A Query-Based SMS Translation in Information Access System

Ademola O. Adesina, Kehinde K. Agbele, Nureni A. Azeez, Ademola P. Abidoye

Abstract*— Mobile technology has contributed to the evolution of several media of communication such as chats, emails and short message service (SMS) text. This has significantly influenced the traditional standard way of expressing views from letter writing to a high-tech form of expression known as texting language. In this paper we investigated building a mobile information access system based on SMS queries. The difficulties with SMS communication were explored in terms of the informal communication passage and the associated difficulty in searching and retrieving results from an SMS-based web search engine under its non-standardization. The query is a pre-defined phrase-based translated English version of the SMS. The SMS machine tool normalization algorithm (SCORE) was invented for the query to interface with the best ranked and highly optimized results in the search engine. Our results, when compared with a number of open sources SMS translators gave a better and robust performance of translation of the normalized SMS.

Index terms – Short messaging service (SMS), spell checker, spell corrector, spell error

I. INTRODUCTION

Language usage over the computer-mediated discourses differ significantly from the language of communication from person-to-person or tribe-to-tribe, the same applies to information retrieval mechanism. Few years ago, the primary technologies used to construct large information systems were database systems, information retrieval systems, and information filtering approaches. Database systems were used to handle large volumes of structured data and to provide guarantees of reliability and consistency despite systems failures and high volumes of update transactions. Information retrieval systems were used to search large databases of text, such as encyclopedia materials with hundreds of thousands of articles, biographies, videos, abstracts, legal materials, or newspaper stories. Information filtering approaches provide periodic updates in the form of text stories, mostly in the business domain, based on user profiles and it is always specific. Relatively for this short period, there have been many developments that have affected how information technology is talked about and used nowadays, the growth of the Internet and the availability of cheap hardware are among these factors influencing the technologies for the large information systems [1].

The spread of Internet and mobile devices have brought about the web and mobile search retrieval mechanism respectively to compliment the traditional retrieval system and they all follow completely different path in the way of their execution [2]. Mostly the traditional retrieval system may include the application of natural language understanding which involves the extensive knowledge of the outside world and the ability to establish the context awareness of the intention of what user is actually being referred to. The web search retrieval mechanism is based on the use of search engines to retrieve information but the quality of retrieved information varies from low to high, in terms of relevance to the users' needs [3].

Illustration of the request for "*University of the Western Cape*" in some major search engines like *Google Search, Yahoo search, Alta vista* and *Alltheweb* gave different results in the ranking position even with the same query. This query was later re-arranged and different results emerged from the former search. The execution time of returns of results for each query supplied gave a great difference in terms of results generated. It is important at this stage to confirm the best web search information retrieval mechanism through a standard metrics of measurement, in order to deduce the most efficient algorithm among the search engines besides the precision and recall or hit and misses being used until now [4].

The results of the search engines, in most cases, are influenced by text transformation such as stemming, for instance, the arrangement of the words "*medical", "medicine", "medicinal", "medically*" return different answers under different timing. With this, there is no guarantee of the correctness of the information on the web, because different results are the function of the type of search engines used for the conduct of the query.

Mobile search is a reflection of the development of mobile technology with difference in the way of giving immediate response to a request. It is mostly available in a pre-defined field like navigational query, entertainment-related query, music, or ringing tone, sports, local knowledge, shopping, and references. Other major consideration of the mobile search engine is the size available which may not allow much navigation of the web pages to be achieved especially for interactive search process, improbable, so only a single SMS should be returned, for any given query. As users will be forced to rephrase or reformulate the query if their answers are not made available in the preliminary pages of the mobile web search;

Manuscript Received October 09, 2011.

Ademola O. Adesina, Department of Computer Science, University of the Western Cape, Private Bag X17, Bellville 7535, South Africa, (inadesina@gmail.com)
[2]Department of Computer Science, Lagos State University PMB 001 Ojo, Lagos Nigeria
Agbele K. Kehinde, Department of Computer Science, University of the Western Cape, Private Bag X17, Bellville 7535, South Africa, (agbelek@yahoo.com)
Azeez N. Ayofe, Department of Computer Science, University of the Western Cape, Private Bag X17, Bellville 7535, South Africa, (nurayhn1@gmail.com)
Ademola P. Abidoye, Department of Computer Science, University of the Western Cape, Private Bag X17, Bellville 7535, South Africa (ademolaabidoye@gmail.com)
Department of Computer Science, Lagos State University PMB 001 Ojo, Lagos Nigeria.

unlike in desktop where there is possibility of accommodating ten or more results per page. This method will be disadvantageous to the newly created pages, making popular pages even more frequently visited and new pages completely ignored; as user gives preference to the front pages during web information searching.

