Global Journals LaTeX JournalKaleidoscope™
Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

Dynamic vs Static Term-Expansion using Semantic Resources in Information Retrieval

Ramakrishna kolikipogu[1] and Padmaja Rani B[2]

[1] Sridevi Womens Engineering College

Received: 14 December 2012 Accepted: 4 January 2013 Published: 15 January 2013

Abstract

Information Retrieval in a Telugu language is upcoming area of research. Telugu is one of the recognized Indian languages. We present a novel approach in reformulating item terms at the time of crawling and indexing. The idea is not new, but use of synset and other lexical resources in Indian languages context has limitations due to unavailability of language resources. We prepared a synset for 1,43,001 root words out of 4,83,670 unique words from training corpus of 3500 documents during indexing. Index time document expansion gave improved recall ratio, when compared to base line approach i.e. simple information retrieval without term expansion at both the ends. We studied the effect of query terms expansion at search time using synset and compared with simple information retrieval process without expansion, recall is greatly affected and improved. We further extended this work by expanding terms in two sides and plotted results, which resemble recall growth. Surprisingly all expansions are showing improvement in recall and little fall in precision. We argue that expansion of terms at any level may cause inverse effect on precision. Necessary care is required while expanding documents or queries with help of language resources like Synset, WordNet and other resources.

Index terms— information retrieval, query expansion, semantics, indexing, document expansion, information retrieval in indian languages.

introduction nformation Retrieval in local languages is getting more popularity in developing countries like India. Use of Internet and other Information Accessing Systems plays major role in Education, Medical, Business, Agriculture and other significant domains. Information Retrieval Systems in Local Languages that are getting popular among the Netigens, who prefer to access their information needs in their mother tongue language. Availability of digital documents in native languages creates interest to the user to access the information by typing query in local languages. India is multilingual country and people across the country speak more than 400 languages, but all the languages are not recognized due to lack of scripts and rules. The government of India has given "languages of the 8 th Schedule" official status for 22 languages. Telugu is one of the recognized languages of India. Processing of Telugu digital items is more difficult when compared to European languages and other Indian Languages.

Building efficient Information Retrieval System for Telugu is a challenging task due to richness in Morphology and conflational features of the language. In this paper we studied the effect of Document Expansion and Query Reformulation Techniques with the help of synset lexical resource. Naïve users prefer to give one time query and expect adequate results in the first glance. In general lot of fuzziness involved in user query and it is difficult to match the relevant items. There is a necessity to reduce the vocabulary mismatch between naïve user query and repository prepared by domain experts. This paper is an attempt to study the impact of terms expansion during Indexing and searching on a sample Telugu text corpus.

When we expand the terms during indexing and supplied normal un-expanded queries to the Information Retrieval System, we observed that, there is a great fall in precision. Based on the synset length for each

term the recall is positively affected. We then tested the same system by expanding query terms and keeping unexpanded items as source. Surprisingly the effect is similar and found improvement in recall and negative effect on precision. Expansion of terms either from document or query, the precision and recall are inversely proportional in growth rate. Main Objective of Information Retrieval is to retrieve relevant information from huge repository and preset top ranked items to the end user by reducing overhead in terms of time. Naïve users may not give strong queries to represent the concept in which, they expected to retrieve. Terminology of user for writing a query is always simple and vague; it may not resemble the concept of an item to be retrieved. We mean naïve user's vocabulary may generally drawn from day to day usage language, where as resources are drafted content in expertise vocabulary. Sometimes user may fail to use expertise vocabulary to write queries and represent the concept.

