

# Pseudo Relevance Feedback by linking WordNet for Expanding Queries in Information Retrieval Process

Ramakrishna Kolikipogu, B. Padmaja Rani, and N. Swapna

**Abstract**—Vocabulary and Word mismatches are common problems in Information Retrieval Systems. Query Expansion(QE) gives a solution to these problems. Selection of key terms to expand the root query is a challenging task. Semantic relationships of a particular term in WordNet will improve the performance. Pseudo Relevance Feedback is one of the proven methods of QE which consider relevant terms to be expanded. Initial retrieved Items are considered as relevant to the query and further used to expand the query. Whenever the sense of key term is found, its lexical relationship synonym is considered for QE. The specific terms to be included in the initial query are selected automatically and the terms selected for the initial query are connected before expansion with OR Boolean operator. We found there is no Telugu WordNet available hence the proposed system is aimed to test on hand-crafted WordNet on limited text collection for Telugu language.

**Index Terms**—Query expansion, pseudo relevance feedback, wordnet, synset, word sense disambiguation, parts of speech, normalization.

## I. INTRODUCTION

Web is the best Source for Information and to Retrieve Information, where different techniques are used to give exact information needed by the users. Naive users are not familiar in writing structured queries. They submit short queries that do not consider the variety of terms used to describe a topic, which resulting in poor recall power [1]. Word mismatch is a common problem in Information Retrieval System. Vocabulary mismatch between naïve user and the author makes the search process more difficult. Due to lack of knowledge users face difficulties while searching for information in a knowledge repository to find the words that may retrieve the relevant and interested items or precise answers. On one hand, the vocabulary users employ in their queries may be different from the vocabulary within particular Internet resources; on the other hand, user's vocabulary may not be discriminating enough. Both cases lead to retrieval failure [2].

From the last three decades the research brings rich methods to improve results to the user query, where in PRF is one of the proven methods of query expansion for guaranteed better results. Even though the different approaches to PRF method are not in satisfactory levels, this paper introduces a novel approach to PRF method which aims to improve

precision without compromising recall growth. The proposed framework is aimed to test on Telugu item sets. Telugu is one of the Indian Languages belongs to Dravidian Family. Indian languages are rich in morphology [3]. Words in Telugu language have more variants, this identical feature is also a major problem to apply query expansion techniques. The Term selection plays a key role in retrieving relevant results for user query. In this method WordNet is used to extract candidate words for query expansion. We perform query expansion by generating lexical paraphrases of queries. These paraphrases replace content terms in the queries with their synsets. The resources used for selecting such term required Telugu POS Tagger or Morphological Analyzer, Stemmer and WordNet. This paper is organized into five chapters with introduction as Chapter 1. Chapter 2 is all about Query Expansion and the way how we use Query Expansion. Complete framework has been discussed in Chapter 3. Stop words are removed from each of these questions, and are passed to a shallow stemming program. The shallow stemmer is a rule based stemmer which stems the topic words to return the stem [4]. Use of only global models for query expansion may not affect the results. The integrated approach for query expansion gives observable growth in performance. In this paper our approach uses PRF (Local Model) with WordNet (Global Model). This provides a new platform for query expansion, which reduces the user interaction and preserves the meaning of the query. Off course using WordNet alone may not give effective results in terms of precision as the web is drastically increasing with new terminology day by day. In addition to this approach, to overcome the limitation of WordNet we also propose an online WordNet that would give updated lexical relations to the query terms. This paper aimed to find the difficulties in expanding the Telugu query and propose a solution to improve the Search results in Indian Information Retrieval (IIR). Once the framework is built the same model can be extended for other Indian Languages. Throughout the Paper examples are given in Telugu Language with phonetic representation.

## II. QUERY EXPANSION

Query expansion is a step in deriving more relevant information to the interest of end user by adding new terms to the original query. Query Expansion is one of the promising approaches to deal with the vocabulary mismatch problem in information retrieval. Queries are often not well-formulated, but may be ambiguous, insufficiently precise, or use terminology that is specific to a country. Query Expansion is a process of reformulating the root query by adding an optimal set of terms that improves recall and precision. The

Manuscript received May 22, 2013; revised July 18, 2013.

Ramakrishna Kolikipogu is with Computer Science and Information Technology, Hyderabad (e-mail:Krkrishna.csit@gmail.com).

