

Experiments in Mobile Content Enrichment

Karen Church and Barry Smyth

Adaptive Information Cluster,
Department of Computer Science,
University College Dublin, Belfield, Dublin 4, Ireland
{Karen.Church, Barry.Smyth}@ucd.ie

Abstract. Mobile content, by its concise nature, offers limited indexing opportunities, which makes it difficult to build high-quality mobile search engines and indexes. In this paper we address this problem by evaluating a heuristic content enrichment framework that uses standard Web resources as a source of additional indexing knowledge. We present an evaluation using a mobile news service that demonstrates significant improvements in search performance compared to a benchmark mobile search engine.

1 Introduction

Recent positive developments in handsets, infrastructure, content quality, and charging models are fueling renewed optimism about the potential of the Mobile Internet. For example, a recent report by Informa put the number of mobile subscribers at 2.7 billion at the end of 2006 ¹. Ipsos Insight published a study in 2006 which shows that 28% of mobile subscribers worldwide have browsed the Internet using their mobile phone ². This pattern of growth was driven primarily by older users (age 35+) indicating that the traditional early adopter segment, i.e. young males, no longer dominates wireless Internet access.

Until recently, the Mobile Internet, for the majority of subscribers, has meant the “walled-garden” of content that has been available through their operator’s portal. However, as off-portal content has grown so too has the interest of users in accessing this content, which has led to a sharp increase of interest in the potential for mobile search engines to provide users with fast and efficient access to this content. For example, major players within the search engine industry like Google and Yahoo continue to release exciting new improvements to their mobile search services. One recent study reports that 31% of users started using mobile search in 2006 with another 48% expecting to start in the next few months ³.

¹ IT Week: *Mobile industry 'bullish' for the new year*,
<http://www.itweek.co.uk/vnunet/analysis/2171436/mobile-industry-bullish-2007>

² Ipsos Insight: *Mobile phones could soon rival the PC as worlds dominant Internet platform*, <http://www.ipsos-na.com/news/pressrelease.cfm?id=3049>

³ Mobile Marketing Association: *Mobile Marketing Association announces Mobile Search study key findings*,
<http://www.primezone.com/newsroom/news.html?d=103374>

There are many significant issues to be addressed if current search engine technologies are to deliver the type of search experience that is necessary to engage mobile users. For a start there are the many well-documented challenges of delivering information to small screen devices [5]. In addition, restricted text input capabilities inevitably have a significant impact on the use of these handsets as search devices, limiting the type of queries that will likely be provided. These device limitations are but one aspect of the larger problem and the nature of mobile content itself is such that it introduces additional challenges from an indexing and search standpoint. Mobile pages are typically much shorter in length than their Web counterparts; most mobile gateways and handsets are limited by the size of the pages they can process, and traditionally WML (Wireless Markup Language, the mobile equivalent of HTML) decks are restricted to just a few KB. Mobile Internet access is mainly about *content snacking* but smaller pages mean that there is less content for a search engine to index and thus less information available to inform retrieval. This exacerbates the so-called *vocabulary gap* [4] that plagues Web search — referring to the tendency for searchers to often choose query terms that do not correspond to those used to index their target document — because there are even fewer terms available for indexing thus reducing the likelihood of a query match during future searches.

We consider this issue of limited page content in this paper and describe a heuristic *context enrichment* strategy to extend the indexing knowledge that is used to characterise mobile content by discovering additional indexing terms. To do this for some item of content we automatically transform the content item in to a set of *enrichment queries* which are used to retrieve a set of *enrichment results* from a standard Web search engine. These related result pages then act as a source of *enrichment terms* which are extracted, ranked and added to the mobile index. We have previously described and evaluated a basic content enrichment (CE) framework, focusing on the selection of enrichment terms, and demonstrating a significant improvement in retrieval accuracy [2]. However, this work highlighted a number of opportunities for improving the way in which enrichment queries are generated, and the way in which enrichment results are selected prior to enrichment term extraction and in this paper we propose improvements in both of these areas and demonstrate additional performance gains.

