W3C Incubator Group Report

Multimedia Annotation Interoperability Framework

W3C Incubator Group Editor's Draft 14 August 2007

This version:
http://www.w3.org/2005/Incubator/mmsem/XGR-interoperability-20070814/

Latest version:
http://www.w3.org/2005/Incubator/mmsem/XGR-interoperability/

Previous version:
This is the first public version.

Editor:
Vassilis Tzouvaras, IVML, National Technical University of Athens
Raphaël Troncy, Center for Mathematics and Computer Science (CWI Amsterdam)
Jeff Z. Pan, University of Aberdeen

Also see Acknowledgements.

Abstract

Multimedia systems typically contain digital documents of mixed media types, which are indexed on the basis of strongly divergent metadata standards. This severely hamplers the inter-operation of such systems. Therefore, machine understanding of metadata comming from different applications is a basic requirement for the inter-operation of distributed Multimedia systems. In this document, we present how interoperability among metadata, vocabularies/ontologies and services is enhanced using Semantic Web technologies. In addition, it provides guidelines for semantic interoperability, illustrated by use cases. Finally, it presents an overview of the most commonly used metadata standards and tools, and provides the general research direction for semantic interoperability using Semantic Web technologies.

Status of This Document

This section describes the status of this document at the time of its publication. Other documents may supersede this document. A list of Final Incubator Group Reports is available. See also the W3C technical reports index at http://www.w3.org/TR/.

This document was developed by the W3C Multimedia Semantics Incubator Group, part of the W3C Incubator Activity.

Publication of this document by W3C as part of the W3C Incubator Activity indicates no endorsement of its content by W3C, nor that W3C has, is, or will be allocating any resources to the issues addressed by it. Participation in Incubator Groups and publication of Incubator Group Reports at the W3C site are benefits of W3C Membership.

Incubator Groups have as a goal to produce work that can be implemented on a Royalty Free basis, as defined in the W3C Patent Policy. Participants in this Incubator Group have made no statements about whether they will offer licenses according to the licensing requirements of the W3C Patent Policy for portions of this Incubator Group Report that are subsequently incorporated in a W3C Recommendation.

Scope

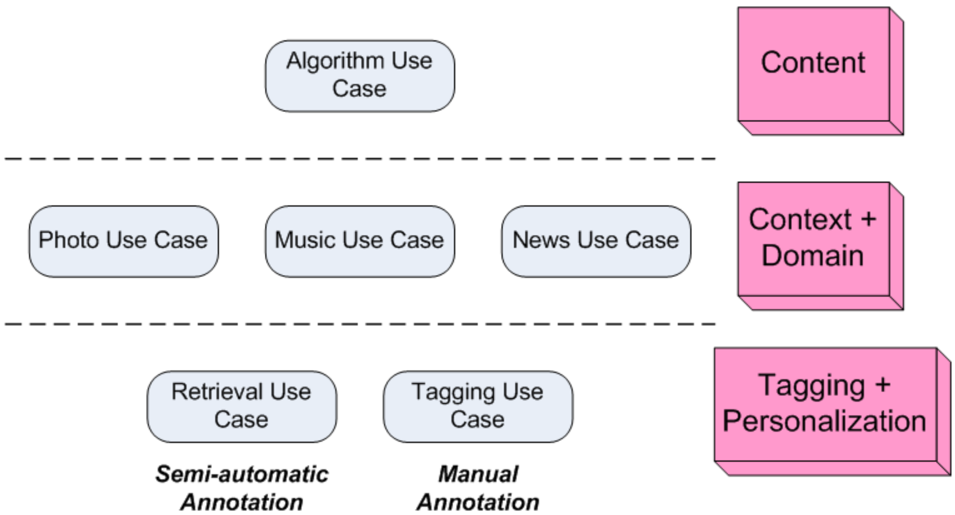
This document targets at people with an interest in semantic interoperability, ranging from non-professional end-users that have content are manually annotating their personal digital photos to professionals working with digital pictures in image and video banks, audiovisual archives, museums, libraries, media production and broadcast industry, etc.

Discussion of this document is invited on the public mailing list public-xg-mmsem@w3.org (public archives). Public comments should include "[MMSEM-Interoperability]" as subject prefix .

Table of Contents

1. Introduction in Semantic Interoperability in Multimedia Applications

This document uses a bottom-up approach to provide a simple extensible framework to improve interoperability of applications related to some key use cases discussed in the XG.



2. Use Cases and Possible Solutions

In this section, we present several use cases showing interoperability problems with multimedia metadata formats. Each use case starts with an example illustrating the main problem, and proposed after a possible solution using Semantic Web technologies.

2.1 Use Case: Photo

Introduction

Currently, we are facing a market in which, for example, more than 20 billion digital photos are taken per year in Europe [GFK]. The number of tools, either for desktop machines or web-based, that perform automatic as well as manual annotation of the content is increasing. For example, a large number of personal photo management tools extract information from the so called EXIF [EXIF] header and add this information to the photo description. These tools typically allow to tag and describe single photos. There are also many web-based tools that allow to upload photos to share them, organize them and annotate them. Web sites such as [Flickr] allow tagging on the large scale. Sites like [Riya] provide specific services such as face detection and face recognition of personal photo collections. Photo community sites such as [Foto Community] allow an organization of the photos in categories and allow rating and commenting on them. Even though our photos today find more and more tools to manage and share them, these tools come with different capabilities. What remains difficult is finding, sharing, reusing photo collections across the borders of tools and sites. Not only the way in which photos are automatically and manually annotated is different but also the way in which this metadata is described and represented finds many different standards. At the beginning of the management of personal photo collections is the semantic understanding of the photos.

Motivating Example

From the perspective of an end user let us consider the following scenario to describe what is missing and needed for next generation digital photo services. Ellen Scott and her family were on a nice two-week vacation in Tuscany. They enjoyed the sun at the beaches of the Mediterranean, appreciating the great culture in Florence, Siena and Pisa, and traveling on the traces of the Etruscans through the small villages of the Maremma. During their marvelous trip, the family was taking pictures of the sightseeing spots, the landscapes and of course from the family members. The digital camera they use is already equipped with a GPS receiver, so every photo is stamped not only with the time when, but also with the geo-location where it has been taken.

Photo annotation and selection

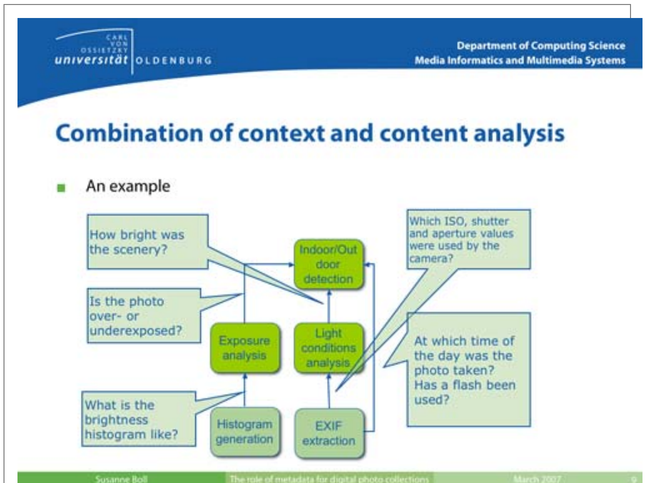
Back home, the family uploads about 1000 pictures from the camera to the computer and wants to create an album for grand dad. On this computer, the family uses a photo management tool which both extracts some basic features such as the EXIF header, but also allows for entering tags and personal descriptions. Still fulfilled with the memories of the nice trip the mother of the family labels most of the photos. With a second tool, the tour of the GPS receiver and the photos are merged using the time stamp. As a results, each of the photos is geo-referenced with the GPS position stored in the EXIF header. However, showing all the photos would take an entire weekend. So Ellen starts to create an excerpt of their trip with the highlights. Her photo album software takes in the 1000 pictures and makes suggestions for the selection and the arrangement of the pictures in a photo album. For example, the album software shows her a map of Tuscany and visualises, where she has taken which photos and groups them together making suggestions which photos would best represent this part of the vacation. For places for which the software detects highlights, the system offers to add information about the place to the album, stating that on this Piazza in front of the Palazzo Vecchio there is the copy of Michelangelo's famous David statue. Depending on the selected style, the software creates a layout and the distribution of all images over the pages of the album taking into account color, spatial and temporal clusters and template preference. So, in about 20 minutes Ellen has finished the album and orders a paper version as well as an online-version. The paper album is delivered to her by mail three days later. It looks great, and the explaining texts that her software has almost automatically added to the pictures are informative and help her remembering the great vacation. They show the album to grandpa and he can take his time to study their vacation and the wonderful Tuscany.

Exchanging and sharing photos

Selecting the most impressive photos, the son of the family uploads a nice set of photos to Flickr, to give his friends an impression of the great vacation. Unfortunately, all the descriptions and annotations from the personal photo management system are lost after the Web upload. Therefore, he adds a few own tags to the Flickr photos to describe the places, events, persons of the trip. Even the GPS track is lost and he places the photos again on the Flickr map application to geo-reference them. One friend finds a cool picture from the Spanish Stairs in Rome by night and would like to get the photo and its location from Flickr. This is difficult again as a pure download of the photo does not retain the geo-location. When aunt Mary visits the Web album and starts looking on the photos she tries to download a few onto her laptop to integrate them into her own photo management software. Now aunt Mary would like to incorporate some of the pictures of her nieces and nephews into her photo management system. And again, the system imports the photos but the precious metadata that mother and sun of family Miller have already annotated twice are gone.

The fundamental problem: semantic content understanding

What is needed is a better and more effective automatic annotation of digital photos that better reflects one's personal memory of the events captured by the photos an allows different applications to create value-added services on top of them such as the creation of a personal photo album book. For understanding the personal photos and overcoming the semantic gap, digital cameras leave us with files like `dsc5881.jpg`, a very poor reflection of the actual event. Is is a 2D visual snapshot of a multi-sensory personal experience. The quality of photos is often very limited (snapshots, over exposed, blurred, ...). On the other hand digital photos come with a large potential for semantic understanding the photos. Photographs are always taken in context. In contrast to analogical photography, digital photos provide us with explicit contextual information (time, flash, aperture, ...), a "unique id" such as the timestamp allows to later merge contextual information with the pure image content.



Indoor/Outdoor detection with signal analysis and context analysis.
Image courtesy of Susanne Boll, used with permission.

However, what we want to remember along with the photo is where it was, who was there with us, what can be seen on the photo, what the weather was, if we liked the event and so on. In recent years, it became clear that signal analysis alone will not be the solution. In combination with the context of the photo, such as the GPS position or the time stamp, some hard signal processing problems can be solved better. So context analysis has gained much attention and became important for photos and very helpful for photo understanding. In the opposite figure, a simple example is given of how to combine signal analysis and context analysis to achieve a better indoor/outdoor detection of photos. And, not only with the advent of the Web 2.0 the actual user came into focus. The manual effort of single user annotations but also collaborative effects are considered to be important for semantic photo understanding.

The role of metadata for this usage of photo collections is manyfold:

- Save the experience: The central goal is to overcome the semantic gap and represent as much of the humans impression of the moment when the photo was taken.
- Browse and find previously taken photos: Allow searching for events and persons, places, moments in time, etc.
- Share photos with the metadata with others: Give your annotated photos from Flickr or from Foto Community to your friends' applications.
- Use comprehensive metadata for value-added services of the photos: Create an automatic photo collage or send a flash presentation to your auntâ€™s TV, notify all friends that are interested in photos from certain locations, events, or persons, etc.

The opposite figure illustrates the use of photos today and what we do with our photos at home but also in the Web.

So the social life of personal photos can be summarized as:

- Capturing: one or more persons capture and event, with one or different cameras with different capabilities and characteristics
- Storing: one or more persons store the photos with different tools on different systems
- Processing: post-editing with different tools that change the quality and maybe the metadata
- Uploading: some persons make their photos available on Web (2.0) sites (Flickr); different sites offer different kinds of value-added services to the photos (Riya)
- Sharing: photos are given away or are given access to via email, Web sites, print, ...
- Receiving: photos from others are received via MMS, email, download, ...
- Combining: Photos from own and different sources are selected and reused for services like T-Shirt, Mugs, mouse pads, photo albums, collages, ...

Photos usage.
Image courtesy of Susanne Boll, used with permission.

For this, metadata plays a central role at all times and places of the social life of our photos.

The multimedia semantics interoperability problem

Different levels and types of metadata for photos

The problem we have here is that metadata is created and enhanced by different tools and systems and follows different standards and representations. Even though there are many tools and standards that aim to capture and maintain this metadata, they are not necessarily interoperable. So on a technical level, we have the problem of a common representation of metadata that is helpful and relevant for photo management, sharing and reuse. Metadata and end user typically gets in touch with descriptive metadata that stem from the context of the photo. At the same time, in more than a decade many results in multimedia analysis have been achieved to extract many different valuable features from multimedia content. For photos for example, this includes color histograms, edge detection, brightness, texture and so on. With MPEG-7, a very large standard has been developed that allows to describe these features in a standardized way. However, both the size of the standard but also the many optional attributes in the standard have lead to a situation in which MPEG-7 is used only in very specific applications and has not been achieved as a world wide accepted standard for adding (some) metadata to a media item. Especially in the area of personal media, in the same fashion as in the tagging scenario, a small but comprehensive shareable and exchangeable description scheme for personal media is missing.

Different standards for photo metadata and annotations

What is needed is a machine readable description that comes with each photo that allows a site to offer valuable search and selection functionality on the uploaded photos. Even though approaches for Photo Annotation have been proposed they still do not address the wide range of metadata, annotations that could and should be stored with an image in a standardized fashion.

