

Mobile Content Enrichment

Karen Church

Adaptive Information Cluster,
Computer Science & Informatics,
University College Dublin,
Belfield, Dublin 4, Ireland
+353-1-7165355
karen.church@ucd.ie

Barry Smyth

Adaptive Information Cluster,
Computer Science & Informatics,
University College Dublin,
Belfield, Dublin 4, Ireland
+353-1-7162473
barry.smyth@ucd.ie

ABSTRACT

Delivering an effective mobile search service is challenging for many reasons. Certainly small-screen mobile handsets with limited text input capabilities do not make ideal search devices. In addition, the brevity of Mobile Internet content hampers effective indexing and limits retrieval opportunities. In this paper we focus on this indexing issue and describe an approach that leverages Web search engines as a source of content enrichment. We present an evaluation using a mobile news service that demonstrated significant improvements in search performance compared to a standard benchmark system.

ACM Classification: H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval, H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing - Indexing methods

General terms: Experimentation, Measurement, Performance

Keywords: Mobile Search, Mobile Web, Mobile Internet, Content Enrichment, Index Enhancement

INTRODUCTION

The Mobile Internet offers anytime, anywhere access to a wealth of information, services and applications to millions of users across the globe. Improvements in mobile data services and billing models, developments in wireless infrastructure, and advances in mobile handset technology have lead to rapid growth in the mobile space. Indeed there is now evidence that mobile phones will become the primary gateway to digital content, as mobile handsets will soon vastly out-number PCs, especially in developing markets. At the same time, there has been rapid growth in mobile content and today millions of users are accessing content through their mobile handsets via walled-garden operator portals. But these walled gardens are unlikely to hold user interest for long and recent predictions point to the rapid growth of mo-

bile search technologies in the coming years as a way to help users discover new content and services online. If search engine technologies are to succeed, however, they must overcome a number of significant challenges presented by the Mobile Internet. In this paper we will consider one such problem, related to the limited content that is available for indexing mobile pages, and describe a technique for enriching this content to improve index quality.

Mobile Internet Access

The latest estimates from Wireless Intelligence indicate that the total number of cellular subscribers reached 2.5 billion in September 2006 [4] highlighting the phenomenal growth of the mobile sector: it took the cellular industry twenty years to reach a total of 1 billion subscribers, three more years to reach 2 billion, and by 2007 there will be an estimated 3 billion subscribers as current growth rates achieve 40 million new subscribers per month. The lion's share of this growth stems from emerging markets such as China, India and Russia, where traditional forms of Internet access have been limited [4]. Thus, across the globe mobile devices will soon surpass PCs as the primary Internet access device, and in these emerging markets it will bring the Internet to the masses for the first time [11].

The largest mobile carriers in the US, namely, Cingular Wireless, Sprint Nextel, T-Mobile and Verizon Wireless, are currently experiencing huge growth in wireless data usage. Together they generated more than \$6.3 billion in wireless data revenues for the first half of 2006 and they look set to generate more than \$15 billion in data revenue by the end of 2006. This represents a staggering 75% increase from 2005, when data services for the entire year accounted for 8.6 billion [27]. A recent study by Ipsos Insight shows that 28% of mobile subscribers worldwide have browsed the Internet using their mobile phone. 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 dominate wireless Internet access. In Japan 40% of subscribers use the Mobile Internet while in Europe, France and the UK are leading the way [11].

Towards Mobile Search

Recent evidence shows that users are beginning to venture beyond the walled-garden of the operator portal in search of interesting mobile content leading to increase interest in the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IUT'07, January 28–31, 2007, Honolulu, Hawaii, USA..

Copyright 2007 ACM 1-59593-481-2/07/0001 ...\$5.00.

potential for mobile search engines to dominate as the primary mode of information access for users, as they have on the World Wide Web. Certainly there has been significant search-related activity within the mobile sector. For example, major players within the search engine industry like Google, Yahoo, AOL, etc. continue to release exciting new improvements to mobile search services. The latest statistics from the US also indicate that mobile search usage is increasing. New findings published by the Mobile Marketing Association (MMA) show that 31% of users surveyed have used mobile search for the first time in July/August 2006, with existing users doing so an average of nine times. Another 48% said they expect to start in the next few months and estimate they would do so about once a week [14]. Therefore, the potential impact of the Mobile Internet and in particular mobile search is profound.

However, there are significant issues that need to be addressed if mobile search is to prove to be a successful information access methodology in the mobile sector. There are of course many well-documented challenges when it comes to leveraging a mobile handset as an information access device. Certainly small screen sizes and limited text input capabilities will have a significant impact on the use of these handsets as search devices, limiting the type of queries that will likely be provided and the number of results that can be displayed. See Figure 1 for an illustrative example of Google’s mobile search service.

These device limitations, however, are but one side of the coin. The nature of mobile content itself is such that it introduces new problems from an indexing and search standpoint. Mobile pages are typically much shorter in length than their Web counterparts; most WAP gateways and mobile phones limit the size of the pages they can process, and traditionally WML decks are typically restricted to just a few KB. Mobile Internet access is mainly about *content snacking* but this means 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* that plagues Web search — referring to the tendency for searchers to often chose 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.

