# Multimedia Annotation using Semantic Web Technologies

TOMO SJEKAVICA[1], GORDAN GLEDEC[2], MARKO HORVAT[2]
[1] University of Dubrovnik
Department of Electrical Engineering and Computing
Ćira Carića 4, Dubrovnik
CROATIA
[2] University of Zagreb
Faculty of Electrical Engineering and Computing
Unska 3, Zagreb
CROATIA
tomo.sjekavica@unidu.hr, gordan.gledec@fer.hr, marko.horvat2@fer.hr

*Abstract:* - Due to the progressively increasing amount of multimedia on the Web, the need for efficient metadata formats describing that content has become increasingly evident. This paper gives an overview of the different approaches and methods for creation and retrieval of semantic rich multimedia metadata. Semantic web and its most important technologies XML, RDF and ontologies used for multimedia annotation are defined. An overview of various multimedia metadata vocabularies and formats that vary in their size and purpose is provided. Multimedia metadata is a type of metadata used for describing different aspects of multimedia content. All formats of multimedia metadata are not compatible with each other and most of it do not provide enough semantics. New Semantic Web technologies provide well-defined information meaning so different multimedia metadata can be more easily processed by computers.

*Key-Words:* - multimedia, semantic annotation, metadata, Semantic Web, ontologies

## 1 Introduction

With the expansion of web technologies, Internet is becoming more accessible to a large number of users. On various websites every day it is possible to find progressively increasing amounts of data, information and diverse content. Multimedia is steadily increasing its share in web-available content, be it in the form of images, video or audio clips. Multimedia content needs to be annotated for easier and efficiently use.

Multimedia metadata provide added value both to users and computers that use multimedia content. The simplest form of multimedia metadata is plain text, easily readable by humans, but the formal semantics of that metadata is very poor and it is very hard for computers to process those annotations. Another form of multimedia metadata is obtained by adding keywords that describe some specific part or the whole multimedia content. These keywords are usually entered manually by web users, but generally that metadata also lacks formal semantics. Meaning of multimedia metadata and their semantics should be converted into a formal language that is understandable to computers. A possible solution is to create a common vocabulary for a specific domain.

Created vocabularies are the basis for ontologies construction. Ontologies have usage in many areas of computer science, which includes usage in Semantic Web to enhance the usefulness of the Web and its resources. The Semantic Web is not a separate Web but an extension of the existing one in which information is given well-defined meaning, thus facilitating collaboration of humans and computers [1]. Ontologies define a list of terms and concepts and their relationships within a particular domain of use [2]. Besides ontologies which are third major component of the Semantic Web, the first two, XML and RDF can also be used for multimedia annotation. XML allows all users to create their own tags, and RDF defines a specific meaning in the form of RDF statement that consists of three elementary parts (subject, predicate and object) [1].

Many different standards for describing multimedia content have been developed. Some of multimedia standards were developed before Semantic Web so those standards are mainly based on XML and among them lacks formal semantics. Due to the lack of appropriate applications and the complexity of most standards, manual annotation is a time-consuming and an expensive process. To

solve these problems, there is a need to merge good practices in multimedia industry with the benefits of Semantic Web technologies [3]. This way of integration will immediately payoff to providers of multimedia metadata because they will directly benefit from the Semantic Web applications that are public available. Besides, integration would enable the development of intelligent applications that could understand multimedia metadata, which is not possible with XML syntax based standards. Semantic Web open approach would enable easier integration of multiple vocabularies from different communities. Finally, extensible small and simple vocabularies could be defined. These vocabularies should be suitable for personal use, but at the same time flexible enough for extension in order to be used in more complex and professional tasks for multimedia annotation.

## 2 Semantic Web

The usage of Semantic Web technologies is recommended for creating new and enriching existing annotations due to a large number of multimedia metadata formats and standards and their incompatibility.

Semantic Web is an extension of the World Wide Web and not a separate Web. With the Semantic Web, information and content on the Web gets a well defined meaning that computers facilitate understanding of the meaning, semantics and information [1]. Semantic Web describes properties of the content and dependencies between different content, which allows unambigous exchange of information between people and computers. The first form of semantic data on the Internet were the metadata that represent data about the data. Multimedia metadata is type of metadata used for describing multimedia content.

