Automatic people tagging for expertise profiling in the enterprise

Pavel Serdyukov[1*], Mike Taylor[2], Vishwa Vinay[2], Matthew Richardson[3], and Ryen W. White[3]

[1] Yandex, Moscow, Russia
pserdukov@yandex.ru
[2] Microsoft Research, Cambridge, UK
[3] Microsoft Research, Redmond, USA
{mitaylor,vvinay,mattri,ryenw}@microsoft.com

Abstract. In an enterprise search setting, there is a class of queries for which people, rather than documents, are desirable answers. However, presenting users with just a list of names of knowledgeable employees without any description of their expertise may lead to confusion, lack of trust in search results, and abandonment of the search engine. At the same time, building a concise meaningful description for a person is not a trivial summarization task. In this paper, we propose a solution to this problem by automatically tagging people for the purpose of profiling their expertise areas in the scope of the enterprise where they are employed. We address the novel task of automatic people tagging by using a machine learning algorithm that combines evidence that a certain tag is relevant to a certain employee acquired from different sources in the enterprise. We experiment with the data from a large distributed organization, which also allows us to study sources of expertise evidence that have been previously overlooked, such as personal click-through history. The evaluation of the proposed methods shows that our technique clearly outperforms state of the art approaches.

1 Introduction

Members of large organizations frequently search for other people rather than documents. The need for well-informed colleagues is often critical, but manual expert identification through browsing documents or via professional social connections becomes more challenging with every new-hired employee. Expert finding algorithms have been developed to address this problem and retrieve ranked lists of people (rather than documents) in response to a search query. The task of expert finding has recently drawn significant attention from the academic community. However, there is still little reported research on industrial application scenarios of expert finding, which present new challenges and opportunities for the task.

* The work was done while this author was visiting Microsoft Research, Cambridge

One such challenge is the need to summarize the output of an expert finding system by presenting a concise description of expertise for each returned employee. Surprisingly, this task, referred to as *expertise profiling*, has been almost completely neglected by researchers, yet the information presented in such profiles is important to users in deciding which people results to select.

It is challenging to summarize the expertise of a person, since evidence about an employee's knowledge could be spread over many disparate sources in the enterprise. It is also difficult to find pieces of text that would summarize even parts of personal expertise in just a few sentences. At the same time, *tagging* of resources traditionally allows for their concise and meaningful summarization. Automatic *tagging* (*tag suggestion*, or *tag recommendation*) methods have been recently proposed for photos [14] or Web pages [15], but are lacking for people.

In this paper, we propose the novel task of automatic people tagging for the purpose of expertise profiling. In addition our work makes the following contributions:

- We conduct our study in an operational setting within an organization of more than 100,000 employees, using personal data of more than a thousand of them.
- We approach the problem of tag suggestion with a machine learning algorithm that ranks tags by their probability of being a good descriptor of personal expertise for a given employee. We demonstrate that combining expertise evidence extracted from various sources in the enterprise greatly outperforms a state-of-the-art expert finding method adopted for the same purpose, as well as a static ranking of tags by popularity.
- We demonstrate (first, to our knowledge) the usefulness of the enterprise search system's *personal* click-through data for expertise evidence mining.

The remainder of this paper is structured as follows. The related research on expert finding and tag suggestion is reviewed in the next section. In Section 3 we describe our approach for automatic tagging of employees and provide details about the dataset, including a description of all expertise evidence sources examined in this work. Section 4 demonstrates the evaluation of the proposed method. We discuss these results and our work in general in Section 5. Finally, Section 6 summarizes our findings and outlines directions for future research.