Enriching Mobile Content to Improve Search

Thus, the main problem that we address in this paper is how one might improve a mobile search index to facilitate better retrieval performance — to bridge the vocabulary gap between the user and the target page. Our solution is based on a *content enrichment* strategy, whereby we seek to augment basic mobile content with additional terms for the purposes of indexing. Briefly, the limited content available in mobile pages is elaborated with additional informative terms. To obtain these terms we leverage a standard Web search engine. That is, to elaborate a mobile page p we first convert this page into a set of queries q_1, \dots, q_n , submit these queries to the search engine, and then extract informative terms from the top ranking search results. During the course of this paper we will describe different strategies for identifying the most

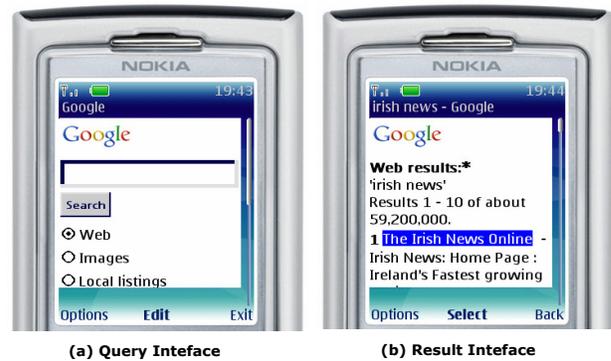


Figure 1: Google’s mobile search - (a) shows Google’s query interface (b) shows the results page generated for the query *irish news*.

informative terms to extract. In addition, we describe the results of a comprehensive evaluation using a mobile news service.

RELATED WORK

We have identified two different areas of related research that are particularly relevant when it comes to content enrichment. Both areas focus on the wider Web. The first research strand is called *query expansion*. We refer to the second strand of research as *index enhancement*.

Query Expansion

Short queries and term mismatching can cause relevant documents to be omitted from search results because they do not contain the terms within a users 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* [18, 23]. The standard approach to relevance feedback 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. This standard approach was enhanced by Salton & Buckley in 1990 [22]. They devised a simple vector modification process that not only adds new terms to the vector but also modifies the weights of existing query terms according to whether the terms are present/absent in previously retrieved documents. Their approach yielded relative improvements of between 50-150%.

Although existing research shows that relevance feedback can achieve significant improvements in retrieval performance, it requires that users provide accurate relevance judgments. However, in real search scenarios, 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 [3, 7].

To date there have been a number of advances in the area of *pseudo relevance feedback*, with researchers making use of past queries and past result selections as efficient sources of information. For example, Fitzpatrick & Dent were able to obtain improvements of approx. 15% in average precision using past query data [8]. Their approach was to identify similar queries to the user's initial query, retrieve the top 100 documents for each of these *similar* queries and using TF.IDF techniques extract candidate terms from these relevant documents.

Hust et al. leveraged past queries and their relevant documents for improving retrieval performance [10]. In their approach, documents that were relevant to previous similar queries are assumed to be relevant to the current query. These *relevant* documents are retrieved and their terms are used to enrich the current query. The authors obtained highly positive results when they combined their new method of query expansion with a pseudo-relevance feedback method similar to that of Fitzpatrick & Dent [8].

Billerbeck et al. utilized *query associations* as an effective source of expansion terms [1]. The idea behind *query associations* is that each document within a collection is associated with a set of queries. The authors extracted the queries associated with the top K documents returned to an initial user query and used these as their source of enrichment terms. Using TREC-10 data, the authors showed that their expanded queries were at least 26% better than standard queries.

Finally, Cui et al. examined the use of past search information to extract query expansion terms [5]. Using Web search logs, the authors calculated probabilistic correlations between query terms and document terms. These correlations were then used to extract enrichment terms used for query expansion. This log-based query expansion technique showed an improvement of approx. 39% over the pseudo-relevance technique and approx. 75% over the baseline method.

Index Enhancement

Index enhancement comes in the form of *summarization* or *enrichment* of indices. These enhanced indices are designed to increase retrieval effectiveness and efficiency. *Summary indices* use summaries instead of full-texts as a source of index terms. Brandow et al. reported on the results of a simple experiment that compared summary versus full-text indexes [2]. They used 12 Boolean queries and a dataset of approx. 20,000 documents for their evaluation. Their results showed that a summary index can improve overall precision but a large loss in recall was also reported. Sumita & Iida carried out a similar smaller study involving 10 queries and 600 news stories. They reported an improvement in the overall effectiveness of retrieving highly relevant documents using the summary index [25].

Sakai & Sparck-Jones [19] tested two hypotheses in their experiments, namely, that a summary index can be as effective as the corresponding full-text index for retrieving highly

relevant documents and that a summary index can be useful for pseudo-relevance feedback (described in the previous section). Using TREC data for their experiments, their results illustrated that a summary index can be as effective as a full-text index for *precision-oriented search of highly relevant documents*. Regarding the pseudo-relevance feedback, they found that using a summary index at the initial search phases and a full-text index at the final search phases yielded the best results.

Wasson compared summary indices to full-text indices using 30 searches and a large collection of documents from the LexisNexis news library [26]. He shows that summary indexes were suitable when users wanted smaller result sets that included highly focused and relevant documents while full-text indexes were suitable when users want to retrieve all references to a particular topic.

Rajashekar and Croft used the INQUERY retrieval system to combine manual (i.e. keywords and thesaurus terms) and automatic index representations of both queries and documents of the INSPEC collection. Their experiments yielded consistent improvements when these *combined* representations were employed while the most significant results were achieved when all available representations were used, i.e. controlled vocabulary, keywords and text [15].

Another form of index enhancement includes *enrichment* of indexes, that is, supplementing an index with useful information. *Semantic indexing* is one such technique that uses lexical-semantic resources like WordNet to disambiguate terms. For example, Gonzalo et al. performed experiments using the SMART retrieval system and the Semcor disambiguated collection [9]. Their results showed up to a 29% increase in recall after using synsets for indexing. Mihalcea & Moldovan used WordNet to add lexical and semantic information to both the query and the documents during pre-processing. Their semi-complete disambiguation algorithm was able to disambiguate approx. 55% of the nouns and verbs and proved highly precise, yielding over 92% accuracy in experiments [13].