On the other hand, mobile web-based keywords are fewer than their desktop-based counterparts. For instance, in the data provided by *Massachusetts-based* mobile search solution provider *JumpTap*, it suggests that the average desktop search contains something between two and three keywords, whereas less than 15% of all searches carried out from mobile devices contain more than two keywords. This has a great effect on the publishing as the content is highly focused, relevant and preferably identifiable by a single keyword [5].

Contextual SMS search is different from traditional contextual desktop search in many ways because of the absence of web page in the former. Web page is very good for reference context as it is inferred automatically. However, there is no advanced search features in SMS search as it is completely neglected by the desktop search users because the users prefer to be taking extra energy in reformulating queries.[6-7]

The characteristic features of the mobile devices have brought about the unique ways of communication that is referred to as “”texting” or “SMSing”. This has become the language of the youths, notably in its usage is the restriction in the number of characters/words engaged by the user to communicate because of the memory space, pad structures of the design and writing skills of the query and bandwidth issue [8]. Its construction is based on convenience of spelling and homophony of the wordings. Regardless of the range of the handsets (*low*, *medium* or *high end)*; SMS stands as the only mode of text communication, therefore service providers have been encouraged to build information based services around this technology. For example, translation of SMS text into different languages so as to communicate with a foreign person with little or no considering of the recipient mother language [9]. The service of the third party provider are sometime needed in translating the languages into other languages but it has the short comings of the high cost to the user, limited message space and users unfriendliness [9]. SMS provides convenient platform where messages are delivered even when the recipient is engaged in voice communication, in a lousy or noisy environment or when the recipient is busy and unable to attend to a call.

In the current dispensation, the invention of the computer machine has brought about computer mediated languages like *emails, SMS, chats etc*. The major factors for the choice of SMS are ease of writing, phonological articulation- as the words sound during pronunciations, constrain in memory space and the design of the keyboards. The number of strikes that will be needed to input “t” in the 14 keypads created handsets will be 4 as compared to the QWERT keyboard structured handsets. SMS language is considered as a technological generated language spreading round even to the area of games, where competition is set for fastest writing text users.

The emergence of computer-mediated communication necessitated the need for machine translations to convert from one (source) language to another (target) language. The input data in the text generation system can be taken to be the source language in the machine translation, the method is the translation process, and the output from the machine translation is the target language. The major challenge in this chain of translation is the generation, understanding of the result of the translation (in terms of the context), and the statistical natural language processing. Statistical modelling using stochastic or probabilistic methods are common ways to resolving the dissimilarities in language translation. This is applicable to many specific pair of languages regardless of the linguistic rules, but the natural language processing (NLP) may involve the use of grammar or rule-based translations systems that will require manual creation of linguistic or grammatical rules. The short coming of this approach is that it is both costly and restricted to the pair of languages it is meant to translate. This has made the statistical modelling to stand the test of time in its application to the solving the problems of language translation in text generation [10].

The rest of the paper is structured as follows: section II talks about the background and the motivation for the SMS translation as coursed by language shift created by the computer-mediated paradigm. Section III discusses the related works on spelling, grammar checker and phrases selection in order to better express the context of the subject. Section IV enumerates the methodology adopted in collection the SMS collection for the research work as well as the processing techniques used for the corpus. Section V delves into the particular algorithm used in the system architecture and design. Section VI describes the experiment testing and its evaluation. Sections VII and VIII discuss the results and conclusion respectively.