2 Related work

Automatic tagging methods were recently proposed for scientific documents/Web pages [15] and photos [14, 7]. This research has been based either on analysis of existing tags with the purpose of extending the given set using tag co-occurrence statistics, or on the "propagation" of tags from tagged to untagged resources using their link graph. Such an approach would be difficult to employ in an enterprise setting, where large volumes of tags and tagged resources are typically unavailable for mining. There is other research that considers tags as topics and treats the tagging task as topical classification using co-occurring tags or terms of tagged documents as features [4]. Such an approach is not only computationally expensive due to the number of classes, but often suffers from the lack of training data for infrequent tags. To the best of our knowledge, there is no research on automatic tag suggestion for people as resources or on automatic tagging of resources whose textual context would be so diffuse as in the case of people. However, the idea of people tagging, as a form of social bookmarking, is not entirely new. Farrel et al. [3] conducted user studies after running a pilot people tagging system at IBM for over a year. They reported the overall satisfaction of users with the application and found that their tags were accurate descriptions of their interests and expertise. However, they did not provide the users with a candidate set of tags potentially relevant for their expertise.

Expert finding is a well-researched problem with a variety of approaches including language-model based [1, 12], data fusion [8] and graph based [13, 6] techniques. The only study to date on using result click-through behavior to enhance expert finding was presented by Macdonald and White [9]. However, they only used clicks as priors for documents, considering employees related to more frequently-clicked documents with respect to a given query as more likely to be experts on its topic. In contrast to their work, we not only analyze all queries leading to the documents authored by the employee under study, but actually focus on analyzing the utility of personal click-through data generated by the employee, including their search queries and clicked documents. Some research has also highlighted the importance of combining expertise evidence of various kind found inside [10] or outside [11] the enterprise using either a linear combination of measured estimates or rank aggregation techniques.

Progress on *expertise profiling* has been slow. Balog and de Rijke [2] experimented with the dataset from the Text Retrieval Conference (TREC 2005) for expert finding task, but instead of ranking candidate experts in response to queries, they ranked queries with respect to candidates. They used only 50 queries, which were actually the names of working groups in the organization under study and (almost) all members of these groups were considered as relevant for the pilot run of the expert task at TREC 2005. In this way, they tried to basically route people to relevant working groups. In their follow-up work, they used a different collection with 1491 test queries [1]. However, their work is limited in that queries were not actually tags (so, were carefully selected by an editor and were not overlapping in meaning) and were ranked using only one feature - the same measure which was used to rank candidate experts for the expert finding task using the same test queries. Our method extends this work by handling more noisy data and employing a more realistic and systematic machine-learning based approach to the problem.

3 Ranking tags for expertise profiling

3.1 Problem definition

We consider a scenario where a representative sample of users in the enterprise has already described their expertise with tags, covering the overall organizational expertise to a reasonable extent. The goal of the proposed system is to automatically assign tags from the *controlled tag vocabulary* created by the initial set of users to other employees in the enterprise.

At the core of our tagging system is a classifier. Given a candidate tag from the controlled vocabulary, the classifier's purpose is to predict whether the tag can be used to describe the expertise of the employee under study. However, instead of a hard prediction, we use the classifier to predict how likely it is that the candidate tag would appear in the employee's personal profile. Using confidence probabilities as tag scores, we generate a ranked list of tags given an employee. The goal of the system is to suggest as many relevant tags at or near the top of the tag ranking as possible. One of the reasons for focusing on the top of the ranking is that employees will most probably inspect the list of expertise tags suggested by the system anyway, but would be much more inclined to see only a short list of tags to select from.

3.2 Data

Fig. 1. Expertise tag cloud

Our research was conducted in a large heterogeneous organization with more than 100,000 employees spread across the globe. We asked volunteers from a range of professions and divisions in the company, including attorneys, administrative staff, software developers, managers and support engineers to provide a list of keywords describing their personal expertise. From this procedure, we acquired 1167 employee profiles, where expertise was described using 4450 unique tags. Profiles contained around 8.5 of these tags on average and 5.5 of these tags were used in more than one profile. Average tag length was 1.47 words. Figure 1 shows the top-100 most popular tags used by employees (font size indicates popularity of a tag).

