value for a learning object or preservation data for archival information. In this respect our approach is inline with the vision of Resource Profiles as described in [8].

4.3 Ontology Framework

The data and metadata model as explained in the previous sections are implemented using a set of ontologies in order to publish and describe ICOs on the Web. The import-graph of the ontologies used is depicted in Figure 2[2]: The main ontology is the RICO

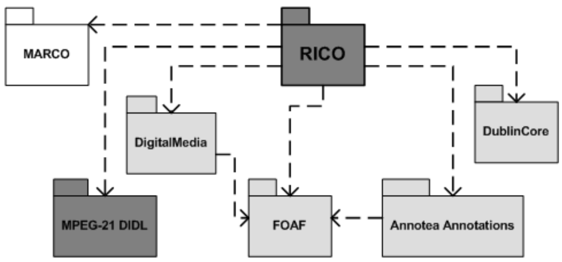


Fig. 2. The Reusable Intelligent Content Objects (RICO) ontology

("Reusable Intelligent Content Objects") - ontology which imports a set of other ontologies. RICO is an OWL-DL ontology. Most notably it makes use of

- the MPEG-21 DIDL ontology which we built to reflect the data model presented in section 4.1,
- the Mindswap Digital Media Ontology which is used to type resources[3]
- an OWL-DL version of the FOAF ontology as provided by Mindswap[4],

[2] The dark grey and white ontologies were built in the course of this work.
[3] The Digital Media Ontology available at `http://www.mindswap.org/2005/owl/digital-media` has been slightly adapted to be in OWL-DL.
[4] `http://www.mindswap.org/2003/owl/foaf`

- the Annotea annotations ontology to represent annotations[5],
- the OWL-Lite version of the Dublin Core ontology, and
- the MARCO ("Metadata for Reusable intelligent Content") - ontology which is currently work in progress and which will cover aspects of the metadata model as presented in section 4.2.

5 Deployment of Intelligent Content Objects on the Web using RDFa

ICOs are published as compound objects on the (Semantic) Web following the data model described in section 4.1, including metadata as described in section 4.2 and marked up with RDFa using the ontologies as described in section 4.3. An ICO includes the resource whose structure is described using the data model and multiple metadata records including the metadata types previously presented.

The compound package information is about to be published inline within an HTML page and will extend the ramm.x ("RDFa based multimedia metadata") model that we suggested in [11] and which can be used to deploy multimedia metadata using RDFa.

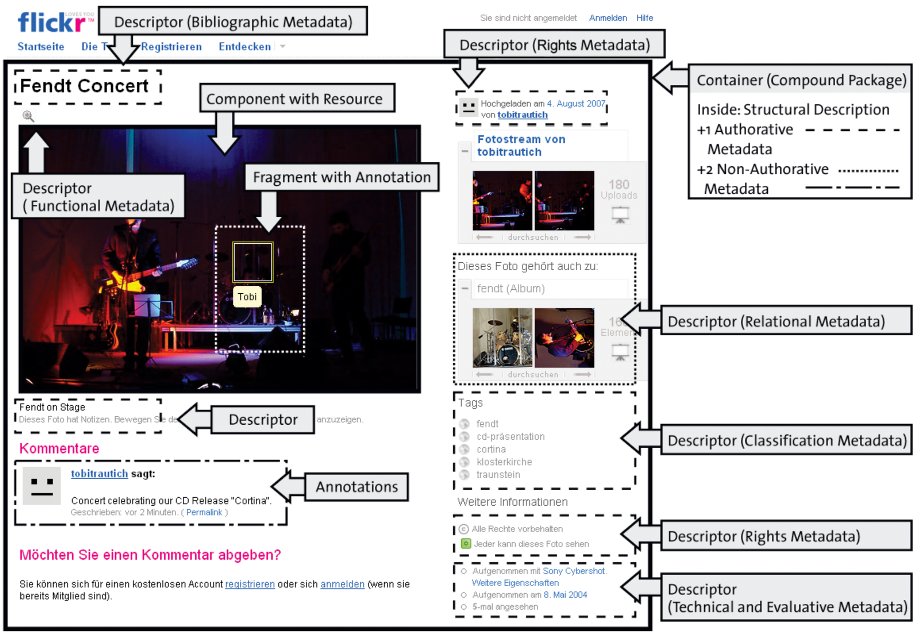


Fig. 3. A image hosted by Flickr deployed as an Intelligent Content package

[5] The Annotea annotations ontology available at `http://www.w3.org/2000/10/annotation-ns#` has been rebuilt in OWL-DL.

Figure 3 shows an example of an ICO, i.e. an image hosted by Flickr for which all descriptions are explicitly marked up and related to the image. The ICO contains three metadata sets: one authorative set (as provided by the owner/creator of the image) and two non-authorative sets (one provided by the hosting platform and one provided by an end user through a commentary). The figure shows only how visible information is related to the image. However further additional (non-visible) metadata could also be provided, e.g. by providing a detailed description of different scenes in a video or further semantic descriptions of the content of an image. Parts of the resulting RDF graph are visualized in Figure 4 which is however not showing the entire graph because of space restrictions (i.e. most descriptors are omitted).

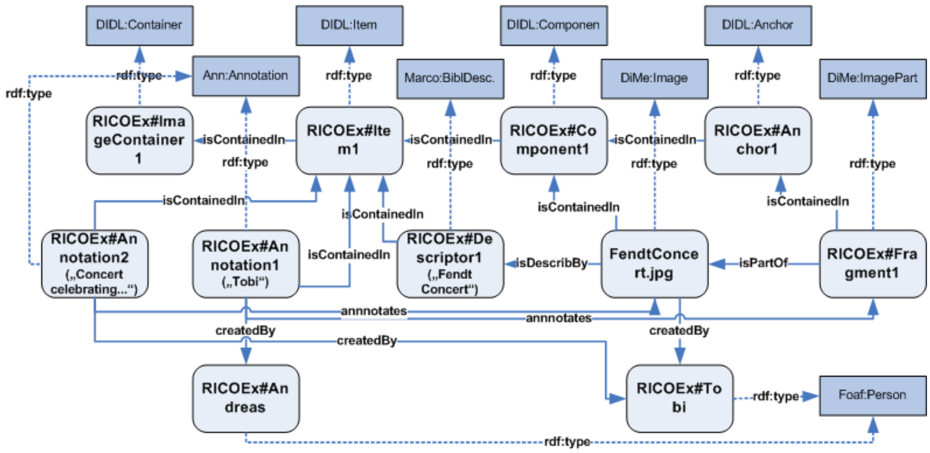


Fig. 4. Partial description of the compound object from Figure 3.

6 Related Work

The model and the ontologies can be used to markup illustrative images in typical web pages, images or videos hosted by social media sharing sites, slides and images embedded in commercial offerings, or multimedia content deployed in blogs and wikis.
Especially in recent years much work has been done on the specification of ontologies that aim to combine traditional multimedia description models, thus trying to develop models that allow reasoning over the structure and semantics of multimedia data (see [5] and [12] for a comprehensive overview). Models with a similar intention, i.e. to publish semantic metadata inline of HTML pages, include lightweight approaches like the hMedia microformat[6] which is a basic vocabulary to mark up media resources on web sites using property value pairs. Furthermore it is related to the ramm.x model which provides a small but extensible vocabulary for marking up resources to include legacy

[6] http://wiki.digitalbazaar.com/en/Media_Info_Microformat

metadata. Ramm.x however does not include something similar to our data model or a detailed metadata model. The intention of the SMIL MetaInformation-Module[7] is to publish RDF-based metadata in SMIL presentations. It is very general and does not prescribe how to use it. We intend to test the applicability of our proposed model with SMIL in the future. The intention of our model is also similar to Adobe's XMP whose intention is to publish RDF-based metadata into PDFs or other document formats. Furthermore we want to acknowledge the work being done by Creative Commons to describe and embed licensing data using RDF which is exploited in searches by Yahoo or in Flickr[8].

More heavyweight approaches include *Intelligent Content* models as previously assessed for example in [6] and which cover a broad range of aspects. Most of these approaches are too heavy for our proposed model.

Traditional models include the standardized framework of MPEG-7 [15] or packaging formats from the archival or E-Learning domain which include the IMS Content Packaging format[9], the Metadata Encoding and Transmission Standard (METS)[10], the MPEG-21 Digital Item Declaration (DID) [4] and most recently the OAI-Object Reuse and Exchange - model (OAI-ORE)[11]. The OAI-ORE model has been designed as an exchange format for scholarly works. Its compound objects model[12] [21] is similar to our model as it also provides facilities to publish semantic descriptions as an overlay graph over web pages. The approach however does not focus on multimedial aspects.

The intention of the presented model is not provide a new standard for the description of the semantic content and content decomposition like it is done by MPEG-7. Thus our approach is only marginally related to endeavors that aim to combine MPEG-7 with semantic technologies like the COMM ontology or other available MPEG-7 ontologies (see cf. [20] for a comparison).

7 Conclusions and Future Work

In this paper we presented a model for deploying multimedia content descriptions, i.e. Intelligent Content Objects, on the Semantic Web with the goal to increase the reusability potential of content in general. The model consists of a data model that supports adaptability of content, a metadata model including properties to explicitly increase findability with respect to reuse (e.g. implicit usage information, ratings, etc.) and a deployment facility to publish content descriptions inline of HTML pages. Deploying resources using this model can have a similar effect like Yahoo's Search Monkey[13] which allows people to mark up their content using Microformats or RDFa whereas the additional information is then used to display and probably rank search results.

[7] http://www.w3.org/TR/2007/WD-SMIL3-20070713/smil-metadata.html
[8] http://search.creativecommons.org/
[9] http://www.imsglobal.org/content/packaging/
[10] http://www.loc.gov/standards/mets/
[11] http://www.openarchives.org/ore/
[12] http://www.openarchives.org/ore/documents/CompoundObjects-200705.html
[13] http://developer.yahoo.com/searchmonkey/

Future work includes the engineering of the MARCO ontology and a qualitative evaluation of our approach. The evaluation of the parts of the model is work in progress. The basic model, which includes the data model and the metadata categorization, meets the requirements of a typical multimedia publishing scenario on the Web and fulfills the criteria of an adaptable data model as defined in [3]. However the effect of the different metadata types on the reuse of content has yet to be validated. We are currently empirically validating the influence of different metadata types on the reuse of content in a study which will be accompanied by an implementation. Different aspect that would also demand special attention but which are beyond the scope of our work include the assessment of the heterogeneity or the consistency of different metadata sets.

References

1. Wernher Behrendt, Guntram Geser, and Andrea Mulrenin. Ep2010 - the future of electronic publishing towards 2010. Information Society DG - Unit E2, EUFO 1-275, Rue Alcide de Gasperi, L-2920, Luxembourg, 2003.
2. Jeroen Bekaert, Emiel De Kooning, and Herbert van de Sompel. Representing digital assets using mpeg-21 digital item declaration. *International Journal on Digital Libraries*, 6(2):159– 173, 2006.
3. Susanne Boll and Woflgang Klas. Zyx – a multimedia document model for reuse and adaptation of multimedia content. *IEEE Transactions on Knowledge and Data Engineering*, 13(3):361–382, 2001.
4. Jan Bormans and Keith Hill. Mpeg-21 overview v.5, 2002. ISO/IEC JTC1/SC29/WG11.
5. Tobias Bürger and Michael Hausenblas. Why real-world multimedia assets fail to enter the semantic web. In *Proc. of the Int. Workshop on Semantic Authoring, Annotation and Knowledge Markup (SAAKM07)*, 2007.
6. Tobias Bürger, Ioan Toma, Omair Shafiq, and Daniel Dögl. State of the art in sws, grid computing and intelligent content objects - can they meet? GRISINO Deliverable D1.1, http://www.grisino.at, October 2006. GRISINO Deliverable D1.1.
7. CCSDS. Reference model for an open archival information system. Blue Book 1, Consultative Committee for Space Data Systems, CCSDS Secretariat Program Integration Division (Code M-3) National Aeronautics and Space Administration Washington, DC 20546, USA, Januar 2002.
8. Stephen Downes. Resource profiles. *Journal of Interactive Media in Education Special Issue on the Educational Semantic Web*, 5:0, 2004.
9. Jane Greenberg. Metadata generation: Processes, people, and tools. *Bulletin of American Society for Information Science and Technology*, 29(2), 2003.
10. Lynda Hardman. Canonical processes of media production. In *MHC '05: Proceedings of the ACM workshop on Multimedia for human communication*, pages 1–6, New York, NY, USA, 2005. ACM Press.
11. Michael Hausenblas, Werner Bailer, Tobias Bürger, and Raphael Troncy. Ramm.x: Deploying multimedia metadata on the semantic web. In *Proceedings of SAMT 2007, Dec 4-7, Genova, Italy*, 2007.
12. Michael Hausenblas, Susanne Boll, Tobias Bürger, Oscar Celma, Christian Halaschek-Wiener, Erik Mannens, and Raphael Troncy. Multimedia vocabularies on the semantic web. W3c incubator group report, W3C, July 2007.
13. Laura Hollink, Guus Schreiber, Bob Wielinga, and Marcel Worring. Classification of user image descriptions. *Int. J. Hum.-Comput. Stud.*, 61(5):601–626, 2004.

14. Harald Kosch, Laszlo Boszormenyi, Mario Doller, Mulugeta Libsie, Peter Schojer, and Andrea Kofler. The life cycle of multimedia metadata. *IEEE MultiMedia*, 12(1):80–86, 2005.
15. José M. Martínez-Sanchez, Rob Koenen, and Fernando Pereira. Mpeg-7: The generic multimedia content description standard, part 1. *IEEE MultiMedia*, 9(2):78–87, 2002.
16. Frank Nack, Jacco van Ossenbruggen, and Lynda Hardman. That Obscure Object of Desire: Multimedia Metadata on the Web (Part II). *IEEE Multimedia*, 12(1), 2005.
17. Mimi Recker and David Wiley. A non-authoritative educational metadata ontology for filtering and recommending learning objects. *Journal of Interactive Learning Environments*, 2001.
18. Ann Rockley. *Managing Enterprise Content: A Unified Content Strategy*. New Riders, 2002.
19. Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1349–1380, 2000.
20. Raphael Troncy, Oscar Celma, Suzanne Little, Roberto Garcia, and Chrisa Tsinaraki. Mpeg-7 based multimedia ontologies: Interoperability support or interoperability issue? In *Proceedings of the 1st Workshop on Multimedia Annotation and Retrieval enabled by Shared Ontologies (Mareso '07)*, 2007.
21. Herbert van de Sompel and Carl Lagoze. Interoperability for the discovery, use, and re-use of units of scholarly communication. *CTWatch Quarterly*, 3(3), August 2007.
22. Jacco van Ossenbruggen, Frank Nack, and Lynda Hardman. That Obscure Object of Desire: Multimedia Metadata on the Web (Part I). *IEEE Multimedia*, 11(4), 2004.

