A Flexible Approach for Managing Digital Images on the Semantic Web

Christian Halaschek-Wiener[1], Andrew Schain[2], Jennifer Golbeck[1], Michael Grove[1], Bijan Parsia[1], and Jim Hendler[1]

[1] University of Maryland, MIND Lab, 8400 Baltimore Ave., College Park, MD 20742, USA
[2] NASA Headquarters, Washington, DC 20546, USA
{halasche,golbeck,hendler}@cs.umd.edu, {andrew.schain}@nasa.gov mhgrove@hotmail.com, bparsia@isr.umd.edu

Abstract. As the volume of digital images available on the Web continues to increase, there is a clear need for more advanced techniques for their effective retrieval and management. Recently, there has been an interest in applying Semantic Web technologies to represent the high level content of digital images in a machine processable format. While progress has been made, through a representative use case, we provide motivation for further work in developing more domain independent techniques for both annotating and managing images on the Web. Following this, we present an approach for publishing (OWL) annotations of image content to the Semantic Web, through the loose coupling of an annotation environment with a Semantic Web portal. Additionally, we present an implementation of the approach and describe a hypothetical use case that resulted in a proof-of-concept designed in collaboration with NASA.

1 Introduction

As the scale and infrastructure of the Internet have dramatically increased over the past years, we have seen the incorporation of various digital media types onto the Web, including images, video, and audio. As production of digital media content continues to grow in the commercial and home use markets, and as Internet access and wider bandwidth become even more pervasive, we can anticipate a continued increase of these complex (non-textual multimedia) data types being made available on the Web. Due to the format of such media, standard indexing techniques commonly used on text-based Web content, such as keyword-based approaches [6], are of little use. Given the volume of unstructured digital media, it is clear additional approaches and techniques must be developed to allow for their effective management and accurate retrieval.

Over the past few years, various approaches have been proposed to effectively retrieve and manage digital image content on the Web. Traditionally, these have included techniques such as building keyword indices based on image content [15, 17], embedding keyword-based labels into images [15], analyzing text immediately surrounding images on Web pages [9], etc. More recently, with the advent of the Semantic Web [3], there has been a research focus, commercially, in academia, and elsewhere, to develop techniques to annotate the content of images on the Web, using Web ontology languages such as RDFS [5] and OWL [2].

Recent efforts have largely focused on mapping low-level features of images to ontological concepts [1,7,19] and have involved the development of tools that are closely tied to domain specific ontologies for annotation purposes [12,14,16] (see Section 5 for additional details). Additionally, past approaches have largely left unaddressed image metadata management and advanced interaction (browsing and search capabilities) that is enabled by employing Semantic Web technologies. While substantial progress has been made, we see the need for further work in defining a more generic approach for annotating and managing digital images on the Web.

In this work, we present an approach that provides generic, domain independent flexibility for publishing annotations of digital image content to the Semantic Web, as well as a mechanism for managing such annotations through a highly customizable, ontology-backed Semantic Web portal. Through the loose coupling of the annotation and management components of our approach, a seamless environment is provided in which users can annotate, share, and manage their digital images on the Semantic Web. Additionally, we present an open source implementation of the framework and describe a hypothetical use case of both the approach and implementation based on discussions that resulted in a proof-of-concept designed in collaboration with the National Aeronautics and Space Administration (NASA).

The remainder of the paper is organized as follows: Section 2 provides some initial motivations and an overview of the proposed approach. Section 3 describes the current implementation of the approach. Section 4 proceeds to present a discussion of the approach. Section 5 presents related work and lastly Section 6 concludes the paper.

2 Motivation and Approach Overview

To understand the generic requirements that have driven the approach presented here, a representative use case based on a subset of NASA requirements is provided. While this motivation is presented in the context of NASA, we feel the model is sufficiently generic, thus capturing the general issues associated with managing metadata of digital images.

As an enterprise, NASA has hundreds of thousands of images, stored in different formats and locations, at different levels of availability and resolution, and with associated descriptive information at various levels of detail and formality. NASA also generates thousands of images on an ongoing basis that are collected and cataloged, often in accordance with needs of the image creator's specific disciplines and domain (preliminary investigators, mission specialists, public affairs, etc.). It is clear that a mechanism is needed to catalog all the different types of image content across different domains. Information is required about both the image itself (creation date, dpi, source, etc.) and also about the content of the picture (contains a satellite, astronaut, etc). The associated metadata must be maintainable and extensible so associated relationships between images and data can evolve cumulatively within a discipline or branching into other disciplines. The service must be available to a global consumer population but should be flexible enough to enforce restriction based on content type, ownership, authorization, or time (we note here we do not address the Web-based policies, as it is out of the scope of this publication).

