Metadata goes where Metadata is: contextual networks in the photographic domain

Rodrigo Fontenele Carvalho

The University of Sheffield, Web Intelligence Technologies lab,
211 Portobello, Sheffield, S1 4DP
R.Carvalho@dcs.shef.ac.uk

Abstract. Realizing the potential of digital media within the Semantic Web, relies on the existence of detailed, semantically unambiguous metadata. However, in the domain of personal photography within family centered communities, the generation of such metadata remains an unsolved issue. This paper introduces a solution to this problem based on exploiting the existence of strong semantic connections among photographs and entities specific to the user's context. In particular, three scenarios where this approach would be useful are discussed such as: (1) the reuse of existing metadata in order to generate more semantic metadata for unannotated photographs, (2) taking advantage of the connections among photographs to enhance search over a user's collection and (3) the support for communities by an external framework allows the use of image-based communications (e.g. sharing) to aid in metadata generation by reinforcing the connections across multiple graphs.

Keywords: photograph, metadata, semantic web, social web, context, graph

1 Introduction

In order to realize the potential of digital media within the Semantic Web, detailed and semantically unambiguous metadata (i.e. data about data) is needed, as it will provide the handles that will enable automated agents to interact with the external world to perform the more tedious tasks of finding, sharing and combining information on the web [1].

In the domain of personal photography, and more specifically within family centered communities, there is a growing need to exploit the existence of semantic metadata for facilitating the management of photographic collections. However the issue of generating such metadata for photographs remains unsolved.

The development of an approach that tackles this problem in the domain of family photography presents one main challenge: the use of appropriate data capture and reuse strategies. This challenge has not only been previously referred to as a crucial step to the wider adoption of the Semantic Web [2] but is also key in personal photography as previous studies by Frohlich et al [3], and Miller and Edward [4], have found that users in this domain are unlikely to make extensive contributions to generating semantic metadata about their ever expanding photographic collections.

Furthermore, offering great affordances for image-based communications (e.g. sharing) is an important requirement for this domain. The importance of this has only become clearer over the past years with the observation of an ever increasing number

of users that take advantage of web applications that enable the sharing of photographs and associated metadata via direct or indirect social interactions[1].

To address the issue of generating semantic metadata for personal photographs in the family domain an approach is proposed that assumes that (1) although users are unlikely to generate semantic metadata for their entire photographic collection, it is likely that some such metadata be generated for key images within a collection [5], and (2) photographs and entities specific to the user's context share strong semantic connections. By taking an approach based on directed graphs to representing and integrating user generated content with their social and physical environment, a **Context Network** can be created for modeling among other things: the content of photographs as well as user input (direct[2] or indirect[3]); user acquaintances (e.g. friends and genealogy tree), familiar locations and social interactions; automatic input from capture devices such as GPS coordinates, Exif and accelerometer data[4].

By taking advantage of the data structure formalization offered by ontologies to represent such context networks, the existing link structure between photographs and other context specific entities can be explored. More specifically, this can be targeted at facilitating the inference of detailed and semantically unambiguous metadata from the reuse of metadata that has been previously associated with other nodes in the graph (i.e. photographs). Another advantage of representing this data as a directed graph include improved searching capabilities over the photographic collection.

While the requirement for image-based communications is not directly addressed by the approach proposed, it can be implemented externally by a supporting framework. The social interactions between users that are triggered by the sharing of photographs can then be explored for the generation of semantic metadata.

The following sections introduce related work, details how the approach addresses the issue of semantic metadata generation as well as how it enhances search and capitalizes on the sharing of photographs for the generation of semantic metadata.

3 Related Work

Recent work in this area demonstrates different approaches for tackling the problem of metadata generation. Systems such as Photocopain [6] and MediAssist [7] rely on cheap contextual information as well as active human input for automatically or semi-automatically inferring textual metadata about the image contents. The use of collective input about a particular resource in order to make similar metadata suggestions for previously unseen resources has also been explored in AktiveMedia [8]. The application of human computation for annotation tasks has also been well researched, and can be exemplified by the game-like platform seen in [9].

