Determining Expert Profiles
(With an Application to Expert Finding)

Krisztian Balog and Maarten de Rijke
ISLA, University of Amsterdam
Kruislaan 403, 1098 SJ Amsterdam
kbalog,mdr@science.uva.nl

Abstract

The profile of an individual is a record of the types and areas of skills of that individual ("topical profile") plus a description of her collaboration network ("social profile"). In this paper we define and formalize the task of automatically determining an expert profile of a person from a heterogeneous corpus made up of a large organization's intranet. We propose multiple models for addressing the topical profiling task. Our main methods build on ideas from information retrieval, while refinements bring in filtering (allowing an area into a person's profile only if she is among the top ranking experts in the area). An evaluation based on the W3C-corpus made available by TREC, shows significant improvements of the refined methods over the baseline. We apply our profiling algorithms to significantly enhance the performance of a state-of-the-art expert finding algorithm and to help users of an operational expert search system find the person they would contact, given a specific problem, topic or information need. Finally, we address the task of determining a social profile for a given person, using graph-based methods.

1 Introduction

Some of the most valuable knowledge in an enterprise resides in the minds of its employees. Enterprises must combine digital information with the knowledge and experience of employees. Expert finding addresses the task of finding the right person with the appropriate skills and knowledge: "Who are the experts on topic X?" The task has recently received increased attention, especially since the launch of an expert finding task as part of the enterprise track at TREC in 2005 [TREC, 2005]. Given a query (describing the area in which expertise is being sought), participating systems have to return a ranked list of person names in response.

Like most tasks assessed at retrieval evaluation platforms such as TREC, NTCIR, CLEF, and INEX, the expert finding task is an abstraction of a real task. The abstractions are important when trying to set up experiments that will lead to stable and re-usable test sets [Voorhees and Harman, 2005]. But when people search for expertise, they are often looking for experts, but not in isolation—the desired output should be more than a ranked list of person names [Hawking, 2004]. Context and evidence are needed to help users of expertise finding systems decide whom to contact when seeking expertise in some area. E.g., given an expert whose name is returned in response to a query, what are her areas of expertise? Who does she work with? What are her contact details? Is she well-connected, just in case she is not able to help us herself?

The main aim of this paper is to introduce the task of determining an expert's profile—i.e., a concise description of the areas in which she is an expert plus a description of her collaboration environment—, and to devise and assess algorithms that address this profiling task. To make matters more concrete, let us look at an expert finding system that is currently being developed; it started out as "a ranked list of person names" system and is now evolving so as to include the type of context and evidence discussed above. Figure 1 provides a screen dump. In Figure 1 we see the information dis-

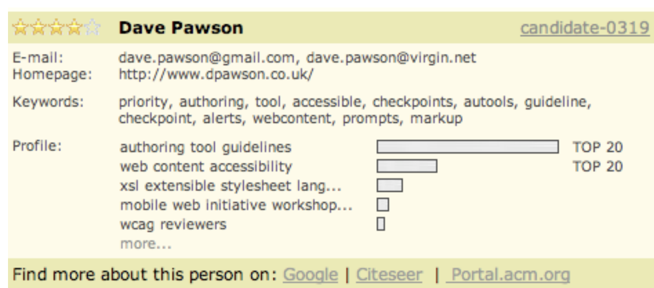

Figure 1: Screen dump of the expert search interface.

played for one person (one of the hits produced in response to a query on "authoring tools"). In the top row we see the person's name (optionally his internal id or username), and his relevance for the given topical query. Below this, the contact details (e-mail, web address, phone, fax number) are shown. The keywords serve as a type of context, in the form of a "tag cloud," describing the general interests and activities of the person; these keywords are extracted from documents that the person is associated with. The candidate's *topical profile* is presented as a list of knowledge areas, and the level of competence in each (which is reflected by the size of the bars in our example). The relative ranking is shown when the person is among the top experts given that field. (As an aside the full interface also includes links to documents associated with the person, as well as a link to a figure depicting the candidate's social profile; see Section 5 for more on the latter.)

Our main contribution in this paper is the introduction of expert profiling as a task, together with algorithms for addressing the task. We first ask whether existing expert finding algorithms can be used for topical profiling—effectively, these algorithms build up a matrix with experts along one dimension, and topical areas along the other: given an area, determine the candidate experts with the highest expertise levels. Can these algorithms be "inverted": given a candidate, find the areas for which her expertise levels are highest?