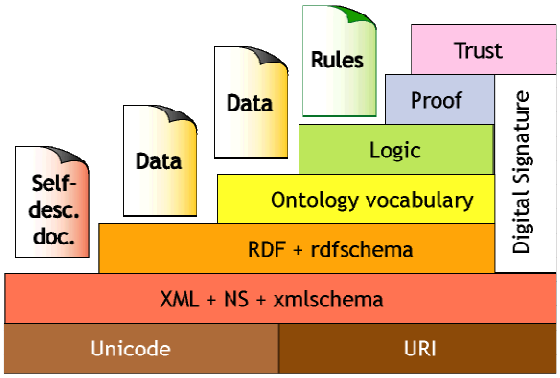


*Fig 1 Semantic Web Stack by Tim Berners-Lee [4]*

Architecture of the Semantic Web can be displayed using the Semantic Web Stack shown in Figure Fig *1*. Three important standards that make architecture of the Semantic Web and that are used in multimedia annotation are XML, RDF and ontologies.

XML is placed on the second layer of Semantic Web stack. Using XML, users can create their own tags for structured web documents. These custom tags can be used as tags of whole or a part of web pages, as well as other content on the Web. XML allows no semantic value for the meaning of XML documents. XML Schema is a language used to define the structure of the XML documents. Two applications that want to communicate with each other can use the same vocabulary or the same definiton of the structure of an XML document.

RDF [5] is a basic data-model used to write simple statements about resources on the Web. RDF data-model does not rely on XML, but uses XML based syntax. Resources, properties and statements are three main concepts of RDF. Anything that can be idetified by URI is a resource. Properties are used in order to define specific characteristics, attributes or relations that describe resources. Specific resource, along with its named property and property value, makes RDF statement. Each RDF statement consists of three parts: subject, predicate and object. Due to the simplicity of the RDF syntax, it has wide use and it can be used for multimedia annotation.

RDF is independent of the domain of use and for describing specific domain RDF Schema [5] is used. A set of classes and their specific properties that define a particular domain of use can be defined with RDF Schema. Inheritance can be used in RDF Schema, so one class can become a subclass of another class. Inheritance also applies to properties, thus, one property can become a subproperty of another property.

Ontologies are formal and explicit descriptions of the concepts within a specific domain [2]. The final list of terms and concepts and relationship between those terms and concepts can be defined using ontologies. Ontologies on the Web are commonly used in web search and in defining the meaning of terms and resources on the Semantic Web.

Web Ontology Language (OWL) [6] is a descriptive language for Web ontologies used for describing properties and classes, as well as relations between classes. It is designed for use by applications that handle the content of information instead of just presenting information to the people. OWL languages provide additional formal vocabulary with added semantics that allows better communication with computers than XML, RDF and RDF Schema provide. Multimedia ontologies created using OWL enable creation of high quality multimedia metadata.

ISBN: 978-960-474-304-9

## 3 Multimedia metadata formats

There are many different standard vocabularies containing elements that describe various aspects of the image. These vocabularies differ in size, granularity and the number of elements. Usually, for a single image more than one vocabulary is used to cover all different aspects of the image. Overview of different multimedia metadata standards and formats for various forms of multimedia content is given in [7]. This chapter provides an overview of the most important standards of multimedia annotations for images and photos.

Exif (Exchangeable image file format) [8] is a standard that defines multimedia metadata formats used for describing images, audio records and tags for digital cameras and other systems using photos and audio records taken with digital cameras. Within the Exif header of the image multimedia metadata is created while taking photos. Exif tags for multimedia metadata includes tags related to image data structure (e.g., image height, image width, orientation, resolution), recording offset (e.g., image data location, number of rows per strip, bytes per compressed strip), image data characteristics (e.g., transfer function, white point chromaticity, color space transformation matrix coefficients), picture-taking conditions (e.g., exsposure time, ISO speed, lens focal length, contrast, sharpness) and general information (e.g., image title, date and time, equipment manufacturer, copryright holder). Newer digital cameras can write GPS information for location of shooting photo.

DCMES (Dublin Core Metadata Element Set) [9] is very small vocabulary containing only fifteen propertis used for describing a variety of resources on the Web. Its elements are: contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title and type. Because of universal elements this vocabulary has a very wide use and it can be used for multimedia annotation.

VRA Core (Visual Resource Association) [10] is a data standard for the description of culture heritage works, as well as photos documenting them. Unlike DCMES which defines small and frequently used elements for resources on the Web in general, VRA Core defines a small vocabulary that focuses specifically on culture heritage works. Vocabulary defines the basic elements for multimedia metadata of which some are identical or similar to elements of DCMES vocabulary. Some of the elements of the VRA Core vocabulary are date, description, location, measurements, rights, style period, subject, technique and title.