The appearance of the Web 2.0 (i.e. support for social interaction and user generated content) has also motivated the creation of many photography applications

1 Flickr (http://www.flickr.com), Zooomr (http://www.zooomr.com), Riya (http://www.riya.com), MyHeritage (http://www.myheritage.com)

2 Direct user input includes any user interaction that has the explicit intent of generating photographic metadata such as adding captions, placing photographs on a map, etc.

3 Indirect user input includes any user interaction whose intent is not directed towards the generation of metadata, such as sharing photographs with other family members.

4 http://en.wikipedia.org/wiki/Accelerometer

(e.g. Flickr, Facebook[5]) that make use of active human input about photographs in the form of tags, comments or other annotation types (e.g. geographic, face, object, audio, etc). Others use input from the capture device or from processing the medium to infer data such as the geographic location (e.g. ZoneTag[6]) or face detection and recognition (e.g. Riya, MyHeritage.com[7]) for creating folksonomies of image metadata.

These approaches however miss out on the integration of the context of images as well as their owner's social environment[8] in generating semantic metadata for representing photographs. This network of data gives essential clues as to what kind of metadata is likely to be relevant and not using it in an integrated manner can be very limiting. For instance, in processing semantic captures such as Ben Nevis, contextual information about the photograph and the user can help determine whether the entity refers to the Scottish mountain or a person (e.g. does GPS data show the photograph was taken in Scotland? Does the user know anyone called Ben Nevis?).

4 Proposed Approach

The approach proposed for facilitating the generation of photographic metadata is founded on the notion that data mining requires taking into account not only the features of the entities of interest, but also the underlying structure of the data [11,10]. Its relation with the Semantic Web comes from a clear parallel that exists between such data structures and those imposed by the latter. Coupling graph based algorithms with the highly interlinked nature of RDF data is therefore not only natural but a sensible approach to developing novel techniques within this field of research.

Metadata Generation. For addressing the problem of metadata generation, the approach relies on a single core principle: ***"a little metadata goes a long way."*** This principle is based on the assumption that users often provide metadata for key photographs in their collections and addresses the challenge of making appropriate reuse of existing semantic metadata for the generation of more semantic metadata for unannotated photographs.

Fig. 1: Only the first image in this series contains textual metadata – *Everest trek*.

Fig. 1[9] exemplifies a situation commonly found in a person's collection of photographs: higher proportion of unannotated images than annotated ones. By taking advantage of strong semantic connections between the seed image and other images

5 http://www.facebook.com/
6 http://zonetag.research.yahoo.com/
7 http://www.myheritage.com/
8 This includes with whom a photograph has been shared, the user's closer acquaintances, familiar locations or even the similarity with other photographs in the same set.
9 Images contributed by Flickr user *Fernweh*.

in the same collection, as well as information in the context network, and applying algorithms that explore not only the links between these images but also the similarities between them [10] (e.g. visual similarity, closeness in time and GPS data), the information that is known about the seed image could be propagated to other neighbouring nodes (i.e. the other two images).

Searching. The existence of a graph with strong semantic connections between entities can also be explored for use in the retrieval of photographs. Queries based on keywords or visual similarities could be used for retrieving base images from a collection, and the context network commonly shared between these could be used to discover other resources that are also likely to be relevant but that may not share the exact same metadata as the base images. This would enable keyword queries to return images that have no textual metadata or content based[10] queries to return images that are conceptually similar but not necessarily visually similar. For instance, a seed image of a predominantly orange sea at sunset would not prevent images from the sea at sunrise to be retrieved in the same result set. Similar approaches have shown promising results in the domain of video retrieval [12].

Sharing. The requirement that any technique or tool developed for the family photography domain offers image-based communications can be turned into a feature of this approach. Intuitively, in a directed graph, data points (i.e. photographs) would be represented as both annotated and unannotated nodes and the edges between them are weighted so that the closer the nodes are according to a distance metric, the larger the weight of their connection. The social aspect of image sharing would play a vital role in controlling the weights between nodes from different graphs by modelling a direct relationship between people's social clique and each edge weight. That is, the stronger the social clique between people, the more likely it is that their images will influence the semantic metadata in each other's collections. The benefits of this would then be twofold: (1) for predicting who is likely to be interested in a specific image; (2) for using the information in other people's RDF graph to increment a user's graph.