A promising strategy for such image management requirements is an annotation environment that enables both providers and users to annotate information about images or regions in images using concepts in ontologies (OWL and/or RDFS). Thus, subject matter experts and consumers (regardless of their location) will be able to assert metadata elements about images and publish their annotations to the Semantic Web. There, such digital image annotations can be harvested and merged, resulting in advanced browsing, searching, and management.

We generalize these (NASA specific) high level requirements into the following application independent requirements: support for *adhoc* ontology-based annotation of images on the Web, enabling support for annotation with respect to *any* domain; the ability to make assertions about images and the contents of specific regions in images; the ability to automatically publish annotations to the Semantic Web, where they can be shared, indexed, and maintained; provide a metadata management facility for interacting with and maintaining image metadata that is accessible to a global community – the Semantic Web; the ability to accumulate metadata about a specific image over a period of time from different sources. To the best of our knowledge, there has not yet been a seamless integration of *all* these capabilities (details provided in Section 6).

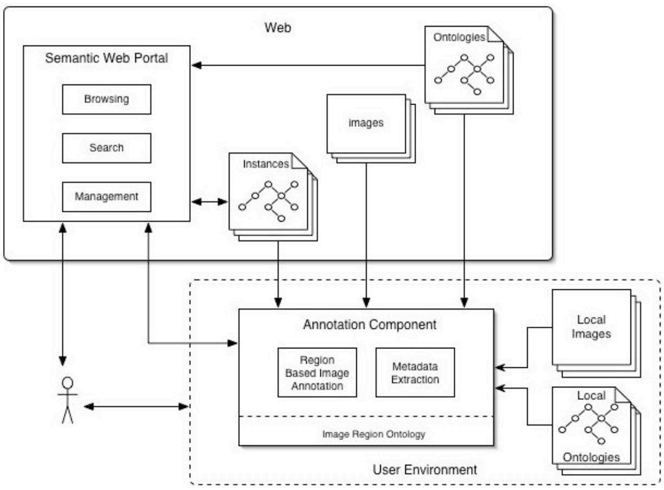

Fig. 1. Image Annotation and Management Approach

Given these requirements, we present a loosely coupled approach (depicted above in Figure 1) that provides generic, domain independent flexibility for creating and publishing annotations of digital image content to the Semantic Web, as well as a mechanism for managing such annotations through a highly customizable, ontology-backed Semantic Web portal. By loosely coupled, we refer to allowing ontologies and instance knowledge bases (KBs) to be used in an interactive and ad hoc manner during image annotation. This allows users to utilize predefined concepts and instances on the Semantic Web in their image annotations. Additionally, resulting annotations can be published to the Semantic Web for future use, where (in the context of this work) they are maintained and managed by a Semantic Web portal.

3 Implementation Details

The first component of the approach presented in this work is a digital image annotation environment. Currently, a prototype, PhotoStuff[1], implementing the annotation capabilities presented in the Figure 1 has been deployed. The following section provides an overview of PhotoStuff, as well as the digital annotation component of the approach in general.

3.1 Digital Image Annotation – PhotoStuff

PhotoStuff is a platform independent (Java-based), open source, image annotation tool that allows users to annotate an image and its regions with respect to concepts from any number of ontologies specified in RDFS or OWL (note that this is an implementation of the annotation component depicted in Figure 1). PhotoStuff provides functionality to import images (and their embedded metadata), ontologies, instance-bases, perform markup, and export the resulting annotations. The tool provides users the ability to load multiple OWL and/or RDFS ontologies, allowing annotation of image content with respect to any concept, defined in any number of ontologies. The ability to annotate images with respect to any ontology is extremely important; this is due to the fact that the content of images can span multiple domains, thus a single ontology often times can not capture the complexity of the content. Thus the approach presented here is completely compatible with the Semantic Web, which heavily hinges on the development of multiple ontologies by various individuals, spanning many domains.

In PhotoStuff, an ontology-based approach has also been adopted in order to make statements regarding the high level concepts depicted in images. An ontology is used to provide the expressiveness required to assert what is depicted within an image, as well information about the image itself (date created, etc.). In this work, an image-region ontology[2] has been specified, using OWL, which defines a set of concepts (and their relations) for images, videos, regions, and depictions.