- EXIF [EXIF] is a standard that comprises many photographic and capture relevant metadata. Even though the end user might use only a few of the key/value pairs, they are relevant at least for photo editing and archiving tools which read this kind of metadata and visualize it. So EXIF is a necessary set of metadata which is needed for photos.
- Tags from Flickr and other photo web sites and tools are metadata of low structure but high relevance for the user and the use of the photos. Manually added they reflect the users knowledge and understanding of the content which can not be replaced by any automatic semantic extraction. Therefore, a representation of these is needed. Depending on the source of tags is might be of interest to relate the tags to their origin such as "taken from an existing vocabulary", "from a suggested set of other tags" or just "free tags". XMP seems to be a very promising standard as it allows to define RDF-based metadata for photos. However, in the description of the standard, it clearly states that it leaves the application dependent schema /vocabulary definition to the application and only makes suggestions for a set of "generic" sets such as EXIF, Dublin Core. So the standard could be a good "host" for a defined photo metadata description scheme in RDF but does not define it.
- PhotoRDF [PhotoRDF] "describes a project for describing & retrieving (digitized) photos with (RDF) metadata. It describes the RDF schemas, a data-entry program for quickly entering metadata for large numbers of photos, a way to serve the photos and the metadata over HTTP, and some suggestions for search methods to retrieve photos based on their descriptions." So wonderful, but the standard is separated into three different schemas: Dublin Core [Dublin Core], a Technical Schema which comprises more or less entries about author, camera and short description, and a Content Schema which provides a set of 10 keywords. With PhotoRDF, the type and number of attributes is limited, does not even comprise the full EXIF schema and is also limited with regard to the content description of a photo.
- The Extensible Metadata Platform or XMP [XMP] and the IPTC-IIM-Standard [IIM] have been introduced to define how metadata (not only) of a photo can be stored with the media element itself. However, these standards come with their own set of attributes to describe the photo or allow to define individual metadata templates. This is the killer for any sharing and semantic Web search! What is missing is an actual standardized vocabulary what information about a photo is important and relevant to a large set of next generation digital photo services has not been reached.
- The Image Annotation on the Semantic Web [MMSEM Image] provides an overview of the existing standard such as those mentioned above. At the same time it shows how diverse the world of annotation is. The use case for photo annotation choses RDF/XML syntax of RDF in order to gain interoperability. It refers to a large set of different standards and approaches that can be used to image annotation but there is no unified view on image annotation and metadata relevant for photos. The attempt here is to integrated existing standards. If those however are too many, too comprehensive, and might even have overlapping attributes is might not be adopted as the common photo annotation scheme on the Web. For example, for the low level features for example, there is only a link to MPEG-7.
- The DIG35 Initiative Group of the International Imaging Industry Association aims "provide a standardized mechanism which allows end-users to see digital image use as being equally as easy, as convenient and as flexible as the traditional photographic methods while enabling additional benefits that are possible only with a digital format." [DIG35]. The DIG35 standards aims to define a standard set of metadata for digital images that can be widely implemented across multiple image file formats. From all the photo standards this is the broadest one with respect to typical photo metadata and is already defined as a XML Schema.
- MPEG-7 is far to big even though the standard comprises metadata elements that are relevant also for a Web wide usage of media content. The advantage of MPEG-7 is that one can define an own description scheme and with it collect a subset of relevant feature related metadata with a photo. But, there is no chance to actually include an entire XML-based MPEG-7 description of a photo into the raw content. For the description of the content the use case refers to three domain-specific ontologies: personal history event, location and landscape.

Towards a solution

The result is clear, that there is not one standardized representation and vocabulary for adding metadata to photos. Even though the different semantic Web applications and developments should be embraced, a photo annotation standard as a patchwork of too many different specifications is not helpful. The opposite Figure illustrates some of the different actitivities, as described aboce in the scenario, that people do with their photos and what different standalone or web-based tools they use for this.

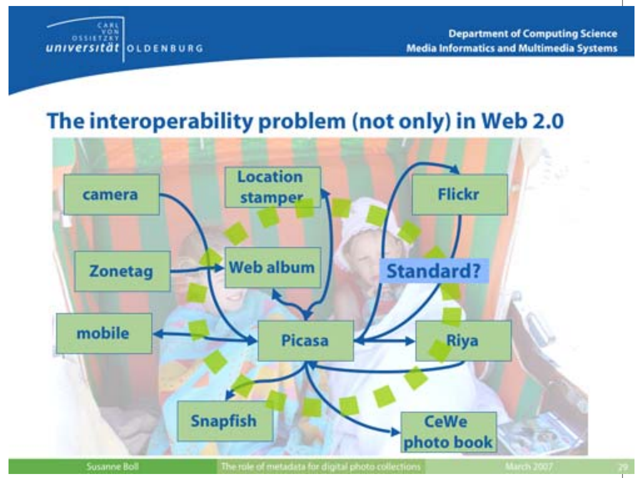
What is missing, however, for content management, search, retrieval, sharing and innovative semantic (Web 2.0) applications is a limited and simple but at the same time comprehensive vocabulary in a machine-readable, exchangeable, but not over complicated representation is needed. However, the single standards described only solve part of the problem. For example, a standardization of tags is very helpful for a semantic search on photos in the Web. However, today the low(er) level features are also lost. Even though the semantic search is fine on a search level, for a later use and exploitation of a set of photos, previously extracted and annotated lower-level features might be interesting as well. Maybe a Web site would like to offer a grouping of photos along the color distribution. Then either the site needs to do the extraction of a color histogram or the photo itself brings this information already in in its standardized header information. A face detection software might have found the bounding boxes on the photo where a face has been detected and also provide a face count. Then the Web site might allow to search for photos with two or more persons on it. And so one. Even though low level features do not seem relevant at first sight, for a detailed search, visualization and also later processing the previously extracted metadata should be stored and available with the photo.

2.2 Use Case: Music

Introduction

In recent years the typical music consumption behaviour has changed dramatically. Personal music collections have grown favoured by technological improvements in networks, storage, portability of devices and Internet services. The amount and availability of songs has de-emphasized its value: it is usually the case that users own many digital music files that they have only listened to once or even never. It seems reasonable to think that by providing listeners with efficient ways to create a personalized order on their collections, and by providing ways to explore hidden "treasures" inside them, the value of their collection will drastically increase.

Also, notwithstanding the digital revolution had many advantages, we can point out some negative effects. Users own huge music collections that need proper storage and labelling. Searching inside digital collections arise new methods for accessing and retrieving data. But, sometimes there is no metadata -or only the file name- that informs about the content of the audio, and that is not enough for an effective utilization and navigation of the music collection.



Toward a solution for photo metadata interoperability.
Image courtesy of Susanne Boll, used with permission.

Thus, users can get lost searching into the digital pile of his music collection. Yet, nowadays, the web is increasingly becoming the primary source of music titles in digital form. With millions of tracks available from thousands of websites, finding the right songs, and being informed of newly music releases is becoming a problematic task. Thus, web page filtering has become necessary for most web users.

Beside, on the digital music distribution front, there is a need to find ways of improving music retrieval effectiveness. Artist, title, and genre keywords might not be the only criteria to help music consumers finding music they like. This is currently mainly achieved using cultural or editorial metadata ("artist A is somehow related with artist B") or exploiting existing purchasing behaviour data ("since you bought this artist, you might also want to buy this one"). A largely unexplored (and potentially interesting) complement is using semantic descriptors automatically extracted from the music audio files. These descriptors can be applied, for example, to recommend new music, or generate personalized playlists.

A complete description of a popular song

In [Pachet], Pachet classifies the music knowledge management. This classification allows to create meaningful descriptions of music, and to exploit these descriptions to build music related systems. The three categories that Pachet defines are: editorial (EM), cultural (CM) and acoustic metadata (AM).

Editorial metadata includes simple creation and production information (e.g. the song C'mon Billy, written by P.J. Harvey in 1995, was produced by John Parish and Flood, and the song appears as the track number 4, on the album "To bring you my love"). EM includes, in addition, artist biography, album reviews, genre information, relationships among artists, etc. As it can be seen, editorial information is not necessarily objective. It is usual the case that different experts cannot agree in assigning a concrete genre to a song or to an artist. Even more diffcult is a common consensus of a taxonomy of musical genres.

Cultural metadata is defined as the information that is implicitly present in huge amounts of data. This data is gathered from weblogs, forums, music radio programs, or even from web search engines' results. This information has a clear subjective component as it is based on personal opinions.

The last category of music information is acoustic metadata. In this context, acoustic metadata describes the content analysis of an audio file. It is intended to be objective information. Most of the current music content processing systems operating on complex audio signals are mainly based on computing low-level signal features. These features are good at characterising the acoustic properties of the signal, returning a description that can be associated to texture, or at best, to the rhythmical attributes of the signal. Alternatively, a more general approach proposes that music content can be successfully characterized according to several "musical facets" (i.e. rhythm, harmony, melody, timbre, structure) by incorporating higher-level semantic descriptors to a given feature set. Semantic descriptors are predicates that can be computed directly from the audio signal, by means of the combination of signal processing, machine learning techniques, and musical knowledge.

Semantic Web languages allow to describe all this metadata, as well as integrating it from different music repositories.

The following example shows an RDF description of an artist, and a song by the artist:

```
<rdf:Description rdf:about="http://www.garageband.com/artist/randycoleman">
  <rdf:type rdf:resource="&music;Artist"/>
  <music:name>Randy Coleman</music:name>
  <music:decade>1990</music:decade>
  <music:decade>2000</music:decade>
  <music:genre>Pop</music:genre>
  <music:city>Los Angeles</music:city>
  <music:nationality>US</music:nationality>
  <geo:Point>
    <geo:lat>34.052</geo:lat>
    <geo:long>-118.243</geo:long>
  </geo:Point>
  <music:influencedBy rdf:resource="http://www.coldplay.com"/>
  <music:influencedBy rdf:resource="http://www.jeffbuckley.com"/>
  <music:influencedBy rdf:resource="http://www.radiohead.com"/>
</rdf:Description>

<rdf:Description rdf:about="http://www.garageband.com/song?|pe1|S8LTM0LdsaSkaFeyYG0">
  <rdf:type rdf:resource="&music;Track"/>
  <music:title>Last Salutation</music:title>
  <music:playedBy rdf:resource="http://www.garageband.com/artist/randycoleman"/>
  <music:duration>T00:04:27</music:duration>
  <music:key>D</music:key>
  <music:keyMode>Major</music:keyMode>
  <music:tonalness>0.84</music:tonalness>
  <music:tempo>72</music:tempo>
```

```
</rdf:Description
```

Lyrics as metadata

For a complete description of a song, lyrics must be considered as well. While lyrics could in a sense be regarded as "acoustic metadata", they are per se actual information entities which have themselves annotation needs. Lyrics share many similarities with metadata, e.g. they usually refer directly to well specified song, but acceptions exists as different artist might sing the same lyrics sometimes even with different musical bases and styles. Most notably, lyrics have often different authors than the music and voice that interprets them and might be composed at a different time. Lyrics are not a simple text; they often have a structure which is similar to that of the song (e.g. a chorus) so they justify the use use of a markup language with a well specified semantics. Unlike the previous types of metadata, however, they are not well suited to be expressed using the W3C Semantic Web initiative languages, e.g. in RDF. While RDF has been suggested instead of XML for for representig texts in situation where advanced and multilayered markup is wanted [Ref RDFTEI], music lyrics markup needs usually limit themselves to indicating particular sections of the songs (e.g. intro, outro, chorus) and possibly the performing character (e.g. in duets). While there is no widespread standard for machine encoded lyrics, some have been proposed [LML][4ML] which in general fit the need for formatting and differentiating main parts. An encoding in RDF of lyrics would be of limited use but still possible with RDF based queries possible just thanks to text search operators in the query language (therfore likely to be limited to "lyrics that contain word X"). More complex queries could be possible if more characters are performing in the lirics and each denoted by an RDF entity which has other metadata attached to it (e.g. the metadata described in the examples above).

It is to be reported however that an RDF encoding would have the disadvantage of complexity. In general it would require a supporting software (for example http://rdftef.sourceforge.net/) to be encoded as XML/RDF can be difficultly written by hand. Also, contrary to an XML based encoding, it could not be easily visualized in a human readable way by, e.g., a simple XSLT transformation.

Both in case of RDF and XML encoding, interesting processing and queries (e.g. conceptual similarities between texts, moods etc) would necessitate advanced textual analysis algorithms well outside the scope or XML or RDF languages. Interestingly however, it might be possible to use RDF description to encode the results of such advanced processings. Keyword extraction algorithms (usually a combination of statistical analysis, stemming and linguistical processing e.g. using wordnet) can be successfully employed on lyrics. The resulting reppresentative "terms" can be encoded as metadata to the lyrics or to the related song itself.

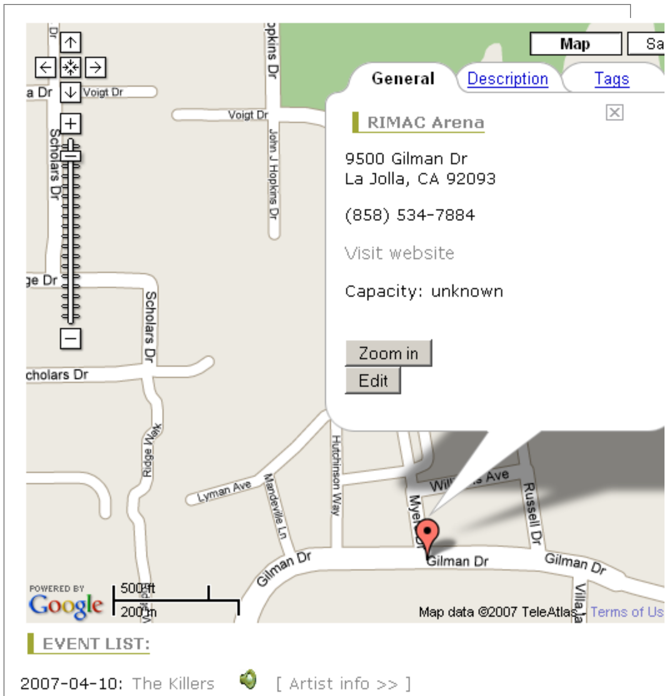
Lower Level Acoustic metadata

"Acoustic metadata" is a broad term which can encompass both features which have an immediate use in higher level use cases (e.g. those presented in the above examples such as tempo, key, keyMode etc) and those that can only be interpreted by data analisys (e.g. a full or simplified representation of the spectrum or the average power sliced every 10 ms). As we have seen, semantic technologies are suitable for reppresenting the higher level acoustic metadata. These are in fact both concise and can be used directly in semantic queries using, e.g., SparQL. Lower level metadata however, e.g. the MPEG7 features extracted by extractors like [Ref MPEG7AUDIODB] is very ill suited to be represented in RDF and is better kept in mpeg-7/xml format for serialization and interchange.