Our answer to this question is a clear "No"—implying that expert finding and expert profiling are two distinct tasks. The next question, then, becomes: How can we compute expert profiles? We propose two models for extracting expert profiles; the first uses information retrieval techniques to obtain a set of relevant documents for a given knowledge area, and aggregates the relevance of those documents that are associated with the given person. Our second model represents both candidates and knowledge areas as a set of keywords, and the skills of an individual are estimated based on the overlap between these sets. We demonstrate that both models perform significantly better than the "inverted" expert search results, which serves as our baseline. Moreover, we introduce a filtering algorithm, which rejects areas to be part of the individual's profile, if there is not enough evidence found to support that the person has reasonably high knowledge on that topic, compared to others within the enterprise.

While profiling algorithms are important for feeding the sort of interface described above, they are also helpful in other ways within expertise finding scenarios: we show that they can be used for re-ranking expert search results. Our re-ranking method is very effective, improving upon state-of-the-art expert finding methods in terms of mean average precision, mean reciprocal rank, as well as precision@5 scores.

Colleagues and collaborators play an important role in the value of a person as an expert: an "isolated" expert might be able to answer specific questions, but a well-connected expert might put us on track to explore new or additional areas [Cross *et al.*, 2001]. A *social profile* of a person contains information about her collaborative network. We propose to capture the social profile of a person by examining the documents associated with her, and determining which other people are also associated with those documents.

In Section 2 we provide background on expert finding and profiling. Section 3 is devoted to topical profiling. In Section 4 we put expert profiles to work to improve expert finding algorithms. Then, in Section 5 we turn to social profiles. We conclude in Section 6.

2 Background

Initial approaches to expert finding employed a database housing the skills and knowledge of each individual in the organization [Maron *et al.*, 1986; Davenport and Prusak, 1998], and were mainly focused on how to unify disparate and dissimilar databases of the organization into one data warehouse that can easily be mined. Most of this early work was performed by the Knowledge Management and Computer Supported Cooperative Work community, usually called yellow pages, people-finding systems or expertise-management [EC-SCW, 1999]. Most of the tools rely on people to self-assess their skill against a predefined set of keywords, and there is a need for intelligent technologies that could enhance the process of updating profiles [Becerra-Fernandez, 2000].

More recently there has been a move to automatically extract such representations from heterogeneous document collections such as those found within a corporate intranet [Craswell *et al.*, 2001]. However, until recently much of the work in this area has been performed in industry with only sketchy solutions, tailored to specific organizational needs, and without formal evaluations.

TREC 2005 [TREC, 2005] has introduced the expert finding task and provided a common platform with the Enterprise Search Track for researchers to evaluate and assess methods and techniques. The following scenario is presented: Given a crawl of the World Wide Web Consortium's web site, a list of candidate experts and a set of topics, the task is to find experts (provide a ranked list of candidates) for each of these topics. This scenario is ideal for measuring the accuracy and effectiveness of different methods, and for making a fair comparison between the performance of expert finder systems.

To the best of our knowledge, the task introduced in this paper—expert profiling—is new. More precisely, *topical* expert profiling has not been previously identified as a task amenable to computational approaches. In contrast, *social* profiling (finding collaboration networks around an expert) is a task that has been gaining increasing attention, especially from researchers with a background in social network analysis. See, e.g., [Abrol *et al.*, 2002] for an industrial example.

3 Topical Profiles

In this section we define what a topical profile is, and propose two methods for creating such profiles.

A topical profile of an individual is a record of the types and areas of skills and knowledge of that individual, together with an identification of levels of 'competency' in each ("What does expert Y know?"). We model the profile of a candidate as a vector, where each element of the vector corresponds the person's skills on the given knowledge area. This skill is expressed by a score (not a probability), reflecting the person's knowledge on the given topic.

3.1 Algorithm

The output of a profiling algorithm is a ranked list of knowledge areas, the estimated level of competency in each as well as the evidence that supports these results (e.g., a list of documents). The task of determining expert profiles is naturally decomposed into two stages.