Mining User Profiles to Support Structure and Explanation in Open Social Networking

Avaré Stewart, Ernesto Diaz-Aviles, and Wolfgang Nejdl

L3S Research Center / Leibniz Universität Hannover
Appelstr. 9a, 30167 Hannover, Germany
{stewart,diaz,nejdl}@L3S.de
http://www.L3S.de/

Abstract. The proliferation of media sharing and social networking websites has brought with it vast collections of site-specific user generated content. The result is a Social Networking Divide in which the concepts and structure common across different sites are hidden.
The knowledge and structures from one social site are not adequately exploited to provide new information and resources to the same or different users in comparable social sites. For music bloggers, this latent structure, forces bloggers to select sub-optimal blogrolls. However, by integrating the social activities of music bloggers and listeners, we are able to overcome this limitation: improving the quality of the blogroll neighborhoods, in terms of similarity, by 85 percent when using tracks and by 120 percent when integrating tags from another site.

Key words: Open Social Networking, Cross Domain Discovery

1 Introduction

The increasingly growing collections of user generated content spread over heterogeneous social networking and media sharing platforms, each supporting specific media types. The content typically has a latent structure and latent, interrelated topics: this has resulted in a Social Networking Divide.

Recent advances toward a more *Open Social Networking* (OSN) paradigm are focused on (*de facto*) standards, and only address part of the problem. Specifically, current OSN efforts attempt to handle issues related to the portability of data, common APIs (e.g., Google OpenSocial[1]), and social graphs, e.g., FOAF[2], XHTML Friends Network[3].

We posit that open social networking is more than an agreed upon "language" for describing relationships and sharing data across systems. In addition, it is the exploitation of social activities in one site, to support the discovery of new interrelationships within a community. This is crucial, given that it is becoming increasingly difficult for seekers to cope with the cognitive challenges of efficiently finding and effectively analyzing relevant information, when inundated with its volume, variety and evolution.

[1] http://code.google.com/opensocial
[2] http://www.foaf-project.org/
[3] http://gmpg.org/xfn/

1.1 Scenario

To motivate the aforementioned ideas, consider the following scenario in which there are two social network sites. In one site, a *Blogger.com*[4], music community, the main activities are writing text about artists, tracks, albums or music videos. Bloggers create explicit links to other participants to express their preferred blogs (via a *blogroll*). In the other site: *Last.fm*[5], the users listen to music tracks, tag these tracks and build friendship relationships.

Symbiotically, the social activities in one site can have an impact on the other site. Blogger.com bloggers do not tag the entities about which they write. However, the tagging activity can better help bloggers see the structure in their community and find new information. Conversely, Last.fm users do not provide prose for the tracks they listen to, but such prose can be a valuable source of metadata for audio tracks.

In Figure 1, a navigation tool is depicted, in which the Blogger.com site has been enriched with information from Last.fm. The graph represents the similarity between (potentially unknown) bloggers on the set of tracks they have written about in their blogs. By selecting a node in the graph, the list if tracks that blogger has mentioned in their blogsite is presented, along with the overall popularity of these tracks. Also depicted in the figure are the tags from Last.fm, which can be used to filter the nodes and edges in the graph. The tool supports navigation and visualization of latent concepts and relations within, and across the sites. One of the challenges in realizing such a

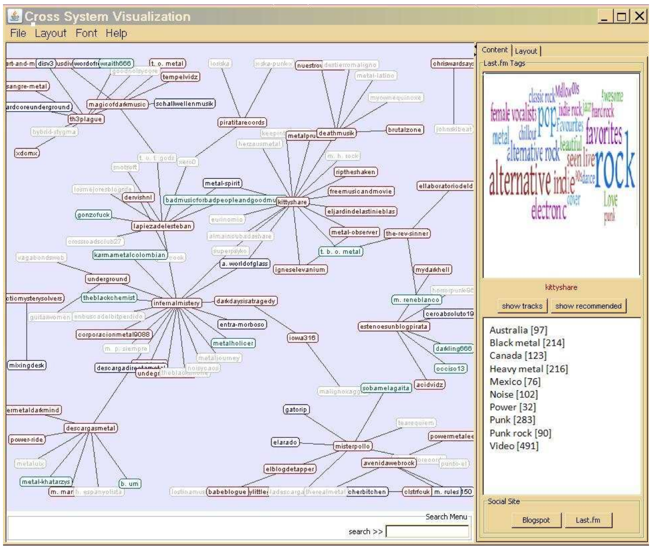


Fig. 1: Cross System Visualization Blogger.com and Last.fm

[4] `http://www.blogger.com`
[5] `http://www.last.fm`

scenario is that the concepts and structure within —and across domains— are latent. Specifically within blogs, the topics to which the blog site is devoted, are very often not made explicit. Furthermore, the readership relationship is typically unobserved; and the blogroll relationship, though observed, is often unexplained. Finally, standards may support, but do not address, how the social practices in one domain may be exploited to support bloggers in another similar social site; particularly when resources of different types are being mapped.

The contributions of this work are: 1) extension of the commonly held view to open social networking: to infer new relationships and resources, and provide support with navigation and visualization across comparable social networks; 2) integration music bloggers with music listeners; bridging the gap between different types of social networking systems; 3) examination of the blogroll relationship in the music domain based an open social networking approach.

In Section 2 we discuss related research in cross domain discovery. In Section 3 we present the conceptual approach to open social networking in the music domain. In Section 4 experimental results are presented and in Section 5 we conclude and discuss future work.

2 Related Work

In machine learning, cross domain discovery [1] or domain adaptation [2] is a body of work in which multiple information sources from comparable, but different domains are combined. Work done in this area has focused on classification tasks, in which labeled data exists in abundance in one domain, but a statistical model that performs well on a related domain is desired. Since hand-labeling in the new domain is costly, one often wishes to leverage the original out-of-domain data when building a model of the new, in-domain data [3].

Also in the area of machine learning, the method of *View Completion* [4], has used collaborative tags to heuristically complete missing or inadequate feature sets (or views). The basic premise underlying view completion, is that for many tasks, combining multiple information sources yields significantly better results than using just a single one alone. Views are used in this case, since that blogs are not typically available on collaborative tagging websites, and as such the tags provided by bloggers suffer from the vocabulary problem and cannot be adequately used as a shared index.

Another related area is *Cross-System Personalization* [5] which enables personalization information across different systems to be shared. In digital libraries, cross system personalization is used to overcome the problem that information needed to support a personalized user experience is not shared among different libraries. In other work, the focus has been on adequate representations of [6, 7], and dependencies between [8] the user's profile, to support a unified representation in the different systems. These approaches are ego-centric in that they assume the same user to exist across different systems; and that the user is interested in an aggregated view of their profile or social networking information. This is not the case in an open social networking environment, where users are assumed to be similar (in some way), but have distinct digital identities.

Our view is a socio-centric one and focuses on common patterns in the community as a whole.

Work exists, in the area of association mining [9, 10]. In [9] the goal is to provide a seamless navigation between tag spaces. The work presented in [10] merges the areas of formal concept analysis and association rule mining to discover shared conceptualizations that are hidden in folksonomies. They present a formalism for folksonomies that includes a set U of users, a set T of tags, and a set R of resources, represented by the ternary relation, Y. In our work, it is the set, T, of tags in the ternary relation that we propagate from one site to enrich another. Furthermore, we propose, that "citizen-defined" structuring (i.e. blogroll, friends, or comment networks) allow other types of ternary relations to be inferred, that are not restricted to the *user-tag-resource* triple.

3 Open Social Networking: the Music Domain

Open social networking, has two-fold goal: 1)improve the structure of information within a single site and 2)exploit the social activities in a different sites to enhance the activities in comparable ones. Toward this end, two aspects are considered: a representation for the activities within each social site; and mapping parts of the data and structures from one social site to another, to augment the social activities therein.

3.1 Terminology

We adopt the definition of a folksonomy as described by Hotho et al. [10, 11], as a four-tuple[6], $\mathbb{F} := (U, T, R, Y)$, where:

- U,T and R are finite sets, whose elements are called users, tags and resources, respectively, and
- Y a ternary relation between them, i.e. $Y \subseteq U \times T \times R$, whose elements are called tag assignments.

For the music domain, the set of resources are considered to be artists, tracks and albums. Additionally, we distinguish the different roles a site may have when describing the mappings between them. A target site, or *in-domain site*, is the one onto which data from another social site is mapped. The *out-of-domain site* is the social site from which data is extracted to augment the comparable, target site. The roles of in- and out-of-domain may be interchanged depending upon integration goals, and there may be multiple out-of-domain sites. For the purpose of this work, we consider $Blogger.com$ and $Last.fm$ to be the *in-domain site* and *out-of-domain site*, respectively.

3.2 Cross-Site Enrichment

We represent the Blogger.com music community conceptually as tuples, $B := \{(u_b, r_b) \mid (u_b, r_b) \in U_{Blogger.com} \times R_{Blogger.com}\}$, since Blogger.com bloggers

[6] In the original definition [10, 11], it is additionally introduced a subtag/supertag relation, which we omit for the purpose of this paper.

do not tag the entities about which they write. On the other hand, the Last.fm social site is ripe with tag data of the form $L := \{(u_l, t_l, r_l) \mid (u_l, t_l, r_l) \in U_{Last.fm} \times T_{Last.fm} \times R_{Last.fm}\}$, representing the tag a given user has applied to a track within Last.fm. Then the mapping of tags onto Blogger.com is computed as follows:

$$Y := \{(u, t, r) \mid (u, t, r) \in \pi_{u_b, t_l, r_b}(\sigma_{r_b = r_l}(B \times L))\} \tag{1}$$

were σ and π are the relational algebra operators for selection and projection, respectively.

First, from the cartesian product $B \times L$, the tuples with equal resources in both sites are selected, and then the projection is taken over the Blogger.com users, the Last.fm tags, and the common resource elements. In general, the user sets are considered to be disjoint, i.e., $U_{Blogger.com} \cap U_{Last.fm} = \emptyset$.

3.3 Site-Specific Enrichment

The (hidden) relationships between blogs and/or bloggers can be exploited to infer relationships between the entities within the blog site. However, even within a single blog community, the relationships between resources, may not be well understood. For this reason, we undertake an exploratory analysis of the blogroll relationship.

4 Experiments

The experimental goals are to first examine the explicit blogroll structure; laying the foundation for further analysis of an ideal or "optimal" resource-specific blogrolls. Resource specific blogrolls are those in which the nature of the blogroll is assumed to be explained in terms similarity in tastes for a given type of resource, i.e., track, or artist. Then, to investigate the extent to which these optimal resource-specific blogrolls: overlap with the explicit blogrolls; and with each other. In the remainder of this discussion "'optimal"' resource-specific blogrolls is referred to as optimal blogrolls.

4.1 Data Set

For Cross System Music Blog Mining, we used two data sets: one data set consisted of personal music blogs from $Blogger.com$, one of the most popular blogsites, whereas the second data set consisted of tagged tracks from $Last.fm$, a radio and music community website and one of the largest social music platforms. The details of each data set are presented in this section.

Blogger.com Community: The blogroll relationship induces a network representing a preferential reading of others people's blogs. The network data was collected by experimentally selecting seed bloggers using several music directories[7] and limiting the

[7] `http://www.musicblogscatalog.com/`
`http://yocheckthisjam.com/music-blog-directory/`
`http://www.blogged.com/directory/entertainment/music/rock`
`http://www.blogcatalog.com/directory/music/rock`

bloggers selected to the genre of pop and rock music in the Blogger.com domain. The blogroll for each seed was traversed, fanning out in a breath-first order, yielding a total number of bloggers equal to $|U_{Blogger.com}| = 976$.

Summary statistics for the overall structure and topological statistics for the largest five weak components are given in Table 1, from there it can be seen that the components exhibit varying structural properties and that the structural view provided by the blogroll is a disjointed one.

In addition to the community data, profiles were built by parsing the tracks in the user's blog and relying upon a dictionary of tracks gathered from MusicBrainz.org[8]. A total of 2196 unique tracks were collected; and for these tracks, a total 147801 Last.fm tags were obtained, which allowed us to construct the triples.

Table 1: Blogger.com Community Statistics and Components

No. Nodes	No. Edges	Weak Components	Avg. Component Size	Max. Component Size	Min. Component Size	Reciprocal Edges
976	2011	72	13.55	662	2	581

Comp. Id	No. Nodes	No. Edges	Avg. Cluster Coef.	Shortest Average Distance	Longest Average Distance	Reciprocal Edges
1	662	1755	0.492	2.797	8.134	568
2	53	62	0.574	2.113	5.811	13
3	28	27	0.964	0.964	1.892	0
4	19	20	0.526	2.0	4.684	0
5	16	15	0.751	1.687	3.187	0

4.2 Experimental Results

Blogroll Quality Based on Tracks and Tags We investigate the extent to which the explicit blogroll relationship, within the in-domain site, can be described by the similarity between track and tag profiles. For each user $u \in U$, i.e. blogger, a track-based vector profile (*track profile*) $\mathbf{u}$ is constructed such that $\mathbf{u} := \{0, 1\}^{|R|}$, with the ith dimension $\mathbf{u}_i$ set to 1 if the track $r_i \in R_u$, appears in the user's blog, and 0 otherwise. Alternatively, after including the tag information from the out-of-domain site, we constructed a profile based on tag annotations, which corresponds to the *tag profile* for the user, i.e., $\mathbf{u} := \{0, 1\}^{|T|}$, with the ith dimension $\mathbf{u}_i$ set to 1 if the tag $t_i \in T_u$, and 0 otherwise[9].

[8] `http://www.musicbrainz.org`
[9] The 20000 most popular tags were used to build the profiles, i.e. $|T| = 20000$

Table 2: Average similarity and overlap of the blogroll B and optimal blogroll B^*

Profile	$AvgSim(B)$	$AvgSim(B^*)$	Improvement (%)	$Avg(\|B \cap B^*\|)$
Track-Based	0.295	0.547	85%	1.48 $\approx$ 1 blogger, with probability = 0.085
Tag-Based	0.293	0.645	120%	1.37 $\approx$ 1 blogger, with probability = 0.081

To evaluate the quality of the explicit blogroll B_u of user u, we computed an average similarity score between a user and all the persons in his blogroll. We perform the similarity computation using the cosine-based measure:

$$sim(\mathbf{u}, \mathbf{v}) := cos(\mathbf{u}, \mathbf{v}) := \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{||\mathbf{u}|| \cdot ||\mathbf{v}||} \quad (2)$$

where **u** and **v** denote either the track or tag profiles of users u and v, respectively.

For the optimal blogroll B^* computation, we constructed a user-track (resp., user-tag) profile matrix and proceeded as follows:

(i) Compute the user similarity matrix $S_{|U| \times |U|} := (sim(\mathbf{u}, \mathbf{v}))$
(ii) Keep the highest k entries in each column of S.
(iii) Set the optimal blogroll based on track profile (resp., tag profiles) for user u, i.e., $B^*_{u\ R}$ (resp., $B^*_{u\ T}$), to be the users in the non-zero columns of the corresponding u's row.