II. BACKGROUND AND MOTIVATION

The invention of mobile technologies has moved the traditional desktop computer and laptops toward more portable and easily accessible handheld devices like cell-phones. This paradigm shift did not neglect the computational capability of the devices as most roles played are incorporated. One of these features is web browsing, web searching, emailing, to mention a few. Searching using mobile devices are completely dissimilar to desktop search because relevant results are returned immediately on the first page using simple search interface. The results gathered in the first instance determine whether the query will have to be interactively refined again from the suggestion of the search engine or using the feedback from the collections of results that were gathered from the initial search. Effective performance of any search engine is a function of its indexing in other to return the appropriate documents during searching. Smartphone have capacity to do iterative search process repeatedly until the best result is generated but the majority of mobile devices are using SMS for search with their cell phones.

A part of the challenge faced with the use of SMS-based search is the cost incurred on the bandwidth and the sketchy exploration by the mobile users of the relevant web pages that are returned. This is different from the traditional search engine in the course of finding the information needed by the users. The information are not explicit enough due to little spaces available for the sentences that are returned as the results of the enquiries.

For this reason, there is adaptability of the mobile device to handle the type of enquiries like finding the addresses, phone numbers, directions, dates, prices, and short phrases that serves as information or coded words. These may not need a whole web page to be able to satisfy the requests.

Our concern in this paper is to develop an SMS-based query format that will be able to access the database and return the best options/results for the enquiry. The way queries are raised has significant influence on the reports generated by the search engine. There are some questions that were prepared in English language for about 150 students of the Computer Science Department, at the University of the Western Cape, South Africa [14]. At this stage it becomes important to note that web searching is incomparable with making enquiries based on SMS in mobile devices. This may be because of the following factors: restriction in the number of letters used by the user to communicate, limitation created by the phone screen size, striking of the keypads and the language choice of the users.

We have been able to develop an algorithm that can take all questions of the mobile users that were used in the course of our research and converge them to what they are meant to stand for in their original English version. This has a major advantage in the parsing of the tokens of the questions. SMS is a computer-mediated language cropped up by the society and level of high technology, therefore as a language there seems to be language dynamism. Consequently, there will always be a need to create room for expansion or upgrade for the enquiries hence; the algorithm gives room for such in our query-based SMS formulation.

Based on the statistical approach and the observation, we shall look at the query-based and SMS translations to bring about the retrieval of the documents in a mobile search engine. The pre-defined queries are phase-based formatted and their translations are the keyword generated model. This will be used for the information extraction in the database to provide results to the queries. It should be noted that the expansion model for the keyword is considered for the upgrading of the keywords. Different form of expressing queries under Frequently Asked Questions (FAQ) were constructed in the database and the provision for expansion in case of new texting languages coming up in the future were considered.

III. SPELLING CHECKER AND CORRECTIONS

Autocorrect, as part of several systems are designed to correct common typing or spelling errors automatically, changing two initial capitals, capitalizing both the first letter of a day and the first letter of a sentence. Several systems can analyse a text file for potential spelling errors, pointing out probable correct spellings. Spelling programs are expected to be a standard part of all text processing systems. Several companies, including Microsoft, have been working on two types of spelling programs: *spelling checkers* and *spelling correctors.* In the former, the input file of text find those words which are incorrect while the latter detects the misspelled words and tries to find the best substitute from the range of correct choices of words provided. This problem contains elements of pattern recognition and coding theory [11]

Spelling errors can be introduced in word processing in many ways either by *user's deliberate* or *in-deliberate attitude.* Such errors can lead to consistent misspellings and are probably related to the difference between how a word sounds and how it is actually spelled, for example *in4mation* for *information*. Typographical errors on typing are not very common in long essays but their occurrence are not ruled out, because of the position of keys on the keypad or keyboard that may generate the errors due to the finger movement [12]. An interactive spelling checker can also be helpful in correction. In the simplest case, the checker remembers all tokens which the user has indicated "should be replaced" and the words with which the tokens are to be replaced. After the first such replacement, future occurrences of the misspelled token can be automatically corrected, although this feature might be undesirable in some instances.

A similar approach is including common misspellings in the dictionary with the correct spelling. Thus, *greatfully* would be tagged as a misspelling of *gratefully* and *prehaps* as a misspelling of *perhaps.* This approach has not been included in any current spellers, probably because of the lack of an obvious source of known misspellings and the low frequency of even common misspellings. Most misspellings can be generated from correct spellings by a few simple rules. Damerau [13] indicates that 80 percent of all spelling errors are as a result of transposition of two letters, one letter extra, one letter missing and one letter wrong.

