# Personalized Tag Suggestion for Flickr

Nikhil Garg
Ecole Polytechnique Fédérale de Lausanne
nikhil.garg@epfl.ch

Ingmar Weber
Ecole Polytechnique Fédérale de Lausanne
ingmar.weber@epfl.ch

## ABSTRACT

We present a system for personalized tag suggestion for Flickr: While the user is entering/selecting new tags for a particular picture, the system is suggesting related tags to her, based on the tags that she or other people have used in the past along with (some of) the tags already entered. The suggested tags are dynamically updated with every additional tag entered/selected. We describe three algorithms which can be applied to this problem. In experiments, our best-performing method yields an improvement in precision of 10-15% over a baseline method very similar to the system currently used by Flickr. Our system is accessible at http://ltaa5.epfl.ch/flickr-tags/.

To the best of our knowledge, this is the first study on tag suggestion in a setting where (i) no full text information is available, such as for blogs, (ii) no item has been tagged by more than one person, such as for social bookmarking sites, and (iii) suggestions are dynamically updated, requiring efficient yet effective algorithms.

## Categories and Subject Descriptors

H.5 [**Information Interfaces and Presentation**]: H.5.2 User Interfaces

## General Terms

Algorithms, Human Factors, Languages, Measurement

## Keywords

tagging systems, tag suggestions, experiments, flickr

## 1. INTRODUCTION

Flickr (http://www.flickr.com) is a popular photo sharing site which hosts about 2 billion images. Users can upload and share their pictures, and they can also search among the public pictures. Search is *not* content-based, e.g., you cannot ask for "all pictures showing a jaguar". Rather Flickr offers keyword search: "find all pictures containing the term 'jaguar' in the title or in the description or which have this tag". Titles are often not very useful, as they are typically very short and/or not descriptive, e.g., "freedom" or "love". Similarly, the descriptions usually do *not* describe the picture, but tend to be more of a commentary or poetic nature, e.g., praising the beauty of something or containing a related poem. Tags are usually the best resource for finding or organizing pictures, both personal ones and pictures taken by others, as tags tend to capture the relevant *keywords*. Encouraging users to add more quality tags thus increases the overall value of the system. In this work, we describe algorithms which help to semi-automate the tagging process by suggesting relevant tags to the user, who can then choose to add them (by clicking) or to ignore them (by adding a different tag manually). More concretely, we propose algorithms for the following *personalized tag suggestion problem*:

Given (i) the identity of a user, (ii) an initial (small or even empty) set of tags, as well as (iii) the tagging history of all (or many) users, suggest a ranked list of related tags to the user. The initial set of tags we refer to as a "query".

This problem is independent of any particular application, but we only evaluated our algorithms in the context of Flickr. We consider it most applicable in an interactive setting, where the user adds/selects tags one after another, and the system presents an updated choice to the user at each step. The personalization component helps to (i) find specialized tags, such as names of friends or places, and to (ii) disambiguate the query, by suggesting "car" to a car lover, when the query is "jaguar". Flickr itself offers a simple service which tries to suggest tags: when a user wants to tag a picture, she can select "Choose from your tags" and is then presented with a subset of tags she used in the past. This list is *sorted lexicographically* and the items displayed seem to be a mix of tags used both very recently and very frequently by the user. The set is *independent* of the query, i.e., the tags already entered by the user for the given picture. This approach is similar to our Method 1.

## 2. RELATED WORK

Automatically assigning/suggesting tags to *blog posts* [3, 4] is related to our problem. But as blog posts come with full text, finding similar posts and hence related tags is considerably easier than for pictures. Similarly, automatic tagging of web pages can be improved/personalized, if the system has access to the surfer's desktop [1]. In settings where several users collaboratively tag *the same* items, such as for social bookmarking sites, yet other techniques can be applied [5]. Algorithms for mining association rules [2] are also relevant. But they (i) are comparably slow, (ii) would not give any result if the initial set of tags does not occur in any picture, and (iii) would not generate a ranked result list. Still, we might explore adaptations of those techniques in the future.

## 3. THE ALGORITHMS

All of the following algorithms take as input an instance of the aforementioned problem and output a ranked list of tags, along with a weight $w_t$ for each tag $t$. Weights $w_t^{(a)}$


*WWW 2008,* April 21–25, 2008, Beijing, China.
ACM 978-1-60558-085-2/08/04.

and $w_t^{(b)}$ can thus be easily linearly combined, assuming that weights are normalized to make them comparable.

## 3.1 Method 1: Local & Query-Independent

Let $u$ be the user asking for tag suggestions. Let $n_u(t)$ be the number of times a user $u$ has used tag $t$ in the past. Simply using $w_t^{(1)} = n_u(t)/\sqrt{(\sum_t n_u(t)^2)}$, gives us our first method, which we will use as a baseline. If a user has used "sky" 9 times and "john" 4 times before, it suggests "sky" before "john", regardless of what the user has already entered. Method 1 (or variants using the recency of a tag), which is similar to Flickr's approach, is a natural candidate to use at the beginning when $T(q)$, the set of tags present in the query, is still empty.

## 3.2 Method 2: Local & Query-Dependent

Let $T(p)$ be the set of tags of a picture $p$. Let $P(u)$ be the set of pictures for user $u$. Method 2 first computes a weight for each picture $p \in P(u)$ and each tag $t \in T(p)$ as $w_p(t) = |T(p) \cap T(q)|$ (and $= 0$ if $p \notin P(u)$ or $t \notin T(p)$). The weight for a tag $t$ is then given by ${w'}_t^{(2)} = \sum_{p \in P(u)} w_p(t)$, normalized to $w_t^{(2)} = {w'}_t^{(2)}/\sqrt{(\sum_t {w'}_t^{(2)2})}$. This method gives a higher weight to "john", if the user has already entered "man", and if "man" and "john" were used together in the past.