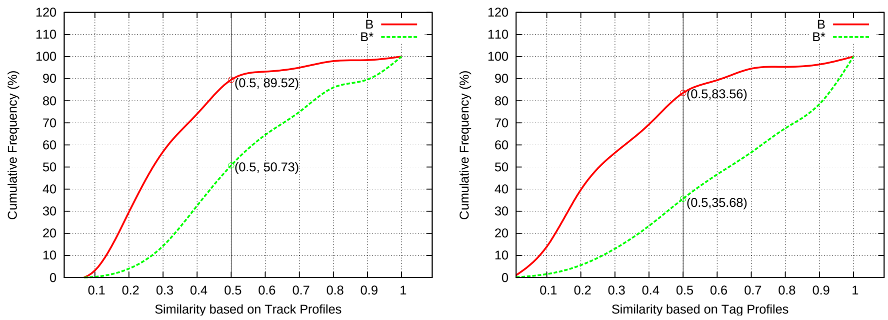
In our experiments we set the value of parameter $k = 10$. Table 2 summarizes the results.

B^*_R and B^*_T Overlap measures the extent to which optimal blogrolls, computed with track and tag profiles, agree on his members. We found that 77.66% of the time they agree on at least one member, and the average of the overlap in these cases is $Avg(|B^*_R \cap B^*_T|) = 4.64 \approx 5$ bloggers out of 10 (the fixed size of the optimal blogrolls). The distribution of the intersection size for the optimal blogrolls is presented in Fig. 4.

4.3 Discussion

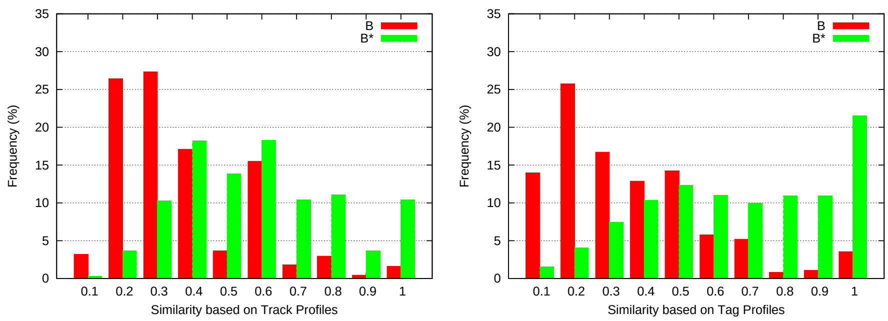
From Table 2, it can be observed that the improvement, in terms of similarity, when computing B^* based on tracks is 85%, and 120% when tag profiles are used. The table also shows that the overlap between the explicit and optimal blogrolls, computed either with track or tag profiles, occurs only 9% of the time, corresponding to an average of a single blogger in those cases.

Furthermore, the optimal blogroll similarity distributions are better than the one produced by the explicit relationships, as shown in Fig. 3, which corresponds to the absolute frequencies of the explicit and optimal blogrols, for different values of similarity. The frequencies of optimal blogrolls, for similarity values over 0.5, are higher than the ones for explicit blogrolls.



(a) (b)

Fig. 2: Cumulative Frequency for the explicit blogroll B and the optimal B^* computed using (a) Track Profiles and (b) Tag Profiles.



(a) (b)

Fig. 3: Frequency Distribution for the explicit blogroll B and the optimal B^* computed using (a) Track Profiles and (b) Tag Profiles.

The cumulative frequency of blogrolls over discrete similarity bins is presented in Fig. 2, which shows that both the track-based (49.27%) and tag-based (64.32%) optimal blogrolls exhibit good similarity quality, i.e., over 0.50, in contrast to the respective explicit blogrolls, where just less than 11% of them in the case of track-based profiles (resp., 16.44% for tag-based) fall in bins corresponding to similarity values over 0.5. Tag based computations perform better than using tracks, i.e., builds optimal blogrolls with higher values of similarity, this can be explained by the fact that tags capture some structure of the domain, e.g., genre, improving the overlap between user profiles when computing the similarity measure.

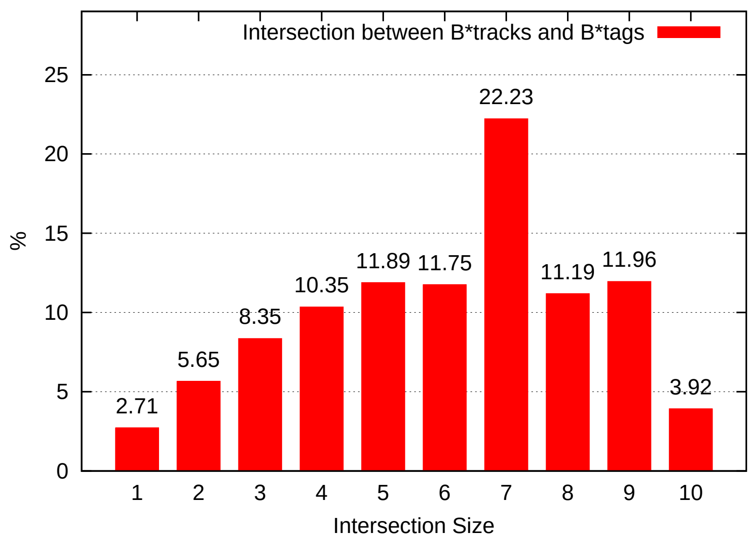


Fig. 4: Distribution of intersection size between the optimals blogrolls $|B^*_R \cap B^*_T|$

5 Conclusions

In this paper, we explored to what extent the knowledge and structures from one social site (out-of-domain) can be adequately exploited to provide new information and resources to the users in a comparable social site (in-domain), in particular for a music blog community within the $Blogger.com$ social network. An examination of the explicit blogroll structure, which is assumed to express a preferential reading of others people's blogs, has revealed that bloggers tend to produce sub-optimal blogrolls when measuring the similarity between users based on track, as well as tags. The implication for this is that if users are interested in learning about tracks, tags or other bloggers, some assistance to guide them is needed. On the other hand, neither tracks nor tags comes close to fully "explaining" the nature of the blogroll.

However, by integrating the social activities of music bloggers and listeners, we were able to overcome this limitation. We have shown the improvement that Open Social Networking can have on the quality of blogrolls: Last.fm offers better optimal blogrolls, than the tracks alone, from Blogger.com, improving the quality of the blogroll neighborhoods, in terms of similarity, by 85% when using tracks and by 120% percent when integrating tags from the site. The higher value of similarities computed based on tags can be explained by the fact that tags capture some structure of the domain, e.g., genre, increasing the overlap probability and size on the user profiles. Though this do not necessarily mean that tags are better predictor of similarity than tracks, it strongly suggests that the kind of information captured by tags can be exploited effectively, complementing tasks or models where tracks are used alone.

Although our investigation has provided promising results, we believe that our contribution is an initial step in the study of Open Social Networking, future work is required to evaluate the usefulness of optimal blogrolls, e.g., in providing recommendations. Furthermore, we plan to investigate the extent to which the explicit community bonds and "citizen-defined" structuring (i.e. blogroll, friends, or comment networks) can be described by mining and inferring associations between the profiles of users across social sites, towards a more general model, that considers new dimensions beyond the ternary relation between users, tags and resources.

Acknowledgements. This work was funded in part by the European Project PHAROS (IST Contract No. 045035), and by the Programme AlBan, the European Union Programme of High Level Scholarships for Latin America, scholarship no. E07D400591SV.

References

1. Swarup, S., Ray, S.R.: Cross-domain knowledge transfer using structured representations. In: AAAI, AAAI Press (2006)
2. III, H.D., Marcu, D.: Domain adaptation for statistical classifiers. J. Artif. Intell. Res. (JAIR) **26** (2006) 101–126
3. Dai, W., Xue, G.R., Yang, Q., Yu, Y.: Co-clustering based classification for out-of-domain documents. In: KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA, ACM (2007) 210–219
4. Shankara Bhargava Subramanya: View Completation And Collaborative Tagging In Blogosphere. Master's thesis, Arizona State University (July 2008)
5. Bhaskar, M., Hofmann, T.: Cross system personalization by learning manifold alignments. **4314/2007** (2006) 244–259
6. Niedere, C., Stewart, A., Mehta, B., Hemmje, M.: A multi-dimensional, unified user model for cross-system personalization. In: Proceedings of Workshop On Environments For Personalized Information Access at Advanced Visual Interfaces. (May 2004)
7. Wang, C., Zhang, Y., Zhang, F.: User modeling for cross system personalization in digital libraries. Information Technologies and Applications in Education, 2007. ISITAE '07. First IEEE International Symposium on (Nov. 2007) 238–243
8. Bhaskar, M., Hofmann, T.: Cross system personalization by factor analysis. (2006)
9. Oldenburg, S.: Comparative studies of social classification systems using rss feeds. In Cordeiro, J., Filipe, J., Hammoudi, S., eds.: WEBIST (2), INSTICC Press (2008) 394–403
10. Jäschke, R., Hotho, A., Schmitz, C., Ganter, B., Stumme, G.: Discovering shared conceptualizations in folksonomies. Web Semant. **6**(1) (2008) 38–53
11. Hotho, A., Jschke, R., Schmitz, C., Stumme, G. In: Information Retrieval in Folksonomies: Search and Ranking. Volume 4011 of Lecture Notes in Computer Science. Springer, Berlin/Heidelberg (2006) 411–426

Interlinking Multimedia – Principles and Requirements

Tobias Bürger[1], Michael Hausenblas[2]

[1] Semantic Technology Institute, STI Innsbruck, University of Innsbruck, 6020 Innsbruck, Austria,
tobias.buerger@sti2.at

[2] Institute of Information Systems & Information Management, JOANNEUM RESEARCH, 8010 Graz, Austria,
michael.hausenblas@joanneum.at

Abstract. The linked data principles have gained a huge momentum by providing means to interlink datasets and by that contributing to a rich user experience on the Web. Methods to interlink data however still do not cover multimedia content in a sufficient way, as *Interlinking Multimedia* requires more than just putting resources globally in relation to each other. In order to close this gap, we propose a set of principles and requirements for interlinking multimedia content on the Web. As one major source we have identified user interaction to establish static or dynamic links between (parts of) multimedia resources.

1 Introduction

In early 2007, the W3C launched the Linking Open Data (LOD) community project[3] whose goal is to bootstrap the Semantic Web by publishing datasets using RDF and to publish and interlink open data on the Semantic Web. This is either done by using already existing sets of open data or by creating new linked datasets. The LOD project currently includes over 30 different datasets: From one billion triples and 250k links in mid-2007 the LOD dataset has grown to over two billion triples and 3 million links in early 2008, representing a steadily growing, open implementation of the Linked Data principles[4].

Several approaches exist for semantically linking data: RDF links can either be set manually or generated by automated linking algorithms for large datasets [1]. Advanced approaches such as those described in [1] make use of extended literal lookups or graph matching algorithms which are used to disambiguate similar matches.

Recent development in the linked data community is well documented by the proceedings of the Linked Data on the Web workshop (LDOW2008) [2], and submissions received by the Triplification challenge[5] including a proposal for "User Contributed Interlinking" [3] (UCI) of multimedia content [11].

What however can be observed is, that recent approaches for linking data mainly focused on the automated integration of textual resources and the interlinking of resources

[3] http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

[4] http://esw.w3.org/topic/LinkedData

[5] http://triplify.org/Challenge

as a whole. However, referring to Sir Tim Berners-Lee *the next generation Web should not be based on the false assumption that text is predominant [...] The Web is a multimedia environment, which makes for complex semantics* [4]. This fact has to be taken into account when we think about future directions for linked data. The envisioned situation is a *Interlinked Multimedia Web* in which objects or sequences in multimedia resources are linked to each other based on their semantic relationships.

Only recently Web 2.0 based applications emerged, in which image metadata is augmented by user generated tags. However, the possibility to set typed links between resources, or objects that are part of these resources, is still immature. YouTube launched a first facility[6] to annotate parts of videos spatio-temporally and to link to particular time points in videos[7] which is a promising start. However typed links between fragments can not be established.

The contribution of this paper is an analysis of the envisioned situation and a proposal of a set of requirements for *Interlinking Multimedia* which has only recently been discussed in the linked data community[8]. Furthermore we propose a set of principles based on the semantics of multimedia content and the interaction with multimedia content that can be used to interlink multimedia resources on a semantic level (section 2). We especially identify intended and monitored user interaction as a source for high quality links (section 3).

2 Interlinking Multimedia – Principles and Requirements

The interlinking of resources and parts of it, shares similarities with Hypermedia research: A hypermedia document such as defined in [5] refers to a collection of information units including information about synchronization between these units and about references between them. Typically temporal and a spatial dimensions are included, whereas references can be made between parts in both dimensions. *Interlinking Multimedia* is not an attempt to resurrect hypermedia but rather a light-weight, bottom-up approach to interlink multimedia content on the Web.

As only recently demonstrated by the BBC[9], interlinking of music related information, which may be publicly available on the Web or in closed archives, can significantly contribute to an enhanced end user experience. Moreover, as summarized in [6], there is a demand in several other communities for annotation tools to specify links between whole objects or segments within these objects and the typing of these links or relationships: Not only media researchers that want to relate and annotate segments between books, or screenplays and different films or film versions demand for facilities to interrelate rich content [6]. In order to realise the envisioned situation in which multimedia resources are semantically interlinked on a fine-granular level, one should take the following principles and requirements into account:

[6] http://youtube.com/watch?v=UxnopxbOdic
[7] http://www.techcrunch.com/2008/10/25/youtube-enables-deep-linking-within-videos/
[8] http://community.linkeddata.org/MediaWiki/index.php?InterlinkingMultimedia
[9] http://www.bbc.co.uk/music/beta

1. In order to be become part of the LOD cloud, ***Interlinking Multimedia* must follow the linked data principles** [7]:
   (a) All items should be identified using URIs;
   (b) All URIs should be dereferenceable and it should be possible to lookup the identified items using HTTP;
   (c) When looking up an URI, that is, an RDF property is interpreted as a hyperlink, it leads to more data;
   (d) Links to other URIs should be included in order to enable the discovery of more data.
2. Solutions should take into account the characteristics of multimedia whose semantics – when watched by a user – are typically derived based on the experiences and background of a human being. Thus **solutions should consider provenance information**; who says what and when is an important contextual aspect to represent the semantics of content (even if statements or references were created by machines).
3. **Metadata descriptions have to be interoperable** in order to reference and integrate parts of the described resources. This issues are discussed in [8], addressed by recent proposals like ramm.x[10] and by the W3C Media Annotations Working Group[11].
4. As discussed in [9] recently, **localizing and identifying fragments** is essential in order to link parts of resources with each other. It is essential to provide means to mark up spatial or temporal fragments, then to provide a mechanism to specify URIs for those fragments and finally to draw links between those fragments. This issue is particularily researched in the recently started W3C Fragments Working Group[12].
5. Furthermore **Interlinking Methods**, which we discuss in section 3, are essential in order to manually or (semi-) automatically interlink multimedia resources.

