Correlating Topic Rankings and Person Rankings to Find Experts

Thijs Westerveld
CWI, INS1
PO Box 94079
NL-1090 GB Amsterdam
The Netherlands

ABSTRACT

Expert search is about finding people rather than documents. The goal is to retrieve a ranked list of candidates with expertise on a given topic. The task is studied in the context of the enterprise track.

We describe an approach that compares topic profiles and candidate profiles directly. These profiles are not based on unordered sets of documents, but on ranked lists. This allows us to differentiate between documents that are highly related to a topic or a candidate and documents that are only marginally related. The ranked lists for topics and candidates are obtained by simple retrieval queries. The correlation between the ranked list of documents for a topic and the ranked list for a candidate is used as an indicator of the candidate's expertise on the topic. We study different ways to rank documents for the candidate profiles as well as various ways of comparing the document and candidate based ranked lists. Experiments show that starting from the right candidate profiles, reasonable results can be obtained. Furthermore, it seems important to take a correlation measure that takes into account the orderings of documents in both the candidate profile and the documents profile.

1. INTRODUCTION

TREC's enterprise track focuses on information retrieval problems in an intranet setting. The track tackles the problems of dealing with a heterogeneous collection containing web pages, email archives, meeting reports, technical documents, memos and more. The tasks studied are the following

Discussion search Retrieve arguments, pro's and con's around a given topic, e.g., *blocking pop-ups* or *Is XHTML more accessible than HTML*

Expert search Retrieve experts on a given topic, e.g., *RDF Data Access*, *SVG*, or *privacy on the web*.

We only participated in the expert search task, a task that is common in enterprises and other organisations. The goal of this task is to retrieve a ranked list of at most 100 experts for each topic, and to provide up to 20 documents with each expert that support their expertise. The experts come from a predefined set of candidates, each candidate has an identifier plus a name and email address. Topics for the 2006 expert search task are created by the participating groups and have the typical TREC format with title, description and narrative.

The approach we take is based on comparing topic and candidate profiles. We produce a ranked list for each topic and one for each candidate expert and correlate them. The lists with highest correlations indicate experts on topics.

The remainder of this paper is organised as follows. Section 2 discusses other approaches to the same task. Section 3 explains how correlations between document rankings can help to identify experts. Experimental results are discussed in Section 4. The paper ends with a summary of the main conclusions and a discussion of possible points for improvement and future research in Section 5.

2. RELATED WORK

Two typical approaches to expert ranking exist, one based on document ranking, the other on candidate profiles. The document ranking approach [8, 1] starts with a document ranking for the topic on which we want to find experts. The top ranked documents are then analysed to locate occurrences of candidates, for example by looking for their names and email addresses. Candidates are subsequently ordered by decreasing number of occurrences. Figure 1-a illustrates this scheme.

The candidate profile approach [7, 6] starts by creating a profile for each candidate, for example by gathering all documents authored by the candidate, or all documents mentioning the candidate. These profiles are then treated as (large) documents and ranked according to the topic of interest. The highest ranked profiles are assumed to correspond to the experts on the topic. Figure 1-b illustrates this strategy.