To demonstrate the use of PhotoStuff, Figure 2 shows a screenshot of the tool in which as user is marking up information about an astronaut taking a space walk. The ontologies are visualized in both a class tree and list, depicted in the far left pane of the tool. In this example, the FOAF (Friend of a Friend) ontology has been loaded, as well as a Shuttle Crew ontology that is expanded in the window. This allows the user to choose concepts from both ontologies to mark up the photograph and its sub-regions. In the example, the Shuttle Crew ontology includes relations between classes such as "hours in space", which (in this example) can be combined with data repre-

[1] PhotoStuff Homepage: http://www.mindswap.org/2003/PhotoStuff/
[2] Image-Region Ontology: http://www.mindswap.org/2005/owl/digital-media

sented in FOAF, such as social network information. Any number of additional ontologies could also be used.

The approach presented here provides region based image annotation. Using a variety of region drawing tools, users are able to highlight regions around portions of images loaded in PhotoStuff. Figure 2 illustrates this with a region drawn around the astronaut. The classes listed in both the tree and list can be dragged into any region, or into the image itself, creating a new instance of the selected class. An instance creation form is dynamically generated from the properties of the selected class (range restrictions are imposed). With region support, metadata can be more closely tied to the depiction it describes. Instead of simply stating that a photograph depicts several people, the metadata will contain coordinates for the regions of the photo that contain the depictions. In Figure 2, the astronaut (Storey Musgrave) has been asserted to be a "Payload Commander", and some information about him has been entered in the form on the right. A full view of the metadata will show that the region depicts this instance. The region is also semantically linked to the image, maintaining the connection between the image and the instance.

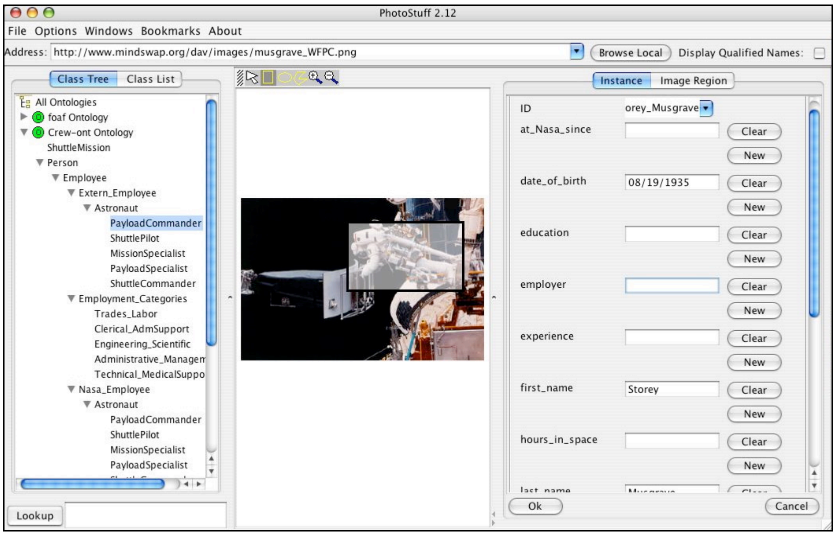

Fig. 2. PhotoStuff Screenshot

Existing instances can be loaded from any URI that references a RDF/XML document on the Web. Using these preloaded instances, depictions can reference existing instances (a drop down list of instance IDs can be used to select the desired instance). For example, if someone were adding information about the astronaut depicted in Figure 2, it would be possible to load Semantic Web data from a NASA website. That information can be tied to a region or the photograph as a whole, pre-populating the forms with data. The user can then add more data or modify the contents.

Additionally, the approach presented here leverages current efforts in multimedia format standardizations that provide support to embed image metadata in actual image files. For example, the JPEG [10] file format provides support for embedding a standard set of markers in the file header, defining metadata elements including file size, width/height, pixel density, etc [10]. Further, there are extensions to this element set, such as the Exchangeable Image File Format (EXIF), which include camera specific information (camera make, model, orientation, etc.)[3]. The approach here takes advantage of such existing metadata by extracting and encoding this information into RDF/XML, thus allowing embedded metadata to be directly incorporated into the framework presented here and the Semantic Web in general.

As mentioned earlier, PhotoStuff, and the approach in general, maintains a loose coupling with a Semantic Web portal. As briefly discussed before, there are three ways in which PhotoStuff interacts with the portal: retrieving all instances that have been submitted to the portal, submitting generated RDF/XML, and uploading local images so they can be referenced by a URI (thus allowing them to be referenced using RDF/XML). The following section outlines the metadata management and browsing functionality provided though the loose coupling of the annotation environment with the Semantic Web portal.