3 Interlinking Methods

Due to the inherent characteristics of multimedia content, the implementation of interlinking methods is far from being trivial. This is mainly due to the Semantic Gap, i.e. the large gulf between the low-level semantics which are derivable by machines and the high level semantics a user is typically interested in. This gap significantly hinders automation in the establishment of high quality links. As only little work is available at time of writing, we propose a set of interlinking methods that could close this gap:

Automatic Interlinking (AI) can be applied in situations in which quality metadata information is available that can be used to identify objects and their semantics. While AI methods[13] have demonstrated to yield fair results for global, textual resources [13], for fine-grained interlinking of multimedia content we doubt that this is the preferred path to follow.

[10] http://sw.joanneum.at/rammx/
[11] http://www.w3.org/2008/WebVideo/Annotations/
[12] http://www.w3.org/2008/WebVideo/Fragments/
[13] http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/EquivalenceMining

Emergent Interlinking (EI) can be based on the principles of Emergent Semantics whose underlying principle is to discover semantics through observing how multimedia information is used [12]. This can be essentially accomplished by putting multimedia resources in context-rich environments being able to monitor the user and his behavior. In these environments, two different types of context are present: (i) static or structural context, which is derived from the way how the content is placed in the environment (e.g. a Web page) and (ii) dynamic context, which is derived from the interactions of the user in the environment (e.g. his browsing behavior, which links he follows, or on which object he zooms). As stated in [12], in appropriate environments, the browsing path of a user is semantically coherent and thus allows to derive links between objects which are semantically close to each other.

User Contributed Interlinking (UCI) – a term which has been coined in [3] – is an approach for manually creating high-quality interlinks. The advantage of the application of UCI for the interlinking of multimedia is that it is based on end users as sources of qualitative information. First steps have already been made for UCI-based interlinking methods, such as available in the still image concept demonstrator CaMiCatzee [11] or in Henry[14] (for interlinking temporal audio fragments).

Game Based Interlinking (GBI) can be based on the principles set forward by Louis van Ahn with his *games with a purpose*[15] [14]. By that, interlinking of resources or parts of these resources could be hidden behind games. This is related to UCI but with the main difference that the user is not aware of him contributing links, e.g. his task is hidden behind a game. GBI seems to be a promising direction for multimedia interlinking. The most interesting examples to build on are Louis van Ahn's *ESP Game* in which users are asked to describe images or *Squigl*[16] in which users are asked to trace objects in pictures. Another interesting approach is followed by *OntoGame* [15] whose general aim is to find shared conceptualizations of a domain. During the game, players are asked to describe images, audio or video files. Users are awarded if they describe content in the same way. These approaches together with appropriate browsing interfaces for multimedia objects could be a promising starting point to let users draw meaningful relations between objects and parts thereof.

The methods can be arranged in a three-dimensional matrix with the dimensions time, quality and amount of annotations as depicted in Figure 1: While UCI might reach the highest quality and needs the highest amount of time from an end user perspective, automatic interlinking might produce the greatest amount of annotation and thus links with the least amount of time and manual effort needed.

4 Conclusion and Further Challenges

In this paper we discussed a future direction for linked data and pointed out to several issues with respect to *Interlinking Multimedia*. Besides the requirements that we

[14] http://dbtune.org/henry/
[15] http://www.gwap.com/
[16] http://www.gwap.com/gwap/gamesPreview/squigl/

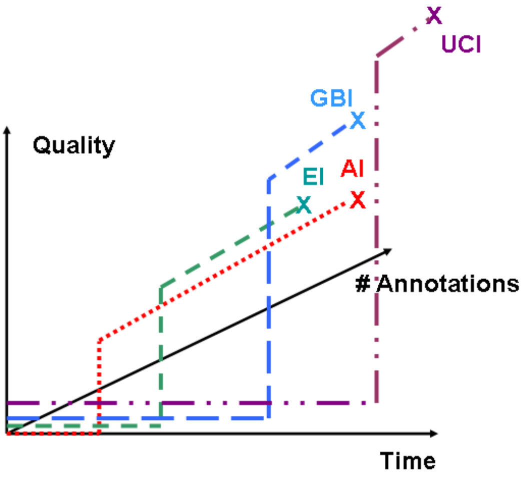


Fig. 1. Multimedia Interlinking – methods

formulated in section 2, a few other challenges have to be faced. These include generally applicable challenges for LOD like *Discovery & Usage* which has recently been addressed with voiD, the vocabulary of interlinked datasets[17], *Performance & Scalability* or *Privacy & Trust* which is addressed in another position paper for this workshop. We particularly identify user interaction is an essential ingredient to address a fourth challenge: *Quality* of links.

We believe that with the realization of the *Interlinking Multimedia* – principles a further step can be taken to a truly rich experience on the *Web of Data*.

Acknowledgements: The research leading to this paper was partially supported by the European Commission under contract IST-FP6-027122 “SALERO”.

References

[1] Raimond, Y., Sutton, C., and Sandler, C.: “Automatic Interlinking of Music Datasets on the Semantic Web” In: Proceedings of Linked Data on the Web (LDOW2008), Beijing, China, 2008.

[2] Bizer, C., Heath, T., Idehen, K., and Berners-Lee, T. (eds.): “Proceedings of the Linked Data on the Web Workshop”, Beijing, China, April 22, 2008, CEUR Workshop Proceedings, ISSN 1613-0073, online CEUR-WS.org/Vol-369/

[3] Halb, W., Raimond, Y., and Hausenblas, M. “Building Linked Data For Both Humans and Machines” In: Proceedings of Linked Data on the Web (LDOW2008), Beijing, China, 2008.

[4] Berners-Lee, T., Hall, W., Hendler, J. A., O’Hara, K., Shadbolt, N. and Weitzner, D. J. “A Framework forWeb Science” In: Foundations and Trends in Web Science, Vol. 1, No 1, 2006, pp. 1130

[17] http://community.linkeddata.org/MediaWiki/index.php?VoiD

[5] Hardman, L. "Modelling and Authoring Hypermedia Documents" PhD thesis, CWI, Amsterdam, 2004. http://homepages.cwi.nl/ lynda/thesis/

[6] Schroeter, R. and Hunter, J. "Annotating Relationships between Multiple Mixed-media Digital Objects by Extending Annotea" In: Proceedings of the European Semantic Web Conference (ESWC2007), 2007.

[7] Berners-Lee, T. "Linked Data", Design Issue Note, 27.07.2006, online http://www.w3.org/DesignIssues/LinkedData.html

[8] Tzouvaras, V., Troncy, R., Pan, J. Z. (eds.) "Multimedia Annotation Interoperability Framework" W3C Incubator Group Editor's Draft, 14 August 2007. http://www.w3.org/2005/Incubator/mmsem/XGR-interoperability/

[9] Troncy, R., Hardman, L., van Ossenbruggen, J., and Hausenblas, M. "Position Paper on Identifying Spatial and Temporal Media Fragments on the Web". W3C Video on the Web Workshop. Dec 2007.

[10] Hausenblas, M., Bailer, W., Bürger, T., Troncy, R. "Deploying Multimedia Metadata on the Semantic Web" In: Posterproceedings of the 2nd International Conference on Semantics And digital Media Technologies (SAMT 07), 2007.

[11] Hausenblas, M., Halb, W. "Interlinking Multimedia Data" Linking Open Data Triplification Challenge at the International Conference on Semantic Systems (I-Semantics08) at TRIPLE-I, September 2008.

[12] Grosky, W. I., Sreenath, D. V., and Fotouhi, F. "Emergent Semantics and the Multimedia Semantic Web". SIGMOD Rec. 31, 4 (Dec. 2002), 54-58.

[13] Hausenblas, M., Halb, W., Raimond, Y., and Heath, T. "What is the Size of the Semantic Web?". I-Semantics 2008: International Conference on Semantic Systems, 2008.

[14] von Ahn, L. "Games with a Purpose'". IEEE Computer 39(6): 92–94 (2006)

[15] Siorpaes, K. and Hepp, M. "Games with a Purpose for the Semantic Web". IEEE Intelligent Systems 23(3): 50–60 (2008)

ImageNotion as a Mashup Service for a Semantic Image Web*

Andreas Walter[1] and Gábor Nagypál[2]

[1] FZI Research Center for Information Technologies , Information Process Engineering, Haid-und-Neu-Straße 10-14, 76131 Karlsruhe, Germany
Andreas.Walter@fzi.de

[2] disy Informationssysteme GmbH, Erbprinzenstr. 4-12, Eingang B, 76133 Karlsruhe, Germany
nagypal@disy.net

Abstract. Most web image archives still use plain text to annotate images. The ImageNotion system ([1] [2]) extends the state-of-the-art by providing semantic annotation of images and their parts. In this paper, we show how to implement a mashup of ImageNotion with popular web image archives such as Flickr. This allows our users to load the images from external image platforms in ImageNotion and to automatically create semantic annotations using ImageNotion. Further, they can use the advanced search features of ImageNotion on those images, including the search for related images. In addition, we also show how to extend this mashup to a semantic web service. Such a service creates semantic image annotations semi-automatically, and thus makes those images available for processing on the Semantic Web.

1 Introduction

State-of-the-art image archives use textual annotations. Those textual annotations may be image caption, description or a set of simple tags, as in Web 2.0 systems. For users of such systems, using semantic image annotations instead of textual ones would provide many advantages. For example, ontologies and also thesauri make it possible to separate homonymous meanings of a tag into different ontology elements, such as "Paris" as a person (Paris Hilton) and "Paris" as a city. Additionally, it is also possible to propose semantically related images in an image archive by using "narrower/broader term" relations of thesauri, such as done in Riya [3]; or "subclass-of" and "is-a" relations of ontologies, such as done in SemSpace [4].

Further, users can explore the contents of an image archive instead of formulating precise queries. They can navigate to related images, following links to topics, persons and events in the context of their current search request. E.g., they may start searching for all "portraits of actors from Germany", refine their search to "images from female actors in the movie Barfuss" (e.g. *Johanna Wokalek* and *Alexandra Neldel*), and finally to "images of the main actors of Barfuss" (*Johanna Wokalek* and *Til Schweiger*) just by following links among ontology elements that are attached to the images or to their parts.

* This work was co-funded by the European Commission within the project IMAGINATION.

Semantic image annotations also allow automated systems, e.g. agents, to query for images, and to "understand" the displayed image contents [5]. E.g., agents can search for images containing persons, and they can understand who are those people based using the semantic image annotations.

Since state-of-the-art systems use textual image annotations, a new system is required that is capable of creating semantic image annotations, of transforming textual annotations to semantic annotations, and of creating semantic annotations on image parts. The ImageNotion application [1] provides a visual methodology that supports collaborative, work-integrated ontology development. ImageNotion allows for semantic search and for the navigation through image archives based on the available image annotations.

In most information systems, it is clear that semantic technologies would help the system to provide a better end-user experience. As the creation of semantic annotation is very time consuming, using only manual annotation methods often make it impossible to employ such technologies in real-world systems. To overcome this problem in ImageNotion, we automatize the generation of semantic annotations to the maximum possible extent.

Users have already uploaded many images to many different image platforms, e.g. to Flickr or Facebook. ImageNotion should be able to use these images directly from their sources. In this paper, we propose a system architecture to use ImageNotion as a mashup service. The service should read images from image sources available on the Web, and should automatically create semantic image annotations for them. This mashup then may work as a middleware between the world of textual image annotations on the one hand, and between the world of Semantic Web relying on semantic annotations on the other hand.

Based on this mashup, we also show how ImageNotion may run as a Semantic Web service for images. This Semantic Web service creates image annotations that use popular ontologies [6] and thus provides semantic interoperability with other image processing applications on the Semantic Web.

The paper is structured as followed: in Section 2, we discuss related work. Section 3 gives an overview about the features of ImageNotion. These features include collaborative and work integrated ontology development, the combination of the results using automated processes to improve the quality of generated metadata, and using the semantic annotations during semantic search. In Section 4, the mashup system based on ImageNotion is introduced. This mashup is extended to a Semantic Web service in Section 5. Section 6 finally concludes and gives some outlook.

2 Related work

In this section, we report on related work that is relevant to our goal to extend ImageNotion to a mashup service and to provide a Semantic Web service for this mashup.

Data integration from image platforms: For our mashup, the integration of image metadata from image platforms is necessary. Image platforms include portals (such as the German portal fotomarktplatz[3] for professional photographers and image agencies),

[3] www.fotomarktplatz.de

images displayed on web pages (e.g., crawled by Google), photo sharing platforms (e.g., MySpace, Flickr and Riya) and social network sites (e.g. Facebook, studiVZ or LinkedIn). We give an overview on possible data integration techniques for these image platforms.

One possibility is to use the OpenSocial API[7]. It defines a set of commonly used, standardized methods for social network sites. Thus, it allows for interchanging and linking among others profile data, friend lists and even images from various sources supporting this standard. E.g. images from MySpace [8] can be accessed this way. Currently, for most sites, e.g. Flickr, Facebook and Riya, OpenSocial is not supported but proprietary APIs are offered and proprietary data exchange formats are provided. For such services, a wrapper is needed. The APIs allow either to retrieve all publicly available images or images of a given user where the the user agreed exchanging data with the service using the API. Via the APIs, it is possible to read the existing image annotations and also to retrieve the images themselves. Image platforms, such as Riya or Facebook, that support the annotation of image parts, also support the retrieval of image part annotations [9].

Integrating automated processes: Creating image annotations, and especially annotations for image parts, is very time consuming. This process should be automatized to as much as possible. We give an overview how automated processes are used in other image platforms. Tag4you[4] uses the Flickr API to allow face detection in Flickr images. The system automatically marks the areas of detected faces and also allows adding tags to those areas manually. The tags are then automatically written back to Flickr. Riya offers face detection and recognition algorithms. Text recognition based on OCR in images is also provided. The ImageSorter application [10] sorts images by similarity of colors and structures and allows for searching for "similar images by structure or color". The face detection and recognition algorithm of Fraunhofer IIS [11] (we use it in ImageNotion) also supports gender classification and can detect moods like happy, angry or surprised.

Semantic image annotation: To create semantic image annotations, users have to be supported by adequate tools. RDFPic [12] and PhotoStuff [4] allow for the generation of semantic image annotations using RDF and imported domain ontologies. Both applications are only available as desktop applications and offer no support for collaborative ontology development and semantic annotation.

Standardized ontologies for semantic interoperability on the Semantic Web: Semantic Web services must use standardized ontologies and multimedia standards so that they can provide semantic description of the images that can be interpreted in different systems [6,5]. The internal ontology of the system therefore should be mapped to commonly used domain ontologies [6] describing people, objects, events and their relations. FOAF [13] is an ontology for the Semantic Web to describe people, their activities and their relations to other people and objects. In our case, this ontology may be used to exchange semantic annotations of people. In a similar way, the CYC ontology [14] can be used to describe objects and CIDOC-CRM is a standard ontology for describing cultural heritage objects. Using these ontologies, it is possible to map the ontology internally used in ImageNotion to these standards, as long as adequate concepts

[4] www.tag4you.com

are available in the internal ontology. Otherwise, the mapping is only possible with core concepts of these ontologies [15].

