# INFORMATION TECHNOLOGY JOURNAL

# Query Suggestion Generation Methods for Mobile Phones

W. Aisha Banu, P. Sheik Abdul Khader and R. Shriram

B.S. Abdur Rahman University, GST Road, Vandalur, Chennai-48, Tamil Nadu, India

**Abstract:** The Internet is a powerful vehicle of communication. It is widely used by billions of users in their daily activities. The large content of information on the web is essentially useless unless this wealth of information can be discovered and consumed by other users. This study, focuses on the problem of information retrieval through mobile phones. As the web becomes more and more accessible through mobile phones, there is a need for better mechanisms for access of the content. The key problem is that apart from the limitations of the domain such as small screen size, limited processing capability etc., there is a need for increasing the relevance of the content shown to the users. One of the causes of the non-relevant information being retrieved is the query ambiguity. The relevance of information retrieved can be improved by the use of query suggestions. The aim is to narrow down the scope of the query. This study focuses on a unique method for improving the quality of suggestions generated for the user queries from the web. The effect of the query suggestions on increasing the relevance of the user queries has been outlined. The results show considerable promise and can potential lead to significant further research in the domain.

**Key words:** Information retrieval, mobile devices, query, suggestions, dependency tree

## INTRODUCTION

The growth in research in the domain of Information Retrieval has been fuelled largely by the web and internet. The domain of Web Information Retrieval focuses on the retrieval and ranking of the content from the web. An objective of the Web Information Retrieval system is to find the relevant content for the given input query term. Another important criterion is to find the content with the minimal number of accesses. Search Engines are used as the primary tools in searching the query term in the web. The explosive growth of wireless and mobile devices such as personal digital assistants and mobile phones has necessitated variants of the Search engines specially designed for these devices. The existing Information Retrieval techniques cannot be applied directly to web browsing through the mobile devices. So these methods (Hasany *et al.*, 2010; Niknafs and Band, 2010) must be modified to overcome the limitations of the mobile devices, such as small display size, low processing speed and un-reliable connectivity. The sheer volume of data available and the challenges described above can pose considerable hurdles for the usage of mobile devices in Information Retrieval applications as discussed by Shriram *et al.* (2008). Studies in Information Retrieval through mobile devices by various authors (Kamvar and Baluja, 2004, 2008; Yi *et al.*, 2008; Church *et al.*, 2008; Kamvar *et al.*, 2009) have thrown up the following interesting facts.

**Query suggestions:**

- Mobile users often modify their queries throughout their session in order to find exactly what they are looking for. Users would like to refine their query first rather than parsing the search results
- Users rely heavily on Query suggestions
- Users accept a correct suggestion quickly
- There is a need for adaptive suggestion generation system that is based on a thesaurus
- Mobile queries have a high rate of spelling errors

**Diversity of queries:**

- Mobile phone users have a low diversity of queries. The majority of queries can be categorized under common terms, such as entertainment, people, retail, sports, technology and travel categories
- Mobile users want quick answers as opposed to efficient ones
- A small set of queries are repeated many times in different forms. The proportion of identical queries and query overlap is very high

**Context of the query:**

- Search queries through mobile devices are shorter (Average: 2.2 Words to 2.5 Words) compared to the search queries through computers (Average: 2.94 Words). They are ambiguous and lack context

---

**Corresponding Author:** W. Aisha Banu, B.S. Abdur Rahman University, GST Road, Vandalur, Chennai-48, Tamil Nadu, India

- Search queries through mobile phones have a high proportion of local queries (based on location)

Thus we can conclude that use of suggestions in the mobile devices can improve the quality of search results. The key question is the method of generating the suggestion. The overall objective of the work is to develop a framework for improving the query results in a mobile environment. To achieve this major goal, several components have to be created, upon which the overall framework will be built. In order to create these building blocks, the specific objective of this study is to propose a method for query suggestion generation. The generation of suggestions using the results retrieved from the web is explored. The objective of the system is to generate a set of related content to the user. Based on the initial query term, the user is taken in search of the target query, which is based on the contents of the ontology or the contents of the web.

