The Multimedia Metadata Community

(http://www.multimedia-metadata.info)

presents the

Proceedings of the

10th International Workshop on Semantic Multimedia Database Technologies (SeMuDaTe'09)

Graz, Austria, December 2, 2009

Edited by

Ralf Klamma, RTWH Aachen University
Harald Kosch, University of Passau
Matthias Lux, University of Klagenfurt
Florian Stegmaier, University of Passau

Foreword

We have the pleasure to organize the 10th Workshop of the Multimedia Metadata Community (http://www.multimedia-metadata.info). This second 2009 workshop has a special focus on semantic multimedia database technologies and is held in conjunction with the 4th International Conference on Semantic and Digital Media Technologies (SAMT 2009), December 2-4, Graz, Austria. Both events are bringing researchers and industry experts to fruitful discussions.

Ontology-based systems have been developed to structure content and support knowledge retrieval and management. Semantic multimedia data processing and indexing in ontology- based systems is usually done in several steps. One starts by enriching multimedia metadata with additional semantic information (possibly obtained by methods for bridging the semantic gap). Then, in order to structure data, a localized and domain specific ontology becomes necessary since the data has to be interpreted domain-specifically. The annotations are stored in an ontology management system where they are kept for further processing. In this scope, Semantic Database Technologies are now applied to ensure reliable and secure access, efficient search, and effective storage and distribution for both multimedia metadata and data. Their services can be used to adapt multimedia to a given context based on multimedia metadata or even ontology information. Services automate cumbersome multimedia processing steps and enable ubiquitous intelligent adaptation. Both, database and automation support facilitate the ubiquitous use of multimedia in advanced applications.

This time we got 21 submissions in as full, position or demonstration papers. Altogether we accepted 7 full papers, 5 position papers and 2 demonstration papers. Our thanks go again to the reviewers, who provided timely and thorough reviews. Their suggestions allowed authors to better their contributions.

Naturally, our thanks also go to the organizers of SAMT 2009, namely to Werner Bailer. Their logistic support has been essential to the organization of our workshop.

We wish you a productive and enriching workshop and an excellent stay in Graz.

Your workshop co-chairs,

Ralf Klamma, RTWH Aachen University
Harald Kosch and Florian Steigmaier, University of Passau
Matthias Lux, University of Klagenfurt

Conference Organization

Programme Chairs

Ralf Klamma
Harald Kosch
Mathias Lux
Florian Stegmaier

Programme Committee

Giuseppe Amato
Werner Bailer
Laszlo Boeszoermenyi
Lionel Brunie
FranÃois Bry
Yu Cao
Yiwei Cao
Anna Carreras
Vincent Charvillat
Savvas Chatzichristofis
Richard Chbeir
Thierry Delot
Mario Doeller
Baltasar Fernandez-Manjon
Romulus Grigoras
William Grosky
Christian Guetl
Pascal Hitzler
GÃnther HÃlbling
Oge Marques
Britta Meixner
Timo Ojala
Vincent Oria
Chris Poppe
Dominik Renzel
Ansgar Scherp
Timothy K. Shih
Marc Spaniol
Markus Strohmaier

Local Organization

M3O: The Multimedia Metadata Ontology

Carsten Saathoff and Ansgar Scherp

ISWeb, University of Koblenz-Landau, Universitätsstr. 1, Koblenz 56070, Germany
{saathoff,scherp}@uni-koblenz.de

Abstract. We propose the Multimedia Metadata Ontology (M3O), a framework for integrating the central aspects of multimedia metadata. These central aspects are the separation of the information conveyed by multimedia items and their realization, the annotation with both semantic and low-level metadata, and the decomposition of multimedia content. M3O bases on Semantic Web technologies and provides the means for rich semantic annotation using further, possibly domain-specific ontologies. Moreover, it can be used to represent other existing metadata models and metadata standards. We introduce the M3O and present its application at the example of a SMIL presentation.

1 Introduction

Multimedia metadata and semantic annotation of multimedia content is the key-enabler for improved services on multimedia content. The archiving, retrieval, and management of multimedia content becomes very hard if not even practicably infeasible if no or only limited metadata and annotations are provided. Looking at the existing metdata models and metadata standards, we find a huge number and variety serving different purposes and goals. In addition, the models are of different scope and level of detail. Typically, the existing models cannot be combined with each other. For example, image descriptions using EXIF [1] can not be combined with MPEG-7 [2] descriptors. In addition, the existing models are semantically ambiguous, i.e., they do not provide a well-defined interpretation of the metadata. For example, in IPTC [3] the location fields are defined to contain the locations the content is "focusing on". However, it remains unclear what this "focusing on" actually means. For instance, consider an image from the atomic bombing of Nagasaki in Japan in 1945. This image is about Nagasaki since it documents an event taking place in that city. But it is also about the world as a whole since the atomic bombing is of global importance. Distinguishing these different roles a location can play is impossible with IPTC. In general, support for semantic annotations using formally defined background knowledge is hardly found. Finally, the models are typically focused on a single media type, ignoring the type's relation to other media types or their context within a true multimedia presentation. As a consequence of this, providing interoperability between different applications that deal with the storage, retrieval, and delivery of multimedia content and single media assets annotated with today's models becomes very hard. However, this is required in many multimedia application scenarios, in particular in the open world of the Web.

What is missing is a representation of the data structures that underlie today's multimedia metadata models and metadata standards. We aim at extracting the common patterns underlying existing metadata models and metadata standards. We provide these patterns as a set of *ontology design patterns (ODPs)* [4]. It provides a comprehensive modeling framework for representing arbitrary multimedia metadata and is called the Multimedia Metadata Ontology (M3O). Basing the M3O on Semantic Web and ontologies particularly provides support for the rich semantic annotation of multimedia content.

2 Annotating Structured Multimedia Content

Using a simple scenario, we show the different requirements that need to be considered when annotating structured multimedia content. We assume that we need to give a lecture on discussing the advantages and disadvantages of nuclear energy. For this lecture, we have prepared a multimedia presentation shown in Figure 1 to start discussions. Both for later retrieval and for descriptive purposes, we would like to annotate the presentation. The multimedia content of our multimedia presentation consists of different single media assets. These media assets are combined in a coherent, structured way. This means that the content provides a spatial layout and a temporal course and also includes interactivity. The multimedia content is encoded using the multimedia presentation format SMIL[1] and rendered using the RealPlayer[2].

(a) (b)

Fig. 1: An image of Albert Einstein combined with an image of the Times Square and an image of the nuclear bomb cloud expressing contrary views on *nuclear energy*.

Our SMIL presentation discussing the advantages and disadvantages of nuclear energy consists of two parts. The first part depicted in Figure 1a shows a picture of Albert Einstein[3] and a photo of the Times Square in New York.

[1] Synchronized Multimedia Integration Language, `http://www.w3.org/TR/2008/REC-SMIL3-20081201/`

[2] RealNetworks, Inc., 2009, `http://www.real.com/realplayer/`

[3] `http://en.wikipedia.org/wiki/File:Einstein1921_by_F_Schmutzer_4.jpg`, from Wikipedia. The image is in the public domain.

This part of the presentation serves as a metaphor for the achievements reached by the discovery of nuclear energy in which Einstein played a central role. By the peaceful use of nuclear energy, it can serve large cities like New York with electricity.

In the second part of our SMIL presentation depicted in Figure 1b, we replace the photo of the Times Square by a picture showing the atomic bombing of the city of Nagasaki[4] in Japan in 1945. The picture of Einstein remains unchanged. However, the contextual use in which the picture of Einstein is shown is completely different. By this change of contextual use, the media assets composed transmit a totally different message and express a different semantics [5]. Instead of showing the advantages of nuclear energy, this part of the presentation serves as metaphor for the risks and the potential destructive power of nuclear energy.

For providing a comprehensive semantic description of this multimedia presentation, there are different kinds of annotations involved. These different annotations put requirements to the metadata model used to represent the semantics of the multimedia content shown in the presentation. We discuss these requirements in the following section.

3 Requirements on a Multimedia Metadata Model

From the scenario above, we can derive three principal requirements that need to be supported for annotating rich, structured multimedia content such as the SMIL presentation in the scenario. These requirements are the separation between information objects and information realization, multimedia annotation, and multimedia decomposition. They need to be reflected by a multimedia metadata ontology.

Separation between Information Objects and Information Realizations. On the conceptual level, multimedia content conveys information to the consumer. As such, the multimedia content plays the role of a message that is transmitted to a recipient. Such a message can be understood as an abstract information object [6]. Examples of information objects are stories, stage plays, or narrative structures. The information object can be realized by different so-called information realizations [6]. The narrative structure of our scenario above is, e.g., realized in a SMIL presentation. The following requirements of multimedia annotation and multimedia decomposition can be applied on both levels of information objects and information realization.

Annotation of Information Objects and Information Realizations. The model needs to support the annotation of multimedia content. This can be annotations in the style of typed key-value pairs as provided, e.g., by EXIF or semantic annotation, i.e., the use of semantic background knowledge for describing the multimedia content. In our example, we could annotate the picture from the

[4] `http://commons.wikimedia.org/wiki/File:Nagasakibomb.jpg`, from Wikimedia Commons. The image is in the public domain.

Times Square with the geo-coordinates where it was taken or annotate the whole presentation with the general topic it discusses. Please note that low-level metadata, such as EXIF, typically is attached to the realization, while the semantic annotation rather applies to the information object.

Decomposition of Information Objects and Information Realizations. Multimedia content can be decomposed into its constituent parts. The SMIL presentation above can, e.g., be decomposed into the two parts it consists of. The parts can be decomposed into the images they contain. The realization of the presentation can be decomposed into the realizations of the contained images. Decomposition can be applied arbitrarily often, i.e., we can can create a hierarchy of parts.

4 Related Work

In research and industry, numerous metadata models and metadata standards have been proposed so far. These models come from different backgrounds and with different goals set. They vary in various aspects such as the domain for which they have been designed. The models can be domain-specific or designed for general purpose. The existing metadata models also focus on a specific single media type such as image, text, or video. In addition, the metadata models differ in the complexity of the data structures they provide. With standards like EXIF [1], XMP [7], and IPTC [3] we find metadata models that provide (typed) key-value pairs to represent metadata of the image media type. Harmonization efforts like in the case of image metadata pursued by the Metadata Working Group[5] are very much appreciated. However, they remain on the same technological level and do not extend their effort beyond the single media type of image. Another metadata model like Dublin Core[6] and its extension for multimedia content[7] support hierarchical modeling of key-value pairs. It can be used to describe almost any resources. However, only entire documents and not parts of it. With MPEG-7 [2], we find a comprehensive metadata standard that aims at covering mainly decomposition and description of low-level features of audiovisual media content. MPEG-7 also provides basic means for semantic annotation. Several approaches have been published providing a formalization of MPEG-7 as an ontology, e.g., by Hunter [8] or the Core Ontology on Multimedia (COMM) [9]. However, although these ontologies provide clear semantics and an integration with Semantic Web standards, they still focus on MPEG-7 as the underlying metadata standard. As a consequence, they do not provide a generic framework for the integration of different metadata standards and metadata models. Furthermore, most metadata models also lack in supporting structured multimedia content. Structured multimedia content means that the content is organized in different discrete media assets such as images and text and continuous media assets like videos and audio. It has a coherent spatial layout, temporal course,

[5] `http://www.metadataworkinggroup.org/`
[6] `http://dublincore.org/`
[7] `http://dublincore.org/documents/dcmi-type-vocabulary/`

and some interaction with the user. Annotation of such structured multimedia content is in principle possible with MPEG-7 using separate media signals for the invidual media assets. However, actually doing it for a complex structured multimedia presentation is not very practical due to the complexity involved with this MPEG-7 annotation. In addition, various studies have shown the need in image retrieval for semantic annotation and conceptual queries [10–12].

This list of metadata models and metadata standards is very far from being complete and is beyond the scope of this work. Some overview of multimedia metadata models and standards can be found in a report [13] by the W3C Multimedia Semantics Incubator Group or in the overview[8] of the current W3C Media Annotations Working Group. The examples mentioned have been selected to show the variety of the different multimedia metadata models that exist today.

5 Multimedia Metadata Ontology

For defining our Multimedia Metadata Ontology (M3O), we leverage Semantic Web technologies and follow a pattern-oriented ontology design approach. We identified five core patterns required to express metadata for multimedia content. These patterns model the basic structural elements of existing metadata formats and conceptual models. In order to realize a specific metadata standard or metadata model in M3O, these patterns need to be specialized. The patterns base on the foundational ontology DOLCE+DnS Ultralight[9] and are formalized using Description Logics [14]. By this, we provide a clear semantics of the patterns and their elements. We achieve an improved formal representation of the metadata compared to existing models. In addition, such a generic model is not limited to a single media type such as images, video, text, and audio but provides support for structured multimedia content as it can be created with today's multimedia presentation formats such as SMIL, SVG[10], and Flash[11].

Furthermore, implementing the M3O using Semantic Web technologies is a promising approach, as it allows for representing rich metadata and multimedia semantics. Thus, it provides the infrastructure to represent both high-level semantic annotation with background knowledge as well as the annotation with low-level features extracted from the multimedia content. In addition, existing standardized multimedia presentation formats such as SMIL and SVG explicitly define the use of the Semantic Web standard RDF [15] for modeling the annotations. Semantic Web technologies ease the use of formal domain ontologies, leverage the employment of reasoning services, and provide the means to exploit the growing amount of Linked Open Data[12] available on the web.

[8] http://www.w3.org/2008/WebVideo/Annotations/drafts/ontology10/WD/mapping_table.html
[9] http://ontologydesignpatterns.org/wiki/Ontology:DOLCE+DnS_Ultralite
[10] http://www.w3.org/Graphics/SVG/
[11] http://www.adobe.com/de/products/flashplayer/
[12] http://linkeddata.org/

In the following, we introduce three basic patterns from DOLCE+DnS Ultralight that we use for our model. Subsequently, we present two patterns provided by M3O for multimedia annotation and multimedia decomposition.

5.1 DOLCE+DnS Ultralight Patterns

The Descriptions and Situation Pattern allows for the representation of contextualized views on the relations of a set of individuals and is depicted in Figure 2a. It provides a formally defined mechanism to view relations among individuals within a context, and assign roles or types that are only valid within this context.

The pattern consists of a Situation that satisfies a Description. The Description defines the roles and types present in a context, called Concepts. Each Concept classifies an Entity. The entities are the individuals that are relevant in a given context. Each Entity is connected to the situation via the hasSetting relation. Furthermore, the concepts can be related to other concepts by the isRelatedToConcept relation in order to express their dependency. The Descriptions and Situations Pattern therefore expresses an n-ary relation among a set of entities. The concepts determine the roles that the entities play within this context.

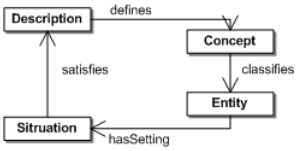
The information realization pattern in Figure 2b models the distinction between information objects and information realizations. An example is the lecture from our scenario and its realization as a SMIL presentation. The lecture would be the information object, while the SMIL presentation is the information realization. The same information can be realized in different ways. The pattern consists of the InformationRealization that is connected to the InformationObject by the realizes relation. Both are subconcepts of InformationEntity, which will make presentation of our M3O patterns easier.

With ontologies, we can use abstract concepts and clearly identifiable individuals to represent data and to perform inferencing over the data. However, at a certain point one will need to represent concrete data values, such as strings or numerical values. The Data Value Pattern (depicted in Figure 2c) assigns a concrete data value to an attribute of that entity. The attribute is represented by the concept Quality and is connected to the Entity by the hasQuality property. The Quality is connected to a Region by the hasRegion relation. The Region models the data space the value comes from. We attach the concrete value to the Region using the relation hasRegionDataValue. The data value is encoded using typed literals, i.e., the datatype can be specified using XML Schema Datatypes [16]. Using the hasPart relation, we can also express structured data values, such as present in MPEG-7.

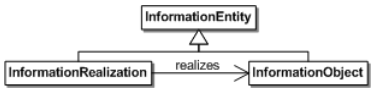
5.2 Annotation Pattern

Annotation denotes the description of some entity in terms of a note or an explanation[13]. In the context of a computer system, annotation usually refers to the description of some document stored on the computer. An example might be the

[13] Merriam-Webster Online, `http://www.merriam-webster.com`.

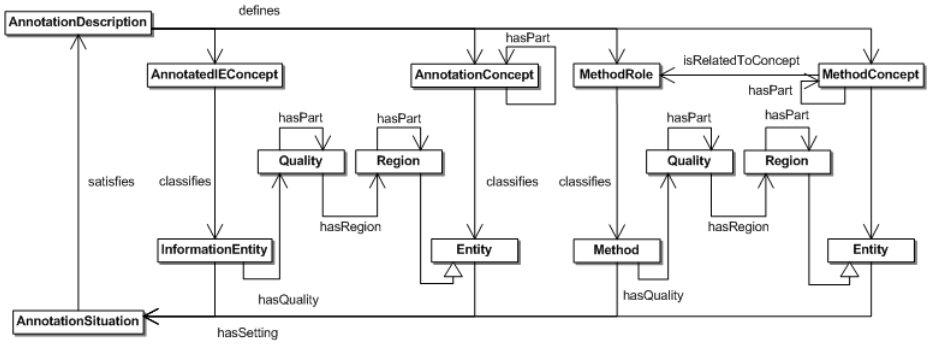


(a) Description and Situation Pattern

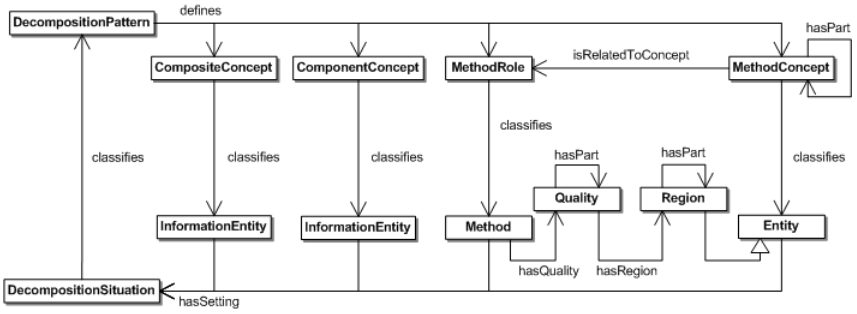


(b) Information Realization Pattern

(c) Data Value Pattern



(d) Annotation Pattern



(e) Composition Pattern

Fig. 2: Ontology Patterns of the Multimedia Metadata Ontology (M3O)

tagging of images on Flickr[14]. More generally speaking, we can define annotation as the attachment of metadata to an information entity on a computer system.

As we have discussed in Section 4, metadata comes in a various forms, such as low-level descriptors obtained by automatic methods, non-visual information covering authorship and technical details, or semantic annotation, aiming at a formal and machine-understandable representation of the contents. We identified that the underlying basic structure of annotation is always the same. Our annotation pattern models this basic structure and allows for assigning arbitrary annotations to information entities, while providing the means for modeling provenance and context.

The Annotation Pattern depicted in Figure 2d is a specialization of the Descriptions and Situations pattern and consists of an AnnotationSituation that satisfies an AnnotationDescription. The description defines at least one AnnotatedIEConcept that classifies each InformationEntity that is annotated by an instance of this pattern. The InformationEntity has as setting the AnnotationSituation. Each metadata item is represented by an Entity that is classified by an AnnotationConcept. Furthermore, we can express provenance and context information using the second part of the pattern. A Method that is classified by some MethodRole might specify how this annotation was produced. An example could be an agorithm or a manual annotation. We can further describe details, such as parameters, of the applied Method using a number of entities included in the IEAnnotationSituation that are classified by MethodConcepts, which are related to the MethodRole. In case of concrete data values for the metadata or the parameters, the Data Value Pattern is used. Please note that in the case of structured data values, also the MethodConcepts might have parts. This is expressed by the hasPart relation that classifies the parts of the Region.

5.3 Decomposition Pattern

Our Decomposition Pattern models the decomposition of information entities, e.g., the decomposition of a SMIL presentation into its logical parts or the segmentation of an image. After a decomposition, there is a whole, the composite, and there are the parts, the components. We decided to call this pattern Decomposition Pattern, since from a metadata point of view we decompose the media into parts, which we want to annotate further. Obviously, the same pattern can also be viewed as a composition of media elements and might be used like that.

The Decomposition Pattern consists of an IEDecompositionDescription that defines exactly one CompositeConcept and at least one ComponentConcept. The CompositeConcept classifies an InformationEntity, which is the whole. Each ComponentConcept classifies an InformationEntity, which are the parts. We can further specify a Method which generated the composition, and which is classified by a MethodRole. The Method can further be described by entities that are classified by Concepts, providing the means to model parameters or more abstract reasons for this decomposition. This part of the pattern is similar to the Annotation Pattern. All classified entities have the IEDecompositionSituation as setting.

[14] `http://flickr.com/`

It is important to note that in cases of structured multimedia content there is already composition information available in the media itself. A SMIL file, e.g., contains information about how single media assets are arranged. However, with M3O we aim at representing metadata about parts of the media that are not necessarily equal to or included in the physical structure defined in the SMIL file.

6 Application of M3O

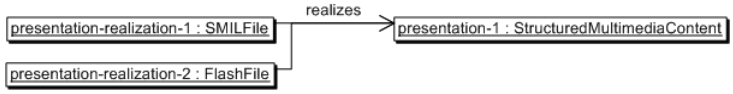
We demonstrate the application of our Multimedia Metadata Ontology at the example of the scenario in Section 2. For reasons of brevity, we present the core aspects of our model, namely the information realization, decomposition, and annotation of multimedia. Decomposition and annotation are only demonstrated on the information object level. More elaborate examples, up-to-date documentation, and discussions will be available from our wiki[15]. In the following, we use the term individual when we refer to concrete objects and the term concept when we refer to concepts of the M3O ontology. Please note that within an instantiation of a pattern only individuals appear. Additionally, we use terms like image or presentation in order to refer to the information object, and terms like image file or SMIL presentation when we refer to their realization.

We start with an example of how to apply the Information Object Pattern in order to represent the two basic levels of our model, i.e., the information object and the information realization. In this example, we consider two realizations of our presentation, namely one based on SMIL and one based on Flash.

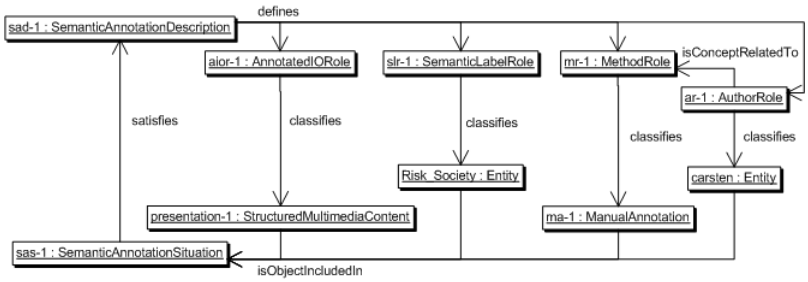
In Figure 3a, we can see that there is one individual presentation-1 of type StructuredMultimediaContent, which is a subclass of InformationObject. The files are represented by the individuals presentation-realization-1 and presentation-realization-2, which realize the presentation. They are of type SMILFile and FlashFile, which are subclasses of InformationRealization. Further information about the realization such as storage location, size, access rights, and others can be added using the annotation pattern.

In Figure 3b, the application of the Annotation Pattern is shown. The description defines four roles. The first two roles are an AnnotatedIORole and a SemanticLabelRole. The former classifies the individual presentation-1 and expresses that this is the information object being annotated. This individual is the same used in the Information Realization pattern in Figure 3a. The latter classifies the individual Risk_Society from DBpedia, which thus represents the semantic label. We exemplify the support of our patterns for context and provenance by including information about the author. The MethodRole classifies a ManualAnnotation, and thus expresses that this image was labeled manually. We specify the author of this annotation by classifying some individual carsten using the AuthorRole. The AuthorRole isConceptRelatedTo MethodeRole, expressing that carsten is the author of this manual annotation.

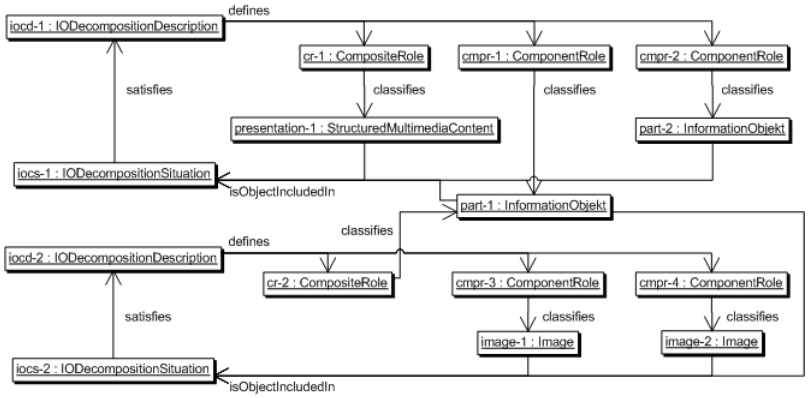
[15] http://semantic-multimedia.org/index.php/M3O:Documentation



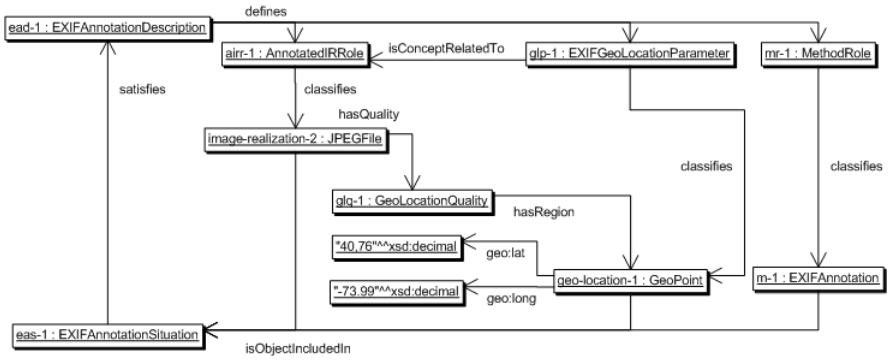
(a) An Example of Information Realization



(b) The Semantic Annotation of the Presentation



(c) A Two-Layered Decomposition



(d) Annotation with Geo-Coordinates based on EXIF

Fig. 3: Example Instantiations of our Patterns Based on the Scenario in Section 2.

Subsequently, we present the decomposition of the presentation into logical components that we want to annotate further. We can describe the decomposition both on the information object level and on the information realization level. However, in this paper we focus on the information object level. In Figure 3c we show the logical decomposition of the presentation into two parts representing the positive and negative aspects of nuclear energy, respectively. We further demonstrate the decomposition of the first part into the two images of Albert Einstein and the Times Square.

The upper part of Figure 3c shows the first composition, the lower half the second one. We see that the IODecompositionDescription defines the CompositeRole and two ComponentRoles. The CompositeRole classifies the individual presentation-1, which is again the information object representing our presentation. The ComponentRoles classify the two InformationObjects named part-1 and part-2, representing the two logical parts of the presentation. The lower half shows how the first part of the image, represented by part-1, which is further decomposed into the two images present in this part, represented by image-1 and image-2. The individual part-1 plays the ComponentRole in the first composition and the CompositeRole in the second one.

Finally, we demonstrate the annotation of an image file with EXIF metadata. Please note that we attach the EXIF descriptor to the realization image-realization-2, which represents the JPEG file realizing the image from the Times Square. The basic pattern is the same as in the example of the semantic annotation. Annotating an information entity with low-level or semantic metadata follows the same underlying structure and only the kind of metadata is different. We use an EXIFAnnotationSituation that satisfies the EXIFAnnotationDescription in order to represent that this annotation is an EXIF descriptor. The description defines a EXIFGeoParameter that parametrizes a GeoPoint, which is the Region. In order to represent the coordinates, we employ the Data Value Pattern, attaching latitude and longitude using the WGS84 vocabulary, i.e., geo:lat and geo:long [17] and use a GeoLocationQuality as the quality of the image.

7 Conclusions and Future Work

In this paper, we presented the Multimedia Metadata Ontology (M3O) that aims at capturing the structural elements of today's multimedia metadata models and metadata standards. The M3O introduces core ontology patterns for annotations and decomposition of multimedia content. It clearly distinguishes between the information object and its realization. It supports both the representation of high-level semantic annotation with background knowledge as well as the annotation with low-level features extracted from the multimedia content. With the M3O, we can better describe multimedia content and integrate the metadata provided with today's models. The current patterns presented are available in OWL at `http://m3o.semantic-multimedia.org/ontology/2009/09/16/`.

Future work is to demonstrate the general applicability and support for the different aspects of today's metadata models by providing a set of default modules covering, e.g., the well established EXIF standard and rich semantic anno-

tation. We also need to integrate further aspects of existing conceptual models [10–12]. It is also aimed at supporting new requirements that may occur in future.

Acknowledgements: We thank Frank Nack for discussing the features and concepts of the MPEG-7 metadata standard. This research has been co-funded by the EU in FP6 in the X-Media project (026978) and FP7 in the WeKnowIt project (215453).

References

1. Technical Standardization Committee on AV & IT Storage Systems and Equipment: Exchangeable image file format for digital still cameras: Exif Version 2.2. Technical Report JEITA CP-3451 (April 2002)
2. MPEG-7: Multimedia content description interface. Technical report, Standard No. ISO/IEC n15938 (2001)
3. International Press Telecommunications Council: "IPTC Core" Schema for XMP Version 1.0 Specification document (2005)
4. Gangemi, A., Presutti, V.: Ontology Design Patterns. In: Handbook on Ontologies. 2nd edn. Springer (2009)
5. Scherp, A., Jain, R.: An ecosystem for semantics. IEEE MultiMedia **16**(2) (2009) 18–25
6. Borgo, S., Masolo, C.: Foundational choices in DOLCE. In: Handbook on Ontologies. 2nd edn. Springer (2009)
7. Adobe Systems Incorporated: XMP – Adding Intelligence to Media (September 2005)
8. Hunter, J.: Enhancing the semantic interoperability of multimedia through a core ontology. IEEE Transactions on Circuits and Systems for Video Technology **13**(1) (January 2003) 49–58
9. Arndt, R., Troncy, R., Staab, S., Hardman, L., Vacura, M.: COMM: designing a well-founded multimedia ontology for the web. In: The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007. (2007) 30–43
10. Markkula, M., Sormunen, E.: End-user searching challenges indexing practices in the digital newspaper photo archive. Information Retrieval **1**(4) (January 2000) 259–285
11. Hollink, L., Schreiber, A.T., Wielinga, B.J., Worring, M.: Classification of user image descriptions. International Journal of Human-Computer Studies **61**(5) (November 2004) 601 – 626
12. Hollink, L., Schreiber, G., Wielinga, B.: Patterns of semantic relations to improve image content search. Web Semantics: Science, Services and Agents on the World Wide Web **5**(3) (2007) 195–203
13. Boll, S., Bürger, T., Celma, O., Halaschek-Wiener, C., Mannens, E., Troncy, R.: Multimedia Vocabularies on the Semantic Web. Multimedia Semantics Incubator Group Report (XGR) (July 2007)
14. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F., eds.: The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press (2003)
15. Manola, F., Miller, E.: RDF Primer (February 2004)
16. Biron, P.V., Malhotra, A.: XML Schema Part 2: Datatypes Second Edition, W3C Recommendation. (October 2004)
17. Brickley, D.: Basic Geo (WGS84 lat/long) Vocabulary (2006)

A Bandit's Perspective on Website Adaptation

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

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

Abstract. Ubiquitous access to websites stresses the importance of adaptation for modern websites. The design and management of decision-taking adaptation engines has become a major research challenge. This paper proposes a bandit-based adaptation decisional model and shows how to describe and manage adaptation policies using a lightweight XML-based description language. The model allows a web marketeer to easily design and deploy adaptation policies. Experimental results on a real website show the effectiveness of our model.

1 Introduction and Problem Statement

A tremendous amount of information is available online. Web sites provide a wealth of services, ranging from information broadcast, to electronic commerce or on-line learning. Web sites have become increasingly complex and are now facing an important challenge: ubiquitous access. Indeed, users reach the web from diverse contexts (from a desktop, on the go etc.), and use a variety of terminals and networks. As humans, users have diverse expectations and behaviors while accessing online services. Providing users with the best experience raises many challenges that are being addressed by web-site and, more generally, multimedia adaptation [1–4]. The user is arguably the most important component of the environment. Therefore, recently a lot of research has focused on user-aware multimedia adaptation.

On the web, marketeers are working hard on increasing the effectiveness of web sites. This means, for information sites, making information easily browsable and provide users with information tailored to their needs. For commercial sites, this means increasing the match between the products or services the user is looking for and their characteristics, quality and quantity. Recommendation or advertising related products on a commercial web site is an example of an adaptation tool that a web marketeers can use [5–8]. Generally speaking, web marketeers aim at optimizing the effectiveness of a given website by taking into account the population of users, the website content and the effectiveness criteria (e.g. sales maximization). They also need to have a measure of impact that provides feedback on the quality of the adaptation strategy (called adaptation policy). This paper proposes a system that helps web marketeers to design and deploy adaptation policies, that can range from fully static (statically defined beforehand and applied without change to a whole group of users) to very dynamic (evolving with time and/or finely tailored to individual users).

A natural way of expressing information about adaptation policies is to model them as website metadata. In this work we propose flexible, XML data structures that allow for easy management of website adaptation policies. The associated policies we introduce in this paper are derived from a simple decisional technique (the so-called Bandit problem).

The next section lays the basis of this work. Our Bandit based framework is presented in Section 3 while Section 4 details a practical implementation. We consider a real-life adaptation problem dealing with optimal delivery of richmedia banners. Section 5 concludes the paper and brings avenues for future work.

2 An adaptation architecture

2.1 Metadata and decision-taking

Achieving interoperable access to distributed richmedia content by shielding users from network and terminal heterogeneity starts with a formal description of the delivery context. Since this description must be interoperable in itself, the role of metadata standards is prominent [9, 10] . Many standards exist. The CC/PP (Composite capabilities/Preference Profiles), the MPEG-21 UED (Usage Environment Description) or the DCO Delivery Context Ontology are compared in the survey of Timmerer and al. [11].

Describing the context is necessary but not sufficient for deciding how to perform adaptation. The decision-taking is a key component for context-aware adaptation [1] and many decisional models have been devised [2, 3]. Picking up a feasible solution is very different than selecting the optimal adaptation decision. Figure 1 introduces a possible architecture for adapting the content in an optimal way.

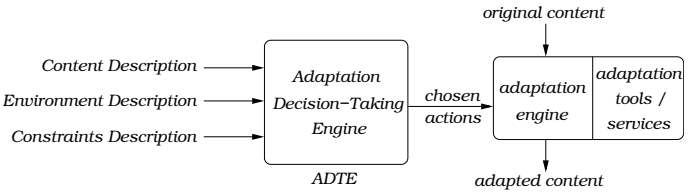


Fig. 1. Decision-taking agent

The adaptation decision-taking engine (ADTE [3]) takes a context delivery descriptor as input. This input descriptor is threefold in figure 1 and divided in content, environment and constraints sub-descriptions. The ADTE outputs a decision that is forwarded to the adaptation engine (AE). The actual transformation of the content is performed thanks to available services on the AE.

2.2 Handling dynamic environment in closed-loop

The previous architecture can be generalized in highly-varying delivery context. In this case many context features vary dynamically: both the available resources and the user intentions may change at any time. The decisional agent, like the ADTE, must dynamically react to context variations.

An even more sophisticated solution would be to use a learning agent that should itself learn from its interactions with the context and improve gradually. We have already proposed a closed-loop approach in that respect [4]. The main idea is to add a feedback channel to control the open-loop architecture shown in Figure 1. In this approach, we consider an adaptation agent and its environment. This environment abstractly integrates the multimedia content, the user, his mobile terminal, the available network and so on. The three steps of the closed-loop are the following:

- **Perception.** At each instant, the decision agent perceives (at least partially) the current characteristics of the delivery context. The current state of the agent in its environment is built upon these perceived characteristics.
- **Action.** Among various possible adaptation decisions it can make in a given state, a learning agent tries to choose the best action to generate as output (in line with Figure 1).
- **Feedback.** The action changes the state of the environment (the adaptation does influence the user, the subsequent resources, etc.) and the value of this transition is communicated to the agent through a reward (a scalar "reinforcement signal"). The agent policy should choose actions that tend to increase the long-term summation of rewards. It can learn to do this by reinforcing decisions that resulted in good accumulation of rewards and, conversely, by trying to avoid unfruitful decisions.