At DIG35 [11] a standard set of elements for digital photos which should improve semantic interoperability between computers and services is defined. This semantic interoperability allows for easy organization, sharing and using digital photos. Vocabulary elements are divided into five basic building blocks that provide information about basic image parameters, image creation, content description, history and intellectual property rights. Metadata properties at DIG35 standard are displayed using XML Schema.

MPEG-7 (Multimedia Content Description Interface) [12] is an international ISO/IEC standard developed by the MPEG working group, which provides important functionalities for managing and manipulating with various tyes of multimedia content and their associated metadata. This standard is suitable for use by people, but also by computers that process multimedia content. MPEG-7 provides a standardized set of descriptive tools that define the syntax and semantic of the metadata elements using descriptors and that define structure and semantics of relationships between them using description schemas. MPEG-7 uses XML format for storing multimedia metadata.

## 4 Semantic Web technologies and multimedia metadata standards integration

High quality of multimedia metadata is essential for their use in multimedia applications for personal and especially for professional use. Significant problems of multimedia metadata are very similar to general problems of ordinary metadata [13]:

1. Cost – Although some metadata can be obtained automatically from some low level features, most applications need higher level annotations that require human labor, which is an expensive and a time consuming process;
2. Subjectivity – Even with a good application for creating metadata, users often interpret those metadata different and that is especially expressed with manual annotation;
3. Restrictiveness – Metadata with strong formal semantics provide computers more relevant information, while users consider them too limited for use. On the other hand, metadata with less formal semantics are often subjective and inconsistent, so computer processing is difficult;
4. Longevity – Defining metadata that would be applicable for short and long periods, and at the same time be specific enough for use within their

ISBN: 978-960-474-304-9

domain and generic enough to be used across different domains is difficult;
5. Privacy – Metadata can include private or confidential data that require special attention;
6. Standardization – Applications for creating metadata often differ from end-user applications, which could cause a shortage of the necessary interoperability between these applications.

In addition, a major problem is the large number of different multimedia metadata standards and formats that are not compatible with one another. Scientific research group W3C Semantic Multimedia Incubator Group has been established due to the need for integration of Semantic Web technologies and various multimedia metadata standards. The goals of this group are [14]:

- Use of Semantic Web technologies for making existing multimedia metadata standards interoperable, so existing metadata formats can be combined;
- Show the added value of formal semantics on the Semantic Web with practical applications and services that provide additional functionality;
- Provide best practices for creating multimedia metadata and using multimedia content on the Web with practical use cases that identify users, type of content and type of metadata they want to enable.

W3C Semantic Multimedia Incubator Group published a report [15] demonstrating the benefits of using Semantic Web technologies for creating, storing, sharing and processing multimedia metadata. Multimedia annotation for professional use is very complex, so this report faced the following related issues:

- Production versus post-production annotation – Annotation during production is better and cheaper access because most of the information required for multimedia annotation is available in production time;
- Generic versus task-specific annotation – With generic access metadata is created without a specific context, so it will not cover all new requirements after development of the target application. Annotation for specific tasks is usually used for a single application so created metadata could be too specific for use in other applications. The best approach for annotation would be specific enough for the application, but with minimal application specific assumptions;
- Manual versus automatic annotation and the Semantic Gap [16] – The Semantic Gap represents the difference between the rich higher level descriptions obtained manually and descriptions of low level features obtained automatically;
- Different types of metadata – Metadata can describe properties of the image, but also theme and objects in the image. There are different vocabularies used for describing various aspects of the image. In most cases, it is necessary to use more than one vocabulary to create metadata for a single image;
- Lack of syntactic and semantic interoperability – Syntactic interoperability is the inability to use metadata created by one application with another application because of different syntax. Semantic interoperability is expressed by assigning different meanings and semantics at different applications for the same annotation. Both problems can be solved by using Semantic Web technologies explicitly determining syntax and semantics for annotations.

# 5 Related work

In last few years a lot of research on multimedia annotation, indexing and retrieval of multimedia content on the Web has been done. Most current approaches use Semantic Web technologies for efficiently multimedia annotation so computers can easier and effectively process that metadata.

In [17] ramm.x (RDFa-deployed Multimedia Metadata) has been proposed. Ramm.x is using RDFa with lightweight formal vocabulary for multimedia annotation on Semantic Web. RDFa is a serilization syntax of RDF data model intended for use in (X)HTML enviroments. Existing multimedia metadata that may be in various formats ramm.x associate with web services that allow converting parts or entire annotations in RDF format. Prerequisites for using ramm.x are: i) multimedia content needs to be published on the Web together with their metadata, ii) multimedia metadata is not free text or presented using RDF based ontology and iii) there is an added value from creating multimedia metadata available to a Semantic Web agent.