2 Related Work

The research area we have identified as most relevant to our current work is the *query expansion* or *relevant feedback* domain. A well known issue in Web search and IR is that short queries and term mismatching can cause relevant documents to be omitted from search results because they do not contain the terms within a user's query. The idea behind query expansion is that if we *enrich* the query using terms from a set of relevant documents, overall retrieval performance increases. One of the most popular query expansion techniques is known as *relevance feedback* [6, 7]. The standard approach involves a user submitting an initial query to the system, receiving a set of results and indicating which results

are relevant. Terms are then extracted from these *relevant* documents and are used to supplement the users initial query. This *iterative* process continues until the users information need is satisfied.

Although existing research shows that relevance feedback can achieve significant improvements in retrieval performance, it requires that users provide accurate relevance judgments. However, users are often reluctant to provide such information. To overcome this difficulty, the concept of *pseudo-relevance* or *blind feedback* was introduced. In this approach the system handles relevance feedback by assuming that the top-ranked documents returned to a given query are relevant. Terms are then extracted from these documents and are used to formulate a new enriched query. Previous studies have shown that pseudo-relevance can lead to significant improvements in retrieval performance [1, 3]. Our approach to content enrichment is similar in spirit to *pseudo-relevance feedback* domain except that top-ranking documents are used as a source of indexing terms (at indexing time) as opposed to a source of query terms at search time.

3 Mobile Content Enrichment

The main contribution of this paper is to describe a technique for enriching mobile content by leveraging existing Web search resources. The objective is to expand the limited content of a typical mobile page to produce an enriched version of this page for the purpose of indexing. The enrichment process involves using elements of the page’s original content as queries to a Web search engine, with the *enrichment terms* extracted from the top ranking results retrieved for these queries. The assumption is that this will lead to enrichment terms that are missing from the original content (document) but that are nevertheless useful for indexing, and that thus provide for additional retrieval opportunities.

Previously we [2] described a basic enrichment technique, focusing on how enrichment terms are extracted from the search engine results, and demonstrating search performance improvements in excess of 20%. In this paper, we will extend this work by focusing on two different stages of the enrichment process— (1) the extraction of query terms from the original content; and (2) the selection of results as a source of enrichment terms — and will demonstrate how these improvements have a further impact on search engine performance.

3.1 The Enrichment Process

To provide a general overview of the basic enrichment process, consider a mobile page S' . Enrichment is then a 5-stage process:

1. *Query Extraction*: Generate a query, $Q(S')$, or rather a set of q queries from S' by extracting k informative terms.
2. *Result Extraction*: Submit each query $Q_i(S')$ to a Web search engine (by default we use Yahoo) to generate a set of r results, $R_{Q_i(S')}$.
3. *Result Selection*: Select the r highest quality results from $R_{Q_i(S')}$ to produce a filtered set of Web search results, $R_{Q_i(S')}^{filtered}$.

4. *Term Extraction*: Analyse the content of these filtered results to extract a set of the n most informative enrichment terms $V(S') = t_1, \dots, t_n$.
5. *Page Indexing*: Index S' using a combination of its own terms and the enrichment terms; that is, $E(S') = S' \cup V_{Q(S')}$.

In this way each mobile page S' is expanded by a set of enrichment terms which have been selected because they appear to be related to the content of S' . The process above can be tuned in a variety of ways, by varying key parameters (such as the number of enrichment terms to add or the number of related results to select as a source of these terms) to facilitate a narrow or broad approach to enrichment. If the enrichment terms are too narrow then new retrieval opportunities may be limited. If they are too broad, while new retrieval opportunities may be readily available, retrieval precision may be reduced.