Such an agent learns over time by reinforcement (or by trial and error) [12]. A difficulty is that the agent must explicitly *explore* its environment to estimate the utility of taking actions in all reachable states. Intuitively each state must be visited, each action must be evaluated before converging to the best policy that maps states to optimal actions. There is then a fundamental trade-off between exploration and exploitation. The dilemma the agent faces at each trial is between "exploitation" of the current "best action" that has the highest expected payoff and "exploration" to get more information about the expected payoffs of the other actions.

2.3 Issues

Formally, a learning agent will be defined by a discrete set of environment states, a discrete set of adaptation actions and a reward function that outputs a value for an action in a given state. The main issue is to model a real-life problem using this formal ingredients. Crucial choices must also be made with respect to the problem structure. Either we have an independent problem for each state or we do not. In the latter case, the transition rules between states must be modeled or learnt.

3 A Multi-Armed Bandit Solution

As explained above, the reinforcement learning approach fits well with dynamic multimedia adaptation. Indeed, we applied this framework to several problems, such as ubiquitous streaming [4] and user-aware adaptation [13, 14]. Until now, we explicitly took into account the relations and the transitions between decisional states. This led us to model our adaptation problems with so-called Markov Decisional Processes [12]. In this paper, we assess a simpler decisional model: the multi-armed Bandit problem. This model, although less general, is expected to be much easier to use in practice. In particular we show that it can be easily declared in a simple XML format.

3.1 A unified decisional state

We first capitalize on previous ideas and introduce a unified decisional state. A state of -(the agent in)- the delivery context is a triplet which integrates:

- some current observations,
- some inferred information,
- an amount of memory.

The current observations are factual elements that allow to describe unambiguously the adaptation target and other features of the adaptation problem. The website content to be adapted, the associated URL, the time within current web session, the characteristics of the terminal may be parts of these observations [4, 14]. The inferred information are composed of partially observable metadata. We called them "subjective semantic descriptors" in [10] (by contrast with "objective" observations). The user "interest level", the "importance" of a given media are contextual metadata that can only be approximately inferred. The memory of the decisional state finally allows to satisfy a self contained property. Intuitively, the aim is to retain all relevant information from the past. If a given adaptation has already been performed, the idea is to adapt the current content in different way. Memorizing previous decisions helps to do this.

It is now straightforward to argue for the use of a Multi-Armed Bandit Problem with such a decisional state.

3.2 The Bandit and the adaptation actions

A Multi-Armed Bandit Problem (MABP) is named by analogy to a slot machine. For example, in the K-Armed Bandit Problem, a gambler has to choose which of the K slot machines to play. At each time step, he pulls the arm of one machine and receives a reward. His aim is to maximize the sum of rewards he perceives over time. This clearly shows the "exploration vs exploitation" dilemma: the purpose of the gambler is to find, as rapidly as possible, the arm that gives the best expected reward.

In order to solve a Bandit problem (i.e. to find the best arm to play), various strategies can be used [12]. Recent research has proposed various solutions to

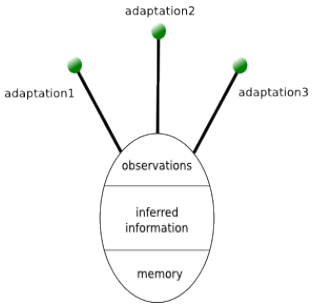


Fig. 2. A decisional state (depicted as an ellipse, containing the information triplet describing the context) and the associated 3-armed Bandit problem (adaptation 1, 2 or 3 can be performed).

solve optimally and online this dilemma, minimizing the number of errors over time. One of the most efficient is called Upper Confidence Bound (UCB, [15]). At each play, it computes a priority index for each arm, based on the previous rewards and the number of times it has already been invoked. The index p_j for arm j is given by

$$p_j = \bar{x_j} + \sqrt{\frac{2ln(n)}{n_j}} \tag{1}$$

where $\bar{x_j}$ is the average rewards obtained from arm j, n_j the number of times arm j has been chosen and n the overall number of plays done so far. The best arm to choose is the one with the highest priority.

For our adapation problem, we use MABPs as follows:

- we strategically define a set of decisional states,
- we associate one MABP to each state of the context,
- each arm corresponds to a possible adaptation action in a given state,
- rewards are given by an impact/utility measure.

Figure 2 depicts a 3-armed Bandit problem associated to a state. Pushing arm i at a play means performing adaptation i at this step.

As a result, we get a collection of independent multi-armed Bandit problems. Each MAPB can be solved independently using the UCB technique, at the expense of neglecting potential relations between states.

3.3 Using a declarative language

The model we propose is really simple, and only states and their associated Bandit have to be defined. Thus, we propose a XML-based language that allows web marketeers to easily design and deploy adaptation policies.

The first thing we have to do is to define and declare a state. A state (figure 3), as said earlier, is composed of three parts: observations, inferred information

and memory. Observations are described as a set of observable elements (current page, terminal information, etc.), memory as a set of past taken actions. For inferred information, a handle (e.g. a an Uniform Resource Identifier (URI) to a service able to produce this information by inference (e.g. by analyzing the logged behaviors of website visitors) must be provided.

```
<state id="s1">
  <observations>
    <observation>terminal</observation>
  </observations>
  <inferences>
    <inference>user-activity</inference>
  </inferences>
  <memory>
    <action>banner3</action>
  </memory>
  <bandit-ref>bs1</bandit-ref>
</state>
```

Fig. 3. An XML description of a state ($s1$)

Then, we need to describe the possible adaptations in this state, that is to say the MABP associated to the state. A MABP (figure 4) contains a set of actions. Each action is composed of a handle to the corresponding adaptation engine and the number of times it has been invoked. This last number allows an ADTE to compute the priority index given by UCB (equation 1) in order to choose a correct action. A reward handle is also given in order to measure the impact (or utility) of the chosen action.

```
<bandit id="bs1">
  <actions>
    <action>
      <engine>banner3-injector</engine>
      <nbInvok>0</nbInvok>
    </action>
    ...
  </actions>
  <reward>session-duration</reward>
</bandit>
```

Fig. 4. An XML description of a Bandit problem associated to state $s1$

3.4 Software Architecture

Based on the previous XML description of our model, we propose in figure 5 a software adaptation architecture.

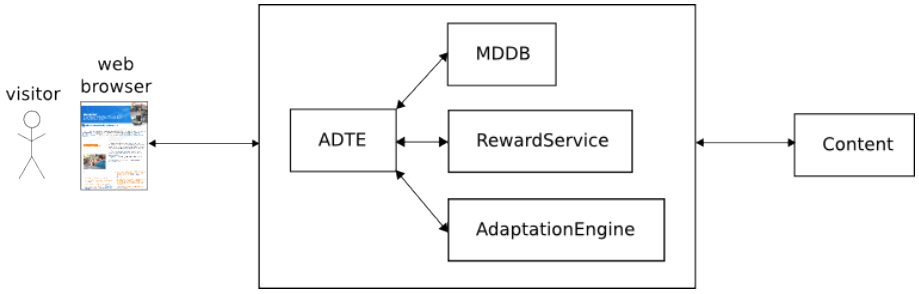


Fig. 5. The software architecture

A visitor is browsing a website using his favorite browser. While browsing the site, he is monitored by the *ADTE*, that observes the usage and builds a state using direct and inferred information and past adaptation actions. Our observation platform has already been presented in our community [16]. When a new state is computed, the ADTE asks *the metadata database (MDDB)* and eventually gets an associated Bandit problem. If so, it computes the priority indexes using UCB (equation 1) and chooses the action to perform. This action is then transmitted to *the adaptation engine* that actually performs it. Finally, using *the reward service*, it gets a reward qualifying the degree of success of the taken adaptation decision. The ADTE updates the Bandit information (mean reward, number of times an action has been played) and sends it to the MDDB. Thus, the updated information will be used for next visitors who would be in this state.

4 Case study

In this section, we describe how to use our adaptation model on a real website[3]. The site we choose is a collaborative website and it is designed to be used in a professional environment. It is organized in different workspaces, composed of various sections (blogs, wikis, portal, file sharing, etc.), allowing information to be produced and shared by company employees.

4.1 User aware banner service

Adaptations on this website consist in adding personalized banners that recommend "hot" parts of the site to users (e.g. a new blog entry, a modified wiki page or an uploaded file).

This adaptation problem can be seen in two dimensions: the content to recommend and the format of the banner. In this study, we intentionally put aside the banner content production issue and consider it, as other authors (e.g. [8]),

[3] http://www.linkforus.org

as a separate question. As a result, we choose to use the existing RSS feed as the content provider for recommendations. Concerning the format of the banner, three types of banner formats are available (figure 6):

- the basic version, only composed of a text (including links to recommended content),
- the video/avatar version. In that case, an avatar in a video serves as a teaser,
- the 3D version. It uses a "carousel" component written in Adobe Flash. Recommendations are included in the different facets.

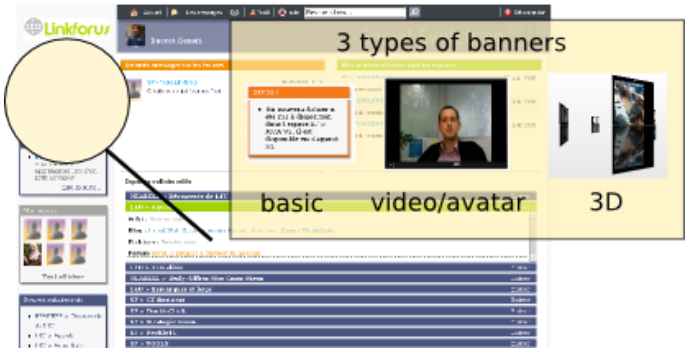


Fig. 6. The different types of banners

In line with the work of [17], we choose to display banners in sequence. A sequence contains exactly one banner of each format. Therefore, six sequences are available (basic/video/3D, video/basic/3D, etc.). Thus, adaptations will consist in displaying the banner in a certain order.

Among many possibilities, we chose to use as an objective/target of the adaptation to get users stay longer on the site, i.e. increase their session duration. We naturally use the session duration as an impact measure (reward) for a given sequence.

4.2 A simple MABP instance

We have to set the different states and the associated Bandit problems.

For the state, the simplest solution is to use a single state. It only contains an inferred information: the user "activity" (i.e. the number of events produced by the user in a given time window) on the website. When the user activity decreases, a banner is displayed on the site according to the chosen sequence of banners.

As for the associated Bandit problems, actions are the different sequences to display, and thus we consider a 6-armed Bandit problem. The (stochastic) reward is given by the session duration.

4.3 Results

In order to get validation, we use a navigation simulator that has been seeded with the data collected from a previous work ([17]) in which we also have sequences of three similar banner types. We have already concluded which sequence of banners is the best (video/3D/basic). Using this "ground truth", we want to determine whether the Bandit problem rediscovers or not this conclusion.

Figure 7 presents the evolution of priority indexes values for each arm of the associated Bandit. As the number of times the Bandit is invoked gets higher, priority indexes decreases. However, while zooming, we notice that the one for sequence video/3D/basic (bold black line) is often greater than the others. It means that the associated arm is pulled more frequently. Using the ground truth, we realize that, indeed, this adaptation action is better that the others with respect to the session duration.

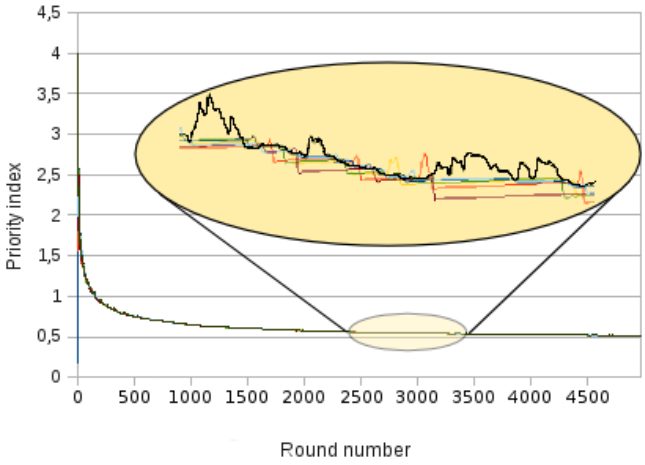


Fig. 7. Evolution of priority indexes for each arm of the Bandit problem.

Figure 8 shows the percentage of the optimal action the Bandit has been chosen in function of the number of plays. Let us recall that we know which action is the best thanks to the original dataset. As the number of plays gets higher, the percentage increases, indicating the efficiency of the Bandit strategy. Interesting results are reached after around 10.000 sessions. On our test website, this can be realized in less than a month.

Bandit problems allow us to draw a conclusion similar to [17]: the usefulness of considering sequence for improving the effectiveness of banners. Results show that an improper format of a banner in the beginning of the sequence wipes out the positive effects of the subsequent banners. These results demonstrate the power and simplicity of a Bandit-based adaptation strategy.

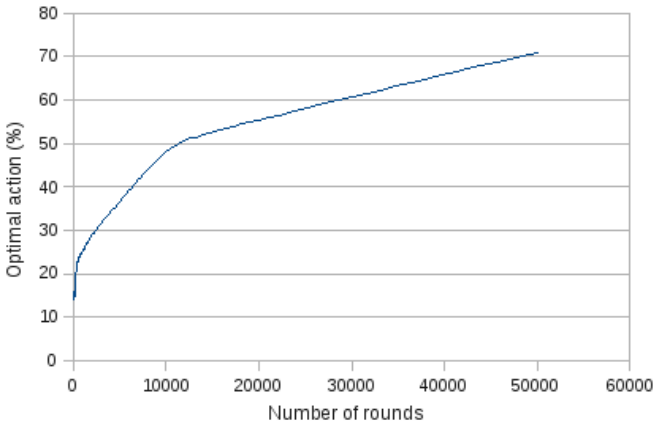


Fig. 8. Percentage of optimal action chosen

5 Conclusion

Decision-taking engines are an important component of adaptive web sites. To help in designing effective adaptation systems, this paper contributed with a bandit-based model for the website adaptation problem. We proposed a declarative, XML-based language for expressing and managing a multi-armed bandit decision model. In order to experimentally validate our model, we presented a case study that shows the use of the model on a real website. Our model proves to be a simple, lightweight decisional model. It provides finely tunable yet simple to use adaptation policies.

Our model gives web marketeers a great flexibility while managing a policy. It is possible to choose/concentrate on a limited number of states (the state space is virtually very large). On these selected states, it is possible to configure the system to only use a selected/useful subset of all possible adaptation actions.

The perspectives of this work are twofold. First we would like to investigate networks or trees of bandit problems. This would be a possible solution to partially relate a given decisional state to subsequent ones while avoiding the complexity of Markov Decision Processes. Secondly, we believe that the Upper Confidence Bound (UCB) algorithm, when applied to trees (the so-called UCT), would be a good candidate to solve our adaptation problems in that case.

References

1. Mukherjee, D., Delfosse, E., Kim, J.G., Wang, Y.: Optimal adaptation decision-taking for terminal and network quality-of-service. IEEE Transactions on Multimedia **7**(3) (2005) 454–462

2. López, F., Martínez, J.M., Valdés, V.: Multimedia content adaptation within the cain framework via constraints satisfaction and optimization. In: Adaptive Multimedia Retrieval. (2006) 149–163
3. Jannach, D., Leopold, K., Timmerer, C., Hellwagner, H.: A knowledge-based framework for multimedia adaptation. Appl. Intell. **24**(2) (2006) 109–125
4. Charvillat, V., Grigoras, R.: Reinforcement learning for dynamic multimedia adaptation. J. Network and Computer Applications **30**(3) (2007) 1034–1058
5. Pandey, S., Olston, C.: Handling advertisements of unknown quality in search advertising. In: Twentieth Annual Conference on Neural Information Processing Systems (NIPS). (2006)
6. McCoy, S., Everard, A., Polak, P., Galletta, D.F.: The effects of online advertising. Commun. ACM **50**(3) (2007) 84–88
7. Attenberg, J., Pandey, S., Suel, T.: Modeling and predicting user behavior in sponsored search. In: KDD. (2009) 1067–1076
8. Hauser, J.R., Urban, G.L., Liberali, G., Braun, M.: Website morphing. Marketing Science **28**(2) (2009) 202–223
9. Kosch, H., Böszörményi, L., Döller, M., Libsie, M., Schojer, P., Kofler, A.: The life cycle of multimedia metadata. IEEE MultiMedia **12**(1) (2005) 80–86
10. Lux, M., Granitzer, M., Spaniol, M., eds.: Multimedia Semantics - The Role of Metadata. Volume 101 of Studies in Computational Intelligence. Springer, Berlin (August 2008)
11. Timmerer, C., Jabornig, J., Hellwagner, H.: Delivery context descriptions - a comparison and mapping model. In: Proceedings of the 9th Workshop on Multimedia Metadata (WMM'09). (2009)
12. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press (1998)
13. Plesca, C., Charvillat, V., Grigoras, R.: A formal framework for multimedia adaptation revisited: a metadata perspective. In: BTW Workshops. (2007) 160–178
14. Plesca, C., Charvillat, V., Grigoras, R.: Adapting content delivery to limited resources and inferred user interest. International Journal of Digital Multimedia Broadcasting **2008** (2008) doi:10.1155/2008/171385
15. Auer, P., Cesa-Bianchi, N., Freund, Y., Schapire, R.E.: The nonstochastic multi-armed bandit problem. SIAM Journal on Computing **32**(1) (2003) 48–77
16. Baccot, B., Charvillat, V., Grigoras, R., Plesca, C.: Visual attention metadata from pictures browsing. In: Ninth International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS'08. (May 2008) 122–125
17. Baccot, B., Choudary, O., Grigoras, R., Charvillat, V.: On the impact of sequence and time in rich media advertising. In: Proceedings of the 17th ACM Conference on Multimedia. (2009)

Intent Tag Clouds: An Intentional Approach To Visual Text Analysis

Fleur Jeanquartier[1,2], Mark Kröll[1], Markus Strohmaier[1,2]

[1]Graz University of Technology. Inffeldgasse 21a, 8010 Graz, Austria and
[2]Know-Center. Inffeldgasse 21a, 8010 Graz, Austria.
{fjeanquartier,mkroell,markus.strohmaier}@tugraz.at

Abstract. Getting a quick impression of the author's intention of a text is an task often performed. An author's intention plays a major role in successfully understanding a text. For supporting readers in this task, we present an intentional approach to visual text analysis, making use of tag clouds. The objective of tag clouds is presenting meta-information in a visually appealing way. However there is also much uncertainty associated with tag clouds, such as giving the wrong impression. It is not clear whether the author's intent can be grasped clearly while looking at a corresponding tag cloud. Therefore it is interesting to ask to what extent, with tag clouds, it is possible to support the user in understanding intentions expressed. In order to answer this question, we construct an intentional perspective on textual content. Based on an existing algorithm for extracting intent annotations from textual content we present a prototypical implementation to produce intent tag clouds, and describe a formative testing, illustrating how intent visualizations may support readers in understanding a text successfully. With the initial prototype, we conducted user studies of our intentional tag cloud visualization and a comparison with a traditional one that visualizes frequent terms. The evaluation's results indicate, that intent tag clouds have a positive effect on supporting users in grasping an author's intent.

Key words: Intent Tag Cloud, Usability Study, Visual Text Analysis

1 Introduction

Ongoing developments in the field of information visualization on the web support visual tasks in various ways. A technique called tag clouds has become a quite familiar technique to visualize textual data on many websites and many users know how to use it. Figure 1 shows two tag clouds. Tag clouds can be used in various ways to help users in getting a quick overview. Imagine that a user wants to visit a website to read an online article or essay but before would like to know what the text is about or moreover even know what the meaning and purpose of the text is. This means we need a simple methodology and application for supporting the task of both understanding a text as well as having an idea what the author(s) intended to communicate. The reader would like to sense what the authors meant or implicated when they created a specific text. This is a common problem that has no clear solution yet. According to [9], an

author's intentions are crucial for understanding the meaning of a (speech) text. Therefore our approach is not only to sum up a text but further trying to explain it. We assume, it is possible to visualize specific information in a way the readers are able to grasp both the meaning and purpose of the text. We therefore developed intent tag clouds as a research prototype for improving the process of successfully understanding a text.

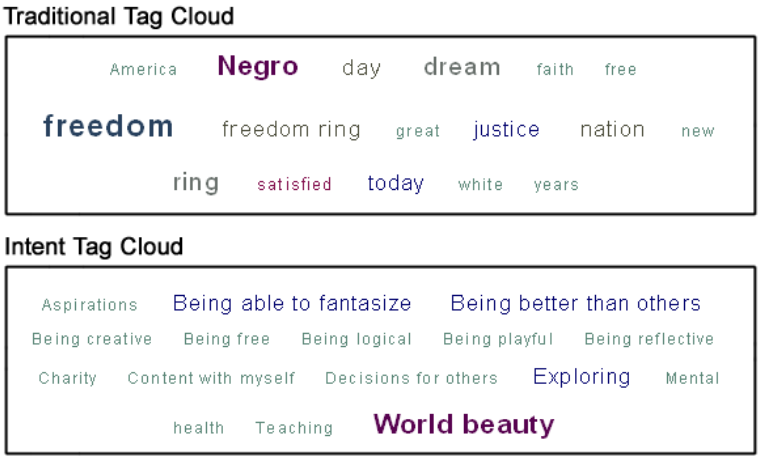


Fig. 1. Two Tag Cloud Versions of M. L. Kings famous speech in 1963

However, the use of traditional tag clouds is also a controversial topic due to the fact, that tag clouds may provide a wrong impression or just do not fullfill the task of giving a quick impression of an author's intentions of a text. Moreover, research within this field still lacks user evaluations. This paper addresses the question, how to support the user in understanding a text and its intentions and explores the usage of tag clouds to provide an intentional perspective on the textual content.

2 Related Work

A tag cloud is a non-hierarchical presentation of linked terms [10]. A tag cloud is also described as a visualization of word frequencies [16].
The author of [15] recapitulates the history of tag clouds insofar as he argues, that the basic look of a tag cloud, namely a combination of many different type sizes in a single view, goes back to the early 20th century. This visualization technique has first been introduced outside of academic circles, namely on the popular website called Flickr[1] as described in [10].
The authors of [1] state that the motivation for tagging also changed with the flickr online community. They show that users can be motivated to annotate

[1] Flickr - Photo Sharing: `http://www.flickr.com`

content. Both the before mentioned aspects as well as the increasing incentives for tagging result in an increasing number of online annotations.
For now, there are many different kinds and variations of tag clouds that are currently available: Such as improvements over traditional tag clouds presented by [16], include the ability to measure the frequency of two word tags in a text and to dynamically filter the tag cloud by entering query strings. The work of [6] presents a different tag cloud layout to improve information retrieval based on clustering of similar tags. The paper [15] shortly presents some non-traditional tag clouds such as a time-based one.Yahoo Research created the geographic tag visualization Tagmaps, a world exploration tool as described in [18]. The same authors also created the so called Taglines[2] which is an online tool demonstrating some novel contributions for expressing timescales to generate the possibility to navigate through the interesting tags for a particular period of time [5]. Alternative ways, where intent annotations can play a major role in supporting a user's understanding, including results presented in [13] that show how capturing aspects of intent rather than content can support social software. The work of [11] explores the way how users express their intentions in digital photo search. Such works indicate that user intentions may also play a role in multimedia retrieval and context different from textual content as well.

2.1 Discussion Of Tag Cloud Visualizations

Research has shown, that tag clouds can have positive effects on basic visual tasks due to the layout's compactness, due to it's ability to show more dimensions (alphabetically, size and items) at once due to the fact, that within tag clouds, users are able to quickly identify the most frequent term etc [7]. Therefore tag clouds are scannable, offering good overview. Compared to [17] where it is shown that users read about 20% of the text on the average page, these positive effects of tag clouds appear useful. Moreover, the work by [12] shows that tag clouds can support many user tasks such as providing an overview and general impression of the underlying data set. [10] also shows that tag clouds are good for prototyping because of the easy implementation. Other visualization techniques are more complex. Last but not least research, such as [8] has shown that tag clouds are useful for social information such as showing human behavior and reflect human mental activities.
Next to the already stated positive effects of tag clouds there are also a few drawbacks. The authors of [7] and [8] show that longer words grab more attention than shorter ones. Moreover, there is also no meaning in visual proximity and therefore meaningful associations are lost. Last but not least visual comparisons are difficult, The work by [8] even suggests to compare also other research results such as proposals by E.Tufte.
As summarized before, there are points of criticism for tag cloud discussions. However, many of these can be addressed simply by visualization enhancements regarding tag positioning, tag sorting, tag normalization as well as aesthetic

[2] TagLines: `http://research.yahoo.com/taglines/`

considerations. The author of [3] also states that tag clouds are only one specific kind of weighted lists. There are many kinds of mappings from visual features to underlying data that have not yet been exploited. Bumgardner [3] suggests trying out different mappings such as mapping font size to time or using older-fashioned fonts for older data. The authors of [4] describe a Yahoo project that makes use of the Flickr service. Their approach was that any user may append a tag to any photo in the system. There are also existing guidelines for tag cloud construction and comparisons between semantic arrangements, alphabetic and random tag layouts, such as described in [12]. Enhanced tag clouds then guarantee scannability and visual appeal. Some of these methods were also used while usability inspections on our intent tag cloud prototype revealed some needs for improvements.

2.2 Research Rationale & Setup

However, all these studies did not try to clarify whether the user clearly grasps the author's intent of a text, nor tried to support the user in understanding a text successfully. Furthermore no user-based evaluations of intent annotation approaches have been conducted yet. Therefore it is our aim to to answer the question, how to best support the user in successfully understanding a text respectively in determining the author's intentions corresponding to a given text. As knowledge of intentions is relevant for interpreting text, we try not only to sum up a text, but also visualize information in a way the readers are able to grasp the meaning and purpose the author intended to communicate with the given text.
We explore the usage of tag clouds to provide an intentional perspective on the textual content. The authors of [14] demonstrated how to automatically annotate textual resources with human intent. We try to make use of this novel idea of intent annotation and present an approach making use of tag clouds for presenting the author's intentions of a speech text. In other words, we propose visualizing such intent annotations instead of traditionally visualizing a tag cloud based on term frequency. Understanding these design enhancements may allow interface customization that could further improve the task of keeping the user informed.

3 Intent Tag Cloud Prototype

In figure 1 we show two versions of tag clouds. The top tag cloud shows a traditional tag cloud consisting of frequent terms. The bottom tag cloud is containing intent tags. These tag clouds are one part of the output of our prototypical implementation that can be seen in figure 2. The implementation and benefits of the intentional visualization approach are further described below:

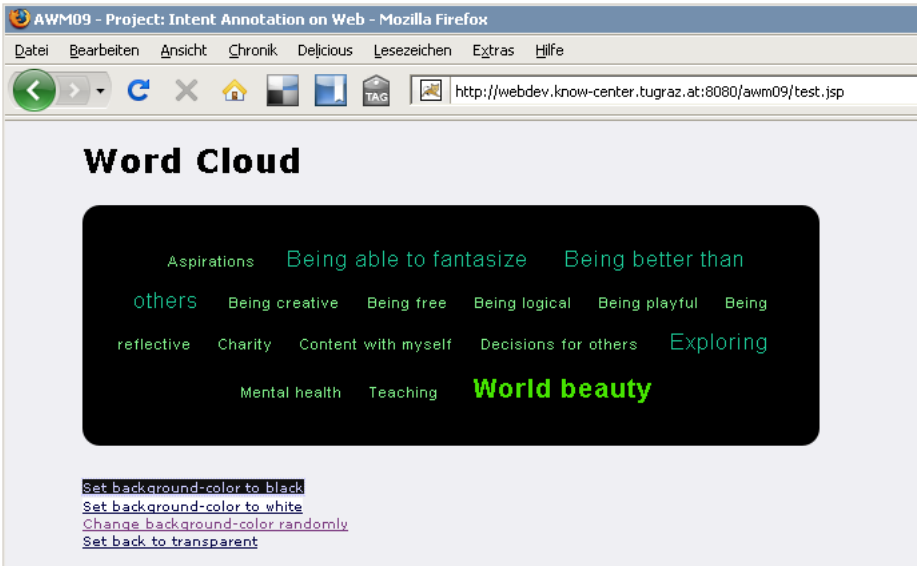


Fig. 2. Screenshot of the prototype used in the experiment

3.1 Intent Annotation

Existing tag suggestion approaches mainly focus on annotating a document according to its most predominant subject matter such as using frequency terms to show what a text is about (e.g. 'sport', 'politics'). In contrast, the authors of [14] describe the annotation of resources according to the intentions, such as showing what goals a resource is about (e.g. 'Achieve Happiness' or 'Maintain Good Health'). According to [14], intent annotations deal with future states of affairs that someone would like to achieve (in contrast to topic, sentiment or opinion tags). In [14], the authors explore the use of indicative actions as a proxy for inferring intentions from textual resources. Therefore intent annotation can be understood as the problem of identifying a set of adequate intent annotations for each and every action indicative of intent in a given textual resource. The basic concepts of intent annotation itself and the automatic extraction approach is summarized within [14].
The algorithm described there is only one possible way how intent annotations can be generated. Also the already mentioned work of [13] shows another possibility. However this paper focuses mainly on exploring the usage and benefits of visual interfaces for intent annotations; the generation of the intent tags is not the focus of our investigations here.

3.2 Implementation of the Intent Tag Cloud

The simplified prototype in Figure 2 has been used in the experiment. The original prototype also includes some other interaction possibilities such as not only

using sample speech texts but also using own text-input and changing the visualization in size and color. First of all the interface takes a speech text as input. After making use of the automatic extraction mechanism described in [14], a weighted list of intent annotations is created. The list's terms are organized insofar as the listed intent annotations are tag pairs, consisting of both the tag-name as String and a weight-level of type Double. This tag-list is then used for the creating the intent tag cloud visualization. For the later formative testing [2] a traditional tag cloud is also generated, making use of a web-based tool for generating a cloud of frequent terms from a given text.
The Design Process was an iterative one, including prototyping and usability evaluation. While usability inspections methods have been applied, several designs for the prototype have been created and the intent tag cloud has been upgraded consistently. At the very beginning, when the first prototype version went online, the visual interface has been inspected using web usability heuristics. In the first prototype versions, no user tests have been conducted due to the early stages. These inspections have led to enhancements of the tag cloud visualization itself, such as the font-family, letter-spacing, positioning and colouring, but also including interface enhancements, such as the web form's usability including the possibility of changing parts of the visualization dynamically as well as providing an example-text.
Last but not least, after further enhancing the tag cloud visualization, a usability test was planned and conducted. Test focus was the visualization's usability, its effectiveness and its benefits to answer the research question addressed. Selected participants did comparisons between a simple traditional tag cloud and an intentional version. After successfully executing the formative test, using a simple questionnaire, a final analysis and conclusion was produced.

3.3 Evaluation Setup

To test the the intent tag cloud prototype, a formative test has been conducted to evaluate the prototype's usability, its precision as well its as informative completeness. The formative test was planned and performed in the following manner:
First of all a testplan has been written and 4 test users have been chosen for participating in a questionnaire. Two tag clouds versions have been created, one version making use of the presented intent tags, the other version visualizing the common frequency tags. Figure 1 shows both of them.
The tag clouds have been created using Martin Luther King's famous speech 'I have a dream' (given on August 28, 1963). Both tag cloud versions are using the same mechanisms for visualization. The different tag levels were represented as XHTML, whereas the tag level was an integer value that ranges from 1 to 11. Depending on the tag's level a CSS selector has been assigned to the various tags. According to the CSS selector, the tag has been styled with a different font-size and -color value as well as given a varying letter-spacing value.
Tag cloud differences existed in particular with regard to term-length and level-variation. In more detail, the traditional tag cloud's tags have been shorter and

consisted only of one word, whereas some of the intent tags have been represented by two word combinations such as 'being playful'. The other difference occured for the tag levels insofar as word frequency calculated levels from 1 to 9 (namely 1,2,3,4,5,7,9) as output, whereas the intent tags varied only between 1, 3 and 7. As a result the traditional tag cloud looked more colourful and dynamic than the intent tag cloud.
After completing the test setup, participants have been chosen and invited to join the questionnaire sessions. All four participants were Austrian, therefore the questionnaire's language was German. The questionnaire included an introductory text, a statement of agreement and several task descriptions and questions. The participants were asked to speak aloud what they think. Some of the questions have been designed as close ones (yes/no) and (1-to-10-selections) as well as some open ones to get qualitative feedback. For detailed information on usability inspection and evaluation methods we used - in addition to our own experiences - primarily the between-groups description of [2].

4 Results

For answering the research questions that have been stated aerlier, i.e. how to support the user in understanding a text and its intentions expressed successfully and are the intentions always scannable for the reader within such a tag cloud, we make use of both initial findings from related work as well as qualitative studies. The results have been collected with an excel sheet and include both qualitative feedback such as a list of interesting participant quotations and suggestions as well as closed answers such as yes/no and numerical answers from one to ten. Figure 3 to 5 show charts that summarizes the collected answers and data.
Formative testing methods usually involve observing a small number of test users [2] using an interface in order to gain more qualitative feedback and insights why something does (not) work as planned. Four participants have been chosen to join the formative testing. Table 1 shows the participants' distribution. All participants are using Computers on daily basis, but all are working in different areas, ranging from medicine and chemistry over design up to administrative fields. participants have been asked to speak aloud and tell us what they are thinking while trying to solve the stated tasks or respectively answer the stated questions [2].

Participants:	**1**	**2**	**3**	**4**
Traditional Tag Cloud Version	x		x	
Intentional Tag Cloud Version		x		x
Age	27	26	29	27
Gender	female	male	male	female
Educational level	Academic	Student	Technical College	Academic

Table 1. Participant Distribution

Some questions were asked during the interview. For instance, the participants asked, when looking on the tag cloud terms: 'What is the striking point?' and 'Could that be a political speech because it is a quite spongy one?'. Table 2 shows a list of all questions.

Introductory Text	**Question**
Look at the tag cloud for 15 seconds.	Please tell us, after this short time, in one or two words, what do you think the author intended to communicate?
Study the tag cloud for at least one minute.	Do you think you understand what the purpose and meaning of the text is? - What is your impression? (unclear = 0, clear = 10)
Please answer in short the following more specific questions:	
	Do you perceive the tag cloud being of avail and helpful? (0 = no, not at all, 10 = yes, quite helpful)
	Would you wish to see such a visualization of meta informationen more often? (1 = yes, 0 = no)
	What do you perceive as positive within the tested tag cloud and what did you perceive as disturbing?

Table 2. Questionnaire Extract (Translated from German)

For example by asking the first question, we tried to understand what is the participants perceived value of the displayed meta information in general.
Do the participants think they have a clear impression of what the author intended to communicate, or are they rather unsure about it? We also tried to answer the question whether the participants thought, they understand the text's meaning and purpose by asking to name us those one or two terms, they think the speech text obviously describes. Figure 3 shows the recapitulated answers as a block chart. As can be seen in this figure the participants 1 and 2 had a quite clear impression of what the text is about, but participant 3 and 4 stated that they had a adequately clear impression. For this particular test case there are no noticeable differences between the tag clouds and their performances regarding the quality of information.

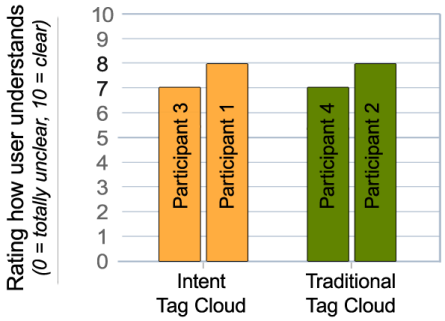


Fig. 3. Do the participants understand what the author intended to communicate?

To learn more about the general tag cloud's readability, another closed question was asked. We wanted to know more about the fact, whether the tag cloud and its tags are easy to read and therefore may support or rather interfere with getting a clear understanding of what the text is about. Figure 4 shows the answers as a horizontal chart.

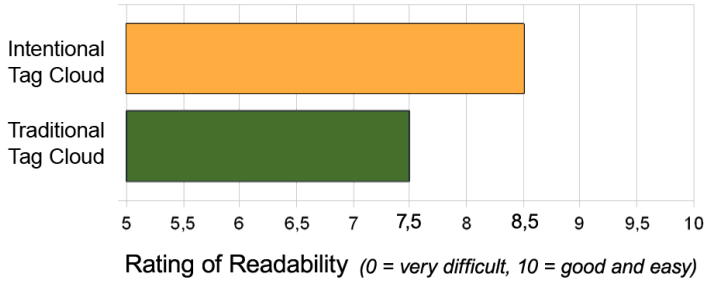


Fig. 4. What is the participants' impression of the readability?

This chart illustrates that the participants mainly agreed on the fact that tag clouds are both readable, while the intent tag cloud testers rated the readability a little bit higher. A participant testing the traditional tag cloud argued that some terms are clearer with a more specific meaning than others. That is why the participant felt not comfortable when deciding which term fits best for describing the meaning and purpose of the associated speech text.
This perception of imbalance when comparing the different terms while trying to choose an appropriate one shows one of the limitations of the term frequency method.