1 I

Abstract -Information Retrieval in a Telugu language is upcoming area of research. Telugu is one of the recognized Indian languages. We present a novel approach in reformulating item terms at the time of crawling and indexing. The idea is not new, but use of synset and other lexical resources in Indian languages context has limitations due to unavailability of language resources. We prepared a synset for 1,43,001 root words out of 4,83,670 unique words from training corpus of 3500 documents during indexing. Index time document expansion gave improved recall ratio, when compared to base line approach i.e. simple information retrieval without term expansion at both the ends. We studied the effect of query terms expansion at search time using synset and compared with simple information retrieval process without expansion, recall is greatly affected and improved. We further extended this work by expanding terms in two sides and plotted results, which resemble recall growth. Surprisingly all expansions are showing improvement in recall and little fall in precision. We argue that expansion of terms at any level may cause inverse effect on precision. Necessary care is required while expanding documents or queries with help of language resources like Synset, WordNet and other resources. Expansion techniques sometimes lead to poor performance and may miss the concept too. This increases overhead on naïve users to decide relevancy of outcome. Exhaustivity must be low to control adverse effect of precision and balance the recall as well. The same approaches are adapted to huge document collection from Wiki-Telugu and studied the effect. Most of the systems work on syntactic base; it requires exact matching of terms from query to document. Syntactic patterns are words in text mining. Word mismatch is a severe problem in Information [1].

2 1.

1) Simple IR System using statistical Indexing with nterms length query. 2) Query Reformulation at runtime using term expansion with synset in Pseudo Relevance Feedback (PRF) Approach. 3) Item Reformulation based on query terms using PRF approach. 4) Query Expansion and item expansion using synset with blind retrieval approach.

These approaches were discussed in Chapter 3 and Results are given in Chapter 5.

Information is growing in an exponential manner on World Wide Web, the problem of finding useful information and knowledge from abundant source becomes one of the most important topics in information retrieval and storage [4], [5]. Information retrieval support systems are being developed in supporting users to find necessary information and knowledge [3]. Information Retrieval System is a multidiscipline area of research, which involves text processing, speech processing, image processing, video processing and other mode of information processing. Retrieval of any kind of information mainly aims at satisfying end user to his query. Usually naïve users search with text query by limited vocabulary. Representation of source in order to facilitate matching against user query plays major role and having equal importance with query structure. In this paper we limited to text documents as resource to retrieve for the given query. Many of the documents retrieved for general queries are totally irrelevant to the subject of user interest, due to insufficient keywords supplied in the search [6]. Sometimes the words entered by user may not express the interest of the user. Vocabulary of users may far from the expert's terminology in documents and it is difficult to match the same. The word mismatch can be solved by rewriting queries with new terms called as query expansion [7]. Our objective in this paper is to select a suitable term for expansion and to improve precision and recall as well. Level of query expansion varies from model to model. Expansion Terms can be selected in many ways 1) Suggested terms are provided to select by user and expand the query without missing concept. This is more accurate way of term expansion called manual expansion, but it requires knowledge to judge the term relevance, which increases overhead on user. Naïve users are not familiar in writing queries; hence the word miss match comes into the picture. User can not be given burden to use retrieval system, that's why automatic query expansions techniques are regular practice in IR Systems. Relevance Feedback [8] method considers user selection out of retrieved as relevant and reformulate the query to repeat the search by adjusting weights of initial query terms. Users who are familiar with query expansion takes maximum benefit of query reformulation [9] with relevance feedback. In expertise user will better serve with Pseudo Relevance Feedback called Automatic Query Expansion [10].

Information retrieval using Language models are used to improve relevance of a query outcome by document set feedback [11]. Cluster Feedback (CFB) is another way of term selection to find more similar terms by clusters. If relevant clusters are identified, then combining them to generate a query model that is good at discovering

documents belonging to these clusters instead of the irrelevant one [12]. In Automatic Query Expansion (AQE), terms are given new weights to score the terms. Sum of weights will represent final score of terms, which is statistically good for item selection. Pure statistical weights may not functionally useful to represent query terms. Different functions have been proposed to assign high scores to the terms that best discriminate relevant from non-relevant documents [10]. A disadvantage of Query Expansion is the inherent inefficiency of reformulating a query [13]. The query is expanded using words or phrases with similar meaning to those in the query and the chances of matching words in relevant documents are therefore increased. This is the basic idea behind the use of a thesaurus in query formulation [15]. To improve the relatedness of the terms to documents, lexical resources Thesaurus, WordNet or Dictionaries usage promising little improvements in search results [16]. While global analysis mechanisms are inherently much more efficient than local ones (only dictionary lookups are performed during query time, rather than costly document retrieval and parsing), they are also likely to be less successful [1]. Document expansion by modifying Vector Space is to bring closer the query Vectors [14]. Good thesaurus for whole language is difficult to obtain. Synonyms are used to extract from thesaurus [18] for query expansion. Expansion terms are selected based on query association, where queries are stored with documents that are highly similar statistically. Falk Scholer and others [17] claimed that adding query associations to documents improves the accuracy of Web topic finding searches by up to 7%, and provides an excellent complement to existing supplement techniques for site finding. The studies are showing that, the query expansion improves the results of Information retrieval system. Statistical relatedness may not work properly and choose correct alternate terms to reformulate the query. Document expansion during indexing reduces search time. In this paper we studied the effect of Query In this paper we present various term selection methods for query reformulation and item expansion with implementation along with results as listed:

