Advantages of Semantic Web Technologies Usage in the Multimedia Annotation and Retrieval

Tomo Sjekavica, Gordan Gledec, and Marko Horvat

Abstract**—**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 retrieval 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. Using Semantic Web technologies such as XML, RDF and ontologies is recommended for creating new and enriching existing annotations due to a large number of multimedia metadata formats and standards and their incompatibility.

Keywords**—**image retrieval, metadata, ontologies, semantic annotation, Semantic Web

I. 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. Due to the lack of appropriate applications manual annotation is both time and money consuming process, so researchers are looking for automatic image annotation solutions. In [1] architecture of the image retrieval system and four automatic image annotation techniques for images published on the Web are shown. Author proposes automatic annotation of images by gathering annotation from the hosted web pages, from the web pages structural blocks, from anchor text through the link structure and by sharing annotation from images with same visual signature.

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 [2]. Ontologies define a list of terms and concepts and their relationships within a particular domain of use [3]. Ontologies also contain rules for using defined terms and concepts. 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) [2].

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. To solve these problems, there is a need to merge good practices in multimedia industry with the benefits of Semantic Web

T. Sjekavica is with the Department of Electrical Engineering and Computing, University of Dubrovnik, Cira Carica 4, 20000 Dubrovnik, CROATIA (phone: +385 20 445 793; e-mail: tomo.sjekavica@unidu.hr).

G. Gledec is with the University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, CROATIA (e-mail: gordan.gledec@fer.hr).

M. Horvat is with the University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, CROATIA (e-mail: marko.horvat2@fer.hr).

technologies [4]. 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.

This paper is organized as follows: in the next section Semantic web and its main components are shown and explained. In third section more information about existing standards, vocabularies and formats of multimedia metadata for images and photos are discussed, while the fourth section shows the integration of those mutlimedia metadata standards and formats with Semantic Web technologies. Overview of related researches that show various methods and approaches for creating semantically rich multimedia metadata using different Semantic Web technologies is shown in the fifth section. Then in the sixth section, last one before conclusion future chalenges in semantic multimedia annotation and our ongoing research are discussed.

II. Semantic Web

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 [2]. Semantic Web describes properties of the content and dependencies between different content, which allows unambiguous exchange of information between people and computers. The first form of semantic data on the Internet was the metadata that represent data about the data. Multimedia metadata is type of metadata used for describing multimedia content.

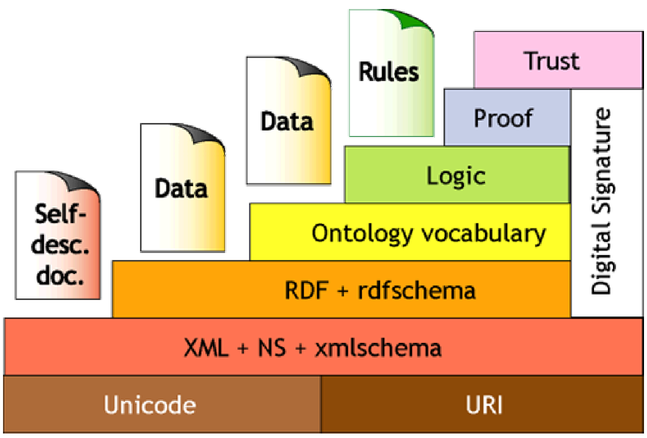

Fig. 1 Semantic Web Stack by Tim Berners-Lee [5]

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

At the bottom of the Semantic Web Stack are Unicode [6] and URI (Uniform Resource Identifier) [7]. Unicode is a universal standard for coding multilingual characters allowing easy exchange of text on a global level. Older versions of HTML supported only ISO Latin-1 character set that supports only Western European languages. Today, HTML and XML use Unicode as default standard for text coding, which allows use of larger set of characters. All content on the Web can be defined by URI providing simple and expandable meaning which identifies particular resource on the Web.

A. *XML and XML Schema*

XML is placed on the second layer of Semantic Web stack. Using XML, users can create their own tags for structured web documents. Tags in the XML document can be nested. 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 the XML documents. Syntax of newer languages for exchanging data on the Web is mostly XML based.