Furthermore the participants have been asked, whether they think that tag clouds were of avail and helpful. They were also asked to state reasons why (not) and how exactly the tag cloud is helpful in successfully understanding the text in their opinion. The participants showed a positive attitude regarding assistance. In the end, the participants were asked to summarize their positive and negative findings. We wanted to get an impression of what are the participant's overall thoughts and feelings about the particular tag cloud. We also used the opportunity to get feedback for future improvements. Figure 5 shows the tag cloud comparison illustrating the sum of Pros and Cons.

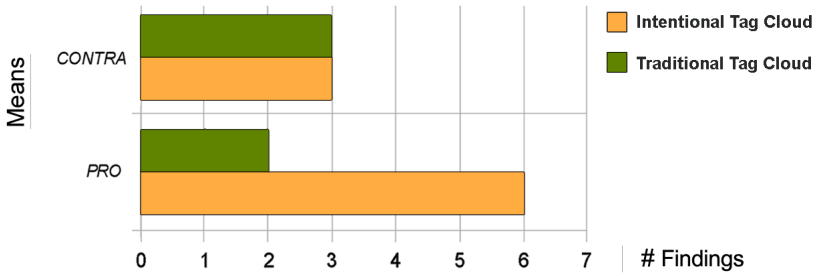


Fig. 5. Amount of Pros and Cons stated by the participants.

This figure illustrates quite concisely the variations in positive findings. Taking all the collected answers into account we assume that the intent tag cloud benefits from interpretation issues. The intent tag cloud terms seem to be clearer and more similar. Especially regarding these differences in semantic density, it can be assumed that a tag cloud of intent tags is a useful approach for describing textual content.
Regarding the experiences made with the formative test, especially the consistently positive answers to the question whether tag clouds were of avail and helpful or not, we can answer the main research question of how to support the user by understanding a text's meaning: Namely the approach of using (intentional) tag clouds appears to support the process of successfully understanding a text. To answer the subsequent question, namely whether the intentions are always scannable for the reader within such a tag cloud or not, we again refer to the formative test results: The answers indicate that intentions are scannable for the reader.
Additionally, the gathered statements during the test also give a good insight to answer the question 'Compared to a content tag cloud, is there a clear benefit in using an intent tag cloud?' and also the subsequent question of 'What are possibilities of further improving the intent annotation visualization?'. First of all due to the varying answers regarding the topic precision, we argue that the traditional tag cloud version, compared to the intent annotation version, is best used when giving a quick overview of what a text is about, whereas enhanced intent tag clouds grant the possibility of spotting and recognizing more precisely a speech text's intentions. The participants who have been using the intent tag cloud version answered in a more focused and specific way and the topic guesses were well chosen. Nevertheless, participants also argued that they are not sure whether the visualization may lack the most important term(s). On the other hand, the traditional tag loud users were quite satisified with their mostly general impressions of what the speech text is about, because they perceived the tag cloud visualization itself as a kind of funny and motivating type of presenting meta information in a concise way. That may be why the intentional version performed well in precision whereas the traditional tag cloud version performed well in delivering a quick and motivating general glance about the text's topic.

4.1 Future Improvements

During the project's evaluation phase a number of ideas have been generated that will be a focus of future work. Among these ideas, answering the question of how to understand a topic change over time has become mostly prominent. Therefore we consider further enhancing our intent annotation visualization and develop a mechanism for visualizing changes over time. We will continue with studying approaches like the one shown in [5] and [15]. Due to the fact that we only investigated one possible way how intent annotations may support the user in understanding a textual content, the studies can be extended to other multimedia content as well.

5 Conclusions & Outlook

In this paper we explored the usability of intentional tag clouds by implementing a prototype and conducting a user study. Results from our user study suggest that intent tag clouds are accepted and support users in analyzing textual content visually. We described one possible way how intent annotations may be used in a supportive way. Though we used a particular automatic extraction technique, it is not essential how intent tags are produced. We hope that this paper is used as an inspiration, how intent annotations can help users in understanding. Referring to the user studies, intent tag clouds might be applied to other online- as well as offline applications. For instance online magazines might benefit by using the proposed tag cloud enhancements for summarizing articles; however, these text summaries may also work with all other kind of text as well as multimedia, such as articles and advertisement in print media and also digital image databases. The intent annotation visualization can also further be extended by integrating new features such as a mechanism for visualizing changes over time.

Acknowledgments.

This work is in part funded by the FWF Austrian Science Fund Grant P20269 *TransAgere*. The Know-Center is funded within the Austrian COMET Program under the auspices of the Austrian Ministry of Transport, Innovation and Technology, the Austrian Ministry of Economics and Labor and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

References

1. Morgan Ames and Mor Naaman. Why we tag: motivations for annotation in mobile and online media. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 971–980, New York, NY, USA, 2007. ACM Press.

2. Keith Andrews. Evaluating information visualisations. In *BELIV '06: Proceedings of the 2006 AVI workshop on Beyond time and errors*, pages 1–5, New York, NY, USA, 2006. ACM.
3. Jim Bumgardner. Design tips for building tag clouds, June 2006.
4. Micah Dubinko, Ravi Kumar, Joseph Magnani, Jasmine Novak, Prabhakar Raghavan, and Andrew Tomkins. Visualizing tags over time. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 193–202, New York, NY, USA, 2006. ACM Press.
5. Micah Dubinko, Ravi Kumar, Joseph Magnani, Jasmine Novak, Prabhakar Raghavan, and Andrew Tomkins. Visualizing tags over time. *ACM Trans. Web*, 1(2), 2007.
6. Yusef Hassan-Montero and Victor Herrero-Solana. Improving tag-clouds as visual information retrieval interfaces. In *Proceedings of Multidisciplinary Information Sciences and Technologies, InSciT2006*, Merida, Spain, October 2006.
7. Marti Hearst. Taxonomy bootcamp '07 keynote talk: Thoughts on social tagging, 2007.
8. Marti A. Hearst and Daniela Rosner. Tag clouds: Data analysis tool or social signaller? In *HICSS '08: Proceedings of the Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, Washington, DC, USA, 2008. IEEE Computer Society.
9. Eric Donald Hirsch. *Validity in Interpretation.* New Haven and London: Yale University Press, 1967.
10. Owen Kaser and Daniel Lemire. Tag-cloud drawing: Algorithms for cloud visualization, May 2007.
11. C. Kofler and M. Lux. An exploratory study on the explicitness of user intentions in digital photo retrieval. In Klaus Tochtermann and Hermann Maurer, editors, *Proceedings of I-KNOW 09*, pages 208–214. Know-Center, Graz, Journal of Universal Computer Science, September 2009.
12. A. W. Rivadeneira, Daniel M. Gruen, Michael J. Muller, and David R. Millen. Getting our head in the clouds: toward evaluation studies of tagclouds. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 995–998, New York, NY, USA, 2007. ACM.
13. Markus Strohmaier. Purpose tagging: capturing user intent to assist goal-oriented social search. In *SSM '08: Proceeding of the 2008 ACM workshop on Search in social media*, pages 35–42, New York, NY, USA, 2008. ACM.
14. Markus Strohmaier, Mark Kroell, and Christian Koerner. Automatically annotating textual resources with human intentions. In *HT '09: Proceedings of the Twentieth ACM Conference on Hypertext and Hypermedia*, New York, NY, USA, July 2009. ACM.
15. Fernanda B. Viegas and Martin Wattenberg. Timelines - tag clouds and the case for vernacular visualization. *interactions*, 15(4):49–52, 2008.
16. Fernanda B. Viegas, Martin Wattenberg, Frank van Ham, Jesse Kriss, and Matt Mckeon. Manyeyes: a site for visualization at internet scale. *Transactions on Visualization and Computer Graphics*, 13(6):1121–1128, 2007.
17. Harald Weinreich, Hartmut Obendorf, Eelco Herder, and Matthias Mayer. Not quite the average: An empirical study of web use. *ACM Trans. Web*, 2(1):1–31, 2008.
18. Faris Yakob and Noah Brier. Ways of seeing. *Contagious Magazine*, (15), June 2008.

Evaluation of Current RDF Database Solutions

Florian Stegmaier[1], Udo Gröbner[1], Mario Döller[1], Harald Kosch[1] and Gero Baese[2]

[1] Chair of Distributed Information Systems
University of Passau
Passau, Germany
forename.surname@uni-passau.de
[2] Corporate Technology
Siemens AG
Munich, Germany
gero.baese@siemens.com

Abstract. Unstructured data (e.g., digital still images) is generated, distributed and stored worldwide at an ever increasing rate. In order to provide efficient annotation, storage and search capabilities among this data and XML based description formats, data stores and query languages have been introduced. As XML lacks on expressing semantic meanings and coherences, it has been enhanced by the Resource Description Format (RDF) and the associated query language SPARQL.
In this context, the paper evaluates currently existing RDF databases that support the SPARQL query language by the following means: general features such as details about software producer and license information, architectural comparison and efficiency comparison of the interpretation of SPARQL queries on a scalable test data set.

1 Introduction

The production of unstructured data especially in the multimedia domain is overwhelming. For instance, recent studies[3] report that 60% of today's mobile multimedia devices equipped with an image sensor, audio support and video playback have basic multimedia functionalities, almost nine out of ten in the year 2011. In this context, the annotation of unstructured data has become a necessity in order to increase retrieval efficiency during search. In the last couple of years, the Extensible Markup Language (XML) [16], due to its interoperability features, has become a de-facto standard as a basis for the use of description formats in various domains. In the case of multimedia, there are for instance the well known MPEG-7 [13] and Dublin Core [12] standards or in the domain of cultural heritage the Museumdat[4] and the Categories for the Description of Works of Art (CDWA) Lite[5] description formats. All these formats provide a

[3] http://www.multimediaintelligence.com
[4] http://museum.zib.de/museumdat/museumdat-v1.0.pdf
[5] http://www.getty.edu/research/conducting_research/standards/cdwa/cdwalite.html

XML Schema for annotation purposes. Related to this, several XML databases (e.g., Xindice[6]) and query languages (e.g., XPath 2.0 [2], XQuery [20]) have been introduced in order to improve storage and retrieval capabilities of XML instance documents.

The description based on XML Schema has its advantages in expressing structural and descriptive information. However, it lacks in expressing semantic coherences and semantic meaning within content descriptions. In order to close this gap, techniques emerging from the Semantic Web[7] have been introduced. The main contribution is RDF [19] and its quasi standard query language SPARQL [11]. Both, are recommendations of W3C[8], just as XML.

In this context, the paper provides an evaluation of currently existing RDF databases that support the SPARQL query language. The evaluation concentrates on general features such as details about software producer and license information as well as an architectural comparison and efficiency comparison of the interpretation of SPARQL queries on a scalable test data set.

The remainder of this paper is organized as follows: Section 2 covers some basic informations about accessing and evaluating RDF data. The definition of evaluation criteria is done in section 4. Section 5 provides an architectural overview of the triple stores in scope. Details about the test environment and the results of the performance tests are part of section 6. The paper is concluded in section 7.

2 Related work

This chapter covers basic information about related paradigms and technologies/standards required to perform the evaluation.

2.1 RDF data representation and storage approaches

Recent work already investigated several approaches concerning the storage of RDF data. In general, RDF data can be represented in different formats:

- *Notation 3 (N3)* [3] is a very complex language in order to store RDF-Triples, which was issued in 1998.
- *N-Triples* [17] was a recommendation of W3C, published in the year 2004. It is a subset of N3 in order to reduce its complexity.
- *Terse RDF Triple Language (Turtle)* [1] was invented in order to enlarge the expressiveness of N-Triples. The Turtle syntax is also used to define graph patterns in the query language SPARQL [8].
- *RDF/XML* [18] defines an XML syntax for representing RDF-Triples.

Three fundamental different storage approaches can be identified at present:

[6] `http://xml.apache.org/xindice/`
[7] `http://www.w3.org/2001/sw/`
[8] `http://www.w3.org`

– *in-memory storage* allocates a certain amount of the available main memory to store the given RDF data. Obviously this approach is intended to be used for few RDF data.
– *native storage* is a way to save RDF data permanently on the file system. These implementations may fall back on (in this terms) well investigated index structures, such as B-Tree.
– *relational database storage* makes use of relational database systems (e.g., PostgreSQL) to store RDF data permanently. Like the *native storage*, this approach relies on research results in the database domain (e.g., indices or efficient processing). Two different mapping strategies have been considered: The first is an universal table, which contains all RDF triples. The second solution is to create a mapping of the ontology into a table structure. Apparently, this leads to a (potentially) large number of tables.

2.2 RDF databases

An overview of frameworks and applications with the ability to store and to query RDF data is provided in Table 1. To retrieve the stored data, (quasi–) standards can be used, in names RDF Query Language (RQL) [10], RDF Data Query Language (RDQL) [15] and finally the W3C Recommendation SPARQL Protocol and RDF Query Language (SPARQL) [21]. A comparison of RDF query languages of the year 2004 can be found in [14].

2.3 RDF performance benchmarks

In addition to the huge efforts necessary to provide RDF database systems and defining query languages, appropriate evaluation methodologies[9] for triple stores have been introduced recently.

This section gives an overview of three promising performance benchmarks:

Berlin SPARQL Benchmark (BSBM)[10] [5] provides an benchmark using SPARQL. This benchmark includes a data generator and a test suite. The data generator is able to build a scalable amount of test data in RDF/XML format, which is based on an e-commerce use case. For example, a search for products from different suppliers can be performed or comments on the product can be provided. The mode of operation of the test suite is based on a use–case taken from real life. An automtic execution of miscellaneous queries is imitating the behavior of human operators.

Lehigh University Benchmark (LUBM)[11] [9] specifies the test data by an ontology named *Univ-Bench*. It represents an university with professors, students, courses and so on. The test data set can be constructed with the associated data generator [6]. The benchmark contains 14 test queries written in a KIF[12]–like

[9] `http://esw.w3.org/topic/RdfStoreBenchmarking`
[10] `http://www4.wiwiss.fu-berlin.de/bizer/BerlinSPARQLBenchmark/`
[11] `http://swat.cse.lehigh.edu/projects/lubm/`
[12] `http://www.csee.umbc.edu/kse/kif/`

Table 1. Overview of available RDF Triple Stores (abbreviations: o. = ongoing, disc. = discontinued, e.d.s. = early developing stage, u. = unknown)

Name	State	Programming language	Supported query language	Supported storage	Part of eval.	License
3Store	o.	C	SPARQL, RDQL	MySQL, Berkley DB	no	GPL
AllegroGraph	o.	Lisp	SPARQL	– (native disk storage)	yes	commercial
ARC	o.	PHP	SPARQL	MySQL	no	open source
BigOWLIM	o.	Java	SPARQL	– (plug-in of Sesame)	no	commercial
Bigdata	o.	Java	SPARQL	distributed databases	no	GPL
Boca	disc.	Java	SPARQL	relational databases	no	Eclipse Public License
Inkling	disc.	Java	SquishQL	relational databases	no	GPL
Jena	o.	Java	SPARQL, RDQL	in–memory, native disk storage, relational backends	yes	open source
Heart	e.d.s.	u.	u.	u.	no	u.
Kowari metastore	disc.	Java	SPARQL, RDQL, iTQL	native disk storage	no	Mozilla Public License
Mulgara	o.	Java	SPARQL, TQL & Jena bindings	integrated database	no	Open Software License v3.0
Open Anzo	o.	Java	SPARQL	relational database	yes	Eclipse Public License
Oracle's Semantic Technologies	o.	Java	SPARQL	relational database	yes	BSD-style license
RAP	o.	PHP	SPARQL, RDQL	in–memory, relational database	no	LGPL
rdfDB	o.	Perl	SQLish query language	Sleepycat Berkeley DB	no	open source
RDFStore	o.	Perl	SPARQL, RDQL	relational database	no	open source
Redland	o.	C	SPARQL, RDQL	relational databases	no	LGPL 2.1, GPL 2 or Apache 2
Semantics.Server 1.0	o.	.NET	SPARQL	MySQL	no	commercial
SemWeb – DotNet	o.	.NET	SPARQL	in–memory, relational database	no	GPL
Sesame	o.	Java	SPARQL, SeRQL	in–memory, native disk storage, relational database	yes	BSD-style license
Virtuoso	o.	Java	SPARQL	relational database	no	open source & commercial & open source
YARS	o.	Java	subset of N3	Berkeley DB	no	BSD-style license

language and a test suite called *UBT*, which manages the loading of data and the query execution automatically.

SP²B SPARQL Performance Benchmark (SP²B)[13] [7] benchmark consists of two major components. The first component is a (command line driven) data generator, which can automatically create the evaluation data. The amount of triples in this data set is scalable and based on the DBLP Computer Science Library[14]. In this case the data generator uses several well known ontologies, such as Friend of a Friend (FOAF)[15]. The second component consists of SPARQL queries, which are specifically designed for the DBLP use case.

3 Preselection of technologies in scope

This section provides the reasoning for the chosen databases and evaluation benchmark.

All technologies, which are discontinued or in a too early state of development, are excluded. As the development of Boca, Inkling, Kowari and RDFStore is discontinued and the Heart project is not yet implemented, a closer examination is not possible.

Furthermore, all databases shall have the ability to interpret SPARQL queries. As the overview in section 2.2 shows, rdfDB and YARS do not support SPARQL, these databases will not be part of the further evaluation.

Based on the evaluation in [7] the achieved evaluation of ARC, Redland and Virtuoso are insufficient, thus a further examination of these databases is not part of this paper. Our paper extends this previous work by highlighting architectural facets and general information of the tested databases (see section 4 for details). Furthermore, we collected yet available databases in table 1, which takes the current technologies and implementation efforts (e.g., Oracle's Semantic Technologies) into account. Schmidt et al. investigated in [7] the execution times for in–memory and native storage. In contrast to that, our evaluation is based on the relational storage approach.

The evaluation is based on SP²B, because it is most up–to–date and SPARQL specific. In order to use LUBM, a translation of the queries into SPARQL must be conducted, which is not satisfactory. Comparing the test data structure of BSBM to the data of SP²B, the SP²B data uses already well known ontologies, which is an additional advantage.

4 Evaluation criteria

The evaluation of RDF databases is based on three categories. The first category focuses on general information about the technologies:

[13] `http://dbis.informatik.uni-freiburg.de/index.php?project=SP2B`
[14] `http://www.informatik.uni-trier.de/~ley/db/`
[15] `http://www.foaf-project.org`

Software producer provides details about the company implementing the framework.

Associated licenses shed light on the usage of the frameworks, whether it can be used in business applications or not.

Project documentation should be rather complete. Furthermore, tutorials should be available supporting the work with these systems especially in the period of vocational adjustment.

Support is the last basic criteria. Support should be covered for example by an active forum or a newsgroup.

The aspects of the second category examine architectural facets of the considered frameworks, such as:

Extensibility is a very important criteria for the integration of new features, e.g., to optimize the existing working process. One of these features could be the implementation of new indices, which accelerate the performance and advance the efficiency of the entire system.

Architectural overview provides an insight into the structure of the framework and the used programming language.

OWL should be supported by the databases, because it enlarges the semantic expressiveness of RDF especially as far as reasoning is concerned.

Available query languages is another point of interest, is there support for other RDF addressing query languages in addition to SPARQL.

Interpretable RDF data formats are not part of central focus. The most important formats (as mentioned in section 2.1) should be covered by the frameworks from the point of completeness.

The evaluation of these two categories can be found in Chapter 5.

The third category is based on the expressiveness of SPARQL queries and the performance of the frameworks / applications. SPARQL consists of four different query forms: *SELECT*, *ASK*, *CONSTRUCT* and *DESCRIBE*. This evaluation is restricted to the *SELECT* query type. It is discussed in Chapter 6. Further details about the test environment are provided there, too.

5 Evaluation of considered databases

This section covers the evaluation of AllegroGraph, Jena, Open Anzo, Oracle's Semantic Technologies and Sesame following the reasoning in section 3.

5.1 AllegroGraph

The *software producer* of AllegroGraph RDF Store[16] is Franz Inc.[17]. The company has been founded in 1984 and is well known for its Lisp programming

[16] `http://www.franz.com/agraph/allegrograph/`

[17] `http://franz.com/`

language expertise. Recently, they also started developing semantic tools, like AllegroGraph.

The *associated licenses* of AllegroGraph come in two different flavors. The version evaluated in this paper is the free edition, which is limited to 50 million triples maximum. In contrast to that, the enterprise version has no limits regarding to the number of stored triples but underlies a commercial license.

The *product documentation* delivered by Franz Inc. is rather complete. Several useful example Java classes can be found on the companies website alongside the Javadoc of the Java binding.

Support for AllegroGraph is offered by Franz Inc. in a commercial way. In detail, they offer training for the software, seminars and consulting services, which also includes application-specific coding if needed.

AllegroGraph is not *extensible*. It is closed source and stores data as well as the database indices inside its particular storage stack.

Because of its closed source, an *architectural overview* is not possible. Therefore, figure 1 shows a client server architecture of AllegroGraph. The software is developed especially for 64 Bit systems and runs out of the box, as it doesn't need any other databases or software. Storage, indexing and query processing is performed inside AllegroGraph. The software can be accessed using Java, C#, Python or Lisp. There are bindings for Sesame or Jena integration available and also an option to access AllegroGraph via HTTP.

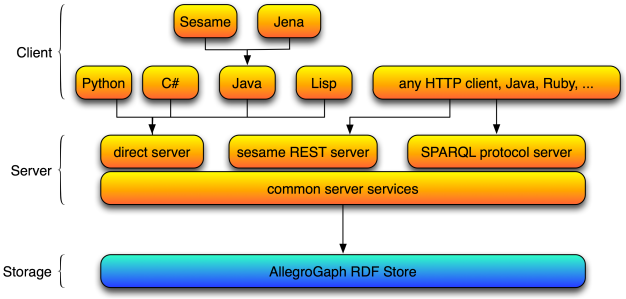


Fig. 1. AllegroGraph client server architecture

Franz Inc. suggests using TopBraid Composer[18] by TopQuadrant Inc. for *OWL support*.

The *available query language* of the software is SPARQL, but it also supports low level API calls for direct access to triples by subject, predicate and object. With those API calls, it is possible to retrieve all datasets matching a certain triple. The API calls provide functionality, which can be compared to SQL SELECT statements.

[18] `http://www.topquadrant.com/topbraid/composer/index.html`

The *interpretable RDF data formats* of AllegroGraph are RDF/XML and N-Triples. Other formats are planned to be supported in future versions.

5.2 Jena

The *software producers* of Jena[19] are the HP Labs[20], which are a part of the Hewlett-Packard Development Company. This department was founded in 1966 by Bill Hewlett and Dave Packard. Jena was developed in the terms of the HP Labs Semantic Web Research.

The *associated license* of the Jena project is completely open source. This implies that redistribution and use in source and binary forms with or without modification are permitted[21].

The Jena *product documentation* can be found on the project page and is widely complete. The documentation covers the central parts of Jena providing basic information about the framework, Javadocs and several tutorials respectively HowTos. The downloadable version of Jena also includes code examples, which underline the basic steps in the working process of Jena.

The *support* focuses on a newsgroup[22], which is founded in the Yahoo! Groups[23]. It may be considered unsatisfactory that support is primarily limited to a newsgroup. But due to the fact that there is a large amount of registered members[24] the activity of the newsgroup and therefore the delivered support is excellent.

The Jena download package includes the source files of the entire Jena project implemented in Java. This provides a basis for implementations *extending* the framework, for instance with new indices.

Figure 2 illustrates an *architectural overview* of Jena. The framework offers methods to load RDF data into a memory based triple store, a native storage or into a persistent triple store. In order to build a persistent triple store a variety of relational databases, for example MySQL, PostgreSQL or Oracle, can be used. The stored data may be retrieved through SPARQL queries. A standard implementation of the SPARQL query language is encapsulated in the ARQ package of Jena. SPARQL queries can be executed using Java applications or by the use of the graphical frontend Joseki[25]. The Ontology API provides methods to work on ontologies of different formats, like OWL or RDFS. Jena's Core RDF Model API offers methods to create, manipulate, navigate, read, write or query RDF data. The remaining major components are on the one hand the Inference API, which allows the integration of inference engines or reasoners into the system. On the other hand the Reification API is a proposal to optimize the representation of reification.

[19] `http://jena.sourceforge.net/`
[20] `http://www.hpl.hp.com/`
[21] `http://jena.sourceforge.net/license.html`
[22] `http://tech.groups.yahoo.com/group/jena-dev/`
[23] `http://groups.yahoo.com/`
[24] Members of the Jena newsgroup (at time of writing): 2752
[25] `http://www.joseki.org/`

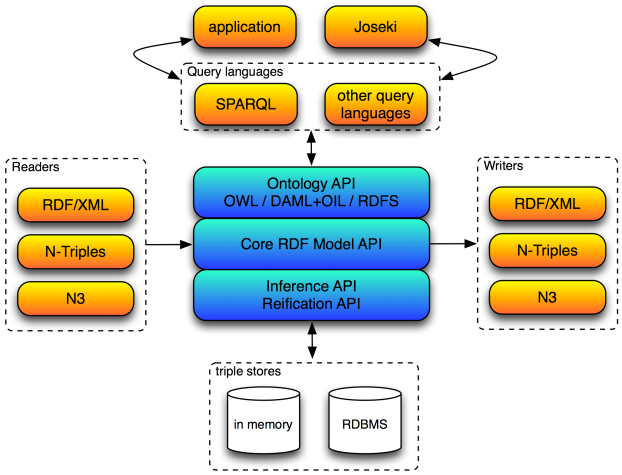


Fig. 2. Architectural overview of Jena

OWL support is given in form of the Ontology API. The inference subsystem[26] enables the use of inference engines or reasoners in Jena.

Besides SPARQL, RDQL is a *supported query language*. In a tutorial about RDQL it is recommended that new users of Jena should use SPARQL instead.

Jena uses readers and writers for RDF/XML, N-Triples and N3, which are commonly known *RDF data formats*.

5.3 Open Anzo

Open Anzo[27] is the prosecution of Boca[28] and other components produced by the IBM Semantic Layered Research Platform[29].

The Open Anzo project offers a good *product documentation*. The key topics are architectural facets of the current version, programmer guides and design documents. There are also documents available describing key features of an upcoming version of Open Anzo.

The *support* is based on several tutorials and a Google group[30] with about 63 members at time of writing.

As already mentioned, Open Anzo is complete open source, underlying the Eclipse Public License. So it is possible to *extend* the given framework by needed functionalities.

[26] `http://jena.sourceforge.net/inference/`
[27] `http://www.openanzo.org/`
[28] `http://ibm-slrp.sourceforge.net/`
[29] `http://ibm-slrp.sourceforge.net/`
[30] `http://groups.google.com/group/openanzo`

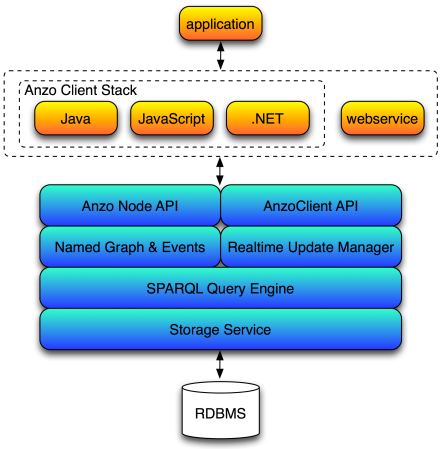


Fig. 3. Architectural overview of Open Anzo

Figure 3 highlights the main components of the Open Anzo *architecture*. Open Anzo can be used with three modes of operation. It is possible to embed it in an application, run it as a remote server or use it locally. The entry points to the framework are the Anzo Client Stack (offers API implementations in Java, Javascript and .NET) or a webservice. The Anzo Node API is the basis to describe the structure of RDF data. The named graph component enables user to access the RDF data. Beside that, the AnzoClient API encapsulates transaction preconditions and connectivity events to the database. The purpose of the Realtime Update Manager is to deliver messages about certain processing states. In order to execute SPARQL queries in Open Anzo, the SPARQL Query API is needed. The Storage Service is used to save and retrieve RDF data using a relational database (like DB2 or Oracle). This is the center of any mode of operation in an Open Anzo system.

There are OWL related classes in the project, but further information is missing in the documentation regarding the coverage of *OWL* functionalities. The producers claim on the product page that other semantic web technologies (3^{rd} party components) could easily be plugged into the system.

Open Anzo supports SPARQL queries and typed full-text search capabilities, which also use an index system in order to improve the retrieval process.

N3, N-Triples, RDF/XML and TriX[31] are the supported *RDF data formats*.

[31] `http://www.w3.org/2004/03/trix/`

5.4 Oracle's Semantic Technologies

Software producer Oracle[32] is one of the major players in database business. The company comprises relational database knowledge of 30 years and has added support for semantic technologies to its products lately. The evaluated Semantic add-on is the Jena Adapter 2.0 for Oracle Databases. It implements the Jena Graph and model APIs as described earlier. The add-on requires Oracle Database 11g Release 11.1.0.6 (or higher) or Oracle Database 10g Release 10.20.0.1 or 10.2.0.3.

Licensing options can be found at the Oracle page[33]. The Jena Adapter is provided from Oracle for free as closed source.

Product documentation can be found at Oracle Semantic Technologies Center[34] and offers code samples, usage scenarios, training material and documentation for administrators as well as developers. The documentation provides a good overview, but the structure of the website could be improved for usability reasons.

Support is available via the Oracle forum[35] for free, with excellent answer times. Paid support is also available from several partners[36] and from Oracle itself.

An overview of the semantic capabilities of Oracle's add–ons is illustrated in figure 4.

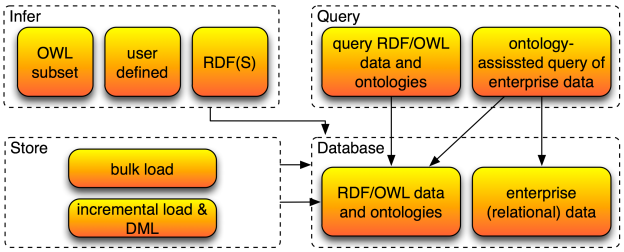


Fig. 4. Oracle's Semantic Technologies capabilities

Oracle supports large graphs of billions of triples, which can be queried by SPARQL-like syntax and/or SQL. Complete SPARQL support is at the time of this writing only available via the Jena adapter but native support for SPARQL is planned. The RDF data model includes capabilities for inference via RDFS, its

[32] `http://www.oracle.com`
[33] `http://www.oracle.com/us/corporate/pricing/index.htm`
[34] `http://www.oracle.com/technology/tech/semantic_technologies/index.html`
[35] `http://forums.oracle.com/forums/forum.jspa?forumID=269`
[36] `http://www.oracle.com/technology/tech/semantic_technologies/htdocs/semtech_partners.html`

subset RDFS++, OWL, its subsets OWLSIF and OWLPrime, and user–defined rules.

RDF data formats are RDF/XML, N-Triples and N3 because Jena is being utilized. Semantic data can also be compressed by using the advanced compression option to reduce needed disk space.

5.5 Sesame

The *software producer* of Sesame[37] is Aduna[38]. This company sets the focus of their work in revealing the meaning of information. Sesame was started as a prototype of the EU project On-To-Knowledge[39] and is now developed by Aduna in a cooperation with NLNet Foundation[40].

Like Jena, Sesames *associated license* is open source underlying the BSD-style license.

The *product documentation* of Sesame is well organized. There is a large user guide available for every version of Sesame. Users can also refer to Javadocs and tutorials completed with example code.

Aduna provides *support* in form of an active forum accessible on the project page and a mailing list based on SourceForge[41]. Commercial consulting services are also provided.

Sesame's download package is shipped with the Java source files. Therefore, a basis for *extending* the framework is provided similar to Jena.

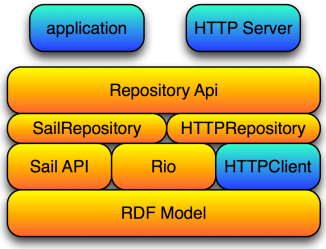


Fig. 5. Architectural overview of Sesame

Figure 5 shows an *architectural overview* of Sesame. In order to use Sesame, Apache Tomcat is recommended. The Sesame package also contains two web applications, the Sesame server which stores the RDF data and the OpenRDF

[37] `http://www.openrdf.org/`
[38] `http://www.aduna-software.com/`
[39] `http://www.ontoknowledge.org/`
[40] `http://www.nlnet.nl/`
[41] `http://www.sourceforge.net`

Workbench as a graphical frontend for the server. This workbench can manage repositories, load RDF data and execute queries. Sesame is able to handle all three in section 2.1 discussed approaches to store RDF data. The RDF Model implements basic concepts about RDF data. The component RDF I/O (Rio) consists of a set of parser and writer for the handling of RDF data. This is for instance used by the Storage And Inference Layer (Sail) API for initializing, querying, modifying and the shut down of RDF stores. On the topmost layer constitutes the Repository API the main entrance to address repositories. Compared to Sail, which is rather a low level API, the Repository API is the associated high level API with a larger amount of methods for managing RDF data. The HTTPRepository is an implementation that acts like a proxy in order to connect to a remote Sesame server via the HTTP protocol.

In order to achieve *OWL support* a Plug-In is available called BigOWLIM[42]. It is implemented as a high performance semantic repository for Sesame and packaged as a Sail.

Alternatively to SPARQL Sesame is able to interpret the Sesame RDF Query Language (SeRQL) [4] integrated for enhancing the functionality of RQL and RDQL.

Sesame offers parsers for various well known *RDF formats* N3, N-Triples, RDF/XML, Turtle and two new formats TriG[43] and TriX.

6 Performance tests

The performance tests of AllegroGraph 3.3.1, Jena (SDB 1.1), Open Anzo 3.1.0, Oracle's Semantic Technologies (Jena Adapter v.2.0)and Sesame 2.2.4 are conducted in the following test environment. It consists of a client and a server connected over a 1 Gb LAN network. The main parts of the server are two Intel Xeon 3,8GHz Single-Core CPUs, 6 GB RAM and two 136GB Ultra320-SCSI HDDs in a Hardware-RAID-1 with a Ubuntu 8.04.1 operating system running on top. The client is a MacBook Pro with a 2,4 GHz Intel Core 2 Duo CPU, 2 GB Ram and a 150 GB Fujitsu HDD and the Mac OS 10.5.7 operating system. In order to create persistent triple stores in Jena and Sesame, PostgreSQL is used. All performance tests are conducted with the standard configurations of the frameworks and database backends.

The queries of the SP^{2}B benchmark can be classified into two groups according to the expected complexity. On the one hand *FILTER*, *OPTIONAL* and *UNION* are very similar to well known SQL paradigms (*SELECT*, left outer joins, relational *UNION*). Only minor influence on the performance of query execution is assumed, because efficient implementations can be used [7]. On the other hand keywords like *DISTINCT*, *LIMIT* or *OFFSET* will seriously affect the query execution [7] (*pipeline breaker*). The queries will indicate the correctness of this assumptions, as they insist on at least one of the keywords or a combination of them. The graph structure, which will be build by the queries can

[42] `http://ontotext.com/owlim/big/`
[43] `http://www4.wiwiss.fu-berlin.de/bizer/TriG/`

be distinguished into long path chains[44], bushy patterns[45] or the combination of these two structure types.

The evaluation data was created in the N3 data format with the SP^{2}B data generator. A data set with about 100.000 triples (10.3 MB) another with 1.000.000 triples (107 MB) and a last one with 5.000.000 triples (538 MB) have been created. In order to import the N3 data into AllegroGraph, CWM[46] has been used to parse the N3 data into RDF/XML, which AllegroGraph is able to process. The parser was not able to parse the dataset with 5.000.000 triples. Therefore, this data set could not be tested with AllegroGraph.

The following part shows the results of the evaluation focusing on the query execution time. This time only includes the query execution and the transfer of the result set from the server to the client (opening and closing of the connection to the repository not included). The time unit given in the figure 6 are milliseconds. A value of 1.000.000 milliseconds indicates a timeout of the query.