Padmaja Rani is with a professor & HOD of CSE, Jntuhcoe, Hyderabad. N. Swapna is with VREC, Hyderabad.

motivation for query expansion is to reduce the mismatch between query and documents by expanding the query terms using words or phrases which are synonymous to query terms or share other statistical relationships with the terms contained in the set of relevant documents [5]. Major research on query expansion is proven in improving search results by short queries. Short queries generally consist of two or three terms to expand. Sometimes two to three terms are not discriminating enough to understand the expectations of end user and fail to express topic of search. Various methods are being used to expand the query. In our study, Relevance Feedback and Pseudo Relevance Feedback methods (i.e. Local Analysis) were used frequently in reformulating the initial query with the help of Global techniques. Global techniques are used to expand or reformulate the terms in independent of the initial query and resultant items in first iteration. Inducing of Global Resources like Dictionaries, Thesaurus, WordNet, Ontology and other semantic resources may cause the new query to match other semantically similar terms and used to expand the query. The problem of word mismatch is fundamental obstacle in information retrieval process.

#### A. Global Analysis

The global Analysis techniques examine word occurrences and relationships in the corpus as a whole, and use this information to expand any particular query [6]. In Global methods individual query terms will be taken to expand. Based on regular usage, new Global techniques are being added to the list. In general the global techniques use various resources to expand the initial query of user, which include “Word Dictionaries, Thesaurus, WordNet and Ontologies etc.”. Method of Use of Global resources, which are created dynamically, is a prominent technique to improve precision of the search. The study on Global Techniques shown that there is an improvement in precision by matching more concepts through repeated query expansion, but sometimes the retrieved item does not contain any query terms given by the user. This may be the one of the reason, that the use of global techniques may decline the progress in recall. Use of External resources will cause to search in out of context called query drift. To overcome this problem controlled vocabulary is required while expanding query terms by global analysis.

#### A. Thesaurus

A Thesaurus is a collection of related words based on synonyms and homonyms. There are different ways to develop a thesaurus. Automatic Thesaurus generation is one of such methods. The Thesaurus is developed automatically using co-occurring terms or grammatically related terms. The terms which co-occur frequently in a corpus are more likely to be related. The other approach is to analyze the corpus for grammatical dependencies. Handcrafted thesaurus describes the synonymous relationship between words [7]. It is expected that using of words from Thesaurus will improve precision, but it may causes some problems in expansion of Queries. E.g1: వొంఱ(viMta)/Novelty, Curiosity, wonder; oddity, oddness; an odd thing, a rarity, a thing that causes wonder, a marvel, have many forms such as ఆశ్చర్యము(AScaryamu)/surprise,wonder,

అద్భుతము(adButamu) / miracle E.g2: బండో (baMDi)/trans / cart / wagon. “ఒక(oka) చోటనుండో(cOTanuMDi) మరొక(maroka) చోటికి(cOTiki) తోసుకు(tIsuku) వాళ్ళో(veLLE) సాదనము(sAdanamu)”/Mode of Transportation. It is a transport or an instrument or a vehicle to take anything from one place to another place. వాహనము(vAhanamu)/vehicle etc. It’s observed that the thesaurus-based QE techniques have not been as successful [8].

#### B. WordNet

WordNet is manually developed lexical resource, which is also called as lexical semantic database that provides tremendous support for information retrieval. The English WordNet is developed by Princeton University to model the lexical knowledge of a native speaker of English [9]. It’s a base for all WordNets of different languages in the world. WordNet is organized around the notion of sets of synonyms (synsets) with the words with the same meaning. These synsets have different relations between them. The relation of hypernymy/hyponymy (is-a relation) is the principal relation and creates a hierarchic structure. There are also relations of meronymy / holonymy (part-of relation). WordNet is divided into different classifications by the type of word: nouns, verbs, adjectives and adverbs (i.e. Parts Of Speech (POS), some POSes are not useful in building index such as stop list: articles, prepositions and conjunctions ect. ). Query Expansion terms are selected from the correct Synset for each key term of root query. The basic Problem is that the number of synonyms for each word is excessive, this causes for poor performance. In order to overcome this problem we use word sense frequency measure of Synset.Synset is formed by grouping the synonyms with same meaning for a word. Each Synset represents one sense. Associated with each synset of a term t, there is a definition and frequency measure that indicates the extent the term t is utilized in this sense. For example, the frequency values of “బిడ్డ(biDDa)/baby” in Synset { బిడ్డ(biDDa) / baby, పాప (pApa) / baby girl } and { బిడ్డ(biDDa) / baby, బాబు (bAbu) / baby boy } word representations may vary based on the sense. Then, the noun “బిడ్డ(biDDa) / baby” is more likely to be used in the sense of “పాప (pApa) papa/baby girl” or “బాబు (bAbu) / baby boy” than the sense of “ కొడుకు (koDuku) / son” or “కూతురు (kUturu) / daughter”. The Synsets in WordNet are linked to each other through different relations including hyponyms, part of and member of etc. Whenever the sense of a given term is determined to be the Synset S, its synonyms, words or phrases from its definition, its hyponyms and compound words of the given term are considered for possible addition to the query.