XML Schema [8] is a language used to define the structure of the XML documents. Its syntax is XML based. Two applications that want to communicate with each other can use the same vocabulary or the same definition of the structure of an XML document, which is provided in the related XML Schema.

B. *RDF and RDF Schema*

RDF [9] 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. RDF is located above XML layer in the Semantic Web Stack. Resources, properties and statements are three main concepts of RDF. Anything that can be identified by URI is a resource. Properties are used in order to define specific characteristics, attributes or relations that describe resources. Properties can also be defined by URI. Specific resource, along with its named property and property value, makes an RDF statement. Each RDF statement consists of three elementary 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.

Graph representation of RDF statement in a form of RDF triple which includes subject, predicate and object is shown in Fig. 2.

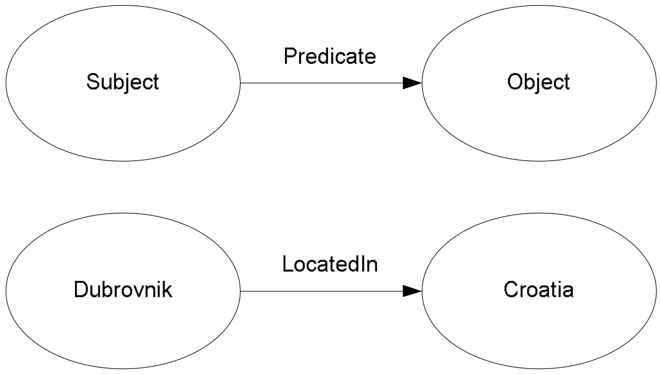

Fig. 2 Graph representation of RDF triple

RDF is independent of the domain of use and for describing specific domain RDF Schema [9] 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.

C. Ontologies

Ontologies are formal and explicit descriptions of the concepts within a specific domain [3]. The final list of terms and concepts and relationship between those terms and concepts can be defined using ontologies. With XML and RDF ontologies represent the third major component of the Semantic Web. Ontologies on the Web are commonly used in web search and in defining the meaning of terms and resources on the Semantic Web.

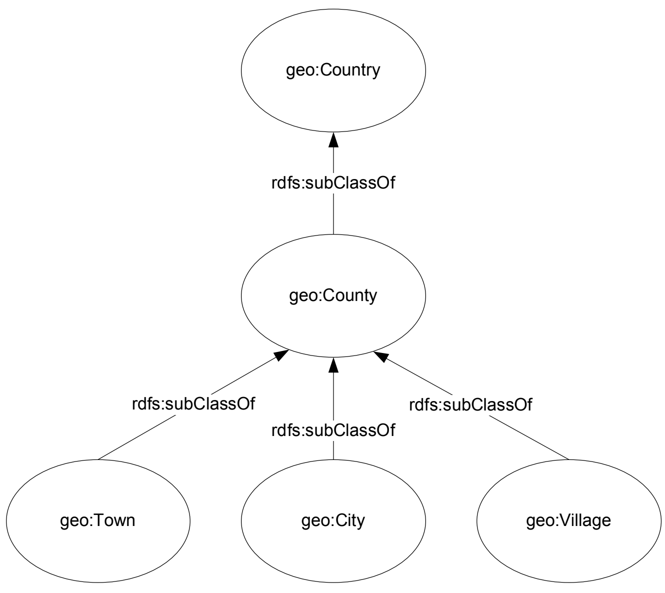

Fig. 3 Hierarchy of ontology classes

Relationships in ontologies usually include hieararchy of concepts (classes), which specifies that a class C1 can be subclass of another class C2 if every object in class C1 is included in the class C2. In Fig. 2 is shown an example for the hieararchy of ontology classes in geographical domain. In this example one country can be divided into counties, and those counties can contain towns, cities and villages.

D. OWL

Web Ontology Language (OWL) [10] is a formal and descriptive language for Web ontologies used for describing properties and classes, as well as relations between classes. Characteristics of properties can also be defined by OWL. Description Logic (DL) [11] provides a formal basis for the definition of the OWL. It is designed for use by applications that handle the content of information instead of just presenting information to the people.