The execution times clearly show a great difference in the query execution between Jena, Open Anzo, Oracle's Semantic Technologies, Sesame and AllegroGraph and are similar to the execution times achieved in [7] for in–memory and native storage. For instance the execution of query 4 regarding the 100.000 triple test set lasts 28 milliseconds in Jena and 18 milliseconds in Oracle. In contrast, this query on the same test set took 14478 milliseconds in Sesame, 141155 milliseconds in Open Anzo and 176496 milliseconds in AllegroGraph. There are also queries, where Sesame's execution times are smaller than Jena's or Oracles, for example Query 1 and 2 (also in the two bigger data sets). A reason for this behavior comparing Jena, Oracle and Sesame is the diverse import strategy of these two frameworks. The import of data in Sesame leads to the creation of 69 tables for the 100.000 triples data set, 79 tables for the 1.000.000 triples data set and 87 tables for the 5.000.000 triples data set. Jena creates constantly 4 tables (universal table approach as discussed in section 2.1). Oracle's Semantic Technologies is using the Jena framework, the storage approach is the identical. Sesame performs a mapping of the different entities in the N3 data sets directly into tables of the database while building several other tables to save the RDF triples data. Jena doesn't use a mapping like this. Obviously, queries consisting of a great amount of *dots*[47] increase the execution time on a database with about 70 tables compared to a database with only 4 tables. The other way round Sesame is able to minimize the number of cross products during query execution because it is able to address the elements of a special entity saved in a particular table. AllegroGraph is saving the triples directly on the hard disk. It creates one data file containing the RDF data and several other files, which purpose is not deducible. Although AllegroGraph uses some kind of indices on the repository the execution lasts much longer than in the other frameworks.

[44] Similar to joins over a few tables in a relational database.
[45] For example queries on a Star Schema
[46] `http://www.w3.org/2000/10/swap/doc/cwm.html`
[47] *dots* are similar to joins.

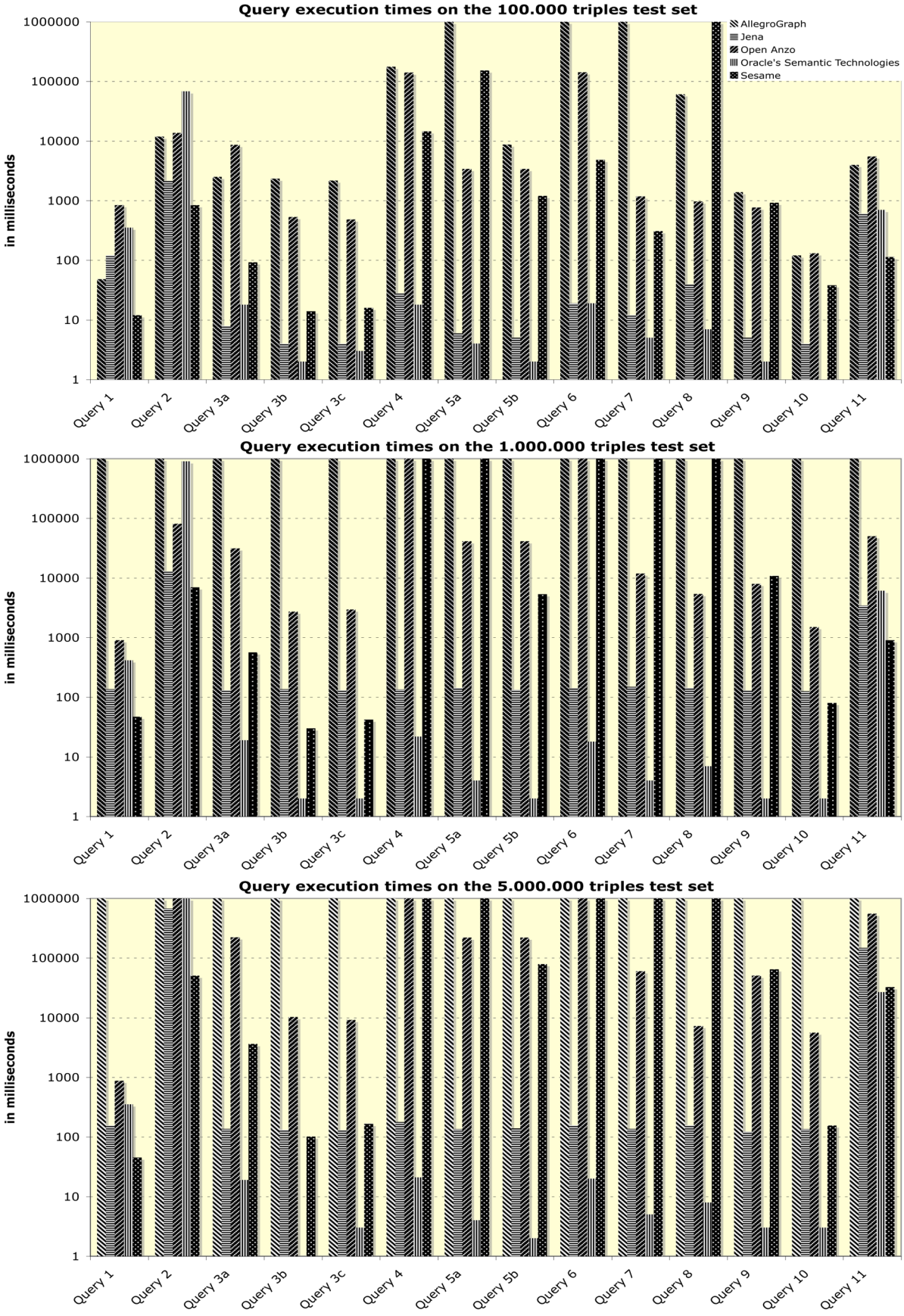


Fig. 6. Query execution times on the three different test sets

Figure 6 also shows the results of the evaluation for the 1.000.000 triples data set and for the 5.000.000 triples data set. The execution time of AllegroGraph was exceeding the time limit (terminated after 30 minutes per query) for the 1.000.000 triples data set. There is also an ascent of the execution times and timeouts observable for the other triple stores.

7 Conclusion & Outlook

The architectural overview of chapter 5 and the performance tests of chapter 6 shows that AllegroGraph is not fulfilling the criteria defined in chapter 4. It is neither extensible nor are the execution times satisfying. Jena and Sesame are both API extensible but Jena obtained continuous evaluation times at the moment. Oracle's Semantic Technologies is using the Jena framework but it comes with database procedures, which have an impact on the performance. In contrast to that, Open Anzo serves well for small data but is not very good in handling big amounts of RDF data. Jena and Oracle Semantics Technologies are fulfilling the chosen criteria best. However, a decision to use one or the other framework must be based on the domain to be addressed by such a system and on the query structure expected. A deeper analysis of these two factors helps finding the answer, what kind of storage approach would be appropriate.

This paper, especially section 2.2 shows that huge efforts were done in the field of accessing RDF data. This trend is still ongoing as the development of new RDF triple stores (e.g., HEART) is indicating. Up to now, only relational databases or XML databases are in scope of these technologies. Only one database, namely Bigdata, is able to operate on a distributed database. Enlarging the set of accessible backends may improve the performance issues of certain query paradigms in a good way. Future work could focus on the mapping of SPARQL to SQL. Here, already well known database techniques could seriously enhance the processing of queries.

8 Acknowledgments

This work has been supported in part by the THESEUS Program, which is funded by the German Federal Ministry of Economics and Technology.

References

1. David Beckett. Turtle - terse rdf triple language. http://www.dajobe.org/2004/01/turtle/, November 2007.
2. Anders Berglund, Scott Boag, Don Chamberlin, Mary F. Fernandez, Michael Kay, Jonathan Robie, and Jerome Simeon. XML Path Language (XPath) 2.0. *W3C Recommendation, http://www.w3.org/TR/xpath20/*, 2007.
3. Tim Berners-Lee. Notation 3. http://www.w3.org/DesignIssues/Notation3, March 2006.

4. Jeen Broekstra and Arjohn Kampman. SeRQL: A Second Generation RDF Query Language. http://www.w3.org/2001/sw/Europe/events/20031113-storage/positions/aduna.pdf, November 2003.
5. Christian Bizer et al. Benchmarking the performance of storage systems that expose sparql endpoints. In *Proceedings of the 4th International Workshop on Scalable Semantic Web knowledge Base Systems (SSWS2008)*, 2008.
6. Kurt Rohloff et al. An evaluation of triple-store technologies for large data stores. In Robert Meersman et al., editor, *OTM Workshops (2)*, volume 4806 of *Lecture Notes in Computer Science*, pages 1105–1114. Springer, 2007.
7. Michael Schmidt et al. SP2Bench: A SPARQL Performance Benchmark. *CoRR*, abs/0806.4627, 2008.
8. Pascal Hitzler et al. *Semantic Web*. Springer, 2008.
9. Yuanbo Guo et al. Lubm: A benchmark for owl knowledge base systems. *J. Web Sem.*, 3(2–3):158–182, 2005.
10. Gregory Karvounarakis, Sofia Alexaki, Vassilis Christophides, Dimitris Plexousakis and Michel Scholl. RQL: a declarative query language for RDF. In *WWW*, pages 592–603, 2002.
11. Eric Prud hommeaux and Andy Seaborne. SPARQL Query Language for RDF. *W3C Recommendation, http://www.w3.org/TR/rdf-sparql-query/*, 2008.
12. Dublin Core Metadata Initiative. Dublin core metadata element set - version 1.1: Reference description. *http://dublincore.org/documents/dces/*, 2008.
13. J. M. Martinez, R. Koenen, and F. Pereira. MPEG-7. *IEEE Multimedia*, 9(2):78–87, April-June 2002.
14. Peter Haase, Jeen Broekstra, Andreas Eberhart and Raphael Volz. A Comparison of RDF Query Languages. In *International Semantic Web Conference*, volume 3298, pages 502–517, 2004.
15. Andy Seaborne. RDQL - A Query Language for RDF. http://www.w3.org/Submission/2004/SUBM-RDQL-20040109/, January 2004.
16. W3C. Extensible Markup Language (XML) 1.1, W3C Recommendation. *http://www.w3.org/XML/*, February 2004.
17. W3C. Rdf test cases. http://www.w3.org/TR/rdf-testcases/, February 2004.
18. W3C. RDF/XML Syntax Specification (Revised). http://www.w3.org/TR/rdf-syntax-grammar/, February 2004.
19. W3C. Resource Description Framework (RDF). *http://www.w3.org/RDF/*, 2004.
20. W3C. XQuery 1.0: An XML Query Language. *W3C, http://www.w3.org/TR/2007/REC-xquery-20070123/*, 2007.
21. W3C. SPARQL Query Language for RDF. http://www.w3.org/TR/rdf-sparql-query/, January 2008.

How to Align Media Metadata Schemas? Design and Implementation of the Media Ontology

Florian Stegmaier[1], Werner Bailer[2], Tobias Bürger[3], Mario Döller[1], Martin Höffernig[3], Wonsuk Lee[4], Véronique Malaisé[5], Chris Poppe[6], Raphaël Troncy[7], Harald Kosch[1] and Rik Van de Walle[6]

[1] Chair of Distributed Information Systems, University of Passau, Germany
[2] Institute of Information Systems, JOANNEUM RESEARCH, Graz, Austria
[3] Semantic Technology Institute (STI), University of Innsbruck, Austria
[4] Standards Research Center, ETRI, Daejeon, Korea
[5] Web and Media Group, VU University, Amsterdam, Netherlands
[6] Department of Electronics and Information Systems, Ghent University - IBBT, Belgium
[7] EURECOM, Sophia Antipolis, France

Abstract. Multimedia data is generated, shared, stored and distributed worldwide at an ever increasing rate. This huge amount of content comes with metadata represented in different formats which hardly interoperate although they partially overlap. The W3C Media Annotations Working Group is chartered to recommend a Media Ontology compatible with most of these schemas. In this paper, we present the process for modeling this ontology and we discuss various approaches for explicitly representing the mappings between the core set of annotation properties defined in the Media Ontology and some major deployed metadata standards. We highlight the benefits and drawbacks of each approach and conclude on future work for the implementation of the Media Ontology.

1 Introduction

The publication and consumption of multimedia data on the Web has grown heavily thanks to the multiplicity of photo and video sharing platforms, usually embedded within social networks, along with the spread of multimedia enabled mobile devices. This huge amount of content can be generally accessed either via standardized and proprietary metadata formats, or more directly via APIs attached to web sites. As a result, the content is often locked in within silos preventing an effective search across these sites and making it complicated to create mashable applications.

While the multimedia metadata formats used on the web largely overlap in their coverage, they are at the same time dissimilar in many ways. **Coverage**: MPEG-7 [9] for example aims to be domain independent while DICOM [10] focuses on medical images, videos and workflows; **Comprehensiveness**: For example, MPEG-7 aims to provide comprehensive descriptions of multimedia content ranging from low-level features that can be extracted automatically to

fine-grained semantic description of a scene, while Dublin Core [6] provides a simple list of general annotation properties and EXIF focuses on the technical aspects of the media; **Complexity**: Metadata formats also differ in the complexity of their description syntax. For example, the Dublin Core `dc:creator` property is a simple name or an URI identifying an agent whereas the creator's name in MPEG-7 is divided into a complex nested structure of `Title`, `FamilyName` and `GivenName` along with the definition of his or her `Role`.

Designing multimedia systems nowadays often amounts to choose a subset of these various formats and implements manually their correspondence which severely hampers their interoperability. In this paper, we report on the design and implementation of the Media Ontology developed by the W3C Media Annotations Working Group (MAWG)[8] which aims at defining a set of minimal annotation properties for describing multimedia content along with a set of mappings between the main metadata formats in use at the moment. This ontology being described in prose, we investigate and discuss different options of formalization and implementation of its core annotation properties and the defined mappings with other standard formats.

The remainder of this paper is organized as follows. Section 2 presents multimedia metadata formats between which interoperability is necessary, and an overview of interoperability approaches for XML or RDF/OWL-based schemas. Section 3 presents the Media Ontology and the process of its elaboration. Section 4 discusses various implementation approaches for representing the ontology itself and the mappings between multimedia formats. Finally, Section 5 concludes the paper and outlines some future work.

2 Related work

Several standards have been created to improve the interoperability between different systems within one domain or application type. In this section, we describe some image and video metadata standards (*i.e.* schemas), and discuss some approaches for combining them. An exhaustive list of multimedia metadata formats has been produced by the W3C Multimedia Semantics Incubator Group[9].

2.1 Many Standards for Different Needs

Photos taken by digital cameras come with Exchangeable Image File (EXIF[10]) metadata directly embedded into the header of image files. It provides technical characteristics such as the shutter speed or aperture, and contextual information (date and time) of the captured image. Two RDFS ontologies of this specification have been proposed by Kanzaki and Norm Walsh. The Extensible Metadata Platform (XMP[11]) is a specification published by Adobe for attaching metadata

[8] `http://www.w3.org/2008/WebVideo/Annotations/`
[9] `http://www.w3.org/2005/Incubator/mmsem/XGR-vocabularies/`
[10] `http://www.digicamsoft.com/exif22/exif22/html/exif22_1.htm`
[11] `http://www.adobe.com/devnet/xmp/`

to media assets in order to enable a better management of multimedia content. The specification standardizes the definition, creation, and processing of metadata by providing a data model, a storage model, and formal predefined sets of metadata property definitions. XMP makes use of RDF in order to represent the metadata properties associated with a document. The DIG35[12] specification of the International Imaging Industry Association (I3A) defines a standard set of metadata for digital images including basic image parameter, image creation (à la EXIF), content creation and intellectual property rights and represented in XML. The IPTC Photo Metadata standard[13] developed by the International Press Telecommunication Council (IPTC) provides also a set of metadata properties being administrative, descriptive or related to the image rights. Largely based on XMP, this specification allows to represent as well complex semantic descriptions of the subject matter (e.g. persons, organizations, events).

EBUCore[14] is an XML-based metadata standard created by the European Broadcasting Union (EBU) consisting in a set of metadata properties specializing Dublin Core for describing radio and television content. MPEG-7 [9] is the Motion Pictures Expert Group (MPEG)[15] standard for the description of audio, video and multimedia content designed for document retrieval. The standard is based on XML Schema but MPEG-7 ontologies expressed in OWL have been proposed and compared among each other [12]. The standard is composed of many descriptor tools for diverse types of annotations on different semantic levels, ranging from very low-level features, such as visual (e.g. texture, camera motion) or audio (e.g. melody), to more abstract descriptions. The flexibility of MPEG-7 is based on structuring tools, which allow descriptions to be associated with arbitrary multimedia segments or regions, at any level of granularity, using different levels of abstraction.

Numerous metadata standards exist for annotating multimedia resources, all with their own merits and community usage. It is undesirable to enforce a single multimedia metadata standard that would satisfy all use cases. Some additional steps are needed to combine these formats and interoperability can be achieved by the means of mappings or relationships between the different schemas. In the next section, we review approaches for structural (*i.e.* syntactic) and semantic integration of multimedia metadata schemas.

2.2 Interoperability Approaches between Metadata Schemas

JPSearch is a project issued by the JPEG standardization committee to develop technologies that enable search and retrieval capabilities among image archives, consiting of five parts. While the first part focus on describing use cases and the overall architecture of image retrieval systems, the part 2 introduces an XML-based core metadata schema and transformation rules for mapping descriptive

[12] http://xml.coverpages.org/FU-Berlin-DIG35-v10-Sept00.pdf
[13] http://www.iptc.org/std/photometadata/2008/specification/IPTC-PhotoMetadata-2008_2.pdf
[14] http://tech.ebu.ch/docs/tech/tech3293-2008.pdf
[15] http://www.chiariglione.org/mpeg/

information (e.g., core metadata to MPEG-7 or core metadata to Dublin Core) between peers [2]. Part 3 adapts a profile of the MPEG Query Format [3] for ensuring standardized querying. Part 4 adopts the well known image data formats (JPEG and JPEG 2000) for embedding metadata information. The benefit of such an integration and combination of metadata with raw data is the mobility of metadata and its persistent association with the image itself. By embedding the metadata into the image raw data file format, one improves the flexibility within the annotation life cycle. However, the interchange of image data between JPSearch compliant systems remains an open issue. For this purpose, Part 5 concentrates on the standardization of a format for the exchange of image or image collections and its metadata and metadata schema between JPSearch compliant systems.

Xing et al. [13] present a system for automating the transformation of XML documents using a tree matching approach. However, this method has an important restriction: the leaf text in the different documents has to be exactly identical. This is hardly the case when combining different metadata standards. Likewise, Yang et al. [14] propose to integrate XML Schemas. They use a more semantic approach, using the ORA-SS data model to represent the information available in the XML Schemas and to provide mappings between the different documents. The ORA-SS data model allows to define objects and attributes to represent hierarchical data, however more advanced mappings involving semantic relationships cannot be represented.

Cruz et al. [1] introduced an ontology-based framework for XML semantic integration. For each XML source integrated, a local RDFS ontology is created and merged in a global ontology. During this mapping, a table is created that is further used to translate queries over the RDF data of the global ontology to queries over the XML original sources. The authors assume that every concept in the local ontologies is mapped to a concept in the global ontology. This assumption can be hard to maintain when the number and the degree of complexity of the incorporated ontologies increases. Poppe et al. [11] advocates a similar approach to deal with interoperability problems in content management systems. An OWL upper ontology is created and the different XML-based metadata formats are represented as OWL ontologies and mapped to the upper ontology using OWL constructs and rules. However, the upper ontology is dedicated to content management system and, as such, is not as general as the approach proposed in this paper.

The W3C Multimedia Semantics Incubator Group[16] elaborated on the inherent problems of using XML-based metadata standards[17]. The goal of the group was to investigate the usage of Semantic Web Technologies to overcome interoperability issues. The group discussed the advantages and open issues regarding the use of Semantic Web technologies but was not chartered for providing one common ontology for metadata annotation.

[16] `http://www.w3.org/2005/Incubator/mmsem/`

[17] Such metadata standards consist generally of an XML schema defining a syntax and a textual description specifying in prose the semantics of the standard

3 The W3C Media Ontology

The W3C Media Annotations Working Group (MAWG) has the goal of improving the interoperability between media metadata schemas. The proposed approach is to provide an interlingua ontology and an API designed to facilitate cross-community data integration of information related to media resources in the web, such as video, audio, and images.

The set of core properties that constitute the Media Ontology 1.0 is based on a list of the most commonly used annotation properties from media metadata schemas currently in use. This set is derived from the work of the W3C Incubator Group Report on Multimedia Vocabularies on the Semantic Web and a list of use cases [7], compiled after a public call. The use cases involve heterogeneous media metadata schemas used in different communities (interactive TV, cultural heritage institutions, etc.). In this section, we describe the content of this ontology and how this content is related to other metadata formats.

3.1 The Media Ontology Core Properties

The set of core properties defined in the Media Ontology 1.0 (`ma` namespace) consists of 20 descriptive and 8 technical metadata properties. This distinction has been made as the descriptive properties are media agnostic and also apply to descriptions of multimedia works that are not specific instantiations, e.g. the description of a movie on IMDB in contrast to a particular MPEG-4 encoded version of this movie broadcasted of the RAI Italian TV channel. The technical properties, specific to certain media types, are only essential when describing a certain instantiation of the content[18].

All properties are defined within the `ma` namespace since we have tried to clarify and disambiguate their definitions in the context of media resources description. However, whenever these properties exist in other standards, we try to explicitly define how they are related. Additionally, for many of the descriptive properties, we have foreseen subtypes that optionally further qualify the property, e.g. qualify a title as main or secondary.

The descriptive properties contain identification metadata such as identifiers, titles, languages and the locator[19] of the media resource being described. Other properties describe the creation of the content (the creation date, creation location, the different kinds of creators and contributors, etc.), the content description as free text, the genre, a rating of the content by users or organizations and a set of keywords. There are also properties to describe the collections the described resource belongs to, and to express relations to other media resources, e.g. source and derived works, thumbnails or trailers. As we consider digital rights management out of our scope, the set of properties only contains a copyright statement and a reference to a license (e.g. Creative Commons or

[18] This distinction is also present in the FRBR model where a `Work` is distinguished from a `Manifestation`.

[19] The locator is the physical place where the resource can be accessed.

MPEG-21 licenses). The distribution related metadata includes the description of the publisher and the target audience in terms of regions and age classification. Annotation properties can be attached to the whole media or to part of it, for example using the Media Fragments URI specification for identifying multimedia fragments.

The set of technical properties has been limited to the frame size of images and video, the duration, the audio sampling rate and frame rate, the format (specified as MIME type), the compression type, the number of tracks and the average bit rate. These were the only properties that were needed for the different use cases listed by the group.

This set of annotation properties is not considered final and properties might be added if it turns out to be useful. However, the aim is to keep the size of the ontology limited. If necessary, profiles can be defined, e.g. to group the properties that apply to a certain media type.

3.2 Expressing Mappings with other Standards

This core set of annotation properties has often correspondences with existing metadata standards. The working group has therefore further specified a mapping table that defines *one-way* mappings between the Media Ontology core properties and the metadata fields from 24 other standards [8].

The mappings that have been taken into account have different semantics, which can be characterized as:

- Exact matches: the semantics of the two properties are equivalent in most of the possible contexts. For example, `ma:title` matches exactly `dc:title`.
- More specific: the property of the vocabulary taken into account has a semantic that covers only a subset of the possibilities expressed by the property defined in the Media Ontology. For example, `ipr_names@description` and `ipr_person@description` defined in in DIG35 are more specific than the property `ma:publisher`.
- More generic: the inverse of the above, the property of the vocabulary taken into account has a semantic that is broader than the property defined in the Media Ontology. For example, `location` defined in the DIG35 is more general than `ma:location`.
- Related: the two properties are related in a way that is relevant for some use cases, but this relation has no defined semantics. For example, `media:credit` defined in MediaRSS[20] is related to `ma:creator`.

We discuss in the next section how these mappings can be represented.

4 Implementation Approaches

The W3C Media Ontology has been designed to be a meaningful subset of common annotation properties defined in standards used on the Web (see Section 2).

[20] `http://search.yahoo.com/mrss/`

The question is therefore how to implement or serialize the mapping relationships between the core set of properties defined by the Media Ontology and the other standards. This section discusses two classes of approaches: expressing a direct mapping using a more or less epxressive semantic web language (Sections 4.1 and 4.2) , or using a pivot upper ontology (Sections 4.3 and 4.4).

We illustrate each approach with a simple and a complex mapping between a property defined in the Media Ontology and its correspondence in another standard. These mappings concern the `ma:title` property which value is a simple string and the `ma:frameSize` property which value is composed of two integers representing the width and height of the video frames. The example 1.1 lists the prefixes we use for representing these mappings though all ontologies are not yet dereferencable.

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix ma: <http://www.w3.org/2009/09/mediaont#> .
@prefix ebu: <http://www.ebu.ch/metadata/ontologies/> .
```

Precondition 1.1. Declaration of prefix used in the examples.

4.1 Expressing Mappings in SKOS

SKOS[21] is a W3C Recommendation that defines a vocabulary for representing Knowledge Organization Systems (i.e. vocabularies) and relationships amongst them. SKOS provides constructs to formalize how concepts are related to each other. These constructs include `skos:exactMatch`, to express that two concepts are equivalent in *most* cases, `skos:closeMatch`, to express an equivalence valid in *some* cases, `skos:narrowMatch` and `skos:broaderMatch`, to express hierarchical relationships between concepts, and `skos:relatedMatch`, to express any other type of relatedness.

The first approach consists in applying these constructs to express all mapping relationships considered by the working group[22].

```
ma:title skos:exactMatch dc:title .
```

Example 1.2. A simple mapping represented in SKOS

[21] `http://www.w3.org/TR/skos-reference/`

[22] The mapping tables are available from `http://www.w3.org/2008/WebVideo/Annotations/drafts/ontology10/WD/mapping_table.html`

The use of these properties has a first implication: it entails that the properties `ma:title` and `dc:title` become instances of `skos:Concept` per definition of the `skos:exactMatch` construct. Second, we use the `skos:Collection` construct to group and list items, enabling the representation of a mapping between a simple property on the one hand, and multiple ones on the other hand. `skos:OrderedCollection` represents an ordered list of properties, enabling a more precise matching if necessary, but complex operations cannot be expressed. For example, the creator property defined in the Media Ontology has a simple value, whereas other vocabularies such as MPEG-7 define people with multiple properties: first name, last name, role, etc. SKOS cannot be used to represent that these values must be aggregated and concatenated to be used as value in the Media Ontology.

```
ma:frameSize skos:closeMatch [
  skos:Collection [
        skos:member ebucore:formatHeigth , ebu:formatWidth
    ] ;
 ] .
```

Example 1.3. A complex mapping represented in SKOS

Benefits of this approach:

- Scalability: new properties can be added to the mapping list;
- Fuzziness: mappings are created between properties that are more loosely related than a strict equivalence, which is often the case across schemas designed for specific applications.

Drawbacks:

- Assume that schemas and ontologies to be aligned have been formalized in RDF;
- Inference possibilities are limited;
- No formal complex rule can be attached to this representation.

4.2 Expressing Mappings in OWL and SWRL

Another approach consists in using a more expressive knowledge representation language to express direct mappings between the Media Ontology and other standards. The authors in [11] propose to use OWL and SWRL constructs as shown in the example 1.4 for defining a formal semantic equivalence between the title property defined in EBUCore and Dublin Core and in the Media Ontology.

Additionally, logical rules can be employed to do any type of conversion (including syntactic ones) and transformation of values (e.g., convert bps to kbps). Example 1.5 expresses in SWRL [5] that the value of `ma:frameSize` property

```
ma:title owl:equivalentProperty dc:title .
```

Example 1.4. A simple mapping represented in OWL.

```
[r1: (?res rdf:type ebu:ResourceManifestation)
     (?res ebu:width ?width)     (?res ebu:height ?height)
     (?width ebu:unit "pixels")  (?width  ebu:value ?w1)
     (?height ebu:unit "pixels") (?height ebu:value ?h1)
     -> (?size1 rdf:type ma:Size)
        (?size1 ma:width ?w1) (?size1 ma:height ?h1)
        (?res ma:size ?size1)]
```

Example 1.5. A complex mapping represented in SWRL.

can be filled from the values of the `ebu:width` and `ebu:height` properties. Benefits of this approach:

- Scalability: new properties can be added to the mapping list;
- Formalization: all sort of mappings can be formally represented, including complex ones, allowing inferences to be performed.

Drawbacks:

- Not all metadata standards have formal representations. Sometimes, there are even multiple formalizations of the same standard (e.g. MPEG-7 [12]);
- Complexity: the use of OWL constructs and complex rules can yield in undecidable reasoning.

4.3 Expressing Mappings Using a Format Independent Ontology

An alternative approach is to mediate the mappings through a pivot ontology. The following proposal extends an approach for mapping metadata elements between different stages of the production process of audiovisual media. Different metadata formats and standards are used in the workflow, containing metadata elements with similar and partly overlapping semantics, though not fully identical. In the context of the 2020 3D Media project[23], it has been attempted to model the metadata elements used throughout the production process in a format independent way by creating an ontology that models these elements and the relationships between them [4]. Modeling is done at a meta level, considering grouping and definition relations between the elements. The work considers three problems: (i) verify whether a given metadata element is defined by another given metadata element, (ii) find all metadata elements that are defined by a given metadata element and (iii) find all metadata elements that define a given metadata element. A demo application that addresses the first of these problems for a small set of production metadata items is available at `http://meon.joanneum.at`.

[23] `http://www.20203dmedia.eu`

OWL-DL is used to formally capture the semantics of the metadata elements and their relations. The ontology is format independent and contains the classes `Concept`, with subclasses `AtomicConcept` and `CompoundConcept`. Specific metadata properties are instances of these concepts. The relation `contains` exists between `CompoundConcept` and a set of concepts, the relation `defines` between concepts (bidirectional `defines` relations express identity of concepts). Additionally logical rules are used to infer implicit knowledge about relations between metadata elements. The existing implementation ignores specific data types of the metadata properties.

This approach can be extended for expressing mappings between multimedia metadata schemas and the Media Ontology. In addition to the schema independent ontology, schema specific ones are created for each standards following the same pattern. A new relation type is introduced, which relates concepts between the two ontologies. The relation is modeled as a class, that has properties for qualifying the relation (similar to the SKOS properties) and mapping instructions for data format conversion. The classes representing concepts in the schema specific ontology can be extended to carry additional information needed for mapping, e.g. XPath or binary key of the metadata element. The same rules can be used in both the generic and schema specific ontology for inference.

Figure 1 shows a schematic example for aligning some properties from EBU Core to the Media Ontology. The generic `meon` ontology represents the set of concepts, in that case `title`, `resolution`, `lines` and `columns`. It also models their relations, i.e. the compound of `lines` and `columns` is equivalent to the `resolution`. Relations are introduced to link concepts from the different ontologies. Hence, both `dc:title` and `ma:title` are completely aligned with `meon:mainTitle`. The value for these three properties being a literal, the mapping instruction is the identity function operating on simple datatypes.

The example of the frame size is more interesting. `ebu:formatWidth` (resp. `ebu:formatHeight`) is identical to `meon:colums` (resp. `meon:lines`) with potentially the help of a conversion of the number format. `ma:frameSize` is also equivalent to `meon:resolution`, again with a possible conversion of the format (which is specified by a function name in the relation). Using rules, we can infer from the relations within the `meon` ontology and between the ontologies that `ma:frameSize` defines both `ebu:formatWidth` and `ebu:formatHeight`, but not vice versa. In addition, because of modeling resolution as a compound concept in `meon`, we can also infer that `ebu:formatWidth` and `ebu:formatHeight` *together* define `ma:frameSize`. From the relations along the path between the elements we can collect the format mapping instructions to obtain a chain of functions that maps data types from EBU Core to the Media Ontology. These instructions are applied to the instances of the concepts encountered in the input document.

Benefits of this approach:

- Clean separation between generic concepts and schema specific concepts;
- Formal representation of the semantics of the properties in one format, which can e.g. also be used for validation;
- Inference is used to generate implicit relations and compound concepts.

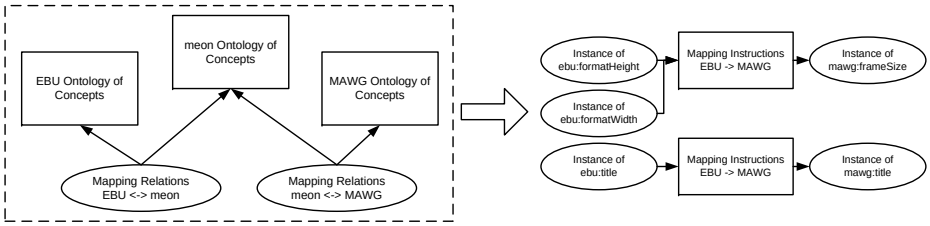


Fig. 1. Mapping using format independent ontology.

Drawbacks:

- Requires building ontology of properties for each schema, which may not be trivial;
- Scalability might be an issue with hundreds of concepts;
- Data type conversions might need built-in functions in the rule engine or external code to be executed.

4.4 Expressing Mappings with Built-in Properties

We present finally an alternative to the approach presented in the Section 4.3. The mappings are still mediated through a pivot ontology, but this ontology is directly related to the Media Ontology. This pivot ontology can be described as followed. Instances of the `MAWGMetadataProperty` class are described by the core set of annotation properties of the Media Ontology, while instances of the `StandardMetadataProperty` class are described by annotation properties of multimedia metadata schemas to be mapped. The `MetadataProperty` class is a superclass of these two classes. The `MetadataPropertyRelation` class characterizes the nature of the mapping relationship. It provides further information such as the transformation rule to operate on the values, the type of the mapping (e.g. exact) or whether it is a compound relationship or not. A priority operator can also be defined, in case various metadata properties from various standards can be aligned to a particular annotation property from the Media Ontology. This operator aims at defining a priority hierarchy for implementing a *SET* functionality in a API built on top of the Media Ontology.

The examples 1.6 and 1.7 illustrate this approach for the `ma:title` and `ma:frameSize` properties. Benefits of this approach:

- No specific representation format (e.g., OWL) of metadata standards is needed.

Drawbacks:

- No distinction between different versions of metadata formats. This issue could produce inconsistencies;
- No inference (e.g. between properties) is possible;

```
:MAWGMetadataProperty_21 a :MAWGMetadataProperty ;
  rdfs:isDefinedBy ma:title ;
  skos:inScheme <http://www.w3.org/2009/09/mediaont#> ;
  :hasMetadataPropertyRelation [
    :isCompositeRelation false ;
    :relationSemantic "exact" ;
    :hasStandardMetadataProperty [
      skos:inScheme <http://purl.org/dc/elements/1.1/>;
      rdfs:isDefinedBy dc:title ] ] .
```

Example 1.6. A simple mapping.