## STATE OF THE ART AND MOTIVATION FOR THE PRESENT WORK

A good deal of literature is available pertaining to Information Retrieval through mobile devices. In spite of the tremendous amount of publications in this area, it was felt that there is a need for research enhancement and development of algorithms that focus on search results re-ranking. Many research papers have reported delineating aspects such as search result clustering, user interface adaptation, use of ontologies, context awareness and query suggestions. The review of work in query suggestion is given in Table 1.

Due to the limitations of the domain, it is helpful for the system to narrow down the input word given by the users, before sending it to the search engine. A query mismatch problem frequently occurs as the exact needs of information for the query may be tough to judge due to the problem of synonymy and hypernymy. A query term can have multiple meanings (synonym). The key challenge here is to accurately identify the correct meaning of the query. Consider the term 'mobile phone' it can refer to the history of mobile phones, different brands of mobile phones, architecture and operation of mobile phones, etc. Each of these terms has a different set of content associated with it. Hence, the challenge here is to identify the proper content for the term quickly. Another problem is that multiple words can have the same meaning (hypernym). Thus, 'hand phone', 'mobile phone', 'handset' all have the same meaning, but different phrasal structures. Such words have to be mapped in terms of the relationships accurately, so that the appropriate meaning of each term is shown clearly. Query suggestions can help in this respect (Rajan and Rajagopalan, 2008). A genetic algorithm based approach for disambiguating the hypernyms and synonyms has been explained by Bergstrom *et al.* (2000).

The suggestions for the query process can come from the (a) Database dictionary, (b) Ontology, (c) Search results, (d) Past history and (e) Other users. There are various dimensions of context-the profile, location and temporal aspects that are relevant in a mobile domain. These dimensions of the context can be used to give suggestions to the users, as well as to filter the information available to the users. The context approaches (Lee, 2009) are used in two ways: location extraction and concept extraction. These context aspects can be used for providing valuable suggestions to the system for the search process. Location based methods (Mountain, 2007) can be used to eliminate the results not relevant to the user. Lane *et al.* (2010) outlined a context based location dependent system for search and Information Retrieval through mobile devices. Paek *et al.* (2009) have introduced a Phrase Builder and a Real Time Query Expansion interface. The Phrase Builder reduces the keystrokes by facilitating the selection of individual words in addition to whole phrases and by leveraging back-off query techniques to offer suggestions for out-of-index queries. Banu *et al.* (2011) focused on the use of ontology for mobile phones for generating the suggestions. In Wu (2008) XML techniques have been used in the design of Personal Information systems.

Table 1: Literature status

| Status | References |
| --- | --- |
| **Information retrieval literature relevant to queries** | |
| Context based location dependent system for search | Lane *et al.* (2010) |
| A Real time query expansion interface | Paek *et al.* (2009) |
| Patterns for relation extraction from web-texts | Bergstrom *et al.* (2000) |
| Transforming user clicks into conceptual interests | Lee (2009) |
| Spatial filters for mobile Information Retrieval | Mountain (2007) |
| Use of ontologies for Mobile phones | Banu *et al.* 2011) |
| XML techniques | Wu (2008) |
| Indexing method | Agbele *et al.* (2011) |
| Vector space models | Prasannakumari (2010) |
| Ambiguity | Rajan and Rajagopalan (2008) |
| Information scent | Bedi and Chawla (2010) |
| **Ontologies** | |
| Location information | Gao *et al.* (2011) |

Using the query log of the user and reusing the content already used, is a form of predictive proxy caching. The method relies on the user's being consistent in their information needs. The problem will be when the user has different contexts. The method by Agbele *et al.* (2011) uses a term frequency method and corpus indexing for SMS queries. Similarly, Bani-Ahmad and Al-Dweik (2011) have focused on the use of term frequency approaches for the digital libraries. This approach can work where there is a corpus documents. This is not possible for all dynamic queries as in our application. The work on multimedia databases uses vector space indexing (Prasannakumari, 2010). The vector space indexing method will be a part of our future work. Erba *et al.* (2011) have proposed a method that uses explicit gain measures for the information. A variant of this approach is used in our work in that the query personalization in effect is a function of the users personal preference. Hence the gain that the users will get is necessarily a function of this. However, the measures of explicit gain used in this work will be a part of our proposed work to make this correlation explicit. The information scent based approach is a new algorithm proposed by Bedi and Chawla (2010). The work is a variant of the frequency measures for personalizing the search. Gao *et al.* (2011) have proposed the use of ontologies in information retrieval for identifying the location information.