The approach discussed here can be viewed as a mix of the document ranking and candidate profiling approaches. While we start from a document ranking, we do not directly identify candidates in the retrieved documents. Instead, we investigate whether the retrieved documents are included in the candidate's profile. In our approach, candidate profiles are rankings rather than sets. That means documents can be included in a candidate's profile to different degrees. In a sense the relation between documents and candidates is weighted. Others have studied this weighting on the side of topical ranking: candidates matching documents retrieved at a higher rank are more likely to be experts than candidates matching documents further down the topical ranking, see for example [5]. The use of a similar weighting on the candidate side of the coin is something we have not seen
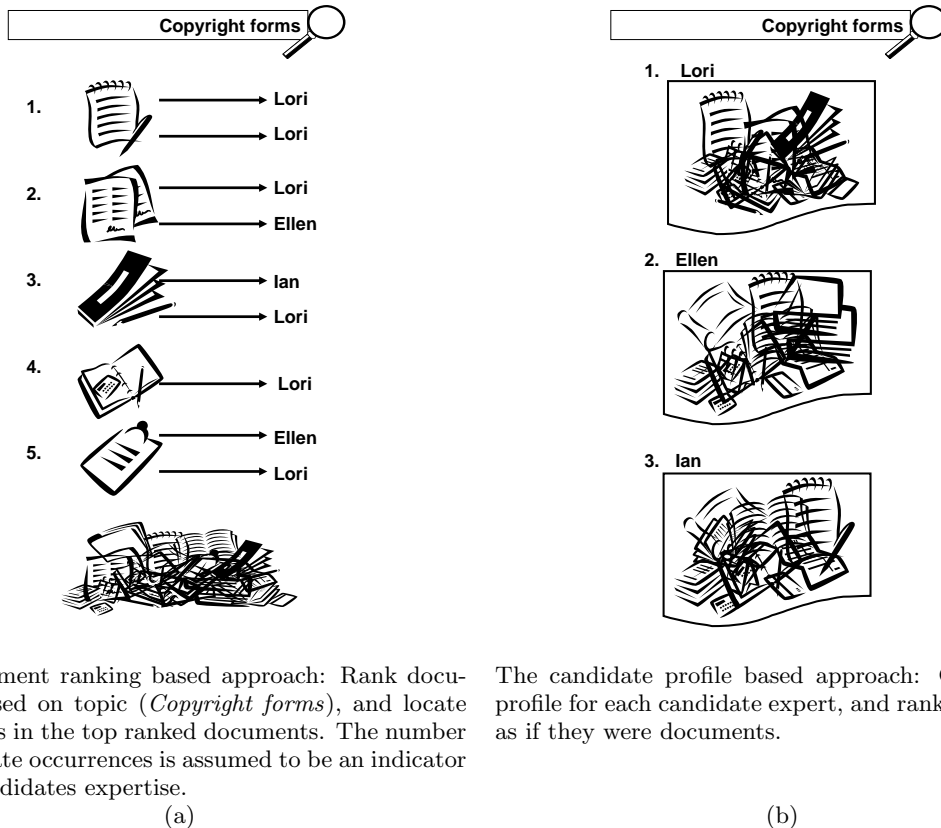
The document ranking based approach: Rank documents based on topic (*Copyright forms*), and locate candidates in the top ranked documents. The number of candidate occurrences is assumed to be an indicator of the candidates expertise.

(a)

The candidate profile based approach: Construct a profile for each candidate expert, and rank the profiles as if they were documents.

(b)

Figure 1: Two typical approaches to expert search.

before.

3. CORRELATING RANKINGS

Our approach to expert search is based on correlating candidate profiles and topic profiles. Each of these profiles is a simple document ranking. The correlation between the rankings is an indication of the candidate's topical expertise. The intuition behind this approach is the following. If person X is an expert on topic Y, the documents retrieved when searching for person X can be expected to overlap with the documents retrieved when searching for topic Y. Of course, not all documents on topic Y will mention person X, and not all documents that mention the person will be on topic, but a reasonable overlap can be expected, at least an overlap bigger than for non experts on Y. We use this intuition and compare the topic based ranking and the candidate based ranking directly.

In our setting, a topical profile is the document ranking obtained from a query on a language modelling based IR system using only the topic's *title* field. For candidate profiles, we experimented with different variants as explained in the following sub-section.

3.1 Candidate Profiles

The candidate profiles we use, are ranked lists of documents. We investigate various ways of constructing these. Our base-

line is to rank the documents using simple queries based on the names and email addresses of the candidates. No special processing of names or email addresses in collection or queries is used. This means these are tokenised as any other string. This makes this approach a relatively poor baseline since persons sharing a first or last name, or a domain name in their email address are easily confused.

In a second document ranking for the candidates, we tackle these problems by using the W3C corpus with candidate annotation as created by Jianhan Zhu at the Open University. This corpus is a copy of the original W3C collection with the occurrences of the candidates annotated with special tags. A small sample is shown in Figure 2. We transform this collection to a collection of only candidates. In this candidate collection, all text from the documents is removed and the candidate *tags* are replaced by candidate *terms* (see Figure 3). Candidate queries against this candidate collection can easily retrieve the most relevant documents for a given candidate based on term frequency and inverse document frequency of the candidate term.