1. Discovering and identifying possible *knowledge areas*.
2. Measuring the person's competency in these areas.

We assume that a list of possible knowledge areas ($KA = \{ka_i | i = 1, \ldots, n\}$) is given, and restrict ourselves to stage 2). We represent the profile of a person ca as a vector of n values, where the ith element of the vector corresponds to the knowledge area ka_i, and $score(ca, ka_i)$ reflects the individual's knowledge in the given area:

$$profile(ca) = \langle score(ca, ka_1), score(ca, ka_2), \ldots, score(ca, ka_n)\rangle$$

Baseline

As a baseline topical profiling method we use the results generated by an (existing) expert search method [Balog *et al.*, 2006]. The expert search method estimates $p_{ES}(ca|q)$: what is the probability of a candidate ca being an expert given the query topic (knowledge area) q? Let $rank_{ES}(ca, q) = 1, \ldots, m$ be the rank of candidate ca on topic q, where the ranking of candidates is proportional to $p_{ES}(ca|q)$:

$$p_{ES}(ca_i|q) \geq p_{ES}(ca_j|q) \Rightarrow rank_{ES}(ca_i, q) \leq rank_{ES}(ca_j, q).$$

We can use both the scores and the ranking generated by the expert search method to estimate the level of competence (or knowledge) of a given candidates. That is

- Baseline(probability): $score(ca, ka) = p_{ES}(ca|ka)$
- Baseline(rank): $score(ca, ka) = 1/rank_{ES}(ca)$

A shortcoming of forming scores this way is that they are relative to the capabilities of other candidates, and do not reflect the individual's absolute knowledge. Hence, given a knowledge area, we assume the candidate's skill to be higher if she is a higher ranked expert on the corresponding topic.

Method 1

We now describe the first of two profiling methods that go beyond the baseline. The intuition behind this first method is that a person's skill can be represented as a score over documents that are relevant given a knowledge area.

For each knowledge area ka a query-biased subset of documents D_{ka} is obtained by using the top n documents retrieved for the query ka. We iterate over the relevant documents, and sum up the relevance of those that are associated with the given candidate. Formally, the score of an individual ca given the knowledge area ka is:

$$score(ca, ka) = \sum_{d \in D_{ka}} relevance(d, ka) A(d, ca). \quad (1)$$

The association method $A(d, ca)$ returns 1 if the name or the e-mail address of person ca appears in the document d; otherwise it returns 0. The recognition of candidates is approached as a (restricted and) specialized named entity recognition task. We do not differentiate between the roles of the person (author, contact person, mail recipient, etc.), or the level of the contribution the person may have made to d.

To estimate the relevance of a document, we use standard generative language model techniques [Baeza-Yates and Ribeiro-Neto, 1999]. Specifically, $relevance(d, ka) = p(ka|\theta_d)$ expresses how likely the document d would generate a certain knowledge area ka. The document model is constructed by taking a product of a linear combination of the background model $p(t)$ and the smoothed estimate for each term $t \in q$:

$$p(q|\theta_d) = \prod_{t \in q} \Big\{(1-\lambda)p(t|d) + \lambda p(t)\Big\} \quad (2)$$

The specific choice of language modeling (LM) techniques is pragmatic. For the topical profiling task, we could have used any other information retrieval method for ranking documents. Our recent work on expert search used LM techniques, and by re-using that work we can analyze, and obtain a better understanding of, differences and similarities between the expert finding and profiling tasks.

Conceptually, this method is similar to the expert finding method Model 2, introduced by [Balog *et al.*, 2006], but associations are not turned into probabilities, thus their strength is not estimated—practically (and realistically) speaking, we simply cannot capture the extent to which the candidate is responsible for a document's content, compared to other individuals that may also be associated with the same document.

Method 2

Our next method takes a completely different approach, where the profiling scores are estimated using keyword similarity of candidates and knowledge areas. We first tokenize documents, and remove standard stopwords (no stemming is applied), then extract the top 20 keywords for each document, using the TF·IDF weighting formula [Baeza-Yates and Ribeiro-Neto, 1999]. Let $KW(d)$ be the set of keywords that are extracted from document d.