Another technique which relates to *Semantic Indexing* is *Latent Semantic Indexing (LSI)* [6]. LSI is an intelligent retrieval/indexing technique which uses singular value decomposition to reduce the dimensions of the term-document space. By investigating how terms co-occur across documents, LSI can model the associations between terms and documents. By exploiting the underlying semantic relationships that exist between terms and documents, LSI has the ability to solve many of the synonymy and polysemy issues that plague traditional IR systems. LSI has previously been shown to improve overall retrieval performance when compared to standard approaches. For example, Deerwester et al. carried out experiments which compared LSI to SMART [20] using two standard collections (MED and CISI). They found that the average precision for MED increased from 0.45 to 0.51 while no significant differences were found for the CISI collection [6].

MOBILE CONTENT ENRICHMENT

The main contribution of this paper is 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. Typically these terms do not occur in the mobile content but they are used at indexing time in the hope that they will prove to be useful retrieval cues during future mobile searches. In turn we will describe and evaluate a number of different techniques for identifying informative enrichment terms from result-lists.

The Enrichment Process

Consider a mobile page S' . Our enrichment process proceeds in the form of a number of basic steps:

1. Query extraction: generate a query, $Q(S')$, or rather a set of q queries from S' by extracting k informative terms.
2. Query submission: each query is submitted to a Web search engine (by default we use Yahoo) to generate a set of r results, $R_{Q(S')} = R_1, \dots, R_r$.
3. Term extraction: convert $R_{Q(S')}$ into a set of terms and select the n most informative terms using one of our *enrichment techniques* to produce an *enrichment vector* $V_{Q(S')} = t_1, \dots, t_n$.
4. Page indexing: combine S' with its associated enrichment vector ($E(S') = S' \cup V_{Q(S')}$) and use this content to index S' .

Figure 2 illustrates the enrichment process using a sample news story from our dataset. In this example the news article relates to an explosion in Bali in October 2005. Using this news story we generate a query $Q(S')$ and set k , the number of query terms to 3. The query we generate is *bali blasts dead*. Next we submit the query $Q(S')$ to Yahoo and extract a set of r results. In this example, we set $r = 10$. To generate our pool of candidate enrichment terms we concatenate each of the Yahoo results together (i.e. the titles and snippets) to form a single string, we tokenize this string to extract a list of distinct terms and finally remove any stopwords. The resulting set of terms represents our pool of candidate enrichment terms.

Next we generate our enrichment vector $V_{Q(S')}$ by extracting n enrichment terms using one of four term selection techniques. In this example we set $n = 10$ and the enrichment vector we generate is (*dead, people, blasts, bali, bombings, time, ap, terror, Saturday, resort*). We then combine this enrichment vector $V_{Q(S')}$ with the original news story S' to produce our enriched news story $E(S')$. We then index this enriched content and use the resulting index to search.

Thus, in this way each mobile page S' is expanded with a set of enrichment terms. By varying q, k, r and n we can influence the number and quality of enrichment terms. For

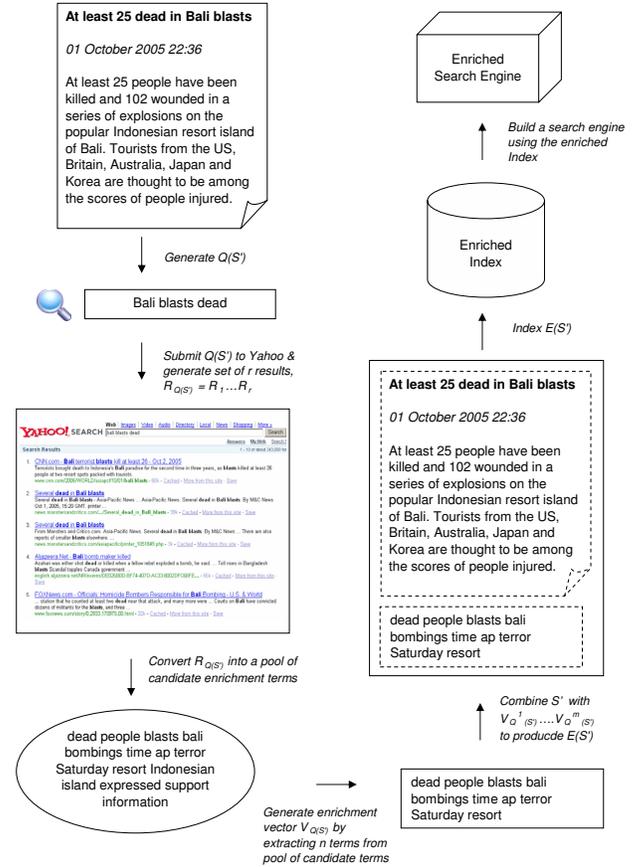


Figure 2: Enrichment example illustrating our enrichment process.

example, by limiting k, r and n we can limit the scope of the enrichment terms to focus closely on the content of S' , but in doing so we may limit the retrieval opportunities under which S' can be retrieved in the future. On the other hand, increasing k, r and n will lead to the generation of a broader set of enrichment terms, thereby facilitating a much broader set of retrieval. The danger now is that the S' may be retrieved for queries that are not so relevant to its content.

In the above procedure there are also a number of opportunities for varying the manner in which queries are generated (Step 1) and the way in which we score, rank and select terms from the result-lists (Step 3). In this paper we adopt a straight forward approach to the enrichment query extraction (Step 1). That is, for each S' we generate a set of enrichment queries, $q = 3$ with 3, 5 or 10 terms from the t most frequent terms in S' . Specifically, we extract the set of distinct terms from S' , remove any stopwords and select the top 1, 2 or 3 terms by their frequency. However, we examine a number of different approaches to ranking and selecting enrichment terms from $R_{Q(S')}$ (Step 3) as we will now discuss.

Selecting Informative Enrichment Terms

The primary aim of this paper is to examine alternative approaches for generating sets of the most *informative* terms to use as enrichment terms, ultimately during the indexing

process. Previous work has been limited to a standard term frequency technique, selecting high frequency terms from result-lists. Despite its simplicity it proved to be very effective at improving index quality and increasing search performance under a variety of evaluation conditions. In the following we describe a number of alternative strategies.