MPEG-7 [16] can describe the structure of an image, including the contours of image parts. In addition, it is also possible to include semantic annotations of image parts or of the whole image. The annotations may be defined in an arbitrary ontology, such as FOAF or CYC. MPEG-7 is therefore suitable for exchanging semantic metadata between multimedia information systems.

3 The ImageNotion application

In this section, we give a brief overview on the ImageNotion application. For further details please refer to [1,2,17]. The ImageNotion application is publicly available at `www.imagenotion.com`.

3.1 Collaborative and work-integrated ontology development

We call our ontology elements (concepts and instances) *imagenotions*, formed from the words image and notion. An imagenotion visually represents an ontology element through corresponding images. The visual representation of ontology elements helps image annotators to get a better understanding of their meaning. Based on the ontology maturing process model [18], a collaborative and work integrated ontology development methodology is implemented in ImageNotion. New imagenotions may be added by users in the first phase. In the next step, imagenotions are consolidated in communities of users (in Fig. 1, the annotators add descriptive information for the current EU president Manuel Barroso). In this phase, a stable definition of the concept emerges as users communicate with each other, or work on the same concept definition. In the third phase, it is possible to add relations between imagenotions (see Fig. 2, where annotators added relations to “president”, “male” and “EU commission”). Imagenotions from each maturing grade may be used for semantic image annotations immediately after creation.

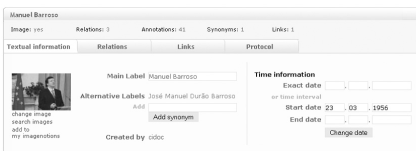


Fig. 1. Adding descriptive information to imagenotions

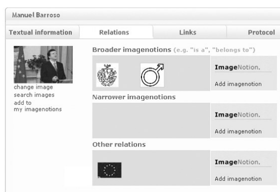


Fig. 2. Adding relations to imagenotions

The ImageNotion methodology allows to start the development of a new ontology either from scratch or by reusing existing ontologies (such as CIDOC-CRM or CYC) or parts of these. A community may then collaboratively add further ontology elements. For example in the EU project IMAGINATION, the group of image annotators decided to start with core parts of the multimedia ontology CIDOC-CRM and to use only the relations broader, narrower and unnamed relations from SKOS [19].

3.2 Combining automated processes in ImageNotion

For the automatic generation of semantic annotations, we combine the results of face detection and recognition (from Fraunhofer IIS, [11]), object and person detection (from NTUA, [20]) and text mining algorithms (e.g. text classification of JSI [21]). The generated semantic annotations use the existing imagenotions in the ontology.

Example: Fig. 3 shows the result of applying the automated processes in ImageNotion for a given image. The text mining algorithm has created the semantic annotations "Romano Prodi" and "Manuel Barroso", based on the textual title of the image: "EU president Barroso meets Prodi". The person and object detection algorithms have created bounding boxes for the shapes of the two persons. The face detection algorithm has created two bounding boxes for the detected faces. The face recognition has identified "Manuel Barroso" with a score of 80 percent and "Jan Figel" with a score of 20 percent for Barroso's face. Romano Prodi's face was identified as male person, but was not recognized by the face recognition algorithm.

The controller now combines the results (see Fig. 4). For the second face, it creates a new image annotation "Romano Prodi" for Prodi's face with a score of 100 percent and sets the score of "Male" to zero. This is possible because the annotations created by text mining state that there must be the persons "Romano Prodi" and "Manuel Barroso" on the image, and the other region was already correctly recognized as Manuel Barroso. Also, it sets the score of the detected areas for the persons to zero, since faces inside those areas were detected, and our end users prefer annotating faces to annotating person bodies.

Training of the automated processes Some of the automated processes needs training to provide acceptable results. This is the most challenging for the face recognition algorithm where each person that should be recognized by the algorithm, needs training images. In ImageNotion, we embed the specification of training images into the process of creating new imagenotions and new semantic annotations as follows:

1. A user uploads images. One or more faces on these images are unknown for the system or although the faces are known, they cannot be recognized correctly for some reasons. The face detection algorithm that operates very reliable, determines the bounding boxes also for such faces. The system notices that the face recognition algorithm failed to recognize the person because the recognition score is too low.
2. The web user interface asks, who the person is. The user may associate the bounding box with an existing imagenotion of a person or create a new imagenotion.
3. If a new imagenotion is added, gender detection is additionally executed the user is asked whether the detected gender is correct. Finally, relations to the imagenotions "gender" and "person" are added automatically.
4. The image part showing the person is added to the training images of her imagenotion.

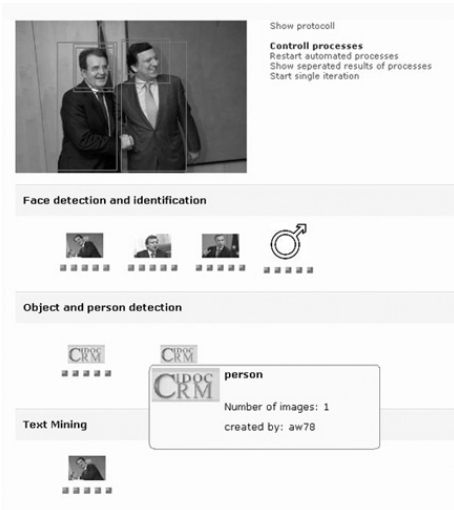


Fig. 3. Results of the automated processes

Fig. 4. Overall result

3.3 Visualization of semantic annotations and semantic search

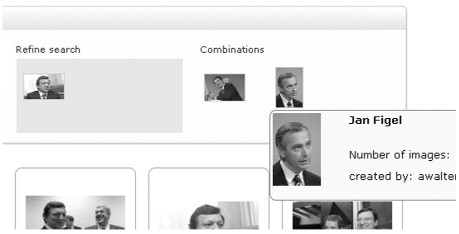


Fig. 5. Visual search refinment

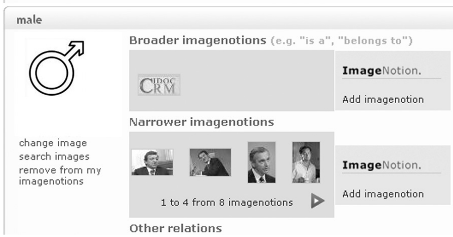


Fig. 6. Browsing the ontology

The manually or automatically created semantic image annotations can be used in ImageNotion for the navigation through the image archive. Here, we give a short overview of the already implemented techniques. For details, we refer to [2]. A user may start with a full text search and visually choose her desired imagenotions to start a semantic search. When a user clicks on an image part, the user may choose between starting a new search, refining an existing search or viewing the details of the imagenotion used for the semantic image annotation.

Fig. 5 shows the image clustering technique in our system. After a semantic search, annotations of the resulting images are analyzed, and the imagenotions that are most frequently used together with the imagenotion(s) forming the semantic search are grouped together in a cluster, e.g. all images displaying “Manuel Prodi” together with “Jan

Figel". Fig. 6 shows an example scenario of our ontology browser. By searching for "male", a user can browse all existing imagenotions for male people to start new semantic search requests.

4 Building mashups with ImageNotion

The ImageNotion application currently stores all images, ontologies and semantic annotations in its local database. Our aim is to integrate images from external image sources by building a mashup system. Then, it would be possible to use the ImageNotion application for semantic search and visualization of the semantic annotations for these images. In addition, we can provide an API to read the generated annotations. In this case, the integration of ImageNotion in existing applications, e.g., in image search engines, is possible. Those applications may then benefit from the semantic annotations generated by ImageNotion.

4.1 System architecture

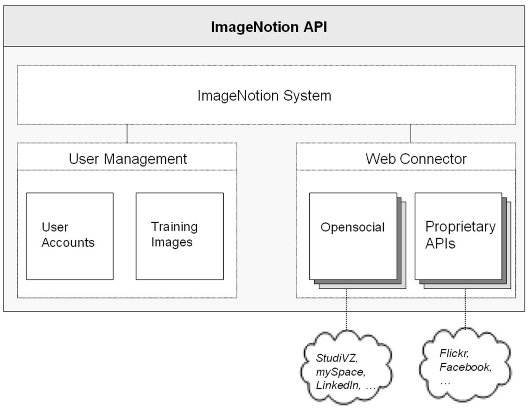


Fig. 7. System architecture for building mashups with ImageNotion

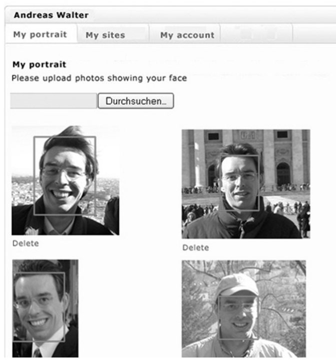


Fig. 8. User inteface of the mashup sytem for the user Andreas Walter

Fig. 7 shows the system architecture for building ImageNotion as mashup system. It consists of a user management module for storing user related information and a web connector module for connecting to the image sources.

User management module: This module stores the information of users accounts in external image platforms in the *user account* component. Consequently, the mashup application can load all images of a user and create semantic image annotations for them. Because face recognition requires training data (see Section 3.2), a user may train some faces using the procedure introduced before. The training images are stored in the *training images* component and passed to the face detection and recognition algorithms as training data.

Web connector module: The web connector module is responsible for loading all images from the desired image platforms. Each platform with a proprietary API requires a wrapper. Each wrapper connects to the corresponding image platform, e.g. Flickr, to load the images. With the *OpenSocial* component, we could load images from all social network sites which implement this standard, such as MySpace.

Subscription for new images: To load new images, the mashup system can be configured in two different ways. In the *user defined* setup, the system loads only images from the image platforms for which a user provided account information. In the *public images* setup, the system regularly polls new images in the connected image platforms (e.g., by searching for "all uploaded images in the last 10 minutes").

Creating semantic image annotations: The *web connector module* loads images from the image platforms in preview size (e.g. 800 pixel picture size), creates a unique identifier (currently the source name, e.g. Flickr, combined with the image name in the source) for the image, and passes the image object together with all available textual annotations or annotations for image parts to the ImageNotion system. The ImageNotion application generates semantic annotations for the image. The image annotations are stored in the ImageNotion system and are linked to the unique identifier of the image.

4.2 Prototype implementation

Our current prototype implementation allows one to read images from Flickr and Riya. We use the *user defined* setup for the subscription of new images. A user may define her accounts to these two sites and the web connector retrieves the data of her images (see Fig. 8). The ImageNotion system then creates semantic image annotations for these images.

4.3 Integration of the mashup system in other applications

By providing an API, we make it possible to extend existing applications with the features of the ImageNotion application. E.g., one may improve an existing web portal of images by embedding the semantic search feature of ImageNotion into the portal. ImageNotion may even be embedded in the case when the portal accesses multiple independent image sources.

The ImageNotion API is currently under construction. We will offer methods for generating semantic annotations, training new faces, adding new images and performing search requests.

5 ImageNotion as a Semantic Web service

In this section, we publish the ImageNotion mashup system as a Semantic Web service. By doing that we can fulfill the following goals. First, the Semantic Web service can automatically generate semantic image annotations for any image on the Web. The service gets an URL of an image in the service request, creates the semantic annotations, and finally returns these annotations in the service response. Second, the already introduced mashup service that accesses and integrates multiple image archives can be globally published as a semantic search service. In this section, we introduce a suitable system architecture for the ImageNotion Semantic Web service. We also discuss a possible solution to achieve semantical interoperability between ImageNotion and other Semantic Web systems.

5.1 System architecture

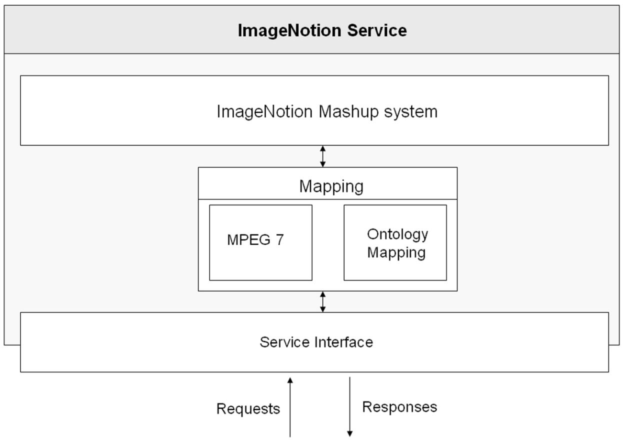


Fig. 9. System architecture for ImageNotion as Semantic Web service

Fig. 9 shows a possible system architecture for the ImageNotion Semantic Web service. The system architecture is based on the mashup architecture introduced in the last section. The architecture is extended with two new components, one for ontology mapping purposes and the other to provide the web service interface.

Service interface: The service interface offers methods for starting search requests and for initiating the annotation of new images. For example the service method *doImageAnnotation* receives as request the URL of an image and forwards this information to the *ImageNotion mashup* component. It creates the image annotations and returns them to the service interfaces. The service interface then returns the generated annotations as its response.

Ontology mapping: In the Semantic Web, standardized ontologies help exchange information between different applications. Therefore, our service should map the internally used ontologies to popular standard ontologies. FOAF is a popular format to describe information about people and thus may be used in our system. Among others, FOAF can also represent gender information. Objects and events could be mapped to core concepts of CIDOC-CRM (as it is done e.g. in the IMAGINATION[5] project).

[5] `http://www.imagination-project.org`

Alternatively, a mapping to matching elements or to the core concepts of the CYC knowledge base is also possible.

Semantic interoperability with other applications: MPEG-7 provides a standardized way to describe multimedia objects and their parts. The *MPEG-7* component receives the information about image regions from the ImageNotion system and it creates the MPEG-7 description of these image regions. The component receives the semantic image annotation mapped to FOAF and to CIDOC-CRM or CYC from the *ontology mapping* component.

5.2 Example use of the service for semantic image annotation

We now give an example how the ImageNotion Semantic Web may be used to create semantic annotations for an image by using the service method *doImageAnnotation*. The method is invoked by a simple GET request, the URL of the image is passed as a standard URI parameter. In our example, the image that is referenced by the URI is shown in Fig. 10.