In order to simulate the above-described scenario (where the tags assigned for the subset of users are predicted for those with no tags), we considered a random sample of 700 profiles (60%) as our training set, 167 profiles as our validation set (14%) to tune parameters in our methods and 300 profiles as our test set (26%), which we tried to build automatically. 1275 tags used by employees in the training set and that appeared in at least two profiles were considered as candidates for ranking. We did not consider tags appearing in only one profile for two reasons. First, we wanted to highlight expertise which is less personal and more specific to the organization, which means that it should be possessed by at least two employees. Second, we wanted to avoid using overly-specific descriptions of a certain expertise area, but rather predict more common tags with a similar meaning, which we hoped would increase the readability of the constructed profiles. In the absence of relevance judgments for each candidate tag, we considered those tags specified by users in their profiles as *relevant* (positive examples) and all other tags not used to describe their expertise as *non-relevant* (negative examples). Our goal was then to predict how likely a tag is to be relevant given an employee using features and expertise evidence sources described later in this section.

3.3 Extracting features from expertise evidence sources

Expert finding methods traditionally rely on aggregates of relevance measures calculated over all informational units related to the employee [1, 8]. In most cases, for example, they regard the sum of relevance scores of documents mentioning the employee as a measure of personal expertise on the given topic. In this work, we decided not to rely only on one stream of textual context of an employee or only on one measure indicating the strength of relation between a person and a tag. Using different definitions of a tag's relevance to an employee's expertise and all accessible streams of evidence, we extracted the features described further in this section.

To align with previous research [1, 2], we used a language modeling approach to information retrieval to obtain an estimate for the probability of relevance $P(e,t)$ of the tag t in respect to the personal expertise of the employee e given an evidence stream S (e.g., a set of authored documents):

$$P(e,t) = \sum_{D \in S} P(e,t|D)P(D) \tag{1}$$

$$P(e,t|D) = P(e|D)P(t|D) = P(e|D)\prod_{w \in t} P(w|D) \tag{2}$$

where w is the word from the tag t, $P(D)$ is the document's D prior probability of relevance, whose distribution is uniform, $P(e|D)$ is the probability of relation between the person e and the document D, which we considered binary in our work, as was often done earlier [8, 6]. The probability to generate the term w from the document's language model [5]:

$$P(w|D) = (1-\lambda_G)\frac{c(w,D)}{|D|} + \lambda_G P(w|G), \tag{3}$$

where $c(w,D)$ is the count of the term w in the document D, $|D|$ is its length, λ_G is the probability that term w will be generated from the global language model $P(w|G)$, which is estimated over the entire set of existing documents for all employees. Following previous studies, we set the λ_G to 0.5 in our experiments. Alternatively, apart from the language model based estimate of tag relevance given a document (***LM***) we considered the simple (***Binary***) model, which assumes that the probability $P(t|D) = 1.0$, when the tag t appears in the document D as *a phrase* at least once, and $P(t|D) = 0$ otherwise.

There is an important difference between scoring tags for expertise profiling and scoring employees for expert finding. It is clear that some tags will appear frequently in sources related to employees, and will not be descriptive of their expertise. Such tags will tend to dominate the top of the ranking just because they are generally very frequent. At the same time, it is intuitively important not only to be "rich in a tag", to be an expert on the topic that it covers, but also to be richer than an average employee. In addition to the features described above, we also calculated their deviations from the averages measured on the set of training profiles:

$$P(e,t)^{dev} = P(e,t) - \frac{1}{|train|} \sum_{e' \in train} P(e',t) \tag{4}$$

Note that such transformation would not affect the rank ordering for expert finding, where employees are ranked given a tag (query), so the subtraction of any tag-specific constant for each employee's score does not affect their rank. However, it does change the ranking for tags, since each tag has a different average score in the training set. As we show in our experiments, such transformation is also beneficial for performance.

As a result, given an employee and a tag, we calculated two scores, based on probabilistic (***LM***) and on binary model of relevance (***Binary***), for each set of informational units in each stream related to the employee and used these *scores* with their *deviations* as individual features for tag-employee pairs.

3.4 Additional features

There are also features that are obtainable for a tag even with no information about a particular employee. We experimented with such features, including profile frequency of a tag in the training set, inverted document frequency, and the tag length in words or characters. According to our preliminary evaluation on the validation set (see Section 3.2), only the frequency of a tag in expertise profiles was predictive of the tag's relevance, which we further used in our experiments as a feature.