with and without Expansion versus Document with and without Expansion using Synset. Proposed work is proven to increase recall and precision as well. In few cases like, document expansion, precision is inversely affected the results.

3 a) Preprocessing of Telugu Text

Telugu is derived from Brahmi family [], one of the Dravidian languages. Telugu is morphologically rich language and word conflation is very high. The language scripts are complex to process, because they are combined syllables when compared to English. So it is difficult to preprocess using language models like stemming, n-gram etc. Romanization called WXNotation standards aim at providing a unique representation of Indian Languages in Roman alphabet [27]. Internally each script is represented UNICODE standard. The Unicode Standard, Version 6.2 assigned a hexadecimal code point for Telugu Scripts in the Range of 0C00-0C7F [28]. In this paper implementation is done by converting text from WX-to-UTF1 and UTF-to-WX before and after processing. This process slower the results, but efficiency in terms of recall and precision are not influenced. Carrying task directly in Unicode give faster results and possible, but processing text in Unicode level is difficult for programming. Our future work is planned to directly process in Unicode to improve the results speed. WX notations for Telugu language are given in Table 1.

4 Query Expasnion

Terms supplied by user may not be sufficient to express the concept and match documents. Terms may be out of bounds or in different vocabulary. Out of bounds problem can be solved by user feedback system. Vocabulary mismatch is common problem in Information retrieval. Vocabulary mismatch is one is of the principal causes of poor recall in Information different subset of words to specify a given topic, causing retrieval techniques based on lexical matching to miss relevant documents [19]. Expansion of query at search time is called run time query expansion. Query Expansion is a process of reformulating the root query by adding an optimal set of terms that improves recall and precision. The motivation for query expansion is rate of failure in retrieving relevant documents by simple queries. Various Query Expansion methods are in regular practice to improve the retrieval performance. Local Analysis and Global analysis.

5 a) Local Analysis

Initial search results of given query are analyzed and used to expand the query called local analysis. The top ranked documents were taken to change weights of query terms and repeat the search [20] [21]. User judge the relevance of top ranked items to the query as Relevance Feedback [8]. The thought of relevance feedback is to involve the user in the retrieval process so as to improve the final result set. The user issues an initial query. The system returns an initial set of relevant documents. In particular, the user gives feedback on the relevance of documents in an initial set of results. The system computes a better representation of the information need based on the user feedback [22]. It may cause the user to endure the process. Pseudo Relevance Feedback (PRF) is viable alternate to void user interaction during feedback. PRF is also called Blind Relevance Feedback or Automatic Relevance Feedback method, which automates the manual part of relevance feedback, so that the user gets improved retrieval performance without an extended interaction. PRF via query-expansion has been proven to be effective in many information retrieval (IR) tasks [23]. In most existing works, the top-ranked documents from an initial search are assumed to be relevant and used for PRF. One problem with this approach is that