#### B. Local Analysis

Local analysis considers the top ranked list of items return for the initial query as a reference to expand the query. Selection of expansion terms depends on the different techniques. Use of these techniques may vary from user to user. It depends on the domain knowledge and vocabulary usage levels of the user who write the initial query. Sometimes user may be good in discriminating the terms

related or what exactly he means to ask the search engine to retrieve. It is difficult to write such a query that includes the expert system vocabulary, then taking initial results as a reference to expand the query may stand good. The Local analysis Techniques analyzes only the information in some initial documents, which are retrieved for the original query [10]. There are two approaches in analyzing the top ranked items. Two well known methods 1) Relevance Feedback and 2) Pseudo Relevance Feedback are more in practice.

A. Relevance feedback

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 in 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 [11]. It may cause the user to endure the process. An alternate for this is Pseudo Relevance Feedback model.

III. SYSTEM ARCHITECTURE

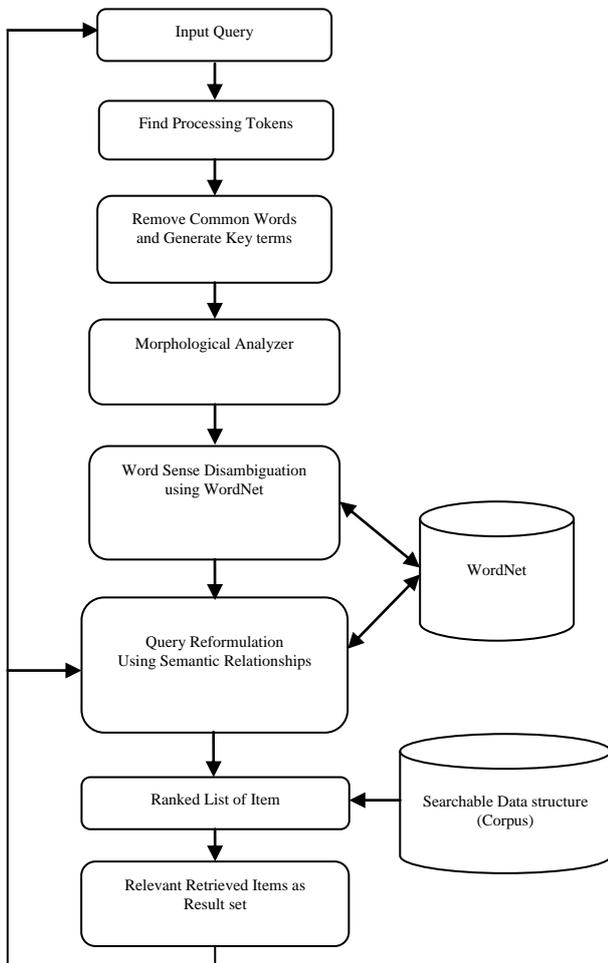


Fig. 1. Query reformulation processes using WordNet.

A. Pseudo Relevance Feedback

Pseudo Relevance Feedback method 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 [12]. 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 non-relevant, 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 [13].

IV. ITEM NORMALIZATION

There is twofold pre-processing in developing a search engine. One is pre-preprocessing input items to create searchable data structure and another one is to collect bag of words to search from user query. Expansion of the terms takes place while collecting bag of words from the initial search query.