Semantic technologies could be of use in describing such "chunks" of low level metadata, e.g. describing what the content is in terms of describing which features are contained and at which quality. While this would be a duplicaiton of the information encoded in the MPEG-7/XML, it might be of use in semantic queries which select tracks also based on the availability of rich low level metadata.

Motivating Example

Commuting is a big issue in any modern society. Semantically Personalized Playlists might provide both relief and actually benefit in time that cannot be devoted to actively productive activities. Filippo commutes every morning an average of 50+-10 minutes. Before leaving he connects his USB stick/mp3 player to have it "filled" with his morning playlist. The process is completed in 10 seconds, afterall is just 50Mb. he is downloading. During the time of his commute, Filippo will be offered a smooth flow of news, personal daily , entertainment, and cultural snippets from audiobooks and classes.



The next gig.
Image courtesy of Oscar Celma, used with permission.

Musical content comes from Filippo personal music collection or via a content provider (e.g. a low cost thanks to a one time pay license). Further audio content comes from podcasts but also from text to speech reading blog posts, emails, calendar items etc.

Behind the scenes the system works by a combination of semantic queries and ad-hoc algorithms. Semantic queries operate on an RDF database collecting the semantic reppresentation of music metadata (as explained in section 1), as well as annotations on podcasts, news items, audiobooks, and "semantic desktop items" that is represting Filippo's personal desktop information -such as emails and calendar entries.

Ad-hoc algorithms operate on low level metadata to provide smooth transition among tracks. Algorithms for text analysis provide further links among songs and links within songs, pieces of news, emails etc.

At a higher level, a global optimization algorithm takes care of the final playlist creation. This is done by balancing the need for having high priority items played first (e.g. emails from addresss considered important) with the overall goal of providing a smooth and entertaining experience (e.g. interleaving news with music etc).

Semantics can help in providing "related information or content" which can be put adjacent to the actual core content. This can be done in relative freedom since the content can be at any time skipped by the user using simply the forward button.

Upcoming concerts

John has been listening to the "Snow Patrol" band for a while. He discovered the band while listening to one of his favorite podcasts about alternative music. He has to travel to San Diego next week, and he is finding upcoming concerts that he would enjoy there, and he asks his personalized semantic web music service to provide him with some recommendations of upcoming gigs in the area, and decent bars to have a beer.

```
<!-- San Diego geolocation -->
<foaf:based_near geo:lat='32.715' geo:long='-117.156'/>
```

The system is tracking user listening habits, so it detects than one song from "The Killers" band (scrapped from their website) sounds similar to the last song John has listened to from "Snow Patrol". Moreover, both bands have similar styles, and there are some podcasts that contain songs from both bands in the same session. Interestingly enough, the system knows that the Killers are playing close to San Diego next weekend, thus it recommends to John to assist to that gig.

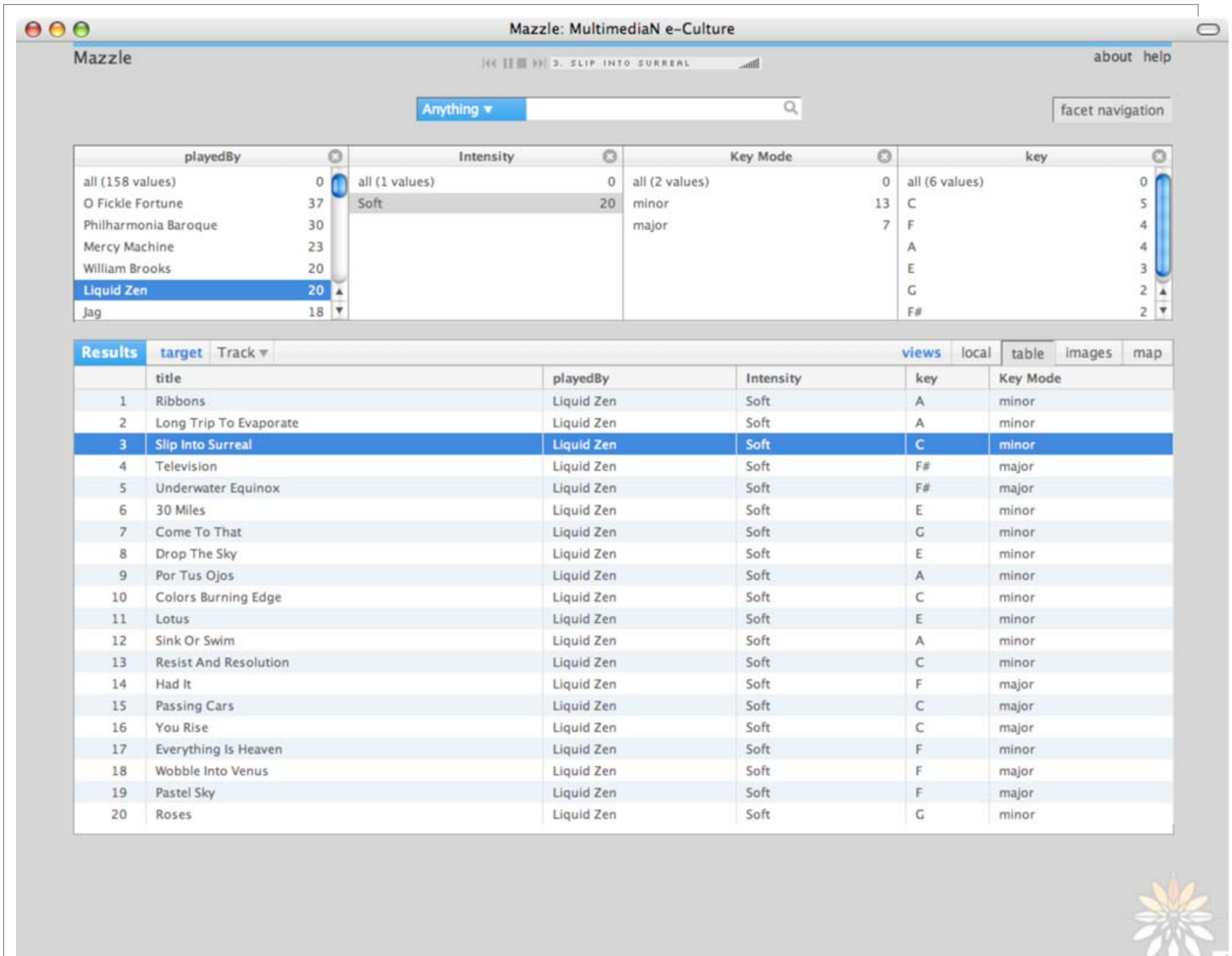
Facet browsing of Music Collections

Michael has a brand new (last generation-posh) iPod. He is looking for some music using the classic hierarchical navigation (Genre->Artist->Album->Songs). But the main problem is that he is not able to find a decent list of songs (from his 100K music collection) to move into his iPod. On the other hand, facet browsing has recently become popular as a user friendly interface to data repositories.

/facet system [Hildebrand] presents a new and intuitive way to navigate large collections, using several facets or aspects, of multimedia assets. /facet extends browsing of Semantic Web data in four ways. First, users are able to select and navigate through facets of resources of any type and to make selections based on properties of other, semantically related, types. Second, it addresses a disadvantage of hierarchy-based navigation by adding a keyword search interface that dynamically makes semantically relevant suggestions. Third, the /facet interface, allows the inclusion of facet-specific display options that go beyond the hierarchical navigation that characterizes current facet browsing. Fourth, the browser works on any RDF dataset without any additional configuration.

Thus, based on a RDF description of music titles, the user can navigate through music facets, such as Rhythm (beats per minute), Tonality (Key and mode), Intensity of the piece (moderate, energetic, etc.)

A fully functional example can be seen at http://slashfacet.semanticweb.org/music/mazzle



The Mazzle Interface.
Image courtesy of Michiel Hildebrand, used with permission.

Music Metadata on the Semantic Web

Nowadays, in the context of the World Wide Web, the increasing amount of available music makes very difficult, to the user, to find music he/she would like to listen to. To overcome this problem, there are some audio search engines that can fit the user's needs (for example: http://search.singingfish.com/, http://audio.search.yahoo.com/, http://www.audiocrawler.com/, http://www.alltheweb.com/?cat=mp3, http://www.searchsounds.net and http://www.altavista.com/audio/).

Some of the current existing search engines are nevertheless not fully exploited because their companies would have to deal with copyright infringing material. Music search engines have a crucial component: an audio crawler, that scans the web and gathers related information about audio files.

Moreover, describing music it not an easy task. As presented in section 1, music metadata copes with several categories (editorial, acoustic, and cultural). Yet, none of the audio metadata used in practice (e.g ID3, OGG Vorbis, etc.) can fully describe all these facets. Actually, metadata for describing music are mostly tags implemented in the Key-Value form [TAG]=[VALUE], for instance, "ARTIST=The Killers".

The following section introduces, then, the mappings between current audio vocabularies within the Semantic Web technologies. This will allow to extend the description of a piece of music, as well as adding explicit semantics.

Integrating Various Vocabularies Using RDF

In this section we present a way to integrate several audio vocabularies into a single one, based on RDF. For more details about the audio vocabularies, the reader is refered to Vocabularies - Audio Content Section, and Vocabularies - Audio Ontologies Section.

This section will focus on the ID3 and OGG Vorbis metadata initiatives, as they are the most used ones. Though, both vocabularies cope only editorial data. Moreover, a first mapping with the Music Ontology is presented, too.

ID3 is a metadata container most often used in conjunction with the MP3 audio file format. It allows information such as the title, artist, album, track number, or other information about the file to be stored in the file itself (from Wikipedia).

The most important metadata descriptors are:

- Artist name <=> `<foaf:name></foaf:name>`
- Album name <=> `<mo:Record><dc:title>Album name</dc:title></mo:Record>`
- Song title <=> `<mo:Track><dc:title>Album name</dc:title></mo:Track>`
- Year
- Track number <=> `<mo:trackNum>Track number</mo:trackNum></mo:Track>`
- Genre (from a predefined list of more than 100 genres) <=> `<mo:Genre>Genre name</mo:Genre>`

OGG Vorbis metadata, called comments, support metadata 'tags' similar to those implemented in the ID3. The metadata is stored in a vector of strings, encoded in UTF-8

- TITLE <=> `<mo:Track><dc:title>Album name</dc:title></mo:Track>`
- VERSION: The version field may be used to differentiate multiple versions of the same track title
- ALBUM <=> `<mo:Record><dc:title>Album name</dc:title></mo:Record>`
- TRACKNUMBER <=> `<mo:trackNum>Track number</mo:trackNum></mo:Track>`
- ARTIST <=> `<foaf:name></foaf:name>`
- PERFORMER <=> `<foaf:name></foaf:name>` ???
- COPYRIGHT: Copyright attribution
- LICENSE: License information, eg, 'All Rights Reserved', 'Any Use Permitted', a URL to a license such as a Creative Commons license
- ORGANIZATION: Name of the organization producing the track (i.e â€˜a record labelâ€™)
- DESCRIPTION: A short text description of the contents
- GENRE <=> `<mo:Genre>Genre name</mo:Genre>`
- DATE
- LOCATION: Location where the track was recorded

RDFizing songs

We present a way to RDFize tracks based on the Music Ontology.