Scientific research group W3C Web Ontology Working Group has defined three different types of OWL languages [10]:

- OWL Lite contains simple constraints and the classification hierarchy. It is used for simple ontologies;
- OWL DL has maximum expressiveness with the restriction that all conclusions can be computed and that all calculations can be completed in a finite time. It is used for expressive ontologies;
- OWL Full has a maximum expressiveness and the syntatic freedom, but with no guarantee of computation. It is used when compatibility with RDF and RDF Schema is primary.

OWL DL sublanguage is an extension of OWL Lite sublanguage, while OWL Full sublanguage is an extension of OWL DL sublanguage. Those extensions refer in what can be validly concluded and legally expressed from its simpler predecessor. Following relations hold for OWL sublanguages, but their inverses do not [10]:

- Every legal OWL Lite ontology is a legal OWL DL ontology.
- Every legal OWL DL ontology is a legal OWL Full ontology.
- Every valid OWL Lite conclusion is a valid OWL DL conclusion.
- Every valid OWL DL conclusion is a valid OWL Full conclusion.

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.

III. 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 needs to be 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 [12]. This chapter provides an overview of the most important standards of multimedia annotations for images and photos.

A. Exif

Exchangeable image file format (Exif) [13] 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 of image, image resolution in height direction, image resolution in width direction), 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., exposure time, ISO speed, lens focal length, contrast, sharpness) and general information (e.g., image title, date and time, equipment manufacturer, copyright holder). Newer digital cameras can write GPS information for location of shooting photo (e.g., GPS tag version, latitude, longitude, North or South latitude, East or West longitude, GPS time).

B. DCMES

Dublin Core Metadata Element Set (DCMES) [14] is very small vocabulary containing only fifteen properties 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. Those fifteen elements are part of larger set of technical specifications and metadata vocabularies that are mainted by Dublin Core Metadata Initiative (DCMI). Because of universal elements this vocabulary has a very wide use and it can be used for multimedia annotation.

C. VRA Core

Visual Resource Association Core (VRA Core) [15] 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, inscription, location, material, measurements, rights, state edition, style period, subject, technique, agent, work type and title.

In latest version of VRA Core 4.0 third type of record for collections has been added to the two existing types of records: works and images. A work presents a unique entity such as cultural event or object, like sculpture, building or painting. Image presents a visual representation of part of the work or the whole work. The collection is a set of works or images, which allows collection-level cataloging. VRA Core 4.0 uses XML and XML Schema for displaying its metadata elements.

D. NISO Z39.87

Standard NISO Z39.87 [16] defines a set of elements for raster digital images metadata to allow users development, exchange and interpretation of digital images. Vocabulary elements are covering a wide range of metadata for images such as basic digital object information, basic image information, image capture metadata, image assessment metadata and change history. Vocabulary is designed to facilitate interoperability between systems, services and applications, as well as uninterrupted access to collections of digital images. This standard is independent of the image file format.

E. DIG35

At DIG35 [17] 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: i) basic image parameters, ii) image creation, iii) content description, iv) history and v) intellectual property rights (IPR). Fundamental metadata types and fields define the forms of the fields defined in above mentioned building blocks. Metadata properties at DIG35 standard are displayed using XML Schema.

F. MPEG-7

MPEG-7 [18] is an international ISO/IEC standard developed by the MPEG working group (Motion Picture Experts Group), which provides important functionalities for managing and manipulating with various types of multimedia content and their associated metadata. MPEG-7 is formally named Multimedia Content Description Interface. This standard is suitable for use by people, but also by computers that process multimedia content, and it is not aimed to any particular application.

MPEG-7 provides a standardized set of descriptive tools that define the syntax and semantic of the metadata elements using Descriptors (Ds), and that define structure and semantics of relationships between them using Description Schemas (DSs). Syntatic rules for creating, combining, refining and extending MPEG-7 descriptive tools Ds and DSs are provided using Description Definition Language (DDL) [19]. MPEG-7 uses XML format for storing multimedia metadata and XML Schema as schema for MPEG-7 DDL. 1182 elements, 417 attributes and 377 complex types are defined in the MPEG-7 XML Schemas.