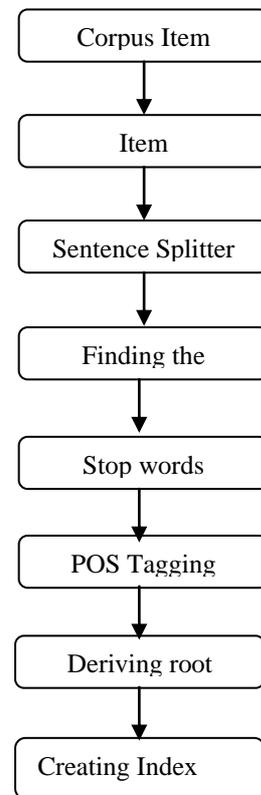


Fig. 2. Creation of searchable data structure using item normalization.

Item Normalization is a process of creating searchable data structure by the following steps:

- Standardizing the input takes the different external format of input data and performs the translation to the formats acceptable to the system.
- Translate text in different languages into any one of the standards (i.e. Unicode, ISCII, Wx, Romanization etc.).
- Divide the item into logical sub-sections that have meaning to the end user to optimize the search results in display area. It is to increase the precision of the search results.
- Find the processing tokens.
- Apply any Stop algorithm against input items to save the system resources.
- Characterize the terms of the item to disambiguate the sense. It depends on the model used by the developers.

- Select searchable terms by windowing with N-gramming technique (or) Find the root words by Stemming algorithms. In the proposed model we use Morphological analyzer to derive the root words with Word Sense Disambiguation.
- Creating Searchable Data Structure with Concept relations using WordNet.

V. QUERY EXPANSION USING WORDNET

In Fig. 1. The Query expansion can be employed to overcome word mismatch problem of the query by adding new terms, which are related to the original query in a particular context. Term co-occurrence can be used as a means to identify the related terms. Usually the normalization process is common for both query and document.

**Step 1:** When a user query is received by IR System, Text pre-processing is usually the first step to extract bag of words to match against indexing terms of the source. In this step, query normalization is done with sentence splitting, tokenization, stop word removal, characterization by using Parts of Speech (POS) Tagging *et al.*

- **Sentence Splitter:** Source items are first normalized to create index. Document is taken as input and divided in to sentences using sentence splitter based on period (dot) operator as sentence separator. If it is query, entire query is taken as single sentence and starts with Tokenization.
- **Tokenization:** The given query is divided in to tokens and they are converted to the representation that is useful Information Retrieval. This process is called as lexical analysis. E.g.: “మన ( mana)/our తొలుగు (telugu)/Telugu సాహిత్యమునకు / to the Literature వల (vEla)/thousands of సంవత్సరముల (saMvatsaramula) / years of ఘన (Gana) /great చరిత్ర ( caritra) /history ఉన్నది (unnadi)” (Telugu literature has thousands of years history). In this sentence every word separated with a space is taken as a processing token.
- **Stop word Removal:** Mostly Prepositions, conjunctions and articles may or may not have much meaning. These words are often removed during the preprocessing of the document collection. A beneficial side effect of stop word removal is that the index size may be reduced by up to 40% [14]. E.g.: మన/our, (“has”, “of” are English stop words in the above sentence).
- **Characterization:** A word may appear in several syntactical variations, such as plural, gender form and past tense with different suffixes. It causes a problem in matching key terms with its inflectional variations. Normalizing the terms by stemming or lemmatization is a solution to this problem. We use Telugu morphological analyzer for this process. E.g.: సాహిత్యమునకు (sAhityamunaku) / “to the literature”, సాహిత్యముతో (sAhityamuto) / “with the literature”, సాహిత్యముచే (sAhityamucE) / “due to literature” (Here we provided English translation to understand the meaning of the Telugu words. Telugu is morphologically rich language, which represents long sentences of English with simple words). సాహిత్యము is the root word of all above variations. And given a tagging to disambiguate the terms by Parts of Speech (POS) Tagger, where words can be found in several different syntactic variations (i.e.

noun, pronoun, verb, adjective etc.) and their meaning may be different depending on the context they appear in. One way of dealing with this is to use a Part-Of-Speech tagger [15]. E.g.: “ మన తొలుగు సాహిత్యమునకు వల సంవత్సరముల ఘన చరిత్ర ఉన్నది ”

TABLE I: POS TAGGING FOR THE ABOVE SENTENCE

POS	Word (Telugu/English)
Noun	తొలుగు/Telugu, సాహిత్యము/literature
Pronoun	మన/our
Verb	ఉన్నది/have
Adjective	ఘన/great