```
:MAWGMetadataProperty_10 a :MAWGMetadataProperty ;
  rdfs:isDefinedBy ma:frameSize ;
  skos:inScheme <http://www.w3.org/2009/09/mediaont#> ;
  :hasMetadataPropertyRelation [
    :isCompositeRelation true ;
    :relationSemantic "exact" ;
    :hasStandardMetadataProperty [
      skos:inScheme <http://www.ebu.ch/metadata/ontologies/>;
      rdfs:isDefinedBy [ owl:unionOf (
        [ a owl:Restriction ; owl:onProperty ebu:formatWidth ;
            owl:allValuesFrom xsd:int]
        [ a owl:Restriction ; owl:onProperty ebu:formatHeigth ;
            owl:allValuesFrom xsd:int] ) ] ; ] ; ] .
```

Example 1.7. A complex mapping.

5 Conclusion and Future Work

This paper addresses the interoperability issue between multimedia metadata formats. The related work described in section 2 and the numerous use cases summarized in [7] show that there is a need for solving this issue. We have presented a core set of annotation properties defined in the Media Ontology developed by the W3C Media Annotations Working Group. Furthermore, we have discussed how mapping relationships between this core set of annotation properties and the multimedia metadata standards can be represented, either directly using semantic web languages (SKOS, OWL, or the forthcoming RIF[24] recommendations) or through a pivot ontology.

Each approach presents benefits and drawbacks that can be grouped in the following criteria: *complexity*, *scalability* and *reasoning capabilities*. The listing of these benefits and drawbacks is currently done ad-hoc. As such, future work consists of an in-depth evaluation in which each of the criteria is measured for the different approaches. Expressing direct mappings is intuitive and provide scalability. However, it requires that the metadata formats to be aligned have been formally represented in SKOS, RDFS or OWL. The use of a pivot ontology tends to be a more generic solution which has the price of complexity in terms of the number of triples generated.

[24] http://www.w3.org/TR/rif-core/

Future work deals primarily with the recommendation of the Media Ontology. Its coverage is still evolving and profiles might be introduced, in particular, for offering a degree of variability in the way mappings with other standards is formalized. Another important milestone planned is the design of an API on top of the Media Ontology. The main purpose of this API will be the implementation of appropriate *GET* and *SET* functionalities. One of the open issues concerns the implementation procedure to follow in case of collision between various semantic mappings. The priority operator introduced in the Section 4.4 is a useful contribution with this respect.

Acknowledgments

The authors would like to thank all the participants of the *W3C Media Annotations Working Group* for their willingness to discuss the core ontology, the mappings between metadata standards, and their implementation. This work has been partially supported by the THESEUS Program, which is funded by the German Federal Ministry of Economics and Technology, and under the 7th Framework Programme of the European Union within the ICT project "Presto-PRIME" (FP7 231161).

References

1. Isabel Cruz, Huiyong Xiao, and Feihong Hsu. An Ontology-Based Framework for XML Semantic Integration. In *International Database Engineering and Applications Symposium*, pages 217–226, Coimbra, Portugal, 2004.
2. Mario Döller, Florian Stegmaier, Harald Kosch, Ruben Tous, and Jaime Delgado. Standardized Interoperable Image Retrieval. In *ACM Symposium on Applied Computing (SAC), Track on Advances in Spatial and Image-based Information Systems (ASIIS)*, 2010, to appear.
3. Mario Döller, Ruben Tous, Matthias Gruhne, Kyoungro Yoon, Masanori Sano, and Ian Burnett. The MPEG Query Format: On the way to unify the access to Multimedia Retrieval Systems. *IEEE Multimedia*, 15(4):82–95, 2008.
4. Martin Höffernig and Werner Bailer. Formal Metadata Semantics for Interoperability in the Audiovisual Media Production Process. In *Workshop on Semantic Multimedia Database Technologies (SeMuDaTe 2009), co-located with the 4th International Conference on Semantic and Digital Media Technologies (SAMT2009)*, Graz, Austria, 2009, to appear.
5. Ian Horrocks, Peter F. Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, and Mike Dean. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. W3C Member Submission, 2004. `http://www.w3.org/Submission/SWRL/`.
6. Dublin Core Metadata Initiative. Dublin core metadata element set - version 1.1: Reference description. `http://dublincore.org/documents/dces/`, 2008.
7. WonSuk Lee, Tobias Bürger, Felix Sasaki, and Véronique Malaisé. Use Cases and Requirements for Ontology and API for Media Object 1.0. W3C Working Draft, 2009. `http://www.w3.org/TR/media-annot-reqs/`.

8. WonSuk Lee, Tobias Bürger, Felix Sasaki, Véronique Malaisé, Florian Stegmaier, and Joakim Söderberg. Ontology for Media Resource 1.0. W3C Working Draft, 2009. http://www.w3.org/TR/mediaont-10/.
9. MPEG-7. Multimedia Content Description Interface. ISO/IEC 15938, 2001.
10. National Electrical Manufacturers Association. Digital Imaging and Communications in Medicine (DICOM). ftp://medical.nema.org/medical/dicom/2008/, 2008.
11. Chris Poppe, Gaëtan Martens, Erik Mannens, and Rik Van de Walle. Personal Content Management System: A Semantic Approach. *Journal of Visual Communication and Image Representation*, 20(2):131–144, 2009.
12. Raphaël Troncy, Óscar Celma, Suzanne Little, Roberto García, and Chrisa Tsinaraki. MPEG-7 based Multimedia Ontologies: Interoperability Support or Interoperability Issue? In *1st International Workshop on Multimedia Annotation and Retrieval enabled by Shared Ontologies*, 2007.
13. Guangming Xing, Zhonghang Xia, and Andrew Ernest. Building automatic mapping between XML documents using approximate tree matching. In *ACM Symposium on Applied Computing (SAC)*, pages 525–526, Seoul, Korea, 2007.
14. Xia Yang, MongLi Lee, Tok Wang Ling, Lee Tok, and Wang Ling. Resolving Structural Conflicts in the Integration of XML Schemas: A Semantic Approach. In *International Conference on Conceptual Modeling (ER)*, pages 520–533, Chicago, USA, 2003.

Formal Metadata Semantics for Interoperability in the Audiovisual Media Production Process

Martin Höffernig and Werner Bailer

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

Abstract. In the different stages of the production process of audiovisual media such as movies, a number of different metadata properties exist. Different metadata formats and standards are used in the stages of the process, containing metadata properties with similar and partly overlapping semantics, though not fully identical. We attempt to model the metadata properties used throughout the production process in a format independent way by creating an ontology that models these properties and the relations between them. Modeling is done on a high level, considering grouping and definition relations between the properties. We apply the proposed approach to the problem of verifying whether a set of metadata properties can be unambiguously derived from another and present a web based demo application for this use case.

1 Introduction

1.1 Motivation

A large number of different metadata properties exist throughout the audiovisual media production process. These properties are produced and consumed at different stages of the process. Typically the different devices and tools used in the chain also make use of different metadata representations. The SMPTE metadata dictionary [11] alone lists nearly 1,500 metadata properties, and there are many more that are not covered by this dictionary [1]. Other relevant standards in the media production process are for example the MXF Descriptive Metadata Scheme 1 [3], MPEG-7 Multimedia Content Description Interface [6] and EBU P_Meta [9]. Current trends such as 3D and multi-view content add additional requirements to the metadata representation, as the relations between different media properties need to be described (from high level information such as relating different views of the same scene down to precise measurements such as camera calibration parameters).

When analyzing the various properties and their definitions in different stages of the process or in different metadata formats, we encounter a number of properties which represent the same information, but modeled differently or only partly overlapping. For example, at capture, a number of parameters are available, such

as the resolution of the sensor, its aspect ratio, the temporal sampling rate, etc. At a later stage, the header of a stream might contain an identifier of the video standard used, which implies the values for a number of these parameters. In order to support content exchange and automation in the production process, it is necessary to establish metadata interoperability between the steps of the process. Due to the multitude of metadata properties and formats, which are often tailored toward the specific needs of a certain step in the process, it is utopian to expect that a single format serving the needs of all steps in the process can be defined, that will also be adopted by all the devices and tools involved. We thus need to deal with the diversity in terms of metadata and establish interoperability by well-defined semantics of the different metadata properties, so that they can be mapped between the different stages of the process.

This work presents a first step in this direction. We aim to model the concepts behind the metadata properties in the process, leaving specifics of formats such as data types aside. These things can be addressed in an additional layer on top. In the remainder of this section we discuss approaches for solving related problems. We then analyze the aspects of this interoperability problem in more detail in Section 2 and present the proposed approach in Section 3. In Section 4 we discuss a prototypical implementation of the approach for an ontology covering a small set of metadata properties and Section 5 concludes the discussion.

1.2 Related Work

In [8] the automation of media production processes by using a workflow management system is discussed. In that work, the open source workflow language YAWL (Yet Another Workflow Language [12]) has been chosen and extended to fit the area of film production (YAWL4Film[1]). YAWL4Film contains workflow patterns that support the production crew in collecting, creating, and distributing required documents and data for certain production tasks. For example, the process for a daily shooting procedure has been modeled, in which documents such as time sheets for cast members or the daily-shooting progress reports are created and distributed automatically by the workflow system. YAWL is based on XML technologies and all the data being processed during the process steps is defined using custom XML Schemata. Due to known limitations of XML Schema describing semantics [7], interoperability to existing metadata standards in the media production process is very limited.

In the different stages of the media production process, different types of metadata are needed, such as descriptive, technical, structural, composition, and editing metadata. In [10] these metadata types are listed and allocated to the concerning production stages. Furthermore, relevant metadata standards, for describing these different types of metadata have been identified. In [1] the requirements for a metadata model for audiovisual media production have been developed and discussed, and existing standards have been analyzed. The conclusion of this work is that none of the current metadata standards is able to

[1] http://www.yawl4film.com/

achieve all the defined requirements. However, interoperability between metadata standards needs to be enabled in order to exchange metadata properties between different standards.

The Simple Knowledge Organization System Reference (SKOS) provides a vocabulary to classify concepts and to describe how they relate to others concepts [5]. Semantic relations, such as narrower, broader, and related are available for describing relations between SKOS concepts. It is of course possible to build up a classification scheme using the SKOS relations. However, we want to model more complex relations, for example, if a concept defines other concepts or if a concept can be substituted by others. Describing such relations are out of the scope of SKOS[2].

The W3C Media Annotations Working Group[3] also deals with the interoperability issue of metadata formats. Their goal is to develop a simple ontology of core metadata properties for audiovisual content and an API for accessing these properties from descriptions in a range of formats. This clearly needs mappings between the considered formats and the proposed set of properties. The working draft containing the core vocabulary to describe media resources is available at [4]. In contrast to this work we do not want to define mappings between different metadata standards, but rather to describe semantic relations between format independent metadata properties.

2 Problem Definition

Different metadata properties represented in different metadata formats exist in the media production process. However, we encounter a number of properties which represent – at least partially – the same information, but modeled differently or with only partly overlapping semantics. In order to support content exchange and automation in the production process, it is necessary to establish metadata interoperability between different metadata models and representations being used in the steps of the process. As first step to solve this interoperability issue we model the relations between metadata properties or groups of metadata properties. We then define a set of queries that our system needs to be able to answer. These queries yield information about the how metadata properties in the different stages of the process are related, they do not yet implement conversion between metadata formats.

2.1 Relations

Definition A metadata property (or group of properties) A defines another metadata property (or group of properties) B, if B can be derived without any semantic ambiguity from A by some mapping/conversion.

[2] Compare the discussion about the usage of SKOS mappings for this purpose: http://lists.w3.org/Archives/Public/public-media-annotation/2009Mar/0067.html

[3] http://www.w3.org/2008/WebVideo/Annotations/

Equivalence A metadata property (or group of properties) B is equivalent to another metadata property (or group of properties) B, if A defines B and B defines A.

At this point it is not relevant which specific metadata formats are used and what kind of data type is used to represent a specific metadata property. It is only important to model the concept represented by a metadata property and the relations between them.

In this paper we use the following notation. To formally express a group of metadata properties, the conjunction ($\wedge$) is employed. Additionally, the implication operator ($\rightarrow$) is used to express the definition relation between metadata properties (or group of) properties, and the equivalence operator ($\leftrightarrow$) describes the equivalence relation between metadata properties (or group of) properties.

As an example, assume that there are the following metadata properties of a video: number of lines, number of columns, spatial resolution, frame rate, and a video payload identifier of a container file format, that describes the video standard by an identifier[4]. Spatial resolution is equivalent to a metadata properties group containing lines and columns. Furthermore, the payload identifier defines lines, columns, and frame rate. These two statements can be formally expressed as equations 1 and 2. Since spatial resolution is equivalent to the group of lines and columns, the payload identifier also defines the resolution which is expressed in equation 3.

$$resolution \leftrightarrow lines \wedge columns. \tag{1}$$

$$payload\ identifier \rightarrow lines \wedge columns \wedge frame\ rate. \tag{2}$$

$$\begin{aligned}(payload\ identifier \rightarrow\ & lines \wedge columns \wedge frame\ rate) \wedge \\ & (lines \wedge columns \leftrightarrow resolution) \rightarrow \\ & (payload\ identifier \rightarrow resolution)\end{aligned} \tag{3}$$

2.2 Queries

Expressing definition and equivalence relations between metadata properties enables to infer information about interoperability between metadata properties. The following three types of queries have been identified[5]. It holds for all types of queries, that some queries may be answerable directly by the facts represented in the ontology while others need inference.

1. Verify whether a given metadata property is defined by another given metadata property or not. For example, it should be verified if there exists an definition relation between payload identifier and resolution (equation 4). Since there is an inferred definition the response to this query is yes.

[4] Such an property exists e.g. in the header of an MXF file.

[5] Note that in the description of the queries "metadata property" stands for single metadata properties as well as groups of metadata properties.

$$payload\ identifier \rightarrow resolution\ ? \tag{4}$$

2. Find all metadata properties that are defined by a given metadata property. An example is the query in equation 5. Lines, columns, and frame rate are direct results, while resolution is an inferred result to this query.

$$payload\ identifier \rightarrow ? \tag{5}$$

3. Find all metadata properties that define a given metadata property. As an example all metadata properties which imply lines should be listed (cf. equation 6). In this case, payload identifier and resolution are the results.

$$?\ \rightarrow lines. \tag{6}$$

To simplify matters, in all of the examples above only single metadata properties have been used as query parameters. In addition, only single metadata properties are listed as results. However, the result set could be expanded to include groups of properties, e.g. the group (lines, columns) in addition to resolution. When dealing with groups of metadata properties we can distinguish two types: those explicitly defined in our ontology and those not explicitly defined in the ontology but stated in the query or emerging from the result. For example, equation 7 contains an explicitly defined group of metadata properties since this group has been explicitly expressed in equation 1. On the other hand, the metadata group contained in equation 8 is only created in the query.

$$payload\ identifier \rightarrow (lines \wedge columns) \tag{7}$$

$$payload\ identifier \rightarrow (resolution \wedge frame\ rate) \tag{8}$$

Additionally, query results can also contain groups of metadata properties that are not explicitly defined. For example, valid results of the query in equation 5 are among others (lines $\wedge$ columns) and (lines $\wedge$ resolution).

3 Proposed Approach

In this section we propose an approach for expressing the required relations between metadata properties (as discussed in Section 2). As a proof of concept the approach is applied to solve the verification query task (cf. equation 4). We propose the use of an ontology which is called *meon*[6] for the formal representation of metadata properties and the relations between them. Furthermore, logical rules are applied to infer new knowledge.

OWL-DL [2], which is a subset of the Web Ontology Language, is used to formally capture the semantics of the metadata properties and their relations. The class `Concept` models the general concept represented by a metadata

[6] prefix `meon`: http://www.20203dmedia.eu/meon#

property in the ontology. Subclasses of class `Concept` are `AtomicConcept` and `CompoundConcept`. Class `AtomicConcept` represents all single metadata properties, while class `CompoundConcept` describes groups of metadata properties containing at least two metadata properties. Property `contains` describes that an instance of class `Concept` is part of a group of metadata properties (i.e. part of an instance of class `CompoundConcept`). In order to model the definition relation between two metadata properties (or group of), the transitive property `defines` is used. This property can be applied between instances of class `Concept`. Since the equivalence relation is just the result of a bidirectional definition relation, an appropriate usage of this property is expressive enough to model also the equivalence relation. The metadata properties in the example presented in Section 2 are represented by the ontology in Figure 1[7]. In this ontology `PayloadIdentifier`, `FrameRate`, `Lines`, `Columns`, and `Resolution` are instances of class `AtomicConcept`, `CC_1` and `CC_2` are anonymous instances of class `CompoundConcept`. `CC_1` represents a group of metadata properties containing frame rate, lines, and columns. Additionally the metadata properties lines and columns form another metadata group described by `CC_2`.

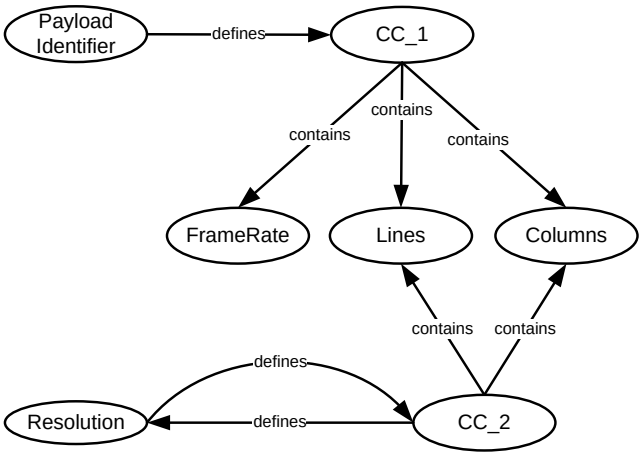


Fig. 1. Example of an ontology for describing metadata properties and their relations between.

In order to model the verification query task (cf. equation 4) as a proof of concept of the proposed approach, we split groups of metadata properties into all possible combinations and then infer new relations in the ontology. Therefore logical rules have been created to semantically express the knowledge required to solve this task. These rules make implicit knowledge in the ontology explicit by adding new instances and relations which enable the subsequent rea-

[7] For the sake of simplicity, the `meon` namespace and `rdf:type` relations have been omitted.

soning and querying steps. The Jena rules syntax[8] is used for defining the rules. One set of rules is responsible to split metadata groups into their combinations and to establish definition relations between the parent metadata group and they newly created ones. Another set of rules expresses the equivalence between metadata groups. An important prerequisite for determining the equivalence between metadata groups is that it must be possible to express the number of metadata properties contained in the group. Therefore an additional property (`countContains`) is added to the ontology, and a rule containing a custom procedural builtin[9] is used to compute the number of metadata properties for each metadata properties group. An example rule for expressing the equivalence relation between two metadata groups containing two metadata properties is shown in Figure 2. First, two ambiguous instances of class `CompoundConcept` (`?cc1` and `?cc2`) containing exactly two instances of class `AtomicConcept` are identified. Then it is verified whether in `?cc1` and `?cc2` the same instances of class `AtomicConcept` (`?ac1` and `?ac2`) are included. In case that (`?cc1` and `?cc2`) are equivalent, two new definition relations between them are added to the ontology. Another rule expresses that the `defines` relation between an instance of class `CompoundConcept` and a instance of class `AtomicConcept` also infers a definition relation (`defines`) between them. It is obvious that a metadata property is equivalent to itself. This fact is also explicitly added by a rule.

```
@prefix meon: <http://www.20203dmedia.eu/meon#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

[equivalence_between_compound_concepts_containing_2_atomic_concepts:
  (?cc1 meon:countContains "2"^^xsd:int),
  (?cc2 meon:countContains "2"^^xsd:int),
  notEqual(?cc1, ?cc2),
  (?cc1 meon:contains ?ac1),
  (?cc2 meon:contains ?ac1),
  (?cc1 meon:contains ?ac2),
  (?cc2 meon:contains ?ac2),
  notEqual(?ac1, ?ac2),
  ->
  (?cc1 meon:defines ?cc2),
  (?cc2 meon:defines ?cc1)
]
```

Fig. 2. Rule for infering the equivalence between two metadata property groups consisting of two metadata properties.

After applying the presented rules to the example ontology depicted in Figure 1, it can be derived that payload identifier defines resolution (cf. equation 3).

[8] http://jena.sourceforge.net/inference/index.html#rules
[9] http://jena.sourceforge.net/inference/index.html#RULEextensions

The extended example ontology after applying the rules is shown in Figure 3. First a new anonymous instance `CC_4` of class `CompoundConcept`, which is a subgroup of `CC_1`, is added to the ontology by the according rule. Additional subgroups of `CC_1` are created as well but for simplicity they are not shown in Figure 3. Then the equivalence between `CC_4` and `CC_2` is computed (cf. rule shown in Figure 2). According to the transitive behavior of property `defines` there is now a definition relation between the instances `PayloadIdentifier` and `Resolution`.

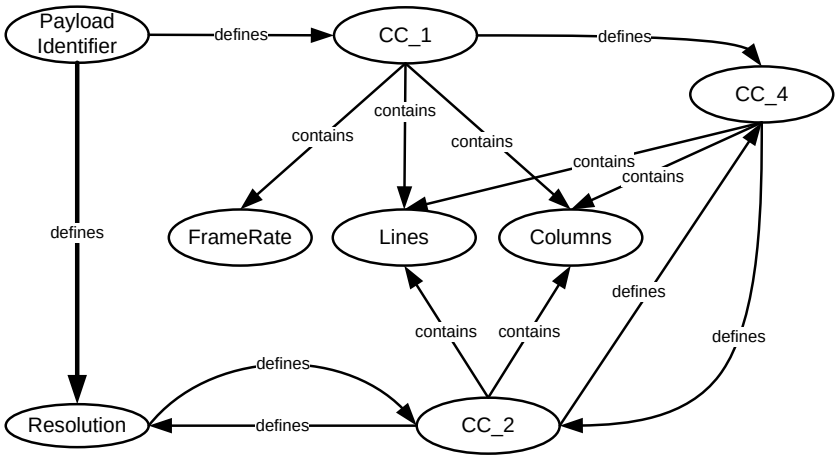


Fig. 3. Extended ontology after applying rules to example ontology (shown in Figure 1) in order to accomplish the verification query task.

4 Demo Application

To demonstrate our approach to solve the verification query task we have implemented a web based demo application. As described in Section 3, the presented Jena rules are applied to the ontology containing the metadata properties. Then new definition relations are inferred. Although transitive reasoning is a basic task for an OWL-DL reasoner[10], this task is also performed by a Jena rule. The reason for this design decision is that it would be the only usage of an OWL-DL reasoner during the whole process, and due to performance considerations this task has been moved to the rule reasoning block. The next step is to perform a SPARQL[11] select query to verify whether there is a definition relation between the selected metadata properties or not. For example, the SPARQL select query shown in Figure 4 is used to query for relations between the metadata properties

[10] e.g. http://clarkparsia.com/pellet
[11] http://www.w3.org/TR/rdf-sparql-query/

payload identifier and resolution. Finally it is checked if the definition relation, represented using the property `defines`, is part of the SPARQL query result in order to determine the result of a verification query task.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX meon: <http://www.20203dmedia.eu/meon#>

SELECT DISTINCT ?property WHERE
{ meon:PayloadIdentifier ?property meon:Resolution }
```

Fig. 4. SPARQL select query to verify equation 4.

The user interface of our web application[12] is shown in Figure 5. The web application has been created using the Google Web Toolkit (GWT)[13] and is deployed on Apache Tomcat 6.0[14]. All OWL processing tasks (including rule reasoning and executing SPARQL queries) are performed using Jena 2.6.0[15]. After loading an ontology all single metadata properties are displayed. The user selects the metadata properties to be mapped and performs the verification.

5 Conclusion and Future Work

We have analyzed the problem of mapping metadata properties between stages of the audiovisual media production process with possible (partly) different semantics. Three basic types of queries have been identified. The proposed approach consists of an ontology modeling simple as well as compound metadata properties as well as their relations. Jena rules are used to infer implicit knowledge about the metadata properties. As a proof of concept, the approach has been successfully applied to the verification query type, i.e. verifying whether a metadata property can be unambiguously derived from another. A web application has been implemented as a demonstrator.

The next step is to also implement the other query types. In particular, groups of metadata properties implicitly defined in the query (but not modeled in the ontology) need to be supported. In addition, the ontology will be extended to cover a wide range of metadata properties used in the audiovisual media production process, as well as their relations. Furthermore, if we also describe the relation between the metadata properties in our ontology and specific formats and consider data type issues, the approach could be useful to enable automatic conversion between metadata formats.

[12] The application can be accessed from http://meon.joanneum.at.
[13] http://code.google.com/webtoolkit/overview.html
[14] http://tomcat.apache.org/index.html
[15] http://jena.sourceforge.net/

Fig. 5. meon web application.

Acknowledgements

The research leading to this paper was partially supported by the European Commission under the IST contract FP7-215475, "2020 3D Media: Spatial Sound and Vision" (http://www.20203dmedia.eu).

References

1. Werner Bailer and Peter Schallauer. Metadata in the audiovisual media production process. In Michael Granitzer, Mathias Lux, and Marc Spaniol, editors, *Multimedia Semantics - The Role of Metadata*, volume 101 of *Studies in Computational Intelligence*, pages 65–84. Springer, Jun. 2008.
2. Mike Dean and Guus Schreiber. OWL Web Ontology Language: Reference. W3C Recommendation, 10 February 2004. http://www.w3.org/TR/owl-ref/.
3. Material Exchange Format (MXF) - Descriptive Metadata Scheme-1. SMPTE 380M, 2004.
4. WonSuk Lee, Tobias Bürger, Felix Sasaki, Véronique Malaisé, and Florian Stegmaier. Ontology for Media Resource 1.0. W3C Working Draft, June 2009. Editor's draft at http://www.w3.org/2008/WebVideo/Annotations/drafts/ontology10/ First-Draft/Overview.html, to appear at http://www.w3.org/TR/mediaont-10.
5. Alistair Miles and Sean Bechhofer. SKOS Simple Knowledge Organization System Reference. W3C Candidate Recommendation, 17 March 2009.

6. Information Technology - Multimedia Content Description Interface (MPEG-7). ISO/IEC 15938, 2001.
7. 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.
8. Chun Ouyang, Marcello La Rosa, Arthur H.M. ter Hofstede, Marlon Dumas, and Katherine Shortland. Toward web-scale workflows for film production. *IEEE Internet Computing*, 12(5):53–61, 2008.
9. EBU P_META 2.0 Metadata Library. EBU Tech 3295-v2, Jul. 2007.
10. Konstantin Schinas, Wolfgang Schmidt, Franz Höller, Herwig Zeiner, Werner Bailer, and Michael Hausenblas. D3.2.1 Metadata in the Digital Cinema Workflow and its Standards. Public deliverable, IP-RACINE (IST-2-511316-IP), 2005. http://www.ipracine.org/documents/Del_3_2_1_metadata.pdf.
11. Metadata dictionary registry of metadata element descriptions. SMPTE RP210.11, 2004.
12. W. M. P. van der Aalst, L. Aldred, M. Dumas, and Ter A. H. M. Hofstede. Design and implementation of the YAWL system. *Proceedings of the 16th International Conference on Advanced Information Systems Engineering (CAiSE'04)*, 2004.

Automatic Annotation of Web Images combined with Learning of Contextual Knowledge

Thomas Scholz, Sadet Alčić, and Stefan Conrad

Institute for Computer Science
Databases and Information Systems
Heinrich Heine University
D-40225 Düsseldorf, Germany
thomas.scholz@uni-duesseldorf.de
{alcic,conrad}@cs.uni-duesseldorf.de

Abstract. This paper introduces an approach for automatic image annotation which is based on contextual knowledge extracted from semantic rich web documents containing images. The knowledge itself is organized in ontologies and extended by learning algorithms for new contextual information. For this purpose, the contextual background of a picture is used for the annotation process and later in the image retrieval process. The paper shows the design of our system and how the different parts work together to enable and improve the annotation procedure. We created learning algorithms for harvesting new contextual information and thus improve context analysis. Finally, we evaluate our methods with a set of sports web pages.

1 Introduction

Today, we are facing a very huge and rapidly increasing number of web images. Controlled access to this large repository is challenging. Content-based Image Retrieval (CBIR) uses low-level feature extraction for retrieval modes. Compared to the human way to handle images and pictured objects this presents a totally different approach, which lacks in high-level semantics. The problem is known as the *Semantic Gap* [4]. There are many low-level features that can be extracted from images, but in general it is difficult to find the corresponding interpretation. Any additional information to the context of the image can improve the retrieval quality. Unfortunately, such information is very difficult to be estimated only based on low-level features.

An approach on a higher level are manually applied annotations provided by human annotators. They are very useful, but expensive in time and human effort. Further, the problem with the Semantic Gap still is not solved but relinquished to human.

In a web environment text contents often provide semantic information on a higher level. According to [1] contextual information can increase the quality of annotations which are made by human. This applies more than ever when ontologies with hierarchical structures build the backbone of the contextual knowledge.

Having considered this advantages of contextual information for manual image annotation, our basic idea is using contextual information for automatic image annotation. In this way we want to get additional information and thus improve the retrieval quality for web images.

The preliminary considerations of our approach can be summarized as follows:

1. The algorithms and data structures for the whole annotation process should be simple.
2. Only a few start information should be needed.
3. The created system should be able to learn new information.

The remainder of this paper is as follows: In Sect. 2 we review some related works and highlight the differences to our work. After that we introduce our image annotating approach in Sect. 3 by giving a short overview the system's components before going into detail. In Sect. 4 we evaluate our methods by checking the resulting metadata and putting the automatic annotated images into a retrieval situation. And finally, the paper is concluded with the evaluation results and a short outline to future works.

2 Related Works

The early approach to searching in image databases was the Content-based Image Retrieval (CBIR) where a selection of low-level features formed all capabilities for query answering [10]. Although CBIR can be enhanced by relevance feedback techniques [11, 12], this way of browsing in image databases is limited by the *Semantic Gap*. While such representation is manageable for computers, handling of low-level features is very difficult for humans, e.g. in the retrieval situation, where a query has to be specified.

More promising approaches are image annotation techniques where image content is described by textual keywords which later can be used as basis for image retrieval. There the annotations are either associated with the whole image or with regions. In the latter case some kind of image segmentation is needed. There are different approaches to generate the resulting annotations which vary from manual annotations given by human annotators and semi automatic approaches to full automated approaches using relevance models between annotated training sets and their low-level descriptions [8].

One of the problems which occurs in image annotation approaches is the word disambiguation. A possible solution is often the use of an dictionary to extends keywords.

Another kind of extension is ontology-based image annotation and brings a new architecture of conceptual image annotations [6, 7]. The new semantic information are gained by the conceptual structures and relationships or leads to new models to describe images [9]. Approaches like [5] are learning from ontologies or discover knowledge in ontologies.

In our application we combine these two approaches: on the one hand an ontology model is used to improve automatic image annotation and on the other hand for storing new information which are the results of a learning procedure. Thus we are able to gain more contextual information and generate a growing and dynamical ontology.

3 System Design

3.1 Overview

The proposed system consists of two main modules: The DUNCAN component, which extracts the image context and annotations from the article, while the PAGANEL component learns new context information from the results of the DUNCAN module. An overview is shown in Fig. 1. In the following section all parts of the system will be discussed in detail.

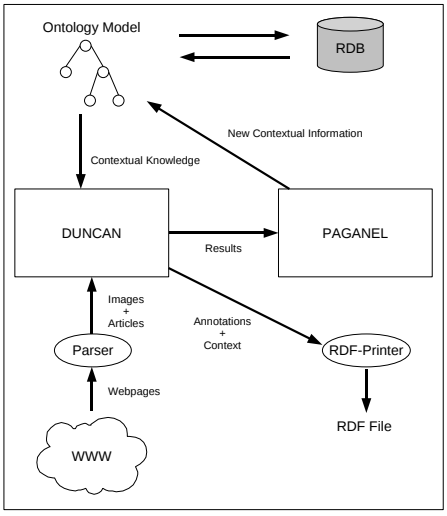


Fig. 1. System Overview

3.2 System Components

WWW and Parser At the beginning of the automatic annotation process, a parser considers pairs of images and corresponding articles of a given web page like proposed in [13].

Then the stopwords of the article are removed. Now three different types of textual information are available: *Metadata* of the image (e.g., alternative text),

image caption (if the webpage has such design, otherwise this text is empty) and the *full text* of the article.

Ontology Model Our ontology model for image annotation is quite simple. It has two types of entities: classes and attributes. Classes represent a context, while attributes are extra information which allow to determine a context and contain extra information for the annotation process. In the tree representation the attributes are leaves while the classes are inner nodes.

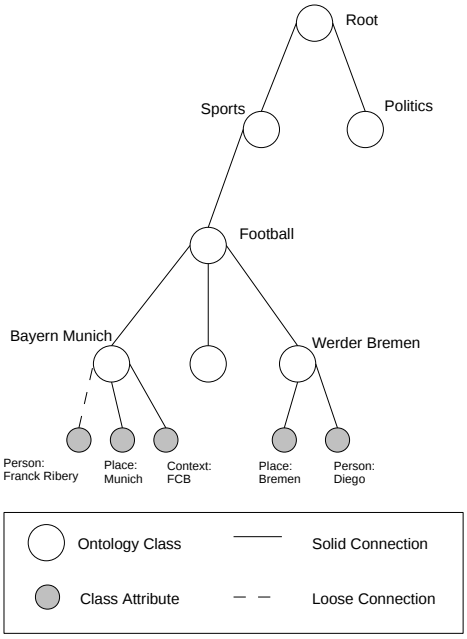


Fig. 2. A sample ontology

The connection between the classes of our ontology are solid, and thus reliable. Loose connections can occur between classes and new learned attributes. This distinction is very important for the learning process, because only the attributes with a higher occurrence will get a solid connection to their corresponding class and are used in the context detecting step.

DUNCAN Module The main tasks of the DUNCAN module are to find a context class, to extract the image annotations of the text (person, place, object and action) and to send the results to the RDF-Printer and the PAGANEL component.

DUNCAN uses the ontology model to determine a concrete context class for an image. Thus all three text types are needed in order to search for class attributes in the text. To every context class a score value is maintained. This value is increased if the text has attributes which belong to the context class. The score value is provided by the following function p

$$p(a, c) = \sum_{i=1}^{n} f(w_i, c) * (n - i) \tag{1}$$

where w_i represents the *i-th* word in the article a (if an expression is larger than one word, the first word defines the position), n is the number of words in the whole article and c the context class of the ontology.

The function f

$$f(w, c) = \begin{cases} 1 & : \quad w \subseteq \Phi(c) \\ 0 & : \quad w \not\subseteq \Phi(c) \end{cases} \tag{2}$$

is an indicator function which shows if the word w is an attribute of the context c or not. $\Phi(c)$ is the set of attributes which solidly belong to the context c. p is further designed in such way that words at the beginning of an article assign more importance for the context than words at the end. The class with the highest score provides the final context of the image. The people names and location information are extracted from articles using the OpenNLP library [14]. The objects and actions were extracted from the surrounding text of the image (alternative text, image caption and heading of the article) using a dictionary to unmask words as actions or objects.

Finally, the results (context class with annotations) are sent to the RDF-Printer to create a RDF-File and to the PAGANEL component for the learning process.

PAGANEL Module The main task of the PAGANEL component is to extend the ontology by learning new contextual information. The results of the DUNCAN module form the basis of this learning process. New attributes are obtained from the last annotations while the ontology class is given by the determined context. For example, if the last image is annotated with the person "Franck Ribery" and the context "Bayern Munich", PAGANEL extends the ontology class "Bayern Munich" with the attribute "Franck Ribery" of type *person*.

Additionally, PAGANEL retrieves the header of the article, which is used to extend the ontology class with attributes of the general type *context*. Tab. 1 shows an excerpt of class attributes of our example ontology.

Fig. 3 shows how the knowledge stored in the ontology is extended during the learning process. Elements, that are new in the ontology, get a loose connection to a corresponding class. This loose connection becomes solid if this relationship is learned more often (appearance count is over a specific threshold). The threshold depends on the length of the learning period.

Generally, the learning process consists of two periods: A class level period for each ontology class and a general level period for the whole ontology. PAGANEL

Table 1. Class Attribute Table

Class Attribute	Ontology Class	Type	Appearance
Franck Ribery	Bayern Munich	Person	2
Corner Kick	Football	Action	3
Jentson Button	Brawn GP	Person	1

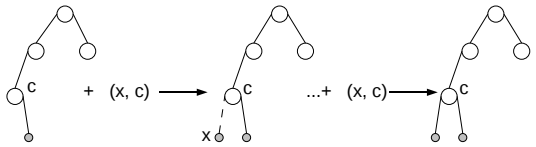


Fig. 3. Learning Process: Attributes are learned in 2 steps.

maintains the number of annotated images with a certain context. Hence, the system can establish learning periods for every context class. After a period PAGANEL deletes all attributes with loose connections. Moreover an attribute can loose the status of a solid connection and become a loose connection when its occurrence in the database is very low. For example all attributes with an occurrence score lower than 3 will get a score decreased by one after a class level learning period.

It must also be noted that this threshold is relative to the total number of annotated images because every context class has its own counter for the class level period. The counter for the whole ontology counts all annotated images. When this counter signals that a learning period is over, the knowledge of the ontology is getting reorganized.

At the beginning, the general distributed contextual information are collected at the upper class. This means: When a learned information appears with a certain percentage in all lower classes, the information (the class attribute) changes its connection to the higher class (see Fig. 4). The percentage can be e.g. 50%. In our football example the "header" can be learned in the context of the different football classes, but it is only an indication of football in general.

After that, contextual information, which appears in more than one ontology class, is moving to the class with the most appearance value. Some pieces of knowledge are learned in the wrong context, but the basic idea is that the growing number of results effects a more and more increasing accuracy in the learning process.

Finally, we add the possibility for attributes to change their context class (see Fig. 5). In our sports example a football player could play for a new football club. In this case our system increases appearance to the new context class and decreases the appearance in the old context class. So the old connection can get weaker and disappear at the end. Thus, we take into account that knowledge is time dependent and dynamic.

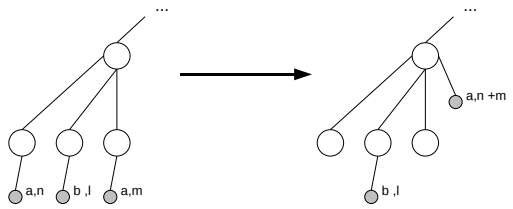


Fig. 4. Learning Process: The same distributed information moves to the upper class.

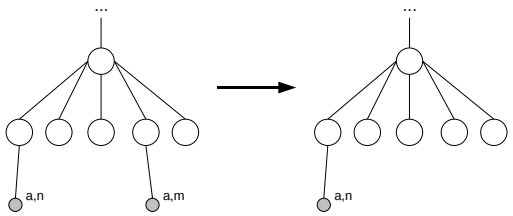


Fig. 5. Learning Process: The class attribute with the greater appearance takes the attributes in the same level

The change of a context also reduces the possibility of keeping wrong learned information for a long time in the ontology. This, and the implementation of loose and solid connections decelerate the learning progress, but increase the quality of new extracted knowledge. Eventually repetitions are even necessary for human, if they want to learn new things.

"Hoffenheim" is "Sinsheim" (this club has a small hometown so that they play in a bigger neighbor city).