1. Highest Term Frequency (HTF)

This is our baseline technique and involves selecting n terms from a *pool* of candidate enrichment terms in decreasing order of their term frequency. For example if n is set to 10 we select the top 10 terms with the highest term frequency from the pool of candidate terms.

2. Lowest Term Frequency (LTF)

An alternative to selecting the highest frequency terms is to look for low frequency terms on the grounds that such terms are often more discriminating of a particular piece of content when compared to more popular terms. With this in mind, our second technique is essentially the inverse of HTF, selecting the n least frequent terms among the top ranking result-list documents.

3. Term Frequency Inverse Document Frequency (TF.IDF)

TF.IDF is a well-known term weighting technique in information retrieval research that argues that useful terms are those that occur frequently in the target document but infrequently in the document collection as a whole [21]. Thus each term is weighted as a combination of its term frequency within the target document and its frequency across the entire document collection; see Equation 1

$$w_{ij} = tf_{ij} * idf_j = tf_{ij} * \log(N/df_i) \quad (1)$$

where tf_{ij} is the frequency of term i in document j , N is total number of documents, df_i is the number of documents that contain term i , idf_i is the inverse document frequency is the logarithm of N divided by df_i . In our evaluation the *documents* are represented by the Yahoo results, $R_{Q(S')}$.

4. Robertson Selection Value (RSV)

Our final technique is borrowed directly from the query expansion / relevance feedback domain. RSV is a well-known approach for ranking candidate query expansion terms used in relevance feedback [16]. The *RSV* of a term i is defined in Equation 2, where $r(i)$ is the number of relevant documents containing term i , and $rw(i)$ is the standard Robertson/Sparck Jones relevance weight [17]:

$$rsv(i) = r(i) \times rw(i) \quad (2)$$

$$rw(i) = \frac{(r(i) + 0.5)(N - n(i) - R + r(i) + 0.5)}{(n(i) - r(i) + 0.5)(R - r(i) + 0.5)}$$

where $n(i)$ is the total number of documents containing term i , R is the total number of relevant documents for this query, and N is the total number of documents.

In our experiments we explore an alternative approach to the standard *rsv*. Firstly, each document is represented by a Yahoo result $R_{Q(S')}$, i.e the title and snippet text returned by Yahoo, rather than the actual web page associated with each Yahoo result. This means that when we want to determine the presence of term i we examine the title and snippet rather than the full document. Secondly, we set R the number of relevant documents to be the number of Yahoo results we examine i.e. 10, 20 or 50. Finally we restrict the total number of documents N to a maximum of 100. This was mainly due to limitations regarding how many results Yahoo returns per query.

EVALUATION

The motivation behind the work presented in this paper is the construction of mobile search technologies that are better adapted to the challenges presented by the Mobile Internet and its content. Specifically we have argued that, because Mobile Internet pages tend to be much shorter than their Web counterparts, there are significant problems when it comes to indexing mobile pages in a way that facilitates flexible retrieval. As part of a proposed solution to this particular problem, we have described a strategy for enriching mobile content by locating additional indexing terms with a view to improving index quality and increasing search performance. In this section we will describe the results of an initial investigation to assess the potential of this approach to deliver meaningful increases in retrieval performance under a variety of experimental conditions.

Setup

To test our enrichment approach we developed 5 versions of a mobile news search engine that indexes a recent-news database of 3532 news stories harvested from a local Web-based news service during October 2005-May 2006. The different search engine versions each corresponded to a different approach to indexing and enrichment: *SE1* used no enrichment and news stories were indexed using their existing content only; *SE2* – *SE5* implemented different version of our enrichment strategies (HTF, LTF, TD.IDF, and RSV, respectively) so that each story was indexed using a (potentially) different set of index terms. All of the search engines were implemented using the *Lucene*¹ platform.

News stories were chosen for a couple of important reasons. First of all, news is a good example of the type of content that is popular on the Mobile Internet. Mobile news stories tend to provide short synopses of current news that is well suited to mobile content delivery. Providing search support over an archive of news content is one important, and often requested, feature that is missing from many existing mobile news services.

The second reason for choosing news content is that it facilitates an automatic approach to search engine performance testing. Each news story that we harvested represented a full-story — typically about a page of Web content — and was truncated to produce a shorter version (the *mobile form*) that was suitable as a mobile page. Typically these shorter versions were about 10-20% of the original and, in the case

¹<http://lucene.apache.org>

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 that did not make it into the mobile form of the story was available as a plausible source of potential target test queries. In this way, during performance testing, for each indexed story we generated a test query using the same technique for generating enrichment queries. These test queries were submitted to each search engine and we measured the percentage of times that the target news story was retrieved among the top 10 results.

Thus, each experimental run involved submitting test queries for each of the 3532 news stories to each of the 5 test search engines, producing 5 average percentage test scores. This was repeated for a variety of different parameterizations of our content enrichment technique — varying $k = 3, 5, 10$ (the number of terms used to generate the enrichment $Q(S')$ queries), $r = 10, 20, 50$ (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. We also varied p (the number 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' . This allowed us to determine the effect of varying the test queries on the performance of each technique.

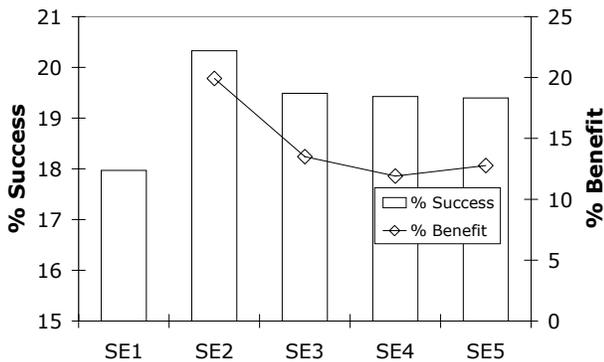


Figure 3: Average search success rates and average relative benefit compared to SE1 in terms of percentage success rates.