**Step 2:** When a key words has different meanings in different senses. We use Word sense Disambiguation (WSD) with help of WordNet to reducing complexity of Information Retrieval process. Synset is built over the WordNet. A Synset is a set of words with the same part of speech that interchangeably in certain contexts. The lexical relations built on Synset for POS are Synonymy, hypernymy/hyponymy (is-a relation meronymy /holonymy (part-of relation). Whenever the sense of a given term is determined to be the Synset  $S_i$ , its semantic relations are considered for possible addition to the query. In this approach, all the senses of a word to be disambiguated are retrieved from the WordNet. Each of these senses is then compared to the WordNet definitions of all the remaining words in context. The sense with highest overlap with these context words is chosen as the correct sense. E.g.: If **Noun** is considered as POS then Semantic relations of a word  $w$  is represented as  $\{\text{holonymy } (S_i) \cup \text{hypernym } (S_i) \cup \text{meronym } (S_i) \cup \text{hyponym } (S_i)\}$ . Similarly for other POS we apply Word sense Disambiguation. This process reduces the complexity of selecting relevant expansion terms.

**Step 3:** The final query is a Boolean combination of the original user query and the expansion terms, using the AND and OR operators. The terms selected for the initial queries could be connected before expansion with LOGICAL operators **AND** or **OR**. For example, for the original query ( $t_1$  **AND**  $t_2$ ) if we have only a term  $t_1'$  related with  $t_1$  the expanded query will be ( $t_1$  **OR**  $t_1'$ ) **AND**  $t_2$  or ( $t_1$  **AND**  $t_1'$ ) **AND**  $t_2$ . Where ( $t_1, t_2$ ) are two terms and ( $t_1', t_2'$ ) are inverse terms. This would be repeated even for every iteration of Query Expansion.

**Step 4:** Query Expansion is a process of reformulating the root query by selecting key terms in order to improve the precision and recall. Pseudo relevance Feedback is one of the proven methods of Query Expansion. Pseudo Relevance Feedback method automates the manual part of relevance feedback, so that the user gets improved retrieval performance without an extended interaction. The term weighting is updated using Rocchio Algorithm. Pseudo relevance feedback attempts to automate the manual part of relevance feedback. It retrieves an initial set of relevant documents. And assume that top  $m$  ranked documents are relevant. It is found to improve performance in TREC ad-hoc task. We propose to implement for Telugu text Documents collection.  $Q_{init}$  → Initial Query weighted term vector,  $Q_{new}$  → Expanded Query weighted term vector.  $R, R'$  are sets of relevant and non relevant documents and  $r, r'$  are term

weighting vectors extracted from  $R$  and  $R'$ , respectively. The weights in each vector are computed by a weighting scheme applied to the whole collection.

$$Q_{new} = \alpha \frac{\beta}{|R|} Q_{init} + \frac{\gamma}{|R'|} \sum_{r \in R} \gamma - \sum_{r' \in R'} \gamma' \quad (1)$$

In case of term ranking and automatic query expansion rely on a set of top scored retrieved documents, (i.e. supervisory learning) then above formula (1) is represented as:

$$Q_{new} = \alpha \frac{\beta}{|R|} Q_{init} + \frac{\gamma}{|R'|} \sum_{r \in R} \gamma \quad (2)$$

where  $R$  is the set of top retrieved documents which are assumed to be relevant.  $\alpha$ ,  $\beta$ , and  $\gamma$  are expansion quotients.  $\alpha = 1$ ,  $\beta = 1$  and  $\gamma =$  are fixed for this work. The weights obtained with Eq. (2) are typically used to both rank and reweight the query terms [16]. With due considerations of all the above steps this Automatic Query Expansion improve the performance of Information Retrieval.

#### VI. ALGORITHMIC APPROACH

**Input** : Initial Query  $Q_i$  by User.

**Output** : More Relevant Information Retrieval.

#### Procedure:

**Step1:** User enters Initial Query. From this query keywords are taken into a sets ' $Q_{init}$ ' using tokenization, stop word removal and normalization process.  $Q_{init} = \{q_{w1}, q_{w2}, q_{w3}, \dots, q_{wn}\}$ , where ' $n$ ' is no. of keywords in a query  $q$ . ' $w$ ' indicates query word.

**Step 2:** Search the key words in WordNet and based on sense and semantic relation the selection of expansion word is done.