The same query may necessitate different results depending on the context. The Word-net can be used for generating the query suggestions. This approach is suitable when an intermediate server, is used as the amount of information that can be generated will be exponential. The personal profile can guide the context of the users. This work generates query suggestions using an algorithm that trawls the search engine. These suggestions can also be added to the ontology. The Word-Net is not used for the query suggestions in this system as it generates a large amount of data and takes a lot of time for processing. However, the use of word-net will be a part of the future work. In addition, the use of predictive text systems will be explored as part of the future work.

## MATERIALS AND METHODS

The major objective of this work is to improve the efficiency of the search process in a mobile device environment through the query suggestions. This method is called expanded mode. Expanded mode relies on the web for generating the suggestions. The operation of this method is depicted in Fig. 1. The idea here is that, a process by which the content terms are extracted from the
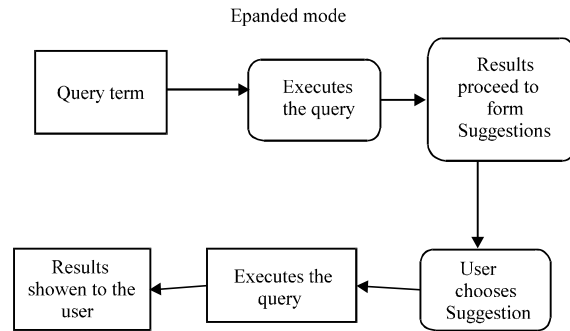


Fig. 1: Web based query suggestions

web search results to form categories is far preferred by the users than a complex meaning extraction stage. This stems from the literature, which suggests that suggestions must focus on categories. Hence, the expanded mode is used. Once the user finalizes the query, the search results are retrieved and shown.

The objective of the overall system is to cluster the content in the snippets retrieved from the search engine and generate semantic suggestions for the expanded context mode. The outcome is a set of clustered words that are related to each other. The contents retrieved are aligned with reference to the ontological structure wherever possible and the relationships identified and are generated. There are two stages in this: Preprocessing and Dependency tree modeling. In the preprocessing stage a tokenizer and stop word remover are used. The stop words list is a customized set of unwanted words, which are commonly present in snippets. Preprocessing reduces the size and number of the input documents considerably, which is essential in any information retrieval system. Preprocessing removes all types of stop words, special characters, extensions, etc., to reduce the processing overhead created by including the stop words into the system's preprocessing framework. The lexical analysis is used to divide a stream of characters into a stream of words. In the dependency tree modeling stage, the contents are initially modeled in the form of $S_{ij}$, where a single $S_{i1}$ represents the first word of the snippet $S_i$. Thus, $S_{11}$ represents the first term of the first snippet. Now, the query represents the root. The first term $S_{11}$ is aligned with reference to the query. If there are any relationship exists (meaning, relevant, related), they are modeled first (for example term t and the connector relation). If no relationships exist, then a new connector is created (new 1, 2, 3...) and the term is placed in order. Now, for every subsequent term the alignment is done accordingly, namely, if there is an existing relationship in the ontology, then the terms are aligned in line with these
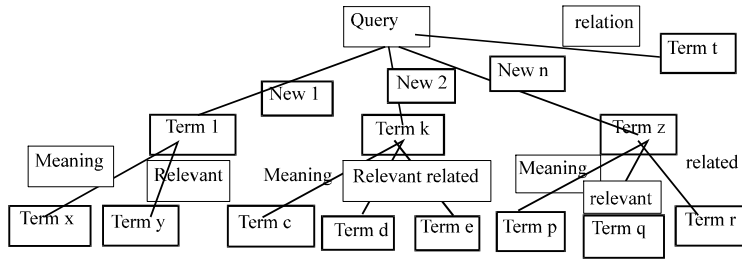
Fig. 2: Query dependency tree-priority order-balanced mode