Profiles are used in order to reduce the complexity of MPEG-7 descriptions. Three standardized MPEG-7 profiles are: Simple Metadata Profile (SMP), User Description Profile (UDP) and Core Description Profile (CDP). SMP can be used in images, music and mobile applications arreas for creating simple metadata for single instances of image, audio or video clip. UDP describes user personal preferences and usage history of multimedia content, thus enabling discovering, selecting and recommendation of multimedia content automatically. CDP allows describing different multimedia content such as images, audio and video, as well as collections of such multimedia content.

G. MXF

Material Exchange Format (MXF) [20] is an open file format whose primary purpose is the exchange of multimedia content with its metadata. MXF presents a wrapper that is used for encapsulation of multimedia content in the form of still images, video and audio clips. This format is independent of audio and video coding of the source multimedia files. MXF header contains enough structural information that allows applications sharing the essence of multimedia content without any a priori information. MXF metadata is divided in two categories: structural and descriptive. Structural metadata defines the picture size, aspect ratio, picture rate and other essence description parameters for images, audio or video clips. Descriptive metadata is created during the planning of production or during the production.

H. SVG

Scalable Vector Graphics (SVG) [21] is a modular language for describing two-dimensional vector and mixed vector-raster graphics in XML format. Three types of graphical objects SVG allows to describe are various forms of vector graphics, images (raster graphics) and text. For those objects style can be changed and objects can be grouped and transformed into previously rendered objects. SVG drawings can be dynamic and interactive. Animations can be defined by embedding elements of SVG animation within SVG content or by scripting. Metadata within the SVG content are defined by DCMES vocabulary elements.

IV. Semantic Web technologies and multimedia 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 the general problems of ordinary metadata [22]:

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 – Longevity is problem with all electronic documents. Defining metadata that would be applicable for short and long periods, and at the same time be specific enough for use within their 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, such as metadata of medical documents that can contain personal information of the patient;
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. In order to solve above mentioned problems Semantic Web technologies can be used. 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 [23]:

- 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 such as using rule-based approaches;
- 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 [24] 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 [25] – 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.

In another report about multimedia vocabularies on the Semantic Web [12] same scientific research group has shown an overview of different formats and vocabularies for multimedia annotation, that can be distributed by type and the category of multimedia content. This report is focused on integration of those multimedia vocabularies into the Semantic Web.

Relevant multimedia retrieval systems with semantic approach are discussed in [26]. Semantic approach deals with:

1) annotation, relevant feedback and concept based multimedia retrieval systems,
2) ontology mediated multimedia retrieval systems,
3) intrinsic semantic framework for recognizing image objects, and
4) adaptive architecture for automatic multimedia retrieval composition of media ontologies with domain specic ontologies.

For retrieval of the images on the Web two methods can be used: text-based imege retireval and content-based image retrieval (CBIR). Semantic Web technologies can be used in multimedia retrieval systems in order to achieve better understanding of multimedia content with its metadata and for bridging the Semantic Gap. Ontology based image retrieval approach based on the content of image is proposed in [27]. That approach allows better understanding of semantic content and aims to provide better standardization of image metadata.

V. 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 [28] RDFa-deployed Multimedia Metadata (ramm.x) has been proposed. Ramm.x is using RDFa [29] with lightweight formal vocabulary for multimedia annotation on Semantic Web. RDFa is serialization syntax of RDF data model intended for use in (X)HTML environments. Existing multimedia metadata that may be in various formats ramm.x associates with web services that allow converting parts or entire annotations in RDF format. Prerequisites for using ramm.x are that: i) multimedia content needs to be published on the Web together with their metadata, ii) multimedia metadata is not in form of 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 [30] on famous people photos use case. Platform uses web services and various sources of semantic knowledge on the Web such as Dbpedia (http://www.dbpedia.org/) and Freebase (http://www.freebase.com/) for finding 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. Supervisor connects blackboard with Web 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 [31] authors propose a generic algorithm that automatically creates additional annotations in several Semantic concepts based on existing manually created 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 incorrect annotations and regarding only annotations related to actual content of the image, inaccuracy of annotations is decreased. Incompleteness is decreased by extending current annotations with related and similar terms.