We can obtain the set of keywords of a knowledge area ka by looking at a query-biased subset of documents D_{ka}, using the top n documents retrieved using ka as the query and the W3C corpus as the document collection. Then we use all keywords from these documents to form a list of keywords of the given area. Formally:

$$KW_{ka} = \bigcup_{d \in D_{ka}} KW(d) \quad (3)$$

Similarly, we can obtain keywords of individuals, by taking keywords of documents that they are associated with:

$$KW_{ca} = \bigcup_{d \in D, A(d,ca)=1} KW(d) \quad (4)$$

Having the set of keywords both for knowledge areas and for candidates in hand, we estimate the profile scores with the ratio of co-occuring keywords:

$$score(ca, ka) = |KW_{ka} \cap KW_{ca}| / |KW_{ka}|. \quad (5)$$

Filtering

We introduce a filtering technique that is to be applied on top of an existing profiling method. In the experimental section we present its performance both on Method 1 and on Method 2. The intuition behind the method is to provide us with a natural cut-off point that will restrict the number of people returned in a profiling setting so as to only retain experts whose topical profile we are highly confident about: a knowledge area can be part of an individual's profile if and only if the person is among the top ranked experts on that field. This means that she has a reasonable knowledge on the given topic both individually and compared to others.

We allow a given knowledge area to appear in a candidate's profile only if the individual scores within the top f among all candidates on that field. Actually, we create a ranking of experts, using the profile scores, and hence solve an expert finding task. To do that, we need a list of potential expert candidates, which we assume to be given. Then we use the

Method	MAP	MRR
Baseline(probability)	0.320	0.397
Baseline(rank)	0.203	0.244
Method 1	**0.407**	**0.503**
Method 2	0.397	0.486

Table 1: Results of the topical profiling methods. The columns are MAP: Mean Average Precision, and MRR: Mean Reciprocal Rank. Best scores are in boldface.

results of the expert finding task to refine the output of the profiling method. Formally,

$$score'(ca, ka) = \begin{cases} score(ca, ka). \\ \quad \text{if } |\{ca' | score(ca', ka) < score(ca, ka)\}| < f \\ 0, \qquad\qquad\qquad\qquad \text{otherwise} \end{cases}$$

3.2 Evaluation

We performed experiments to answer the following questions: Is "inverted expert finding" a viable solution to the profiling problem? How well do Method 1 and Method 2 perform? And what is the impact of filtering?

Experimental Setup

The document collection we use is the W3C corpus [W3C, 2005], a heterogenous document repository containing a mixture of different document types crawled from the W3C website. The corpus contains 330,037 documents, adding up to 5.7GB. A list of 1,092 candidate experts was made available, where each candidate is described with a unique candidate_id, name(s) and one or more e-mail addresses. We took the topics created within the expert finding task of the 2005 edition of the TREC Enterprise track: 50 in total; these are the topics for which experts have to be sought.

Profiling

In our evaluation methodology we utilize the fact that the TREC 2005 topics are actually names of W3C working groups, and experts on a given topic are members of the corresponding working group. A person may be member of multiple working groups, meaning that she is an expert on multiple topics. We use the W3C working group names as knowledge areas, and consider a knowledge area to be part of the individual's profile if the person is a member of the corresponding working group. That is, we reverse the original TREC expert search assessments.

Table 1 reports the results of our experiments. The answer to our first question (Can an expert search algorithm be "inverted" to obtain an effective topical profiling method?) is a clear "No" since both profiling methods significantly outperform both baselines; the probability baseline significantly outperforms the rank baseline.[1] Hence, expert finding and topical profiling are two different tasks. Their mathematical foundation is common, since in both cases a candidate-topic

[1]To determine whether the observed differences are statistically significant, we use the two-tailed Wilcoxon Matchedpair Signed-Ranks Test, and look for improvements at a significance level of 0.05.

	retrieved		filtered	
Method	MAP	MRR	removed	error rate
Method 1	0.407	0.503	–	–
f=150	0.403	0.509	0.330	0.100
f=100	0.395	0.511	0.508	0.116
f=50	0.395	0.535	0.737	0.156
f=30	0.392	0.558	0.846	0.193
f=20	0.388	0.584	0.900	0.232
f=15	**0.408**	0.649	0.926	0.277
f=10	0.385	**0.677**	0.951	0.309
f= 5	0.350	0.654	0.978	0.392
Method 2	**0.397**	0.486	–	–
f=200	0.355	0.463	0.625	0.058
f=150	**0.383**	**0.511**	0.718	0.065
f=100	0.372	0.511	0.813	0.075
f=50	0.303	0.476	0.907	0.082
f=30	0.266	0.450	0.944	0.094
f=20	0.247	0.461	0.962	0.108
f=15	0.249	0.466	0.972	0.118
f=10	0.229	0.438	0.981	0.132
f=5	0.286	0.463	0.990	0.176