As mentioned above, our previous work [2] has focused on *Step 4* of the enrichment process, while using very straightforward approaches in the other steps. In the following sections we will focus on *Steps 1 & 3*. The assumption is that by improving the way in which we generate queries from the source content we can more reliably identify result pages that are likely to provide a good source of enrichment terms. Similarly, by being more selective in the result pages that we use as a source of these terms, we can produce a better set of final enrichment terms; for example, eliminating pages that are among the top retrieved results but that do not appear to be relevant to S' will help to improve the final quality of the enrichment terms.

3.2 Query Generation

The default approach to generating a query from S' , as reported in [2], was based on generating a single query for each S' from the top k most frequently occurring terms in S' (after stop-word removal). While this approach appeared to work reasonably well it is clear that there is considerable room for improvement. As such we propose two new strategies that involve the generation of multiple queries, instead of a single query, using two different query extraction techniques:

1. *MTF (multiple queries, term frequency extraction)*: Instead of generating a single query, we generate q ($q = 10$) queries of size k ($k = 5$) from the terms with the highest frequency in S' . Specifically we extract the top 10 terms with the highest frequency in S' and generate a list of q unique 5-term query combinations from these top 10 terms.
2. *MYH (multiple queries, Yahoo term extraction)*: Generate a set of q ($q = 10$) queries by submitting S' to the *Yahoo Term Extraction Tool*⁴. This tool returns a ranked list of key terms from a body of text (S'), which we use as a source to extract q k - term queries as above (with $q = 10$ and $k = 5$).

⁴ <http://developer.yahoo.com/search/content/V1/termExtraction.html>

3.3 Result Selection

In the enrichment technique employed by Church and Smyth [2], the results chosen as a source of enrichment terms were simply all of the top-ranking results returned by the underlying search engine for some query $Q_i(S')$. The problem with this approach is that it can lead to the inclusion of a result that has little or no relevance to the source content (S'), especially in the face of a vague query.

What is needed is a technique for filtering out result pages that are unlikely to serve as a good source of enrichment terms. One way to do this is to compare each result r_j to S' . If there is a significant overlap between the result and S' then we can reasonably infer a level of similarity or relatedness. Overlap on its own does not go far enough however because we would like to favour results that are related but that also offer extra terms that are missing from S' ; after all r_j may be very related to S' , but if $S' - r_j$ is essentially empty then r_j will not act as a source of new enrichment terms. For this reason we evaluate each result r_j by considering its overlap with S' and the availability of new terms that are missing from S' . We prefer results that have a significant overlap *and* that offer a significant number of new terms by using a harmonic mean of both factors to evaluate result quality; see Equations 1, 2, and 3. Thus, in this paper we will examine a result selection approach (*Sel*) which selects the top m Yahoo results with the highest quality scores. In this way high-ranking results that fall below this threshold will not be considered as a source of enrichment terms.

$$Quality(r_j, S') = \frac{1}{\frac{1}{Overlap(r_j, S')} + \frac{1}{Diff(r_j, S')}} \quad (1)$$

$$Overlap(r_j, S') = \frac{|r_j \cap S'|}{|S'|} \quad (2) \quad Diff(r_j, S') = \frac{|r_j - S'|}{|r_j|} \quad (3)$$

4 Evaluation

We have argued that enrichment is one way to improve the retrievability of mobile content and that existing search engines can be a valuable source of enrichment knowledge. In this section we will evaluate the impact of our new query generation and result selection components compared to a standard mobile search engine and the basic approach to enrichment described by [2].

4.1 Test Data

In this experiment we use a database of 1999 recent news stories, harvested from a Web-based news service during October 2005 - February 2006, as the basis for a mobile news search engine. News stories were chosen for two important reasons. First of all, news is a good example of the type of content that is popular on the mobile Internet. Second, it is relatively easy, for the purpose of this experiment, to convert a long-version of a news story into a shorter, mobile version that will form the basis of the mobile news content. Typically these shorter versions

were about 10-20% of the original and, in the case of our news content, this truncation process was straightforward because each story was preceded by a concise summary of the longer text. The point of this is that the additional story content, which did not make it into the mobile form of the story, was then available as a plausible source of potential target test queries. In this way, during performance testing, for each indexed mobile story (*target story*), we generated test queries (using the same technique for generating enrichment queries) from the content that was missing from the mobile form; thus test queries will often contain terms that are not present in the mobile version of the story, but that are nonetheless relevant as potential query terms. These test queries were submitted to each test search engine and we measured the percentage of times that the target story was retrieved among the top 10 results.

