

Research Article

SMS Influence on the Text Based Communication and an Approach for Domain Specific Information Retrieval Using SIR Model

¹S.A. Harish Rajaraman, ²T.F. Michael Raj. and ²B. RajaKumar

¹Department of English,

²Department of CSE, SRC, SASTRA University, Thanjavur, Tamil Nadu 613401, India

Abstract: Now a days technological advancements influence human language. In fact some such advancement has taken its toll on the written language as well. One of the most important messaging formats is SMS that influences the communication as well as the written contents. The representation of information in SMS may differ in different languages. This study proposes an approach that explains how information retrieval can be made easy when the search query is in the form of SMS text in domain specific environment. This approach also considers the existing message representation and IR technologies. The evaluation of the results shows that this method can be employed as one of the search techniques in IR.

Keywords: Information retrieval, SIR model, SMS process

INTRODUCTION

The development of new technology in mobile devices makes significant changes in communication through SMS (Bollen *et al.*, 2004; Mitchell and Doherty, 2003; Kristine and Fintan, 2005). This helps identify people based on their area of interest (Chin-Yeh *et al.*, 2003). Real time information transformations are using emerging trends in technologies like mobile communications which makes significant changes in IR (Bookstein, 1985).

First of all, it is important to note how different dictionaries give or explain the meaning of SMS. For example, The Oxford Advanced Dictionary of Current English (Seventh Edition) defines SMS as a system for sending short written messages from one mobile phone/cell phone to another (the abbreviation for 'Short Message Service'). The Macmillan English Dictionary for Advanced Learners-International Students Edition says that SMS is a technical short message service: a method of sending a TEXT MESSAGE TO A MOBILE PHONE. As the Shorter Oxford English Dictionary (Deluxe Edition) has it, as an abbreviation it is Short Message Service, a system that enables mobile phone users to send and receive text messages, as a noun it is a text message that is sent or received using SMS.

The BBC defines it as "SMS stands for short message service and is also called texting or text messaging. SMS messages or 'texts' are usually sent from one mobile phone to another, but can also be sent from some home phones. It is a quick and convenient way of sending a short message to someone."

The first ever text message, it is believed, was an engineer's Christmas wish "Merry Christmas", sent on December 3 1992, from a PC to a mobile device using Vodafone's UK network. It has now become one of the most popular forms of communication, with hundreds of thousands being sent every day over both traditional mobiles and smartphones. BBC News BBC news report recorded that a random selection of Londoners to explain when it was appropriate to text and what life was like before the invention. Texting didn't get a really positive feedback, despite its prevalence then, the report added.

That Lee describes text messaging as a 'technology-assault' and Thurlow says that it affects the language of the young, makes one wonder whether formal English would stand such stern tests in the long run. True. Though it is not that technology will fully does it, any language must be prepared to face the damage, because the damage done is done and is irrevocable, even if the users of that language pacify themselves by calling it a 'paradigm shift' or revolutionizing communication!

There is a widespread argument that the use of abbreviations in sundry ways stimulate creativity benefitting literacy. Besides, studies have proved that the young demonstrate that their grammar remains unaffected even after becoming 'textperts' (texting experts) if one may call them. Thus text messaging is viewed as a form of relaxation that offers its users ample freedom.

But there are instances when teenagers become clueless in a formal situation, because even without

Corresponding Author: S.A. Harish Rajaraman, Department of English, SRC, SASTRA University, Thanjavur, Tamil Nadu 613401, India

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

their realizing it they lose track of grammar. Having allowed texting to creep in, the users now have the responsibility to use it to their advantage. It is believed that there is every possibility of language functions losing their significance with computer-conference-participants confining themselves to a comparatively narrow vocabulary range. It is welcome that techies don't feel the need for an expansive vocabulary set to help them convey their ideas. But care should be taken not to over-simplify, or avoid using the exact word so that the functions of language remain intact.

Thanks to SMS language-it has led to the development of a certain common history whereby 'SMSsers' (SMS users) understand the use of textspeaks. The SMS-users population is on the rise. Though this common history doesn't mean that textspeaks can be literal substitutes for the original words, it offers us an exciting study.

LITERATURE REVIEW AND RELATED CONCEPTS

A formal retrieval technique utilizes the concept of document retrieval and indexing. Various information modelling approaches have been developed and used namely, vector space, probability and fuzzy set (Bookstein, 1985). An ad hoc method uses the similarity measures for comparing the SMS text based queries. Logic based IR system deals with Boolean queries (Walker and Hobbs, 1981).