Example: Search a song into MusicBrainz and RDFize results. This first example shows how to query the MusicBrainz music repository, and RDFize the results based on the Music Ontology. Try a complete example at http://foafing-the-music.iua.upf.edu/RDFize/track?artist=U2&title=The+fly. The parameters are song title (The Fly) and artist name (U2).

```
<mo:Track rdf:about='http://musicbrainz.org/track/dddb2236-823d-4c13-a560-bfe0ffbb19fc'>
  <mo:puid rdf:resource='2285a2f8-858d-0d06-f982-3796d62284d4'/>
  <mo:puid rdf:resource='2b04db54-0416-d154-4e27-074e8dcea57c'/>
  <dc:title>The Fly</dc:title>
  <dc:creator>
    <mo:MusicGroup rdf:about='http://musicbrainz.org/artist/a3cb23fc-acd3-4ce0-8f36-1e5aa6a18432'>
      <foaf:img rdf:resource='http://ec1.images-amazon.com/images/P/B000001FS3.01._SCMZZZZZZZ_.jpg'/>
      <mo:musicmoz rdf:resource='http://musicmoz.org/Bands_and_Artists/U/U2/'/>
      <foaf:name>U2</foaf:name>
      <mo:discogs rdf:resource='http://www.discogs.com/artist/U2'/>
      <foaf:homepage rdf:resource='http://www.u2.com/'/>
      <foaf:member rdf:resource='http://musicbrainz.org/artist/0ce1a4c2-ad1e-40d0-80da-d3396bc6518a'/>
      <foaf:member rdf:resource='http://musicbrainz.org/artist/1f52af22-0207-40ac-9a15-e5052bb670c2'/>
      <foaf:member rdf:resource='http://musicbrainz.org/artist/a94e530f-4e9f-40e6-b44b-ebec06f7900e'/>
      <foaf:member rdf:resource='http://musicbrainz.org/artist/7f347782-eb14-40c3-98e2-17b6e1bfe56c'/>
      <mo:wikipedia rdf:resource='http://en.wikipedia.org/wiki/U2_%28band%29'/>
    </mo:MusicGroup>
  </dc:creator>
</mo:Track>
```

Example: The parameter is a URL that contains an MP3 file. In this case it reads the ID3 tags from the MP3 file. See an output example at http://foafing-the-music.iua.upf.edu/RDFize/track?url=http://www.archive.org/download/bt2002-11-21.shnf/bt2002-11-21d1.shnf

/bt2002-11-21d1t01_64kb.mp3 (it might take a little while).

```
 <mo:Track rdf:about='http://musicbrainz.org/track/7201c2ab-e368-4bd3-934f-5d936efffcdc'>
  <dc:creator>
    <mo:MusicGroup rdf:about='http://musicbrainz.org/artist/6b28ecf0-94e6-48bb-aa2a-5ede325b675b'>
      <foaf:name>Blues Traveler</foaf:name>
      <mo:discogs rdf:resource='http://www.discogs.com/artist/Blues+Traveler'/>
      <foaf:homepage rdf:resource='http://www.bluestraveler.com/'/>
      <foaf:member rdf:resource='http://musicbrainz.org/artist/d73c9a5d-5d7d-47ec-b15a-a924a1a271c4'/>
      <mo:wikipedia rdf:resource='http://en.wikipedia.org/wiki/Blues_Traveler'/>
      <foaf:img rdf:resource='http://ec1.images-amazon.com/images/P/B000078JKC.01._SCMZZZZZZZ_.jpg'/>
    </mo:MusicGroup>
  </dc:creator>
  <dc:title>Back in the Day</dc:title>
  <mo:puid rdf:resource='0a57a829-9d3c-eb35-37a8-d0364d1eae3a'/>
  <mo:puid rdf:resource='02039e1b-64bd-6862-2d27-3507726a8268'/>
</mo:Track>
```

Example: Once the songs have been RDFized, we can ask last.fm for the latest tracks a user has been listening to, and then RDFize them. http://foafing-the-music.iua.upf.edu/draft/RDFize/examples/lastfm_tracks.rdf is an example that shows the latest tracks a user (RJ) has been listening to. You can try it at http://foafing-the-music.iua.upf.edu/RDFize/lastfm_tracks?username=RJ

2.3 Use Case: News

Introduction

More and more news is produced and consumed each day. News generally consists of mainly textual stories, which are more and more often illustrated with graphics, images and videos. News can be further processed by professional (newspapers), directly accessible for web users through news agencies, or automatically aggregated on the web, generally by search engine portal and not without copyright problems.

For easing the exchange of news, the International Press Telecommunication Council (IPTC) is currently developping the NewsML G2 Architecture (NAR) whose goal is *to provide a single generic model for exchanging all kinds of newsworthy information, thus providing a framework for a future family of IPTC news exchange standards* [NewsML-G2]. This family includes NewsML, SportsML, EventsML, ProgramGuideML and a future WeatherML. All are XML-based languages used for describing not only the news content (traditional metadata), but also their management and packaging, or related to the exchange itself (transportation, routing).

However, despite this general framework, interoperability problems can occur. News is about the world, so its metadata might use specific controlled vocabularies. For example, IPTC itself is developing the IPTC News Codes [NewsCodes] that currently contain 28 sets of controlled terms. These terms will be the values of the metadata in the NewsML G2 Architecture. The news descriptions often refer to other thesaurus and controlled vocabularies, that might come from the industry (for example, XBRL [XBRL] in the financial domain), and all are represented using different formats. From the media point of view, the pictures taken by the journalist come with their EXIF metadata [EXIF]. Some videos might be described using the EBU format [EBU] or even with MPEG-7 [MPEG-7].

We illustrate these interoperability issues between domain vocabularies and other multimedia standards in the financial news domain. For example, the Reuters Newswires and the Dow Jones Newswires provide categorical metadata associated with news feeds. The particular vocabularies of category codes, however, have been developed independently, leading to clear interoperability issues. The general goal is to improve the search and the presentation of news content in such an heterogeneous environment. We provide a motivating example that highlight the issues discussed above and we present a potential solution to this problem, which leverages Semantic Web technologies.

Motivating Example

XBRL (Extended Business Reporting Language) [XBRL] is a standardized way of enconding financial information of companies, and about the management structure, location, number of employes, etc. of such entities. XBRL is basically about "quantitative" information in the financial domain, and is based on the periodic reports generated by the companies. But for many Business Intelligence applications, there is also a need to consider "qualitative" information, which is mostly delivered by news articles. The problem is therefore how to optimally integrate information from the periodic reports and the day to day information provided by specialized news agencies. Our goal is to provide a platform that allows more semantics in automated ranking of creditworthiness of companies. The financial news are playing an important role since they provide "qualitative" information on companies, branches, trends, countries, regions etc.

There are quite a few news feeds services within the financial domain, including the Dow Jones Newswire and Reuters. Both Reuters and Dow Jones provides an XML based representation and have associated with each article metadata with date, time, headline, full story, company ticker symbol, and category codes.

Example 1: NewsML 1 Format

We consider the news feeds similar to that published by Reuters, where along with the text of the article, there is associated metadata in the form of XML tags. The terms in these tags are associated with a controlled vocabulary developed by Reuters and other industry bodies. Below is a sample news article formatted in NewsML 1, which is similar to the structural format used by Reuters. For exposition, the metadata tags associated with the article are aligned with those used by Reurters.

```
<?xml version="1.0" encoding="UTF-8"?>
<NewsML Duid="MTFH93022_2006-12-14_23-16-17_NewsML">
  <Catalog Href="..."/>
  <NewsEnvelope>
    <DateAndTime>20061214T231617+0000</DateAndTime>
    <NewsService FormalName="..."/>
    <NewsProduct FormalName="TXT"/>
    <Priority FormalName="3"/>
  </NewsEnvelope>
  <NewsItem Duid="MTFH93022_2006-12-14_23-16-17_NEWSITEM">
    <Identification>
      <NewsIdentifier>
        <ProviderId>...</ProviderId>
        <DateId>20061214</DateId>
```

```
        <NewsItemId>MTFH93022_2006-12-14_23-16-17</NewsItemId>
        <RevisionId Update="N" PreviousRevision="0">1</RevisionId>
        <PublicIdentifier>...</PublicIdentifier>
      </NewsIdentifier>
      <DateLabel>2006-12-14 23:16:17 GMT</DateLabel>
    </Identification>
    <NewsManagement>
      <NewsItemType FormalName="News"/>
      <FirstCreated>...</FirstCreated>
      <ThisRevisionCreated>...</ThisRevisionCreated>
      <Status FormalName="Usable"/>
      <Urgency FormalName="3"/>
    </NewsManagement>
    <NewsComponent EquivalentsList="no" Essential="no" Duid="MTFH92062_2002-09-23_09-29-03_T88093_MAIN_NC" xml:lang="en">
      <TopicSet FormalName="HighImportance">
        <Topic Duid="t1">
          <TopicType FormalName="CategoryCode"/>
          <FormalName Scheme="MediaCategory">OEC</FormalName>
          <Description xml:lang="en">Economic news, EC, business/financial pages</Description>
        </Topic>
        <Topic Duid="t2">
          <TopicType FormalName="Geography"/>
          <FormalName Scheme="N2000">DE</FormalName>
          <Description xml:lang="en">Germany</Description>
        </Topic>
      </TopicSet>
      <Role FormalName="Main"/>
      <AdministrativeMetadata>
        <FileName>MTFH93022_2006-12-14_23-16-17.XML</FileName>
        <Provider>
          <Party FormalName="..."/>
        </Provider>
        <Source>
          <Party FormalName="..."/>
        </Source>
        <Property FormalName="SourceFeed" Value="IDS"/>
        <Property FormalName="IDSPublisher" Value="..."/>
      </AdministrativeMetadata>
      <NewsComponent EquivalentsList="no" Essential="no" Duid="MTFH93022_2006-12-14_23-16-17" xml:lang="en">
        <Role FormalName="Main Text"/>
        <NewsLines>
          <HeadLine>Insurances get support</HeadLine>
          <ByLine/>
          <DateLine>December 14, 2006</DateLine>
          <CreditLine>...</CreditLine>
          <CopyrightLine>...</CopyrightLine>
          <SlugLine>...</SlugLine>
          <NewsLine>
            <NewsLineType FormalName="Caption"/>
            <NewsLineText>Insurances get support</NewsLineText>
          </NewsLine>
        </NewsLines>
        <DescriptiveMetadata>
          <Language FormalName="en"/>
          <TopicOccurrence Importance="High" Topic="#t1"/>
          <TopicOccurrence Importance="High" Topic="#t2"/>
        </DescriptiveMetadata>
        <ContentItem Duid="MTFH93022_2006-12-14_23-16-17">
          <MediaType FormalName="Text"/>
          <Format FormalName="XHTML"/>
          <Characteristics>
            <Property FormalName="ContentID" Value="urn:...20061214:MTFH93022_2006-12-14_23-16-17_T88093_TXT:1"/>
            ...
          </Characteristics>
          <DataContent>
            <html xmlns="http://www.w3.org/1999/xhtml">
              <head>
                <title>Insurances get support</title>
              </head>
              <body>
                <h1>The Senate of Germany wants to constraint the participation of clients to the hidden reserves</h1>
                <p>
                 DÃœSSELDORF The German Senate supports the point of view of insurance companies in a central point of the new law
                 defining insurance contracts, foreseen for 2008. In a statement, the Senators show disagreements with the proposal
                 of  the Federal Government, who was in favor of including investment bonds in the hidden reserves, which in the
                 next future should be accessible to the clients of the insurance companies.
                 ...
                </p>
              </body>
            </html>
          </DataContent>
        </ContentItem>
      </NewsComponent>
    </NewsComponent>
  </NewsItem>
</NewsML>
```

Example 2: NewsML G2 Format

If we consider the same data, but expressed in NewsML G2:

```
<?xml version="1.0" encoding="UTF-8"?>
<newsMessage xmlns="http://iptc.org/std/newsml/2006-05-01/" xmlns:xhtml="http://www.w3.org/1999/xhtml">
  <header>
    <date>2006-12-14T23:16:17Z</date>
    <transmitId>696</transmitId>
    <priority>3</priority>
```

```
    <channel>ANA</channel>
  </header>
  <itemSet>
    <newsItem guid="urn:newsml:afp.com:20060720:TX-SGE-SNK66" schema="0.7" version="1">
      <catalogRef href="http://www.afp.com/newsml2/catalog-2006-01-01.xml"/>
      <itemMeta>
        <contentClass code="ccls:text"/>
        <provider literal="Handelsblatt"/>
        <itemCreated>2006-07-20T23:16:17Z</itemCreated>
        <pubStatus code="stat:usable"/>
        <service code="srv:Archives"/>
      </itemMeta>
      <contentMeta>
        <contentCreated>2006-07-20T23:16:17Z</contentCreated>
        <creator/>
        <language literal="en"/>
        <subject code="cat:04006002" type="ctyp:category"/>        #cat:04006002= banking
        <subject code="cat:04006006" type="ctyp:category"/>        #cat:04006006= insurance
        <slugline separator="-">Insurances get support</slugline>
        <headline>The Senate of Germany wants to constraint the participation of clients to the hidden reserves</headline>
      </contentMeta>
      <contentSet>
        <inlineXML type="text/plain">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>Insurances get support</title>
  </head>
  <body>
    <h1>The Senate of Germany wants to constraint the participation of clients to the hidden reserves</h1>
    <p>
    DÃœSSELDORF The German Senate supports the point of view of insurance companies in a central point of the new law defining
    insurance contracts, foreseen for 2008. In a statement, the Senators show disagreements with the proposal of the Federal
    Government, who was in favor of including investment bonds in the hidden reserves, which in the next future should be accessible
    to the clients of the insurance companies.
    ...
    </p>
  </body>
</html>
        </inlineXML>
      </contentSet>
    </newsItem>
  </itemSet>
</newsMessage>
```

Example 3: German Broadcaster Format

The terms in the tags displayed just above are associated with a controlled vocabulary developed by Reuters. If we consider the internal XML encoding that has been proposed provisionally by a running European project (the MUSING project) for the encoding of similar articles in German Newspapers (mapping the HTML tags of the online articles into XML and adding others), we have the following:

```
  <ID>1091484</ID>                # Internal encoding
  <SOURCE>Handelsblatt</SOURCE>   # Name of the newspaper we get the information from
  <DATE>14.12.2006</DATE>         # Date of publication
  <NUMBER>242</NUMBER>            # Numbering of the publication
  <PAGE>27</PAGE>                 # Page number in the publication
  <LENGTH>111</LENGTH>            # The number of lines in the main article

  <ACTIVITY_FIELD>Banking_Insurance</ACTIVITY_FIELD>   # corresponding to the financial domain reported in the article
  <TITLE>Insurances get support</TITLE>
  <SUBTITLE>The Senate of Germany wants to constraint the participation of clients to the hidden reserves</SUBTITLE>
  <ABSTRACT></ABSTRACT>
  <AUTHORS>Lansch, Rita</AUTHORS>
  <LOCATION>Federal Republic of Germany</LOCATION>
  <KEYWORDS>Bank supervision, Money and Stock exchange, Bank</KEYWORDS>
  <PROPERNAMES>Meister, Edgar Remsperger, Hermann Reckers, Hans Fabritius, Hans Georg Zeitler, Franz-Christoph</PROPERNAMES>
  <ORGANISATIONS>Bundesanstalt fÃ¼r Finanzdienstleistungsaufsicht BAFin</ORGANISATIONS>
  <TEXT>DÃœSSELDORF The German Senate supports the point of view of insurance companies in a central point of the new law
  defining insurance contracts, foreseen for 2008. In a statement, the Senators show disagreements with the proposal of the
  Federal Government, who was in favor of including investment bonds in the hidden reserves, which in the next future should
  be accessible to the clients of the insurance companies....</TEXT>
```

Example 4: XBRL Format

Structured data and documents such as Profit & Loss tables can finally be mapped onto existing taxonomies, like XBRL, which is an emerging standard for Business Reporting.