4.2 Test Search Engines

In all we evaluated 6 different test search engines, each corresponding to a different approach to indexing and enrichment: *SE1* used no enrichment and news stories were indexed using their existing content only; the remaining 5 search engines implement different version of our enrichment strategies. All of the search engines were implemented using the *Lucene*⁵ platform and Yahoo was used as the underlying search engine responsible for providing enrichment results. The summary details for each search engine are as follows:

1. *SE1*: This search engine provided a baseline as it used only standard mobile content to index the news stories.
2. *TF*: This search engine uses a basic enrichment technique (described by [2]) in which each news story is enriched from the top 10 ranking results returned from Yahoo based on a single enrichment query made up of the top k most frequent terms in the mobile story.
3. *MFT*: This search engine uses the same enrichment technique as *TF* except that 10 enrichment queries are used to provide enrichment results.
4. *MYH*: This search engine uses the same enrichment approach as *MTF* except that the enrichment queries are generated using the Yahoo Term Extraction Tool as discussed in Section 3.2.
5. *MTFSel*: This variation operates in the same way as *MTF* except that in addition, the enrichment results are filtered using the quality metric discussed in Section 3.3.
6. *MYHSel*: This variation operates in the same way as *MTFSel* except that the enrichment queries are generated using the Yahoo Term Extraction Tool as discussed in Section 3.2 and results are filtered using the quality metric as discussed in Section 3.3.

Note that k , the number of terms used in our mobile query $Q(S')$, is set to 5 for each search engine in our experiments.

⁵ <http://lucene.apache.org>

4.3 Methodology

An experimental run involves testing the retrievability of each of the 1999 news stories. For each *test story*, we generate a set of test queries q , where $q = 3$ from its non-mobile content. To generate the test queries we vary p (the number of terms used in test queries) between 1, 2 or 3 terms by selecting the t most frequent terms in the remaining portion of news content that is not part of the original news stories S' . Next we submit each test query to the 6 test search engines and compute the average percentage accuracy (we will refer to this as the *success rate*) for each of the test search engines as the percentage of times that the target result (the current test story) is found in the top 10 results returned. This is repeated for a variety of different parameterizations of our content enrichment technique by varying $r = 3, 10$ (the number of Yahoo results extracted) and $n = 10, 50, 100, 200$ (the number of terms used for enrichment), to understand how retrieval performance varied under different experimental conditions.

4.4 Overall Successful Rates

Figure 1 presents the overall success rate for each of the 6 test search engines averaged over all parameter settings. Although the average success rates are reasonably low—e.g., the results show that indexing mobile news stories by their own content delivers successful retrievals approx. 20% of the time, on average—we must remember that the procedure we are using enforces a strict notion of relevance, in the sense that only one news story is considered to be relevant for each query. In reality, searchers are likely to entertain a variety of relevant results for their queries. As such it is reasonable to interpret the results as lower bounds on search engine performance. The average searcher is likely to enjoy better search engine performance in practice, but we predict that comparable relative differences between the search engines will remain under such weaker relevance conditions.

Returning to the success rates for each of our test search engines, we find that the basic search engine *SE1* (which uses no enrichment), fails to find the target news story in 79% of cases. In contrast, we see that on average, all of the enriched engines do significantly better. For example, the basic enrichment technique (TF) offers improvements of around 23% over *SE1*. However, the results also show that each of our enhancements to the basic enrichment technique have a positive effect on retrieval performance. For example, *MTF* and *MYH* take advantage of multiple enrichment queries to offer improvements of 27% over *SE1*. *MTFSel* which incorporates a quality-based result selection phase offers an even greater overall benefit (34%). Overall, the best performing technique, *MYHSel*, combines the use of multiple queries, the more sophisticated Yahoo term extraction service to generate enrichment queries, and quality-based result selection, to achieve a relative success rate increase of 37% over *SE1*.