Error types are many hence there are different algorithms to solving each. Effort of this research is to resolve or reduce the orthographic nature of the SMS messages in text entry; mostly *word processor* use *string manipulation or matching technique* to resolve this much early work at finding and correcting errors brought about by specific input devices in specific contexts. For instance, the concern may be to detect the potential of misspelled names of passengers for a specific airline flight, hotel reservation, population census registration, examination registration, and banking transaction as either the stored or inquiry name (or both) might be misspelled.

The era of integration of computer into technology is fast fading away as highly technological, modernized cell-phones have been designed to integrate the roles of computer, and this is fast becoming an integral part of the society. Most cell phones users do not create time to read or study the *user's manual* that come with the cell-phones but rely on their peers for instructions. In making appreciable spelling error correction many academic studies on the general problem of string matching and correction algorithms have been conducted. In spite of this, none was aimed at producing a working spelling program for general text. however, several spelling checkers have been written for the sole purpose of checking text [11].

The uniqueness of SMS is the strong deviation it has from the formal language or normal spelling for example, a phrase like "what is" can have over ten SMS versions like "*Wat is", "Wats", "Watz", Whats", "Whts", "Wots", "Wt s", "Wt's", "Wts", "Wtz", "Wht is", "Wat's"* [14]. Autocorrect fails to recognize some of these words as they are not included in the dictionary of the open source. They completely deviated from the standard of the English language. The spelling checkers see almost all the SMS writing as being mistakes.

But this is not considered as mistake by texting language users as they are able to understand this type of communication. This then means there is a complete communication passage of information across between the source and the target. There is a great advantage of spelling and grammar checkers as it could help users' correct spelling and grammar errors conversely the tools may not be relevant in organizing and composing any article from the beginning.

IV. METHODOLOGY

A. SMS Corpus Collection

Short Messaging Service (SMS) messages were collected from a group of first year computer science students totalling 150 and a set of 25 questions were administered. The SMS communication appears in different forms because of flexibilities associated with this form of communication. A set of predefined questions related to HIV/AIDS were provided for all participants. The participants were then required to re-write the same questions, assuming they were personally sending the same question via an SMS technique. Three different methods used in collecting question data sets: the use of the MXit platform, transfer of data via blue-tooth and hand-written on paper. MXit [15] is a free, instant messaging application developed in South Africa that runs on 3G General Packet Radio Service (GPRS/3G) mobile phones and on PCs. This technology allows the user to send and receive one-on-one text and multimedia messages to and from other users, as well as in general chat rooms. For all the methods used, laptop was configured to serve as database server. It received all forms of text messages from the participants based on their responses.

Cell phones that have capacity for data transfer via blue tooth or infra red were used for the data transfer. The participants saved the SMS as a draft in their handsets, later transfer to a dedicated cell phone through Bluetooth or Infra red. From this, the text messages are deposited into the database server. The FAQs were written in hard copy and the participants were asked to write the SMS languages' equivalent. This was later transferred into the database server. This approach was widely acceptable because it was considered very fast to complete as no one struggled with the key-pad or the tiny screen of the cell phones.

B Pre-Processing the Corpus

After the end of the data collection there was a partial parsing by the use of the spelling checkers to confirm the regularity of the typographic errors, ungrammatical abbreviations, acronyms, colloquial words, duplicated words and so on. The in-depth study of the SMS collected showed the great variability of the words form. Getting a clear meaning of the words in the corpus may be difficult to the extent that one has to do some guess work in order to confirm the correct interpretations of the words that are under search. This perhaps makes the reading difficult or impossible-because of lack of vowels (e.g. 'nyt' for 'night') or many consonants combining (e.g. 'trtmnt' for 'treatment'), even sometimes diphthong (e.g. 'ao' for 'how') appearing in the SMS formation.

The query readability is another major phenomenon that needs to be taken into consideration as some words for example, 'ask4trtmnt' meaning 'ask for treatment' are written without white space or delimiter, the upper and lower cases are sometime mixed together e.g. 'disCHg' for 'discharge' and mixing letters and figure together of some words like '8s' for 'AIDS' and '4rm' for 'from' are languages that one needs to understand the codes, usages and habits for perfect translation of the SMS language. Otherwise, this may have been taken as errors unless one is familiar with the usage to avoid misrepresentation.