XBRL definition in Wikipedia: "XBRL is an emerging XML-based standard to define and exchange business and financial performance information. The standard is governed by a not-for-profit international consortium XBRL International Incorporated of approximately 450 organizations, including regulators, government agencies, infomediaries and software vendors. XBRL is a standard way to communicate business and financial performance data. These communications are defined by metadata set in taxonomies. Taxonomies capture the definition of individual reporting elements as well as the relationships between elements within a taxonomy and in other taxonomies.

The relations between elements supported, for the time being, (at least for the German Accounting Principles expressed in the corresponding XBRL taxonomy, see http://www.xbrl.de/) are:

- child-parent
- parent-child
- dimension-element
- element-dimension

In fact the child-parent/parent-child relation haves to be understood as part-of relations within finanical reporting documents rather than as sub-class relations, as we noticed in an attempt to formlize XBRL in OWL, in the context of the European MUSING R&D project (http://www.musing.eu/).

The table below shows how a balance sheet looks like:

structured P&L	2002 EUR	2002 EUR	2002 EUR
Sales	850.000,00	800.000,00	300.000,00
Changes in stock	171.000,00	104.000,00	83.000,00
Own work capitalized	0,00	0,00	0,00
Total output	1.021.000,00	904.000,00	383.000,00
...			
Net income/net loss for the year	139.000,00	180.000,00	-154.000,00
	2002	**2001**	**2000**
Number of Employees	27	25	23
....			

There is a lot of variations in both the way the information can be displayed (number of columns, use of fonts, etc.) but also in the terminology used: the financial terms in the leftmost column are not normalized at all. Also the figures are not normalized (clearly, the company has more than just "27" employees, but it is not indicated in the table if we deal with 27000 employess). This makes this kind of information unable to be used by semantic applications. XBRL is a very important step in the normalization of such data, as can be seen in the following example displaying the XBRL encoding of the kind of data that was presented just above in the table:

```
<group xsi:schemaLocation="http://www.xbrl.org/german/ap/ci/2002-02-15 german_ap.xsd">
   <numericContext id="c0" precision="8" cwa="false">
      <entity>
         <identifier scheme="urn:datev:www.datev.de/zmsd">11115,129472/12346</identifier>
      </entity>
      <period>
         <startDate>2002-01-01</startDate>
         <endDate>2002-12-31</endDate>
      </period>
      <unit>
         <measure>ISO4217:EUR</measure>
      </unit>
   </numericContext>
   <numericContext id="c1" precision="8" cwa="false">
      <entity>
         <identifier scheme="urn:datev:www.datev.de/zmsd">11115,129472/12346</identifier>
      </entity>
      <period>
         <startDate>2001-01-01</startDate>
         <endDate>2001-12-31</endDate>
      </period>
      <unit>
         <measure>ISO4217:EUR</measure>
      </unit>
   </numericContext>
   <numericContext id="c2" precision="8" cwa="false">
      <entity>
         <identifier scheme="urn:datev:www.datev.de/zmsd">11115,129472/12346</identifier>
      </entity>
      <period>
         <startDate>2000-01-01</startDate>
         <endDate>2000-12-31</endDate>
      </period>
      <unit>
         <measure>ISO4217:EUR</measure>
      </unit>
   </numericContext>
   <t:bs.ass numericContext="c2">1954000</t:bs.ass>
   <t:bs.ass.accountingConvenience numericContext="c0">40000</t:bs.ass.accountingConvenience>
   <t:bs.ass.accountingConvenience numericContext="c1">70000</t:bs.ass.accountingConvenience>
   <t:bs.ass.accountingConvenience numericContext="c2">0</t:bs.ass.accountingConvenience>
   <t:bs.ass.accountingConvenience.changeDem2Eur numericContext="c0">0</t:bs.ass.accountingConvenience.changeDem2Eur>
   <t:bs.ass.accountingConvenience.changeDem2Eur numericContext="c1">20000</t:bs.ass.accountingConvenience.changeDem2Eur>
   <t:bs.ass.accountingConvenience.changeDem2Eur numericContext="c2">0</t:bs.ass.accountingConvenience.changeDem2Eur>
   <t:bs.ass.accountingConvenience.startUpCost numericContext="c0">40000</t:bs.ass.accountingConvenience.startUpCost>
   <t:bs.ass.accountingConvenience.startUpCost numericContext="c1">50000</t:bs.ass.accountingConvenience.startUpCost>
   <t:bs.ass.accountingConvenience.startUpCost numericContext="c2">0</t:bs.ass.accountingConvenience.startUpCost>
   <t:bs.ass.currAss numericContext="c0">571500</t:bs.ass.currAss>
   <t:bs.ass.currAss numericContext="c1">558000</t:bs.ass.currAss>
   <t:bs.ass.currAss numericContext="c2">394000</t:bs.ass.currAss>
</group>
```

In the XBRL example shown just above, one can see the normalization of the periods for which the reporting is valid, and for the currency used in the report. The annotation of the financial values of the financial items is then proposed on the base of a XBRL tag (language independent) in the context of the uniquely identified period (the "c0", "c1" etc), and with the encoded currency.

The XBRL representation is marking a real progress compared to the "classical" way of displaying financial information. And as such XBRL allows for some semantics, describing for example various types of relations. The need for more semantics is mainly driven by applications requiring merging of the quantitative information encoded in XBRL with other kind of information, which is crucial in Business Intelligence scenarios, for example merging balance sheet information with information coming from newswires or with information in related domain, like politics. Therefore some initiatives started looking at representing information encoded in XBRL within OWL, as the basic ontology language representation in the Semantic Web community [Declerck], [Lara].

Potential Solution: Converting Various Vocabularies into RDF

In this section, we discuss a potential solution to the problems highlighted in this document. We propose utilizing Semantic Web technologies for the purpose of aligning these standards and controlled vocabularies. Specifically, we discuss adding an RDF/OWL layer on top of these standards and vocabularies for the purpose of data integration and reuse. The following sections discuss this approach in more detail.

XBRL in the Semantic Web

We sketch how we convert XBRL to OWL. The XBRL OWL base taxonomy was manually developed using the OWL plugin of the Protege knowledge base editor [Knublauch]. The version of XBRL we used together with the Accounting Principles for German consists of 2,414 concepts, 34 properties, and 4,780 instances. Overall, this translates into 24,395 unique RDF triples. The basic idea during our export was that even though we are developing an XBRL taxonomy in OWL using Protege, the information that is stored on disk is still RDF on the syntactic level. We were thus interested in RDF data base systems which make sense of the semantics of OWL and RDFS constructs such as rdfs:subClassOf or owl:equivalentClass. We have been experimenting with the Sesame open-source middleware framework for storing and retrieving RDF data [Broekstra].

Sesame partially supports the semantics of RDFS and OWL constructs via entailment rules that compute "missing" RDF triples (the deductive closure) in a forward-chaining style at compile time. Since sets of RDF statements represent RDF graphs, querying information in an RDF framework means to specify path expressions. Sesame comes with a very powerful query language, SeRQL, which includes (i) generalised path expressions, (ii) a restricted form of disjunction through optional matching, (iii) existential quantifiation over predicates, and (iv) Boolean constraints. From an RDF point of view, additional 62,598 triples were generated through Sesame's (incomplete) forward chaining inference mechanism.

For proof of concept, we looked at the freely available financial reporting taxonomies(http://www.xbrl.org/FRTaxonomies/) and took the final German AP Commercial and Industrial (German Accounting Principles) taxonomy (February 15, 2002; http://www.xbrl-deutschland.de/xe news2.htm), acknowledged by XBRL International. The taxonomy can be obtained as a packed zip file from http://www.xbrl-deutschland.de/germanap.zip.

xbrl-instance.xsd specifies the XBRL base taxonomy using XML Schema. The file makes use of XML schema datatypes, such as xsd:string or xsd:date, but also defines simple types (simpleType), complex types (complexType), elements (element), and attributes (attribute). Element and attribute declarations are used to restrict the usage of elements and attributes in XBRL XML documents. Since OWL only knows the distinction between classes and properties, the correpondences between XBRL and OWL description primitives is not a one-to-one mapping:

However, OWL allows to characterize properties more precisely than just having only a domain and a range. We can mark a property as functional (instead of being relational, the default case), meaning that it takes at most one value. This clearly means that a property must not have a value for each instance of a class on which it is defined. Thus a functional property is in fact a partial (and must not necessarily be a total) function. Exactly the distinction functional vs. relational is represented by the attribute vs. element distinction, since multiple elements are allowed within a surrounding context. However, at most one attribute-value combination for each attribute name is allowed within an element:

XBRL	**OWL**
simple type	class
complex type	class
attribute	functional property
element	relational property

Simple and complex types differs from one another in that simple types are essentially defined as extensions of the basic XML Schema datatypes, whereas complex types are XBRL specifications that do not build upon XSD types, but instead introduce their own element and attribute descriptions. Here are simple type specifications found in the base terminology of XBRL, located in the file xbrl-instance.xsd:

Since OWL only claims that "As a minimum, tools must support datatype reasoning for the XML Schema datatypes xsd:string and xsd:integer." [OWL, p. 30] and because "It is not illegal, although not recommended, for applications to define their own datatypes ..." [OWL, p. 29], we have decided to implement a workaround that represents all the necessary XML Schema datatypes used in XBRL. This was done by having a wrapper type for each simple XML Schema type. For instance, "monetary" is a simple subtype of the wrapper type "decimal": `<restriction base="decimal"/>`. Below we show the first lines of the actual OWL version of XBRL we have implemented:

```
<?xml version="1.0"?>
<rdf:RDF xmlns="http://xbrl.dfki.de/main.owl#"
         xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
         xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
         xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
         xmlns:owl="http://www.w3.org/2002/07/owl#"
         xml:base="http://xbrl.dfki.de/main.owl">
  <owl:Ontology rdf:about=""/>
  <owl:Class rdf:ID="bs.ass.fixAss.tan.machinery.installations">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="Locator"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="nt.ass.fixAss.fin.loansToParticip.net.addition">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#Locator"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="nt.ass.fixAss.fin.loansToSharehold.net.beginOfPeriod.endOfPrevPeriod">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#Locator"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="nt.ass.fixAss.fin.gross.revaluation.comment">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#Locator"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="nt.ass.fixAss.fin.securities.gross.beginOfPeriod.otherDiff">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#Locator"/>
    </rdfs:subClassOf>
  </owl:Class>
   ...
</owl:Ontology>
```

```
</rdf:RDF>
```

The German Accounting Principles taxonomy consists of 2,387 concepts, plus 27 concepts from the base taxonomy for XBRL. 34 properties were defined and 4,780 instance fnally generated.

Besides the ontologization of XBRL, we would propose to build an ontology on the top of the taxonomic organization of NACE codes. Then we need a clear ontological representation of the time units/information relevant in the domain. And last but not least, we would also use all the classification/categorization information of NewsML/IPTC to use more accurate semantic Metadata for the encoding of the (financial) news articles.

EXIF in the Semantic Web

One of today's commonly used image format and metadata standard is the Exchangeable Image File Format [EXIF]. This file format provides a standard specification for storing metadata regarding image. Metadata elements pertaining to the image are stored in the image file header and are marked with unique tags, which serves as an element identifying.

As we note in this document, one potentional way to integrate EXIF metadata with additinoal news/multimedia metadata formats is to add an RDF layer on top of the metadata standards. Recently there has been efforts to encode EXIF metadata in such Semantic Web standards, which we briefly detail below. We note that both of these ontologies are semantically very similar, thus this issue is not addressed here. Essentially both are a straightforward encodings of the EXIF metadata tags for images. There are some syntactic differences, but again they are quite similar; they primarily differ in their naming conventions utilized.

The Kanzaki EXIF RDF Schema provides an encoding of the basic EXIF metadata tags in RDFS. Essentially, these are the tags defined from Section 4.6 of [EXIF]. We also note here that relevant domains and ranges are utilized as well. It additionally provides an EXIF conversion service, EXIF-to-RDF, which extracts EXIF metadata from images and automatically maps it to the RDF encoding. In particular the service takes a URL to an EXIF image and extracts the embedded EXIF metadata. The service then converts this metadata to the RDF schema and returns this to the user.

The Norm Walsh EXIF RDF Schema provides another encoding of the basic EXIF metadata tags in RDFS. Again, these are the tags defined from Section 4.6 of [EXIF]. It additionally provides JPEGRDF, which is a Java application that provides an API to read and manipulate EXIF meatadata stored in JPEG images. Currently, JPEGRDF can can extract, query, and augment the EXIF/RDF data stored in the file headers. In particular, we note that the API can be used to convert existing EXIF metadata in file headers to the schema. The resulting RDF can then be stored in the image file header, etc. (Note here that the API's functionality greatly extends that which was briefly presented here).

Putting All That Together

Some text showing how this qualitative and quantitative information benefits to interoperate ...

2.4 Use Case: Tagging

Introduction

Tags are what may be the simplest form of annotation: simple user-provided keywords that are assigned to resources, in order to support subsequent retrieval. In itself, this idea is not particularly new or revolutionary: keyword-based retrieval has been around for a while. In contrast to the formal semantics provided by the Semantic Web standards, tags have no semantic relations whatsoever, including a lack of hierarchy; tags are just flat collections of keywords.
There are however new dimensions that have boosted the popularity of this approach and given a new perspective on an old theme: low-cost applicability and collaborative tagging.

Tagging lowers the barrier of metadata annotation, since it requires minimal effort on behalf of annotators: there are no special tools or complex interface that the user needs to get familiar with, and no deep understanding of logic principles or formal semantics required â€" just some standard technical expertise. Tagging seems to work in a way that is intuitive to most people, as demonstrated by its widespread adoption, as well as by certain studies conducted on the field [Trant]. Thus, it helps bridging the 'semantic gap' between content creators and content consumers, by offering 'alternative points of access' to document collections.