These results clearly demonstrate that we can significantly increase retrieval performance by (1) using multiple queries at the query generation phase, (2) using more sophisticated term extraction techniques, and (3) employing the quality metric to filter out any *low-quality* enrichment results.

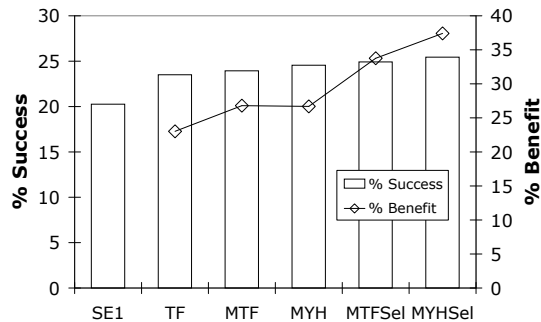


Fig. 1. Average overall percentage success rates for each of the 6 search engines and the relative benefit of the 5 enriched search engines compared to SE1.

4.5 Varying n : the number of enrichment terms

Varying the number of additional enrichment terms is likely to have a significant impact on retrieval performance. The more relevant the terms selected for indexing, the better able the search engine will be to return the target document for a relevant query. The risk, however, is that if too many irrelevant (or at least less relevant) terms are used to index a document then this will increase the likelihood of false reminders at search time. If this happens we can expect to find many irrelevant documents being retrieved during a typical search, thereby increasing the likelihood that the target document will be pushed further down the result-list and thus reduce search success.

The results obtained by varying n , the number of enrichment terms used during indexing, are presented in Figure 2 and 3, by averaging the experimental runs for each particular value of n . Once again, we find a significant advantage accruing to the enriched search engines. Indeed we see that as the enrichment terms are increased from 10 to 200 there is a steady improvement in overall search engine performance, relative to *SE1*. At only 10 enrichment terms there is little significant advantage for the enriched search engines but the use of 200 enrichment terms sees a minimum relative improvement of 33% for the basic enrichment technique *TF*, while *MYHSel* achieves a 73% relative improvement over *SE1*; note 73% is the average of the individual relative benefits found for *MYHSel* during the $r = 3$ and $r = 10$ runs, at $n = 200$.

Again it is clear that the use of multiple queries, the more sophisticated term extraction technique for query generation, and the quality filter for result selection, all have an independently positive effect on overall retrieval performance. For example, the use of multiple queries contributes up to 20% in relative benefit terms, whereas the use of the more sophisticated term extraction technique adds another 6%, and the result selection filter adds a further 14%.

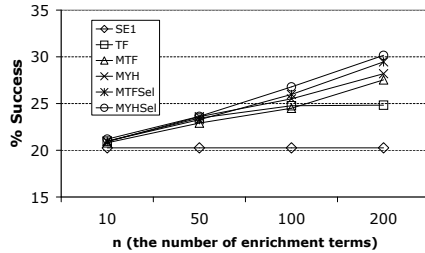


Fig. 2. Search success rates

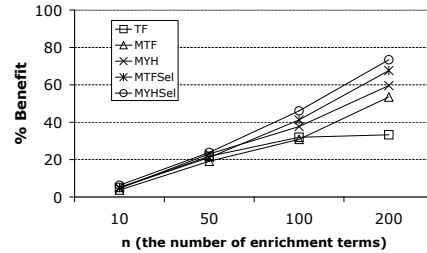


Fig. 3. Average benefit relative to SE1

4.6 Varying r : the number of enrichment results

Another critical parameter in the enrichment process is the number of results that are retrieved in response to each enrichment query. These results provide the raw material from which the enrichment terms are extracted. If too many results are retrieved then it is likely that less relevant sources of enrichment terms will be considered during the term selection phase. This will potentially reduce the quality of enrichment terms. At the same time, if we retrieve too few results then we are limiting the enrichment process to a reduced set of potential enrichment terms and this could lead to missed indexing opportunities.

