Extending Folksonomies for Image Tagging

Roman Kern, Michael Granitzer, Viktoria Pammer
Know-Center Graz, Styria, Austria
{mgrani, rkern, vpammer}@know-center.at

Abstract

Due to the unsatisfactory results of content based image retrieval methods, organisation and retrieval of multimedia data strongly relies on metadata and free text description. Folksonomies, collaboratively created sets of metadata, emerged recently and help for organising web based multimedia information. This contribution addresses the question how to extend a classical Folksonomy with additional metadata and discusses the quality of the extended Folksonomy and its application for tag recommendation in particular. We show that some relations of the original Folksonomies can be replaced by the extended Folksonomy while others are unique. In addition our analysis shows, that for 40% the correct tag is in the first 10% of the set of tag recommendations.

1 Introduction

Organisation and retrieval of multimedia data strongly relies on metadata and free text description. Content based retrieval methods, like for example for images, exist, but they mostly rely on low level descriptors. Hence, retrieval based on low level descriptors yields most often to unsatisfactory results [6].

Recently Folksonomies, collaboratively created sets of metadata, become more and more important for organising web based multimedia information. Folksonomies can be seen as graph consisting of user, tag and resource nodes which are interconnected with each other (see [3] for a formal definition). The graph structure is created by users assigning keywords, called tags, which are not restricted to a vocabulary, to a resource like for example photos. This collaboratively collected assignments of tags do not only offer new means in managing multimedia, but provide also a rich foundation of human generated knowledge about multimedia items and personal preferences - knowledge valuable for retrieval and to support multimedia annotation.

Research group have exploited Folksonomy structures for retrieval [2], analyzed emergent semantics in Folksonomies [5, 4] or use Folksonomie structures for tag recommendation [3].

Our contribution extends those prior research and addresses the question how to extend a classical Folksonomy with this additional metadata and discusses the quality of the extended Folksonomy for tag recommendation in particular and Folkonosmies in general. The work presented in this paper starts at extending the "classical" Folksonomie graph with additional metadata available for each photo. Depending on the kind and modality of the metadata, preprocessing methods are applied to extend the Folksonomy. Relying on spreading activation techniques for querying the graph we are analysing the properties of the extended Folksonomy and compare it to the "classical" Folksonomy. While our analysis mainly address tag recommendation for images, findings may also be applied to different applications of Folksonomies like for example retrieval.

The dataset on which our evaluation is run, was collected using the Flickr[1] web service. Out of the group *fruits* & *vegs* 13651 photos, 4336 people and various additional data, like the favorite photos of the users were downloaded and stored in a relational database.

2 Extending the Folksonomy

A folksonomy is defined as tripartite graph where each node identifies either a user, a tag or a resource (i.e. resources are photos in our case) and relationships among each other. In addition Flickr provides further metadata like title, description etc. associated with an photo. Our extensions starts at viewing this additional data togehter with the original folksonomy data as directed bipartite graph G_s, depicted in detail in Figure 1, left. This graph consists of nodes of different types and weighted edges between the nodes. Each node type is assigned to one of two classes, the *connector nodes* and the *value nodes*. G_s is build from a relational database of Flickr data, where rows of the tables have been transformed into connector nodes, one for each row, and columns have been used to create the value nodes.

[1] http://www.flickr.com

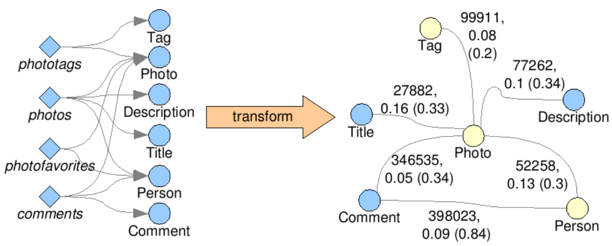

Figure 1. Transformation of G_s (left) into G_t (right). Diamonds in G_s represent the database nodes which build the connections between the value nodes depicted as circles. Relations and node types of G_t are shown right. The edges are labeled with the total number of associations between each node type and its average weight. The average weight of the 10000 topmost associations is shown in brackets

The values for the *Description*, *Title* and *Comment* columns - which are containing textual content - have been tokenized and a value node was created for each token [2]. By using the cosine similartiy on all pairs of row vectors in the adjacency matrix of G_s we are determining the edge weights between nodes. So for example the similarity of a Tag with a Photo is calculated based on the shared connections with the phototags connector nodes, built using the phototags database table.