Overall Successful Rates

To begin with Figure 3 presents the overall *success rate* and *relative benefit* for each of the 5 test search engines averaged over all parameter settings. The first thing to notice is the average success rates are reasonably low; for example, the results show that indexing mobile news stories by their own content delivers successful retrievals less than 18% of the time, on average. However, we must remember here that the procedure we are using encodes a fairly tough relevance criterion in the sense that only one news story is considered 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 here as reflecting reasonable lower bounds on search engine performance. The

average searcher is likely to enjoy better search engine performance in practice but we believe that relative differences between the different search engines will remain under such weaker relevance conditions.

Returning to the success rate for SE1: we find that when we generate queries by extracting terms from an expanded version of these truncated mobile stories — a reasonable source of relevant content terms — that most of the time (82%), these queries do not result in the retrieval of the target story among the top 10 results. In contrast, we see that, on average, all of the enrichment techniques do significantly better. For example, the best performing technique, SE2 (which uses the HTF strategy), achieves a success rate of more than 20%. While this overall increase is hardly dramatic, it does represent a relative increase of almost 20% over SE1. Moreover, as we shall see in the following sections this average disguises some much more significant potential benefits under a number of reasonable parameterizations that are likely to correspond to realistic mobile search scenarios.

Varying k : the number of enrichment query terms

To begin with we will look at the influence of the number of terms used to generate enrichment queries on the success rates of the enriched search engines. During the experiment we varied k to allow for enrichment terms that contained 3, 5, and 10 query terms. Intuitively, too few query terms will likely produce vague queries that are unlikely to result in the retrieval of relevant search results and this will produce enrichment sets that contribute little to the indexing of the target document. On the other hand, too many query terms are likely to produce queries that are too specific and may limit the number of results that are retrieved, which can also impact the quality of the enrichment sets.

The success rate results for varying k presented in Figure 4 (a) and (b) confirm these intuitions. These results reflect average success rates computed across the subset of experimental runs corresponding to each particular value of k . Enrichment queries with 3 or 5 terms result in significant improvement for each of the search engines SE2-SE5 compared to SE1, but using 10 terms per enrichment query degrades this advantage significantly. For example, we see that SE2 once again performs best, achieving success rates in excess of 21% for enrichment queries with 3 and 5 terms (a relative benefit over SE1 of 25%), but this drops to just over 19% for 10-term queries.

Varying n : the number of enrichment terms

Obviously, varying the number of additional terms is likely to have a significant impact on retrieval performance. The more relevant terms selected and used 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.

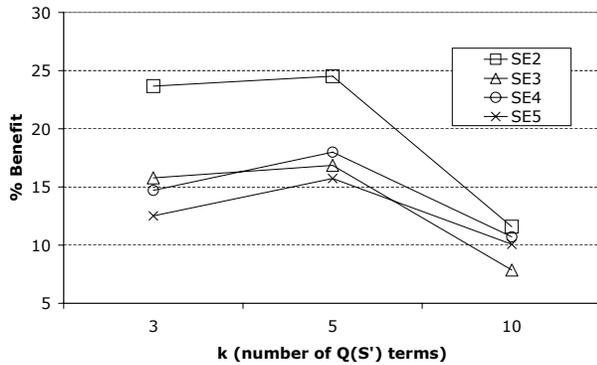
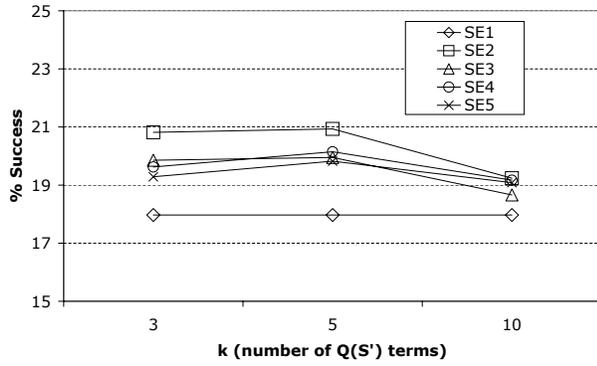


Figure 4: (a) Search success rates and (b) average benefit relative to SE1, for varying enrichment query sizes (k).

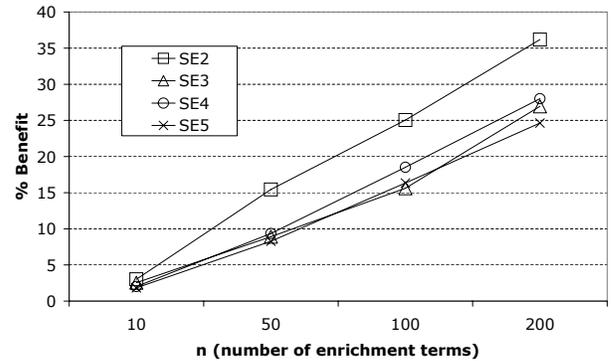
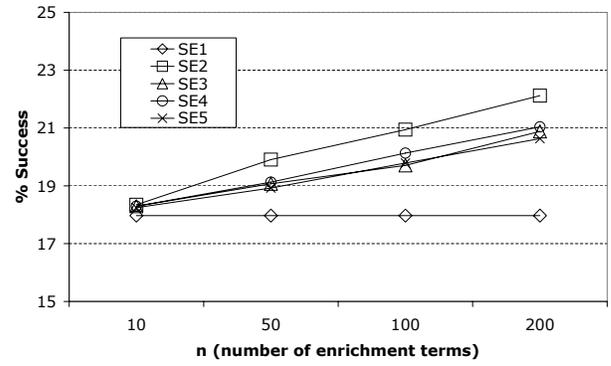


Figure 5: (a) Search success rates and (b) average benefit relative to SE1, for varying the number of enrichment terms.