The results for 2 different settings for r (3 and 10) are presented in Figure 4 and 5. As before we see that the enriched search engines all do significantly better than *SE1* across the different values for r . Interestingly, using fewer results ($r = 3$) is seen to have a positive impact on overall retrieval performance. For $r = 10$ each of our enriched search engines offers improvements over *SE1* of between 23%-33%. However, when $r = 3$ we see improvements in the range 23%-42%. The results suggest that higher ranking results are a better source of enrichment material and while the result-selection filter helps to control the quality of the results used in enrichment, at $r = 10$ some lower quality results are being incorporated. Remember that in this experiment we extract the top r results with the highest threshold. At $r = 3$ the average quality of selected results is approximately 0.4, but at $r = 10$ this average quality decreased to 0.36. Thus the higher quality results selected for $r = 3$ have contributed to an overall increase in performance for this setting.

5 Conclusions

Mobile search is challenging because of fundamental device and content limitations. Mobile content, by its very nature, tends to be brief and can lack the richness that is needed to build a good search index and, in this paper, we have addressed this problem by presenting a heuristic content enrichment framework that leverages traditional Web search engines and tools as a way to enrich mobile content prior to indexing. We have described a number of enhancements to

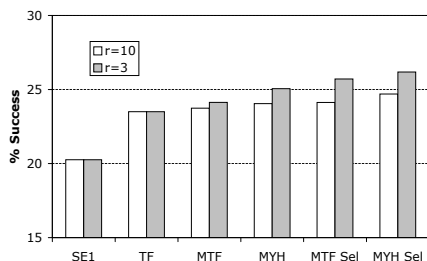


Fig. 4. Search success rates

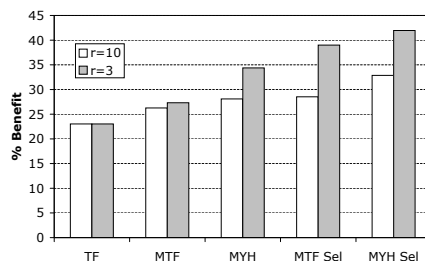


Fig. 5. Average benefit relative to SE1

our basic content enrichment strategy and evaluated the performance of each of these enrichment strategies, relative to a benchmark mobile search engine, and under a variety of experimental conditions. The results demonstrate significant improvements in overall search engine performance.

This content enrichment framework provides many opportunities for further developments beyond the term-based selection and weighting techniques used to date. For example, a logical next step is to take advantage of linguistic knowledge and simple natural language processing techniques in order to guide more meaningful enrichment. Moreover, there are a number of opportunities to look at the application of machine learning techniques to the extraction of suitable enrichment terms and the classification of indexing knowledge, for example.

References

1. C. Buckley, G. Salton, J. Allan, and A. Singhal. Automatic Query Expansion Using SMART: TREC 3. In *Proceedings of the TREC-3*, 1994.
2. K. Church and B. Smyth. Mobile Content Enrichment. In *Proceedings of the International Conference on Intelligent User Interfaces (IUI'07)*, pages 112–121, 2007.
3. E. N. Efthimiadis and P. V. Biron. UCLA-Okapi at TREC-2: Query Expansion Experiments. In *Proceedings of TREC-2*, pages 278–290, 1993.
4. G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais. The Vocabulary Problem in Human-System Communication. *Communications of the ACM*, 30(11):964–971, 1987.
5. M. Ramsay and J. Nielsen. The WAP Usability Report. Neilson Norman Group. 1983.
6. J. J. Rocchio. Relevance Feedback in Information Retrieval. *Gerard Salton, editor, The SMART Retrieval System - Experiments in Automatic Document Processing*, pages 313–323, 1971.
7. G. Salton and M. McGill. Introduction to Modern Information Retrieval. 1983.