Features of SMS text: To develop an independent mobile based IR model through SMS text should consider the following characteristics. Query may contain short form of text like abbreviation of the information to be searched and developing a knowledge base for the domain specific IR is an essential task in this approach.

The performance of the cluster based search technique will produce the similar results as that of the feature based search techniques (Croft, 1980, 1982,

1983). Most of the learners found that using of SMS text based IR is easy and there is a difficult situation of forming the input query for the system (Nurhizam, 2004). The purpose of choosing SMS language for our approach is that it contains the following features (Jesús *et al.*, 2011):

- Phonetic abbreviation and vowel elimination- (e.g., Bk, buk-book)
- Non-phonetic abbreviation- (e.g., C++- C plus plus)
- Orthographic rules- (e.g., Hi! Good morning-Hi! GM)

Issues in the IR: Identifying the issues in the IR will strengthen our approach. There exists some problems in IR and have been described in Helen (1965) and Dagobert (1989). It is determined that query formation is one of the important requirements to develop IR model and some assistance needed for the user. People must understand the usage and processing of information for the domain. It includes defining the dataset for domain:

- Representation of information
- Comparing the SMS query with texts to achieve the objectives
- Identifying the effective technique for information retrieval

Classification of IR techniques: There are two major classification in IR techniques viz. traditional IR and semantic based IR techniques. The traditional IR techniques are straight forward methods. They find search string in a given document and support space complexity and requires only less effort for insert-and-update operations. It covers full text scanning, inversion, signature files and clustering models.

The semantic based IR plays major role even when the user gives minimum information to search the

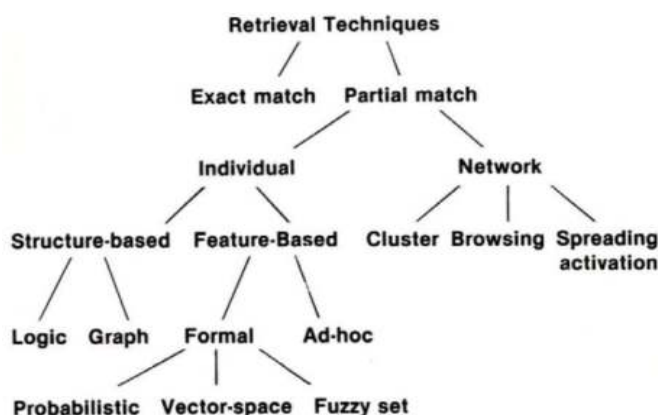


Fig. 1: Classification retrieval techniques

contents. It includes NLP, latent semantic indexing and neural networks (Faloutsos and Oard, 1995) and integrates the concepts of terms and their relevancy to identify the appropriate terms. The term-document relations and semantic relations can be easily determined using NLP and NLP can improve the performance and accuracy of search results (Nagao, 2005). Figure 1 shows the classification techniques.

The process of IR depends upon choosing the IR technique which comprises various features such as speed, consistency, accuracy and ease of use. The following Table 1 distinguishes the IR techniques.

PROPOSED METHODOLOGY

The proposed information retrieval technique uses SMS text as query and it searches for the contents which also refers to the term in the form of SMS. This model is implemented in a very small-scale. The model implements partial match technique and compares queries with the contents represented as a set of index terms or features.

The indexing process is done either by automated process or manual process (Salton and Voorhees, 1985). The architecture for the proposed SIR model is shown in the Fig. 2. It shows the various phases of the approach. The results of the approach may consist of a single word or multiple words but not the phrases or concepts.

The domain specific information is converted into SMS recognizable information and is derived from the domain knowledge of the system. The frequently used terms in this category are considered as knowledge base. The NLP concept has been incorporated into this model to identify the appropriate term related to the SMS text query (Shreyas *et al.*, 2012; Venkat *et al.*, 1997).

The entire model is classified into two phases. Phase-I deals with processing user queries and responding to them with suitable search results and phase-II manages the Knowledge base for the system by deriving it from the past records which comprises the SMS text which is depicted in the steps 5 and 6.