The two ways of constructing candidate profiles are topic independent. Since a person may be an expert on more than a single topic, this may be problematic. Therefore, we try to focus the candidate profiles by making them topic specific. This third candidate ranking is based on queries consisting of the original title queries as provided by the

```
<p>The following people have contributed either through feedback or
beta-testing  (hope everyone is cited):</p>
<ul>
  <li> <candidate-0161>tim berners-lee</candidate-0161> </li>
  <li> <candidate-0056>bert bos</candidate-0056> </li>
  <li>Al Gilman</li>
  <li>Daniel La Liberte</li>
  <li><candidate-0023>gerald oskoboiny</candidate-0023></li>
  <li><candidate-0010>ralph swick</candidate-0010><candidate-0010>swick</candidate-0010></li>
  <li>Ir<E8>ne Vatton</li>
  <li>.. and everyone on the hypermail public mailing list.</li>
</ul>
```

Figure 2: A sample from the tagged W3C collection.

```
<DOC docno="lists-000-0000000">
candidate0002
candidate0002
candidate0161
candidate0056
candidate0023
candidate0010
candidate0010
candidate0018
</DOC>
```

Figure 3: A sample document from the W3C based candidate collection.

track organisers, augmented with a single special candidate term. For example, for topic EX67 and candidate-0001, the query would be *privacy on the web candidate0001*. This query is issued against a modified version of the tagged W3C corpus, where candidate tags are replaced with candidate terms and the rest of the text is left untouched.

3.2   Comparing Rankings

To compare topic profiles and candidate profiles, or rather the two corresponding document rankings, we use two basic approaches, Spearman's rank correlation, and a blind feedback like approach that assumes the top N for one of the rankings (topic based or candidate based) are relevant and measures average precision on the other ranking. The techniques are detailed below.

3.2.1   Spearman Rank Correlation

Spearman's rank correlation, or Spearman's ρ is a non-parametric test that measures the strength of the correlation between two variables:

$$\rho = 1 - 6 \sum \frac{D^2}{N(N^2-1)}, \qquad (1)$$

where D is the difference in rank of corresponding variables, and N is the number of paired values. Our variables are the topic based score and candidate based score for a given document. To limit computational costs, we compute Spearman's ρ based only on the top K returned documents for topic queries. Documents that do not occur in the candidate rankings are assumed to be at rank $K+1$. Note that this is a best case scenario assumption, thus we over-estimate ρ.

Spearman's ρ measures the correlation for the complete ranked list. In our setting, we would like to emphasise correlations at the top of the ranked list and down-weight the correlations further down. To this end, we use a log based variant of Spearman's *rho*. We first perform a log transform on the *ranks* before computing the correlation. Note that while a log transformation of the scores would not influence the ranking, such a transformation on the ranks does influence the correlation. We refer to the traditional Spearman correlation as *SP* and to the log-based variant as *log-SP*

3.2.2   Blind Feedback AP

Another approach to compare the rankings is similar to blind feedback. Here, we assume the top K documents for the topic based ranking are relevant, and we compute the average precision (AP) of the candidate ranking based on that. This will reward candidate queries that have retrieved many of the documents in the topic ranking, especially if these are retrieved at the top ranks.

We also tested the reverse approach: assuming the top K documents of the candidate based rankings are relevant and computing AP on the topic based ranking. In this setting, we directly compare absolute performance of a single topic ranking on different tasks. The tasks are to retrieve documents relating to candidate-0001, relating to candidate-0002, etc. In a normal retrieval setting, comparing absolute numbers across tasks does not make sense. Here we keep the system fixed (the topic run), and effectively try to find the easy topics for this system (what are the easy candidates for this topic?). This is less elegant than the reverse approach, but it may still help to locate the experts.

In the following we will refer to these AP based variants as *AP-Trel* and *AP-Crel*, for the runs with Topic rankings assumed relevant and Candidate rankings assumed relevant respectively.

4.   EXPERIMENTAL RESULTS

For our experiments, we use the open source pf/tijah system [4], a flexible blend of an XQuery engine and an XML IR retrieval systems that supports many information retrieval models. To be able to use this system with the W3C data, we cleaned the collection to make sure all documents are valid XML. While the pf/tijah system supports retrieval at the element level (e.g., retrieve paragraphs or sections), we only use full document queries. Documents and queries are

stemmed and stopped and language models are used to perform the ranking of documents. For the topic specific ranking pf/tijah was not able to perform the 60,000+ queries (55 topics times 1092 candidates) in the available time. For this variant, we had to use the X100 based retrieval system as used by our group in the terabyte track[3].