Based on the extracted value nodes and calculated edge weights, our extended folksonomy graph G_t is created. G_t contains every value node of G_s and introduces a weighted edge for each non-zero similarity between two nodes. The weight of this edge is the similarity between the two nodes and due to the symmetry property of the cosine similarity those edges are undirected. The right side of figure 1 shows the model of the extended folksonomy graph G_t, whereas the nodes *Tag*, *Photo* and *User* as well as the relations between them comprise the original folksonomy.

3 Associative Querying

Our analysis is based on querying the created graphs and compare results, wherefore we rely on spreading activation [1]. Spreading activation allows considering edge weights and is widely used in querying graph based structures like folksonomies. Therefore, a set starting nodes is selected and assigned with an initial activation, usually 1. From these nodes all outbound edges are traversed and the destination nodes are assigned with the product of the source node activation and the edge weight. The results serves as activation for the newly found nodes, i.e. the destination nodes, which are the new starting nodes for the next hop. That way G_t is searched in a spreading activation like manner. If a node is reached via more than one node, the highest weight wins and is used as activation of the node. In this case, the node will be used again for the next association step. Because of performance considerations the number of start nodes for each hop is limited to the 1024 highest weighted nodes. However, due to the sparse nature of the G_t, this limit is seldomly reached.

Figure 2 shows the highest weighted nodes of G_t for a search starting at the user *pizzodisevo* and terminating at Tag nodes, that are reachable within 5 hops. The top Tag nodes are then displayed. The color of the graph nodes encode the node type, the connection length and the stroke strength is proportional to the edge weight. The font size of each node represents the number of connections within the source-graph, which can be interpreted as popularity of the tag or word.

As it can be seen in figure 2 some tags are extracted via description and title, like for example the *courgette* nodes or the *norica* nodes. In contrast, additional nodes are added by description and title raising the question on the differences between the original and the extended Folksonomy.

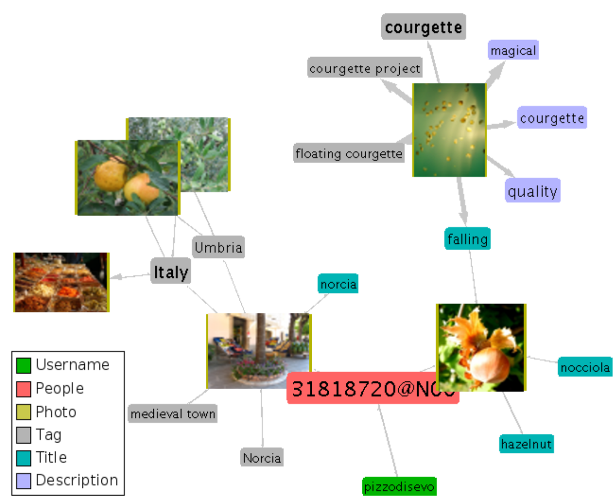

Figure 2. Screenshot of a visualization for a query starting at the user *pizzodisevo*. From there various paths are used to find tags, e.g. the title node *falling* connects the photo of a hazelnut (nocciola [it.]) with the zucchini photo (courgette [fr.]).

[2]No stemming or stop-word removal has been applied

4 Analysis

The main goal of our analysis has been to analyse relevant properties of the extended Folksonomy in comparison to the original Folksonomy for application in an recommender system. In order to achieve this, our analyis has been divided into three separate sub-tasks. The focus of the first task is to measure the *overlap* between a traditional Folksonomy with the extended Folksonomy G_t. The second task aims at determining the importance of single edge types by using a leave-one-out approach within G_t, whereas the final task evaluates the predictive quality for each edge type required by an tag recommender system.

