Expanding Folksonomy Search with Ontologies

Jeff Z. Pan
Department of Computing Science
University of Aberdeen, UK
jeff.z.pan@abdn.ac.uk

Stuart Taylor
Department of Computing Science
University of Aberdeen, UK
staylor@abdn.ac.uk

Edward Thomas
Department of Computing Science
University of Aberdeen, UK
ethomas@csd.abdn.ac.uk

ABSTRACT

In this paper, we propose an approach to expand folksonomy search with ontologies, which are completely transparent to users. Preliminary implementations and evaluations of this approach are promising.

1. INTRODUCTION

There are advantages and disadvantages of the tagging approach [3]. The folksonomy based systems allow great malleability and adaptability, and the simplicity of adding and using tags allows them to be more accessible to users than well defined (more complex) classification systems. However, these systems also surfer from problems [8], such as ambiguity in the meaning of tags and flat organisation of tags.

It has been suggested [4, 8, 1] that Semantic Web technologies could help improve the internal structure of folksonomies. Passant proposed to encourage users to link tags with concepts and individuals in ontologies [8]. However, as suggested by Hotho et al. [4], a key question remains open: how to exploit the benefits of ontology without bothering untrained users with its rigidity.

In this paper, we propose to use ontology entailments to expand folksonomy search, without bothering untrained users to use ontologies directly. In Section 2 we present our approach to increase search precision over folksonomies, Section 3 outlines our current implementations and Section 4 presents a case study of an application making use of our approach in the music domain.

2. FOLKSONOMY SEARCH EXPANSION

This section presents the idea of folksonomy search expansion based on ontologies.

Let us first revisit the formal model of folksonomy to illustrate some limitations of tagging from the aspect of search. Given a folksonomy based system $\mathcal{S}$ that uses the folksonomy $\mathbf{F} = \langle \mathcal{U}, \mathcal{T}, \mathcal{R}, A, \Theta \rangle$, there are three main limitations of tagging regarding search in $\mathcal{S}$ [8]. Firstly, tag variation (users using different tags for the same meaning - e.g., "cat" and "kitty"). Secondly, the flat organisation of tags (e.g., a search for "animal" will not return resources which are tagged with "cat" or "dog") and finally tag ambiguity.

The above limitations suggest that ontologies could be helpful since they contain important concepts and their relationships, as well as some background assumptions of some specific domain.

Now the question is: how to associate tags/keywords with ontologies? Keywords appear in annotations and datatype properties of ontologies. These keywords are associated with classes, properties and individuals (in ontologies), with scores based on ranking factors[1]. We then use these keywords to match each keyword from a set of tags to a class, property or individual in an ontology (depending on the application). For example, if a user searches for "animal", the keyword "animal" can be matched to the class Animal in the ontology. The search for "animal" can then be expanded to include all *types* of animal by including keywords related to sub-classes of Animal.

3. IMPLEMENTATIONS

ONTOSEARCH2[2] [7] is an ontology infrastructure; its basic components include an ontology repository, where users can submit and query ontologies. When users submit their ontologies, ONTOSEARCH2 computes the semantic approximation of the ontologies as described in [6] and computes keyword-ontology association as presented in the previous section.

Taggr provides an ontology-enabled common interface for folksonomy based systems[3]. It stores a basic ontology (which we refer to as the "tagging database") in ONTOSEARCH2, capturing the relationships between users, tags and resources in the folksonomy based systems that it supports. It provides functionality for gathering resources and their related tags from the tagging systems that it supports, and then populate them to its tagging database from time to time.

MusicMash2 is a sample application which we have developed to show how Taggr can be used. Details are given in the following section.

4. CASE STUDY: MUSICMASH2

MusicMash2 is a semantic mashup application which is intended to integrate music-related content from various folksonomy based tagging systems and music meta-data Web services. MusicMash2 has two main functions. Firstly, it gathers information from several music-meta Web services

[1] http://www.seomoz.org/article/search-ranking-factors
[2] http://www.ontosearch.org/
[3] Taggr currently supports YouTube and Flickr.

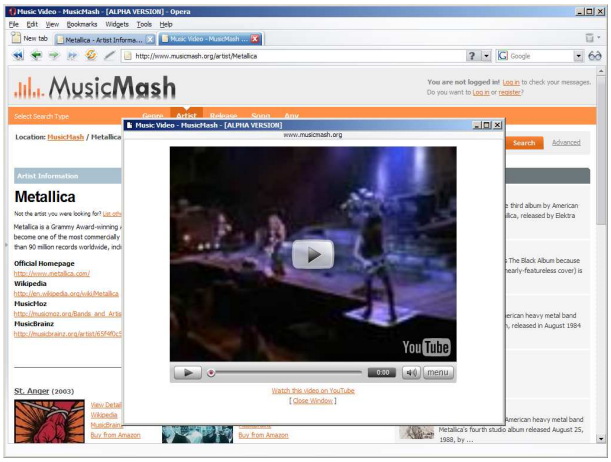

Figure 1: MusicMash2 Alpha: http://www.musicmash.org/

and uses this to populate The Music Ontology[4]. The populated Music Ontology is then stored in the ONTOSEARCH2 repository. The second function is to allow users to search for images and video from Flickr and YouTube. A naive approach to folksonomy search, such as those provided most tagging systems[5], results in unacceptable precision in domain specific searches. Therefore, MusicMash2 makes used of the proposed folksonomy search expansion methods.

