Usefulness of Click-through data in Expert Search

Craig Macdonald
Department of Computing Science
University of Glasgow
Glasgow, G12 8QQ, UK
craigm@dcs.gla.ac.uk

Ryen W. White
Microsoft Research
One Microsoft Way
Redmond, WA 98052
ryenw@microsoft.com

ABSTRACT

The task in expert finding is to identify members of an organisation with relevant expertise on a given topic. Typically, an expert search engine uses evidence from the authors of on-topic documents found in the organisation's intranet by search engines. The search result click-through behaviour of many intranet search engine users provides an additional source of evidence to identify topically-relevant documents, and via document authorship, experts. In this poster, we assess the usefulness of click-through log data for expert finding. We find that ranking authors based solely on the clicks their documents receive is reasonably effective at correctly identifying relevant experts. Moreover, we show that this evidence can successfully be integrated with an existing expert search engine to increase its retrieval effectiveness.

Categories & Subject Descriptors: H.3.3 [Information Storage & Retrieval]: Information Search & Retrieval

General Terms: Performance, Experimentation

Keywords: Intranet, Expert finding, click-through log data

1. INTRODUCTION

Expert finding is an important task within enterprise organisations. Organisational members may have expertise needs, where they require the knowledge of another member to assist them with a project, or to understand a tool or information item. Research has shown that people often manually perform expert search by examining intranet documents retrieved in response to a query, and then contacting the authors of these documents [3]. Expert search engines have been developed to aid people in this process. These systems typically use the authors or people mentioned within documents ranked highly by the intranet search engine in response to the query as candidate experts [2, 4].

A significant advance in search technology in recent years has been the use of click-through log data generated by the millions of users of popular Web search engines to learn which documents should be retrieved for a given query. The implicit endorsement that documents receive through user visitation has been shown to be useful to rank Web search results [1]. The novel research reported here combines expert search and click-through log data mining to examine how documents visited frequently by many users of an intranet search engine can be used as expertise evidence to rank people. To our knowledge, this is the first work to investigate the use of click-through log data for expert search.

We used the click-through logs of an large organisation's intranet search engine as a source of expertise evidence. The logs comprised query-document click pairs gathered over a three-month period. To determine the effectiveness of using behavioural evidence for expert search, we construct a test collection. A standard collection such as that used in [2, 4] is not suitable since we require both queries and clicks.

2. BUILDING THE TEST COLLECTION

The two steps in building the test collection were the selection of queries and judging the relevance of documents pertaining to those queries.

2.1 Query Selection

The organisation where this research was conducted has over 150,000 employees, spread across sites throughout the world. Its intranet search engine has a high daily query volume, with query distributions following the expected power-law distribution. Many queries are infrequent with only a few occurrences and even fewer result clicks. Since the value of log data lies in aggregation across many events, we focused on common information needs with queries for which there were at least 100 clicks on retrieved documents in the three-month duration of our logs. This resulted in a set of around 3,000 popular queries, from which we manually selected 20 for which we could perceive informational or expertise needs (*e.g.*, interviewer training, netmeeting, virtual machines). Queries such as [company store] were not selected, as these were known-item queries, with most users clicking on the intranet homepage of the internal company store.

2.2 Relevance Judging

For each of the 20 selected queries, we pooled likely relevant candidates from two sources: click-through logs and existing expert search engine results. From the log data, we identified the documents visited by intranet search engine users. Then, for as many documents as possible, we used an internal web service API to identify the authors of each document. All authors identified *that were still members of the organisation* were pooled as potentially relevant candidates for that query. We also identified the top 10 candidates from the organisation's existing expert search engine that uses many sources of expertise evidence (*e.g.*, shared documents, group memberships, document co-authorship) *but not behavioural evidence*, and added these to the pool[1].

We emailed each candidate expert in the pool, asking if they believed they had higher than what they deemed would

---

[1]Note that this expert search engine is currently in development, and is based on statistically different document ranking techniques to that from which we obtained query logs.

Statistic	Value
Number of queries	20
Number of candidates	340
(expert search engine)	291
(click-through documents)	61
Number of responses	199
(YES)	147
(NO)	52

Table 1: Statistics of the created test collection. Note that some candidate were pooled for more than one query, and by more than one source.

be average employee expertise in the described query. We requested binary responses (yes/no) to simplify judging. Of the 340 candidate experts who were contacted, 199 (58.5%) responded with a judgement. Table 1 presents the statistics of the generated test collection. In addition to their binary assessment, several responses noted that while they had topic knowledge, they had changed their organisational role, and would rather not be contacted on this topic.