3.5 Expertise evidence sources

We used a variety of information streams related to an employee to calculate the above mentioned features:

– **Authored and Related enterprise documents**. We crawled all documents authored by each person and also related documents which contained the employee's full name and email address. This approach for discovering relationships among employees and documents is well known in expert finding research and we follow the traditional path here [8, 13]. Authored documents are found by examining their metadata. As a result, we had 226 authored and 76 related documents per person on average, including articles, presentations, spreadsheets, etc. We considered each document field as an individual source of expertise evidence: ***Title***, ***File Name***, ***Summary***, ***Content***[4].

[4] Bolded and capitalized names are used later to refer to the specific type of evidence.

– **Web documents**. Recent studies have shown that the evidence located outside the enterprise might be even more valuable than that found by analyzing internal sources [11]. We used the API of a major Web search engine to search for full names and email addresses of employees on the Web. Unfortunately, we could find only on average four related Web documents per person. The same above-mentioned document fields served as evidence sources.
– **Discussion lists**. Each employee in the enterprise under study is able to ask for assistance by sending emails to one of a set of existing discussion lists within the organization, which are monitored by other employees interested or knowledgeable on the topic of the list. While we did not have access to the content of questions and answers (while they are disseminated via discussion lists, their largest part stays strictly personal and the rest is not stored permanently), we had access to the employees' subscriptions - 172 per person on average - and used the ***Name*** of the related discussion list as a source of evidence.
– **Enterprise search click-through**. We used six months of search logs from January 2010 through June 2010 inclusive, obtained from thousands of users querying the Intranet of the organization under study. Click-through expertise evidence was extracted from three sources for each employee: 1) All ***Personal Queries*** issued by the employee: 67 unique queries per person on average; 2) Above-mentioned fields (title, file name, summary, content) of documents ***Clicked*** for these queries: 47 unique documents on average per person; 3) ***Queries*** of any employees that led to clicks on the documents authored by the employee: 12 unique queries on average per person.

4 Experiments

4.1 Baselines and evaluation measures

We consider two baselines in our work, also used as features to train our learning algorithm.

– **Application-specific baseline**. In the case, when there is no information about employees (for example, about those just entering the company), the only source of evidence about tag relevance is its prior probability in the training set. That is why we regard the ***ProfileFrequency*** of a tag to be one of our baselines.
– **State-of-the-art baseline**. Another baseline is taken from the only existing work on expertise profiling and represents a sum of scores of related documents with respect to the tag [2]. During preliminary experimentation with this baseline, we observed that it became relatively competitive only when using authored documents. Hereafter, we refer to this baseline and feature as ***AuthoredContentLM***.

Since we regard our task as a problem of ranking tags, we evaluate our performance using the set of standard IR measures:

- Precisions at 1, 5 and 10 ranked candidate tags (P@1, P@5 and P@10),
- Success at 5 (S@5), showing the ability of the system to predict at least one relevant tag among the top ranked 5,
- Average Precision (AP).

We focus mainly on precision and success measures, since we consider them to be correlated with user satisfaction under the setup described in Section 3.1. In other words, since we assume that the purpose of the tags is to help users know more about each other, errors in the top ranks may lead to a greater effect on the impression of an employee's personal expertise.

4.2 Learning for tag ranking

We experimented with a number of state-of-the-art classification algorithms on our validation set (see Section 3.2) and finally decided to use the output of the logistic regression function to rank tags, based on its performance and stability. Each pair of a tag and the profile where the tag appears served as a positive training example and all other non-matching pairs served as negative examples. In total, we used 4,098 positive and 60,000 negative examples. We sampled our negative examples from the entire set containing 900,000 negatives to avoid severe imbalance and the size of the sample was tuned on the validation set as well. Each combination of one of two expertise measures and their deviations (see Section 3.3) and one of the expertise evidence streams (see Section 3.5) resulted in a feature. In total, we trained using 78 features. Later in this section, we analyze the influence of features and streams on the performance of our learning model. Please note that we refer to specific features using bolded names of measures, evidence sources and specific document fields mentioned in Sections 3.3 and 3.5. For example, feature ***RelatedSummaryLM*** is obtained by summing the ***LM*** based probabilities of relevance of a tag in respect to ***Summary*** fields of ***Related*** enterprise documents.