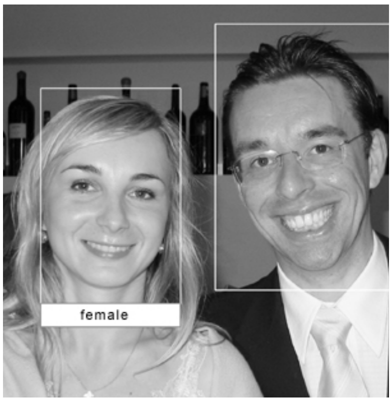


Fig. 10. Example image for the service


```
<Mpeg7>
 <MultimediaContent xsi:type="ImageType">
  <Image>
   <MediaLocator><MediaUri>
    URL of image
   </MediaUri></MediaLocator>
   <SpatialDecomposition>
   <StillImage>
   <StillRegion>
    <Region>(10,10) (100,100)</Region>
     <Semantic>
      <foaf:Person xmlns:foaf=
      "http://xmlns.com/foaf/0.1/"
       xmlns:rdf="http://www.w3.org/1999/02/22-rdf-
       syntax-ns#" rdf:about=
       "http://www.imagenotion.com/foaf/andreaswalter" />
      </Semantic>
     </StillRegion>
     <StillRegion>
      <Region>(140,10) (240,9)</Region>
      <Semantic>
       <foaf:Person xmlns:foaf="http://xmlns.com/foaf/0.1/">
        <foaf:gender>female</foaf:gender>
       </foaf:Person>
       </Semantic>
      </StillRegion>
    </StillImage>
    </SpatialDecomposition>
   </Image>
  </MultimediaContent>
 </Mpeg7>
```


Fig. 11. Service response in MPEG-7 format

Fig. 11 shows the service response to the request in MPEG-7 format. The result in this example contains two semantic image annotations. For the first annotation box, the person “Andreas Walter” was recognized correctly and is mapped to FOAF. For the second annotation box, the face detection has detected a female person but she could not be recognized.

6 Conclusions and future work

State of the art image repositories still use textual image annotations. However, image services on the Semantic Web must be able to provide semantic image annotations instead of textual ones so that their content is processable by automatic agents. ImageNotion helps bridge the gap between text based systems and the Semantic Web by automatically transforming textual image annotation to semantic ones. ImageNotion can automatically generate semantic image annotations using a combination of face detection and recognition, person and object detection, and text mining algorithms. Moreover, the collaborative manual development of semantic annotations is also supported.

In this paper, we have first proposed a mashup system based on ImageNotion. This mashup extends ImageNotion with the ability to use images from external sources, e.g., from popular image platforms like Riya, Flickr or Facebook. In such a system, users can benefit from getting all their images annotated with semantic annotations in ImageNotion, without the need to upload them again.

In the second step, we proposed a system architecture to publish ImageNotion as a Semantic Web service. Such a service may support various scenarios. First, it is possible to send an image URL to the service and get the semantic image annotations for the image. In addition, it is also possible to install a mashup service over different image platforms and to offer semantic search functionalities via this service. To be interoperable with other services and agents on the Semantic Web, it is important to use established Semantic Web standards. Therefore, we proposed to use FOAF for the description of people and CIDOC-CRM or CYC for describing objects, places and events. In addition, we propose to use the MPEG-7 standard for structural description of images. An important feature of MPEG-7 that it supports the description of image regions and their semantic annotations.

Currently, the mashup system is only implemented as a small prototype and the Semantic Web service layer is not yet implemented. In our current and future work, we will connect our system to a higher number of image platforms, we will implement the API and will add the Semantic Web service layer. In addition, we are already working on a mashup which uses the ImageNotion API. In the PRIMO system [22], this API is used to build a privacy-aware Web 2.0 image sharing application.

References

1. Walter, A., Nagypal, G.: Imagenotion – methodology, tool support and evaluation. In: GADA/DOA/CoopIS/ODBASE 2007 Confederated International Conferences DOA, CoopIS and ODBASE, Proceedings, LNCS. Springer. (2007)
2. Walter, A., Nagypal, G., Nagi, K.: Evaluating semantic techniques for the exploration of image archives on the example of the imagenotion system. In: Alexandria Engineering Journal (AEJ), Vol. 47, No. 4. ISSN: 1110-0168 Springer. (2008)
3. Riya: Riya - Visual search. http://www.riya.com/ (2007) (accessed 2007-12-09).
4. van Ossenbruggen, J., Troncy, R., Stamou, G., Pan, J.Z.: Image Annotation on the Semantic Web. W3C working draft, W3C (2006)
5. Bashir, A., Khan, L.: A framework for image annotation using semantic web. In: MSW2004, located at the 10th International ACM SIGKDD Conference on Knowledge Discovery and Data Mining KDD 2004, 22nd August 2004 - Seattle, WA, USA. (2004)

6. Stamou, G., van Ossenbruggen, J., Pan, J., Schreiber, G.: Multimedia annotations on the semantic web. In: IEEE Multimedia, 13(1):86–90, January-March 2006. (2006)
7. OpenSocial: Opensocial. http://code.google.com/apis/opensocial/ (2008) (accessed 2008-09-09).
8. MySpace: Myspace specific extensions on opensocial. http://tinyurl.com/ytpfgl (2008) (accessed 2008-09-09).
9. riyaApi: Riya API . http://www.riya.com/riyaAPI (2008) (accessed 2008-09-09).
10. ImageSorter: ImageSorter V3. http://mmk1.f4.fhtw-berlin.de/Projekte/ImageSorter (2008) (accessed 2008-09-09).
11. Küblbeck, C., Ernst, A.: Face detection and tracking in video sequences using the modified census transformation. In: Image and Vision Computing, Editor: K.D. Baker, Issue No. 6, Elsevier. (2006)
12. Lavon, Y., Boss, B.: Describing and retrieving photos using RDF and HTTP. W3C working draft, W3C (2006)
13. FOAF: The Friend of a Friend (FOAF) project. http://www.foaf-project.org/ (2008) (accessed 2008-09-09).
14. CYC: OpenCyc for the Semantic Web. http://sw.opencyc.org/ (2008) (accessed 2008-09-09).
15. Walter, A., Nagypál, G.: Semantical interoperability with IMAGINATION content using standardized ontologies. In: Workshop on Ontology-Driven Interoperability for Cultural Heritage Objects, Joint DELOS-MultiMatch event, Tirrenia, Pisa (Italy) (2007)
16. Martinez, J. M.: MPEG-7 Overview Version 10 . http://tinyurl.com/gcpfh (2005) (accessed 2008-09-09).
17. Walter, A., Nagypal, G.: The combination of techniques for automatic semantic image annotation generation in the imagination application. In: 5th European Semantic Web Conference, ESWC 2008, Tenerife, Canary Islands, Spain, June 1-5, 2008 Proceedings. (2008)
18. Braun, S., Schmidt, A., Walter, A., Nagypal, G., Zacharias, V.: Ontology maturing: a collaborative web 2.0 approach to ontology engineering. In: Proceedings of the Workshop on Social and Collaborative Construction of Structured Knowledge at the 16th International World Wide Web Conference (WWW 07), Banff, Canada. (2007)
19. Brickley, D., Miles, A.: SKOS Core Vocabulary Specification. W3C working draft, W3C (2005)
20. Kapsalas, P., Rapantzikos, K., Sofou, A., Avrithis, Y.: Regions of interest for accurate object detection. In: TProc. of Sixth International Workshop on Content-Based Multimedia Indexing (CBMI 2008), London, UK. (2008)
21. Blaz Fortuna, Marko Grobelnik, D.M.: Semi-automatic data-driven ontology construction system. In: Proc. of the 9th International multi-conference Information Society IS-2006, Ljubljana. (2006)
22. Burghardt, T., Walter, A., Buchmann, E., Böhm, K.: Primo - towards privacy aware image sharing. In: to appear in proceedings of CISWSN. (2008)

gFacet: A Browser for the Web of Data

Philipp Heim, Jürgen Ziegler and Steffen Lohmann

University of Duisburg-Essen
Lotharstr. 65, 47057 Duisburg, Germany
{Philipp.Heim, Juergen.Ziegler, Steffen.Lohmann}@uni-due.de

Abstract. This paper introduces a new approach to browsing the Web of data by combining graph-based visualization with faceted filtering techniques. The graph-based visualization of facets supports the integration of different domains and an efficient exploration of highly structured and interrelated datasets. It allows to access information from distant user-defined perspectives and thereby enables the exploration beyond the borders of Web pages.

Keywords: Web of Data, Graph-based Visualization, RDF Data, Semantic Web, Hierarchical Facets, Faceted Browsing.

1 Introduction

The Web is commonly thought of as a Web of documents. With the steady growth of the Semantic Web, this Web of documents is more and more transformed into a Web of data. Information is no longer stored as monolithic blocks in Web pages, but is fragmented into many pieces that can be assigned to ontological concepts. This goes along with a fragmentation of the linking structure, transforming the untyped links between Web pages into typed relationships between information pieces.

The information pieces along with the concepts linked to them and the typed relationships are stored in RDF [1], leading to large and highly interrelated RDF datasets. Several such datasets for different domains are already available on the Web; examples include DBpedia[1], FOAF profiles[2] or MusicBrainz[3]. The semantic annotation stored in RDF datasets allows Web content to be found, shared and combined more efficiently and hence fosters the integration of data from distributed sources. These developments require a new generation of applications that provide a single access point to information on the Web in order to facilitate the exploration beyond the borders of Web pages.

In this paper we introduce gFacet, a browsing approach that supports the exploration of the Web of data by combining graph-based visualization with faceted filtering functionalities. The graph-based visualization facilitates a comprehensible integration of different domains; the use of facets supports a controlled filtering of information.

[1] http://dbpedia.org
[2] http://foaf-project.org/
[3] http://musicbrainz.org/

With gFacet, users are enabled to browse the Web of data efficiently and to retrieve information from different user-defined perspectives.

The remainder of this paper is structured as follows. In section 2, we give an overview of existing approaches to visualize and browse RDF data. In section 3, we introduce our approach and describe a prototypical implementation. We proceed with an evaluation of the prototype in section 4 and close the paper with a conclusion and outlook on future work in section 5.

2 Overview of Existing Approaches

Tools to visualize and browse RDF data have been developed within several projects. We can classify these tools into two main groups, each using a different approach. The first group uses a graph-based approach that explicitly visualizes the structure of the RDF datasets by nodes and edges. The second group divides the screen into several areas, each containing a subset of the data, allowing for faceted browsing.

2.1 Graph-based Approach

The RDF syntax is based on triples in the form of subject-predicate-object expressions. An RDF graph is a set of such triples with nodes being subjects or objects and labeled edges being predicates [1]. Thus, using a visualization that directly presents this RDF graph structure to the users seems obvious and is suitable to show the structural complexity of the relations within the data. For instance, RDF Gravity [2], Welkin [3], IsaViz [4] and Paged Graph Visualization [5] are such tools that directly visualize the RDF graph.

Since RDF datasets are often large and highly interconnected [6], visualizing all the relations that exist within the graph structure can quickly cause a high number of edge crossings and hence hinder an understandable visualization of the data. In order to prevent the graph visualization from getting over-cluttered, different mechanisms to reduce the displayed information are provided. RDF Gravity, for example, applies local and global filter techniques to hide certain nodes and edges. The Paged Graph Visualization by contrast starts with only a small set of information and let the user gradually expand this set to explore specific parts and fragments of the graph.

Even though these filter mechanisms can help to reduce the displayed information, they are not appropriate to visualize complex interrelations within large datasets. The direct presentation of all the existing relations between entities only scales well to small sets of data with a limited number of interrelations [7]. The visualization of a large dataset with many interrelations on one screen will still tend to result in a complex graph, which is hardly manageable or understandable by the user [8].

2.2 Facet-based Approach

Another way to visualize large and complex RDF datasets is based on the concept of faceted exploration [9]. In faceted exploration, the data gets partitioned using orthogonal conceptual dimensions. One of the dimensions serves as the result set and the others are used as facets to filter the result set by different attributes that can be selected independently from each other.

Visual models of faceted exploration mostly apply an approach where the different facets and the result set filtered by the current facet selections are shown as lists in different screen areas. A popular example is Apple's music and media player iTunes [10] that uses the faceted filtering approach to let the user access music according to different attributes, such as the artist, album, or genre. The relation between the attribute values and the target objects is not directly visible but can only be recognized indirectly based on the filtering behavior exhibited by the system.

The faceted-browsing paradigm is especially applicable to large datasets with many interrelations. Several tools, such as mSpace [11], Longwell [12], /facet [13] or Haystack [14], use this way of browsing for the visualization and exploration of RDF data. The facets are automatically constructed and can be added and removed to the screen allowing the user to select facets that are of specific interest. However, Haystack, mSpace and Longwell are limited only to facets that are directly related to the result set. With the use of hierarchical facets users become empowered to access information from distant perspectives also.

Hierarchical Facets. Hierarchical facets are facets that are indirectly related to the result set. The simplest case of a hierarchical facet is a facet that has a direct relation to a facet that is in turn directly related to the result set. However, hierarchical facets can be indirectly related to the result set by more than one facet and hence allow to access the result set also from distant perspectives. A common way to visualize them is a tree that can be gradually expanded by the user. Both, Tabulator [15] and the Nested Faceted Browser [16] use trees to present hierarchical facets to the users.

If we extend the iTunes example by hierarchical facets, the filtering is no longer restricted to related facets such as artists, albums or genres, but can also include facets of facets such as the nationality of the artists. By providing hierarchical facets, users become enabled to access information from a broad range of different, user-defined perspectives.

3 Our Approach

The basic idea of our approach for the visualization and exploration of the Web of data is to combine the graph-based and the facet-based approach by arranging facets as nodes in a graph structure. With this specific combination of both visualization paradigms, we are able to counterbalance their respective disadvantages and facilitate the exploration of large and highly interrelated RDF datasets by the user. The major aims of our approach are:

1. **Prevention of an over-cluttered graph:** Using a facet-based visualization groups the instances contained in the Web of data into facets according to their conceptual structure. This prevents relations between single instances to be directly visualized by edges in the graph. Relations between these information pieces become only

indirectly visible when certain instances in a facet get selected by the user in order to filter the result set.

2. **Representation of relations between facets:** Using a graph-based visualization for the presentation of facets allows the relations between facets to be visualized as labeled edges. In doing so, the relations become explicit and dependencies as well as the filtering of the result set are more easily recognizable and traceable. This is especially useful for the visualization and understanding of hierarchical facets: Visualizing the dependencies and relations by labeled edges between facets can facilitate the understanding of how hierarchical facets can be used to find certain information.
3. **Single coherent visualization:** Using a graph-based visualization allows the information to be displayed in a single visualization instead of being spread over several screens or windows. Thus, information from distributed sources can be presented at once and hence reduce the cognitive load of the users and prevents them from getting "lost in hyperspace".

3.1 Prototype

We developed a first prototype to evaluate our proposed approach for graph-based faceted exploration of RDF data. The prototype is implemented in Flash and can be accessed via the Internet[4]. The RDF data is requested by the use of SPARQL[5], a protocol and RDF query language. In the following, the design of the visualization is demonstrated by an example from the field of music, where song information is represented as RDF statements.

Initially, the graph consists of a single list-valued node that can gradually be expanded by other nodes that serve as facets for filtering the list in the original node. In the upper left corner of Figure 1, the initial node contains a list of songs. All directly related facets are shown in a pull down menu below the list. If the user selects one of these relations, a new node is added to the graph and gets connected to the original node by a labeled edge. The label represents the predicate that connects the different types of information in the RDF data, in this case "partOf".