3. FINDING EXPERTS USING CLICKS

The extent to which a document is query relevant can determine its usefulness in estimating the expertise of those associated with it. Indeed, many models of expert search - such as Balog's *Model 2* [2] or the *Voting Model* [4] - consider the retrieval score of the document for the query. Instead, in this section, we describe an alternative approach. We study whether experts can be accurately predicted using the click-through logs of an intranet search engine. For a given query, we rank experts by two simple approaches:

1. ClickVotes: Rank by number of documents clicked on for each candidate expert.

2. ClickCount: Rank by number of total clicks for each expert. This can be calculated as $p(C|Q) = \sum_D p(C|D,Q)$, where $p(C|D,Q) = \frac{associated(C,D) \cdot clicks(D,Q)}{totalClicks(Q)}$ is the maximum likelihood of a user clicking on a document D retrieved for query Q that is associated to candidate C.

It is of note that the approaches proposed here are based on voting techniques from the Voting Model, namely Votes and CombSUM [4]. However, in this application, they are based on clicks, and not document scores or ranks.

4. EXPERIMENTS & RESULTS

In the following, we experiment to determine how useful the click-through log data is for expert search. In particular, we experiment by comparing the *ClickVotes* and *ClickCount* approaches to the existing expert search engine (denoted *Existing*) described above, and to determine whether these approaches can be combined with the existing expert search engine. We evaluate using mean reciprocal rank (MRR), precision across the top-three retrieved documents (P@3), and mean average precision (MAP). For all techniques, results are reported in Table 2. We see that *ClickCount* outperforms *ClickVotes*. This is likely because clicks are more voluminous than clicked documents, facilitating more reliable expertise estimates and finer-grained expert rankings. We also see that using click-through data alone (*ClickVotes* or *ClickCount*) as documentary evidence is less effective than *Existing* by a statistically significant margin (Wilcoxon Signed Rank test). *Existing* uses multiple sources of evidence that may be of higher quality in combination than the click evidence. In addition to using the sources independ-

Technique	MRR	P@3	MAP
Existing	0.7976	**0.4833**	0.4479
ClickVotes	0.4417	0.2167**	0.1324**
ClickCount	0.4917*	0.2833*	0.1624**
Existing + ClickCount ($w_1 = 1, w_2 = 0.1$)	**0.8018**	**0.4833**	**0.5014***

Table 2: Results for ranking experts. Significant differences (Wilcoxon Signed Rank test) from *Existing* are shown (* ($p \leq 0.05$) and ** ($p \leq 0.01$)).

dently, we can merge *ClickCount* with *Existing* through a simple linear combination of the ranks of the candidate expert from the search engine and their scores from the click-through approach. Two parameters w_1 and w_2 control the respective influence of *Existing* and *ClickCount* in the final ranking. The values used in our combination were $w_1 = 1$ and $w_2 = 0.1$. Due to the lack of training data, these were chosen to empirically maximise retrieval performance and therefore demonstrate the potential utility of clicks for this task. As we see in Table 2, adding clicks to *Existing* enhances its retrieval performance, significantly so for MAP. This shows that the click-through log data can indeed enhance a content-based expert search engine.

5. CONCLUSIONS

We studied the usefulness of click-through data for expert search in a large organisation. We first developed a test collection comprising: (i) a subset of popular queries with informational or expertise needs issued to the organisation's intranet search engine, and (ii) relevance judgements for each of these queries from a significant proportion of identified candidate experts. We mined click-through log data for the engine and used the clicked documents for each query to identify candidate experts. The test collection is used to measure the effectiveness of an existing expert search engine, an engine based on click-through data alone, and an engine using a combination of the expert search engine and click-through data. The reasonable performance of the expert search engine based on the click-through data is promising, as this shows that for large intranets, the clicked documents can be used to identify the relevant experts for the same or similar queries. As far as we know, this is the first work in which document clicks are used for ranking experts. In future work, we intend to develop a larger expert search test collection with click-through data, and experiment with more advanced ways in which document-click features can be directly integrated into an expert search model.

6. REFERENCES

[1] E. Agichtein, E. Brill, and S. Dumais. Improving web search ranking by incorporating user behavior information. In *Proc. SIGIR*, pp 19–26, 2006.

[2] K. Balog, L. Azzopardi, and M. de Rijke. Formal models for expert finding in enterprise corpora. In *Proc. SIGIR*, pp 43–50, 2006.

[3] M. Hertzum and A. M. Pejtersen. The information-seeking practices of engineers: searching for documents as well as for people. *Inf. Proc. Manage.*, 36(5):761–778, 2000.

[4] C. Macdonald and I. Ounis. Voting for candidates: adapting data fusion techniques for an expert search task. In *Proc. CIKM*, pp 387–396, 2006.