It is imperative that the corpus be translated into structured language in order to facilitate the exploration of the messages. With this, the meaning can come forth, as the parsing of the unstructured SMS language to a more structured for instance, English language. Without the structured list it is difficult to get the utterance of a given grammar, but it becomes very easy to find the translation of the corresponding SMS.

V. ALGORITHM OF THE SCORE ARCHITECTURE

In Fig. 1, the architecture of the system consists of a server connected to the Internet and to a mobile phone. The client is a user with a mobile phone who sends an SMS message to the server which then processes the search query. The search query is dispatched to a general search engine and result pages are downloaded. The server extracts the results from the downloaded pages, and distils them down to 140 bytes maximum because of the capacity of the mobile phone and the bandwidth measurements. Finally, the server returns the results to the number that issued the request. The extraction process at the server is the heart of SMS-query.

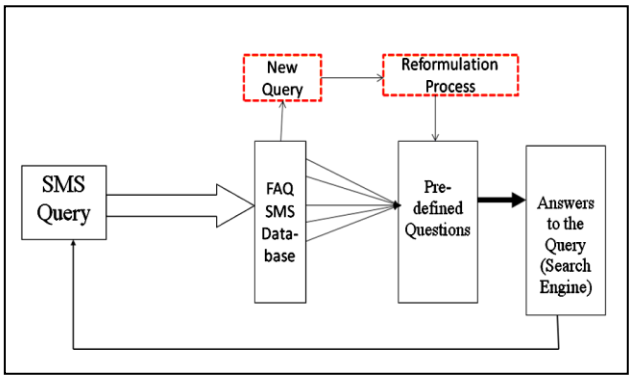

Fig. 1 System architecture of an SMS-Query reformulation process

There is provision for reformulation of the query in case it is not available in the FAQ SMS database and it is updated to be part of the pre-defined questions as this serves as an area of supervised learning for the system architecture.

The input to the system is a search query, in form of request where the query represents the actual search terms and the context specifies the type of contextual information that the user expects the system to extract. During the extraction process, the system can gather results in the form of N-grams from a corpus of web pages, where an N-gram is simply any set of N space delimited terms found within those web pages. The N-grams are measured according to several metrics and then ranked. The most highly ranked result is then returned to the user. The query is sent into the centralised server and the string matching is performed within the database server and then the best option is picked.

This is sent back through the network under the short message service centre (SMSC). Introducing the translator at the user's end will make it to be unfriendly; the users will from time to time interprete the SMS language even after that answer has been provided. This particularly will have drawback effect on the performance of the mobile device in the following area:

1. High cost to the user because of the cost of harbouring the intelligent software/dictionary to do the frequent translation
2. Not user friendly because of the frequent change of the menu to translate the SMS
3. Reduction in the available space to be able to communicate since it will carry the recipient codes or source/target sources
4. Inefficient because of the network traffic

The frequent send and receive of the SMS involve the recipient's number to be carried along with the source and target languages or information with this available character spaces are shortened. Hence becomes less user friendly with the need to remember language code pairs. But the benefit of the system is that there could be a social networking among peers which will lead to better understanding and educational attainment of the treated issues each time the operation is performed among the peers.

VI. TESTING AND EVALUATION

We have the opinion that search queries can be made normalized to enhance the results in the SMS search information access database. There is a need to normalise SMS web based information retrieval disambiguating query from which search is being conducted. With a pre-defined query there may not be need to form query expansion or reformulation at various stages in the search process to improve the recall or precision. The search and results delivery is all about extraction of the topic or keyword from context of the search to represent the relative importance of terms in a vector space model [16]

Accessing the data base is very important in the IR system that we are building but the server may not be able to give the expected results unless the SMS that have appeared in so many forms based on the convenience of the users and the restriction caused by the bandwidth. There is a need to normalize the different forms, in which the language spreads. To implement translation in mobile devices software like Java Micro Edition (J2ME), Symbian, Windows Mobile Operating System are required. The capacity of the cell phone to be able to implement translation is a function of the memory and processing capability. These are all spread on the function of cost on the individual cell phones and the economic capability of the users.