Table 2: Results of filtering on the topical profiling methods. The columns are MAP: Mean Average Precision, MRR: Mean Reciprocal Rank, *removed*: fraction of the original results that is filtered out, and *error rate*: fraction of incorrectly filtered out profile elements. Best scores are in boldface.

matrix is calculated and then the values are sorted along one of the dimensions. However, when one is developing algorithms for filling that matrix, different aspects of the data may need to be taken into account: documents on a topic are examined for expert finding, and documents associated with a specific person are considered for expert profiling.

Finally, both Method 1 and Method 2 achieved fairly high MAP and MRR scores on the profiling task, with Method 1 significantly outperforming Method 2.

Filtering

Recall that filtering (as defined above) allows a knowledge area to appear in the individual's profile only if the person is among the top f ranked experts on that topic. Low f values imply that the profile will contain only the expertise areas of a person. As a consequence of filtering, an individual's profile may become empty, in which case we get a smaller, but far more precise result set. The questions we should ask are: how much do we gain for the experts that we retain, and how many valid profile elements do we lose (i.e., what is the error ratio)? See Table 2: for Method 1 and 2 we can obtain significant improvements (in terms of MRR scores) when filtering, where the original method (with no filtering) acts as a baseline.[2] Thus, filtering has an early precision enhancing effect. However, in most cases the MAP scores drop, which suggests that recall drops: we lose knowledge areas. This is confirmed by an inspection of the error ratios. The two profiling methods behave somewhat differently w.r.t. this measure,

[2]Here, we used the (non-paired) Mann-Whitney U test (two tailed) to establish significance.

	#rel	MAP	MRR	P@5	P@10	P@20
EF (baseline)	576	0.196	0.531	0.336	**0.332**	**0.269**
+ Method 1:						
(A)	576	**0.209***	**0.659***	**0.396***	0.326	0.267
(B) $\lambda = 0.5$	576	**0.197**	**0.584***	**0.376***	0.324	0.267
+ Method 2:						
(A)	576	0.181	**0.576***	**0.340**	0.292	0.242
(B) $\lambda = 0.7$	576	0.188	**0.559***	**0.344**	0.306	0.254

Table 3: Results of reranking expert finding results using individual's profile. The columns are: reranking method, number of relevant retrieved candidates, mean average precision, mean reciprocal rank, precision after 5, 10 and 20 candidates retrieved. Best results bold face; * denotes significant improvements over the baseline.

when filtering is applied. But in both cases filtering substantially reduces the size of the result set, and at the same time, they keep the error ratio of the filtering low. The drop in MAP scores indicates that we lose appropriate knowledge areas that were initially ranked highly, while the rise in MRR indicates that the remaining appropriate ones are ranked higher.

4 An Application to Expert Finding

In this section we describe an application of the extracted topical profiles to expert finding. Our approach takes the results of an existing expert finding method—treated as a black box—as input, and adjust the results using the individuals' profiles: if a knowledge area ranks low on a person's profile, we push the candidate down on the list of experts returned to the user. We expect this idea to have a precision enhancing effect, possibly hurting recall. Since we take the expert finding method used to be a black box, we do not have any information about the applied scoring method, which leaves us no other option than to use the ranking of the expert finding results: we do not make any assumptions about the scores. We combine the reciprocal of these ranks either in a multiplicative or in an additive way: i.e., $rank'_{EF}(ca, ka) =$

$$\begin{aligned}(A) &= \frac{1}{rank_{EF}(ca,ka)} \frac{1}{rank_{PR}(ca,ka)}\\ (B) &= \lambda + \frac{1}{rank_{EF}(ca,ka)} + (1-\lambda)\frac{1}{rank_{PR}(ca,ka)}\end{aligned}$$

Table 3 shows the results of our reranking methods; we only include scores with the best performing λ parameter for (B). Method 1 achieved the highest scores with equal weights on both the expert finding and profiling rankings, while Method 2 operated best when less changes were allowed on the original expert search results. All configurations improve MRR and P@5 over the baseline, and this fact confirms that profiling can be used for improving upon an existing expert finding method in terms of early precision. Moreover, this special application of the extracted profiles allows us to make further comparisons between the profiling methods: Method 1 outperforms Method 2, here and on the original profiling task.