The main idea behind collaborative tagging is simple: collaborative tagging platforms (or, alternatively, distributed classification systems - DCSs [Mejias]) provide the technical means, usually via some sort of web-based interface, that support users in tagging resources. What is the important aspect of this is that they aggregate collections of tags that an individual uses, or his tag vocabulary, called a personomy [Hotho], into what has been termed a folksonomy: a collection of all personomies [Mathes, Smith].

Some of the most popular collaborative tagging systems are Delicious (bookmarks), Flickr (images), Last.fm (music), YouTube (video), Connotea (bibliographic information), steve.museum (museum items) and Technorati (blogging). Using these platforms is free, although in some cases users can opt for more advanced features by getting an upgraded account, for which they have to pay. The most prominent among them are Delicious and Flickr, for which some quantitative user studies are available [HitWise, NetRatings]. These user studies document a phenomenal growth, that indicates that in real-life tagging is a very viable solution for annotating any type of resource.

Motivating Scenario

Let us view some of the current limitations of tag-based annotation, by examining a motivating example:

Let's suppose that user Mary has an account on platform S1, that specializes in images. Mary has been using S1 for a while, so she has progressively built a large image collection, as well as a rich vocabulary of tags (personomy).

Another user, Sylvia, who is Mary's friend, is using a different platform, S2, to annotate her images. At some point, Mary and Sylvia attended the same event, and each one took some pictures with her own camera. As each user has her reasons for choosing a preferred platform, none of them would like to change. They would like however to be able to link to each other's annotated pictures, where applicable: it can be expected that since the pictures were taken at the same time and place, some of them may be annotated in similar way (same tags), even by different annotators. So they may (within the boundaries of word ambiguity) be about the same topic.

In the course of time Mary also becomes interested in video and starts shooting some of her own. As her personal video collection begins to grow, she decides to start using another collaborative tagging system, S3, that specializes in video, in order to better organise it. Since she already has a rich personomy built in S1, she would naturally like to reuse it in S3, to the extent possible: while some of the tags may not be appropriate, as they may represent one-off ('29-08-06') or photography-specific ('CameraXYZ') use, others might as well be reused across modalities/domains, in case they represent high-level concepts ('holidays'). So if Mary has both video and photographic material of some event, and since she has already created a personomy on S1, she would naturally like to be able to reuse it (partially, perhaps) on S2 as well.

Issues

The above scenario demonstrates limitations of tag-based systems with respect to personomy reuse:

- A personomy maintained at one platform cannot easily be reused for a tag-based retrieval on another tagging platform.
- A personomy maintained at one platform cannot easily be reused to organize further media or resources on another tagging platform.

As media resides not only on Internet platforms but is most likely maintained on a local computer at first, local organizational structures can also not easily be transferred to a tagging platform. The opposite holds as well, a personomy maintained on a tagging platform cannot easily be reused on a desktop computer.

Personomy reuse is currently not easily possible as each platform uses ad-hoc solutions and only provides tag navigation within its own boundaries: there is no standardization that regulates how tags and relations between tags, users, and resources are represented. Due to that lack of standardization there are further technical issues that become visible through the application programming interfaces provided by some tagging platforms:

- Some platforms prohibit tags containing space characters while other allow such tags
- Different platforms provide different functionality for organizing tags themselves, e.g. some platforms allow to summarize tags in tag-bundles

Possible Solutions

When it comes to interoperability, standards-based solutions have repeatedly proven successful in enabling to bridge different systems. This could also be the case here, as a standard for expressing personomies and folksonomies would enable interoperability across platforms. On the other hand, use of a standard should not enforce changes in the way tags are handled internally by each system - it simply aims to function as a bridge between different systems. The question is then, what standard?

We may be able to answer this question if we consider a personomy as a concept scheme: tags used by an individual express his or her expertise, interests and vocabulary, thus constituting the individual's own concept scheme. A recent W3C standard that has been designed specifically to express the basic structure and content of concept schemes is SKOS Core [SKOS]. The SKOS Core Vocabulary is an application of the Resource Description Framework (RDF), that can be used to express a concept scheme as an RDF graph. Using RDF allows data to be linked to and/or merged with other RDF data by semantic web applications.

Expressing personomies and folksonomies using SKOS is a good match for promoting a standard representation for tags, as well as integrating tag representation with Semantic Web standards: not only does it enable expression of personomies in a standard format that fits semantically, but also allows mixing personomies with existing Semantic Web ontologies. There is already a publicly available SKOS-based tagging ontology that can be used to build on [Newman], as well as some existing efforts to induce an ontology from collaborative tagging platforms [Schmitz].

Ideally, we would expect existing collaborative tagging platform to build on a standard representation for tags in order to enable interoperability and offer this as a service to their users. In practice however , even if such a representation was eventually adopted as a standard, our expectation is that there will be both technical and political reasons that could possibly hinder its adoption. A different strategy that may be able to deal with this issue then would be to implement this as a separate service that will integrate disparate collaborative tagging platforms based on such an emergind standard for tag representation, in the spirit of Web2.0 mashups. This service could either be provided by a 3rd party, or even be self-hosted by individual users, in the spirit of [Koivunen, Segawa]

2.5 Use Case : Semantic Media Analysis for Intelligent Retrieval

Introduction

Semantic Media Analysis seen from a multimedia retrieval perspective is equivalent to the automatic creation of semantic indices and annotations based on multimedia and domain ontologies to enable intelligent human-like multimedia retrieval purposes. An efficient multimedia retrieval system [Naphade], must:

i. Be able to handle the semantics of the query,
ii. Unify multiple modalities in a homogeneous frameworkb and
iii. Abstract the relationship between low level media features and high level semantic concepts to allow the user to query in terms of these concepts rather than in terms of examples, i.e. introduction the notion of ontologies.

This Use Case aims to pinpoint problems that arise during the effort for an automatic creation of semantic indices and annotations in an attempt to bridge the multimedia semantic gap and thus provide corresponding solutions using Semantic Web Technologies.

For multimedia data retrieval, based on only low-level features as in the case of "quering by example" and of content-based retrieval paradigms and systems, on the one hand, one gets the advantage of an automatic computation of the required low-level features but on the other hand, such methodology lacks the ability to respond to high-level, semantic-based queries, and evidently loses the relation among low-level multimedia features such as pitch, or zero-crossing rate in audio or color and shape in image and video, or frequency of words in text, to high-level domain concepts that essentially characterize the underlying knowledge in data that a human is capable of quickly grasping, whereas a machine cannot. For this reason, an abstraction of high level multimedia content descriptions and semantics is required based on what can actually be generated automatically, such as low-level features after low-level processing, and on methods, tools and languages to represent the domain ontology and attain the mapping between the two. Tha latter is needed so that semantic indices are extracted as automatic as possible, rather than being produced manually which is a time-consuming and not always efficient task (attains a lot of subjective annotations). To avoid the latter limitations of manual semantic annotations on multimedia data, metadata standards and ontologies (upper, domain, etc.) have to be used and interoperate. Thus, a requirement emerges for multimedia semantics interoperability to further enable efficient solutions interoperation, when considering the distributed nature of the Web and the enormous amounts of multimedia data published there.

An example solution for the interoperability problem stated above is the MPEG-7 standard. MPEG-7, composed of various parts, defines both metadata descriptors for structural and low-level aspects of multimedia documents, as well as high level description schemes (Multimedia Description

Schemes) for a higher-level of descriptions including semantics of multimedia data. However, it does not determine the mapping of the former to the latter based on the addressed application domain. A number of publications have appeared to define the MPEG-7 core ontology to address such issues. What is important is that the MPEG-7 provides the standardised means of descriptors both low-level and high level. The value sets of those descriptions along with a richer set of relationships definitions could form the necessary missing piece along with the knowledge discovery algorithms which will use these to extract semantic descriptions and indices in an almost automatic way out of multimedia data. The bottom line thus is that MPEG-7 metadata descriptions need to be properly linked to domain-specific ontologies that model high-level semantics.

Furthermore, one should consider usually the multimodality feature of multimedia data and content on the Web. The same concept there may be described by different means, that is by news in text as well as an image showing a snapshot of what the news are reporting. Thus, since the provision of cross-linking between different media types or corresponding modalities supports a rich scope for inferencing a semantic interpretation, interoperability between different single media schemes (audio ontology, text ontology, image ontology, video ontology, etc.) is an important issue. This emerges from the need to homogenise different single modalities for which it is possible that:

i. Can infer particular high level semantics with different degrees of confidence (e.g. rely mainly on audio for infering certain concepts than text),
ii. Can be supported by a world modelling (or ontologies) where different relationships exist, e.g. in an image one can attribute spatial relationships while in a video sequence spatio-temporal relationships can be attained, and
iii. Can have different role in a cross-modality fashion â€“ which modality triggers the other, e.g. to identify that a particular photo in a Web page depicts person X, we first extract information from text on the person's identity and thereafter we cross-validate by the corresponding information extraction from the image.

Both of the above concerns, either the single modality tackled first or the cross-modality (which essentially encapsulates the sinlge modality), require semantic interoperability which will support a knowledge representation of the domain concepts and relatioships, of the multimedia descriptors and of the cross-linking of both, as well as a multimedia analysis part combined with modeling, inferencing and mining algorithms that can be directed towards automatic semantics extraction from multimedia to further enable efficient semantic-based indexing and intelligent multimedia retrieval.

Motivating Examples

In the following, current pitfalls with respect to the desired semantic interoperability are given via examples. The discussed pitfalls are not the only ones, therefore, further discussion is needed to cover the broad scope of semantic multimedia analysis and retrieval.

Example 1: Single modality case: Lack of semantics in low-level descriptors

The linking of low-level features to high-level semantics can be obtained by the following two main trends:

i. Using machine learning and mining techniques to infer the required mapping, based on a basic knowledge representation of the concepts of the addressed domain (usually low-to-medium level inferencing) and
ii. Using ontology-driven approaches to both guide the semantic analysis and infer high-level concepts using reasoning and logics. This trend can include the first one as well and then be further driven by medium-level semantics to more abstract domain concepts and relationships.

In both trends, it is appropriate for granularity purposes to produce concept/event detectors, which usually incorporate a training phase applied on training feature sets for which ground-truth is available (apriori knowledge of addressed concepts or events). This phase enable optimization of the underlying artificial intelligence algorithms. Semantic interoperability cannot be achieved by only exchanging low-level features, wrapped in standardised metadata descriptors, between different users or applications, since there is a lack of formal semantics. In particular, a set of low level descriptors (eg. MPEG-7 audio descriptors) cannot be semantically meaningful since there is a lack of intuitive interpretation to higher levels of knowledge - these have been however extensively used in content-based retrieval that relies on similarity measures. The low level descriptors are represented as a vector of numerical values, and thus, they are useful for a content-based multimedia retrieval rather than a semantic multimedia retrieval process.

Furthermore, since a set of optimal low level descriptors per target application (be it music genre recognition or speaker indexing) can be conceived by only multimedia analysis experts, this set has to be transparent to any other user. For example, although a non-expert user can understand the color and shape of a particular object, he is unable to attribute to this object a suitable representation by the selection of appropriate low level descriptors. It is obvious that the low level descriptors do not only lack semantics but also limit their direct use to people that have gained a particular expertise concerning multimedia analysis and multimedia characteristics.

The problem raised out of this example that needs to be solved is in *which way low level descriptors can be efficiently and automatically linked and turned into an exchangeable bag of semantics*.

Example 2: Multi-modality case: Fusion and interchange of semantics among media

In multimedia data and web content, cross-modality aspects are dominant, a characteristic that can be efficiently exploited by semantic multimedia analysis and retrieval, when all modalities can be exploited to infer the same or related concepts or events. One aspect, is again motivated from the analysis part, that refers to particular concepts and relationships capturing, which require a priority in the processing of modalities during their automatic extraction. For example, to enhance recognition of a face of a particular person in an image appearing in a Web page, which is actually a very difficult task, it seems more natural and efficient that initially inferencing is based on the textual content, to locate the identity (name) of the person, and thereafter, the results can be validated or enhanced by related results from image analysis. Similar multimodal media analysis benefits can be obtained by analysing synchronized audio-visual content to semantically annotate it. The trends there are:

i. To conscruct combined feature vectors from audio and visual features and feed those to machine learning algorithms to extract combined semantics
ii. To analyse each single modality separately towards recognizing medium-level semantics or the same concepts and then fuse results of analysis (decision fusion) in usually a weighted or ordered manner (depending on the underlying single modality cross-relations towards the same topic) to either improve the accuracy of semantics extraction results or enrich them, towards higher level semantics.

For the sake of clarity, an example scenario is described in the following which is taken from the â€˜sportsâ€™ domain and more specifically from â€˜athleticsâ€™.

Let's assume that we need to semantically index and annotate, in the most possible automatic way, the web page shown at Figure 1, which is taken from the site of the International Association of Athletics Federation. The subject of this page is "the victory of the athlete Reiko Sosa at the Tokyoâ€™s marathon". Let's try to answer the question: What analysis steps are required if we would like to enable semantic retrieval results for the query "show me images with the athlete Reiko Sosa" ?

One might notice that for each image in this web page there is a caption which includes very useful information about the content of the image, in particular the persons appearing in it, i.e. structural (spatial) relations of the media-rich web page contents. Therefore, it is important to identify the areas of an image and the areas of a caption. Let's assume that we can detect those areas (it is not useful to get into details how). Then, we proceed in the semantics extraction of the textual content in the caption which identifies:

- Person Names = {Naoko Takahashi, Reiko Sosa},
- Places = {Tokyo}, Athletics type = {Womenâ€™s Marathon} and
- Activity = {runs away from} (see Figure 1, in yellow and blue color).