The translation can be implemented in two important places: users and the centralized database server that we have control on instead of the proprietary network provider database server. The processing speed and the memory utilization, it is highly suggested that the backend to the user does this responsibilities. Translation service can be very fast in the centralized server even with a very low cost handset if the translation is performed in the server.

SMS is never advisable to be translated at the sender's end for it to be able to access the information at the server end but rather at the server end because of the capability of the server and its processing potentiality. In translation there is need to have the dictionary where string matching will be executed. The user hand set will need intelligent software to translate the SMS to real standard English for it to be able to access the database. This is at the expense of installing application translator at the users' side hence increases the cost of the handset. To afford this, a centralized server is made available to provide the translation services within the network setting [17]

The set of query are later tested with SMS translators and transferred to the database server where they are returned as results. In returning results under web-based search engine, the results may be poor because most of the web based search engines have pre-defined topic. This means any query that is outside the scope of this database will return void as an answer to the request[8]. This is similar to the pre-defined query that we are investigating. This same pre-defined topic or keyword will be used in information extraction for the best results during the web searching.

The main contribution of this work is the SMS normalization before information searching/retrieval which assist in identification of the pre-defined topic or keywords. We built our own toolkit system using Java programming language to implement the algorithms. Using the algorithm referred to as **S**earch **CO**mpare and **RE**place "**SCORE**" we were able to search, then compare the strings and finally we replace the string in order to be able to get the formal English version of the SMS English. This is very important so that there is uniformity in the request being made by the mobile users. This process is called SMS normalization and it is very important for information retrieval mechanism. There is need to converge to one English version of all the several ways of representing the same predefined SMS query. However, mobile users write SMS texts in several ways even though the query is standing as one.

Our preliminary findings disclose great discrepancy in the translation of the SMS in the open sources even when the same SMS language was used for the same open source. For instance, "*btw"* is translated as "*by the way"* by transl8it! but as "*between"* in Lingo2Word. A phrase-based query like "*What is"* is taken has "*Waz"* by both open sources.

Notably in our algorithm is the issue of capitalization, there is no word that is prioritized in the area of capitalization. Such is given priority in the open sources. Therefore, our system is a compliment to the existing system that needed specialized parsers and data sources. From the dataset we were able to isolate some noticeable complex cases of existence of words, particularly in Health sector in which we are working on, which are strictly uppercase (such as *HIV* and *AIDS)*. Can *aids* or *hiv* be considered abbreviations, misspellings or separate words? Where *aids* is not standing as *help* but as *Acquired Immune Deficiency Syndrome.* These problems are minor but require thought in designing these sets of word as being a keyword in the corpus. Our toolkit was able to decide that tokens which consist solely of uppercase characters should (optionally) not even be considered by normalization since they will most likely be keywords, proper names, variable names, or acronyms.

The basic algorithm used in Spelling Checker use an external list (a *dictionary)* to confirm the correctly spelled words and suggest how replacement or substitution can be introduced. The spelling checker algorithm using a dictionary is

Initialize. Changing the cases, all to lower case
Identifying token like AIDS, HIV

Build List. Construct a list of all distinct tokens in the SMS query file from different mobile users.

Search. Look up each token of the list in the Dictionary in the database server.

If the token is in the dictionary, it is correctly spelled, then it is compared with the correct version of the SMS.

If the token is not in the dictionary, it is **not** known to be part of the SMS database, then update

Replace. Replace with the correct English and ready for the information search

VII. RESULTS OF COMPARISON

The result from Table 1, confirm that our toolkit, base on the SCORE algorithm, was able to normalise the phrase and gives the exact translation for different SMS versions.