3.2 Image Metadata Management

Upon the completion of image annotation, the approach provides the capability for publishing resulting markup to the Semantic Web. This is accomplished through the coupling of the annotation environment with an ontology-backed, Semantic Web portal. Our existing work on a Web portal based on Semantic Web technologies (OWL) has been extended to provided communication with PhotoStuff. It is noted here that the Web portal's functionality extends what is presented here and is an ongoing project within the MINDSWAP[4] research group. Details are provided here regarding the portal implementation through one of its configurations in the context of a proof-of-concept, SemSpace[5], developed as an experiment with NASA. It is noted here that a variety of other domain configurations have been developed at MINDSWAP, including SWINT[6] (counter-terrorism), and the MINDSWAP research group homepage. All configurations provide the same functionality, only differing by the ontologies and instances maintained by the system.

The portal technology is flexible enough to be used in a variety of domains, as it is not limited in the number of ontologies that it can manage; thus for the purpose of this work, any ontology can be used to annotate an image. The portal is designed to use information from the various ontologies to guide the display of and interaction with metadata and the site in general. The main interface for browsing images is driven by the underlying class of each instance, thus providing a high level view of all the metadata of images that have been annotated using PhotoStuff.

When an instance is selected, the user is presented with all images in which the instance is depicted (illustrated below in Figure 3). All of the metadata regarding that instance is presented to the user as well (currently in tabular form). In addition, re-

[3] EXIF Homepage: http://www.exif.org/
[4] MINDSWAP Research Group: http://www.mindswap.org/
[5] SemSpace Homepage: http://semspace.mindswap.org/
[6] SWINT Homepage: http://swint.mindswap.org/

gion based annotations of images can be browsed. Since existing instances can be used during the annotation process, images are linked to pre-existing metadata. Thus, both metadata published via the image and additional sources can be integrated and browsed through this environment. This presents a unique way in which the data is visualized and interacted with. Using these capabilities, the user can browse the content of images by traversing various regions and/or following links through associated metadata related to the image/region. Note from Figure 3, that specific regions are highlighted. By selecting an image region, the various co-regions of the selected image region are displayed (also shown in Figure 3). This allows browsing of the metadata associated with the various regions depicted in the image. Additionally, the portal provides support for searching image metadata. Currently, images are searchable/retrievable at the instance and class level via keyword indexes built from the instance data.

Lastly, the portal component provides various management capabilities. Metadata submissions can be audited, edited, or removed. Due to the distributed manner in which multiple users can annotate images, it is common for duplicate instances to be created. Because the portal is based on Semantic Web technologies, such problems are easily maintained through management interfaces (in this case, duplicate entities can be equated using *owl:sameAs*). Lastly, provenance information (submitter name, email, etc.) from all submissions is maintained and editable.