While the official track results include measures based on retrieved experts and measures based on experts retrieved *with* supporting documents, we ignore the supporting documents and only report on the ability of the various approaches to identify the experts. The metric used is mean average precision (MAP) over the ranked list of experts. We use $K = 1500$, that means the topical profiles consist of the top 1500 documents retrieved for the topic. These 1500 documents are the basis for computing the Spearman correlation measures and they are regarded relevant in the AP-Trel and AP-Crel approaches. The influence of K on the results is studied below. Table 1 shows the results.[1].

One notable result is the low score across the board for the topic specific candidate rankings. Intuitively, these rankings are expected to be better, more focused, and more likely to yield appropriate correlations. The low scores in practise can be explained by the fact that the specific candidate terms that were added to the topic query often have little influence on the ranking. One reason for this is that it is just a single extra term on a title query of on average 2.9 terms. More important though, is that some candidates appear only in very few documents, these few may end up on the top of the ranked list for title+query, but the rest of the list remains unchanged. The effect is that the title+query rankings are often very similar to the title only rankings in particular in the cases where the candidate term matches few or no documents. This means the correlation between the two rankings is higher for candidates that hardly appear in the collection. Clearly, these are not necessarily the experts.

Also, between the other two candidate rankings, there is a remarkable difference. The annotated candidate run clearly outperforms the name and email variant. This was expected as the annotated variant does not suffer of confusion between people with similar names or email addresses.

Finally, the *log-SP* runs give lower scores than the other correlation variants. A reason could be that the top ranks are emphasised too much now. Further analysis is needed to confirm this.

In comparison to the other expert search submissions, these numbers are relatively low. One reason for this is that we did not use any collection specific knowledge. Many runs that performed well treated emails different than other documents and included special treatment or weighting for specific fields like an email's sender, receiver or subject. Also, many successful approaches ranked and analysed information at the sub-document level, creating candidate profiles based on smaller windows around candidate occurrences rather than on complete documents.

---

[1]The official (lower) *SP* and *log-SP* results are based on Spearman ρ's for varying set sizes N, and hence not directly comparable; here we report the numbers for the runs after fixing this mistake
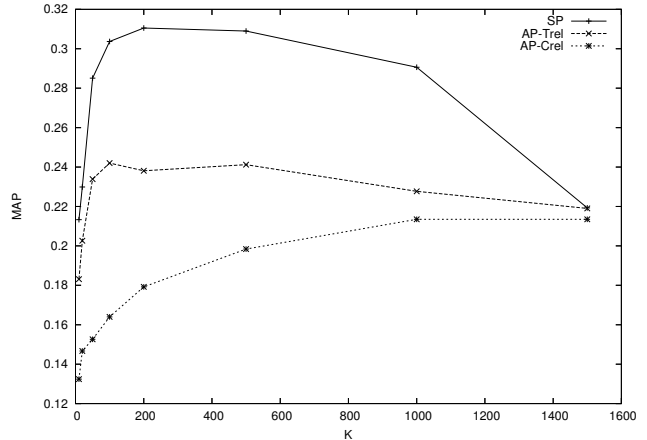

Figure 4: The influence of K on mean average precision (MAP) for different correlations (candidate annotation profile)

4.1 Varying the size of the profiles

The number of documents in the topic and candidate profiles is likely to have an impact on the results. The top ranked documents are probably better indicators of expertise than the ones further down the ranked list. Our comparison of rankings takes this into account to some extend, and we tried to emphasise this by using a log-transform on the ranks. Still, it may be useful to limit the size of the profiles and look only at the top documents. Figure 4 shows the influence of K on mean average precision values for the SP, AP-Trel and AP-Crel correlations.

First of all we notice that while the performance for the different correlation measures on the top 1500 (as reported in Table 1) are the same, large differences exist for smaller values of K. The AP-Crel line shows that taking fewer than 1000 documents as a candidate's profile harms the MAP (Note that for the two other correlation measures K relates to the size of the topic profile; the size of the candidate profile for these is fixed at 1500). The size of the topic profile can be significantly smaller; the optimum seems to be around 100 or 200 documents. It seems the experts need to have only the top ranked topical documents in their profile, but it could also be the case that for many topics fewer than 1500 documents are relevant and that the profile for these includes random documents to reach a profile size of 1500 documents. Furhter investigation is needed to check this.