TABLE 1: Comparison of results for queries between translators

English	SMS Query from participants	Transl8t!	Lingo2Word	SCORE Algorithm translation
What is HIV	*Wat is hiv*	What is	What is	What is HIV
	Wats hiv	whats	What's	
	Watz hiv	whats	What is	
	Whats hiv	wats	What is	
	Wht s hiv	whats	What is	
	Wots hiv	whats	Whats	
	Wt s hiv	whats	wt is	
	Wt's hiv	whats	Wt's	
	Wts hiv	Whats	Wts	
	Wtz hiv	whats	Whats	
	Wht is hiv	What is	What is	
	Wat's hiv	whats	What's	

VIII. CONCLUSION

Using an open source language translation package and a database server we implemented automatic language translation for SMS query based text. We showed that our mechanism will allow mobile operators to provide textual message delivery in users' form of sending SMS. We compare the two open sources and the performance of our toolkit in the query-based information extraction. Results of applying this on our datasets show that FAQ datasets can be better normalized. The adaptation of search to mobile devices is not completely straightforward. There is reason to normalize the SMS query for effective information retrieval. mechanism, but it has been a quite challenging affair due to inherent noisy level of the type of communication. This paper compares our algorithm (**SCORE**) to show how effective query-SMS can be made to access the database for effective retrieval of the best ranked answers.

IX. FUTURE WORK

The future work is on how to apply text steganography to the retrieved documents as it is sent back to the mobile users without much ado to the network operator even if such transmission is noticed. Steganography is the art of hiding information within information such that it appears as if nothing is hidden**.** It promotes the potential capability to hide the existence of confidential data, difficulty of detecting the hidden (i.e. embedded) data, and strengthening of the secrecy of the encrypted data

ACKNOWLEDGMENT

The researchers wish to appreciate the constructive remarks of everyone in the research group during the earlier presentation of this research.

REFERENCES

1. Dr. Hadeel Showket Al-Obaidy, "Building Ontology Web Retrieval System Using Data Mining," ed.
2. R. S. Monojit Choudhury, Vijit Jain, Animesh Mukherjee, Sudeshna Sarkar and Anupam Basu, "Investigation and modeling of the structure of texting language," *IJDAR*, vol. 10, pp. 157-174, 2007.
3. S. P. Sumalatha Ramachandran, Sharon Joseph, Vetriselvi Ramaraj, "Enhanced Trustworthy and high-Quality Infromation Retrieval System fro Web Search Engines," *IJCSI International Journal of Computer Science Issues*, vol. 5, 2009.
4. M. k. Wadee S. Alhalabi, Moiez Tapia, "Search Engine ranking Efficiency Evaluation Tool," *Inroads-The SIGCSE Bulletin*, vol. 39, pp. 97-101, 2007.
5. Available: http://www.seochat.com/c/a/Search-Engine-Optimization-Help/Mobile-Search-Engine-Optimization/
6. G. Pass, *et al.*, "A Picture of Search," *First Intl. Conf. on Scalable Information Systems*, 2006.
7. D. E. Rose and D. Levinson, "Understanding User Goals in Web Search," 2004.
8. J. Chen, *et al.*, "SMS-Based Contextual Web Search," presented at the Mobiheld '09 Barcelona, Spain, 2009.
9. S. K. Samanta, *et al.*, "Automatic language translation for mobile SMS," *International Journal of Information Communication Technologies and Human Development (IJICTHD)*, vol. 2, pp. 43-58, 2010.
10. Chien-Liang Liu, *et al.*, "Computer assisted writing system," *Expert systems and applications*, vol. 38, pp. 804-811, 2010.
11. J. L.Peterson, "Computer Programs for Detecting and Correcting Spelling Errors," *Communication of the ACM*, vol. 23, pp. 676-687, December 1980 1980.
12. Y. How and M. Kan, "Optimizing Predictive Text Entry for Short Message Service on Mobile Phones," in *In Proceedings of human computer interfaces international*, 2005.
13. F. J. Damerau, "A technique for computer detection and correction of spelling errors.," *Comm. ACM*, vol. 7, pp. 171-176, 1964.
14. Ademola O. Adesina, *et al.*, "Text Messaging: a tool in e-Health Services," in *Southern Africa Telecommunication Networks and Applications Conference (SATNAC)*, Spier Estate, Stellenbosch, South Africa, 2010.
15. As cited by en.wikipedia.org/wiki/MXit.
16. C. Yu, *et al.*, "Term Weighting in Information retrieval using the term precision model," *Journal of ACM (JACM)*, vol. 29, pp. 152-170, 1982.
17. Chava V., *et al.*, "System and method for in-transit SMS language translation," United States Patent 7272406.