However, first, we demonstrate the individual performance of the top most predictive features from each stream, including baselines, to give an idea about which of them appeared to be the most useful for the learning algorithm (Table 1, left part). We also show the performance of deviations from the average for these features (see Section 3.3, Equation 4), which prove that such a feature transformation is useful in almost all cases (Table 1, right part). As we see from Table 1, neither of the two baselines is able to outperform the strongest features extracted from distribution lists, as well as filenames and titles of authored documents. Features extracted from the set of personal queries appear to be the most useful among click-through features. They outperform the state-of-the-art performance baseline feature, but still perform worse than the application-specific baseline. Features of the streams that failed to provide competitive evidence for expertise mining are not included in the table due to space constraints.

As we also see in Table 1, our learning algorithm using ***ALL*** features is able to learn a ranking function which greatly outperforms both baselines. It also greatly outperforms the strongest feature ***ListNamesBinary*** (see Table 1), improving

	Feature performance			Deviation performance		
Stream	P@5	AP	S@5	P@5	AP	S@5
ProfileFrequency (baseline)	0.066	0.046	0.253	-	-	-
AuthoredContentLM (baseline)	0.044	0.030	0.180	0.081	0.057	0.310
ListNamesBinary	**0.122**	**0.086**	**0.437**	**0.125**	**0.087**	**0.437**
AuthoredFileNamesBinary	0.071	0.058	0.3	0.093	0.066	0.367
AuthoredTitlesLM	0.072	0.053	0.283	0.085	0.059	0.313
PersonalQueriesBinary	0.055	0.040	0.210	0.059	0.041	0.220
QueriesToAuthoredBinary	0.059	0.038	0.230	0.069	0.043	0.307
RelatedSummaryLM	0.035	0.024	0.163	0.059	0.041	0.257
ClickedTitlesBinary	0.021	0.018	0.087	0.033	0.025	0.133
WebTitlesLM	0.023	0.012	0.093	0.023	0.013	0.097
ALL features	**0.171**	**0.124**	**0.543**	-	-	-

Table 1. Performance of feature groups

P@5 by 40%, AP by 42% and S@5 by 24%. It demonstrates the importance of combining all sources of evidence in the enterprise into one inference mechanisim for the high-quality expertise mining and tag suggestion.

We also studied the contribution of features from each stream via feature ablation, removing the next most useful stream (according to P@5) at each step (until only one stream is left) and observing the change in performance. Rows with (**-**) in Table 2 demonstrate the order in which the groups of features were removed. As appeared, profile frequency was the most important feature in this regard, since it caused the most severe drop in performance, followed by distribution lists, authored documents and personal queries. Since features can interact in the learning model, we also experimented with removing each feature group at a time while leaving the other ones and observing the drops in performance, but it resulted in the same relative importance of streams.

Stream	P@1	P@5	P@10	AP	S@5
ALL	**0.266**	**0.171**	**0.122**	**0.124**	**0.543**
- *ProfileFrequency*	0.240	0.138	0.102	0.110	0.460
- *List*	0.146	0.096	0.073	0.074	0.370
- *Authored*	0.130	0.078	0.065	0.063	0.300
- *PersonalQueries*	0.090	0.057	0.046	0.047	0.250
- *Related*	0.060	0.053	0.044	0.030	0.220
- *Clicked*	0.033	0.046	0.039	0.025	0.180
- *Web*	0.010	0.034	0.032	0.019	0.143
- *QueriesToAuthored*	0	0.025	0.023	0.007	0.123

Table 2. Performance of feature groups

4.3 Importance of click-through

One of the goals of this work was to study the importance of the expertise evidence mined from queries and clicks. Although Tables 1 and 2 show that some of the click-through streams are useful and certainly not inferior to many other streams, it was still important to understand how the system would perform entirely without the click-through data, including personal queries, fields of clicked documents and queries to the documents authored by the employee. As we see from Table 3, the performance indeed drops when we remove all click-through features, for almost all measures by around 6-9%. It confirms the intuition that personal clicktrough history is suitable for mining personal expertise and its contribution is valuable for our inference mechanism.