Approach relying on Hybrid Probabilistic Model (HPM) is proposed in [32]. 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 [33]. Necessary ontologies for image annotation are specified using ontology editor Protégé-2000 [34]. 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 distinguishes three viewpoints: i) subject matter feature, ii) photo feature and iii) medium feature. Subject matter feature is part of photo annotation ontology that connects that ontology with domain-specific ontology. Photo features define metadata from which can be learn when, how and why the photo was taken. Medium feature determine the way how the photo is stored, like a file format or photo resolution. Domain-specific ontology in this example is the animal domain that contains vocabulary and background knowledge that describes domain specific image features.

System for automatic annotation of multimedia files, which is based on implicit user interaction via search engines is proposed in [35]. Authors present a framework which is based on associating user query keywords to selected multimedia content. That framework is recording and analyzing every user search session. When user submits a query to the search engine related images are retrieved. Then user selects some of those retrieved images. Then each selected image is assigned query keywords according to a proper keyword frequency measure and to a voting scheme. With this method user perception is implicitly included in automatic image annotation.

VI. Future challenges and ongoing research

Lack of widely accepted vocabularies that could be used for multimedia annotation of images and photos is still one of the main unresolved challenges for the researchers dealing with multimedia annotation on the Semantic Web [24]. 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 challenge for researchers is the difference between descriptions of the low level image features and higher level image features also known as Semantic Gap [25]. Most users manually enter high level image metadata, while the descriptions of the low level features are generated automatically. Manual annotation is time consuming and tedious process for users, which is why 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. This would facilitate the use and retrieval of multimedia content on the Web so search results for user's queries would be semantically richer.

The quality of annotations created by the users is also one of the future challenges for the researchers, as this may cause misinterpretation if existing annotations are inaccurate or too subjective in later automatic generation of new annotations [22]. When creating annotations, users often interpret multimedia content differently, which can result in inconsistent annotations in the same file. In addition, annotators often have different views on the content and the context within the content is used, which is problem to end users at a later use and interpretation of those annotations. Existence of inaccurate annotations complicates 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 creation of new annotations.

Our ongoing research is directed towards the multimedia on the Semantic Web and the practical use of the acquired knowledge for multimedia annotation of photos and images in order to deal with above mentioned problems and challenges. We are enquiring multimedia ontologies and their use in semantic annotation of photos and images in order to develop multimedia ontology that would be suitable for creation of semantically rich multimedia annotations.

VII. Conclusion

Due to the progressively increasing amount of multimedia on the Web, the need for efficient metadata formats describing that content has become increasingly evident. Multimedia metadata is a type of metadata used for describing different aspects of multimedia content that can be in form of pictures, video or audio files. All formats of multimedia metadata are not compatible with each other and most of it do not provide enough semantics. New Semantic Web technologies like RDF and ontologies provide well-defined information meaning so different multimedia metadata can be more easily combined and processed by computers and applications.

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 technologies in multimedia annotation and retrieval of multimedia content. We have presented different approaches and methods that show progress in the creation of semantically rich multimedia annotations. Especially the process of creating annotations has become largely automated. Despite the joint efforts of the S-emantic 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 semantically rich multimedia metadata and semantic retrieval of multimedia content are left open, there is yet a plenty of room for further work.

REFERENCES

[1] H.-T. Chang, "Automatic web image annotation for image retrieval systems", *Proceedings of the 12th WSEAS international conference on Systems* (ICS'08), Stevens Point, Wisconsin, USA, 2008, pp. 670-674.

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

[3] 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.

[4] 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.

[5] T. Berners-Lee, "Semantic Web – XML2000", Talk, [Online]. Available: http://www.w3.org/2000/Talks/1206-xml2k-tbl

[6] The Unicode Consortium, "The Unicode Standard Version 6.2 – Core Specification", Unicode Inc., Mountain View, California, USA, September 2012.

[7] T. Berners-Lee, R. Fielding, L. Masitner, "Uniform Resource Identifier (URI): Generic Syntax", Network Working Group, January 2005. [Online]. Available: http://www.ietf.org/rfc/rfc3986.txt

[8] H. S. Thompson, D. Beech, M. Maloney, N. Mendelsohn, P. Hayes, XML Schema Part 1: Structures Second edition, W3C Recommendation, October 2004. [Online]. Available: http://www.w3.org/TR/xmlschema-1/

[9] P. Hayes, "RDF Semantics", W3C Recommendation, February 2004. [Online]. Available: http://www.w3.org/TR/rdf-mt/

[10] D. L. McGuinness, F. van Harmelen, "OWL Web Ontology Language overview", W3C Recommendation, February 2004. [Online]. Available: http://www.w3.org/TR/owl-features/

[11] F. Baader, D. Calvanese, D. L. McGuinness, D, Nardi, P. F. Patel-Schneider, *The Description Logic handbook: Theory, implementation and applications*, 2nd ed., Cambridge University Press, 2010.