A force-directed layout algorithm [17] is performed that assigns spring-like forces to position the nodes of the graph so that all the edges are of more or less equal length and there are as few crossing edges as possible. The forces are iteratively applied to the nodes, pulling and pushing them until a stable state is reached. This allows the user to follow how the graph evolves and to understand how a new node or relation affects the graph layout.

The permanent movement of all nodes, however, is also a disadvantage because already found information can move to other locations and hence must be retrieved again by the user. Therefore, we introduce a pinning mechanism that forces already shown information to hold its position. When a new node is added to the graph, the pinning of this node is executed after a short period of time. This delay allows the force-directed algorithm to position new nodes in an appropriate way and at the same time prevents already existing information to change its location.

[4] http://www.gFacet.org

[5] http://www.w3.org/TR/rdf-sparql-query/

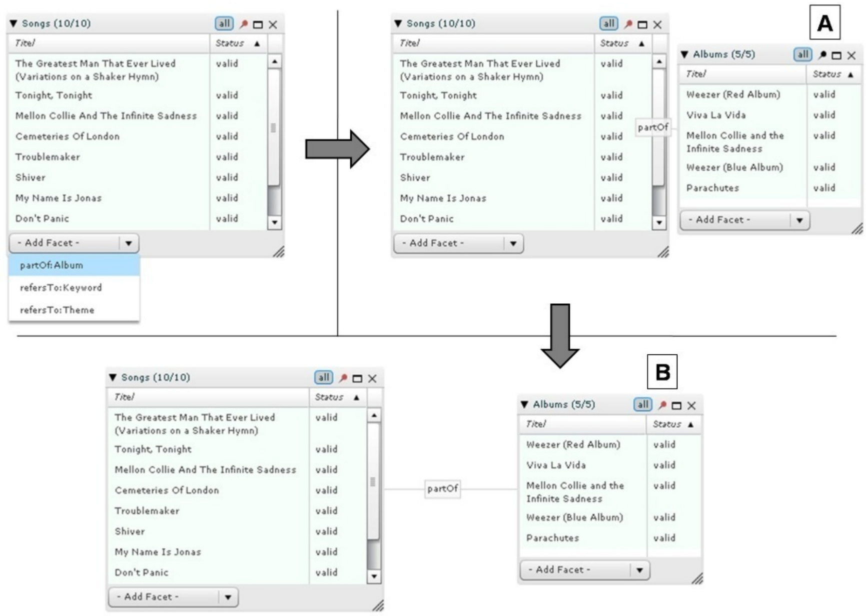


Fig. 1. The user adds new facets to the graph by selecting them from a menu below the list. Facets are automatically arranged by a force-directed layout.

In Figure 1, the new node can move away from the existing node until both do not overlap anymore, because of the delay in pinning its position. After the delay, the pinning is automatically executed and the node is fixed. The color of the needle symbol at the upper right corner of each node indicates whether a node is pinned or not. If the symbol is grey (Figure 1, A), the node is movable, if it is colored (Figure 1, B), the node is pinned. This pinning can also be controlled by the user by clicking at the needle symbol. That way, the user can decide whether a node should stay on a fixed position or should be rearranged by the force-directed algorithm in order to improve the overall appearance of the graph.

Graph-based Facets

The new node can be used as a facet to filter the connected nodes. By selecting a certain row in the list of albums, the list of songs gets filtered to only songs that are part of this album (see Figure 2). In order to prevent hiding possibly valuable information, all other songs as well as all other albums are still visible but marked invalid. The information about the validity of each row is shown in the column "status" and is additionally represented in the background color (see Figure 2). Initially, the rows are arranged according to their status in the following order:

1. **Selected:** Rows that are selected by the user.
2. **Valid:** Rows that are exclusively related to valid or selected rows and are not selected itself.
3. **Invalid:** Rows that are related to at least one invalid row and are not selected itself.

The number of valid rows together with the number of all rows is shown within brackets behind the type of information that is contained in a node, for example “Songs (2/10)” in Figure 2. Deselecting the by default selected button “all” at the top of the node reduces the visible rows in the list to only valid ones.

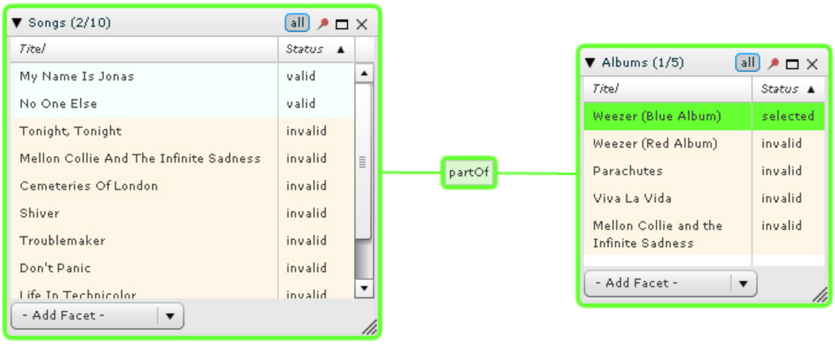


Fig. 2. Information can be filtered by selecting elements in a node.

In order to facilitate users’ understanding of how the elements in the nodes are filtered by certain selections, changes are highlighted by different colors (see Figure 2). Whenever a certain selection causes a decrease of valid rows in at least one node, this is reported by a specific color. The color depends on the node where the selection takes place and provides three different kinds of information:

1. **Selections:** Selected rows are highlighted in the distinct color of their nodes (see “Weezer (Blue Album)” in Figure 2).
2. **Restrictions:** Nodes that are restricted by a selection get surrounded by a ring in the corresponding color. If they are restricted by several selections from different nodes they get surrounded by several rings in different colors.
3. **Distributions:** Labeled edges that connect restricted nodes with the source of their restriction are colored correspondingly (see “partOf” in Figure 2).

The user can click on an already selected row to deselect it and thereby remove the restrictions and the corresponding coloring that were caused by this selection.

Hierarchical Facets

Our graph-based approach is particularly suitable for the presentation of relationships. Even nested relationships can be easily represented by nodes and labeled edges. In order to access information from distant perspectives, the graph can be gradually expanded by adding new nodes along the available relations.

Such indirectly related nodes can be used for filtering in the same way as directly related nodes. If the user selects certain rows to filter directly connected nodes as described above, the filtering of these nodes automatically leads to an update of again all directly related nodes and so on. The changes are iteratively propagated through the graph structure until no further reduction is caused and the elements in the nodes reach a stable status. A detailed description of this propagation algorithm is beyond the scope of this paper.

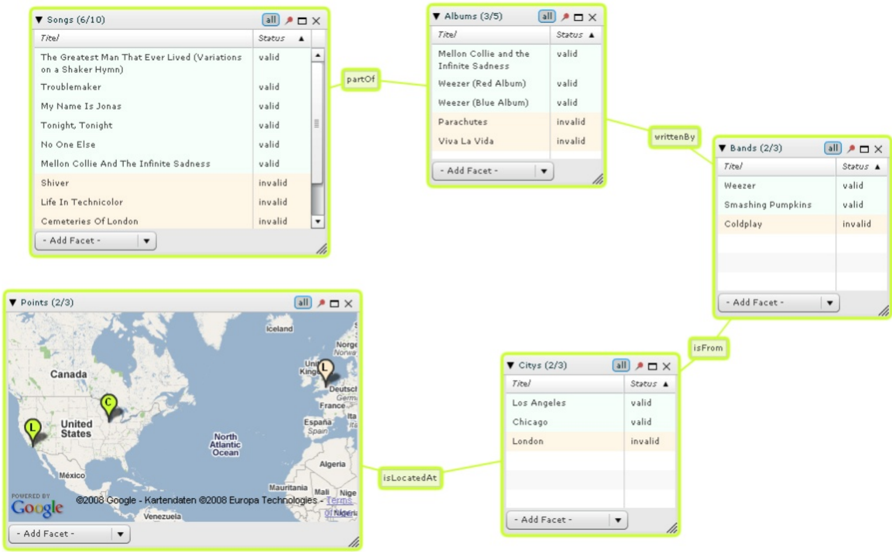


Fig. 3. Hierarchical facets can be used to filter information from distant perspectives.

Figure 3 illustrates how indirect relationships are visualized in our graph-based interface and how they can be used for filtering. The selection of Los Angeles and Chicago in the map restricts the valid music bands to only those that are from these cities. This leads to a restriction of the albums and hence to a restriction of the songs to only those that are part of these albums. That way, geographical locations of cities can be used to filter music songs. Similarly, other domains can be included in the graph to access and filter information from distant perspectives.

Since geographic coordinates expressed as latitude-longitude are not suitable for human processing when presented in a list, we used Google Map's geocoding service[6] to provide a map visualization. On a map, the distribution or distance of cities can be immediately understood and the corresponding areas and countries they are located in can be easily recognized. Our general aim is to provide a data- and task-related visualization where the system initially decides on the presentation form and the user then refines or corrects the system choice if it does not meet his needs. However, the currently implemented forms of presentation are limited to only lists and maps and are automatically selected. We leave the development of a semi-automatic approach with a broader range of presentation forms to future work.

4 User Study

We conducted a small user study to get some first insights on how well users can find information contained in RDF data using gFacet. Our primary interest was in how well users could apply graph-based facets to filter information. To perform this evaluation,

[6] http://code.google.com/apis/maps/

we asked users to find answers to three different questions. To assess users' understanding of the visual interface, we used an eye tracking system[7] in combination with a questionnaire. The results from the eye tracking helped us to better understand the actions performed and the answers given by the users. Note that the aim of this study was to examine whether our approach of graph-based facets is understood and purposefully applied by the user to find information contained in RDF data. We leave a comparative study as future work. 10 participants with an average age of 31, ranging from 21 to 54 years, took part in the study. Three of the participants were female and seven were male. Four participants had already heard about faceted browsing, six had not.

Search Tasks. To determine how well the participants can find certain information, we defined three tasks. With every task, they had to understand an additional functionality of our tool in order to find the information and answer the question. Since prior knowledge about faceted browsing can be supposed to have an impact on the results, we divided the participants into two groups according to their previous knowledge about faceted browsing. In the following we call the group without previous knowledge "Group noPK" and the group with previous knowledge "Group PK".

Task 1. For the first task, the users had to understand how new facets are added to the initially visible list of songs and how these facets can be used to filter the songs to only those that refer to the theme "rock". Three of the four participants of Group PK were able to find these songs. By contrast, only one of the six participants of Group noPK found them. The rates of success for each of the three tasks can be seen in Figure 6.

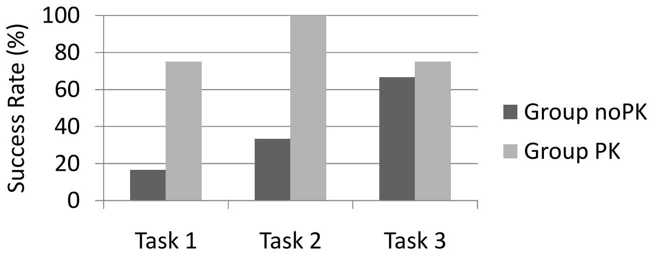


Fig. 6. Success rates by tasks.

Task 2. In order to solve the second task, the users had to understand the principle of hierarchical facets and how it can be used in our tool. To confirm their understanding, the participants were asked to find all songs that are part of albums written by the band "Smashing Pumpkins". All the participants of Group PK were able to use the hierarchical facet to find the right answer. The eye tracking results in Figure 7 suggest that the participants understood the way the connected nodes were filtered based on the selection they made in the hierarchical facet. The eye movement starts at the selected row and follows the highlighted restrictions through the graph structure. It seems as if the edges between the nodes together with the highlighting have guided the eyes to

[7] Tobii T60 Eye Tracker

look at the right spots. This can be also seen in the heatmap, where red areas indicate the highest number of eye fixations and green areas the fewest. However, only two participants of Group noPK found the right answer within the given time period.

Fig. 7. Typical eye movements and areas of interest.

Task 3. In the third task we wanted to evaluate if the participants were able to use the songs to filter the list of albums. They had to select a certain song in order to get the album it belongs to. All in all, only three participants did not succeed this task, the other seven did.

5 Conclusion and Future Work

In this paper, we presented a new approach for browsing the Web of data that combines a graph-based visualization with faceted filtering techniques. The approach supports the dynamic and user-defined integration of distant domains in order to allow the user to access information from different perspectives. This is facilitated by the separation of content and presentation as well as the fragmentation of the information and its relations that comes along with the transformation of the Web of documents to a Web of data.

We use a graph-based visualization to facilitate an understandable integration of different domains and a facet-based filter mechanism to support an efficient exploration of highly structured and interrelated datasets. The combination of a graph and facets is of particular benefit when dealing with complex dependencies as for example within hierarchical facets.

As a general conclusion of our user study, it can be stated that the direct representation of the relationships between facets by labeled edges in a graph facilitates users' understanding of their dependencies. However, the study also showed that participants with previous knowledge about faceted browsing performed better than participants without.

In future work we will extend the visual appearance of the facets and include further presentations forms next to lists and maps. Furthermore, we plan to conduct a comparative study of our tool in order to draw more general conclusions.

References

1. Resource Description Framework (RDF). http://www.w3.org/RDF/.
2. Goyal, S., Wetenthaler, R.: RDF Gravity. Salzburg Research (2004). http://semweb.salzburgresearch.at/apps/rdf-gravity/.
3. SIMILE: Welkin (2004-2005). http://simile.mit.edu/welkin/.
4. IsaViz: A Visual Authoring Tool for RDF (2001-2006). http://www.w3.org/2001/11/IsaViz/.
5. Deligiannidis, L., Kochut, K., Sheth, A.: RDF data exploration and visualization. In: Proc. of the first workshop on CyberInfrastructure 2007, ACM Press, pp. 39-46. (2007)
6. Angles, R., Gutierrez, C.: Querying RDF data from a graph database perspective. In: Proc. ESWC 2005, pp. 346-360. (2005)
7. Frasincar, F., Telea, A., Houben, G.-J.: Adapting graph visualization techniques for the visualization of RDF data. In: Visualizing the Semantic Web 2006, pp. 154-171. (2006)
8. Schraefel, m. c., Karger, D.: The Pathetic Fallacy of RDF. In: International Workshop on the Semantic Web and User Interaction (SWUI) 2006. (2006)
9. Yee, K.-P., Swearingen, K., Li, K., Hearst, M.: Faceted metadata for image search and browsing. In: CHI 03, ACM Press, pp. 401-408. (2003)
10. A. Inc. itunes. http://www.apple.com/itunes/.
11. Schraefel, m.c., Smith, D., Owens, A., Russell, A., Harris, C., Wilson, M.: The evolving mSpace platform: leveraging the semantic web on the trail of the memex. In: Proc. of Hypertext 2005, ACM Press, pp. 174-183. (2005)
12. SIMILE: Longwell RDF Browser (2003-2005). http://simile.mit.edu/longwell/.
13. Hildebrand, M., Ossenbruggen, J., Hardman, L.: /facet: A browser for heterogeneous semantic web repositories. In: Proc. ISWC 2006. (2006)
14. Quan, D., Huynh, D., Karger, D.: Haystack: A Platform for Authoring End User Semantic Web Applications. In: Proc. of ISWC 2003, pp. 738-753. (2003)
15. Berners-Lee, T. Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., Lerer, A., Sheets, D.: Tabulator: Exploring and Analyzing linked data on the Semantic Web. In: Proc. of the 3rd International Semantic Web User Interaction Workshop 2006. (2006)
16. Huynh D.: Nested faceted browsing. http://people.csail.mit.edu/dfhuynh/projects/nfb/.
17. Fruchterman, T., Reingold, E.: Graph drawing by force-directed placement. In: Softw. Pract. Exper. 1991, John Wiley & Sons, pp. 1129-1164. (1991)

select * where { :I :trust :you }
How to Trust Interlinked Multimedia Data

Wolfgang Halb and Michael Hausenblas

Institute of Information Systems & Information Management,
JOANNEUM RESEARCH, Steyrergasse 17, 8010 Graz, Austria
firstname.lastname@joanneum.at

Abstract. Finding, accessing and consuming multimedia content on the Web is still a challenge. In this position statement we discuss three still widely neglected issues arising when one is interacting with multimedia content in social media environments: provenance, trust, and privacy. We will introduce a generic model allowing us to identify potential risks and problems, further discuss this model regarding multimedia content and finally outline how Semantic Web technologies can help.

