Enhancing Folksonomy-Based Content Retrieval with Semantic Web Technology

Rachanee Ungrangsi, Shinawatra University, Thailand
Chutiporn Anutariya, Shinawatra University, Thailand
Vilas Wuwongse, Asian Institute of Technology, Thailand

ABSTRACT

While Flickr, a widely-known photo sharing system, allows users to describe their own photos with tags (aka. folksonomy tags) for indexing purposes, its tag-based photo retrieval function is severely hampered by the inherent nature of folksonomy tags. This paper presents SemFlickr, an application which enhances the search in Flickr with its semantic query suggestion feature. SemFlickr employs SQORE, an ontology retrieval system, to retrieve relevant ontologies from the Semantic Web and then derives query term suggestions from those ontologies. To ensure that the highly related photos will appear at the top of the results, SemFlickr takes the ontological relations among the given query terms to assign tag scores and then generates its ranked results. Experimental outcomes are encouraging and reveal a number of useful insights for developing applications that integrate the Semantic Web and Web 2.0 together.

Keywords: Folksonomy, Photo Retrieval, Semantic Media Adaptation, Semantic Web, Web 2.0

1 INTRODUCTION

Flickr – http://www.flickr.com (20092009Yahoo! Inc., 2009!!) is one of the very first successful and widely-used Web 2.0 applications. It allows users to freely upload, tag and share photos and videos with others. With its self-tagging mechanism, different users can annotate the same resource with different tags depending on the social communities they belong to, their background, their perception and areas of expertise, etc. To date, there are approximately at least 3 million items uploaded per day with 20 million unique users visiting Flickr per month. In spite of its rapid growth and (meta) data richness, its users are experiencing increasing difficulties of searching for a particular photo or video.

Consider, for example, that Flickr currently has more than 2 million photos tagged with the term baby whereas its full-text search engine returns approximately 6 million photos relevant to the term baby. Simply searching Flickr with the query term baby will not return only the
photos of very young children, but also those about baby-related products, toys, young animals, pregnant women and even adults. Unlike a concept defined in an ontology, the semantics of a particular tag is ambiguous, and may carry more than one specific meaning depending on each user’s viewpoint and context. This can answer why Flickr’s most interesting photos of baby contain a lot of animal photos (See Figure 1). Therefore, users must put an extra effort to think of additional precise query terms to clarify a search and obtain the desired results. Assisting the searching process by providing users with query suggestions can help reduce the searching effort and improve the search’s precision.

This paper, therefore, presents the SemFlickr application which employs Semantic Web technologies to enhance the search in Flickr with its query suggestion feature. In particular, SemFlickr is a prototype semantics-driven photo retrieval application that supports the user during a photo search by suggesting additional query terms derived from on-line ontologies available in the Semantic Web. Its search interface is designed to enable the user who has limited knowledge about the searching domain to explore more related query terms and search results. It is available on-line at URL: http://research.siu.ac.th/SemFlickr. In addition, to improve the retrieval performance, a semantics-based ranking mechanism is also proposed. A comparison of Flickr’s and SemFlickr’s search results has been conducted and shows that SemFlickr can yield better search results. Furthermore, a user-based evaluation implies clear benefits of semantic query suggestion and indicates a few issues which require careful consideration in Semantic Web application development.

The paper is organized as follows. Sect. 2 reviews related works. Sect. 3 introduces SQORE, the exploited ontology retrieval engine. Sect. 4 describes SemFlickr and its underlying mechanism for deriving semantic query suggestions and ranking the search results. Sect. 5 and 6 discuss the two conducted experiments and their results. Conclusions and future works are presented in Sect. 7 and followed by acknowledgement in Sect. 8. Finally, Appendix

Figure 1. Top twelve most interesting Flickr’s photos tagged with “Baby” (retrieved on 27 July 2009)