**Step 3:** Perform Similarity measure and closeness of concepts. Consider  $N$  high relative words as pre-expansion words. Use them in query ' $q$ ' as expansion words. The resultant query is marked as  $Q_{new}$ ,  $Q_{new} = \{t_{i1}, t_{i2}, \dots, t_{iN}\}$ , where  $i=1,2,3, \dots, N$ . Goto Next Step.

**Step 4:** Reformulate expanded query as  $Q_{new} = Q_{new} + Q_{init}$ , where  $Q_{new}$  is taken as new query and submit to search engine. The retrieved result is more relevant to the user requirement.

**Step 5:** Repeat Step 2, Step 3 and Step 4 to improve the Precision and Recall.

#### VII. RESULTS ANALYSIS

A set of 15000 Telugu News text articles were used to test this work. Performance of Initial Information Retrieval System for 150 queries was calculated using precision and recall measures. Fig. 3 shows the difference in performance of Information Retrieval System with and without Query Expansion Model. In general recall will get increase with and precision will drop with query expansion. From the proposed model, we combined both Pseudo Relevance Feedback and

WordNet terms for expanding the query. It is interesting that the recall has not affected by this method, where as precision is greatly increased when compared to initial search on 15000 Telugu document collection in an average from 150 topics or queries. Around there is 12.59% growth in Precision. Vector Space Model with TF-IDF based weighing methods are used to represent documents and queries. Cosine similarity is used to find the relevance of the item to the given query and Ranked accordingly.

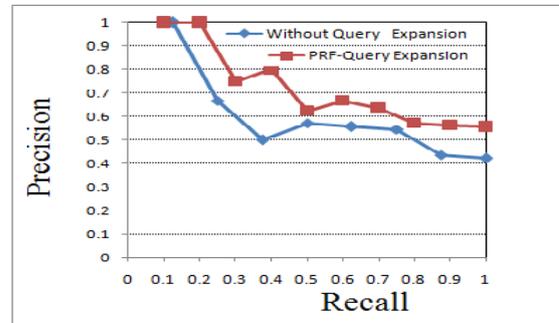


Fig. 3. Precision and recall trade-off with and without query expansion

#### VIII. CONCLUSION AND FUTURE SCOPE

Query Expansion is an already proven method of Information Retrieval for Internet and other information Resource browsing. From the long back the internet and other information resources are available in English and other European languages. Due to education and usage of digitized information in different domains there is a necessity to adopt the IR techniques to other languages, But the complexity of languages like Indian languages required more care to preprocess the information resources. This is the one of the reason that the effect of the retrieval approaches are not implemented and tested in most of the other languages. India is one of the statistically rich countries in internet usage in Asia. There is a lot of scope for the researchers to adapt new retrieval methods that improves precision and recall. Information Retrieval in Indian languages is in a nascent stage. Telugu is one of the Indian languages and 2<sup>nd</sup> biggest spoken language in India. We investigate the effect of Pseudo Relevance Feedback model for Telugu language. We observe the results for the above framework is positively scaled when we tested manually. As part of our research work in Indian Information Retrieval, we proposed to atomize the process and plot the results. There is no Telugu WordNet like Princeton WordNet for English, which initiated us to work in this direction by taking Hindi WordNet developed by IIT Bombay as reference. The online support for the WordNet is also proposed to enhance the performance of Information Retrieval System.

#### REFERENCES

- [1] X. Y. Wang, *User Ontology and Word Sense Disambiguation for Query Expansion*, presented at ICCASM , 22-24 Oct. 2010.
- [2] I. Zukerman, *Bhavani Raskutti Query Expansion and Query Reduction in Document Retrieval*, presented at ICTAI03ZRW, Australia, 18-Mach, 2010.
- [3] D. Patel and D. P. Madalli, "Information Retrieval in Indian Languages: A Case Study of Plural Resolution in Telugu Language," Presented in ICSD-2007, pp. 397-404, 2007.
- [4] J. Jagarlamudi and P. Pingali, "A Dictionary Based Approach with Query Expansion to Cross Language Query Based Multi-Document

Summarization: Experiments in Telugu – English,” presented at National Workshop on Artificial Intelligence, June 2006, Mumbai, India.