one or more of the top retrieved documents may be nonrelevant, which can introduce noise into the feedback process. For all query expansion methods, pseudo relevance feedback (PRF) is attractive because it requires no user input [24]. Major problem with local analysis is that queries have an increased risk of query drift, as the top ranked documents are assumed to be relevant, while they may in fact not be [21]. In this paper we studied both Relevance Feedback and Pseudo relevance Feedback methods on a limited corpus. Top one document is considered as relevant and its terms are given more weight in-line with query terms and repeated search on same collection. Even though, sometimes original queries are totally modified with new terms and missing the concept of original query. Still it is found to be the best approach among alternate methods including global analysis, which is discussed in the next subsection. ? [a] ?[A] ?[i] ?[I] ?[u] ?[U] ?[q] ?[e] ?[eV] ?[E] ?[o] à°?"[oV] ??[aM] ??[aH] ?[ka] ?[Ka] ?[ga] ?[G] ?[fa] ?[ca] ?[Ca] ?[ja] ?[Ja] ?[Fa] ?[ta] ?[Ta] ?[da] ?[Da] ?[Na] ?[wa] ?[Wa] ?[xa] ?[Xa] ?[na] ?[pa] ?[Pa] ?[ba] ?[Ba] ?[ma] ?[ya] ?[ra] ?[la] ?[va] ?[sa] ?[Sa] ?[Ra] ?[ha] ?[lYa] [kRa] ?[rY]

Retrieval. Indexers and searchers invariably choose b) Global Analysis

The global analysis considers term cooccurrences and their relationships in the corpus as a whole, which is used to expand the query independent from the old query. Expansion by global analysis does not rely on initial query terms and the results retrieved from it, so that refinements in the query will cause the new query to match other semantically similar terms. A common problem with these query expansion methods is that the relationships between the original query terms and the expanded query terms are not considered [25]. In this paper our directions are to use synset words for query expansion and study the effect on training corpus. Recall ratio is improved in this direction.

6 i. Synset based Query Expansion

Our research is continuing in Information Retrieval in Indian Languages, as Telugu is one of the most spoken languages in India as well as all over the world. Language resources are limited in Telugu language and cross language attempt are facing many challenges, where the features are different from one language to other language. For this work we collected and manually created Telugu-Telugu synset of «<»> for whole corpus consisting of around 3 laksh words. Query preprocessing is done using similar process as applied to document indexing in section 4.1. Query Terms are expanded using Synset and combined terms with OR Boolean operator. Expanded query is used for search on static Indexed documents. The results were given in Section 5. Somehow vocabulary miss match problem is addressed by synset inclusion through run time terms expansion. E.g. one word query is

[amma]mother. First the root word is verified with word corpus of dictionary look-up, if found all synonyms from synset are connected using OR operation to generate new query. i.

7 e ([amma] OR [mAwa] OR

[walli]synonyms of mother in Telugu) treated a single term and weighted accordingly in query vector. This works good as we collected all possible synonyms in the corpus. Recall is greatly improved, at the same time precision is compromised due to deviating the concept as well as query drift.

8 ii. Synset based Document Expansion

It is impossible to predict the query from user and terms used by him. Instead of expanding terms during run time, as user need to wait for results, index terms of document set are expanded during indexing using synset. Off course the practice is not new, even though it is new to Indian languages especially for Telugu. Telugu Information Retrieval suffers from language resources. There is a demand for language resource to be developed for all Indian languages for public use. We created a synset for our training corpus of 40000 words with 1 .375 synonyms in an average and extracted and created a dictionary file. A hash is maintained to list synonyms of a document term to match against query term during searching. Similar to Query Expansion this attempt deprecate the process and resulted precision loss.

9 c) Relatedness Measurements

Relevance Feedback: Vector Space Model A New query with added terms from the top retrieved documents D is given as expanded query to research. Weighting can be taken either Boolean value, in which 0 represents deletion of old terms and 1 represents addition of new terms to the query. Similarity of Query and document is measured by: Rocchio [8] proposed a Relevance Feedback algorithm which better suggest a new query as:

Where Qnew is Reformulated query and Qold is initial query with Dr Relevant returned Documents, |Dr| number of relevant documents. Dnr is non-relevant returned documents and |Dnr| is total non-relevant documents in terms of vectors. With ? is original query weight, ? is related document weight and ? is weight of non-relevant documents. Less importance terms are represented with 0 in Boolean vector models. Concept of a query may depends on less weighted terms too, hence it is important to equally consider less weighted terms in sorting order. An alternate term weighting method call probabilistic approach better serve the purpose. The documents are ranked in decreasing order of rank as per the expression:IV.