Algorithm Derive_Knowledge (input Q [])

```

{
    //Pre-process the usual text
    1. Read the normal text from the Domain (D)
       related contents
    2. Convert into terms T {ti} where i = 0...n
    3. for each t in T [ ]
       {
           if t ∉ D-Domain then
               insertintoKB (pop (q, Q [ ]))
       }
}
    
```

Table 1: IR techniques and features

Techniques	Features	Advantages
Vector space	Usage of n-dimensional space	Term indexing
Probability	Considers past dataset	Leads to develop new techniques
Fuzzy set	Based on boolean queries and ranking	Uses term dependencies
Logic	Information is represented using formal logic	Rule based IR
Graph	Graph-like representation	Semantic relations

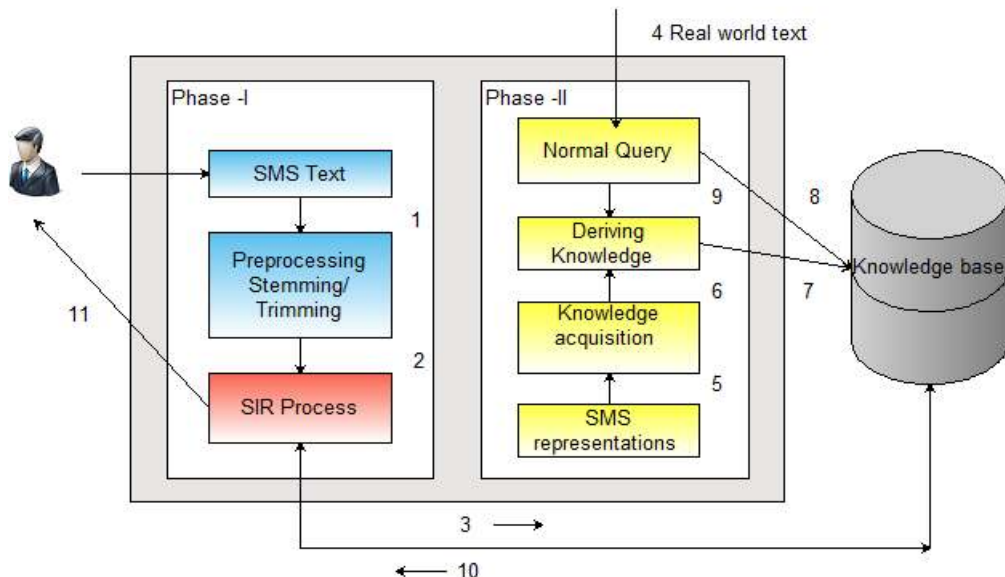


Fig. 2: SIR model

Table 2: Performance ratio in distributed environment

Query	Response time in msec	Response time in msec	Performance ratio	Result
Gr8	5.0003	1585.0907	317	Excellent
bk	13.0008	1492.0854	114.77	Excellent
c++	14.0008	50.0028	3.57	Excellent
C u	13.0007	46.0026	3.54	Excellent
Gd luk	4.0002	1058.0605	264.50	Excellent
Tc	8.0004	27.0015	3.38	Excellent

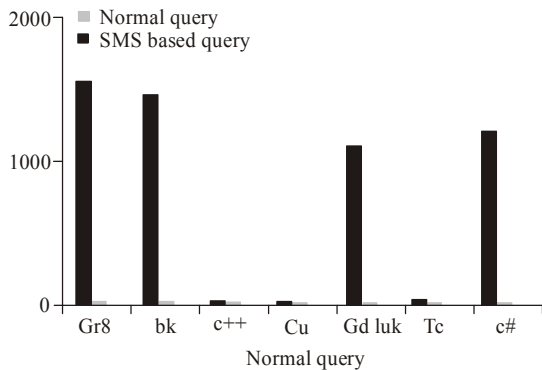


Fig. 3: Performance in distributed environment

Algorithm SIR_Search (input T [], Output resultList [])

```

{
  //Pre-process the user query
  //Query may contain single/multiple terms
  1. Process the query Q into TinQ {qtj} where j = 0...n
  2. for each qt in TinQ [ ]
    {
      //check for domain relevancy
      if qt ∉ D-Domain then
        Irrelevant terms list-Non related terms-
        NRT [ ]
        Derive_Knowledge (input NRT [ ])
      else
        resultList [j] = searchKB (qt, KB)
    }
}

```

RESULTS AND DISCUSSION

Design GUI based interface is an essential solution for any IR technique. SIR model is designed to process the SMS based query as input and is evaluated with thousands of the records which are collected from the various background.