In the case of the semantics extraction from images, we can identify the following concepts and relationships:

- In the image at the upper part of the web page, we can get the â€˜athleteâ€™s facesâ€™ and with respect to the spatial relationship of those faces we can identify which face (athlete) takes lead against the other. Using only the image we cannot draw a conclusion who is the athlete.
- In the image at the lower part of the web page, we can identify that there exist a person after a face detection but still, we cannot ensure to whom this face belongs to.

If we combine both the semantics from textual information in captions and the semantics from image we may give a large support to reasoning mechanisms to reach the conclusion that "we have images with the athlete Reiko Sosa". Nonetheless, in the case that we have several athletes like in the image on the upper web image part, reasoning using the identified spatial relationship can spot which particular athlete between the two, is Reiko Sosa.

Example of a web page about athletics.

Another scenario involved multimodal analysis of audio-visual data, distributed on the web or accessed through it from video archives, and concerns automatic semantics extraction and annotation of video scenes related to violence, for further purposes of content filtering and parental control [Perperis]. Thus, the goal in this scenario is automatic identification and semantic classification of violent content, using features extracted from visual, auditory and textual modalities of multimedia data.

Let's consider that we are trying to automatically identify violent scenes where fighting among two persons takes place with no weapons involved. The low-level analysis parts will lead to different low-level descriptors separately for each modality. For example, for the visual modality the analysis will involve:

- Shot cut detection and video segmentation.
- Human body recognition and motion analysis.

- Human body parts recognition (arms, legs).
- Human body parts movement and tracking (i.e. "Fast horizontal hand movement")
- Interpretation of simple "visual" events/concepts based on spatial and temporal relations of identified objects (medium-level semantics).

On the other hand, the analysis of the auditory modality will involve:

- Audio signal segmentation.
- Segment classification in sound categories, including speech, silence, music, scream, etc. which may relate to violence events or not (medium-level semantics).

Now, by of course fusing medium-level semantics and results from the single modality analysis, taking under consideration spatio-temporal relations and behaviour patterns, we evidently can automatically extract (infer) higher level semantics. For example, the "punch" concept can be automatically extracted based on the initial analysis results and on the sequence or synchronicity of audio or visual detected events such as two person in visual data, the one moving towards the other, while a punch sound and scream of pain is detected in the audio data.

To fulfil such scenarios as the ones presented above, we should solve the problem *how to fuse and interchange semantics from different modalities*.

Possible Solutions

Example 1

As it was mentioned in Example 1, semantics extraction can be achieved via concept detectors after a training phase based upon feature sets. Towards this goal, recently there was a suggestion in [Asbach] to go from a low level description to a more semantic description by extending MPEG-7 to facilitate sharing classifier parameters and class models. This should occur by presenting the classification process in a standardised form. A classifier description must specify on what kind of data it operates, contain a description of the feature extraction process, the transformation to generate feature vectors and a model that associates specific feature vector values to an object class. For this, an upper ontology could be created, called a classifier ontology, which could be linked to a multimedia core ontology (eg. CIDOC CRM ontology), a visual descriptor ontology [VDO] as well as a domain ontology. A similar approach is followed by the method presented in [Tsekeridou], where classifiers are used to recognize and model music genres for efficient music retrieval, and description extensions are introduced to account for such extended functionalities.

As to these aspects, the current Use Case relates at some extend to the Algorithm Representation UC. However, the latter refers mainly to general purpose processing and analysis and not to analysis and semantics extraction, based on classification and machine learning algorithms, to enable intelligent retrieval.

In the proposed solution, the visual descriptor ontology consists of a superset of MPEG-7 descriptors since the existing MPEG-7 descriptors cannot always support an optimal feature set for a particular class.

A scenario that exemplifies the use of the above proposal is given in the following. Maria is an architect who wishes to retrieve available multimedia material of a particular architecture style like â€˜Art Nouveauâ€™, â€˜Art Decoâ€™, â€˜Modernâ€™ among the bulk of data that she has already stored using her multimedia management software. Due to its particular interest, she plugs in the â€˜Art Nouveau classifier kitâ€™ that enables the retrieval of all images or videos that correspond to this particular style in the form of visual representation or non-visual or their combination (eg. a video on exploring the House of V. Horta, a major representative of Art Nouveau style in Brussels, which includes visual instances of the style as well as a narration about Art Nouveau history).

Necessary attributes for the classifier ontology are estimated to be:

- The name and category of the Classifier
- The list and types of input parameters
- The output type
- Limitations on data set, on value ranges for parameters, on processing time and memory requirements
- Permormance metrics
- Guidelines of use
- Links to class models per domain/application and feature sets

In the above examples, the exchangeable bag of semantics is directly linked to an exchangeable bag of supervised classifiers.

Example 2

In this example, to support reasoning mechanisms, it is required that apart from the ontological descriptions for each modality, there is a need for a cross-modality ontological description which interconnects all possible relations from each modality and constructs rules that are cross-modality specific. It is not clear, whether this can be achieved by an upper multimedia ontology or a new cross-modality ontology that will strive toward the knowledge representation of all possibilities combining media. It is evident though, that the cross-modality ontology, along with the single modality ones, greatly relate to the domain ontology, i.e. to the application at hand.

Furthermore, in this new cross-modality ontology, special attention should be taken for the representation of the priorities/ordering among modalities for any multimodal concept (eg. get textual semantics first to attach semantics in an image). This translates to sequential rules construction. However there are cases, where simultaneous semantic instances in different modalities may lead to higher level of semantics, that synchronicity is also a relationship to be accounted for. Apart from the spatial, temporal or spatio-temporal relationships that need to be accounted for, there is also the issue of importance of each modality for identifying a concept or semantic event. This may be represented by means of weights.

The solution is composed also by relating visual, audio, textual descriptor ontologies with a cross-modality ontology showcasing their inter-relations as well as a domain ontology representing the concepts and relations of the application at hand.

2.6 Use Case: Algorithm Representation

Introduction

The problem is that algorithms for image analysis are difficult to manage, understand and apply, particularly for non-expert users. For instance, a researcher needs to reduce the noise and improve the contrast in a radiology image prior to analysis and interpretation but is unfamiliar with the specific algorithms that could apply in this instance. In addition, many applications require the processes applied to media to be concisely recorded for re-use, re-evaluation or integration with other analysis data. Quantifying and integrating knowledge, particularly visual outcomes, about algorithms for media is a challenging problem.

Solution

Our proposed solution is to use an algorithm ontology to record and describe available algorithms for application to image analysis. This ontology can then be used to interactively build sequences of algorithms to achieve particular outcomes. In addition, the record of processes applied to the source image can be used to define the history and provenance of data.

The algorithm ontology should consist of information such as:

- name
- informal natural language description
- formal description
- input format
- output format
- example media prior to application
- example media after application
- goal of the algorithm

To achieve this solution we need:

- a sufficiently detailed and well-constructed algorithm ontology;
- a core multimedia ontology;
- domain ontologies and;
- the underlying interchange framework supplied by semantic web technologies such as XML and RDF.

The benefits of this approach are a modularity through the use of independent ontologies to ensure usability and flexibility.

State of the Art and Challenges

Currently there exists a taxonomy/thesaurus for image analysis algorithms we are working on [Asirelli] but this is insufficient to support the required functionality. We are collaborating on expanding and converting this taxonomy to an OWL ontology.

The challenges are:

- to articulate and quantify the â€˜visualâ€™ result of applying algorithms;
- to associate practical example media with the algorithms specified;
- to integrate and harmonise the ontologies;
- to reason with and apply the knowledge in the algorithm ontology (e.g. using input and output formats to align processes).

Possible Applications

The formal representation of the semantics of algorithms enables recording of provenance, provides reasoning capabilities, facilitates application and supports interoperability of data. This is important in fields such as:

1. Smart assistance to support quality control and defect detection of complex, composite, manufactured objects;
2. Biometrics (face recognition, human behaviour, etc.)
3. The composition of web services to automatically analyse media based on user goals and preferences;
4. To assist in the formal definition of protocols and procedures in fields that are heavily dependent upon media analysis such as scientific or medical research.

These are applications that utilise media analysis and need to integrate information from a range of sources. Often recording the provenance of conclusions and the ability to duplicate and defend results is critical.

For example, in the field of aeronautical engineering, aeroplanes are constructed from components that are manufactured in many different locations. Quality control and defect detection requires data from many disparate sources. An inspector should understand the integrity of a component by acquiring local data (images and others) and combining it with information from one or more databases and possibly interaction with an expert.

Example

Problem:

> Suggest possible clinical descriptors (pneumothorax) given a chest x-ray.

Hypothesis of solution :

> 1) Get a digital chest x-ray of patient P (image A).
> 2) Apply on image A a digital filter to improve the signal-to-noise ratio (image B).
> 3) Apply on image B a region detection algorithm. This algorithm segments image B according to a partition of homogeneous regions (image C).
> 4) Apply on image C an algorithm that 'sorts' according to a given criterion the regions by their geometrical and densitometric properties (from largest to smallest, from darkest to clearest, etc.) (array D).
> 5) Apply on array D an algorithm that searching on a database of clinical descriptors detects the one that best fits the similarity criterion (result E).

However, we should consider the following aspects:

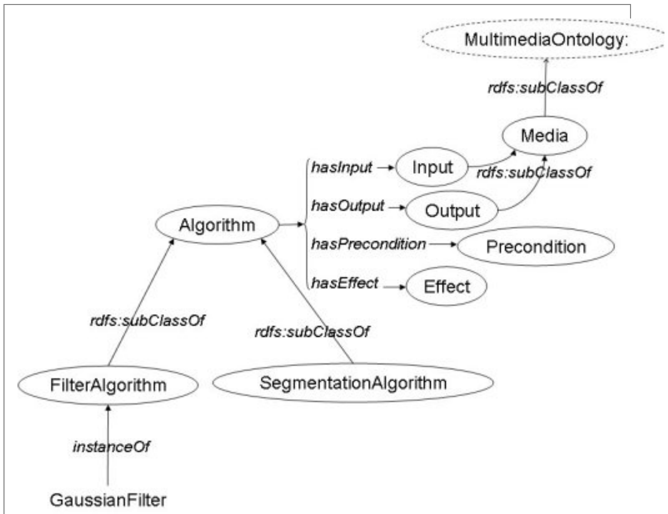
> step 2) Which digital filter should be applied on image A? We can consider different kinds of filters (Fourier, Wiener, Smoothing, etc.) each one having different input-output formats and giving slightly different results.
> step 3) Which segmentation algorithm should be used? We can consider different algorithms (clustering, histogram, homogeneity criterion, etc.).

step 4) How can we define geometrical and densitometric properties of the regions? There are several possibilities depending on the considered mathematical models for describing closed curves (regions) and the grey level distribution inside each region (histogram, Gaussian-like, etc.).
step 5) How can we define similarity between patterns? There are multiple approaches that can be applied (vector distance, probability, etc.).

Each step could be influenced by the previous ones.

Goal: to segment the chest x-ray image (task 3)

A segmentation algorithm is selected. To be most effective this segmentation algorithm requires a particular level of signal-to-noise ratio. This is defined as the precondition (Algorithm.hasPrecondition) of the segmentation algorithm (instanceOf.segmentationAlgoritm). To achieve this result a filter algorithm is found (Gaussian.instanceOf.filterAlgorithm) which has the effect (Algorithm.hasEffect) of improving the signal-to-noise ratio for images of the same type as the chest x-ray image (Algorithm.hasInput). By comparing the values of the precondition of the segmentation algorithm with the effect of the filter algorithm we are able to decide on the best algorithms to achieve our goal.



Excerpt of an Algorithm Ontology.

Interoperability aspects

Two types or levels of interoperability to be considered:

1) low-level interoperability, concerning data formats and algorithms, their transition or selection aspects among the different steps and consequently the possible related ontologies (algorithm ontology, media ontology);
2) high-level interoperability, concerning the semantics at the base of the domain problem, that is how similar problems (segment this image; improve image quality) can be faced or even solved using codified 'experience' extracted from well-known case studies.

In our present use case proposal we focused our attention mainly on the latter.

Considering for instance the pneumothorax example, this can be studied starting from a specific pre-analyzed case in order to define a general reference procedure: what happens if we have to study a pneumothorax case starting from an actual arbitrary image of a patient? Applying simply the general procedure will not give in general the right solution because each image (i.e. each patient) has its own specificity and the algorithms have to be bound to the image type. Thus, the general procedure is not the one which fits for any case because the results depend on the image to be processed. And also in the better case, the result would be supervised and it would be necessary to apply another algorithm to improve the result itself. High-level interoperability would involve also a procedure able to take trace of a specific result and how it has been obtained starting from a particular input.

The open research questions that we are currently investigating relate to the formal description of the values of effect and precondition and how these can be compared and related. The interoperability of the media descriptions and ability to describe visual features in a sufficiently abstract manner are key requirements.

3. Open Issues

3.1 Semantics From Multimedia Authoring

Introduction

Authoring of personalized multimedia content can be considered as a process consisting of selecting, composing, and assembling media elements into coherent multimedia presentations that meet the userâ€™s or user groupâ€™s preferences, interests, current situation, and environment. In the approaches we find today, media items and semantically rich metadata information are used for the selection and composition task.

For example, Mary authors a multimedia birthday book for her daughter's 18th birthday with some nice multimedia authoring tool. For this she selects images, videos and audio from her personal media store but also content which is free or she own from the Web. The selection is based of the different metadata and descriptions that come with the media such as tags, descriptions, the time stamp, the size, the location of the media item and so on. In addition to the media elements used Mary arranges them in a spatio-temporal presentation: A welcome title first and then along "multimedia chapters" sequences and groups of images interleaved by small videos. Music underlies the presentation. Mary arranges and groups, adds comments and titles, resizes media elements, brings some media to front, takes others into the back. And then, finally, there is this great birthday presentation that shows the years of her daughter's life. She presses a button, creates a Flash presentation and all the authoring semantics are gone.