The results obtained for varying n , the number of enrichment terms used during indexing, are presented in Figure 5 (a) and (b); once again we average the experimental runs for each particular value of n . And once again, we find a significant advantage accruing to the enriched search engines SE2-SE5. 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 SE2-SE5, but the use of 200 enrichment terms sees SE2, for example, achieving a very significant 36% relative improvement over SE1. Once again SE2 is seen to produce consistently better retrieval performance than the other enrichment strategies and, perhaps surprisingly, there is no evidence (up to 200 enrichment terms) of a decline in search engine performance due to the inclusion of less relevant enrichment terms.

Varying r : the number of enrichment results

The next 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 the 3 different settings for r (10, 20, and 50) are presented in Figure 6 (a) and (b) and show some interesting behaviours. As before we see that the enriched search engines all do better than SE1 across the different values for r , and even when only 10 results are retrieved per enrichment query we still see a significant improvement in search engine performance (10%-17%) for SE2-SE5.

Interestingly, there seems to be an inflection point at $r = 20$, with two of the enrichment strategies (SE2 and SE3) suffering from a reduction in search performance for $r = 50$ while the other two strategies (SE4 and SE5) enjoy a significant improvement in search engine performance at $r = 50$. For instance, we can see that SE3 experiences a decline in its benefit over SE1, from just over 15% to below 10%, as r is increased from 20 to 50 results. In contrast both SE3 and SE4 see their advantages over SE1 increase significantly from about 11% to just under 17% for the same values of r . Indeed we see that SE3 and SE5 are particularly sensitive to variations in r with the former's advantage dropping consistently from $r = 10$ and the latter's increasing consistently in the same range.

These results suggest that the simpler term weighting strategies used by SE2 and SE3 become increasingly less suc-

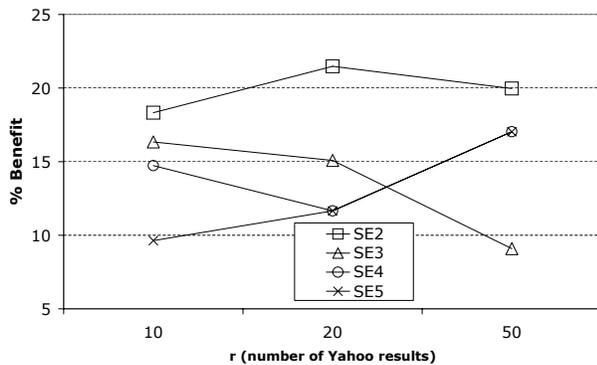
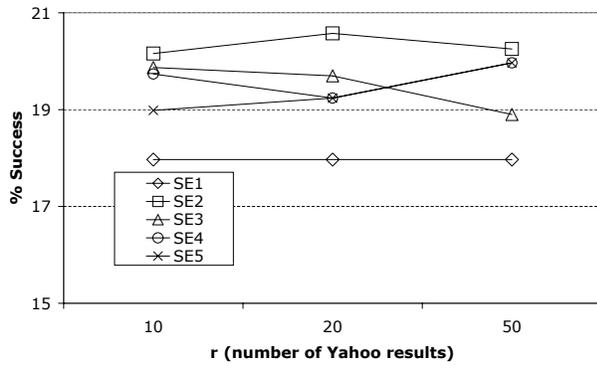


Figure 6: (a) Search success rates and (b) average benefit relative to SE1, for varying the number of enrichment results examined.

successful as more enrichment results are retrieved, whereas the more sophisticated strategies used by SE4 and SE5 appear to come good as the number of r increases. As future work it remains to be seen whether SE4 and SE5 eventually come to beat SE2 and SE3 for values of r greater than 50, as the graphs presented here suggest may happen.

Varying p : the number of test query terms

In the above sections we varied important parameter settings (k , n , and r) used during the enrichment process itself. In addition to these parameters we also must examine the impact of the evaluation test function used in this experiment. This function relies on artificially created test queries, extracted from news content that is not part of the indexed news stories. These queries are generated using 1, 2, or 3 query terms, mainly because these are reasonable length queries that are likely to occur in practice during mobile search; for recent evidence supporting this see [24, 12]. Once again we are interested in a trade-off: too few query terms make for vague queries and poor retrievals, but too many make for over-specified queries which may also limit retrieval performance.

The results obtained by varying p are presented in Figure 7 (a) and (b). Perhaps as expected, we find that for test queries made up of a single term (that is, $p = 1$), the enriched search engines actually do marginally worse than SE1. It seems that adding enrichment terms to their indices affords more oppor-

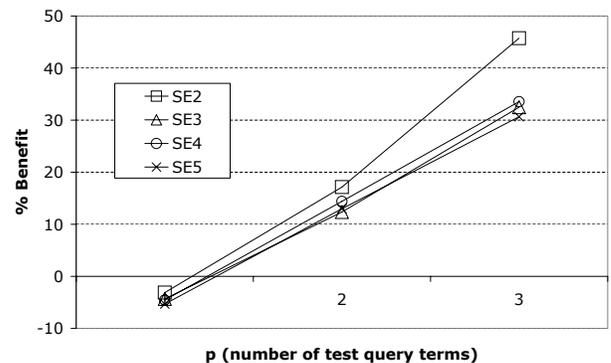
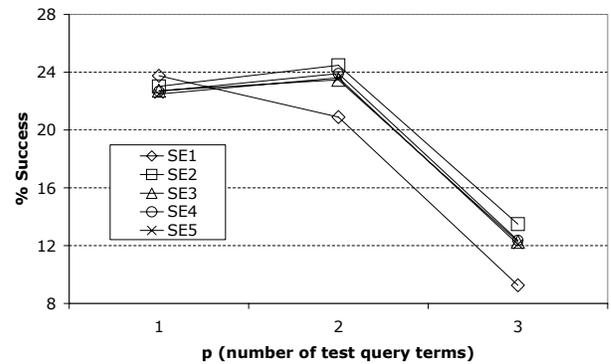