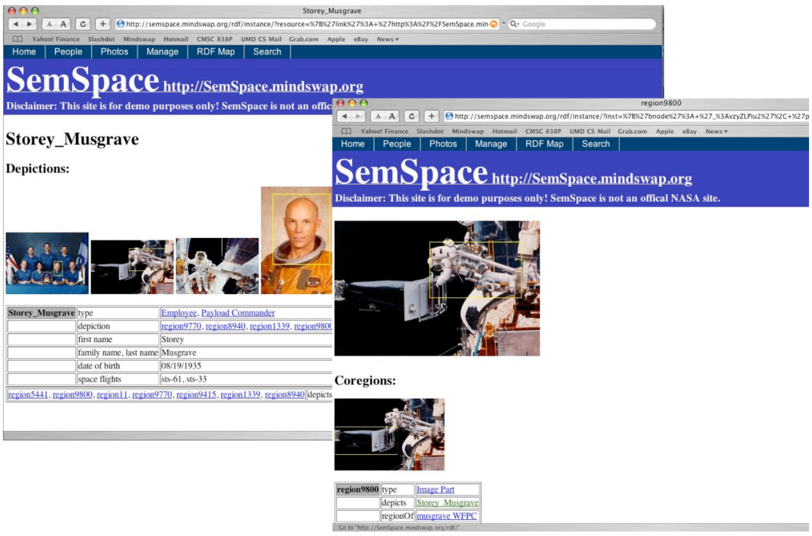

Fig. 3. Instance Depictions and Co-Region Browsing

4 Discussion and Future Directions

The approach discussed in this paper allows ad hoc, manual annotation of image content. This provides a cumulative technique where metadata can be incrementally added or repurposed for future users on a per-need basis. While manual annotation is essential for such ad hoc additions or edits, it can prove to be quite time consuming. This may be slightly alleviated through use of various image processing and automated vision techniques. First, region segmentation techniques may be used to suggest possible regions of interest. Thus, users may avoid having to manually draw regions on the images. Additionally, image-processing techniques could potentially be used to recognize similar regions (based on low level image features), allowing the tool to suggest potential instances that may be depicted in the image. The intuition here is that images loaded from personal use, may be cataloged as albums. In many cases, each "film role" may contain multiple pictures of the same objects, e.g., the same person [18]. This can be exploited through the use of automated recognition techniques. Once images are automatically labeled, users can then simple verify the resulting annotations [18].

Additionally, in this work it has been observed that generating effective, yet generic forms based on class definitions can be quite difficult (in this context, instance creation forms are generated when classes are dragged into image). We have adopted an approach in which the form is directly built from the underlying properties of the class. While this approach is a plausible first step, it can result in a very messy or congested form. We would like to explore allowing the user to create custom forms for classes. Additionally, we would like to investigate allowing ontology creators to embed HTML forms or XForms[7] into comments on class definitions.

5 Related Work

There has been recent work in annotating image content with respect to ontological concepts. In [11] an approach for image annotation using ontologies specified in RDFS is presented. In the work, a photo annotation ontology (capturing subject matter, medium, and photo features) and domain specific ontology are used to annotate images. In the paper, a mapping is used to link both ontologies, thus allowing them to be used for annotation purposes. A tool is presented which supports the annotation of images. In their follow on work, [12], an extension to their previous approach is presented in which four ontologies are supported. Again, a mapping between the ontologies is used to enable annotation. Additionally, the annotation template is restricted to these four domain ontologies. Our work here differs from these two approaches in that we allow annotation with respect to *any* ontology and do not require a user to provide a mapping between an annotation template ontology and domain ontologies. Additionally, our approach allows region-based annotations, full support of OWL, and provides a Web based management environment. [19] presents the M-OntoMat-Annotizer, which provides ontology based image (and video frame) annotation (at both the image and image-region level). Additionally, the tool supports automatic, low-level MPEG-7-based feature extraction from annotated regions, thus providing visual descriptors of the annotated regions. While, our current approach does not provide this type of functionality, we view it as a future work (as discussed in Section 4). Our work differs from [19], in that our annotation environment is coupled with the

[7] XForms 1.0: http://www.w3.org/TR/xforms/

Semantic Web, thus providing automated functionality to publish image annotations to the Web. Additionally, our approach exploits existing metadata embedded in images loaded into the annotation environment.

In [14], an approach is presented in which three domain specific RDF schemas are used to annotate digital images. Resulting RDF/XML is then embedded in the header of the image files (only supports JPEG file format). Our approach differs in that we do not restrict the ontologies for annotation, support OWL, and do not embed the resulting annotation into the actual file. Additionally, we provide functionality for region-based annotation of image content. [4,13] present a similar approach to [14], except they use the PNG image format. There has additionally been effort to semi-automatically map low-level image features to ontological concepts. [7] presents an approach where users can select regions of images, from which low-level features are extracted (e.g., shape and color). Using pre-trained Bayesian networks, these low level features are classified as ontological concepts. In [1], feature vectors (color histograms) of images are extracted to populate concepts defined in domain specific ontology. Our approach differs from [1,7] in that we manually make assertions about image regions. While their approach is more automated, we feel that ours will be more accurate. The manual approach of region drawing also allows for more fine-grained regions that automated techniques may not detect. For example, a PhotoStuff user may want to add information about the watch a person is wearing or about a patch on their clothing – something that an automated technique would likely not detect. In [11], image segmentation techniques are used to segment digital images, and a technique to semi-automatically add spatial information about the segmented regions is presented. While the approach in [11] is automated, spatial information regarding regions can be added in our approach as well. This can achieved using the spatial ontology presented in [11] to manually make the assertions.

6 Conclusions

In this work we have presented a generic, domain independent framework for annotating and managing digital image content using Semantic Web technologies. In the approach, an annotation component is loosely coupled with a Semantic Web portal that supports browsing, searching and managing digital image annotations. Additionally, we have provided details of an open source implementation of this framework and an overview of a representative proof-of-concept designed as an experiment in collaboration with NASA. Potential future work includes automating portions of the annotations process, possibly by using image processing and computer vision techniques. Additionally, we plan to extend our work here to support annotation of additional digital media types, including video and audio.