However, as we noticed from the performance analysis of individual features and streams, some of them are very strong, while actually being very specific to the enterprise under study. We can imagine an enterprise with no such thing as an initial set of user profiles (considering that the controlled vocabulary was extracted from a different source, e.g., a query log) and distribution lists. For example, no enterprises studied previously in related academic research (see Section 2) actually considered these sources of expertise evidence. Such a "typical" enterprise would however be able to gather additional click-through evidence by monitoring users of its Intranet search system. We simulate the performance of our algorithm in such an enterprise by removing lists and profile frequency features. As we see in Table 3, we observe a severe drop in performance when we also exclude click-through features at the next step after that: over P@1 by 37%, over P@5 by 19%, over P@10 by 13%, over AP by 32%, over S@5 by 16%.

Stream	P@1	P@5	P@10	AP	S@5
ALL	0.266	0.171	0.122	0.124	0.543
ALL - Click-through	0.266	0.160	0.112	0.117	0.513
Typical enterprise: *ALL - Lists - ProfileFrequency*	0.146	0.096	0.073	0.074	0.37
Typical enterprise - *Click-through*	0.093	0.078	0.056	0.050	0.310
Click-through (only)	0.09	0.061	0.047	0.048	0.247

Table 3. Performance with/without click-through features

5 Discussion

The findings presented in the previous section show that it is possible to effectively provide members of an organization with a relevant ranking of tags describing their expertise. Building such a system has multiple advantages, the main one of which is that we now have information for people who have not gone through the effort of tagging themselves. Encouraging people to use such a system will have benefits that go beyond the profiling, i.e., related task like expertise search will improve as well. It was also clear that none of the text streams of expertise evidence is sufficient alone to attain the maximum performance of the tag ranking. For example, features extracted from personal click-through, while not being the most predictive on their own, appeared to be highly valuable as a part of the complete feature set.

Some of our findings contradicted previous research. We were particularly surprised by the poor performance of the features extracted from Web documents, considering that Serdyukov and Hiemstra [11] found that Web documents provided the most valuable source of expertise evidence when they experimented with expert finding in a large research institution. It seems that it is possible to obtain a sufficient amount of Web-based expertise evidence only for organizations motivating their employees for regular public activities (e.g., research labs advocating the publication of research results), and this is not often the case for most commercial organizations.

We also analyzed the tags that our algorithm failed to predict over all test employees. In most cases, these were tags whose relevance was hard to evaluate using enterprise data. There were generally two classes of these tags: 1) too personal tags (e.g., "*ice cream*","*cooking*","*dancing*","*judaism*"), and 2) abstract tags (e.g., "*customer satisfaction*", "*public speaking*", "*best practices*"). While, in the first case, such personal tags simply could not be found in work-related evidence sources, in the second case, abstract tags were too general to be often used in the text or even in personal queries.

Another source of "errors" in prediction was the consequence of our definition of relevance used due to the origin of our evaluation set. We tried to predict the exact tag used by the employee to describe own expertise, so only these tags were regarded as relevant. However, this is a much more difficult task than the task of prediction of any relevant tags. In our case, for example, "*machine learning*" and "*data mining*", or "*networking*" and "*networks*", were not necessarily both relevant for the same employee, while it is unusual to possess expertise on only one of these subjects. The success measure used in our work (S@5) is more robust to this issue, since it gives full credit for those test employees for whom the model predicted at least one of correct tags in their profile.