4 Evaluation

4.1 Experiment Design

To evaluate our approach we focus on two aspects: First, we inspect the results of the annotation process. Secondly, the annotated images are tested in an image retrieval scenario.

At first we start with a training set of 130 images which are used to extend the initial context descriptions of the ontology classes. Then we crawled 500 images about German football from 10 different web domains (sports portals, news pages, pages from broadcast stations etc.) for our experiments. For the

contextual background an initial ontology to German football was designed with only a few class attributes per ontology class. This means that every context class has equal or less than 5 start information. One of these information contains the location. For example the ontology class "Bayern Munich" has got the class attribute "Munich" with type "place". Other start information are e.g. aliases of the football club. Persons are not inserted in the initial ontology. The PAGANEL component should learn persons by itself.

4.2 Annotation Quality

In the first part of the evaluation we review the concrete annotations and the allocation of the images to a context class. There we check if the image is annotated with the correct person, has the right location, context and so on. In addition we measure the quality of the learning process by proving the classification of a person into the right context class.

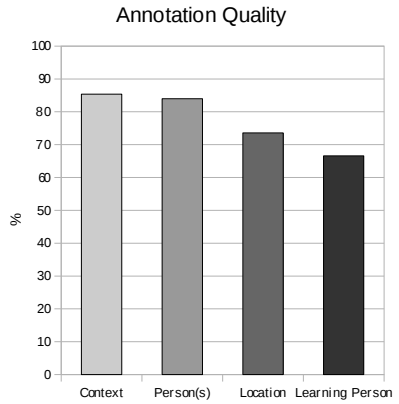


Fig. 6. Annotation Quality

The results of the annotation quality (Fig. 6) are remarkable: The correct context is found with a precision of 85,4%, while 84,0% of the images are annotated with the correct person names. In 73,6% of the cases our system determines the right location of the image.

The quality of the context analysis is very interesting because there are some articles which are quite difficult to analyze in view of the image's contextual environment. Sometimes articles concern a game of two football clubs or speculations which tell about a player who is going to change his football club and the image shows him in the actual club jersey or the player plays for the national team or something like that.

We think that the evaluation process validates the design of formula 1 which controls our context determination. Sometimes articles are quite long and only the beginning of the article is about or belongs to directly to the article's image and later other topics are mentioned in the article.

The efficiency for finding locations can be explained by the following example: Sometimes, e.g. when two teams play against each other, the probability of choosing the right location is 50% (when the right context was determined of course). But the articles reports not only about the games, they tell about press conferences and trainings or are interviews and portraits. For some reason the authors take more home pictures and this explains the performance.

Also the object and action annotations performed very well (they were right in the most cases), but unfortunately only 3,8% and less than 8% of the images get object respectively action annotation from the article. Here is the reason that these annotation are not in the article text which surrounds the image (image caption, head of the article). Certainly one reason is that the author has not to describe that for example the player is running and that the ball is on the pitch. The extraordinary things and actions are mentioned. Another reason is that the images do not have an object or an action. So these results are not so convincing.

At a first look the results of the person learning (66,6%) seem to be worse, but there are regarded two things. On the one hand, the context and the person annotation have to be both correct for a useful class attribute. And on the other hand, a failure in person learning does not immediately lead to bad knowledge in the ontology model. The same person must appear a second time in the same context to get a solid connection. Further, after a learning period solid connections may get loose and loose connections are deleted from the ontology.

Table 2. Class Attribute Table of "Werder Bremen"

ID	Class Attribute	Ontology Class	Type	App.
380	diego	werder bremen	person	5
413	bremen	werder bremen	context	6
525	claudio pizarro	werder bremen	person	2
1538	thomas schaaf	werder bremen	person	7
1595	point	werder bremen	context	3
1899	werder bremen	werder bremen	context	9

To get an impression of the whole learning success, Tab. 2 shows all new learned class attributes which have a solid connection to the ontology class "werder bremen". PAGANEL learned persons like the football players "Diego" and "Claudio Pizarro" and the coach of the team "Thomas Schaaf". It obtained new contextual information from the heads of the articles like "werder bremen", "bremen" and "point", too (the attribute "point" is of course to general for this context).

The evaluation procedure makes clear that the person learning and the insertion of new contextual information helps the context determination.

4.3 Image Retrieval Quality

After looking at the concrete results of the annotation process, we took the collected images with the created RDF files for our retrieval experiments because the results shall fuse that our annotation files serve the purpose in practice. Besides, we want to illustrate that the combination of contextual classification and automatic annotations improve image retrieval.

Fig. 7. The GLENARVAN Retrieval System [1]

We make use of the GLENARVAN retrieval system (Fig. **??**) to combine annotations and contextual information. GLENARVAN works with contextual queries q as a tuple which has the form:

$$q = (s, l) \tag{3}$$

where s denotes a query string consisting of a set of keywords, while l defines the context. Two similarity values are computed, the first one based on a lexical similarity and the second one on a contextual similarity. A dictionary and string comparison (e.g. the edit distance) are used for lexical similarity, while contextual similarity calculates the distance of two ontology classes (see [1]). The multiplication of the two values results in the total similarity. GLENARVAN has a result threshold (not shown in Fig. **??**). The result images must have at least 50% of the best picture's similarity value.

For the evaluation 50 queries are formulated which involve different people in their contexts. We keep the annotation type "persons" tight because we want to relate this results to the results of part one.

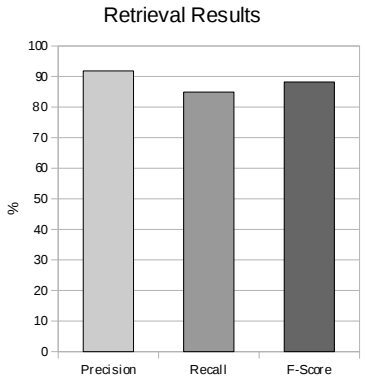


Fig. 8. Retrieval Quality

The results of the retrieval evaluation are summarized in Fig. 8. With 91,84% the precision performs very well and the recall value of 84,91% is also considerable. The F-Score amounts 88,24%.

The evaluation makes clear that the created RDF-files are very useful in a retrieval process especially when the additional contextual information are considered.

5 Conclusion and Future Works

In this paper we have summarized the problems of automatic image annotation and the handling contextual background knowledge. We have presented a new approach which combines automatic annotation and annotations based on ontologies. Besides, we added learning algorithms to obtain new contextual information. Finally, the evaluation illustrates that our system produces advantageous RDF-files which retrieve reliably image data.

The paper shows that the combination of ontology based annotation and learning of new contextual information is a favorable solution for automatic image annotation which can stand in a real retrieval situation.

Prospectively we will work on a way to combine this approach with a large set of external knowledge. We plan to achieve two things:

1. We want to double check the context.
2. We want to develop more possibilities of person, objects and action recognition.

The advantage of a large database would be the co-occurrence of specific expressions which appear again in the same context. By a comparison the determination of the context could be verified. Here links between the separate pieces of knowledge may be a second method. In this way the automatic and conceptual image annotation and contextual learning can be more improved.

References

1. J. Vompras, T. Scholz, and S. Conrad. Extracting contextual information from multiuser systems for improving annotation-based retrieval of image data. In *MIR '08: Proceeding of the 1st ACM international conference on Multimedia information retrieval*, pages 149–155, New York, NY, USA, 2008. ACM.
2. D. Brickley and R. V. Guha. Resource Description Framework (RDF) Schema Specification. World Wide Web Consortium. 2000.
3. T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. In *Scientific American*, page 279, 2001.
4. A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1349–1380, 2000.
5. A. Mallik, P. Pasumarthi, and S. Chaudhury. Multimedia ontology learning for automatic annotation and video browsing. In *MIR '08: Proceeding of the 1st ACM international conference on Multimedia information retrieval*, pages 387–394, New York, NY, USA, 2008. ACM.
6. A. T. G. Schreiber, B. Dubbeldam, J. Wielemaker, and B. Wielinga. Ontology-based photo annotation. *IEEE Intelligent Systems*, 16(3):66–74, 2001.
7. E. Hyvönen and K. Viljanen. Ontogator: combining view- and ontology-based search with semantic browsing. In *In Proceedings of XML*, 2003.
8. J. Liu, M. Li, W.-Y. Ma, Q. Liu, and H. Lu. An adaptive graph model for automatic image annotation. In *MIR '06: Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, pages 61–70, New York, NY, USA, 2006. ACM.
9. T. Osman, D. Thakker, G. Schaefer, and P. Lakin. An integrative semantic framework for image annotation and retrieval. In *WI '07: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, pages 366–373, Washington, DC, USA, 2007. IEEE Computer Society.
10. Y. Ishikawa, R. Subramanya, and C. Faloutsos. MindReader: Querying databases through multiple examples. In *Proceedings of 24th International Conference on Very Large Data Bases, VLDB'98*, pages 218–227, 1998.
11. T. S. Huang, Y. Rui, M. Ortega, and S. Mehrotra. Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 25–36, 1998.
12. Y. Rui, T. S. Huang, and S. Mehrotra. Content-Based Image Retrieval with Relevance Feedback in MARS. In *Proceedings of the 1997 International Conference on Image Processing (ICIP '97)*, pages 815–818, 1997.
13. S. Alcic and S. Conrad. 2-DOM: A 2-dimensional Object Model towards Web Image Annotation In *3rd International Workshop on Semantic Media Adaptation and Personalization (SMAP 08)* December 15-16, 2008. Prague, Czech Republic.
14. OpenNLP Project Website. 2009 `http://opennlp.sourceforge.net`

How MPEG Query Format enables advanced multimedia functionalities

Anna Carreras, Ruben Tous, Jaime Delgado

Distributed Multimedia Applications Group,
Universitat Politècnica de Catalunya,
c/Jordi Girona, 1-3, Mòdul D6, 08034 Barcelona, Spain
{annac, rtous, jaime.delgado}@ac.upc.edu

Abstract: In December 2008, ISO/IEC SC29WG11 (more commonly known as MPEG) published the ISO/IEC 15938-12 standard, i.e. the MPEG Query Format (MPQF), providing a uniform search&retrieval interface for multimedia repositories. While the MPQF's coverage of basic retrieval functionalities is unequivocal, it's suitability for advanced retrieval tasks is still under discussion. This paper analyzes how MPQF addresses four of the most relevant approaches for advanced multimedia retrieval: Query By Example (QBE), Retrieval trough Semantic Indexing, Interactive Retrieval, and Personalized and Adaptive Retrieval. The paper analyzes the contribution of MPQF in narrowing the semantic gap, and the flexibility of the standard. The paper proposes several language extensions to solve the different identified limitations. These extensions are intended to contribute to the forthcoming standardization process of the envisaged MPQF's version 2.

1 Introduction

In today's Multimedia Information Retrieval (MIR) systems, one of the main concerns is how to bridge the semantic gap between the machine-level audio-visual feature descriptors and the semantic-level descriptors directly interpretable by humans. The algorithms currently available in literature are not yet sufficient to assure good results, exploitable in commercial solutions. Of course, the problem arisen by the semantic gap is really difficult to solve since it is intrinsically embedded in the nature of digital contents and strictly related to human interpretation (for example, a picture of a beach at the sunset could be categorized as "sea" or "sunset", according to the mood and the sensitivity of the user). In this paper, we will try to address this issue from two different points of view.

First, from the "content provider side", we will focus on two main retrieval approaches. On the one hand, we will consider QBE, which involves using an example of content to illustrate users' needs (Section 3.1). QBE is one of the most matures approaches for multimedia retrieval and it is based on similarity measures of specific Low Level Features (LLF) that have already been proved to give interesting results [Lux09]. On the other hand, the use of Semantic Indexing will also be addressed (Section 3.2). In this case, links between text-based search terms and semantic extracted descriptors need to be established; although this is a more recent area of research, a lot of work is currently being done on the automatic extraction of these descriptors using complex machine learning and pattern classification techniques.

Second, more related to the subjective perception of the user than to the nature of the digital content, the use of interactive retrieval based on Relative FeedBack (RFB) is the third multimedia retrieval approach that we are going to consider in this paper (Section 3.3). Finally, and because nowadays it is not enough to identify the right content but it is required to be presented in the most suitable way to the user, personalized and adaptive content retrieval will also be addressed in Section 3.4.

In the following section will briefly present the novel MPQF standard [MPQF07] as a possible unified query language. Subsequently, we will identify and separately evaluate its possible application in the four retrieval approaches previously identified: QBE, Semantic Indexing, Interactive Retrieval, and Personalized and Adaptive Content Retrieval. We consider that these four approaches adequately represent today's multimedia scenarios as they cover a broad part of the most relevant work done in this area of research.

Thus, in this paper we will not intend to present a survey on Multimedia Search and Retrieval (interesting works can also be found in [Ha08]), and neither describe the MPQF standard (as in [Dö08]); but analyze its use in some of the most relevant retrieval approaches.

2.1 MPEG Query Format Overview

The MPEG standardization committee (ISO/IEC JTC1/SC29/WG11) has developed a new standard, the MPEG Query Format (MPQF) [MPQF07], which aims to provide a standardized interface to multimedia document repositories. MPQF is an XML-based language which defines the format of queries and replies to be interchanged between parties in a multimedia information search and retrieval environment. MPQF can be used in standalone MMDBs, but it has been specially designed for scenarios in which several MMDBs and content aggregators interact (Fig. 1). Furthermore, MPQF does not make any assumptions about the metadata formats used by the target MMDBs, which can be MPEG-7 but also any other format (Dublin Core for example).

MPQF allows combining Information Retrieval (IR) criteria with Data Retrieval (DR) criteria. Regarding IR-like criteria, MPQF offers a broad range of possibilities that include but are not limited to Query By Example, Query By Feature Range, Query By Spatial or Temporal Relationships, and Query By Relevance Feedback. Regarding the DR criteria, MPQF offers its own XML query algebra, but also offers the possibility to embed XQuery expressions.

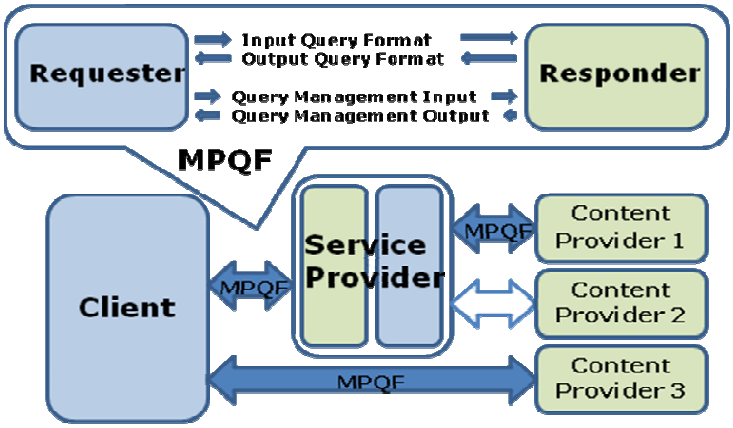


Fig. 1. Possible scenario of use of MPEG Query Format

3 Advanced Multimedia Functionalities with MPQF

3.1 MPQF for Query By Example

The main objective of this section is to report and evaluate the use of MPQF for QBE retrieval. The first thing we need to take into account is that there exist many different ways of implementing QBE algorithms. The detailed analysis of these algorithms is out of scope of this paper, nevertheless, a detailed publication about the QBE algorithms that have been used to raise our conclusions can be found in [Vis07]. In general terms, these algorithms can be based on different types of Low Level Features (LLF) more or less suitable depending on the multimedia content type. For example, when working with entire videos, Temporal LLF and LLF Distributions should be used, while, *ColourStructure* and *HomogeneousTexture* descriptors would be more useful when dealing with image frames. Furthermore, depending on the computation power or even the storage capacity of the system, it would also involve different types of pre-processing techniques along the retrieval process.

In this sense, it is important that service providers are able to query for desired capabilities, and that in turn, content providers are able to communicate their capacities to the service provider. This issue is suitably addressed by the MPQF through Query Management tools which include service discovery, querying service capability, and service capability description (see Query Management Input and Output in Fig. 1).

Moreover, depending on the application scenario, different types of QBE may be required. For example, in a Video Surveillance application scenario it would be interesting to detect similar faces (Query By Region Of Interest), while a user browsing movies similar to his/her favorite ones may require a completely different QBE algorithm (Query By Temporal or Spatial Similarity).

According to ISO-15938-12:2008, the technique of QBE is understood as the combination of different condition expressions such as *QueryByMedia*, *QueryByDescription*, *QueryByROI, SpatialQuery and TemporalQuery*. All these MPQF's condition types are based in the provision of an example (media, media region or media metadata description) in order to express the user information need. These condition types are selected or combined depending on each situation in order to return the best results.

QBE similarity searches are techniques of *content based multimedia retrieval* (CBIR, etc.) which allow expressing the user information need with one or more example digital objects (e.g. an image file). Even though the usage of low-level features description instead of the example object bit stream is also considered QBE, in MPQF these two situations are differentiated, naming *QueryByMedia* (or *QueryByROI*) to the first case (the digital media itself) and *QueryByDescription* the second one. This differentiation is important because in the first case is the query processor who decides which features to extract and use, and in the second case is the requester who perform the feature extraction and selection. In this work we will focus on the first one, as we consider that the *QueryByDescription* is sufficiently well addressed in [Dö08].

The MPQF's *QueryByMedia* type offers multiple possibilities to refer to the example media, as just including the media identifier (a locator such as an URL pointing to an external or internal resource) or directly embedding the image bit stream in Base64 encoding within the XML Query. When the *QueryByMedia* type is used, it's up to the query processor to extract the proper low-level features to perform a similarity search over the index. One of the limitations identified in this work is that MPQF does not standardise a set of parameters or algorithms to be used, leaving this totally open with the consequent lack of interoperability. One possibility could be using MPEG-7 descriptors such as *ScalableColor*, *ColorLayout* or *EdgeHistogram*. The standard should allow expressing different weights to each one of the different descriptors in order to tune the similarity algorithm. Currently these weights are sent to the query processor within non-standard attributes in the MPQF query. The inclusion of non-standard parameters is allowed in MPQF.

Overall, we can conclude that the MPQF offers the necessary tools for performing effective QBE, while maintaining the system network agnostic and media agnostic. Furthermore, it covers all the possible application scenarios we could think of. However, we consider a limitation the fact that MPQF does not standardise a set of parameters or algorithms to be used, leaving this totally open. Currently this information is sent to the query processor within non-standard attributes in the MPQF query, which severely constraints query interoperability.

3.2 MPQF for Retrieval through Semantic indexing

As introduced in Section 2, a MIR system has the particularity that it must combine Information Retrieval (IR) techniques, with techniques for querying metadata, which belong to the Data Retrieval (DR) area within the Databases discipline. Though there is a solid research basis regarding the Information Retrieval challenge, the necessity to face such problem appears, in fact, because of the difficulty of annotating the content with the necessary metadata and the difficulty of formalizing the end-user's semantic-level criteria. As a result, from the multimedia retrieval point-of-view, measures are needed to deal with uncertainty and the potential lack of search precision. However, in a vast number of scenarios, simple IR-like mechanisms like keywords-based search use to offer pretty satisfactory results even when the size of the target collections is big. There are, nevertheless, situations in which the end-user requirements, and/or the circumstances, motivate the efforts of producing higher-level semantic metadata descriptors and formalizing parts of the user's semantic-level criteria moving them to the Data Retrieval realm. An example could be the video surveillance scenario, in which a huge quantity of information is stored, and the query expressiveness and results precision are critical. This formalization task requires enhancing the metadata production layer but also implies offering to the user a richer interface or, in subsequent layers, post-processing the initial non-formalized query. This enrichment of the querying process is related to the improvement of the metadata-level query capabilities. The result is the starting point of what is known as semantic-driven MIR, whose evolution leads to the usage of semantic-specific technologies as those from the Semantic Web initiative.

Current practices in the metadata community show an increasing usage of Semantic Web technologies like RDF and OWL. Some relevant initiatives are choosing the RDF language (e.g. Dublin Core) for modelling semantic metadata because of its advantages with respect to other formalisms. RDF is modular; a subset of RDF triples from an RDF graph can be used separately, keeping a consistent RDF model. So it can be used in presence of partial information, an essential feature in a distributed environment. The union of knowledge is mapped into the union of the corresponding RDF graphs (information can be gathered incrementally from multiple sources).

As introduced in Section 2.1, MPQF is an XML-based language in the sense that all MPQF instances (queries and responses) must be XML documents, i.e. it has an XML serialization format. However, this fact is independent of the target metadata data model. Initially MPQF was designed to only address XML-enabled databases, Formally, MPQF is Part 12 of MPEG-7, which is an XML application, and at the very beginning MPQF was meant to target MPEG-7 repositories. Nevertheless, soon the query format was technically decoupled from MPEG-7 and became metadata-neutral, i.e. MPQF is not coupled to any particular metadata standard. However, the final standard (12/2008) still assumed that queries refer to metadata, at a logical level, as XML trees. The *EvaluationPath* element is probably the most important part of the standard as it identifies the results of the query based on the selected "branch" of this tree. Thus, MPQF expresses conditions and projections over the metadata using XPath expressions, i.e. privileging XML-enabled metadata repositories but restraining those based in other models, especially those based in RDF metadata.

This limitation was already identified in [TD08], and subsequently, and amendment to the MPQF entitled "Semantic Enhancement" [Amd08] was initiated during the 88th MPEG meeting (April 2009), and will be probably finalized during the next meeting (90th MPEG meeting, October 2009). This amendment is the necessary extension to allow the MPQF not only to manage metadata modelled with Semantic Web languages like RDF and OWL, but also to query constructs based on SPARQL.

3.3 MPQF for Interactive Retrieval

When retrieving multimedia content, an important issue that needs to be considered is the subjective perception of the user. Through the use of Relevance Feedback (RFB), the query is refined over stages in which the user indicates which retrieved examples match or do not match the user's need. Based on this feedback, the system modifies its retrieval mechanism in an attempt to return a more desirable instance set to the user.

Once again, depending on the application scenario, the interaction between the user and the system may be different. For example, while a doctor could be very patient to find the most similar medical image within a database in order to make a diagnostic, a user browsing multimedia content in the web would be bothered in the early stages of the interaction.

The MPQF specifies the *QueryByRelevanceFeedback* type which describes a query operation that takes the result of the previous retrieval into consideration. It contains two elements: the *answerID* which identifies the result set where the relevance feedback should be performed; and the *ResultItem* which identifies the good examples that will be used as input for the next query. Althought it is also possible to discard bad results by combining the boolean NOT with the Query By Relevance Feedback operation, we miss the possibility of scoring the results. The MPQF offers the possibility of weighting the query conditions combined within the query "tree", but it would be also interesting to score the different elements of the list of results.

We believe the MPQF is a little bit too simplistic when addressing interactive retrieval as it only allows distinguishing between good and bad results, while it would be much more interesting to know which has or have been the most relevant result/s in order to refine the query. Of course, this should be only an optional attribute suitable for some specific domains or application scenarios.

Nevertheless, we miss an important element we have pointed out earlier in this section: the number of iterations. We believe that the user (or even the Service Provider in some scenarios) should be able to specify the number of iterations she/he is going to perform beforehand, as this would facilitate the application of the most effective matching algorithms.

3.4 MPQF for Personalized and Adaptive Content Retrieval

The main idea on personalized and adaptive retrieval is to use contextual information (from the user or usage environment) in order to provide effective multimedia information retrieval. Of course, this can be considered under the big umbrella of context-awareness area of research. On the one hand, user preferences can help in the identification of retrieved multimedia content, and on the other hand, information about the characteristics and capabilities of the terminal, the network or the natural environment may be used to improve the user's Quality of Experience (QoE) by adapting the content efficiently.

MPQF allows expressing few preferences on the presentation of multimedia content results set. This is done through the *OutputDescription* descriptor included in the Input Query. Nevertheless, it only specifies few listing and sorting options that could easily be extended. We believe personalization is a complex multimedia retrieval service, and as such, it should be considered in the MPQF management tools first. The management part of the MPQF copes with the task of searching for and choosing desired multimedia services for retrieval. This part includes service discovery, querying for service capabilities, and service capability descriptions. We miss the possibility of detecting and selecting a context-aware adaptation service. The MPQF standard can detect services such as authentication, or billing, but does not include context-aware services.

For example, a content provider may offer an integrated service including multimedia contents and the adaptation service. The delivery of most of the contextual information could be done in a separate channel than the query itself as proposed in [ETH09], but probably it would be more useful to integrate the user preferences inside the input query. This could be done by specifying a new query type named "*QueryByUserPreferences*", or even "*QueryByUserContext*" if we think of extending the user preferences with the user historical data for example. Of course, different standards representing contextual information (i.e. MPEG-21, UAProf, etc.) could be used, in the same way MPQF is metadata neutral. Another possibility would be to include this information on the Output Description element.

4 Conclusions

This paper has presented an analysis of the MPQF standard in four relevant areas of research within multimedia search and retrieval applications, namely, query by example, query by semantic indexing, interactive retrieval, and personalized/adaptive retrieval. We can conclude that the first one, QBE, is well addressed, but we consider a limitation the fact that MPQF does not standardise a set of parameters or algorithms to be used. Currently these data are sent to the query processor within non-standard attributes in the MPQF query, which severely constraints query interoperability.

The other three retrieval approaches require further extensions of the standard in order to fully exploit today's application scenarios (the second one is already being addressed trough an Amendment). Probably, the detected limitations are due to the fact that the editors of MPQF have tried to maintain a quite simple standard in order to potentiate its use within the research community. Nevertheless, we believe the specification of MPQF profiles for concrete application scenarios could help to further develop the parts that have been identified as too simplistic, such as personalization, and interactive retrieval.

Finally, it would be very interesting to give the opportunity to the users (or even to the Service Provider) of deciding whether they allow or not the use of advanced retrieval functionalities, such as personalization, or semantic indexing. Usually, these kinds of techniques involve the use of personal information (previous queries, preferences, etc.) that the user may want to protect.

All the identified limitations and proposed solutions are intended to contribute to the forthcoming standardisation process of the envisaged MPQF's version 2. And of course, some evaluation work will be done as soon as a finalised version of a software module based on MPQF exists, which for the moment is not the case.

Acknowledgements

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

References

[Lux09] Lux, M.; Caliph & Emir: MPEG-7 Photo Annotation and Retrieval; Winer of the 2009 ACM Multimedia Open Source Software Competition, October, 2009, Beijing, China. ACM 978-1-60558-608-3/09/10.

[MPQF07] Gruhne, M., Tous, R., Doeller, M., Delgado, J., Kosch, H.; MP7QF: An MPEG-7 Query Format. 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution (AXMEDIS 2007), Barcelona, November 2007. IEEE Computer Society Press. ISBN 0-7695-3030-3. p. 15-18.

[Ha08] Hanjalic, A., Lienhart, R., Ma, W.-Y., Smith, J. R.; The Holy Grail of Multimedia Information Retrieval: So Close or Yet So Far Away?; IEEE Proceedings, Special Issue on Multimedia Information Retrieval, vol.96, no.4, pp. 541-554, 2008.

[Dö08] Döller, M., Tous, R., Gruhne, M., Yoon, K., Sano, M., Burnett, I. S.; The MPEG Query Format: Unifying Access to Multimedia Retrieval Systems; IEEE Multimedia, vol. 15, no.4, pp. 82-95, 2008.

[Vis07] IST-1-038398 - Networked Audiovisual Media Technologies - VISNET II, "Deliverable D2.2.5: First set of developments and evaluation for search systems for distributed and large audiovisual databases". November 2007.

[TD08] Tous, R., Delgado, J.; Semantic-Driven Multimedia Retrival with the MPEG Query Format; in proc. of the Third International Conference on Semantic and Digital Media Technologies (SAMT'08), LNCS 5392, pp.149-163, 2008.

[Amd08] WD 1.0 of ISO/IEC 15938-12:2008 AMD2 Semantic Enhancement.

[ETH09] Eberhard, M., Timmerer, C., Hellwagner, H.; Fully Interoperable Streaming of Media Resources in Heterogeneous Environments. MXM Developer's day 2009.

Context-aware Mobile Multimedia Services in the Cloud

Dejan Kovachev, Ralf Klamma

Chair for Computer Science 5 (Information Systems), RWTH Aachen University
Ahornstr. 55, D-52056, Aachen, Germany
{kovachev, klamma}@dbis.rwth-aachen.de

Abstract. Mobile devices become widely accepted computing paradigms; but the mobile services need to be aware of the dynamical user environment and adapt accordingly to the context. With the increasing amount of multimedia, ontologies can add value to the new semantic multimedia services, by considering the contextual information. Our goal is to provide new concepts for mobile multimedia computing in certain domains like cultural heritage data management. We propose to model the mobile, user and multimedia context with the use of ontologies. We take cloud computing as service infrastructure for supporting complex semantic multimedia tasks for the mobile clients.

1 Introduction

With the massive production of multimedia content nowadays, the usefulness of this content depends largely on the creation, sharing, reuse, discovery, access and delivery of the multimedia. Obviously, in many multimedia applications a semantic approach for knowledge representation and processing for the complete multimedia life cycle is needed. Using ontologies for domain knowledge representation can be identified as a promising tool that supports formal, explicit, machine-processable semantics definition and further knowledge discovery. Personalization brings benefits for the user by matching his stated and learned preferences, thus matching more the user's wishes and needs.

Most important concept of mobile computing is the "anytime, anywhere" computing by decoupling users from smart, intelligent device and viewing applications as entities that perform tasks instead of the user [10]. Using the contextual information in the multimedia value chain brings the possibility to provide value-added services or to execute more and complex tasks. Context-awareness takes an important role in the pervasive computing. Mobile phones that contain the basic building blocks for context awareness such as physical sensors, GPS, compass, accelerometers, light sensors and Internet access are seeing explosive adoption. On the other hand, the diversity of ways that the user context can be used by different services or context consumers is growing fast. This is due to the increasing number of service delivery or provider entities that can be accessed by the user [4]. Mobile phones enable new, rich user experiences, but their hardware

is still very limited in terms of computation, memory, and energy reserves, thus limiting potential applications [6]. But their limitations can be exceeded by off-loading the execution of the hardware-intensive computations into the cloud. A recent study released by ABI Research [1] says that limited processing power, battery life, and data storage will limit mobile application growth in the mass market, even among smart phones like Apple's iPhone. But, applications that connect to cloud resources are much more likely to be successful than those that run only on the mobile device.

The problems are the following. New mobile phones provide a lot of contextual information, but this is not completly explotied for enriching the multimedia services on mobile platforms. The media context needs to be matched with the user's context. Other issues are exploiting the contextual information for adaptation for different devices and interoperability with the existing resources on the web.

In this paper we explore the possibilities for context-aware services for semantic multimedia targeted towards mobile devices in application domains like cultural heritage documentation. We consider also taking community context and other context information into consideration [5].

2 Background

Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between the user and the application, including the user and the applications itself [7]. An application or system is *context-aware* if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task. Examples of context information in mobile applications can be seen, e.g., spatial information - location, orientation, speed; and acceleration, temporal information - time of the day, date and season of the year; environmental information; social information - who you are with, and people that are nearby; resources that are nearby - accessible devices, and hosts; availability of resources - battery, display, network, and bandwidth; physiological measurements - blood pressure, hart rate, respiration rate, muscle activity, and tone of voice; activity - talking, reading, walking, and running; schedules and agendas. Many prototypes of context-aware applications are done [7, 15].

Context-aware systems use context models expressed as ontologies, in order to formalize and limit the notion of context and that relevant information differs from a domain to another and depends on the effective use of this information. The Web Ontology Language (OWL) is used to explicitly formalize the properties and structure of contextual information to guarantee common semantic understanding among different architectural components. OWL has well-defined syntax, formal semantics, reasoning support, and enhances information retrieval and interoperability. Ontologies also can well model the semantics of multimedia [8, 3].

A cloud-computing-based infrastructure for context-aware semantic multimedia services is promising [12]. One major benefit is to enhance interoperability between heterogeneous context-sources and applications. Other benefit are scalable resources on demand. There is also the need to manage massive amounts of diverse user-created data, synthesize it intelligently, and provide it as real-time services. Essential features of such visions are comprehensive context awareness, personalized user interfaces, and multimedia content adaptation. Ontology processing requires a lot of computing resources, especially for ontology reasoning that performs poorly according to the size of the ontology. Chun et al. [6] propose off-loading mobile applications in the cloud for resource demanding computations.

3 Application Areas of Context-aware Mobile Services

Possible applications areas of context-aware mobile applications are tourist guides, mobile advertisement, context-aware proactive news service, cultural heritage, technology enhanced learning and many others [15].

The *tourism* domain is widely considered to be one of the emerging industrial sectors where mobile services are highly demanded. In fact, in 2015 there will be more than 3 billion travelers around the globe and they will demand more ubiquitous services, specific to the situation of each individual, as well as to their personal preferences in specific circumstances. Surveys reveal that over 90% of travelers carry a mobile device with them. Time is a very scarce resource and connectivity to all kinds of services in mobility is highly demanded and required [4].

In the *cultural heritage* domain, a human expert relies on a number of properties to annotate artifacts at the capturing stage. The expert knowledge used in the process of archaeological investigation is then embedded in and integrated with the multimedia including the context information during the archaeological site documentation. This is an example of concurrent engineering where semantic multimedia can play an important role. Effective concurrent engineering systems should be based on knowledge management and sharing mechanisms and standards that are able to provide comprehensive formalization and reasoning infrastructure that supports the design and productions processes [13]. There are many attributes and properties of multimedia that scientists and professionals are using to exchange, process, and share content, and all these have to be classified and formalized thoroughly. Great value of the semantic multimedia is carried out also by the creation and annotation process, the intermediate steps that contributed to the definition of the final product and experts' knowledge used at the various stages of its development.

4 Analysis of Context-aware Mobile Multimedia Services

Conceptualization and realization of mobile context-aware multimedia systems face several design challenges that afford to cope with highly dynamic environments and changing user requirements.

Context-awareness can only be researched in relation to certain application domains or communities. A generic context management approach will not be manageable because of the inherent complexity of the context models as well as the sheer amount of context information. The problems here are related to gathering, modeling, storing, distributing and monitoring context.

We intent to create a set of web services that will allow devices to interface with applications anywhere in the cloud of accessible data sources, services and applications. A level of domain-specific and community-specific middleware glues all parts together, joining data from the sensors and applications with user input, storing contextual information, and allowing the mobile device to share that data across applications or even between different devices [16].

4.1 Architectural Design

The architecture should provide the foundations for the different entities to deal with context (how to discover it, how to store it, how to access it and how to take advantage of the information it provides) in a mobile environment.

Using the service-oriented computing paradigm will broaden the variety of accessible applications for mobile environments. Tim O'Reilly [11] believes that "the future belongs to services that respond in real time to information provided either by their users or by nonhuman sensors." Such services will be attracted to the cloud not only because they must be highly available, but also because these services generally rely on large data sets that are most conveniently hosted in large datacenters. This is especially the case for services that combine two or more data sources or other services, e.g., mashups. While not all mobile devices enjoy connectivity to the cloud 100% of the time, the challenge of disconnected operation has been addressed successfully in specific application domains, so this is not a significant obstacle to the appeal of mobile applications [2].

4.2 Data Management

An overview of the data organization for a cultural heritage documentation use case is given in Figure 1,. The system needs to cope with heterogeneous data from many sources, integrate contextual information from different sensors, cameras, 3D scanners and user input. The cloud of services provides interface to many applications and avoids cross-platform problems while easing the data sharing. The content delivery is performed using adaptation services while taking the context of the user and multimedia in consideration.

4.3 Context Modeling

For systems that provide context-aware mobile multimedia services we need to use context models in order to formalize and limit the notion of context and the relevant information from a domain. Ontology-based models propose a semantic modeling of context information, enhanced by appropriate reasoning mechanisms.

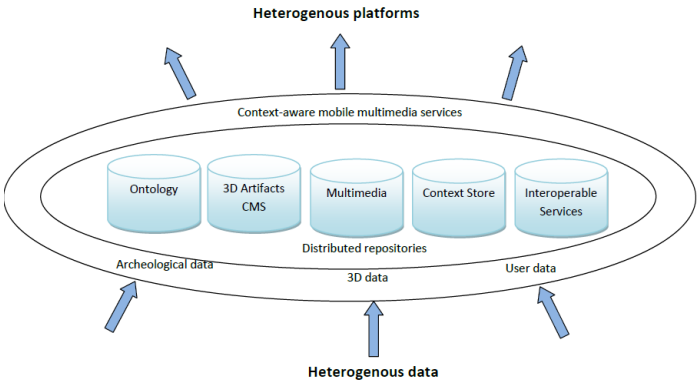


Fig. 1. Data organization in cultural heritage data management scenario

Ontologies are the basis and foundation for new intelligent multimedia applications, but we need further tools in order to make these applications a reality and to create commercially viable new businesses around these applications [9]. Reasoning enables formalization of the media interpretation process. The requirements that the model needs to fulfill are:

- **Appropriate multimedia and context ontology.** Higher-level context can be composed by low-level context. But dependencies of context items cause difficulties, which need to be resolved.
- **Exchange context information with other models.** Sharing conceptualized knowledge between different systems is the underlying idea of the Semantic Web.
- **Reuse existing related ontologies:** Ontonym, YAGO, DBpedia, CIDOC Conceptual Reference Model (CRM), aceMedia, Delivery Context Ontology, COMM and etc. Ontology-based models propose a semantic modeling of context information, enhanced by appropriate reasoning mechanisms [4].
- **Bridge the gap between social and Semantic Web.** Gather information about users from their Social Web identities and enrich with ontological knowledge.