This work was supported in part by grants from Fujitsu, Lockheed Martin, NTT Corp., Kevric Corp., SAIC, the National Science Foundation, the National Geospatial-Intelligence Agency, DARPA, US Army Research Laboratory, and NIST. We would like to thank NASA for their help in documenting requirements for this effort. We would also like to thank Daniel Krech, Ron Alford, Amy Alford, Grecia C. Lapizco-Encinas, and Aditya Kalyanpur for all of their contributions to this work.

References

1. Addis, M., Boniface, M., Goodall, S., Grimwood, P., Kim, S., Lewis, P., Martinez, K. and Stevenson, A. SCULPTEUR: Towards a New Paradigm for Multimedia Museum Information Handling. In Proceedings of the Second International Semantic Web Conference (2003) 582 -596 *[Addis]*
2. Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F., Stein, L. A. OWL Web Ontology Language Reference, W3C Candidate Recommendation, (2003).
3. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American, May 2001.
4. Brickley, D. RDF for Self-Describing Images. http://www.tasi.ac.uk/2000/09/rdfmeta/ (2001)
5. Brickley, D., and Guha, R. V. Resource description framework (RDF) schema specification 1.0. Candidate recommendation, W3C Consortium, 27 March 2000. See: http://www.w3.org.
6. Brin, S., and Page, L. The Anatomy of a Large Scale Hypertextual Web Search Engine, In the Proceedings of the 7th International World Wide Web Conference (1998)
7. Dupplaw, D., Dasmahapatra, S., Hu, B., Lewis, P., and Shadbolt, N. Multimedia Distributed Knowledge Management in MIAKT. ISWC 2004 Workshop on Knowledge Markup and Semantic Annotation. Hiroshima, Japan, November 2004
8. Exchangeable Image File Format for Digital Still Cameras: Exif Version 2.2. Standard of Japan Electronics and Information Technology Industries Association, (2002)
9. Frankel, C., Swain, M., and Athitsos, V. Webseer: An Image Search Engine for the World Wide Web, Tech. Report TR-96-14, Computer Science Dept., Univ. of Chicago, July (1996)
10. Hamilton, E. JPEG File Interchange Format. (1992)
11. Hollink, L., Nguyen, G., Schreiber, G., Wielemaker J., Wielinga. B., and Worring, M. Adding Spatial Semantics to Image Annotation. In Proceedings of Third International Semantic Web Conference - Knowledge Markup and Semantic Annotation Workshop (2004) *[Hollink04]*
12. Hollink, L., Schreiber, G., Wielemaker J., and Wielinga. B. Semantic Annotation of Image Collections. In Proceedings of Knowledge Capture - Knowledge Markup and Semantic Annotation Workshop (2003) *[Hollink]*
13. Hunter, J., Zhan, Z. An Indexing and Querying System for Online Images Based on the PNG Format and Embedded Metadata. In Proceedings of ARLIS/ANZ Conference (1999)
14. Lafon, Y., and Bos, B. Describing and Retrieving Photos Using RDF and HTTP. W3C Note available at: http://www.w3.org/TR/photo-rdf/ (2002)
15. Rui, Y., Huang, T. S., and Chang, S. F. Image Retrieval: Current Techniques, Promising Directions, and Open Issues. Journal of Visual Communication and Image Representation, Volume 10 (1999), pp. 39-62
16. Schreiber, G., Dubbeldam, B., Wielemaker, J., and Wielinga, B. Ontology-Based Photo Annotation. IEEE Intelligent Systems, 16(3) (2001) 66-74. [GuusOnt]
17. Smith, J. R., and Chang, S. F. An Image and Video Search Engine for the World Wide Web. Proc. SPIE 2670 Storage and Retrieval for Still Image and Video Databases IV, SPIE, Bellingham, Wash., (1996) pp. 84-95.
18. Suh, B., and Bederson, B. Semi-Automatic Image Annotation. University of Maryland Computer Science Department Technical Report, HCIL-2004-15, CS-TR-46 (2004)
19. Bloehdorn, S., Petridis, K., Saathoff, C., Simou, N., Tzouvaras, V., Avrithis, Y., Handschuh, S., Kompatsiaris, I., Staab, S., and Strintzis, M. G.: "Semantic Annotation of Images and Videos for Multimedia Analysis", Proc. 2nd European Semantic Web Conference, ESWC 2005, Heraklion, Greece, May 2005.