6 Conclusions and Further Work

In this paper we have proposed a technique for automatic tagging of employees in the enterprise for the purpose of expertise profiling. Assuming that an initial set of employees have described themselves with tags, we investigated the utility of different sources of evidence of a person's expertise as predictors of tags they are likely to use. We were able to train a classifier that produced candidate expertise terms for an as-yet untagged person, thereby providing an automatic profiling mechanism. We have thoroughly tested this technique using data from a large organization and a variety of expertise evidence sources, including those that have not been studied before in the scope of the tasks focused on expertise location, such as personal click-through history.

Our experiences suggest that when asked to describe their own expertise, the words chosen by employees of an enterprise can only partially be inferred from the enterprise content that can be associated with them. Modeling of the remaining *knowledge* of a person remains a challenge. Our experiments also indicated that the problem of ranking of tags for people involve considerations that are familiar to other retrieval tasks, such as the need to diversify the ranked list. Besides, it is important to investigate not only how to build personal profiles, but also how to make these summaries query-dependent to dynamically appropriate them as result snippets for a people search engine. Addressing such concerns will be the subject of our future research.

Acknowledgments. We would like to thank Milad Shokouhi, Emine Yilmaz and Gjergji Kasneci for lots of helpful comments and inspiring discussions.

References

1. K. Balog, T. Bogers, L. Azzopardi, M. de Rijke, and A. van den Bosch. Broad expertise retrieval in sparse data environments. In *SIGIR '07*, pages 551–558. ACM, 2007.
2. K. Balog and M. de Rijke. Determining expert profiles (with an application to expert finding). In *IJCAI-2007*, pages 2657–2662. Morgan Kaufmann Publishers, 2007.
3. S. Farrell, T. Lau, S. Nusser, E. Wilcox, and M. Muller. Socially augmenting employee profiles with people-tagging. In *UIST '07: Proceedings of the 20th Annual ACM symposium on User Interface Software and Technology*, pages 91–100. ACM, 2007.
4. P. Heymann, D. Ramage, and H. Garcia-Molina. Social tag prediction. In *SIGIR '08*, pages 531–538. ACM, 2008.
5. D. Hiemstra and F. M. G. de Jong. Statistical language models and information retrieval: Natural language processing really meets retrieval. *Glot international*, 5(8):288–293, 2001.
6. M. Karimzadehgan, R. W. White, and M. Richardson. Enhancing expert finding using organizational hierarchies. In *ECIR '09*, pages 177–188. Springer-Verlag, 2009.
7. D. Liu, X.-S. Hua, L. Yang, M. Wang, and H.-J. Zhang. Tag ranking. In *WWW '09*, pages 351–360. ACM, 2009.
8. C. Macdonald and I. Ounis. Voting for candidates: adapting data fusion techniques for an expert search task. In *CIKM '06*, pages 387–396. ACM, 2006.
9. C. Macdonald and R. W. White. Usefulness of click-through data in expert search. In *SIGIR '09*, pages 816–817. ACM, 2009.
10. D. Petkova and W. B. Croft. Hierarchical language models for expert finding in enterprise corpora. In *ICTAI '06: Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence*, pages 599–608. IEEE, 2006.
11. P. Serdyukov and D. Hiemstra. Being Omnipresent to Be Almighty: The Importance of the Global Web Evidence for Organizational Expert Finding. In *FCHER'08: Proceedings of the SIGIR'08 Workshop on Future Challenges in Expertise Retrieval*, 2008.
12. P. Serdyukov and D. Hiemstra. Modeling documents as mixtures of persons for expert finding. In *ECIR*, pages 309–320. Springer-Verlag, 2008.
13. P. Serdyukov, H. Rode, and D. Hiemstra. Modeling multi-step relevance propagation for expert finding. In *CIKM '08*, pages 1133–1142. ACM, 2008.
14. B. Sigurbjörnsson and R. van Zwol. Flickr tag recommendation based on collective knowledge. In *WWW '08*, pages 327–336. ACM, 2008.
15. Y. Song, Z. Zhuang, H. Li, Q. Zhao, J. Li, W.-C. Lee, and C. L. Giles. Real-time automatic tag recommendation. In *SIGIR '08*, pages 515–522. ACM, 2008.