---

•**Vegetarianism-Wikipedia, the free encyclopedia**
en.wikipedia.org/wiki/Vegetarianism
Semi-vegetarian diets consist largely of vegetarian foods, but may include fish or poultry, or other meats on an infrequent basis. Those with diets containing fish ...

C**Vegetarian cuisine-Wikipedia, the free encyclopedia**
en.wikipedia.org/wiki/Vegetarian_cuisine
A variety of vegetarian food ingredients that are also vegan.Vegetarian...
C**List of vegetarians - Wikipedia, the free encyclopedia**
en.wikipedia.org/wiki/List_of_vegetarians
This is a list of notable people who have claimed or been reported to have ...

---

Fig. 3: Three Snippets

---

$S_1$:  Semi-vegetarian diets vegetarianfoodsfish poultry other meats diets fish ...
$S_2$:  vegetarian food ingredients vegan. Vegetarian...
$S_3$:  notable people ...

---

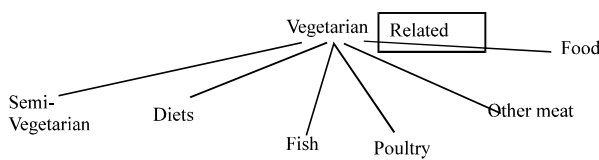Fig. 4: Stop word removed snippets for 'Vegetarian'



Fig. 5: Dependency tree expansion for snippet $S_1$ in balanced mode
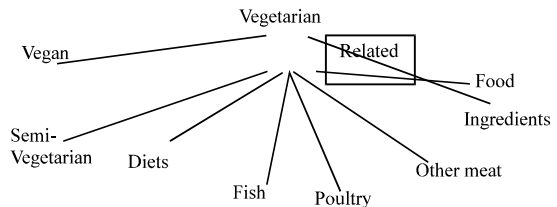


Fig. 6: Dependency tree expansion for snippet $S_1$ and $S_2$ in balanced mode

relationships. The outcome, thus, is a set of clustered terms. The cluster head is chosen as the term which is most strongly connected. Over time as the ontology can become very large techniques like those proposed by Hoque and Avery (2010) can be used. The operation is shown below. The query term is vegetarian and the ontology structure used above in Fig. 2 is used as the reference. For sample, 3 snippets are taken.

The snippets are shown in Fig. 3.

The snippet contents, after stop word removal are show in Fig. 4.

The tree structure for Snippet $S_1$ is shown in Fig. 5.

The tree structure for Snippets $S_1$, $S_2$ is shown in Fig. 6. The tree structure for Snippets $S_1$, $S_2$ and $S_3$ is shown in Fig. 7.

Now, the suggestions are generated as (a) Food, (b) Semi-vegetarian and (c) in the order of the Snippet generation. Thus, the first preference is for a connected structure. The next preference is for the order of terms in the order of frequency. The last preference is for the order of the snippets.
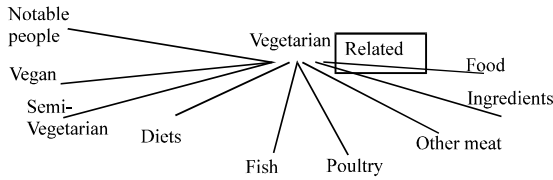
Fig. 7: Dependency tree expansion for snippet $S_1$, $S_2$ and $S_3$ in balanced mode



Fig. 8: Implementation scenario-expanded context mode

## PROTOTYPE IMPLEMENTATION AND RESULTS

The proof of concept prototype was implemented in J2ME. J2ME was selected as the platform of choice as it can be deployed on the prevalent models of mobile devices easily, is widespread and contains a rich library of routines that can be applied for various functions. The Mobile Information Device Profile (MIDP) was used for the functional aspects. MIDP 2.0 was the version used. Figure 8 shows the implementation of the expanded context mode. The suggestions for this mode are obtained from the web and shown to the user. The user can refine the query using these suggestions.