Ontologies offer new possibilities regarding knowledge management, retrieval effectiveness, and online collaboration compared to conventional technologies and techniques. Development of ontologies for multimedia applications are needed by taking care of defining a comprehensive schema for documenting and sharing the media repositories, to be linked and further specialized by experts in different domains. Context ontologies should be designed in a two-level hierarchy. We divide a pervasive computing domain into several sub-domains, and define individual low-level ontology in each domain. We also define a generalized ontology which

describes the general concepts in upper level to link up all the low-level context ontologies. Domain-specific ontologies can be dynamically "bounded" or "re-bounded" with the upper ontology when the domain is changed [14].

4.4 Other Challenges

The other aspects that need to be addressed in the construction of context-aware mobile multimedia services are:

- **Privacy.** Sending the current location information into the cloud could lead to difficulties in establishing trust. The system need to be capable of preserving the user's privacy.
- **Sensing.** A big challenge is to sense context changes and establishing relations between context entities.
- **Context processing and classification.** Deducing information form context can be done in several ways, where the most common are semantic reasoning, interpretation of context, and aggregation of context.

5 Conclusions and Future Work

In this paper, we presented an approach to context-aware mobile services with focus on semantic multimedia. The semantic multimedia can create beneficial opportunities for new mobile applications, since these add value to multimedia assets. Ontologies expressed in OWL can be used for modeling the user and media context. We point the expected benefits of the use of multimedia semantics and describe two applications areas, i.e. cultural heritage documentation or tourist guides. We identify several challenges in the system construction.

In the future, we aim to develop extendable service-oriented infrastructure following the cloud computing paradigm that will provide services for mobile semantic multimedia. The service architecture needs to cope with all the context-awareness issues required for the domain. Prototype will be implemented in cultural heritage domain that will prove the expectations of the described approach.

Acknowledgments. This research work has been supported by the Research School within Bonn-Aachen International Center for Information Technology (B-IT).

References

1. Abi Research. Mobile cloud computing. http://www.abiresearch.com/research/1003385, 2009. Last accessed on 21.09.2009.
2. M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. Above the Clouds: A Berkeley View of Cloud Computing. Technical report, EECS Department, University of California, Berkeley, February 2009.
3. R. Arndt, R. Troncy, S. Staab, L. Hardman, and M. Vacura. *COMM: Designing a Well-Founded Multimedia Ontology for the Web*, volume 4825 of *Lecture Notes in Computer Science*, chapter 3, pages 30–43. Springer Berlin Heidelberg, Berlin, 2008.
4. A. Cadenas, C. Ruiz, I. Larizgoitia, R. García-Castro, C. Lamsfus, I. n. Vázquez, M. González, D. Martín, and M. Poveda. Context management in mobile environments: a semantic approach. In *Proceedings of the 1st Workshop on Context, Information and Ontologies*. ACM, 2009.
5. Y. Cao, R. Klamma, M. Hou, and M. Jarke. Follow Me, Follow You - Spatiotemporal Community Context Modeling and Adaptation for Mobile Information Systems. In *The Ninth International Conference on Mobile Data Management (MDM 2008)*, pages 108–115. IEEE, April 2008.
6. B. Chun and P. Maniatis. Augmented Smartphone Applications Through Clone Cloud Execution. In *12th Workshop on Hot Topics in Operating Systems*. USENIX, 2009.
7. A. K. Dey. Understanding and Using Context. *Personal and Ubiquitous Computing*, 5:4–7, 2001.
8. R. Garcia and O. Celma. Semantic Integration and Retrieval of Multimedia Metadata. *2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media*, 2005.
9. Y. Kompatsiaris and P. Hobson. *Semantic Multimedia and Ontologies: Theory and Applications*. Springer London, London, 1 edition, 2008.
10. K. Kwang-Eun and S. Kwee-Bo. Development of context aware system based on Bayesian network driven context reasoning method and ontology context modeling. In *International Conference on Control, Automation and Systems*, pages 2309–2313. IEEE, 2008.
11. T. O'Reilly. What Is Web 2.0. http://oreilly.com/web2/archive/what-is-web-20.html, 2005. Last accessed on 26.10.2009.
12. S. Schenk, C. Saatho, and A. Scherp. SemaPlorer-Interactive Semantic Exploration of Data and Media based on a Federated Cloud Infrastructure. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7, 2009.
13. M. Spagnuolo and B. Falcidieno. *Semantic Multimedia and Ontologies*, pages 185 – 205. Springer London, London, 2008.
14. G. Tao, P. Hung Keng, and Z. Da Qing. A middleware for building context-aware mobile services. In *59th Vehicular Technology Conference*, volume 5, pages 2656–2660. IEEE, 2004.
15. D. Weiss, M. Duchon, F. Fuchs, and C. L. Popien. Context-aware personalization for mobile multimedia services. In *6th International Conference on Advances in Mobile Computing and Multimedia*, pages 267–271. ACM, 2008.
16. A. Wright. Get smart. *Communications of the ACM*, 52:1, 2009.

Global vs. Local Feature in Video Summarization: Experimental Results.

Marian Kogler, Manfred del Fabro, Mathias Lux, Klaus Schöffmann, and Laszlo Böszörmenyi

Institute for Information Technology
Klagenfurt University
Universitätsstrasse 65–67
9020 Klagenfurt, Austria
{mkogler, manfred, mlux, ks, laszlo}@itec.uni-klu.ac.at

Abstract. We investigate the usefulness of local features in generating static video summaries. The proposed approach is based on bag of visual words using SIFT features. In an explorative experiment we compare this approach to summaries generated with the help of global features. As a resume we conclude that the local feature based approach does not outperform the other ones, however, it seems to be more stable.

1 Introduction

In the last decade the importance of videos conveying information has increased, which is accompanied by the need to store, organize and index the multimedia content appropriately, in order to support the user in retrieving videos. A lot of video clips are produced, broadcasted, shared and stored every day by professionals, amateurs and hobbyists. Finding videos matching the actual information need of a user proves to be a hard problem. *Video abstracts*, or video summaries, aim at presenting the semantics and content of a clip in minimized time and space to allow fast assessment of video clip relevance. In this paper we focus on static methods: still image summaries showing keyframes from the video.

Generally speaking a video abstract should maximize the (semantic) information transported by the summary while minimizing time and pixels needed to show, store and assess the summary. We have created a keyframe selection tool (discussed in detail in [13]), which implements summarization of video clips by keyframe extraction based on global image features.

We further extended the tool to support extraction based on local features in order to find out, whether the summarization process leads to a better summarization of video clips. We apply SIFT features proposed by Lowe [12] to extract feature vectors from salient keypoints of an image. The salient points and their 128 dimensional feature vectors are interpreted as local features describing a video frame. For pairwise comparison of frames we employ the *bag of visual words* approach (see e.g. [5], [10], [20], [18], [16]). All local features are clustered using K-Means [9]. The cluster centers are interpreted as reference feature for the

whole cluster and are called *visual word*. A single frame is then represented by a histogram, called *local feature histogram* [6], denoting the occurrence of visual words within the frame. Figure 1 depicts the described approach, in order to get a better understanding. For keyframe selection the local feature histograms are k-medoid clustered [7] and cluster medoids are selected as representative keyframes of a frame cluster. Cluster medoids are ranked based on the cluster size.

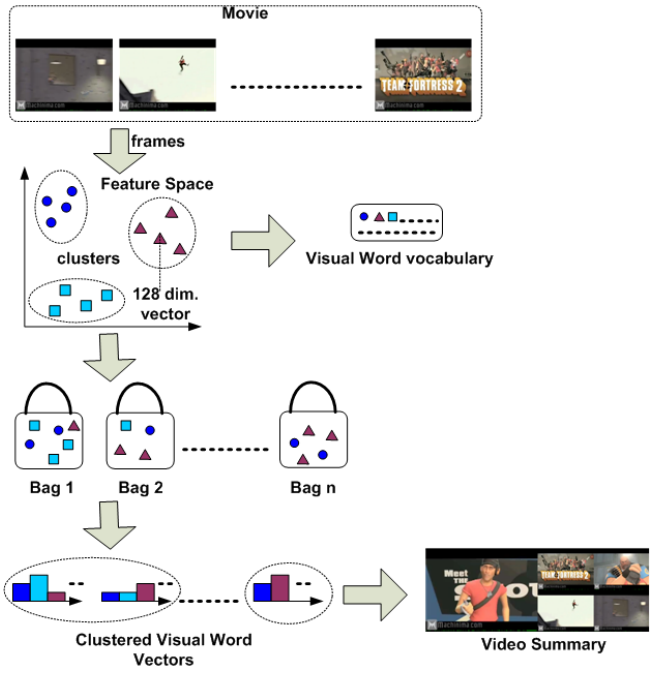


Fig. 1. Bag of Visual Words for Video Summarization

We apply the mentioned bag of visual words technique in our video clip summarization, in order to achieve a video summary more suitable for the user to assess the videos relevance. This should facilitate the search process of the user in a huge multimedia database, by depicting more meaningful images in a video summary.

The remainder of this paper is composed as follows: in Section two we give a short overview of video summary techniques. Section three covers our exploratory study which was conducted in order to test the performance of global vs. local features. Section four concludes our paper, discusses contributions and lessons learned and gives an outlook on future work.

2 Related Work

The main idea behind video summarization is to take the most representative and most interesting segments of a long video in order to concatenate it to a new, smaller, sequence. Video summarization has been investigated for several years already. Nevertheless, in a recent review [17] – that contains more than 160 references! – the authors conclude that "video abstraction is still largely in the research phase".

Proposed methods can be classified by the low-level features which are used for content analysis. In general, video summarization is either performed by image features (e.g. [2]), audio features (e.g. [19]), textual features (e.g. [4]), or a combination of several features (*multi-modal* methods, e.g. [14]).

A further classification can be performed on the presentation of the summary. *Static methods* use representative keyframes (e.g. in a storyboard visualization). *Dynamic methods* use video skims (e.g. a slide-show of keyframes). The static method has the advantage that a user can more quickly watch the entire summary, while the dynamic approach may allow a more comprehendible summary not only because usually audio playback is also available. In addition, *interactive video summarization* methods allow a user to selectively see parts of the summary according to a query.

Another classfication has been presented by Money et al.[15]. They classified video summarization methods into *internal*, *external*, and *hybrid* ones. The most common ones are internal methods, where content analysis is performed directly on the video stream. External methods (e.g. [21]) use information not directly contained in the video stream (e.g. manual annotation), and hybrid methods use a combination of both.

Recent efforts try to create personalized video summaries by integrating the users' interest. For instance, Matos et al.[14] use multimodal analysis together with a model of the *users' arousal*. Lie et al. [11] propose another personalized video summarization system. Their system allows a user to formulate preferences on semantic events like the appearance of humans, the happening of specific events (explosions, moving objects, zoom-in), and the differentiation between indoor and outdoor scenes. Another interesting approach has been presented by Bailer et al. [1]. They propose a collaborative summarization method, where several methods for content segmentation and segment selection are combined and finally fused together in order to produce the video summary.

3 User Study

We conducted an exploratory evaluation, where users had to choose their favorite summaries depicting the corresponding videos in best manner. We distinguished between four low level features (ACC [8], CEDD [3], RGB color histograms, SIFT), which led to four different summaries for each video clip. In a previous study [13] we investigated a number of global features. Summaries generated on the basis of the ACC, CEDD and RGB features were favored by the users and therefore selected in the actual study to compete with local SIFT features.

One summary consisted of five keyframes extracted by our tool. These keyframes were arranged in a single summary image which was presented to the user. We analyzed four short video clips ranging from news to animations. Because videos longer than five minutes probably cover too much information, which cannot be depicted properly in a video summary consisting of five still images, we only considered short ones. A further reason for selecting short clips is, that video clips recorded by users, in order to retain a moment of attraction, usually do not take longer than three minutes. This assumption is based on the observation that the average length of a video clip posted on YouTube is only 2 minutes and 46.17 seconds[1].

Table 1. videos used for exploratory study

Title	**Length**
iPhone commercial[2]	76 s
dinosaurs vault[3]	48 s
hurricane IKE - news reporter almost washed away[4]	30 s
shrek[5]	48 s

Each video is summarized by a full sized frame of the biggest cluster (the cluster with most frames) on the left, followed by four frames half in width and height on the right representing smaller clusters. Figure 2 shows a sample visualization of a video summary created by our tool for the Shrek video.

Fig. 2. Visualization of our video summary (based on CEDD) depicting a video clip

Each participating user had to assess four summaries (4 points for the best, down to 1 point for the worst) for each video clip, which led to a total of 16

[1] Statistics from http://ksudigg.wetpaint.com/page/YouTube+Statistics
[2] http://www.youtube.com/watch?v=2k3zvI2tyPM
[3] http://www.youtube.com/watch?v=Dim0INyvJdw
[4] http://www.youtube.com/watch?v=SYl9mgFhe2o
[5] http://www.youtube.com/watch?v=uvyelwDA0Ws
(last checked: 2009-09-22)

summaries. The user group consisted of 9 people (5 female and 4 male); ages ranging from 20 to 30 years.

3.1 Results

There was no clear winner in our experiment. All four selected image features got similar ratings from the test persons as Figure 3 shows. The summaries based on CEDD have reached the highest score (104 points), followed by our SIFT based visual bag of words approach (91 points), ACC (87 points) and the color histogram (78 points). The scores for each single video are shown in Table 2.

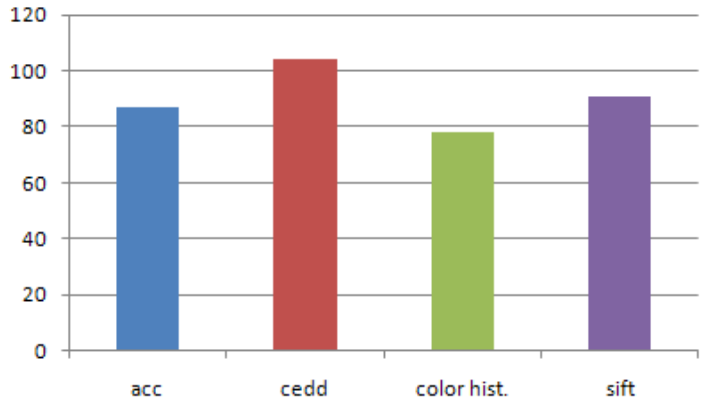


Fig. 3. Summed user ratings of the low-level features used for keyframe selection

Table 2. Rating of the features for each video

	ACC	CEDD	color histogram	SIFT
iPhone	29	23	17	21
News	20	31	16	23
Shrek	11	33	25	21
Dinosaur	27	17	20	26

It can be seen that the type of the chosen features heavily depends on the type of the video. While CEDD produces very good results for the news video and the clip of the movie Shrek, it performs rather poor for the animation with the dinosaurs. On the other side, ACC reaches a high score for the iPhone commercial and the dinosaurs animation, but it is a bad choice for the Shrek clip. Our SIFT-based bag of visual words approach never reached the best score, but also never

performed worst, which can be seen easily in Figure 4. In three cases (iPhone, news and dinosaurs) it reached the second highest score and in the fourth case (Shrek) it reached the third place. Therefore, it seems that this approach based on local image features produces more stable results than the ones based on global image features. This assumption is also supported by the deviation of the samples, given in Table 3. The local feature approach in our experiments features lowest standard deviation (SIFT, 0.84) and can be considered the most stable approach for our test set.

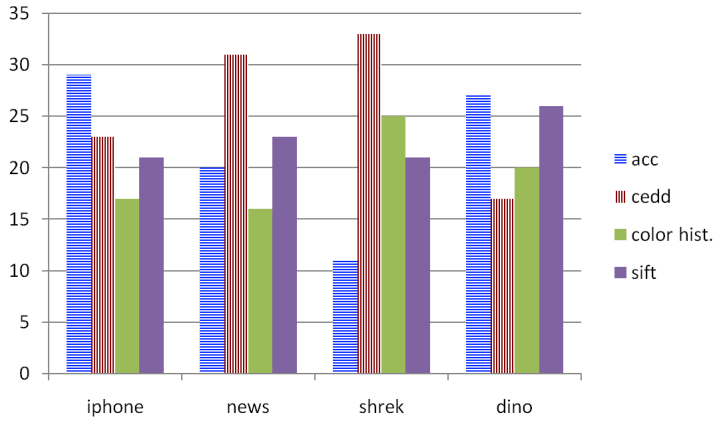


Fig. 4. User ratings of the low-level features used for keyframe selection per video

Table 3. Standard deviation for the ratings of the selected visual features

Feature	Standard deviation
ACC	1.18
CEDD	1.14
color histogram	1.21
SIFT	0.84

4 Conclusion

We presented the results of a study, where users weighted the appropriateness of video summaries based on their ability to describe a short video clip. Main focus of our investigation was the question whether local features achieve better summaries than global features. We employed a visual bag of words approach using

the SIFT descriptor and tested the approach on 4 videos with an exploratory study with 9 users. Results showed that while the local feature approach could not outperform the global feature approaches, it provides for our test data set and the test population the most stable results being ranked second three times and third one time. Even though the test data set and the population of the survey are too small to provide significant results, they allow the hypothesis that local features provide more stability than global features in general use cases and encourage further research on this.

With our implementation we could also see the difference in runtime between the different approaches. Extraction of SIFT features and finding of the visual words took ten times longer than the extraction of global features, say CEDD (70 vs. 700 ms per frame on a 2 GHz dual CPU workstation). While this can be further reduced using faster and optimized local features the whole process of extraction, clustering of the salient points and creation of the visual word vocabulary is significantly slower than a global feature based approach.

In the near future we will test the local feature approach in different domains including medical videos and user captured single shot videos. We hope that we gain insights on the applicability of local feature histograms and the overall performance in and throughout different domains.

5 Acknowledgments

This work was supported by the Lakeside Labs GmbH, Klagenfurt, Austria and funding from the European Regional Development Fund and the Carinthian Economic Promotion Fund (KWF) under grant 20214/17097/24774.

References

1. W. Bailer, E. Dumont, S. Essid, and B. Merialdo, *A collaborative approach to automatic rushes video summarization*, 15th IEEE International Conference on Image Processing, 2008. ICIP 2008, 2008, pp. 29–32.
2. D. Borth, A. Ulges, C. Schulze, and T.M. Breuel, *Keyframe Extraction for Video Tagging and Summarization*, Proc. Informatiktage, 2008, pp. 45–48.
3. Savvas A. Chatzichristofis, Yiannis S. Boutalis, and Mathias Lux, *Selection of the proper compact composite descriptor for improving content based image retrieval*, The Sixth IASTED International Conference on Signal Processing, Pattern Recognition and Applications SPPRA 2009, 2009.
4. G. Ciocca and R. Schettini, *An innovative algorithm for key frame extraction in video summarization*, Journal of Real-Time Image Processing **1** (2006), no. 1, 69–88.
5. T. Deselaers, L. Pimenidis, and H. Ney, *Bag-of-visual-words models for adult image classification and filtering*, Proc. 19th International Conference on Pattern Recognition ICPR 2008, December 8–11, 2008, pp. 1–4.
6. Thomas Deselaers, Daniel Keysers, and Hermann Ney, *Features for image retrieval: an experimental comparison*, Inf. Retr. **11** (2008), no. 2, 77–107.

7. Youssef Hadi, Fedwa Essannouni, and Rachid Oulad Haj Thami, *Video summarization by k-medoid clustering*, SAC '06: Proceedings of the 2006 ACM symposium on Applied computing (New York, NY, USA), ACM, 2006, pp. 1400–1401.
8. Jing Huang, S. Ravi Kumar, Mandar Mitra, Wei-Jing Zhu, and Ramin Zabih, *Image indexing using color correlograms*, CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97) (Washington, DC, USA), IEEE Computer Society, 1997, p. 762.
9. A. K. Jain, M. N. Murty, and P. J. Flynn, *Data clustering: a review*, ACM Comput. Surv. **31** (1999), no. 3, 264–323.
10. Yu-Gang Jiang and Chong-Wah Ngo, *Bag-of-visual-words expansion using visual relatedness for video indexing*, SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval (New York, NY, USA), ACM, 2008, pp. 769–770.
11. W.N. Lie and K.C. Hsu, *Video summarization based on semantic feature analysis and user preference*, Proc. IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing 2008, 2008, pp. 486–491.
12. David G. Lowe, *Object recognition from local scale-invariant features*, ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2 (Washington, DC, USA), IEEE Computer Society, 1999, p. 1150.
13. Lux, M.; Schöffmann, K.; Marques, O.; Böszörmenyi, L., *A novel tool for quick video summarization using keyframe extraction techniques*, Proceedings of 9th Workshop on Multimedia Metadata(WMM'09), CEUR Workship Proceedings, vol. 441, march 19–20 2009.
14. N. Matos and F. Pereira, *Using MPEG-7 for Generic Audiovisual Content Automatic Summarization*, Image Analysis for Multimedia Interactive Services, 2008. WIAMIS'08. Ninth International Workshop on, 2008, pp. 41–45.
15. A.G. Money and H. Agius, *Video summarisation: A conceptual framework and survey of the state of the art*, Journal of Visual Communication and Image Representation **19** (2008), no. 2, 121–143.
16. C.G.M. Snoek, K.E.A. van de Sande, O. de Rooij, B. Huurnink, J.C. van Gemert, J.R.R. Uijlings, J. He, X. Li, I. Everts, V. Nedovic, M. van Liempt, R. van Balen, F. Yan, M.A. Tahir, K. Mikolajczyk, J. Kittler, M. de Rijke, J.M. Geusebroek, T. Gevers, M. Worring, A.W.M. Smeulders, and D.C. Koelma, *The mediamill trecvid 2008 semantic video search engine*, (2009).
17. B.T. Truong and S. Venkatesh, *Video abstraction: A systematic review and classification*, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP) **3** (2007), no. 1.
18. J. R. R. Uijlings, A. W. M. Smeulders, and R. J. H. Scha, *Real-time bag of words, approximately*, ACM International Conference on Image and Video Retrieval, 2009.
19. M. Xu, NC Maddage, C. Xu, M. Kankanhalli, and Q. Tian, *Creating audio keywords for event detection in soccer video*, Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on, vol. 2, 2003.
20. Jun Yang, Yu-Gang Jiang, Alexander G. Hauptmann, and Chong-Wah Ngo, *Evaluating bag-of-visual-words representations in scene classification*, MIR '07: Proceedings of the international workshop on Workshop on multimedia information retrieval (New York, NY, USA), ACM, 2007, pp. 197–206.
21. D. Zhang and S.F. Chang, *Event detection in baseball video using superimposed caption recognition*, Proceedings of the tenth ACM international conference on Multimedia, ACM New York, NY, USA, 2002, pp. 315–318.

Metadata for Creation and Distribution of Multi-view Video Content

Werner Bailer and Martin Höffernig

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

Abstract. Recently the production of multi-view video content has attracted growing attention. The main driving force is stereoscopic cinema, but also 3D television is an upcoming application. In this paper we review the metadata requirements for multi-view video content and analyze how well these requirements are covered in existing metadata standards, both in terms of the coverage of metadata elements and the capabilities to structurally describe multi-view video content. The SMPTE metadata standards, MPEG-7 and EBU P_Meta are considered in this survey. We outline the issues that need to be addressed in future standardization activities.

1 Introduction

Media production and distribution workflows are increasingly shifting from a linear chain to flexible and dynamic processes. This is fostered by advanced tools for media creation and manipulation that blur the boundary between production and post-production and by the fact that productions are today often made for a broad range of target media and distribution channels. In addition, production workflows become increasingly distributed, involving many contributors located at different sites. Thus automation of workflows and metadata interoperability between different workflow steps is of growing importance. Previous work has analyzed the metadata needed in the audiovisual media production process (e.g. [1, 2]) and workflow automation based on workflow languages has been proposed, e.g. for movie production in the YAWL4Film[1] project [3].

Recently the production of multi-view video content has attracted growing attention. The main driving force is stereoscopic cinema, but also 3D television is an upcoming application. Multi-view production further increases the amount of material to be handled in the production process. Next to different language, subtitle, age, etc. versions, 3D adds another degree of freedom to the versions that need to be packaged and distributed. As with many emerging technologies, there are competing systems for stereoscopic exhibition that need to be supported, and of course 2D versions still need to be provided for the majority of theaters,

[1] http://www.yawl4film.com/

television and DVD viewers. Thus there is need for better asset management in distribution to support automatic packaging of the variety of versions.

In this paper we review the metadata requirements for multi-view video content and analyze how well these requirements are covered in existing metadata standards. Section 2 discusses the metadata that needs to be represented for multi-view video content. In Section 3 we analyze different relevant metadata standards w.r.t. these requirements. Section 4 summarizes the analysis and presents an outlook on possible future standardization activities.

2 Metadata Requirements

Various types of metadata exist throughout the digital cinema production workflow. These metadata are produced and consumed at different stages of the workflow. Typically the different devices and tools used in the chain also make use of different metadata representations. In some cases the same metadata properties are stored several times in different formats. Multi-source content adds additional requirements to the metadata representation, as the relations between different media elements need to be described (from the high-level fact that these are different views of the same scene down to precise measurements such as camera calibration parameters). We consider a wide range of visual, audio and several classes of descriptive metadata elements that are produced or used in the different stages of the 3D cinema production workflow. Our discussion does not include data derived from the essence that is in its structure similar to audiovisual essence, such as proxies, key frames, depth maps, maps of the scene geometry etc. Such data can be referenced from the description using relational descriptive metadata. The different properties can be related to three different granularities of the content: to the *production*, i.e. the entire set audiovisual content related to one movie production, to the *asset*, i.e. a single piece of audiovisual essence and to a *segment*, i.e. a (spatio)temporal part of audiovisual essence.

2.1 Technical Metadata

A wide range of technical metadata for video and audio is captured or created during the production process, mainly describing the sampling properties of the audiovisual essence and parameters of devices (e.g. cameras) and tools (e.g. encoders) used in the process. For multi-view video content the parameters describing the geometry of the scene and the recording process are of crucial importance. These include camera position and orientation, absolute lengths in scene needed for metric reconstruction and intrinsic camera parameters. As lenses introduce a number of distortions, precise parameters of the lens distortion model are also required. Another important kind of technical metadata is synchronization information between the different audiovisual streams.

2.2 Descriptive Metadata

The following types of descriptive metadata are created and used in the production and distribution process.

Identification Identification information contains IDs as well as the titles related to the content (working titles, titles used for publishing, etc.).

General content properties These are general description metadata items, not related to a specific modality, such as file size, checksums etc.

Production This describes metadata related to the creation of the content, such as location and time of capture as well as the persons and organizations contributing to the production.

Rights Basic rights information and references to more detailed description of rights and licenses.

Publication/distribution This describes metadata related to the creation of the content, such as location and time of capture as well as the persons and organizations contributing to the production.

Process-related Describes steps in the production workflow (e.g. applied tools, settings). Some processing steps may only apply to certain views (or use different parameters for each of the views), e.g. when performing color correction to adjust one view to another.

Content-related Content-related metadata is descriptive metadata in the narrowest sense. An important part is the description of the structure of the content (e.g. shots, scenes).

Relational/enrichment information Describes links between the content and external data sources, such as other multimedia content or related textual sources. For multi-view video content relational information is needed to link related views, calibration sequences for certain views and other captured data, such as e.g. depth maps.

Most of them are not specific to multi-view video content. However, some of these properties apply to all views, while others might differ. For example, the annotation might describe the objects present in the scene. In a certain setup, a background object could be placed in a corner of the scene so that it is not visible in one of the cameras.

3 Support in Standards

The following standards have been identified to be relevant in different stages of the digital cinema production process and have thus been considered in this study:

- SMPTE Metadata Dictionary [4],
- MXF Descriptive Metadata Scheme 1 [5],
- MPEG-7 Multimedia Content Description Interface [6], and
- EBU P_Meta metadata exchange format [7].

In the following, we analyze both the structural support for multi-view video content in these standards as well as the coverage of the metadata elements discussed in Section 2.

3.1 Structural Support for Multi-view Video Content

We have analyzed whether these metadata standards provide structural support for representing multi-view video content, i.e. allow to describe a set of audiovisual streams that capture the same scene from different positions in space and need to be synchronized, but may have different start times and durations, i.e. temporal offsets.

In most standards there is no explicit concept for representing different views of a scene, especially if they do not have the same temporal extent. Due to the longer tradition of multi-channel audio the support for it is typically much better. While it is in most standards possible to find a representation for multi-view video content, such a representation typically involves application defined semantics and several options might exist.

SMPTE MXF and DMS-1. The MXF container specification [8] provides means to represent several streams of the same modality. The MXF Generic Container [9] can have up to 127 visual or audio data items. However, the semantics of multi-view video content cannot be clearly represented. Depending on the semantics to be expressed two approaches can be chosen:

- Content play-list or edit item pattern for all streams, indicating the type of audiovisual stream (e.g. view from a certain camera) in the metadata. This approach is agnostic to the stream representation of the content, i.e. it could be multiplexed into a single item or be represented by several parallel items.
- Alternate packages representing the audiovisual content for a viewpoint. This requires that sources for different views are not multiplexed into one stream and allows accessing each stream separately. However, the semantics for playing several or all of the views is lost in the description and thus application defined.

MXF DMS-1 defines three frameworks for descriptive metadata: The production framework contains metadata related to all clips and all tracks, the clip framework contains metadata related to a single clip and the scene framework contains metadata for a set of related clips. Typically, the clips described as one scene are temporal segments of the same track. For multi-view video content the scene framework is the only one that could be used. However, a scene will then describe a set of temporal clips from a number of tracks that represent the different views. The semantics will be defined by the metadata of the individual tracks (e.g. camera ID) and their temporal relation. There are no means to describe metadata relating the different views (e.g. relative position information).

MPEG-7 provides flexible mechanisms for describing spatiotemporal decompositions of content and to attach metadata to each of the segments. However, as has been pointed out in other context (e.g. [10]), MPEG-7 allows to create descriptions that convey the same semantics but use different description tools and thus potentially cause interoperability problems. As there is no specific concept for multi-view video content the same problem applies here. Media source

decomposition tools can be used to describe the decomposition of a content segment into constituent (subsequent) media of tracks (such as views). However, the semantics are not clear due to the following two issues:

- Structural composition: the decomposition of views could happen on any level, i.e. one could decompose the root segment representing the entire production into views and the describe the temporal decomposition (e.g. shots, scenes) separately for each view, or one could create a temporal structure of the content and then decompose each clip into views.
- Specification of decomposition criteria: unfortunately this is not a controlled property but free text, so that the semantics of a media source decomposition (e.g. whether into temporally subsequent media or views) are not well defined.

MPEG-7 provides no standard means to describe metadata relating the different views (e.g. relative position information).

EBU P_Meta The ItemGroup in P_Meta is intended the express the editorial relation of content items. It could be used to describe items representing different views of the content. An explanatory note element is provided to describe the relations informally. P_Meta provides no standard means to describe metadata relating the different views (e.g. relative position information).

3.2 Coverage of Required Metadata Elements

Traditional technical metadata, i.e. properties also needed for single-view content, is well covered by many standards, especially the SMPTE Metadata Dictionary and MPEG-7. P_Meta focuses on content exchange and thus mainly covers the technical properties needed there. The technical properties that are especially relevant for multi-view video content are not yet well supported by existing standards. Some camera calibration metadata elements are included in the SMPTE Metadata Dictionary while lens metadata is largely missing in all the standards investigated. Audio metadata are sufficiently covered in the SMPTE Metadata Dictionary, MPEG-7 and P_Meta.

The general descriptive metadata elements and identification metadata are well covered by all standards. The same holds for production metadata. Basic rights metadata is sufficiently supported by the standards coming from the motion picture and broadcast industries, while MPEG-7 lacks some elements[2], and the situation for publication and distribution metadata is similar. For process related metadata, the SMPTE metadata dictionary provides much better support than the other standards. Basic content description and relational metadata is available in all standards.

[2] Of course MPEG-21 could be used to complement this lack.

	structural	**calibration**	**lens**	**identif., prod.**	**process**	**rights**
SMPTE RP210	n/a	some	no	yes	yes	yes
MXF DMS-1	streams	→ RP210	no	yes	→ RP210	yes
MPEG-7	views (informal)	no	no	yes	no	limited
EBU P_Meta	views (informal)	no	no	yes	no	yes

Table 1. Summary of structural and metadata support for multi-view content in selected standards.

4 Summary and Outlook

We have analyzed the metadata requirements to describe multi-view video content and the coverage of these requirements in existing metadata standards. The analysis has shown that several metadata standards can be used for describing multi-view video content. As shown in Table 1, most of the required elements are covered by at least some of the standards. None of the standards provides supports for lens and some calibration metadata elements, so that one has to revert to proprietary or manufacturer specific solutions in this case. This is of course very unsatisfactory w.r.t. interoperability.

Concerning the structural description of multi-view video content we have identified possible solutions in all of the standards. However, in many cases several possible solutions exist, and the semantics are not defined in the standard. Application specific qualifiers and extensions are required in the structural description, leading to formally standard compliant descriptions, but with application defined semantics. Again, this leads to interoperability issues.

In order to improve the metadata workflow in multi-view content production, and establish interoperability between devices and tools, the following issues need to be addressed in standardization:

- Support the required calibration and lens metadata. These metadata elements are hardware related and need to be embedded with the captured essence. Thus SMPTE RP210 seems to be the appropriate standard for this kind of metadata.
- Structural description. Several standards are capable of describing multi-view content, but the semantics for using the standards' tools for representing multi-view content need to specified. This could for example achieved by defining MPEG-7 profiles.

Acknowledgements

The authors would like to thank their partners in the 2020 3D Media project for their contributions to requirements and information about currently used metadata in the production and distribution workflow. The research leading to this paper has been partially supported by the European Commission under the contract FP7-215475, "2020 3D Media – Spatial Sound and Vision" (`http://www.20203dmedia.eu/`).

References

1. Schinas, K., Schmidt, W., Höller, F., Zeiner, H., Bailer, W., Hausenblas, M.: D3.2.1 Metadata in the Digital Cinema Workflow and its Standards. Public deliverable, IP-RACINE (IST-2-511316-IP) (2005) `http://www.ipracine.org/documents/Del_3_2_1_metadata.pdf`.
2. Bailer, W., Schallauer, P.: Metadata in the audiovisual media production process. In Granitzer, M., Lux, M., Spaniol, M., eds.: Multimedia Semantics - The Role of Metadata. Volume 101 of Studies in Computational Intelligence. Springer (Jun. 2008) 65–84
3. Ouyang, C., Rosa, M.L., ter Hofstede, A.H., Dumas, M., Shortland, K.: Toward web-scale workflows for film production. IEEE Internet Computing **12**(5) (2008) 53–61
4. SMPTE: Metadata dictionary registry of metadata element descriptions. SMPTE RP210.11 (2004)
5. SMPTE: Material Exchange Format (MXF) - Descriptive Metadata Scheme-1. SMPTE 380M (2004)
6. ISO: Information Technology - Multimedia Content Description Interface (MPEG-7). ISO/IEC 15938 (2001)
7. EBU: EBU P_META 2.0 Metadata Library. EBU Tech 3295-v2 (Jul. 2007)
8. SMPTE: Material Exchange Format (MXF) - File Format Specification. SMPTE 377M (2004)
9. SMPTE: Material Exchange Format (MXF) - MXF Generic Container. SMPTE 379M (2004)
10. Troncy, R., Bailer, W., Hausenblas, M., Hofmair, P., Schlatte, R.: Enabling Multimedia Metadata Interoperability by Defining Formal Semantics of MPEG-7 Profiles. In: 1^{st} International Conference on Semantics And digital Media Technology (SAMT'06), Athens, Greece (2006) 41–55

Multimedia Processing on Multimedia Semantics and Multimedia Context

Yiwei Cao, Ralf Klamma, Dejan Kovachev

Informatik 5 (Information Systems), RWTH Aachen University
Ahornstr. 55, D-52056, Aachen, Germany
{cao,klamma,kovachev}@dbis.rwth-aachen.de

Abstract. Context awareness and multimedia are observed together for multimedia retrieval. But multimedia semantics and multimedia context are often researched separately in applied multimedia information systems for communities of practice. As the information explosion on the Internet and different devices, we propose a model to identify the information flow of multimedia processing. We associate multimedia semantics with context information. This model can be further evaluated in mobile multimedia information systems which require context-awareness and multimedia retrieval with higher relevance.