The emphasis on the top ranked documents in the profiles that we wanted to attain with the log-SP run, seems indeed important, but it is not as symmetric as the (failing) log-SP run treated it: for the candidate profiles we need many documents, for the topical profiles we need to focus on the top ranks. Finally, Spearman (SP) based comparison of the profiles gives much higher scores than the average precision (AP) based runs. This indicates that it is important to not only look at the set of documents related to a candidate or a topic, but that the ordering of these documents plays an important role. A candidate is an expert on a topic, if the top ranked documents in the topical profile have a similar ordering in the candidates profile.

Table 1: Expert search results (MAP) for various base candidate rankings (rows) and various correlation measures (columns).

Candidate profile	Correlation Type			
	SP	log-SP	AP-Trel	AP-Crel
Name + email	0.1154	0.0841	0.1336	0.1375
Candidate annotation	0.2193	0.1577	0.2190	0.2135
Topic Specific	0.0017	0.0017	0.0016	0.0016

5. CONCLUSIONS

Directly comparing topic profiles and candidate profiles based on the correlation between document rankings is a novel and interesting approach. It is different from traditional approaches that either construct a descriptive document for each candidate and rank these, or rank documents based on topics only and then extract candidates from the top ranked documents. The correlation based approach directly works on the ranked lists and uses information of the entire lists.

The experiments for the expert search task showed that the approach is feasible given that the initial rankings are good enough. The baseline, name and email address based candidate ranking is clearly not good enough, but a candidate ranking based on the tagged candidates gives good results when used in our correlation method[2]. The exact correlation method is of minor importance in such a setting, but *log-SP* is a particular bad choice.

Topic specific candidate rankings harm results significantly. The main reason for this is the relatively low number of appearances in the collection for many candidates. More research is needed to investigate ways to compensate for this. A possible solution would be to compare the title + candidate based ranking not only to the topic based ranking, but also to the candidate based ranking, or to measure the overlap between all sub-queries cf. Carmel et al [2].

In comparison with other expert search approaches the performance of the approach described here is relatively poor. One reason for this is we do not use any collection specific information, another is we treat documents as a single entity. The same techniques could easily be adapted to work on windows, XML elements or passages. Also in creating the rankings for the topic and candidate profiles, collection specific creation could be incorporated. Correlating rankings on top of the successful expert search approaches is an interesting direction for future research.

As a collection independent approach, the correlation based profile comparison gives good results, provided the parameters are chosen correctly. Spearman based correlation is than better than the average precision based measures, indicating that it is important to take the ordering of documents in both the topical profile and the candidate profile into account.

6. REFERENCES

[1] S. Bao, H. Duan, Q. Zhou, M. Xiong, Y. Cao, , and Y. Yu. Research on expert search at enterprise track of trec 2006. In *The 15th Text Retrieval Conference (TREC-2006) Notebook*, 2006.

[2] D. Carmel, E. Yom-Tov, A. Darlow, and D. Pelleg. What makes a query difficult? In *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 390–397, 2006.

[3] S. Héman, M. Zukowski, A. P. de Vries, and P. Boncz. Monetdb/x100 at the 2006 trec terabyte track. In *The 15th Text Retrieval Conference (TREC-2006) Notebook*, 2006.

[4] D. Hiemstra, H. Rode, R. van Os, and J. Flokstra. Pftijah: text search in an xml database system. In *Proceedings of the 2nd International Workshop on Open Source Information Retrieval (OSIR)*, pages 12–17, 2006.

[5] C. Macdonald and I. Ounis. Voting for candidates: adapting data fusion techniques for an expert search task. In *CIKM '06: Proceedings of the 15th ACM international conference on Information and knowledge management*, pages 387–396, New York, NY, USA, 2006. ACM Press.

[6] Z. Ru, Q. Li, W. Xu, and J. Guo. Bupt at trec 2006: Enterprise track. In *The 15th Text Retrieval Conference (TREC-2006) Notebook*, 2006.

[7] G. You, Y. Lu, G. Li, and Y. Yin. Ricoh research at trec 2006: Enterprise track. In *The 15th Text Retrieval Conference (TREC-2006) Notebook*, 2006.

[8] J. Zhu. Open university at trec 2006 enterprise track expert search task. In *The 15th Text Retrieval Conference (TREC-2006) Notebook*, 2006.

---

[2]At this moment it is unclear how this compares to other approaches on the TREC 2006 enterprise track.