## TESTING AND RESULTS

The prototype application was deployed in different Java enabled mobile devices. The system was tested in

Table 2: Suggestion model-balanced mode

| User | Number of times suggestions generated | |
|---|---|---|
| | Balanced | Used |
| 1 | 1 | No |
| 2 | 1 | Yes |
| 3 | 1 | Yes |
| 4 | 1 | Yes |
| 5 | 2 | No |
| 6 | 2 | Yes |
| 7 | 1 | Yes |
| 8 | 2 | No |
| 9 | 2 | Yes |
| 10 | 1 | Yes |
| 11 | 1 | Yes |
| 12 | 1 | No |
| 13 | 1 | No |
| 14 | 2 | Yes |
| 15 | 1 | Yes |
| 16 | 2 | Yes |
| 17 | 1 | Yes |
| 18 | 1 | Yes |
| 19 | 2 | Yes |
| 20 | 19 | Yes |
| Average | 2.25 | Used | No |
| | | 15 | 5 |

March 2011. The browsing experience of 40 users in the age group of 15 to 40 with twenty mobile devices for a browsing period of 5 to 30 min, was benchmarked. The users were trained in the use of the systems and asked to enter queries of their choice. The results are summarized below. The data was entered by the users in a coding sheet.

Users were assigned at random (twenty users each) and the experiences of web browsing and the proposed system, were benchmarked. Volunteers were assigned to help the users. The users were free to type their own queries and could enter any query. They had to note down the data for each parameter. The overall aim of the experimentation was to observe the data and evaluate the precision, time taken, etc and validate the premises. Seventy percent of the users of the users used mobile devices for browsing regularly. The users were Graduate and Post Graduate students of Engineering. The users were knowledgeable in the process of browsing the content from mobile devices. The users were given the option of browsing the content through mobile browsers. The same users were given this prototype and their responses were tabulated. Research hypothesis was framed to validate the work. The discussions of the research hypothesis are given below. The suggestions are generated by the balanced modes. If the suggestions are used by the users, then it means that the system does not have to go to the web again for searching the contents. Thus, more the usages of the query suggestions better the relevance. The hypothesis tested was "The usage of the suggestions by the users in processing improves the

system." The efficiency of the suggestions generated by the Balanced mode was validated next in Table 2. It was found that 15 users used the suggestions and 5 did not. The average number of times the suggestions were generated was found to be 2.25.

The above result shows that the user has used the suggestions given by the query suggestion system. The more the suggestions are accepted by the system, the better its efficiency. Hence, this validates the hypothesis. These points to the direction at the future research, in that advanced methods for query suggestions can be explored.

## CONCLUSION

The study offers fresh motivation for algorithm to generate the query suggestions in the context of information retrieval for mobile devices. This work has proposed a coherent strategy for the problem. The use of query suggestions has improved the operation of the system and delivered effective results. A unique method of by which query suggestions can be generated from the web is outlined. There are many ways in which this research can be taken forward. The Word-net ontology can be used for generating the query suggestions. This can be implemented in tandem with this module. This will involve design of efficient query processing systems and an interface from the server side as well. This will be a part of our future work.

## REFERENCES

Agbele, K., A. Adesina, N .A. Azeez, A. Abidoye and R. Febba, 2011. A novel document ranking algorithm that supports mobile healthcare information access effectiveness. Res. J. Inf. Technol., 3: 153-166.

Bani-Ahmad, S. and G. Al-Dweik, 2011. A new term-ranking approach that supports improved searching in literature digital libraries. Res. J. Inform. Technol., 3: 44-52.

Banu, W.A., P.S.A. Khader and R. Shriram, 2011. Information retrieval through mobile devices using semantic ontology. Inf. Technol. J., 10: 1747-1753.