- [5] J. Rocchio, “Relevance feedback in information retrieval,” in *The Smart Retrieval System: Experiments in Automatic Document Processing*, G. Salton, Ed. Prentice-Hall, Englewood Cliffs, NJ, pp. 313–323.
- [6] J. X. Xu, “Croft, Query Expansion using Local and Global Document Analysis,” SIGIR96, Switzerland, 1996 ACM, pp. 4–11.
- [7] R. Hemayati, W. Y. Meng, and C. Yu, “Semantic-Based Grouping of Search Engine Results Using WordNet,” APWeb/WAIM’07, 2007, pp. 678–686.
- [8] R. Mandala, T. Tokunaga, and H. Tanaka, “Combining multiple evidence from different types of thesaurus for query expansion,” in *Proc. ACM-SIGIR Annual Int. Conf. on Research and Development in Information Retrieval*, Berkeley, California, 1999, pp. 191 – 197.
- [9] C. Fellbaum, *WordNet: An Electronic Lexical Database*, the MIT Press, Cambridge, London, England, May 1998, pp. 445.
- [10] Z. G. Gong, C. W. Cheang, and L. U. Hou, “Web Query Expansion by WordNet,” in *Proc. 16th International Conference, DEXA 2005*, Copenhagen, Denmark, August 22–26, 2005, pp. 166 – 175.
- [11] K. Ramakrishna and B. P. Rani, “Information Retrieval in Indian Languages: Query Expansion models for Telugu language as a casestudy,” IITA2010, Qinhuangdao, china, Nov. 5-7, 2010, pp. 127-131.
- [12] Y. Xu and G. J. F. Jones, “Query Dependent Pseudo Relevance Feedback based on Wikipedia,” SIGIR’09, July 19–23, 2009, Boston, Massachusetts, USA.
- [13] C. Buckley, G. Salton, J. Allan, and A. Singhal, *Automatic query expansion using SMART: TREC-3*, 1994
- [14] B. Yates and R. Neto, *Modern Information Retrieval*, ACM Press / Addison-Wesley, 1999.
- [15] C. Rocha, D. Schwabe, and M. P. de Arag˜ao, “A hybrid approach for searching in the semantic web,” in *Proc. the 13th international conference on World Wide Web*, 2004, pp. 374 – 383.
- [16] P. Srinivasan, “Query expansion and MEDLINE,” *Inf. Process. Manage*, vol. 32, no. 4, 1996, pp. 431–443. V. Rijsbergen, C. J., “A theoretical basis for the use of co-occurrence data in information Retrieval,” *J. Doc.* vol. 33, no. 2, June 1977, pp. 106–119.



**Ramakrishna Kolikipogu** received B.Tech. in Computer Science and Information Technology from VCE-Warangal, JNT University. M.Tech. in Software Engineering from JNTU Kakinada and he is Pursuing Ph.D. in the area of Information Retrieval Systems from the Department of Computer Science and Engineering, JNT University Hyderabad. He has 10 years of teaching experience in various engineering colleges. Currently

he is working as Associate Professor & Head, Dept. of Information Technology, Sridevi Women’s Engineering College, Hyderabad, India. To his credit Mr.K.Ramakrishna has 25 publications in various National / International Conference and Journals. He is also a Member of Various Technical Bodies including IEEE, ACM, CSTA, ISTE, CSI, IACSIT, IAENG, AIRCC etc., He has reviewed more than 100 Research papers for various International Journal IJCS, IJCL, IJCCT, IJCSI, IJECCE and Conferences, such as WARSE, NCRTCST. He is member Editor of IJECCE, IJETAE and Guest Editor of Interscience Open Access Journals. His area of interest includes Information Retrieval, Text Mining, Web Mining, Machine Learning, Information Security etc.



**B. Padmaja Rani** received B.Tech Electronics Engineering from Osmania University, M.Tech in Computer Science from JNT University Hyderabad, India and she has been awarded Ph.D. in Computer Science from JNT University, Hyderabad, India. At present she is working as Professor and Head, Department of Computer Science and Engineering, JNTUH College of Engineering, JNTU University

Hyderabad. She is having 18 years of experience in Industry and Academia. At present she is a Professor of Computer Science and Engineering Department in JNTUH College of Engineering, JNT University, Hyderabad. Her area of Research includes Information Retrieval, Data Mining, Machine Translation, Computer Networks, Software Engineering etc. She is guiding 6 Research Scholars in the area of Information Retrieval and Computer Networks. To her credit she is having more than 60 publications in reputed International Journals and Conferences. She is a member of various advisory committees and Technical Bodies. She is also a Member of Various Technical Associations including ISTE, CSI, IEEE etc.