1 Motivation

It is a trivial truth that, in order to use any kind of content or service on the Web, one must know how to access it (that is, to know the URI). What is true for the Web of Documents is equally true for the Web of Data. While with the rise of linked data [1, 2] the situation has changed—publishing and consuming data is now possible straight forward—there are still a number of issues in the discovery process. For example, with our multimedia interlinking demonstrator CaMiCatzee [3] we have identified issues around trust and believe of information regarding linked data in general and multimedia content in special. CaMiCatzee allows the FOAF-profile-based search for person depictions on flickr. However, when looking at Fig. 1, how can we find out if one of the depicted persons actually is Sir Tim Berners-Lee?

Fig. 1. Exemplary result from CaMiCatzee.

Motivated by this observation we will address—from a practical point of view—widely neglected issues arising when one is interacting with multimedia content in social platforms:

- **Provenance**. Where does content stem from? Who provided annotations?
- **Trust**. Is a person that provided an annotation trustworthy? Is the interlinking eligible?
- **Privacy**. When interacting with content—what are the consequences?

In the following, we will first introduce a generic model addressing the three above listed issues. Based on our experience with interlinked multimedia data we will have a detailed look at the consequences when this model is applied to multimedia content.

2 The Abstract Provenance-Trust-Privacy Model

In order to identify issues with the usage of content on the Web, we have developed the "Provenance-Trust-Privacy" (PTP) model (Fig. 2). Basically, two aspects are covered by this model: the real life and the online world, the Web. In the PTP model we deal with three orthogonal, nevertheless interdependent dimensions, (i) the **social** dimension, (ii) the **interaction** dimension, and (iii) the **content** dimension.

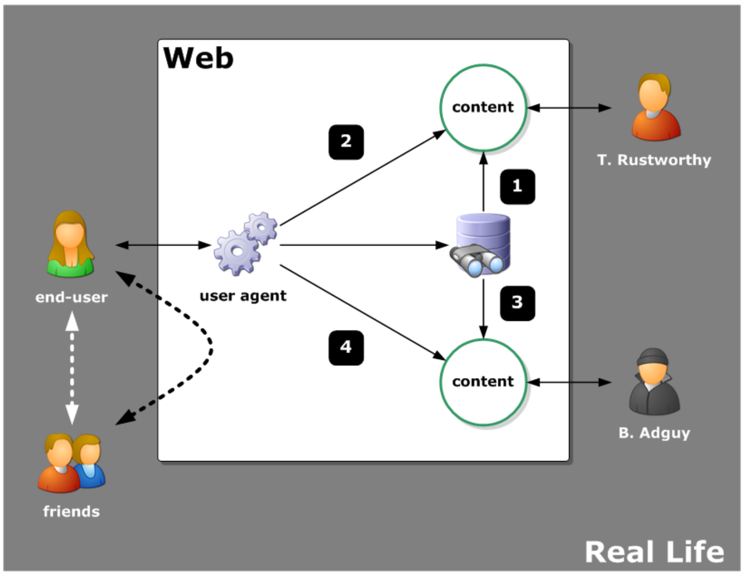


Fig. 2. The abstract PTP model.

The Social Dimension. The dotted arrows in Fig. 2 represent social relations between humans, either in the real life or online. While it is straight-forward to deal with the case when people know each other in real life (and maybe continue this relation in the online world), the other way round can cause substantial problems. For example, only because I have added someone to my "buddy list" on a social platform such as LinkedIn does not mean that I really know the person and that this person also (wants to) know me.

The Interaction Dimension. In the realm of the PTP model we understand the interaction dimension of taking place online-only. Again, referring to Fig. 2, everything that happens between the user-agent (being instructed by a human) and other participants (server, etc.) on the Web. In general, two interaction patterns for the discovery and access of resources on the Web can be observed:

- A **direct access** of the content. In this case the URI of the content is known in advance by the end-user who instructs her user-agent to access the content. The URI may originally stem from a newspaper advertisement, a friend may have pointed it out in an e-mail, etc.
- An **indirect access** of the content by means of consulting an intermediate such as a search engine, a recommendation system, etc.

Based on these two interaction patterns, four possible paths can be identified:

1. The user-agent, equipped with the end-user's profile and her desire consults an intermediate. For example, a user enters a search string into Google and is presented a list of URIs. The end-user happens to select a trustworthy source.
2. The user-agent, equipped with an URI from the end-user, accesses a trustworthy source.
3. The user-agent, equipped with the end-user's profile and her desire consults an intermediate. This time, the end-user happens to select a troublesome source.
4. The user-agent, equipped with an URI from the end-user, accesses a troublesome source.

Obviously, the first two situations are desirable. The end-user has—for example based on previous experiences or trust in the search results—found some content that she can use and which is not causing her troubles (a virus, Trojan horse, etc.). However, equally, we are after avoiding the latter two cases where the end-user actually finds herself using content and/or services that are harmful and/or violate her privacy.

The Content Dimension. Regarding the content dimension one generally can state that the more is known about the content, the easier it is to assess its usefulness and its capabilities regarding a potential damage. Wherever possible, we are after self-descriptive resources, that is we require a minimum level of metadata being available. In our case, we focus on multimedia content. In the next section we will therefore initially discuss this kind of content and along the metadata in greater detail.

3 Multimedia Content in the Provenance-Trust-Privacy Model

In the position paper at hand we focus on multimedia content. We will in the following discuss characteristics of spatio-temporal multimedia content in the context of the emerging interlinking multimedia effort[1]. Further, we apply the above introduced PTP model to multimedia content and try to derive requirements for it.

Characteristics of Multimedia Content and Multimedia Metadata. Multimedia content has some specific characteristics that allow and/or request special treatment. We have reported on this in great detail elsewhere [4]. A basic observation, however, with impact on many parts of the interaction process is that with multimedia content we are dealing almost always with spatio-temporal dimensions.

From the prosumers point-of-view, multimedia content is cheap to produce and available in high volumes (mobile phones, etc.). Further, most of the current content in that regard is publicly and freely available (Flickr, youtube). Business models remain vague. On the other hand, for professionally created content for very specific domains such as broadcaster's archives, adult content, etc. the fees are considerably high.

Multimedia content is in general good for consumption in mobile environments (as opposed to reading longish text on a mobile).

In general it is hard and expensive to create good and detailed multimedia content descriptions (for example in MPEG-7, etc.). This leads to a problem regarding the fine-grained search and automated summaries.

In Table 1 the above discussed characteristics are summarised, and weighted regarding the content itself on the one hand and the metadata on the other hand.

Issue	**Content**	**Metadata**	**Remark**
production (prosumers)	++	-	easy to produce high-volume content (e.g., mobile phone)
production (professionals)	++	-	esp. high-level semantic content descriptions expensive
consumption	++	-	easy to consume (also in mobile environments)
search	-	- -	practically, only global descriptions are available
summaries	- -	- -	little automation possible

Table 1. Overview on Multimedia Content Characteristics.

Applying the PTP Model to Multimedia Content. With the above listed characteristics of multimedia content in mind we claim the following regarding the application of the PTP model:

- Any solution addressing PTP issues must at least avoid accessing "bad" content and should support the discovery and consumption of "good" content.

[1] `http://www.interlinkingmultimedia.info/`

- Existing and deployed multimedia metadata formats (such as Exif, ID3, etc.) have to be taken into account.
- The solution at hand needs to scale to the size of the Web.
- It must be practically relevant in terms of availability in widely used platforms such as Drupal, MediaWiki, etc. (for example as a plug-in, etc.; however it needs to be integrated).
- Provider must be able to easily offer and administer it (e.g., “enable” it with little configuration effort).
- Consumer must be able to use it in a non-disruptive way, for example as a part of their everyday tools.

In the next section we will report on how Semantic Web technologies can be used in combination with other, deployed technologies (such as for identification and authentication) in order to address the above listed requirements.

4 How Semantic Web Technology Can Help

We strongly believe that Semantic Web technologies can help to realise a PTP model for multimedia content. Lots of research is already available[2], however with little practical impact to this end. Apart from avoiding unreliable or even malicious content, the main aim of applying PTP to multimedia content is to help the user in finding trustworthy information sources. In a first step, we consider all content created by a trusted person or authority to be trustworthy. Solving this issue implies that there need to be techniques that can ensure a content’s provenance and the content producer’s identity respectively. Consequently means have to be made available that can decide which person to trust or not.

In the case of multimedia content it also has to be taken into account that information associated with a single content item can potentially have a multitude of contributors. A photo along with some metadata (title, description) on Flickr for instance might be uploaded by the fictitious trusted user “T. Rustworthy”. Another user,“B. Adguy”, could add a fake note about who is depicted in the photo. When accessing the photo and the associated metadata it is thus not sufficient to only consider who contributed the image but we would also need to be able to figure out who contributed the metadata about it. Taking this further to video content it might also be relevant to take into account who contributed which parts of a video (consider, for example, advertisements inserted into a video stream).

In the following we will discuss already available technologies that may be able to address the PTP issues along the three identified dimensions. However, to date only isolated solutions exist and there is still a lack of systems that incorporate all available technologies for the user’s benefit. We envision a framework that would allow to combine the below listed technologies and develop plug-ins for widely used platforms (Flickr, Youtube, etc.) and systems (Drupal, etc.).

[2] `http://www4.wiwiss.fu-berlin.de/bizer/SWTSGuide/`

Social Dimension. For the identification as well as for the authentication several technologies are available. A user can for example provide her OpenID[3] to identify against a service. Further, OAuth[4] can be used for publishing and interacting with protected data. Big players such as Google are already offering support for the above mentioned technologies[5]. It seems advisable to build on this and contemplate what might be missing to align it with the Web of Data, being based on RDF [5].

While FOAF-based white listing and other related approaches [6, 7] are available, there is still a need for an up-front agreed way to deploy it in widely used systems. The same issue can be observed with privacy: there are proposals on the table (for example P3P[6]) but no measurable uptake in terms of users, systems that offer it, etc. can be stated.

Interaction Dimension. Especially for data provenance it seems to us that named graphs [8] offer a solid and scalable solution. With the rise of RDFa[7] one can think of new provenance mechanisms, as the hosting document can actually be understood as the "name" of the graph. Just imagine Flickr (already offering licensing information in RDFa) to include provenance information on both the content and the metadata, based on vocabularies such as the "Semantic Web Publishing Vocabulary" [9, Chapter 6]. Finally, we note that regarding the discovery and usage of linked (multimedia) data we are currently working on VoiD, the "Vocabulary of Interlinked Datasets"[8]—again, provenance and trust issues are in scope, here.

Content dimension. In our understanding, the content dimension of the PTP model requires the most attention. Basic mechanisms proposed to represent the type of multimedia content in RDF[9] are available. Still, practical ways for creating and consuming rich multimedia content descriptions are missing. Recently, we have proposed ramm.x [10] which allows to use existing multimedia metadata formats such as MPEG-7, Exif, ID3, etc. in the Web of Data. However, we expect a fair amount of further research being required to address provenance and trust issues properly and make tools and applications available in a real-world setup.

References

1. C. Bizer, T. Heath, D. Ayers, and Y. Raimond. Interlinking Open Data on the Web (Poster). In *4th European Semantic Web Conference (ESWC2007)*, pages 802–815, 2007.
2. C. Bizer, R. Cyganiak, and T. Heath. How to Publish Linked Data on the Web. `http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/LinkedDataTutorial/`, 2007.

[3] `http://openid.net/`
[4] `http://oauth.net/`
[5] `http://googledataapis.blogspot.com/2008/06/oauth-for-google-data-apis.html`
[6] `http://www.w3.org/P3P/`
[7] `http://www.w3.org/TR/xhtml-rdfa-primer/`
[8] `http://community.linkeddata.org/MediaWiki/index.php?VoiD`
[9] `http://www.w3.org/TR/Content-in-RDF/`

3. M. Hausenblas and W. Halb. Interlinking Multimedia Data. In *Linking Open Data Triplification Challenge at the International Conference on Semantic Systems (I-Semantics08)*, 2008.
4. T. Bürger and M. Hausenblas. Why Real-World Multimedia Assets Fail to Enter the Semantic Web. In *International Workshop on Semantic Authoring, Annotation and Knowledge Markup (SAAKM07)*, Whistler, Canada, 2007.
5. G. Klyne, J. J. Carroll, and B. McBride. RDF/XML Syntax Specification (Revised). W3C Recommendation, RDF Core Working Group, 2004.
6. J. Golbeck. Combining Provenance with Trust in Social Networks for Semantic Web Content Filtering. In *Provenance and Annotation of Data, International Provenance and Annotation Workshop, IPAW 2006, Chicago, IL, USA, May 3-5, 2006, Revised Selected Papers*, volume 4145 of *Lecture Notes in Computer Science*, pages 101–108. Springer, 2006.
7. A. Harth, A. Polleres, and S. Decker. Towards a social provenance model for the Web. In *Workshop on Principles of Provenance (PrOPr), Edinburgh, Scotland*, 2007.
8. J. J. Carroll, C. Bizer, P. Hayes, and P. Stickler. Named Graphs, Provenance and Trust. In *In WWW 05: Proceedings of the 14th international conference on World Wide Web*, pages 613–622. ACM Press, 2005.
9. C. Bizer. *Quality-Driven Information Filtering in the Context of Web-Based Information Systems*. PhD thesis, Freie Universität Berlin, 2007.
10. M. Hausenblas, W. Bailer, T. Bürger, and R. Troncy. Deploying Multimedia Metadata on the Semantic Web (Poster). In *2nd International Conference on Semantics And digital Media Technologies (SAMT 07)*, 2007.