10 Implementation

Telugu language resources are limited for research. We collected 3500 Documents from daily news portals and manually categorized into 10 categories as shown in Table 1. Initially all documents are kept under one set and run the search using 10 ,... 3 , 2 , 1 (qn q q q Qi ? Q i ?

is an initial query as a vector of terms q j . q j ? weight of each query termj in Q i) ... 3 , 2 , 1 (dn d d d Di ? D i ?

is an top Document as a vector of terms d j .d j ?weight of each Document term j in D i ? ? ? ? n i di qi Di Qi Sim 0) , (? ? ? ? ? Dnr Dnr Dr Dr Q Q old new ? ? ?

this process is continuing to create for 1 lakhs words synset. All unique root words from entire corpus are queries and followed by search against categorical sets of documents. There is no difference in results as?? ?? ? ? ?

documents are properly indexed before running search. Little search time varies from search on whole collection and categorical collection of documents. If the documents were categorized the results were bit faster. This time factor is important, but our aim is to improve the precision and recall.

Table **??** : Categorical Documents collection for testing a) Indexing and Searching using Synset Where fqi,j frequency of term i in document j.

ii. Create Inverted List [26] consisting of Document Ids and term frequency against Dictionary lookup. Inverse Document Frequency a term i is calculated using idf Inverted List to give more importance to that file containing a term with synonyms. Weighting factor are greatly affected by synset. 8. Relevance judgment is find using cosine similarity measure between document and query vectors:

11 Results Analysis

In this paper we investigated affect on precision and recall when query is connected with synset. Figure 2 is a simple precision -recall graph as baseline approach to compare the proposed system. Where N total no. of documents and ni is number of document that term i occurs.q d q d q d similarity j j j ? ? ? ? ? ? ? ? ? ?) , (cos? ? ? ? ? ? ? ? ? t i q i j i t i q i j i w w w w 1 2 , 2 , 1 , , cos?

Where ? is angle between i j i q w w , , & , wi,j is weight of term i in document j and wi,q is weight of term i query q. ? varies from 0 to 1. Fig. 3. Shows precision-recall in normal search process without any aids. Query expansion is applied by i analyzing top one document as relevant and adding new terms to the query. This feedback is iterated once and results are plotted in Figure 2.

Recall -Precision@10 variations with baseline, Relevance Feedback (RF) by top 1 document, Query Expansion using Synset (QES) and Document Expansion using Synset (DES)

Recall is not affected much in Figure 3., but precision loss is observed. All methodsare comapred in Figure **??** with Precesion@10. Baseline metho Query Expansion using synset is plotted in figure **??**. When we use synset for query expansion instead of relevance feedback method, the precision is improved along with recall. Whereas Document Expansion with synset instead of Query Expansion, the results were greatly affected by both precision and recall. Query expansion with synset outperforms among all methods and we argue that, use of synset to expansd query is better than document expansion. The proposed system has to be tested on huge corpus so as to claim in universal Information Retrieval. Experiments are taken on TREC using similar methods by meny researchers and found precision loss. As there is no standard corpus for Telugu language we tested on private corpus developed by us. Terms Expansion during indexing using synset gave poor performance as shown in Table **??**. The search results are bit faster in DES when compared to RF & QES methods, but these to are good in relevance calculation.

12 V. Conclusion

Indian is multilingual country stands in 2nd in population. There is observable growth in literacy, but people prefer to use local languages after English. There is necessity for cross lingual information retrieval systems to serve the users according to their information needs. Most of the Indian Languages are having unique language features and it is difficult to translate from one language to other, even Google search engine fails to produce exact cross lingual results. Building monolingual information retrieval is a mandatory task, where compatibility may not be an issue in using language resources. Once identifying all features of a language, it is easy to translate into other language by mapping rules. Information Retrieval in Telugu Language is in inception level, due to lack of language resources like POS Taggers, Entity Recognizers, Morphological Analyzer, Dictionaries, WordNet, Ontologies et. al. The results in this paper were given hope to continue with query expansion. Use of controlled vocabulary may further improve the results. Anyhow Query expansions techniques will have inverse effect on precision and improvement in recall. For the end user precision is more important, as he expects results to be displayed on top in once shot. Exploring concept of the query using synset or WrodNet may give better performance. We need to investigate how Information retrieval system in Telugu Language works by using query concepts. [1] [2] [3]

