Leveraging Multi-faceted Tagging to improve Search in Folksonomy Systems

Fabian Abel
L3S Research Center
Leibniz University Hannover
Appelstr. 4 D-30167 Hannover
abel@L3S.de

Ricardo Kawase
L3S Research Center
Leibniz University Hannover
Appelstr. 4 D-30167 Hannover
kawase@L3S.de

Daniel Krause
L3S Research Center
Leibniz University Hannover
Appelstr. 4 D-30167 Hannover
krause@L3S.de

ABSTRACT

In this paper we present ranking algorithms for folksonomy systems that exploit additional contextual information attached to tag assignments available. We evaluate the algorithms in the TagMe! system, a tagging front-end for Flickr, and show that our algorithms, which exploit categories, spatial information, and URIs describing the semantics of tag assignments, perform significantly better than the FolkRank that does not consider such contextual information.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Search and Retrieval; H.4.m [**Information Systems**]: Miscellaneous

General Terms

Algorithms, Experimentation

Keywords

Social Media, Faceted Tagging, Search, Ranking, Folksonomies, Context

1. INTRODUCTION

Tagging systems like Flickr or Delicious organize and search large item collections by utilizing the Web 2.0 phenomena: Users attach tags to resources and thereby create so-called tag assignments which are valuable metadata. However, imprecise or ambiguous tag assignments can decrease the performance of tagging systems regarding search and retrieval of relevant items.

For disambiguation, approaches like MOAT [5] exist, which support users to attach URIs describing the meaning of a tag to a particular tag assignment analogously to semantic tagging in Faviki[1]. A more sophisticated approach, which exploits Wikipedia and WordNet to detect the meaning of tags, is presented in [4].

In this paper, we extend the common folksonomy model by flexible, contextual tagging facets. We present the TagMe! system that introduces novel tagging facets: Tag assignments are enriched with a DBpedia URI [2] to disambiguate the meaning of a tag. So-called *area tags* enable users to tag a specific part of an image (spatial tagging). Furthermore, a *category* dimension is offered to categorize tag assignments. In our evaluation we examine how the different context facets can be exploited to improve search.

[1]http://faviki.com

HT'10, June 13–16, 2010, Toronto, Ontario, Canada..
ACM 978-1-4503-0041-4/10/06.

2. TAGME! SYSTEM

TagMe![2] [1] is an online image tagging system where users can assign tags to pictures available in Flickr. Users can directly import pictures from their own Flickr account or utilize the Flickr search interface. If users tag their own images in TagMe! then the tags are propagated to Flickr as well. Moreover, TagMe! maps DBpedia URIs to tag assignments by simply selecting the most prominent URI returned by the DBpedia lookup service[3] for a given tag. The mappings of this naive approach, which results in a precision of more than 75%, are finally completed by hand.

TagMe! extends the Flickr tagging functionality in two further facets, specifically *categories* and *area tags*. For each tag assignment the user can enter one or more categories that classify the annotation. While typing in a category, the users get auto-completion suggestions from the pre-existing categories of the user community. TagMe! users can immediately benefit from the categories as it provides a faceted search interface that allows to refine tag-based search activities by category (and vice versa). Additionally, users are enabled to attach a tag assignment to a specific area, which they can draw within the picture similarly to *notes* in Flickr or annotations in LabelMe[4]. When tagging, people usually only tag the main content of the picture, giving less or almost none importance to supplementary scenery images. Area tags motivate the users to do so adding significant semantic value to each annotated image. While the area tags add an enjoyable visible feature for highlighting specific areas of an image and sharing the link to such areas with friends, we consider them as highly valuable to improve search by detecting tag correlations.

3. EVALUATION

In our evaluation we examine the impact of the additional context generated by the multi-faceted tagging on search and mining tag relations. In particular, the key question we would like to answer is: *Does the exploitation of the additional context improve the search and ranking performance?*

[2]http://tagme.groupme.org
[3]http://lookup.dbpedia.org
[4]http://labelme.csail.mit.edu

To answer this question we examine the impact of the advanced semantics provided by the TagMe! context folksonomy on search. In particular, we apply the FolkRank algorithm [3] as well as the Category-, Area-, and URI-based FolkRank adaptions to search and rank Flickr images and investigate how the different context types can help to improve the search performance. We evaluate the algorithms with respect to the following task.

> **Resource Ranking Task.** *Given a keyword query (tag), the task of the ranking strategy is to compute a ranking of resources so that resources that are most relevant to the keyword query appear at the top of the ranking.*

To acomplish this task we proposed three extensions [1] to FolkRank.