Figure 7: (a) Search success rates and (b) average benefit relative to SE1, for varying test query sizes.

tunities for false reminders in the face of vague single-term queries. However, for $p > 1$ we find SE2-SE5 outperforming SE1, with SE2, for example, achieving a relative benefit of almost 45% for $p = 3$. Nevertheless it is worth noting how the success rate of SE1 falls consistently as p increases, with a significant success rate drop (from just over 20% to less than 10%) as the number of test query terms increases from 2 to 3. In contrast, the success rates of the enriched search engines remains stable for 2-term queries (actually their success rates increase slightly) but also suffer from a significant drop for 3-term queries, although they grow their advantage relative to SE1.

A Realistic Configuration

The above experiments have attempted to cover a wide range of parameter settings in order to investigate the relationship between different settings and the ability of the enrichment technique to improve index quality and search engine performance. In large part the averages reported reflect good and bad performance under different configurations. Of course, implementing this approach to mobile search means committing to a particular configuration setting for k , n , and r and in this section we consider a configuration which we argue is likely to be reasonable in practice. As such, selecting parameter settings that we know tend to work well in our experiments, we choose $k = 5$, $r = 10$, $n = 200$. Thus, each news story is enriched by using an enrichment query containing its top 5 terms to retrieve the top 10 results from Yahoo

and generate a list of 200 enrichment terms for use during indexing. In addition we compare the resulting search engine performance in the usual way when using 2 and 3-term test queries; given that, as previously mentioned, mobile queries contain between 2 and 3 terms on average, these are reasonable test query settings.

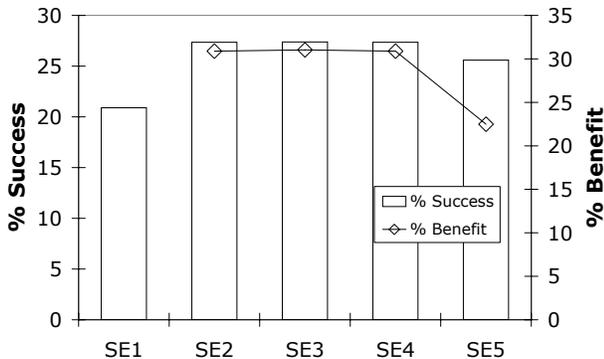


Figure 8: Search success rates and average benefit relative to SE1, for $k = 5$, $r = 10$, and $n = 200$ for 2-term ($q = 2$) test queries.

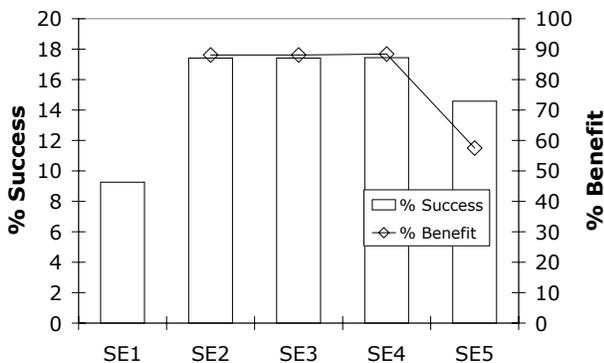


Figure 9: Search success rates and average benefit relative to SE1, for $k = 5$, $r = 10$, and $n = 200$ for 3-term ($q = 3$) test queries.

The results are presented in Figure 8 (for $q = 2$) and Figure 9 (for $q = 3$). In each case we see a significant advantage due to the use of enrichment techniques, with SE2-SE4, in particular, significantly out-performing SE1 by between 26% (for 2-term queries) and almost 90% for 3-term queries. In each set of results we note that the RSV method (SE5) performs significantly worse than the alternatives (SE2-SE4).

These results suggest that under practical configuration settings and realistic query conditions, the enrichment techniques have the potential to deliver significant improvements in search engine performance when compared to a baseline search engine that uses mobile content only for indexing.

CONCLUSIONS

Mobile search faces a number of significant challenges, some because of the inherent limitations of mobile devices, others because of the nature of the Mobile Internet itself. In this paper we have highlighted one issue that is often passed over in

discussion of mobile search and content discovery, namely that the brevity of mobile content offers a very limited indexing vocabulary for mobile search engines, which in turn limits future retrieval opportunities. In this paper we have described a novel approach to addressing this problem, by harnessing Web search engines as a resource for enriching mobile content. We have described a procedure for converting a piece of mobile content into a Web search query and a mechanism for extracting enrichment terms from the search results that can be used to index the document in question. We have also described a number of different ways in which enrichment terms can be scored for selection purposes and evaluated the performance of these enrichment strategies, relative to a benchmark mobile search engine, under a variety of experimental conditions. Overall we have found the results of this evaluation to be promising, with the enriched search engines generally outperforming the benchmark to a significant degree.

Of course this work is just one step forward and really provides a proof-of-concept of the enrichment approach. Further work is required in a number of areas. For sure, a more comprehensive evaluation needs to be conducted, preferably using live users or, at least, real user queries. In addition, the particular enrichment techniques described in this paper are relatively straightforward and there is room for improvement in a number of areas. For example, we have not focused at all on the different ways to produce an enrichment query from a piece of mobile content and there is no doubt that the simple frequency-based approach used here is unlikely to produce very high quality queries. As part of future work we will explore more sophisticated query generation techniques, drawing on related work on query elaboration in the information retrieval literature. In addition, further work is needed on the weighting of enrichment terms prior to selection. In this paper we have found high term frequency (SE2) to be the best indicator of enrichment term value, but it remains somewhat unclear why more sophisticated approaches such as TF.IDF and RSV did not perform better, as expected. One possible explanation is that we used too few Yahoo results in our experiments. Figure 6 in our results section shows a significant increase in the performance of RSV and TF.IDF as r increases. Therefore, perhaps setting the upper bound of r to just 50 results was too restrictive, thus having a negative impact on both the RSV and TF.IDF techniques. This is something we will explore as part of future work.