[1] © 2013 Global Journals Inc. (US) Year

[2] © 2013 Global Journals Inc. (US) Year

[3] © 2013 Global Journals Inc. (US) Year

Figure 1: C

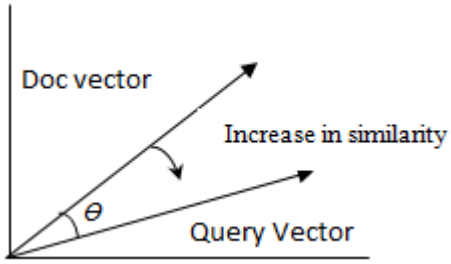

Figure 2: GlobalC

Figure 3: C

Figure 4: C

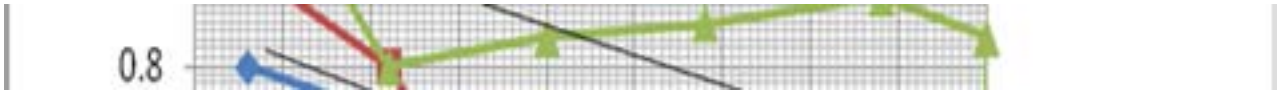

Figure 5:)

Figure 6: 1 .

Figure 7: Figure 1 :

Figure 8: Fig. 2

Figure 9: Figure 3 :

1

Figure 10: Table 1 :

	#Queries 10							
S.No	Category	#Docs						
	Business		150					
2	Devotional	1552						
3 4 5	Editorial Historical Places Literature Politics		150	152	305	332	Total	
6							#Docs	
							3500	
7	Science		152					
8	Songs		298					
9	Sports		294					
10	Stories		155					
	, tf ? j i	max(, , i j i	fq	fq		j	)	
	idf i log ?		ni	N				

[Note: 7. Synset is used to identify synonyms of a term in a document, if found term frequency is incremented in]

Figure 11: C

[Systems (2002)] , Fuzz-Ieee Systems . *IEEE World Congress on Computational Intelligence* 2002. May 12-17, 2002. p. .

[Manning et al. ()] *An introduction to information retrieval*, C Manning , P Raghavan , H Sch Utze . 2009. Cambridge University Press.

[Scholer et al. ()] 'Automatic Keyword Classification for Information Retrieval'. F Scholer , H E Williams , A Turpin . *Journal of the American Society for Information 18. K. Sparck Jones*, (Butterworths, London) 2004. 1971. (Query association surrogates for web search)

[Carpineto and Romano (2012)] 'Automatic Query Expansion in Information Retrieval'. Claudio Carpineto , Gioovanni Romano . *ACM Computing Surveys* 2012. January. 44 (1) . (Publication date)

[Buckley et al. ()] *Automatic query expansion using*, G Buckley , J Salton , A Allan , Singhal . SMART: TREC-3. 1994.

[Bharati et al. ()] Akshar Bharati , Vineet Chaitanya , Rajeev Sangal . *Natural Language Processing: A Paninian Perspective. PHI*, 2010.

[Billerbeck and Jobel ()] 'Document Expansion versus Query Expansion for Ad-hoc Retrieval'. Bodo Billerbeck , Justin Jobel . *Proceedings of the 10th Australasian Document Computing Symposium*, (the 10th Australasian Document Computing SymposiumSydney, Australia) 2005. p. .

[Document Expansion versus Query Expansion for Ad-hoc Retrieval 10th Australasian Document Computing Symposium ()] *Document Expansion versus Query Expansion for Ad-hoc Retrieval 10th Australasian Document Computing Symposium*, 2005. Bodo Billerbeck Justin Zobel; Sydney.