4.1 An Example Scenario

A typical scenario of a user searching for an artist's music videos can be used to illustrate exactly how MusicMash2 makes use of Taggr. In this example, the user is searching for "Focus" (the Dutch progressive rock band), using the search term *Focus* (we have intentionally chosen a suitably ambiguous search term for this example). Firstly, using YouTube directly to search for videos, out of the first page of 20 results, 15 had no relevance to the band Focus. These results clearly have unacceptable precision for a domain-specific application such as MusicMash2. In order to increase the precision of the search, we can be exploit some domain knowledge. Firstly, due to the design of the MusicMash2 user interface, we can easily determine that a user is searching from a Music Artist. Secondly, in the domain of music, artist names are almost always unique (we do accept that there are some exceptions to this, however we allow the user to specify exactly which artist they are looking for by making use of the MusicBrainz search API). This expansion is based on the observation that when searching YouTube, the combination of an artist name and a song title is much less ambiguous than searching using the artist's name alone. More precisely, the search term *Focus* is expanded to a list of search terms of the form $ArtistName\ SongName_1$, $ArtistName\ SongName_2$ ... $ArtistName\ SongName_n$. In this example, the expanded search results have perfect precision[6].

4.2 Scalability of Search Expansions

[4]`http://www.musicontology.com/`
[5]YouTube Developer API: http://www.youtube.com/dev
[6]Focus - http://www.musicmash.org/artist/Focus

First of all, it should be noted that the scalability of this application depends on the performance of ONTOSEARCH2. Evaluations [6, 5] of ONTOSEARCH2 have been made using the Lehigh University Benchmark [2], and these have shown that ONTOSEARCH2 is scalable for large ontologies, containing of millions of individuals. We expect that a full evaluation of Taggr using data generated by MusicMash2 will show that our approach can easily scale to several thousands of artists.

5. CONCLUSIONS AND FUTURE WORK

In the paper, we have investigated the open problem of how to exploit the benefits of ontology to improve folksonomy search, without bothering untrained users with its rigidity. We have proposed to use ontologies to expand folksonomy search transparently for users, and have presented an ontology infrastructure (see Section 3) to enhance Web 2.0 applications that make use folksonomy search.

Our case study, MusicMash2, shows that an application designed to use an OWL DL ontology to express its knowledge has the advantage that the developer can make use of the search expansions provided by Taggr.

Future work includes optimisations to the current implementations of Taggr and MusicMash2, along with a full evaluation covering the usefulness and scalability of our approach. Furthermore, we are working on evaluating our approach with other domains and tagging systems.

6. REFERENCES

[1] A. Ankolekar, M. Krotzsch, T. Tran, and D. Vrandecic. The two cultures: mashing up web 2.0 and the semantic web. In *Proc. of the 16th international World Wide Web conference (WWW2007)*, pages 825–834, 2007.

[2] Y. Guo, Z. Pan, and J. Heflin. LUBM: A Benchmark for OWL Knowledge Base Systems. *Journal of Web Semantics*, 3(2):158–182, 2005.

[3] H. Halpin, V. Robu, and H. Shepherd. The complex dynamics of collaborative tagging. In *Proc. of the 16th International World Wide Web Conference (WWW2007)*, pages 211–220, 2007.

[4] A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme. Information retrieval in folksonomies: Search and ranking. In *Proc. of the 3rd European Semantic Web Conference (ESWC2006)*, pages 411–426, 2006.

[5] J. Z. Pan, G. Stamou, G. Stoilos, S. Taylor, and E. Thomas. Scalable Querying Services over Fuzzy Ontologies. In *Proc. of the 17th International World Wide Web Conference (WWW2008)*, 2008.

[6] J. Z. Pan and E. Thomas. Approximating OWL-DL Ontologies. In *Proceedings of the 22nd Conference on Artificial Intelligence (AAAI-07)*, pages 1434–1439, 2007.

[7] J. Z. Pan, E. Thomas, and D. Sleeman. ONTOSEARCH2: Searching and Querying Web Ontologies. In *Proc. of WWW/Internet 2006*, pages 211–218, 2006.

[8] A. Passant. Using Ontologies to Strengthen Folksonomies and Enrich Information Retrieval in Weblogs. In *Proc. of 2007 International Conference on Weblogs and Social Media (ICWSM2007)*, 2007.