[12] M. Hausenblas, "Multimedia vocabularies on the Semantic Web", W3C Multimedia Semantics Incubator Group Report, July 2007. [Online]. Available: http://www.w3.org/2005/Incubator/mmsem/XGR-vocabularies/

[13] 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.

[14] Dublin Core Metadata Initiative, "Dublin Core Metadata element set", Version 1.1, ISO Standard 15836:2009, February 2009. [Online]. Available: http://dublincore.org/documents/dces/

[15] Visual Resource Association Oversight Committee, "VRA Core 4.0 element description", November 2007.

[16] National Information Standards Organization, "Data dictionary – technical metadata for digital still images", ANSI/NISO Z39.87-2006 (R2011) Standard, May 2011.

[17] DIG35 Phase 2 Initiative Group, "DIG35 specification: metadata for digital images – version 1.1 working draft", Digital Imaging Group, April 2001.

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

[19] J. Hunter, "An overview of the MPEG-7 description definition language (DDL)", *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 11, no. 6, June 2001, pp. 765-772.

[20] Society of Motion Picture and Television Engineers, "Material Exchange Format (MXF) – File Format Specification - Amendment 1", ST 377-1:2011 Am1:2012 Standard, 2012.

[21] E. Dahlström, P. Dengler, A. Grasso, C. Liley, C. McCormack, D. Schepers, J. Watt, "Scalable Vector Graphics (SVG) 1.1 (second edition)", W3C Recommendation, August 2011. [Online]. Available: http://www.w3.org/TR/SVG11/

[22] 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.

[23] R. Troncy, J. Z. Pan, "W3C Multimedia Semantics XG", Talk, [Online]. Available: http://www.w3.org/2005/Incubator/mmsem/talks/SWCG2007/

[24] R. Troncy, J. van Ossenbruggen, J. Z. Pan, G. Stamou, "Image annotation on the Semantic Web", W3C Multimedia Semantics Incubator Group Report, August 2007. [Online]. Available: http://www.w3.org/2005/Incubator/mmsem/XGR-image-annotation/

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

[26] C. Pino, R. Di Salvo, "A survey of semantic multimedia retrieval systems", *Proceedings of the 13th WSEAS international conference on Mathematical and computational methods in science and engineering* (MACMESE '11), Stevens Point, Wisconsin, USA, 2011, pp. 353-358.

[27] W. Zheng, Y. Ouyang, J. Ford, F. S. Makedon, "Ontology-based image retrieval", *Proceedings of Proceedings of WSEAS MMACTEE-WAMUS-NOLASC*, Vouliagmeni, Athens, Greece, 2003

[28] 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.

[29] B. Adida, M. Birbeck, S. McCarron, I. Herman, "RDFa Core 1.1: Syntax and processing rules for embedding RDF through attributes", W3C Recommendation, June 2012. [Online]. Available: http://www.w3.org/TR/rdfa-core/

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] N. F. Noy, R. W. Fergerson, M. A. Musen, "The Knowledge Model of Protégé-2000: Combining Interoperability and Flexibility", *Knowledge Enineering and Knowledge Managment Methods, Models and Tools: 12th International Conference, EKAW 2000, Lecture notes in Artifical Intelligence*, vol. 1937, Springer, Paris, France, October 2000, pp. 17-32.

[35] K. Ntalianis, A. Doulamis, N. Tsapatsoulis, N. Doulamis, "Automatic annotation of multimedia content by user clickthroughs: enhancing the performance of multimedia search engines", *Proceedings of the 10th WSEAS international conference on Mathematical methods, computational techniques and intelligent systems* (MAMECTIS'08), Stevens Point, Wisconsin, USA, 2008, pp. 439-446.