4.1 Overlap

To estimate the additionally information contained in the extended Folksonomy, three different target graphs have been build and queried. The first target graph, G_t^o represents the original Folksonomy and is processed using only the *phototags* and *photos* node types in G_s and restricting the value nodes to *Photo*, *Tag* and *People*. The second target-graph G_t^e is build from the complete source graph G_s. A pre-folksonomy graph G_t^p - excluding all tagging information - is created as third graph using all but the *phototags* common node. Thus, G_t^p should give insights in the relevance of tagging information.

The tests with these three graphs were conducted with two node types: People and Photo. For each instance of these types associated nodes of the same type were searched. The result set of searches within the G_t^e and G_t^p are compared with the result set of G_t^o and the overlap and recall are estimated. The overlap is calculated by dividing the number of common nodes by the sum of all distinct nodes in the search results sets and the recall is the count of common nodes in relation to the size of the result set found using only G_t^p.

Graph Type	Nodes	Overlap	Recall
G_t^e	*user*	15 %	67.5 %
G_t^p	*user*	3.8 %	24.0 %
G_t^e	*photo*	79.8%	100.0%
G_t^p	*photo*	6.0%	11.5%

Table 1. Results of the overlap test for G_t^e and G_t^p in comparison with G_t^o.

The results for the overlap test are presented in table 1. The most obvious difference is that the values for the extended Folksonomy are far higher that the values for the pruned Folksonomy. This leads to the conclusion that the $photo \leftrightarrow tag$ relation plays an important role to associate photos as well as users. The difference between the traditional folksonomy and the extended version is surprisingly low for the $photo \leftrightarrow photo$ association. This again is a hint for the importance of the tag relation. Adding the title and description to a photo appears to add not much information, but it also does not deprive performance. The additional comments relation on the other hand allows the algorithm to detect much more associated users for a given user, lowering the average overlap.

4.2 Leave-one-out

For the second test the top 10000 edges of each relation type have been selected. For each such edge a search in G_t^e using its source node as start node are triggered, whereas traversal over the selected edge type are ignored. During the traversal the path length to the destination node of the selected edge, the activation of the destination node [3] as well as the average weights are calculated. These measures describe the quality of the alternative path in the target graph between the two nodes connected by a specific type of edge.

Edge Type ignored	TAW	AAD	AW	APL
$user \leftrightarrow comment$	*0.84*	0.06	0.24	*2.00*
$photo \leftrightarrow comment$	0.34	*0.13*	*0.37*	*2.00*
$photo \leftrightarrow user$	0.30	0.08	0.31	2.16
$photo \leftrightarrow description$	0.34	0.001	0.16	3.09
$photo \leftrightarrow title$	0.33	0.01	0.22	3.16
$photo \leftrightarrow tag$	0.30	0.001	0.14	3.04

Table 2. Results of leave-one-out test for each edge type with the values for the average weight of the 10000 topmost associations (TAW), the average activation of the destination node (AAD), the average weight (AW) of the edges on this path as well as the average length of the path (APL).

Table 2 shows the averaged measures for the different edge types. The average weight (TAW) of the top 10000 nodes is with 0.84 clearly the highest for the $comment \leftrightarrow person$ edge type. The different vocabulary of individual users used for comments can be attributed for this high value which supports also findings in [4]. The relation between comment and photo nodes appears to carry the least information as the average weight (AW) of the alternative path is the highest of all relation types. On the other side, the connection of a photo node to a description, title or tag node can hardy be compensated once removed since due to

[3] Note that since our initial activation is 1, the activation of the destination node corresponds with multiplying all edge weights on the shortest path between source and destination node

their low activation these nodes would be hardly reached. The average path length (APL) gives a clue for this behavior. On average the number of edges needed to search in the target-graph is more than three (the average weight of the edges on this path is 0.16, 0.22 and 0.14) which yields to a rapid decrease in activation.