Lost multimedia semantics

Metadata and semantics today is mainly seen on the monomedia level. Single media elements such as image, video and text are annotated and enriched with metadata by different means ranging from automatic annotation to manual tagging. In a multimedia document typically a set of media items come together and are arranged into a coherent story with a spatial and temporal layout of the time-continuous presentation, that often also allows user interaction. The authored document is more than "just" the sum of the media elements it becomes a new document with its own semantics. However, in the way we pursue multimedia authoring today, we do not care and lose the emergent sementics from multimedia authoring.

Multimedia authoring semantics do not "survive" the composition

So, most valuable semantics for the media elements and the resulting multimedia content that emerge with and in the authoring process are not considered any further. This means that the effort for semantically enriching media content comes to a sudden halt in the created multimedia document â€“ which is very unfortunate. For example, for a multimedia presentation it could be very helpful if an integrated annotation tells something about the structure of the presentation, the media items and formats used, the lenght of the presentation, its degree of interactivity, the table of contents of index of the presentation, a textual summary of the content, the targeted user group and so on. Current authoring tools just use metadata to select media elements and compose them into a multimedia presentation. They do not extract and summarize the semantics that emerge from the authoring and add them to the created document for later search, retrieval and presentation support.

Multimedia content can learn from composition and media usage

For example, the media store of Mary could "learn" that some of the media items seem to be more relevant than others. Additional comments on parts of the presentation could also be new metadata entries for the media items. And also the metadata of the single media items as well as of the presentation are not added to the presentation such that is can afterwards more easier be shared, searched, managed.

Interoperability problems

Currently, multimedia documents do not come with a single annotation scheme. SMIL comes with the most advanced modeling of annotation. Based on RDF, the head of a SMIL document allows to add an RDF description of the presentation to the structured multiemdia document and gives the author or authoring tool a space where to put the presentation's semantics. In specific domains we find annotation schemes such as LOM that provide the vocabulary for annotating Learning Objects which are often Powerpoint Presentations of PDF documents but might well be multimedia presentations. AKtive Media is an ontology based multimedia annotation (Images and Text) system which provides an interface for adding ontology-based, free-text and relational annotations within multimedia documents. Even though the community effort will contribute to a more or less unified set of tags, this does not ensure interoperability, search, and exchange.

What is needed

A semantic description of multimedia presentation should reveal the semantics of its content as well as of the composition such that a user can search, reuse, integrate multimedia presentation on the Web into his or her system. A unified semantic Web annotation scheme could then describe the thousands of Flash presentations as well as powerpoints presentation, but also SMIL and SVG presentations. For existing presentations this would give the authors a chance to annotate the presentations. For authoring tool creators this will give the chance to publish a standardized semantic presentation description with the presentation.

3.2 Building Multimedial Semantic Web Applications

Introduction

This use case is all about supporting to build real distributed, Semantic Web applications in the domain of multimedial content. It discusses scalability, and interop issues and tries to propose solutions to lower the barrier of implementing such multimedial Semantic Web applications.

Motivation

Shirin is a IT manager at a NGO, called FWW (Foundation for Wildlife in the World) and wants to offer some new multimedial service to inform, alarm, etc. members, e.g.:

Track your animal godchild (TyAG)

> A service that would allow a member to audio-visually track his godchild (using geo-spatial services, camera, satellite, RFID :). DONald ATOR, a contributer of FWW is the godfather of a whale. Using the TyAG service he is able to observe the route that his favorite whale takes (via Geonames) and in case that the godchild is near a FWW-observing point, Donald might also see some video footage. Currently the whales are somewhere around Thule Island. TyAG allows Donald to ask questions like: *When* will the whales be *in my region*? etc.

Video-news (vNews)

> As Donald has gathered some good experiences with TyAG, he wants to be informed about news, upcoming events, etc. w.r.t. *whales*. The backbone of the vNews system is smart enough to understand that *whales* are a kind of *animals that live in the water*. Any time a FWW member puts some footage on the FWW-net that has some *water animals* in it, vNews - using some automated feature extraction utils - offers it to Donald as well to view it. **Note:** There might be a potential use of the outcome of the News Use Case here.

Interactive Annotation

> A kind of video blogging [Parker] using vNews. Enables members to share thoughts about endangered species etc. or to find out more information about a specific entity in a (broadcasted) videostream. Therefore, vNews is able to automatically segment its video-content and set up a list of *objects*, etc. For each of the *objects* in a video, a user can get further information (by linking it to Wikipedia, etc.) and share her thoughts about it with other members of the vNews network.

Possible Solutions

Common to all services listed above is an ample infrastructure that has to deal with the following challenges:

- Using many different (multimedial) metadata (EXIF, GPS-data, etc.) as input, a common internal representation has to be found (e.g. MPEG-7) - INTEROPERABILITY.
- For the domain (animals) a formal description needs to be defined - ONTOLOGY ENGINEERING (also visual to entity mapping).
- Due to the vast amount of metadata, a scaleable approach has to be taken that can handle both *low-level* features (in MPEG-7) and *high-level* features (in RDF/OWL) - SCALABILITY.

We now try to give possible answers to the above listed question to enable Shirin to implement the services in terms of:

- Based on well-known ontology engineering methods give hints on how to modell the domain.
- Giving support in selecting a representation that both addresses low-level as high-level features.
- Supplying an SW-architect with an evaluation of RDF-stores w.r.t. multimedial metadata.

4. Common Framework

In this section, we will propose a common framework that seek to provide both syntactic (via RDF) and semantic interoperability. During the FTF2, we have identified several layers of interoperability. Our methodology is simple: each use case identifies a common ontology/schema to facilitate interoperability in its own domain, and then we provide a simple framework to integrate and harmonise these common ontologies/schema from different domains. Furthermore, the simple extensible mechanism is provided to accommodate other ontologies/schema related to the use cases we considered. Last but not least, the framework includes some guidelines on which standard to use for specific tasks related to the use cases.

4.1. Syntactic Interoperability: RDF

Resource Description Framework (RDF) is a W3C recommendation that provides a standard to create, exchange and use annotations in the Semantic Web. An RDF statement is of the form [subject property object .] This simple and general form of syntax makes RDF a good candidate to provide (at least) syntactic interoperability.

4.2. Layers of Interoperability

[Based on discussions in FTF2]

4.3. Common Ontology/Schema

Individual use case provides its common ontology/schema for its domain.

4.4. Ontology/Schema Integration, Harmonisation and Extension

[Integrate and harmonise the common ontologies/schema presented in the previous sub-section. Based on this, to provide a simple extensible mechanism.]

4.5. Guidelines

Individual use case provides guidelines on which standard to use for specific tasks related to the use case.

5. Conclusion

6. References

[Asbach]
Object detection and classification based on MPEG-7 descriptions â€“ Technical study, use cases and business models. M. Asbach and J-R Ohm. ISO/IEC JTC1/SC29/WG11/MPEG2006/M13207, April 2006, Montreaux, CH.

[Asirelli]
An Infrastructure for MultiMedia Metadata Management. Patrizia Asirelli, Massimo Martinelli, Ovidio Salvetti. *In:* Proceedings of International SWAMM Workshop, 2006.

[Broekstra]
Sesame: A generic archistecture for storing and querying RDF and RDF schema. J. Broekstra, A. Kampman and F. van Harmelen. *In:* Proceedings of The International Semantic Web Conference 2002 (pages 54-68), 2002, Sardinia

[Declerck]
Translating XBRL Into Description Logic. An Approach Using Protege, Sesame & OWL. T. Declerck and H.-U Krieger. *In:* Proceedings of the 9th International Conference on Business Information Systems, 2006

[DIG35]
Digital Imaging Group (DIG), DIG35 Specification - Metadata for Digital Images - Version 1.0 August 30, 2000

[Dublin Core]
The Dublin Core Metadata Initiative, Dublin Core Metadata Element Set, Version 1.1: Reference Description

[EBU]
European Broadcasting Union, http://www.ebu.ch/

[Exif]
Standard of Japan Electronics and Information Technology Industries Association, Exchangeable image file format for digital still cameras: Exif Version 2.2

[Flickr]
Flickr online photo management and sharing application, http://www.flickr.com/, Yahoo! Inc, USA

[Foto Community]
Foto Community, http://www.fotocommunity.com/

[GFK]
Usage behavior digital photography. GfK Group for CeWe Color, 2006

[Hildebrand]
/facet: A browser for heterogeneous semantic web repositories. Michiel Hildebrand, Jacco van Ossenbruggen, and Lynda Hardman. *In:* The Semantic Web - ISWC 2006 (pages 272-285), November 2006, Athens, USA.

[HitWise]
HitWise Intelligence, Del.icio.us Traffic More Than Doubled Since January

[Hotho]
Information Retrieval in Folksonomies: Search and Ranking. A. Hotho, R. Jaschke, C. Schmitz and G. Stumme. *In:* The 3rd European Semantic Web Conference (ESWC), 2006 Budva, Montenegro.

[IIM]
Information Interchange Model, http://www.iptc.org/IIM/, International Press Telecommunication Council (IPTC)

[Koivunen]
Annotea and Semantic Web Supported Collaboration. M. Koivunen. *In:* Proceedings of the European Semantic Web Conference (ESWC), Crete, 2005

[Knublauch]
Editing description logic ontologies with the Protege OWL plugin. H. Knublauch, M.A. Musen and A.L. Rector. *In:* Proceedings of the International

Workshop on Description Logics (DL), 2004

[Lara]
XBRL Taxonomies and OWL Ontologies for Investment Funds. R. Lara, I. Cantador and P. Castells, *In:* ER (Workshops), (pages 271-280), 2006

[Mathes]
Folksonomies - Cooperative Classification and Communication Through Shared Metadata. A. Mathes, Computer Mediated Communication - LIS590CMC, Graduate School of Library and Information Science, University of Illinois Urbana-Champaign, 2004

[Mejias]
Tag literacy. Ulises Ali Mejias, http://ideant.typepad.com/ideant/2005/04/tag_literacy.html, 2005

[MPEG-7]
Information Technology - Multimedia Content Description Interface (MPEG-7). Standard No. ISO/IEC 15938:2001, International Organization for Standardization(ISO), 2001

[MMSEM Image]
Image Annotation on the Semantic Web , Raphaël Troncy, Jacco van Ossenbruggen, Jeff Z. Pan and Giorgos Stamou, Multimedia Semantics Incubator Group Report (XGR), 14 August 2007, http://www.w3.org/2005/Incubator/mmsem/XGR-image-annotation/

[Naphade]
Extracting semantics from audiovisual content: The final frontier in multimedia retrieval, N. Naphade and T. Huang. *In:* IEEE Transactions on Neural Networks, vol. 13, No. 4, 2002.

[NetRatings]
Nielsen/NetRatings, User-generated content drives halfs of US Top 10 fastest growing web brands

[Newman]
Richard Newman, Danny Ayers and Seth Russell. Tag Ontology, http://www.holygoat.co.uk/owl/redwood/0.1/tags/

[NewsML-G2]
IPTC, News Architecture (NAR) for G2-Standards Specifications (released 30th May, 2007)

[NewsCodes]
NewsCodes - Metadata taxonomies for the news industry, http://www.iptc.org/NewsCodes/

[OWL]
OWL Web Ontology Language Reference, S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D.L. McGuinness, P.F. Patel-Schneider and L.A. Stein, Editors, W3C Recommendation, 10 February 2004, http://www.w3.org/TR/owl-guide/

[Pachet]
Knowledge Management and Musical Metadata. F. Pachet, Encyclopedia of Knowledge Management, Schwartz, D. Ed. Idea Group, 2005

[Parker]
Video blogging: Content to the max, C. Parker and S. Pfeiffer. IEEE MultiMedia, vol. 12, no. 2, pp. 4-8, 2005

[Perperis]
Automatic Identification in Video Data of Dangerous to Vulnerable Groups of Users Content, T. Perperis and S. Tsekeridou. Presentation at SSMS2006, Halkidiki, Greece, 2006

[PhotoRDF]
W3C Note 19 April 2002, Describing and retrieving photos using RDF and HTTP

[Riya]
Riya Foto Search, http://www.riya.com/

[Segawa]
Web annotation sharing using P2P, O. Segawa. *In:* Proceedings of the 15th International Conference on World Wide Web, pages 851-852, Edinburgh, Scotland, 2006.

[Smith]
Atomiq: Folksonomy: social classification, G. Smith, August 2004.

[Schmitz]
Inducing Ontology from Flickr Tags, P. Schmitz. *In:* Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland, 2006.

[SKOS]
SKOS Core, http://www.w3.org/2004/02/skos/core/

[Trant]
Exploring the potential for social tagging and folksonomy in art museums: proof of concept, J. Trant. *In:* New Review of Hypermedia and Multimedia, 2006

[Tsekeridou]
MPEG-7 based Music Metadata Extensions for Traditional Greek Music Retrieval, S. Tsekeridou, A. Kokonozi, K. Stavroglou and C. Chamzas. *In:* IAPR Workshop on Multimedia Content Representation, Classification and Security, Istanbul, Turkey, September 2006

[VDO]
aceMedia Visual Descriptor Ontology, http://www.acemedia.org/aceMedia/reference/resource/index.html

[XBRL]
XBRL - eXtensible Business Reporting Language, http://www.xbrl.org/Home/, see also Tim Bray's blog

[XMP]
Adobe, XMP Specification

Acknowledgments

The editors would like to thank all the contributors for the authoring of the use cases (Melliyal Annamalai, George Anadiotis, Patrizia Asirelli, Susanne Boll, Oscar Celma, Thierry Declerk, Thomas Franz, Christian Halaschek Wiener, Michael Hausenblas, Michiel Hildebrand, Suzanne Little, Erik Mannens, Massimo Martinelli, Ioannis Pratikakis, Ovidio Salvetti, Sofia Tsekeridou, Giovanni Tummarello) and the XG members for their feedback on earlier versions of this document.

$Id: Overview.html,v 1.8 2007/08/14 23:54:38 rtroncy Exp $