ACKNOWLEDGEMENTS

This material is based on works supported by the Science Foundation Ireland under Grant No. 03/IN.3/I361.

REFERENCES

1. B. Billerbeck, F. Scholer, H. E. Williams, and J. Zobel. Query Expansion using Associated Queries. In *CIKM '03: Proceedings of the 12th International Conference on Information and Knowledge Management*, pages 2–9. ACM Press, 2003.
2. R. Brandow, K. Mitze, and L. F. Rau. Automatic Condensation of Electronic Publications by Sentence Selection. *Information Processing and Management*, 31(5):675–685, 1995.

3. C. Buckley, G. Salton, J. Allan, and A. Singhal. Automatic Query Expansion Using SMART: TREC 3. In *Proceedings of the 3rd Text Retrieval Conference (TREC-3)*, 1994.
4. Cellular News. 2.5 Billion Mobile Phones In Use. <http://www.cellular-news.com/story/19223.php>, September 2006. Last Checked 28th November 2006.
5. H. Cui, J.-R. Wen, J.-Y. Nie, and W.-Y. Ma. Probabilistic Query Expansion using Query Logs. In *Proceedings of the 11th international conference on World Wide Web (WWW 02)*, pages 325–332, 2002.
6. S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman. Indexing by Latent Semantic Analysis. *Journal of the American Society of Information Science*, 41(6):391–407, 1990.
7. E. N. Efthimiadis and P. V. Biron. UCLA-Okapi at TREC-2: Query Expansion Experiments. In *Proceedings of the 2nd Text Retrieval Conference (TREC-2)*, pages 278–290, 1993.
8. L. Fitzpatrick and M. Dent. Automatic Feedback using Past Queries: Social Searching? In *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 97)*, pages 306–313. ACM Press, 1997.
9. J. Gonzalo, F. Verdejo, I. Chugur, and J. Cigarran. Indexing with WordNet Synsets can Improve Text Retrieval. In *Proceedings of the COLING/ACL '98 Workshop on Usage of WordNet for NLP*, pages 38–44, Montreal, Canada, 1998.
10. A. Hust, S. Klink, M. Junker, and A. Dengel. Query Reformulation in Collaborative Information Retrieval. In *Proceedings of the International Conference on Information and Knowledge Sharing, IKS 2002*, pages 95–100, November 2002.
11. Ipsos Insight. Mobile Phones Could Soon Rival the PC As Worlds Dominant Internet Platform. <http://www.ipsos-na.com/news/pressrelease.cfm?id=3049>, April 2006. Last Checked 28th November 2006.
12. M. Kamvar and S. Baluja. A Large Scale Study of Wireless Search Behavior: Google Mobile Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 06)*, pages 701–709. ACM Press, 2006.
13. R. Mihalcea and D. Moldovan. Semantic Indexing using WordNet Senses. In *Proceedings of the ACL-2000 Workshop on Recent Advances in Natural Language Processing and Information Retrieval*, pages 35–45. Association for Computational Linguistics, 2000.
14. MMA. Mobile Marketing Association Announces Mobile Search Study Key Findings. <http://www.primezone.com/newsroom/news.html?d=103374>, August 2006. Last Checked 28th November 2006.
15. T. B. Rajashekar and W. B. Croft. Combining Automatic and Manual Index Representations in Probabilistic Retrieval. *Journal of the American Society of Information Science*, 46(4):272–283, 1995.
16. S. E. Robertson. On Term Selection for Query Expansion. *Journal of Documentation*, 46(4):359–364, 1990.
17. S. E. Robertson and K. S. Jones. Relevance weighting of search terms. *Journal of the American Society for Information Science*, 27(3):129–146, 1976.
18. J. J. Rocchio. Relevance Feedback in Information Retrieval. *Gerard Salton, editor, The SMART Retrieval System - Experiments in Automatic Document Processing*, pages 313–323, 1971.
19. T. Sakai and K. Sparck-Jones. Generic Summaries for Indexing in Information Retrieval. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 01)*, pages 190–198. ACM Press, 2001.
20. G. Salton. *The SMART Retrieval System - Experiments in Automatic Document Processing*. Prentice-Hall, Inc., 1971.
21. G. Salton. Developments in Automatic Text Retrieval. *Science*, 253:974–979, 1991.
22. G. Salton and C. Buckley. Improving Retrieval Performance by Relevance Feedback. *Journal of the American Society for Information Science*, 41(4):288–297, 1990.
23. G. Salton and M. McGill. Introduction to Modern Information Retrieval. 1983.
24. A. Spink, D. Wolfram, M. B. J. Jansen, and T. Saracevic. Searching the Web: the Public and their Queries. *Journal of the American Society for Information Science*, 52(3):226–234, 2001.
25. E. Sumita and H. Iida. Information Retrieval using a Statistical Abstraction Method (in Japanese). In *Proceedings of the 3rd Annual Meeting of the Association for Natural Language Processing*, pages 353–356, 1997.
26. M. Wasson. Using Summaries in Document Retrieval. In *Proceedings of the ACL-02 Workshop on Automatic Summarization*, pages 27–44. Association for Computational Linguistics, 2002.
27. ZD Net. The Mobile Internet: Are we there yet? http://news.zdnet.com/2100-1035_22-6110100.html, August 2006. Last Checked 28th November 2006.