A generic semantic problem-solving platform for multimedia annotation is presented in [18] on famous people photos use case. Platform uses web services and various sources of semantic knowledge on the Web such as DBpedia and Freebase for findning solutions to complex requirements. Prerequisites for the above mentioned use case are available algorithms for face detection and face recognition, access to a set of rules and ontologies for images, regions and faces, and access to sources of knowledge on Semantic Web. Architecture of

platform is based on: i) blackboard that contains the current requirements and collected informations, ii) a collection of Web services with attached semantic descriptions and iii) supervisor based on compositional algorithm that generates execution plan which combines several algorithms as services. In case of an error or unexpected behavior platform finds an alternative route using Semantic Web technology that leads to the same solution.

In [19] authors propose a generic algorithm that automatically creates additional annotations in several Semantic concepts based on existing annotations. Algorithm uses different strategies based on matching terms and concepts to reduce incompleteness and inaccuracy, while creating new annotations. Using multi phase filtering process that corrects incorect annotations and regarding only annotations related to actual content of the image, inaccuary of annotations is decreased. Incompletness is descreased by extending current annotations with related and similar terms.

Approach relying on HPM (Hybrid Probabilistic Model) is proposed in [20]. HPM is used for automatic multimedia annotation for reducing problem of Semantic Gap combining image low level features such as color, texture and shape with user created higher level features metadata. If the image has user metadata, HPM integrates low level image features and user metadata to create more metadata for the image. If the image does not have any metadata, HPM generate metadata based only on image low level features.

An ontology based approach for creating and searching multimedia metadata is shown in [21]. Necessary ontologies for image annotation are specified using ontology editor Protégé-2000. For multimedia annotation on the images of apes, use case should be defined two ontologies: ontology for structure of photo annotation and domain-specific ontology. Photo annotation ontology distinguish three viewpoints: i) subject matter feature, ii) photo feature and iii) medium feature. Domain-specific ontology in this example is the animal domain that contains vocabulary and background knowledge that describes domain specific image features.

# 6 Research problems

One of the main problems for researchers dealing with multimedia annotation on the Semantic Web is the lack of widely accepted vocabularies that could be used for multimedia metadata for images and photos [15]. Some elements of different vocabularies can have the same name, but do not necessarily have the same meaning, which makes sharing multimedia metadata among different applications and domains of use difficult. The existence of widely accepted vocabularies would facilitate sharing of metadata among different domains and different applications, because then it could not come up with different interpretations of vocabulary elements.

Another problem for researchers is the difference between descriptions of low level image features and higher level image features also known as Semantic Gap [16]. Most users manually enter high level image metadata, while the descriptions of the low level features are generated automatically. Researchers are looking for automatic multimedia annotation solutions for higher level image features that would give semantic meaning to generated metadata so computers can easier process them.

The quality of annotations created by the users can also pose a problem to researchers, as this may cause misinterpretation if existing annotations are inaccurate or too subjective in later automatic generation of new annotations [13]. When creating annotations, users often interpret multimedia content differently, which can result in incosistent annotations in the same file. In addition, annotators often have different views on the content and the context within the content is used. Existence of inaccurate annotations complitcates afterwards automatic creation of new annotations based on the existing ones. Therefore additional algorithms should be used for finding and correcting inaccurate annotations before automatic creattion of new annotations.

# 7 Conclusion

During the last few years, there was a large increase in various forms of multimedia content on the Web, which presents a growing problem for the further use and search of such content. In parallel, with the increase of multimedia content on the Web existing multimedia metadata standards were improved and new standards have been developed. To facilitate the use of multimedia content on the Web, that content is assigned a metadata that describes it. Manually annotation is time-consuming and expensive process. Besides, annotations can be created by different people such as authors, editors, publishers or the end users, which represents a problem, because there may be different interpretations of those annotations. The main disadvantage of such annotations is the lack of well-defined syntax and semantics which is why computers in most cases can hardly process such information.

ISBN: 978-960-474-304-9

Choosing the right vocabulary is the key for creating semantically rich multimedia annotations. In order to obtain high quality annotations, it is usually necessary to use more vocabularies, because a single vocabulary in most cases does not contain all essential elements that can describe all different aspects of an image.