4.3 Prediction

The third and final test evaluates the potential use of the different edge types for predicting future distribution, which is crucial for a recommender system. To accomplish this G_s was divided into two parts. The training set contains 60% of the photos and the test set contains the remaining 40% of all photos. From the training set the extended Folksonomy G_t^e was created. From the test set all photos were iterated and associated metadata was collected. For each metadata type one value is picked out and the others are used as input query for an associated search. Within the ordered result set the position of the picked out value is recoded. For example the Photo with the id *10547374* carries the tags *2002*, *tomato*and *nature*. For the first of three rounds the tag *2002* is removed from the set and the tags *tomato*and *nature*are used to build the search query. The position of the Tag *2002* is recorded and the test is then repeated for both other tags. Not only tag relations are tested this way, but all other related node types related with photos. Although they probably be never be used directly in a real-world recommender system, they still give a better understanding for the available data.

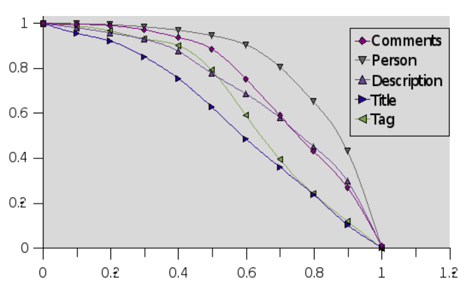

Figure 3. Results of the prediction test that simulates a recommender system. On average the desired node can be found in the first 10% of the result set for about 40% of all nodes (with the exception of the $photo \leftrightarrow description$ relation).

For the final test, the prediction of value nodes, the results are harder to interpret without manual relevance assignment since the quality of predicted tags can not be considered. Still the output of this test is promising as seen in figure 3. On average the correct tag can be found in the first 10% of the result set for about 40% of all searches. This can be seen as baseline performance for building a tag recommender system. Other relation types did perform similar to the $photo \leftrightarrow tag$ relation, especially the connection between photos and users appears to be stronger than for example the relation from a title to photo. This could be exploited by a recommender system to improve its results.

5 Conclusion

In this paper we presented an approach to extend a photo Folksonomy by including other types of metadata into the graph like structure. Our analysis indicate, that not only the tag and the user relation carry information about a photo, but also the additional data as collected by services such as Flickr. These metadata could be used for example in a recommender system that semi-automatically supplies the user with tags. The results, especially these of the prediction test, could be easily improved by using more sophisticated preprocessing techniques. Using a different similarity other than the cosine similarity could also greatly improve performance. These are only small subset of tasks that are needed to be addressed by future research work. [4]

References

[1] F. Crestani. Application of spreading activation techniques in information retrieval. *Artificial Intelligence Review*, 11(6):453–482, December 1997.

[2] A. Hotho, R. Jschke, C. Schmitz, and G. Stumme. Information retrieval in folksonomies: Search and ranking. In *Proceedings of the 3rd European Semantic Web Conference*, LNCS, pages 411–426, Budva, Montenegro, June 2006. Springer.

[3] R. Jäschke, L. B. Marinho, A. Hotho, L. Schmidt-Thieme, and G. Stumme. Tag recommendations in folksonomies. In *Knowledge Discovery in Databases: PKDD 2007*, volume 4702 of *Lecture Notes in Computer Science*, pages 506–514, Berlin, Heidelberg, 2007. Springer.

[4] M. Lux, M. Granitzer, and R. Kern. Aspects of broad folksonomies. In *DEXA '07: Proceedings of the 18th International Conference on Database and Expert Systems Applications (DEXA 2007)*, pages 283–287, Washington, DC, USA, 2007. IEEE Computer Society.

[5] P. Mika. Ontologies are us: A unified model of social networks and semantics. In *International Semantic Web Conference*, LNCS, pages 522–536. Springer, 2005.

[6] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1349–1380, 2000.

[4] **Acknowledgements**: The Know-Center is funded by the Austrian Competence Center program Kplus under the auspices of the Austrian Ministry of Transport, Innovation and Technology (www.ffg.at), and by the State of Styria.