The combination methods we presented here are fairly simple, but still proved to have a positive impact on application of the profiles. Further improvements could be pursued using more sophisticated methods for combining the retrieval results; consult [Kamps and de Rijke, 2004] for an overview.

5 Social Profiles

Collaborators play an important role when one is searching for expertise. People that an individual is working with are part of her personal workspace, and can serve as a background, or context, in which the system's recommendations should be interpreted. This collaboration network can also help us to explore the roles of the individuals within an organization. We might have a specific, well-defined need, where we are looking for an expert, or even for a "specialist." Another typical user scenario is to find a "librarian," who has access to a wide range of knowledge, and can direct us to a specialist, after the request has been narrowed down.

5.1 Collaboration network

We interpret the *social profile* as a collaboration network, and describe an algorithm for building such a network.

A *collaboration network* CN is a directed graph (V, E), where the nodes correspond to people, and a weighted directed edge $(x, y) \in E$ indicates the strength of the collaboration between x and y. Given a topic t, a *topical collaboration network* $CN(t)$ is a special CN where we restrict ourselves to collaborations (between people) relevant for t.

Given our enterprise (personal workspace) setting we can create a topical collaboration network $CN(q)$ given the user's query q. The level of the collaboration between two people x and y is estimated by the relevance of the documents D_{xy} that are associated with both x and y. The weight of the edge between the two individuals is expressed using

$$w(x, y) = \sum_{d \in D_{xy}} relevance(d, q) \quad (6)$$

where $D_{xy} = \{d \in D_q | A(x, d) \wedge A(y, d)\}$. A query-biased subset of documents D_q is obtained by using the top n documents retrieved for the query q, and $A(x, d)$ is a binary function denoting whether the person x is associated with the document d (see Section 3 for details).

The topical collaboration network $CN(q)$ is built using the following algorithm:

Init Obtain a query-biased subset of documents D_q

Step 1 Calculate the topical relevance of each individual

$$\forall x \in V : R(x, q) = \sum_{d \in D_q \wedge A(x,d)} relevance(d, q) \quad (7)$$

Step 2 Calculate the level of collaboration $w(x, y)$ between individuals using Formula 6.

Step 3 Make the edges directed and normalize them using the individuals' topical relevance. We allow nodes to be connected with themselves.

$$\begin{aligned}\forall y \in V_x : w'(x, y) &= \frac{w(x, y)}{R(x, q)} &(8)\\ w'(x, x) &= 1 - \sum_{y \in V_x} w'(x, y) &(9)\end{aligned}$$

where V_x denotes the set of nodes connected to x: $V_x = \{y \in V | w(x, y) > 0\}$.

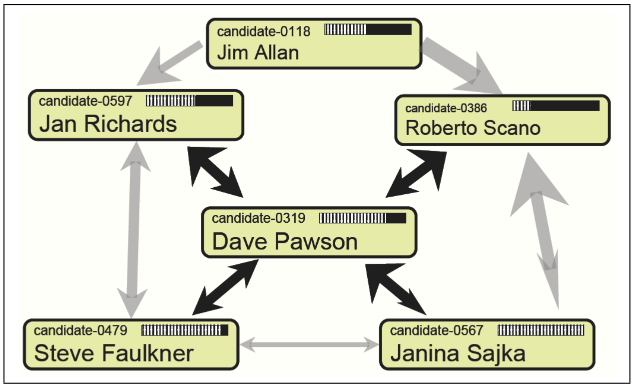

Figure 2: Screen dump of a topical collaboration network displayed for the query "authoring tools."

5.2 Visualization

Once a collaboration network has been built, we need to present it to end-users of the search system. We opted to visualize it centered around a selected candidate, where the thickness of the directed edges reflects the strength of the relations. Figure 2 presents a screen dump from our system, where we search for experts on "authoring tools." The thickness of an edge corresponds to its weight. The full interface also displays individual's profile on mouse-over.

5.3 Evaluation

Topical and social profiling go together and form a complete picture. However, assessment of social profiles is an issue that we are not addressing in this paper. An indirect way of evaluation is to conduct user studies to measure how "useful" a collaboration network is, when one is searching for expertise. A direct evaluation may be performed by using the social collaboration network itself for inferring experts, by using methods from graph-theory.