## 3.3 Method 3: Non-Local & Query-Dependent

Due to space constraints, this method can only be outlined. It proceeds in two phases. First, it identifies a set of promising groups[1]. It does this by building a group-tag matrix and by using cosine similarities both on the user and the group profiles directly (for a query independent user-group similarity) and on the original query which is expanded by Method 2 (for a query-group similarity). Once a small set of promising groups has been identified, it then uses Method 2 for each of these groups to get a list of tags to suggest.

# 4. THE EVALUATION SETUP

Given a set of pictures, which have been tagged by users with $r > k$ tags, we do the following for each picture in the set. First, we remove it from the tag history, so that the system does not know about this combination of tags being used by the particular user. Then we give the first $k$ tags (along with the identity of the user) to the algorithm to evaluate, which has to produce a ranked list of tags. To evaluate the quality of the result list, one can then compute precision-recall measures/graphs, where the remaining $r - k$ tags are treated as *the only* relevant tags. This setup gives an underestimate of the quality of a system, as it assumes that each user has *exhaustively* tagged her pictures, so that any additional tag is considered irrelevant by her.

# 5. EXPERIMENTS

We downloaded roughly 60M publicly accessible pictures for 80K groups. To filter out noise and to reduce the size of the data, we removed tags used less than 50 times. These tags were, however, *not* removed from the pictures on which we ran the evaluation, so that a perfect recall was impossible. We tested the three algorithms on two sets of pictures: (i) 200 pictures with 4-8 given tags and (ii) 200 pictures with 10+ given tags each. Each algorithm was only given the first two tags of each picture, and then produced a ranked list of related tags. For each of the 400 pictures, we also asked a volunteer to evaluate the relevance of about 50 tags, obtained by combining the first 20 results of the three methods, excluding tags present in the original picture. The volunteer used a conservative approach and, e.g., deliberately marked "ears" as irrelevant for a standard portrait or "Europe" for an arbitrary picture taken in a European country.

[1]Groups on flickr are more focused topic-wise than individual users, and groups gave better experimental results.

| Results for 200 pictures with 4-8 tags | | | | |
|---|---|---|---|---|
| | Method 1 | Method 2 | Method 3 | .85·M2+.15·M3 |
| P@1 | .30 (.36) | **.42** (.51) | .24 (.41) | .41 (**.56**) |
| P@5 | .14 (.20) | **.19** (.24) | .12 (.25) | **.19** (**.29**) |
| P@10 | .09 (.13) | **.11** (.15) | .08 (.17) | **.11** (**.19**) |
| P@20 | .05 (.08) | **.06** (.09) | .05 (.11) | **.06** (**.12**) |
| Results for 200 pictures with 10+ tags | | | | |
| | Method 1 | Method 2 | Method 3 | .85·M2+.15·M3 |
| P@1 | .55 (.60) | .67 (**.73**) | .43 (.57) | **.68** (**.73**) |
| P@5 | .35 (.40) | .48 (.54) | .30 (.40) | **.49** (**.57**) |
| P@10 | .25 (.29) | .33 (.39) | .21 (.30) | **.35** (**.42**) |
| P@20 | .17 (.22) | .21 (.26) | .15 (.22) | **.22** (**.29**) |

**Table 1: Precision at tag position 1, 5, 10 and 20 for all three methods and a combination of two schemes. The numbers in parentheses include the additional relevance information from the user study. Numbers in bold indicate the best performing method.**

A combination of Method 2 and 3 gives an improvement of 10-15% for P@1 over Method 1. The fact that the absolute numbers are higher for the second set has two reasons. First, there are more relevant tags to be found. Even with the user evaluation, certain tags such as "vacation" or "friends" were, lacking background information, conservatively chosen as irrelevant.Second, users who add more tags to an individual picture usually have a better tagging history, so that all three methods automatically work better.

Overall, Method 3 benefits the most from the user study, as it suggests relevant tags *outside* the user's vocabulary. Although the improvements of the combined scheme over Method 2 are not dramatic, they are certainly noticable. Currently, we are working on an improved use of the "external knowledge" from groups.

# 6. REFERENCES


[1] P. A. Chirita, S. Costache, W. Nejdl, and S. Handschuh. P-tag: large scale automatic generation of personalized annotation tags for the web. In *The 16th international conference on World Wide Web (WWW)*, pages 845–854, 2007.

[2] J. Hipp, U. Güntzer, and G. Nakhaeizadeh. Algorithms for association rule mining Ů a general survey and comparison. *SIGKDD Explorations Newsletter*, 2(1):58–64, 2000.

[3] G. Mishne. Autotag: a collaborative approach to automated tag assignment for weblog posts. In *The 15th international conference on World Wide Web (WWW)*, pages 953–954, 2006.

[4] S. C. Sood, K. J. Hammond, S. H. Owsley, and L. Birnbaum. TagAssist: Automatic Tag Suggestion for Blog Posts. In *The 1st International Conference on Weblogs and Social Media (ICWSM 2007)*, 2007.

[5] Z. Xu, Y. Fu, J. Mao, and D. Su. Towards the semantic web: Collaborative tag suggestions. *Collaborative Web Tagging Workshop at WWW2006*, 2006.