Since the model is domain specific, we have tested the model to satisfy the students' requirements, which includes only the academic details. The system is evaluated by comparing the normal query and SMS query, by which the performance can be improved in terms of space and time complexity. Table 2 describes the response time and the performance ratio and Fig. 3 gives the pictorial representation of the results under distributed environment.

CONCLUSION

This paper tries to explore the influence of SMS in the non-formal communication and gives an idea to develop an SIR model. The data have been collected from educational background to form a domain specific system.

The model is implemented in a distributed environment, the analysis and evaluation of system is done by comparing the results with non-distributed IR model. The result shows that there is considerable improvement in performance. This model gives ample scope for future enhancement, where extending the domain into an integrated one that would have a new, expansive knowledge-base, which in turn would incorporate the concepts of semantic relations of SMS text message, can't be ruled out.

REFERENCES

Bollen, L., S. Eimler and H.U. Hoppe, 2004. SMS-based discussions: Technology enhanced collaboration for a literature course. Proceeding of the 2nd IEEE International Workshop on Wireless and Mobile Technology in Education, pp: 23-25.

Bookstein, A., 1985. Probability and Fuzzy-sets Applications to Information Retrieval. In: Williams, M. (Ed.), 8th Edn., Annual Review of Information Science and Technology. Knowledge Industry Publications Inc., White Plains, NY, 20: 117-15.

Chin-Yeh, W., L. Baw-Jhiung, C. Kuo-En, H. Jorng-Tzong and C. Gwo-Dong, 2003. Using mobile techniques in improving information awareness to promote learning performance. Proceeding of the 3rd IEEE International Conference Advanced Learning Technologies.

Croft, W.B., 1980. A model of cluster searching based on classification. Inform. Syst., 5(3): 189-195.

Croft, W.B., 1982. An overview of information systems. Inform. Technol. Res. Dev., 1(1): 73-96.

Croft, W.B., 1983. Applications for information retrieval techniques in the office. In: Kuehn, J.J. (Ed.), Proceeding of the 6th Annual International Association for Computing Machinery Special Interest Group on Information Retrieval (ACM'SIGIR) Conference on Research and Development in Information Retrieval, ACM Inc., pp: 18-23.

Dagobert, S., 1989. Important Problems in Information Retrieval. Retrieved from: www.dsoergel.com/cv/G1.html.

- Faloutsos, C. and D.W. Oard, 1995. A survey of information retrieval and filtering methods. Technical Report CS-TR-3514, Department of Computer Science, University of Maryland, College Park.
- Helen, M.T., 1965. Problems of information retrieval. *Aslib Proc.*, 17(7): 210-216.
- Jesús, O., I.S. José, D.D.C. María and I. Ángel, 2011. SMS normalization: Combining phonetics, morphology and semantics. In: Lozano, J.A., J.A. Gámez and J.A. Moreno (Eds.), CAEPIA 2011. LNAI 7023, Springer-Verlag, Berlin, Heidelberg, pp: 273-282.
- Mitchell, A. and M. Doherty, 2003. M-Learning Support for Disadvantaged Youth: Mid-Stage Review. Ultralab, Anglia.
- Nagao, M., 2005. Natural language processing and knowledge. Proceedings of IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE'05).
- Nurhizam, S., 2004. The use of Short Messaging System (SMS) as a supplementary learning tool in Open University Malaysia (OUM). Proceeding of the 18th Annual Conference Association of Asian Open Universities (AAOU). Shanghai, China, pp: 27-30.
- Salton, G. and E.M. Voorhees, 1985. Automatic assignment of soft Boolean operators. Proceeding of the 8th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'85), pp: 54-69.
- Shreyas, M.P., R.J. Rajkalpesh and R.S. Sunil, 2012. Information retrieval system using SMS. Proceeding of the Emerging Trends in Computer Science and Information Technology (ETCSIT, 2012), 1: 28-31.
- Venkat, N.G., V.R. Vijay, I.G. William and K. Rajesh, 1997. Information retrieval on the World Wide Web. *IEEE Internet Comput.*, 1(5): 58-68.
- Walker, D.E. and J.R. Hobbs, 1981. Natural Language Access to Medical Text. Technical Note 240, SRI International, Menlo Park, CA.
- Xristine, F. and C. Fintan, 2005. When fingers do the talking: A study of text messaging. *Interact. Comput.*, 17: 167-185.