This paper presents a survey on the advantages of using Semantic Web tehnologies in multimedia annotation and retrieval. We have presented approaches and methods that show progress in the creation of semantic rich multimedia annotations. Especially the process of creating annotations has become largely automated. Despite the joint efforts of the Semantic Web and multimedia communities, multimedia annotations using Semantic Web technologies are still not fully used in practice. Since this area of research is still insufficiently explored and many questions about creation of semantic multimedia metadata are left open, there is yet a plenty of room for further work.

*References:*


[1] T. Berners-Lee, J. Hendler, O. Lassila, The Semantic Web, *Scientific American*, vol. 284, no. 5, May 2001, pp. 34-43.

[2] N. F. Foy, D. L. McGuinness, Ontology Development 101: A Guide to Creating Your First Ontology, Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.

[3] G. Stamou, J. van Ossenbruggen, J. Z. Pan, G. Schreiber, Multimedia annotations on the Semantic Web, *IEEE MultiMedia*, vol. 13, no. 1, January-March 2006, pp. 86-90.

[4] T. Berners-Lee, Semantic Web – XML2000, Talk, available at: http://www.w3.org/2000/Talks/1206-xml2k-tbl

[5] P. Hayes, RDF Semantics, W3C Recommendation, February 2004.

[6] D. L. McGuinness, F. van Harmelen, OWL Web Ontology Language Overview, W3C Recommendation, February 2004.

[7] M. Hausenblas, Multimedia Vocabularies on the Semantic Web, W3C Multimedia Semantics Incubator Group Report, July 2007.

[8] Camera & Imaging Products Association, Exchangeable image file format for digital still cameras: Exif Version 2.3, CIPA DC-008-2010 & JEITA CP-3451B Standard, April 2010.

[9] Dublin Core Metadata Initiative, Dublin Core Metadata Element Set, Version 1.1, ISO Standard 15836:2009, February 2009.

[10] Visual Resource Association Oversight Committee, VRA Core 4.0 Element Description, November 2007.

[11] DIG35 Phase 2 Initiative Group, DIG35 Specification: Metadata for Digital Images – Version 1.1 Working Draft, Digital Imaging Group, April 2001.

[12] J. M. Martínez, MPEG-7 Overview (version 10), ISO/IEC JTC1/SC29/WG11, ISO/IEC, October 2004.

[13] J. van Ossenbruggen, F. Nack, L. Hardman, That Obscure Object of Desire: Multimedia Metadata on the Web, Part 1, *IEEE Multimedia*, vol. 11, no. 4, October-December 2004, pp. 38-48.

[14] R. Troncy, J. Z. Pan, W3C Multimedia Semantics XG, Talk, available at: http://www.w3.org/2005/Incubator/mmsem/talks/SWCG2007/

[15] R. Troncy, J. van Ossenbruggen, J. Z. Pan, G. Stamou, Image Annotation on the Semantic Web, W3C Multimedia Semantics Incubator Group Report, August 2007.

[16] C. Dorai, S. Venkatesh, Bridging the Semantic Gap with Computational Media Aesthetics, *IEEE Multimedia*, vol. 10, no. 2, April-June 2003, pp. 15-17.

[17] M. Hausenblas, W. Bailer, T. Bürger, R. Troncy, Deploying Multimedia Metadata on the Semantic Web, *Poster proceedings of the 2nd International Conference on Semantics And digital Media Technologies SAMT 07*, Genova, Italy, December 2007, pp. 9-10.

[18] R. Verborgh, D. van Deursen, E. Mannens, C. Poppe, R. van de Walle, Enabling context-aware multimedia annotation by a novel generic semantic problem-solving platform, *Multimedia Tools and Applications*, vol. 61, no. 1, January 2011, pp. 105-129.

[19] M. G. Manzato, R. Goularte, Automatic annotation of tagged content using predefined semantic concepts, WebMedia '12, *Proceedings of the 18th Brazilian symposium on Multimedia and the web*, Sao Paulo, Brazil, October 2012, pp. 237-244.

[20] V. Vijay, I. J. Jacob, Combined Approach of User Specified Tags and Content-Based Image Annotation*, 2012 International Conference on Devices, Circuits and Systems (ICDCS),* Coimbatore, Tamilnadu, India, March 2012, pp. 162-166.

[21] A. T. Schreiber, B. Dubbeldam, J. Wielemaker, B. Wielinga, Ontology-Based Photo Annotation, *IEEE Intelligent Systems*, vol. 16, no. 3, May/June 2001, pp. 66-74.


ISBN: 978-960-474-304-9