[Xu and Croft ()] 'Improving the effectiveness of information retrieval with local context analysis'. J Xu , W B Croft . *ACM Transactions on Information Systems* 2000. 18 (1) p. 112.

[Moldovan and Mihalcea ()] 'Improving the Search on the Internet by using Word Net and Lexical operators'. Dan I Moldovan , Rada Mihalcea . *IEEE Internet Computing* 1998.

[Kolikipogu Ramakrishna et al. ()] 'Information Retrieval in Indian Languages: Query Expansion models for Telugu language as a case study'. Kolikipogu Ramakrishna , . B Dr , Padmaja Rani . *4th IEEE -International Conference on Intelligent Information Technology Applications*, 2010. (china)

[George and Furnas ()] 'Information retrieval using Singular Value Decomposition Model of Latent Semantic Structure'. W George , Furnas . *ACM-SIGIR-1988*, 1988. p. .

[Attar and Fraenkel ()] 'Local Feedback in FW-Text Retrieval Systems'. R Attar , A S Fraenkel . *Journal of the Association for Computing Machinery* 1977. 24 (3) p. .

[Dominich ()] *Mathematical Foundations of Information Retrieval*, S Dominich . 2001. Dordrecht: Kluwer Academic Publishers.

[Renxu Sun Chai-Huat Ong Tat-Seng Chua (ed.) ()] *Mining Dependency Relations for Query Expansion in Passage Retrieval, SIGIR'06*, Renxu Sun Chai-Huat Ong Tat-Seng Chua (ed.) 2006. Seattle, Washington, USA. p. .

[Zhai and Lafferty ()] 'Model-based feedback in the language modeling approach to information retrieval'. C Zhai , J Lafferty . *Proceedings of the 10 th international conference on information and knowledge management*, (the 10 th international conference on information and knowledge management) 2001. p. .

[Baeza-Yates and Ribeiro-Neto ()] *Modern Information Retrieval*, R Baeza-Yates , B Ribeiro-Neto . 1999. New York: Addison Wesley.

[Maron and Kuhns ()] 'On relevance, probabilistic indexing and information retrieval'. M E Maron , J L Kuhns . *J. ACM* 1960. 7 p. .

[Xu and Croft ()] 'Query Expansion using Local and Global Document Expansion'. Jinxi Xu , W.Bruce Croft . *ACM SIGIR Conference Proceedings*, 1996. p. .

[Ruthven ()] 'Re-examining the potential effectiveness of interactive query expansion'. I Ruthven . *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in information Retrieval*, (the 26th Annual International ACM SIGIR Conference on Research and Development in information Retrieval) 2003. ACM Press. p. .

[Kolikipogu et al. ()] 'Reformulation of Web query using Semantic Relationships, International ACM-ICACCI-12'. Ramakrishna Kolikipogu , Padmaja Rani , B ; Y Y Yao . *Information Retrieval Support. Figure* 2012. 2002. 2.

[Rocchio ()] 'Relevance feedback in information retrieval'. J J Rocchio . *The SMART Retrieval System, G. Salton Ed*, (Englewood Cliffs, NJ) 1971. Prentice-Hall. p. .

[Kolikipogu and Rani ()] *Study of Indexing Techniques to Improve the Performance of Information Retrieval in Telugu Language, IJETAE*, Ramakrishna Kolikipogu , Padmaja Rani , B . 2013. 3.

[Tan et al. ()] Bin Tan , Atulya Velivelli , Hui Fang , Chengxiang Zhai . *Term Feedback for Information Retrieval with Language Models, SIGIR-2007 Proceedings*, 2007. p. .

[Salton ()] 'The SMART Retrieval System'. G Salton . 373.393. NJ. *Experiments in Automatic Document Processing*, (Englewood Cliffs) 1971. Prentice-Hall.

[Croft and Harper ()] 'Using probabilistic models of document retrieval without relevance information'. W B Croft , D J Harper . *Journal of Documentation* 1979. 35 p. .

[Xu et al. ()] Yang Xu , J F Gareth , Jones . *Query Dependent Pseudo Relevance Feedback based on Wikipedia, Association of computer Machinery SIGIR'09 Conference Proceedings*, 2009.