6 Conclusions

We presented methods for providing the context and the evidence that supports the recommendations of an operational expertise search system. We introduced the profile of an individual, which is a record of the types and areas of skills of that individual ("topical profile") plus a description of her collaboration network ("social profile"). We proposed multiple models for addressing the topical profiling task, and showed that they significantly outperform an "inverted" expert search baseline. Refinements brought in filtering, allowing an area into a person's profile only if she is among the top ranking experts in the area. Experimental evaluation showed that more than 70% of the retrieved results could be filtered out, while keeping the error rate as low as 6.5%. We applied our profiling algorithms to significantly enhance the performance of a state-of-the-art expert finding algorithm. Our reranking method is very effective in terms of early precision. Finally, we addressed the task of determining a social profile for a given person. We introduced a graph representation of the collaboration network, where nodes represent people, and (weighted, directed) edges reflect the level of collaboration. We described an algorithm for estimating the weights of edges, based on "co-authored" documents.

Possible improvements include more sophisticated algorithms for extracting topical profiles, e.g., taking similarities and overlap between knowledge areas into account. We are also developing methods to assess the quality and "usefulness" of social profiles.

Acknowledgments

Krisztian Balog was supported by the Netherlands Organisation for Scientific Research (NWO) under project numbers 220-80-001, 600.065.120 and 612.000.106. Maarten de Rijke was supported by NWO under project numbers 017.001.190, 220-80-001, 264-70-050, 354-20-005, 600.065.120, 612-13-001, 612.000.106, 612.066.302, 612.069.006, 640.001.501, 640.002.501, and by the E.U. IST programme of the 6th FP for RTD under project MultiMATCH contract IST-033104.

References

[Abrol *et al.*, 2002] M. Abrol, U. Mahadevan, K. McCracken, R. Mukherjee, and P. Raghavan. Social networks for enterprise webs. In *Proceedings WWW 2002*, 2002.

[Balog *et al.*, 2006] K. Balog, L. Azzopardi, and M. de Rijke. Formal models for expert finding in enterprise corpora. In *Proceedings SIGIR 2006*, 2006.

[Baeza-Yates and Ribeiro-Neto, 1999] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. ACM Press, 1999.

[Becerra-Fernandez, 2000] I. Becerra-Fernandez. The role of artificial intelligence technologies in the implementation of People-Finder knowledge management systems. *Knowledge-Based Systems*, 13(5):315–320, 2000.

[Craswell *et al.*, 2001] N. Craswell, D. Hawking, A. M. Vercoustre, and P. Wilkins. P@noptic expert: Searching for experts not just for documents. In *Ausweb*, 2001.

[Cross *et al.*, 2001] R. Cross, A. Parker, L. Prusak, and S.P. Borgatti. Knowing what we know: Supporting knowledge creation and sharing in social networks. *Organizational Dynamics*, 30(2):100–120, 2001.

[Davenport and Prusak, 1998] T. H. Davenport and L. Prusak. *Working Knowledge: How Organizations Manage What They Know*. Harvard Business School Press, Boston, MA, 1998.

[ECSCW, 1999] ECSCW'99 Workshop. *"Beyond Knowledge Management: Managing Expertise"*. 6th European Conference on Computer Supported Cooperative Work, 1999.

[Hawking, 2004] D. Hawking. Challenges in enterprise search. In *Proceedings Fifteenth Australasian Database Conference*, 2004.

[Kamps and de Rijke, 2004] J. Kamps and M. de Rijke. The effectiveness of combining information retrieval strategies for European languages. In *Proceedings 19th Annual ACM Symposium on Applied Computing*, pages 1073–1077, 2004.

[Maron *et al.*, 1986] M. E. Maron, S. Curry, and P. Thompson. An inductive search system: Theory, design and implementation. *IEEE Trans. Systems, Man and Cybernetics*, 16(1):21–28, 1986.

[TREC, 2005] Enterprise track, 2005. URL: `http://www.ins.cwi.nl/projects/trec-ent/wiki/`.

[Voorhees and Harman, 2005] E.M. Voorhees and D.K. Harman, editors. *TREC: Experiment and Evaluation in Information Retrieval*. MIT Press, 2005.

[W3C, 2005] The W3C test collection, 2005. URL: `http://research.microsoft.com/users/nickcr/w3c-summary.html`.