Making predictions about who may be interested in any one specific photograph within a user's collection would involve an analysis on the link strength between nodes in the context network of photographs in two collections. The images with stronger links given by their contextual similarity as well as the user's social clique can be proposed for sharing.

An extension to the metadata generation approach, where nodes that have stronger links should be attributed similar semantic metadata, can be proposed for allowing the graph from two separate users to influence the semantic metadata of each other's photographs. By using the social clique between people as a control variable on the weight of node edges, an approach similar to semi-supervised learning, such as label propagation [10], can be used in the context of the social interaction between people. For instance, if a user tends to share a lot of photographs with his family, the weights between similar nodes across graphs in this community should reflect this fact. The contrary would happen with, for example, work colleagues who you don't usually share anything with.

---

10 CBIR

5 Evaluation and Work Plan

The evaluation of the results obtained for this work will depend largely on the collection of a corpus of photographs that is representative of the domain together with any associated metadata. It is envisaged that the use of a mixture of sources from the web (e.g. Flickr, Facebook, etc.) as well as directly from human users will be enough to provide a good sized heterogeneous corpus. This will not only aid the development of an approach that is valid in the real world, but also the testing of any algorithm that is derived from this work in terms of its accuracy.

It is also possible that some user based testing may be necessary in order to determine the level of satisfaction that is achieved by using this approach.

Some of the work that has been achieved so far involved an experiment related to capturing semantics from photographic descriptions contained within a large corpus of photographs from Flickr. For more details see [5].

Future plans include developing a preliminary model of the context network as well as experimenting with the graph based approach detailed previously for propagating metadata from annotated photographs to unannotated ones according to a user's context network.

References

[1] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific American Magazine*, May. 2001.

[2] M.C. Schraefel, "What is an analogue for the semantic web and why is having one important?," *HT '07: Proceedings of the 18th conference on Hypertext and hypermedia*, Manchester, UK: ACM Press, 2007, pp. 123--132.

[3] D. Frohlich et al., "Requirements for photoware," *2002 ACM conference on Computer supported cooperative work*, New Orleans, Louisiana, USA: ACM Press, 2002, pp. 166-175.

[4] A. Miller and K.W. Edwards, "Give and Take: A Study of Consumer Photo-Sharing Culture and Practice," *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, San Jose, California, USA: ACM Press, 2007, pp. 347--356.

[5] R.F. Carvalho, S. Chapman, and F. Ciravegna, "Extracting semantic meaning from photographic annotations using a hybrid approach.," *MMIU'08: Proceedings of the 1st International Workshop on Metadata Mining for Image Understanding*, Funchal, Madeira - Portugal: 2008.

[6] M.M. Tuffield et al., "Image Annotation with Photocopain," *Proceedings of the Fifteenth World Wide Web Conference (WWW2006)*, 2006.

[7] N. O'Hare et al., "MediAssist: Using Content-Based Analysis and Context to Manage Personal Photo Collections," *Image and Video Retrieval, 5th International Conference*, Tempe, AZ, USA: Springer, 2006, pp. 529-532.

[8] A. Chakravarthy, F. Ciravegna, and V. Lanfranchi, "Cross-media document annotation and enrichment," *1st Semantic Authoring and Annotation Workshop (SAAW2006)*, 2006.

[9] L. von Ahn and L. Dabbish, "Labeling images with a computer game," *ACM Conference on Human Factors in Computing Systems (CHI)*, 2004, pp. 319--326.

[10] X. Zhu, "Semi-supervised learning with graphs," May. 2005.

[11] E. Minkov and W.W. Cohen, "An Email and Meeting Assistant using Graph Walks," *Third Conference on Email and Anti Spam*, Mountain View, California USA: 2006.

[12] W.H. Hsu, L.S. Kennedy, and S. Chang, "Video search reranking through random walk over document-level context graph," *Proceedings of the 15th international conference on Multimedia*, Augsburg, Germany: ACM Press, 2007, pp. 971-980.