AreaFolkRank considers the size and position of an area tag. Our hypothesis is that the larger the size of an area the more important is also the corresponding tag for the given resource. The same holds for the position of the tag where spatial information relevant to the center of a resource are more important for the resource than tag assignments which are associated to the margin.

CategoryFolkRank operates on a context folksonomy where the context is given by categories that are attached to tag assignments. The algorithm relates folksonomy entities via the category assignments and the main hypothesis is that entities sharing the same category are related to each other.

DBpediaFolkRank operates on meaningful URIs instead of tags. This algorithm is therewith resistant against ambiguous tags as well as synonymic tags since the unique URI clearly defines the meaning of the tags.

3.1 Results

Figure 1 shows the precisions within the top 10 (P@10) and top 20 (P@20) search results of the different ranking strategies. Those algorithms that make use of contextual information embedded in the folksonomy perform better than the traditional FolkRank algorithm that considers only the tag assignments without any additional context. Between DBpediaFolkRank and FolkRank there seems to be no remarkable performance difference. The CategoryFolkRank performs good results especially for the precision within the top 20. Hence, the hypothesis that category assignments can be used to relate resources seems to hold. By exploiting the category context, the algorithm detects relevant resources that are not directly related via tag assignments to the given query. The AreaFolkRank algorithm, which exploits the size and position of spatial information attached to the tag assignments, is—with respect to P@10—the best algorithm among the core ranking strategies (P@10 = 52.9%). However, there is no significant difference between the FolkRank and the Area-, Category-, and DBpedia-based FolkRank.

A hybrid strategy "F+C+A+D", which combines all four core ranking strategies (i.e., FolkRank, CategoryFolkRank, AreaFolkRank, and DBpediaFolkRank), is the most successful strategy. It performs significantly better than the FolkRank algorithm regarding the P@10 and P@20 metrics. The combined strategy improves the precision of FolkRank by 20.0% and 21.4% with respect to the precision within the top 10 and top 20 respectively.

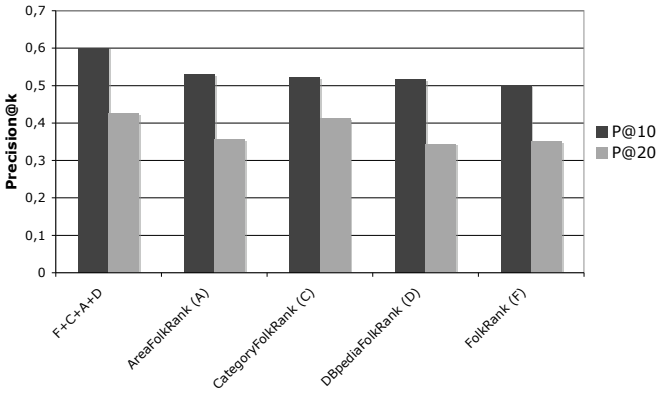

Figure 1: Precisions of FolkRank-based search algorithms.

4. CONCLUSIONS

The algorithm that considered the different contextual facets significantly improved the precision of the baseline FolkRank algorithm by 20.0% and 21.4% with respect to the precision of the search result rankings. Relying on these results we demonstrated that contextual information can significantly improve search. In summary, the exploitation of context embedded in the folksonomy is beneficial for ranking resources. While the size and position of the area helps to improve the precision particularly at the top of the resource rankings, the DBpedia and category context successfully contribute to improve the recall [1]. And by combining the different context types we are able to improve the ranking performance of FolkRank significantly.

5. REFERENCES

[1] F. Abel, N. Henze, R. Kawase, and D. Krause. The impact of multifaceted tagging on learning tag relations and search. In *Extended Semantic Web Conference (ESWC)*, June 2010.

[2] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. DBpedia: A Nucleus for a Web of Open Data. In A. et al., editor, *The Semantic Web, 6th International Semantic Web Conference (ISWC), 2nd Asian Semantic Web Conference (ASWC)*, pages 715–728, November 2007.

[3] 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*, volume 4011 of *LNCS*, pages 411–426, Budva, Montenegro, June 2006. Springer.

[4] A. Marchetti, M. Tesconi, F. Ronzano, M. Rosella, and S. Minutoli. SemKey: A Semantic Collaborative Tagging System. In *Workshop on Tagging and Metadata for Social Information Organization at WWW '07, May 8-12, 2007, Banff, Canada.*, May 2007.

[5] A. Passant and P. Laublet. Meaning Of A Tag: A collaborative approach to bridge the gap between tagging and Linked Data. In *Proceedings of the WWW 2008 Workshop Linked Data on the Web (LDOW2008), Beijing, China*, Apr 2008.