1 Introduction

Context awareness and multimedia are important factors for multimedia retrieval in multimedia applications. But multimedia semantics and multimedia context are often researched separately in applied multimedia information systems for communities of practice. In computer science *context* can be understood as any situational or environmental information with an in depth definition survey in [1]. Multimedia semantics cannot be well processed directly by machines. So multimedia metadata is an crucial approach to computer-processing multimedia semantics [14].

Since the beginning of this century, amount and accessibility of multimedia data have been increased greatly. In comparison to textual information, multimedia information has higher richness. Multimedia creation has been becoming an online activity of everybody who has the Internet access. Meanwhile, handheld devices get more and more compact and multi-functional. The cost of mobile networks gets cheaper. Mobile users can take these advantages to create, process and share multimedia data everywhere and every time. The vision of ubiquitous computing [25] is being realized. With the current research advances, multimedia data accessibility can be enhanced by clear multimedia semantics rather than automatic image processing [19].

There is a great amount of multimedia context information generated together with multimedia creation processes. For example, various information about one image in Flickr on the Web 2.0 can be identified in Figure 1. The context information has its semantics, which can be used for multimedia search and retrieval.

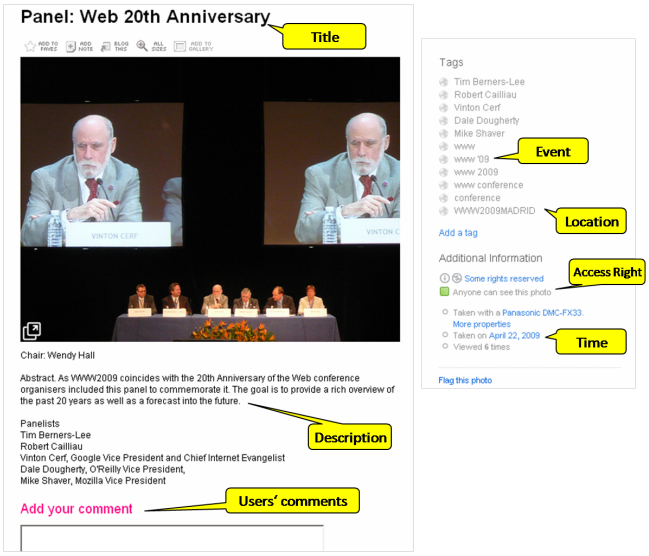


Fig. 1. Multimedia semantics for a photo in Flickr

The problems are obvious. Multimedia semantics and multimedia context are often observed and researched separately in research areas of multimedia information systems. Some of those multimedia information systems focus on multimedia adaptation and personalization, while some of them focus on context-awareness. In fact, both semantic and context have been working together well. An example is the search engine on the Web like Google, which provides suggestions in the search input field. When a song title is typed, often *lyrics* is attached which indicates the context. A further approach to application of context information for multimedia adaptation in the mobile environment is discussed in [15].

In our recent research we associate multimedia context information with multimedia semantics. Semantics information alone can be erroneous. So is context information. We propose a model to identify the information flow and to associate multimedia semantic and context information together, using ontology and impacts of communities of practices. This model can be further evaluated in mobile multimedia information systems which require context-awareness and multimedia retrieval with higher relevance.

Research questions are addressed: how is the complexity and correctness to extend multimedia metadata into ontology with regard to context information and domain information. How effective will it be to use different kinds of ontology?

The rest of this position paper is structured as follows. Section 2 introduces the relevant concepts of multimedia semantics and context. We propose a model for multimedia processing to deliver better multimedia search results by associating multimedia semantics and multimedia context in Section 3. Section 4 addresses

open issues which could arise and need to be dealt with. We conclude the paper with an outlook at future work in Section 5.

2 Terminologies in the Related Work

Semantics is a concept in comparison to *syntax*. Any expression has the semantics so that information is passed. Thus, semantics can be expressed in various formats, under which the most clearly one is in text. In Semantic Web *semantics* is specified as degree of both machine-readability and human-readability. It is stated that machine readable content has quite low semantics [4].

In [9] context is categorized into four groups: *computing context* such as network connectivity, communication bandwidth, display size of the end devices; *user context* such as users' preferences, communities which users belong to; *physical context* such as lighting, location, noise levels, and temperature; and *time context* which can be used as timestamps to identify the records of a context history. Context is widely addressed to device profile, especially referred to those handheld devices with limited capacity. Hence, context-aware adaptation is related to device [23]. Dynamic aspects of context include environmental, spatial or location related, temporal, domain related, and even community related [6].

Le Grand et al. proposed that contextual and semantic information is used together to enrich ontology in order to enhance information retrieval [13]. They employed the concept of *context awareness* to express the relationships among different concepts to complete the ontology. Multimedia semantics and context information together can enhance *information richness*, which is defined as the capacity to clarify ambiguous issues of media communication [10].

Metadata is supposed to fulfill the tasks such as identifying items uniquely worldwide, describing collection items including their *contexts*, supporting retrieval and identification, grouping items into collections within a repository, recording authenticity evidence, facilitating information interchange between autonomous repositories etc. in the domain of digital objects preservation [12].

3 A Model for Multimedia Information Processing

We propose a model depicted in Figure 2 to represent the usage of multimedia semantics and multimedia context information in order to enhance multimedia retrieval. This model is based on the analysis of the impacts of multimedia, metadata, domain information, context, and communities of practice.

A great amount of multimedia information is available. Content description has proved to be an effective way to label or annotate multimedia information [19]. Two approaches are often used to annotate multimedia. One is the Web 2.0 prevalent tagging in free text. The other is adding meta information in line with certain multimedia metadata standards.

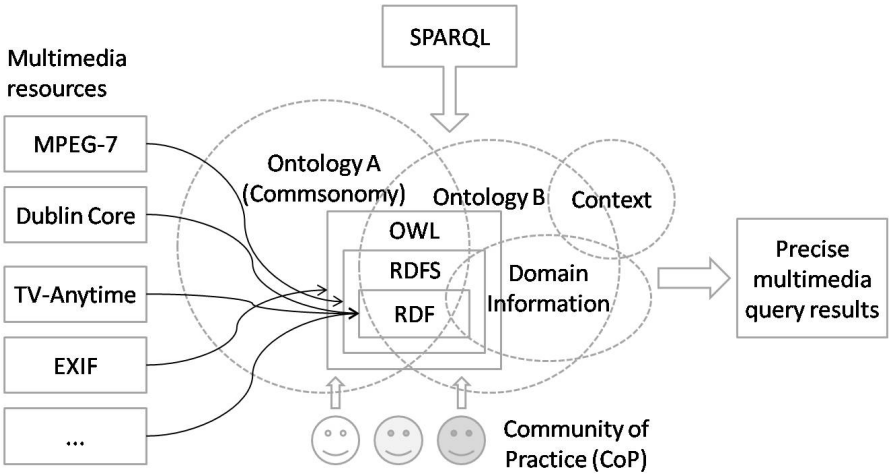


Fig. 2. A multimedia processing model combining multimedia semantics and context

3.1 Metadata Mapping

On the level of metadata standards, a large variability exists again. MPEG-7 standard [17] is one of the richest multimedia content description standards with a comprehensive schema. MPEG-7 is able to express multimedia content covering the most important media aspects including low-level technical information and high-level content semantics. The semantic information expressions may distinguish multimedia creators from the depicted people in a picture or a video clip. MPEG-7 can also be easily used with other metadata standards together, due to its flexible schema. Besides those advantages, MPEG-7 has limitations in semantic expression. Although it has defined many semantic tags, it is still impossible to cover semantic information across different domains. Thus, in different domains several metadata standards can be prevalent in use, such as Dublin Core for digital libraries or digital information preservation [12]. Metadata standards are also used for multimedia adaptation straightforwardly, such as TV-Anytime [11] for adaptive personalized TV programs. The widely spread metadata standard EXIF [21] describes the low-level technical, device, and semantic information such as creation information of images.

Employment of metadata standards aims at enabling data exchange with enhanced data interoperability. However, different metadata standards enhanced data interoperability to certain extent. Metadata standards facilitate data with an effective means to create, describe, search and retrieve multimedia data. Incompatibility and high variety still exist. Terms like *meta-metadata* was coined or crosswalks among different metadata standards have been attempted. It is trivial to specify crosswalks among different metadata standards. A mapping is needed in any two of metadata standards. A transitive mapping can be impossible

theoretically. But information lost and imprecise mapping might lead to many other relevant problems or unexpected consequences.

3.2 Ontology to Bridge Multimedia Semantics and Context

Our approach is to use ontology models to avoid the complexity of mapping among different multimedia metadata standards. The goal is to enrich multimedia semantics with enhanced multimedia interoperability among different multimedia formats and diverse multimedia metadata standards. Ontology represented by a series of concepts which are tightly related to certain domain knowledge.

Context can be modeled by different approaches including key-value, markup scheme, graphical, object-oriented, logic-based, and ontology-based models [22]. Above all, the ontology-based context modeling approach is well evaluated for the purpose to describe context information clearly [6]. Different from the approach in [13], we use concepts specified in certain ontology to represent context information. This context information includes spatial, temporal, community and is modeled in ontology according to domain information.

On the metadata level, RDF, RDFS as well as OWL are proposed as Semantic Web technologies. Resource Description Framework (RDF) [3] provides data model specifications and XML-based serialization syntax. RDF Schema (RDFS) specifies RDF to simplify the process of using Web Ontology Language OWL [2] and also enables the definition of domain ontologies and sharing of domain vocabularies [24]. OWL can be used for the following purposes: (1) *domain formalization*, a domain can be formalized by defining classes and properties of those classes; (2) *property definition*, individuals and assert properties about them can be defined; (3) *reasoning*, one can reason about these classes and individuals. Thus, RDF together with RDFS and OWL can represent context with the information from a certain domain or communities of practice. The SPARQL Protocol and RDF Query Language (SPARQL) can be used for context reasoning [20].

3.3 Commsonomy

We propose *Commsonomy* which is community based folksonomy defined and used within and across communities of practice [16]. Folksonomy come into being as a kind of wide-spread taxonomy with unlimited concepts created by users on social network sites. Commsonomy is a sub set of folksonomy with certain community impacts. Concepts or labels in use could be limited to certain community context.

We employ the concept of *Community of practice*, when we refer to the term *community*. Community of practice is formed because users in communities of practice are engaged with tasks in a mutual way, share a common repertoire, and build up a jointly enterprise [26]. The results from our prior research show that the number of tags or keywords in use decreases as the users attain more expertise knowledge within a community of practice [8].

A suitable commsonomy can cover the knowledge gap which often occurs in Semantic Web. We try to supplement the background knowledge with commsonomy. Mika notifies that lack of background knowledge leads to knowledge gap greatly [18].

With the help of an ontology-based context model using OWL/RDF and the substantially enhanced interoperability, context information can be expressed and reasoned across systems. In summary, the reasoning with SPARQL is carried out on the data set of semantics, context even knowledge or information from communities. The goal is to use multiple dimensions of information to indentify, analyze and reduce the possible information errors.

4 Discussions on Open Issues

In our previous research, we have proposed an approach to multimedia adaptation with regard to context awareness and mobility [7]. Basic queries on context information have been conducted in SPARQL. As the next step proposed in this paper, context model will work together with the multimedia semantic models mapped from different metadata standards.

A potential benefit of this model is targeted for mobile communities. The goal is to deliver mobile users *right* multimedia information on demand on the fly. There might be a lot of scenarios for *Multimedia on the fly*. Users can generate different multimedia with their mobile devices en route. They would also like to search for multimedia for news, local news, and entertainment options etc. People like to contribute and to share information. Ontology is set up in order for multimedia information systems to define rules and apply reasoning on it. Furthermore, some business models should be interesting and useful. In order to get a large set of data, social network sites APIs can be used to collect getagged multimedia originally uploaded across those sites. The tags can be conveyed with the MPEG-7 metadata standards.

5 Conclusions

Semantics, context domain information with certain predefined ontology can help users get better multimedia search results. We analyze the multimedia information flow in community information systems. The information flow includes various multimedia data in different formats, diverse metadata for content or technical description, context information, and the community impacts. Based on this analyze, we propose a model to specify this information and relationships or impacts among these different categories of multimedia related information. In future research, the model can be validated and applied on context-aware mobile multimedia community information systems within the German Excellence Research Cluster UMIC [5].

Acknowledgments. This work has been supported by the UMIC Research Centre, RWTH Aachen University. We would like to thank our colleagues for the fruitful discussions.

References

1. G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles. Towards a better understanding of context and context-awareness. In *HUC '99: Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing*, pages 304–307, London, UK, 1999. Springer-Verlag.
2. S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. A. Stein. OWL Web Ontology Language Reference. [Online], 2004.
3. D. Beckett and B. McBride. RDF/XML Syntax Specification (Revised). [Online], 2004. last access: 1.10.2007.
4. T. Berners-Lee, J. A. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 5 2001.
5. Y. Cao, X. Chen, N. Drobek, R. Klamma, and et al. Virtual campfire - cross-platform services for mobile social software (demo paper). In *Proc. of the Tenth International Conference on Mobile Data Management, May 18-20, 2009, Taipei, Taiwan*, 2009.
6. Y. Cao, R. Klamma, M. Hou, and M. Jarke. Follow me, follow you - spatiotemporal community context modeling and adaptation for mobile information systems. In *Proceedings of the 9th International Conference on Mobile Data Management, April 27-30, 2008, Beijing, China*, pages 108–115. IEEE Society, 4 2008.
7. Y. Cao, R. Klamma, and M. Khodaei. A multimedia service with MPEG-7 metadata and context semantics. In R. Grigoras, V. Charvillat, R. Klamma, and H. Kosch, editors, *Proceedings of the 9th Workshop on Multimedia Metadata (WMM'09), Toulouse, France, March 19-20, 2009, CEUR-WS Vol. 441, http://CEUR-WS.org/Vol-441/*, 2009.
8. Y. Cao, R. Klamma, and A. Martini. Collaborative storytelling in the Web 2.0. In R. Klamma, N. Sharda, B. Fernández-Manjón, H. Kosch, and M. Spaniol, editors, *Proceedings of the First International Workshop on Story-Telling and Educational Games (STEG'08) at EC-TEL 08, Sep. 16, 2008, Maastricht, the Netherlands*. CEUR-WS.org, 2008.
9. G. Chen and D. Kotz. A Survey of Context-Aware Mobile Computing Research. Technical report, Dartmouth College, Hanover, NH, USA, 2000.
10. R. L. Daft and R. H. Lengel. Organizational informations requirements, media richness and structural design. *Management Science*, 32(5):554 – 571, 1986.
11. European Telecommunications Standards Institute. Technical specification. broadcast and online services: Search, select, and rightful use of content on personal storage systems (tv-anytime). ETSI TS 102 822-3-1 V1.3.1, 2005.
12. H. M. Gladney. *Preserving Digital Information*. Springer, 2007.
13. B. L. Grand, M.-A. Aufaure, and M. Soto. Semantic and conceptual context-aware information retrieval. In *Advanced Internet Based Systems and Applications*, volume 4879 of *LNCS*. Springer, Berlin, Heidelberg, April 2009.
14. R. Klamma, Y. Cao, and M. Spaniol. Smart social software for mobile cross-media communities. In M. Granitzer, M. Lux, and M. Spaniol, editors, *Multimedia Semantics - The Role of Metadata*, volume 101 of *Studies in Computational Intelligence*. Springer, 2008.

15. R. Klamma, M. Spaniol, and Y. Cao. Community Aware Content Adaption for Mobile Technology Enhanced Learning. In *Online-Proceedings of the1st European Conference on Technology Enhanced Learning (EC-TEL 2006), Hersonissou, Greece, October 1-3*, pages 227–241. Springer-Verlag, 2006.
16. R. Klamma, M. Spaniol, and D. Renzel. Community-Aware Semantic Multimedia Tagging - From Folksonomies to Commsonomies. In K. Tochtermann, H. Maurer, F. Kappe, and A. Scharl, editors, *Proceedings of I-Media '07, International Conference on New Media Technology and Semantic Systems, Graz, Austria, September 5 - 7*, J.UCS (Journal of Universal Computer Science) Proceedings, pages 163–171, 2007.
17. H. Kosch. *Distributed Multimedia Database Technologies Supported by MPEG-7 and MPEG-21.* CRC Press, Boca Raton et al., 2003.
18. P. Mika. *Social Networks and the Semantic Web.* Springer, New York, NY, USA, 2007.
19. M. Spaniol, R. Klamma, and M. Lux. Imagesemantics: User-Generated Metadata, Content Based Retrieval & Beyond. pages 41–48, 2007.
20. SPARQL. SPARQL Protocol and RDF Query Language. http://en.wikipedia.org/wiki/SPARQL, 2007 [last access: 1.10.2007].
21. Standard of Japan Electronics and Information Technology Industries Association. Exchangeable image file format for digital still cameras: EXIF version 2.2. JEITA CP-3451, 2002.
22. T. Strang and C. Linnhoff-Popien. A Context Modeling Survey. In *First International Workshop on Advanced Context Modelling, Reasoning And Management at UbiComp*, Nottingham, UK, 09 2004.
23. C. Timmerer, J. Jabornig, and H. Hellwagner. Delivery context descriptions - a comparison and mapping model. In R. Grigoras, V. Charvillat, R. Klamma, and H. Kosch, editors, *Proceedings of the 9th Workshop on Multimedia Metadata (WMM'09), Toulouse, France, March 19-20, 2009, CEUR-WS Vol. 441, http://CEUR-WS.org/Vol-441/*, 2009.
24. X. H. Wang, D. Q. Zhang, T. Gu, and H. K. Pung. Ontology Based Context Modeling and Reasoning using OWL. In *PERCOMW '04: Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications Workshops*, pages 18–22, Washington, DC, USA, 2004. IEEE Computer Society.
25. M. Weiser. Some computer science issues in ubiquitous computing. *SIGMOBILE Mob. Comput. Commun. Rev.*, 3(3):12, 1999.
26. E. Wenger. *Communities of Practice: Learning, Meaning, and Identity.* Cambridge University Press, Cambridge, UK, 1998.

Virtual Campfire - Collaborative Multimedia Semantization with Mobile Social Software

Dominik Renzel[1] , Ralf Klamma[1], Yiwei Cao[1], Dejan Kovachev[1]

[1] Chair for Computer Science 5, Information Systems & Databases
RWTH Aachen University,
Ahornstr. 55, 52056 Aachen,
{renzel, klamma, cao, kovachev}@dbis.rwth-aachen.de

Abstract. Apart from automated techniques, collaborative multimedia semantic annotation by people, in particular communities of domain experts, are and will be powerful contributors to multimedia semantization. Recently, we extended our LAS server towards an XMPP server, thereby enabling the real time intertwining of communication and the collaborative utilization of remote services, in particular MPEG-7 multimedia semantic annotation and retrieval services. Within the context of the UMIC Virtual Campfire scenario, this contribution presents a set of mobile multimedia semantic annotation services and tools and their usage in a collaborative multimedia annotation scenario for cultural heritage documentation. In particular NMVX, an MPEG-7 multimedia semantic annotation tool in conjunction with a standard XMPP IM client are demonstrated, both powered by the same LAS XMPP server.

Keywords: Multimedia semantic annotation, MPEG-7, mobile social software, SOA, XMPP, community information systems

1 Introduction to Virtual Campfire

In recent years, great progresses have been made in technologies of mobile network technologies, mobile applications and services, mobile user interfaces, and mobile devices. Like the Linux developer communities in the early years, developers have paid attention to applications and services running on mobile devices such as Java ME devices, iPhones, Google Android devices, etc. User generated services and applications on mobile devices are going ahead together with user generated content on the Web 2.0.

Challenges in developing mobile services and applications are multifold. There is a great variety of mobile standards, operating systems on different devices. Often unfortunately, one application can work on one cell phone well, while it does not work on the other. Meanwhile, Social Software allows users to be content prosumers (consumer and producer in parallel) anywhere at any time. Web 2.0 and Social Software result in a great amount of multimedia content which should be used by mobile communities. How well the mobile services and applications work is also hard to measure.

Virtual Campfire is an approach to providing cross-media and cross-community support for the management of multimedia contents. It serves as a framework for various services enabling communities to share knowledge about multimedia contents.

The core of this framework is a Lightweight Application Server (LAS) [3] serving as the backbone of the Virtual Campfire framework to show its applicability in various application scenarios (cf. Figure 1). It provides communities a set of core services and MPEG-7 semantic multimedia metadata and content processing services to connect to heterogeneous data sources. Furthermore, storytelling services [4], context-aware search services [1] etc. use MPEG-7 services to re-contextualize multimedia content via a non-linear storytelling approach and to search multimedia by giving spatiotemporal and community context information.

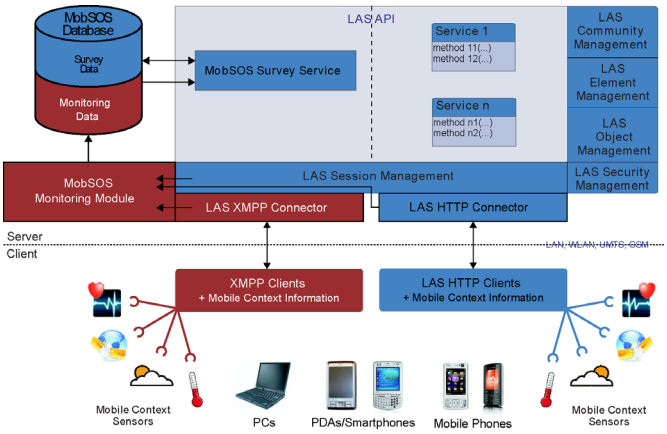


Fig. 1. The LAS Service Architecture for Mobile Applications

Recently, we extended LAS by a connector for XMPP [5], a bidirectional XML-streaming protocol with built-in pull/push/broadcast and server-to-server communication, TLS/SASL encryption, etc. Currently, the protocol core and the standard Jabber RPC XEP were implemented. A LAS XMPP Extension Framework enables the integration of arbitrary extension protocols by the implementation of connection and namespace handlers. The current implementation allows simultaneous utilization of direct user-to-user communication, remote service invocations, etc. Such a scenario is presented in Section 3 with NMVX, a LAS MPEG-7 application enabled to connect over XMPP.

The MobSOS testbed is an extension of LAS primarily designed for the measurement of multimedia service success. The underlying homonymous success model combines qualitative and quantitative measures and takes into account modern requirements for mobile multimedia communities. Model data is collected using the two techniques of monitoring and user surveying. Besides the usage for service success measurement, monitoring data transmitted by mobile devices, in particular context information is used for context-aware services such as automatic MPEG-7 based semantic tagging (e.g. location, time).

The scenario we present here is a cultural heritage documentation scenario of the giant Buddha statues in the Bamiyan Valley. While an on-site researcher team is

actively collecting multimedia with mobile capturing and annotation tools, off-site researchers use desktop applications to immediately access the captured media in order to semantically enrich annotations from their colleagues and to re-contextualize the media using non-linear storytelling tools. All scientists coordinate their work with each other remotely, using synchronous communication tools such as XMPP chat clients.

The following sections present a selection of our prototypes for the demonstration.

2 Mobile Multimedia Capturing & Annotation

NMV Mobile (cf. Figure 2, left) is a multimedia capturing, sharing and annotation system powered by LAS MPEG-7 services. It supports free text annotations, plain keyword tagging as well as MPEG-7 standard compliant community-based semantic tagging to enhance semantic multimedia search and retrieval. NMV Mobile demonstrates the access to mobile context sensors (e.g. GPS) for automatic semantic tagging. The application is realized for J2ME enabled devices compliant with MIDP2.1/CLDC1.1 and demonstrated on a Nokia N95.

ACIS is a GIS enabled multimedia information system hosting diverse user communities [2] and facilitates the intergenerational cooperation among communities on an international level. Similar to NMV Mobile, *iNMV* (cf. Figure 2, right) uses the iPhone GPS sensor to automatically tag photos with spatiotemporal information and upload and retrieve multimedia incl. MPEG-7 metadata using our services.

Fig. 2 NMV Mobile on the Nokia N95 (left) & iNMV on the iPhone (right)

3 Collaborative Multimedia Annotation

The LAS XMPP extension was recently demonstrated in a collaborative multimedia annotation scenario using NMVX, an XMPP enabled version of the NMV desktop version in conjunction with the standard XMPP instant messaging client Pidgin (cf. Figure 3). Both tools connect to the same LAS server via XMPP. While direct communication among users is performed as a chat in Pidgin, multimedia semantic annotations are assigned using NMVX. In further versions of NMVX we intend the integration of direct communication with service invocation in one tool.

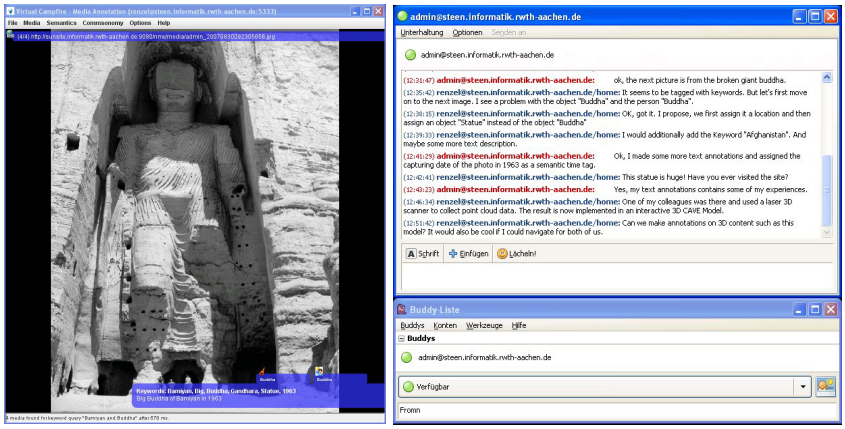


Fig. 3. Collaborative Multimedia Annotation with NMVX/Pidgin

4 Innovative Aspects of Virtual Campfire

Virtual Campfire offers an open architecture that helps professionals flexibly create information systems in versatile application domains. It combines advanced multimedia standards and database technologies that support the creation of mobile information systems on heterogeneous devices.

Acknowledgments. This work has been supported by the UMIC Research Centre, RWTH Aachen University and the EU FP7 IP ROLE. We would like to thank our colleagues for the fruitful discussions.

References

[1] Y. Cao, R. Klamma, M. Hou, M. Jarke: Follow Me, Follow You - Spatiotemporal Community Context Modeling and Adaptation for Mobile Information Systems. Proc. of the 9th International Conference on Mobile Data Management, April 27-30, 2008, Beijing, China, pp. 108-115.

[2] R. Klamma, M. Spaniol, M. Jarke, Y. Cao, M. Jansen and G. Toubekis, “Standards for Geographic Hypermedia: MPEG, OGC and co.”, E. Stefanakis, M.P. Peterson, C. Armenakis, V. Delis (Eds.): Geographic Hypermedia - Concepts and Systems, LNG&C, ISBN 3-540-34237-0, Springer-Verlag, pp. 233-256, 2006.

[3] M. Spaniol, R. Klamma, H. Janßen and D. Renzel: LAS, “A Lightweight Application Server for MPEG-7 Services in Community Engines”, K. Tochtermann, H. Maurer (Eds.): Proceedings of I-KNOW '06, Graz, Austria, September 6 - 8, 2006, J.UCS (Journal of Universal Computer Science) Proceedings, Springer, pp. 592-599.

[4] M. Spaniol, R. Klamma, N. Sharda and M. Jarke, “Web-Based Learning with Non-linear Multimedia Stories”, W. Liu, Q. Li, R. W. H. Lau (Eds.): Advances in Web-Based Learning, Proceedings of ICWL 2006, Penang, Malaysia, July 19-21, Springer-Verlag, Berlin Heidelberg, LNCS 4181, pp. 249-263, 2006.

[5] P. Saint-Andre, Extensible Messaging and Presence Protocol (XMPP): Core, Oct. 2004.

Multimedia Ontology Life Cycle Management with the SALERO Semantic Workbench

Tobias Bürger

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

Abstract. Ontologies are gaining increased importance in the area of multimedia retrieval or management as they try to overcome the commonly known drawbacks of existing multimedia metadata standards for the descriptions of the semantics of multimedia content. In order to build and use ontologies, user have to receive appropriate support. This paper presents the *SALERO Semantic Workbench* which offers a set of services to engineer and manage ontologies throughout their life cycle, i.e., from their (semi-) automatic creation through its storage and use in annotation and search.

1 Introduction

The overall goal of the integrated project SALERO[1], as introduced in [1], is to define and develop "intelligent content" with context-aware behavior for self-adaptive use and delivery across different platforms, building on and extending research in media technologies and web semantics to reverse the trend towards ever-increasing cost of creating media. To support the aforementioned aim, semantic technologies have been identified as a viable solution [2]. Ontologies are commonly acknowledged as being a core ingredient of any solution based on semantic technologies as a means to capture the semantics of a domain of discourse and to provide formally represented machine readable models of it. In that sense, one goal of the SALERO project is to create ontologies which support the annotation and semantic search of media resources.

In order to pave the way for the use of ontologies and semantic technologies in media production, SALERO developed a management framework for multimedia ontologies, tools to annotate existing media resources and semantic search facilities to retrieve resources based on semantic annotations. The framework, which offers these functionalities, is introduced in the following.

2 The SALERO Semantic Workbench

The *SALERO Semantic Workbench* supports the creation, management, and use of domain ontologies which includes the following main functionalities (cf. Figure 1):

[1] http://www.salero.eu

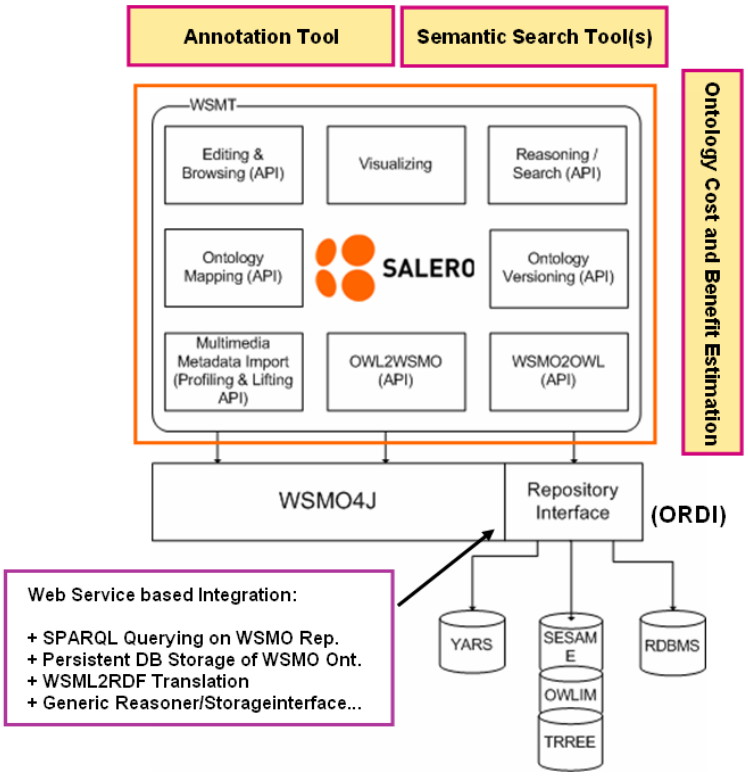


Fig. 1. The SALERO Semantic Workbench – High Level Architecture

1. **Ontology Management** whose central aspects include manual and semi-automatic creation of domain ontologies, alignment of different domain descriptions, translations of ontologies, versioning, or storage of ontologies.
2. **Annotation Support** whose central aspects includes the support for non-technological users with the annotation of media items which is realized in several annotation tools.
3. **Semantic Search Support** which offers advanced retrieval capabilities based on semantic annotations.

To realize this functionality, the workbench not only offers a graphical user interface to engineer ontologies, but also a set of services which provide ontology management functionality to other applications. The workbench acts in the background and its central aspects are thus realized as an API which is designed with the aim to integrate the functionality needed for semantic media annotation and semantic search into plug-ins and interfaces of other applications. This includes most notably storage, querying, or retrieval of annotations.

2.1 Ontology Management

The part of the workbench for the management of ontologies is based on the Web Service Modeling Toolkit (WSMT) that, among others, provides a set of graphical tools for the engineering of WSMO ontologies, for the interaction with external tools such as execution environments and repositories [3].[2] WSMT is a collection of tools for the engineering of Semantic Web Services and ontologies implemented in the Eclipse framework.[3] In SALERO we added the possibility to persistently store and access ontologies in an ontology repository as realized by the *Repository Service* described below.

2.2 Workbench Services

The services offered by the semantic workbench include:

- **The Repository Service** which offers an API for the persistent storage of WSML[4] ontologies and their elements (e.g., concepts, properties, axioms). It supports management of these elements and the execution of SPARQL queries. The service is realized on top of the *Ontology Representation and Data Integration (ORDI)* – framework, which most notably provides a scalable repository implementation, a WSMO-RDF parser, serializer, and access to query and reasoning facilities.[5]
- **The Annotation Service** is concerned with the management of semantic annotations and provides an API to manage and validate annotations against the ontologies stored in the repository.
- **The Semantic Search Service** offers an API to search for ontology elements and additionally offers keyword-based search for annotations which is expanded into full-text queries on a generated index and SPARQL queries.
- **The Ranking Service** offers functionality to rank media resources based on semantic annotations. This service is used by the semantic search and the recommendation service.
- **The Recommendation Service** offers an API for retrieval of ontology elements which are prominently used for annotation and gives recommendations of related results during search.
- **The Profiling and Lifting Service** can be used to extract structural semantic information from existing MPEG-7[6] documents and for their semantic enrichment.

2.3 Tool and Annotation Support

Besides tools for ontology management, the workbench is accompanied with a tool set for annotation and semantic search: the *SALERO Intelligent Media*

[2] `http://sourceforge.net/projects/wsmt/`
[3] `http://www.eclipse.org/jdt/`
[4] `http://www.wsmo.org/wsml/wsml-syntax`
[5] `http://www.ontotext.com/ordi`
[6] `http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm`

Annotation & Search (IMAS) system [4]. The IMAS integrates annotation and search into one application and further provides access to content-based search facilities. Both semantic search and content-based search can be accessed via a single interface and the results are being fused.

In order to adequately support non-experienced users in annotation, the workbench offers a methodology to support the users including (i) the selection of adequate ontology elements and (ii) the extension of ontologies during annotation time (cf. [5]).

3 Conclusions

This paper presented the *SALERO Semantic Workbench* which offers functionalities to manage ontologies throughout their life cycle which most notably includes their manual or semi-automatic engineering and use in annotation and search. In SALERO, some services of the workbench have been specialized to be used for media resources, such as the ranking or profiling and lifting service, while other services, and especially the foundational ontology management functionality, can be used with any ontology and any type of resource.

The services of the workbench have been implemented in the course of SALERO and are underlying the previously introduced *SALERO Intelligent Media Annotation & Search* (IMAS) system which is available online.[7] Preliminary evaluation results of the IMAS annotation functionality are presented in [6].

Acknowledgments The research leading to this paper was partially supported by the European Commission under contract IST-FP6-027122 "SALERO".

References

1. Haas, W., Thallinger, G., Cano, P., Cullen, C., Bürger, T.: Salero: Semantic audiovisual entertainment reusable objects. In: Proceedings of the First International Conference on Semantics And Digital Media Technology (SAMT). (2006)
2. Bürger, T.: The need for formalizing media semantics in the games and entertainment industry. Journal for Universal Computer Science (JUCS) (June 2008)
3. Kerrigan, M., Mocan, A., Tanler, M., Fensel, D.: The web service modeling toolkit – an integrated development environment for semantic web services (system description). In: Proceedings of the 4th European Semantic Web Conference (ESWC), June 2007, Innsbruck, Austria. (2007)
4. Weiss, W., Bürger, T., Villa, R., Swamy, P., Halb, W.: Salero intelligent media annotation & search. In: Proceedings of iSemantics 2009. (2009)
5. Bürger, T., Ammendola, C.: A user centered annotation methodology for multimedia content. In: Poster Proceedings of the 5th European Semantic Web Conference (ESWC) 2008. (2008)
6. Weiss, W., Bürger, T., Villa, R., P., P., Halb, W.: Statement-based semantic annotation of media resources. In: Proceedings of Proceesings of the 4th International Conference on Semantics And digital Media Technology (SAMT) 2009. (2009)

[7] `http://salero.joanneum.at/imas/`