Bedi, P. and S. Chawla, 2010. Agent based information retrieval system using information scent. J. Artif. Intell., 3: 220-238.

Bergstrom, A., P. Jaksetic and P. Nordin, 2000. Enhancing information retrieval by automatic acquisition of textual relations using genetic programming. Proceedings of the International Conference on Intelligent User Interfaces, January 9-12, 2000, ACM Press, pp: 29-32.

Church, K., K. Bradley, P. Cotter and B. Smyth, 2008. A large scale analysis of European mobile search behaviour. Proceedings of the 10th International Conference on Mobile Human Computer Interaction, September 02-05, 2008, ACM Press, Amsterdam.

Erba, F.G., Z. Yu and L. Ting, 2011. Using explicit measures to quantify the potential for personalizing search. Res. J. Inform. Technol., 3: 24-34.

Gao, H., S. Wang and H. Lu, 2011. Local positioning systems for mobile devices based on ontology. Inform. Technol. J., 10: 168-174.

Hasany, N., A.B. Jantan, M.H.B. Selamat and M.I. Saripan, 2010. Querying ontology using keywords and quantitative restriction phrases. Inform. Technol. J., 9: 67-78.

Hoque, M.T. and V.M. Avery, 2010. Novel strategies to speed-up query response. Res. J. Inform. Technol., 2: 11-20.

Kamvar, M. and S. Baluja, 2004. The role of context in query input: Using contextual signals to complete queries on mobile devices. Proceedings of the 6th International Conference on Mobile Human Computer Interaction, September 13-16, 2004, Glasgow, UK., pp: 180-191.

Kamvar, M. and S. Baluja, 2008. Query suggestions for mobile search: Understanding usage patterns. Proceedings of the SIGCHI conference on Human Factors in computing systems, April 5-10, 2008, Florence, Italy.

Kamvar, M., M. Kellar, R. Patel and Y. Xu, 2009. Computers and iphones and mobile phones, oh my! A logs-based comparison of search users on different devices. Proceedings of the 18th International World Wide Web Conference, April 20-24, 2009, Madrid, Spain, pp: 801-810.

Lane, N.D., D. Lymberopoulos, F. Zhao and A.T. Campbell, 2010. Hapori: Context-based local search for mobile phones using community behavioral modeling and similarity. Proceeding of the 12th International Conference on Ubiquitous Computing, September 26-29, 2010, New York, USA, pp: 109-118.

Lee, D.L., 2009. User profiling and personalized information delivery on the static and mobile web. Proceedings of the 11th ACM International Workshop on Web Information and Data Management, November 2, 2009, Hong Kong, China.

Mountain, D.M., 2007. Spatial filters for mobile information retrieval. Proceedings of the ACM 16th Conference on Information and Knowledge Management, November 6-9, 2007, Lisbon Portugal.

Niknafs, A.A. and H.B. Band, 2010. Improved win-win quiescent point algorithm: A recommender system approach. J. Applied Sci., 10: 3084-3090.

Paek, T., B. Lee and B. Thiesson, 2009. Designing phrase builder: A mobile real-time query expansion interface. Proceedings of the 11th International Conference on Human-Computer Interaction, September 15-18, 2009, Bonn, Germany.

Prasannakumari, V., 2010. Contextual information retrieval for multi-media databases with learning by feedback using vector space model. Asian J. Inform. Manage., 4: 12-18.

Rajan, M.S. and S.P. Rajagopalan, 2008. Effective information retrieval using supervised learning approach. Inform. Technol. J., 7: 231-233.

Shriram, R., V. Sugumaran and K. Vivekanandan, 2008. A middleware for information processing in mobile computing platforms. Int. J. Mob. Comm., 6: 646-666.

Wu, C.F., 2008. Design of portable personal information management system with XML technique. Inform. Technol. J., 7: 615-622.

Yi, G., F. Maghoul and J. Pedersen, 2008. Deciphering mobile search patterns: A study of Yahoo! mobile search queries. Proceedings of the 17th International World Wide Web Conference, April 21-25, 2008, Beijing, China.