

Hindi Query Expansion based on Semantic Importance of Hindi WordNet Relations and Fuzzy Graph Connectivity Measures

Amita Jain¹, Sonakshi Vij², Oscar Castillo³

¹ Ambedkar Institute of Advanced Communication Technologies and Research,
Department of CSE,
India

² Krishna Engineering College, Department of CSE,
India

³ Tijuana Institute of Technology, Division of Graduate Studies and Research
Mexico

ocastillo@tectijuana.mx

Abstract. Query expansion refers to the process of adding terms to a given query for improving the performance of information retrieval (IR). The query might consist of polysemous terms, which usually bring down the overall IR performance. To resolve this issue and perform optimized IR, we propose an approach based on fuzzy graphs for Hindi query expansion. To identify additional terms for query, we consider the relative semantic importance of the relations present in Hindi WordNet. The query is represented by the sub-graph extracted from the Hindi WordNet graph. Hindi WordNet is semantically richer due to the presence of a greater number of semantic relations as compared to other WordNets. For all 16 semantic relations present in Hindi WordNet a relative significance score proportional to semantic relatedness is provided. This score acts as the edge weights to the Hindi WordNet graph which is now represented as a fuzzy graph. This assignment helps in moving more semantically related words, closer and recedes away less semantically related words in Hindi WordNet. The selection of significant terms that are to be used for query expansion is done by using local and global fuzzy graph connectivity measures. The proposed method is evaluated on the Forum for Information Retrieval (FIRE) dataset for 3 consecutive years which depicts that the proposed method provides better results than the state-of-art approaches.

Keywords. Fuzzy graph connectivity measures, information retrieval, natural language processing, query expansion, word sense disambiguation.

1 Introduction

Information retrieval (IR) deals with the organization, storage, retrieval and evaluation of information relevant to users' query (Manning et al. 2008, Grossman and Frieder 2004, Salton and McGill 1998). The availability of a large amount of text in electronic form has made it difficult to easily get relevant information. The user submits his/her query to the system in the natural language. Therefore, the system should be able to process and understand the query written in a natural language. This requires building computational models with human language processing abilities, such as knowledge about how human acquire, store and process language.

The major goal of information retrieval is to search a document in a manner relevant to the user query (Tanveer and Tiwari 2008). Users enter the query in a natural language, which generally has impreciseness, vagueness, ambiguity and uncertainty. Many times, the relationships between the terms present in the user query and the terms present in the documents on the web are not precise. These relationships are often approximate.

Despite a high degree of success, the World Wide Web has introduced new problems of its own.

In this context, many times the characterization of information needed by a user is not a simple task. For instance, to satisfy his/her information needs, the user might navigate the space of web links, in a quest to search for information of his interest. However, since the cyberspace is extremely large and almost unknown, such navigation becomes difficult and inefficient. Currently, most of the interactive software systems make exact matches between textual user input and predefined system names of data or commands. Therefore, in such systems the user is bound to adapt to the naming requirements of the system. For a more cooperative behavior, the system should not take the user's input literally. Rather to identify what the user intends; the system should consider the semantic relationships between the concepts of the application domain to find the most likely intended interpretation. Generally, the queries input by users contain terms which do not match the terms used to index the majority of the relevant documents and also, sometimes the un-retrieved relevant documents are indexed by a different set of terms than those in the query or in most of the other relevant documents. In order to solve this problem and to improve the query performance, it is therefore necessary to modify the user's query.

Query expansion is the process of adding additional terms to the original query in order to improve retrieval performance (Manning et al. 2008). It is very difficult to embody all sources of knowledge that human use to process a language. Perhaps the greatest source of difficulty in query expansion is identifying a query's semantics.

Query expansion is often effective in increasing recall, but query expansion may also significantly decrease precision, particularly when the query contains ambiguous terms (Manning et al. 2008).

If we can develop a method which embodies the sources of knowledge that human beings use to process a natural language sentence, then the query expansion process may be enhanced. Since the query framed by user is always in natural language, which may be ambiguous/ imprecise/vague/uncertain, it adds additional problems to the system, as many words can be interpreted in multiple ways depending on the context in which they occur.

In Hindi language, ambiguity creates more problems, as a large amount of the Hindi words

have multiple senses. For instance, the commonly used Hindi words like "मिलना" (milanā) has 21 senses, "खाना" (khānā) has 19 senses, "पाना" (pānā) has 13 senses, "दाल" (dāl) has 12 senses, "बैठना" (baiṭhanā) has 12 senses, "बांधना" (bāndhanā) has 12 senses and the word "अच्छा" (acchā) has 11 senses mentioned in Hindi WordNet. Assume a scenario, where the user frames a query containing 5 words, out of which 3 words are ambiguous. Further, assume that on average, every ambiguous word has 4 possible senses, and then the total number of interpretations of the query is $4^3 = 64$, which is quite large. Hence it is necessary to expand the query in the direction of the correct interpretation of the query. This gives the motivation for the present work, as many times the expanded query can add unintended terms which rather degrade the IR performance, but our approach overcomes this problem. Further, query expansion methods provide high recall but at the cost of low precision (Manning et al. 2008). The reason behind low precision is only the ambiguous nature of the query (Manning et al. 2008). The approach proposed in the present paper focuses on ambiguity resolution thus eradicating the cause of low precision.

In this context we propose a Hindi query expansion method which also disambiguates the ambiguous query terms using fuzzy Hindi WordNet. The method enhances the sense disambiguation and query expansion process by considering the semantic importance of lexical and semantic relations present in Hindi WordNet; thereby assigning strength to each relation. Here we use Hindi WordNet because of its rich link structure.

In Hindi WordNet all the concepts/word-senses are related with each other through various relations like hypernymy, hyponymy, meronymy, holonymy, entailment, troponymy and various parts of speech linkage etc. These relations play an important role in representing concepts/word senses in a semantically enriched way. This representation of words helps in extracting the appropriate terms for query expansion.

In the literature, Hindi WordNet is widely used in many NLP applications and viewed as a graph where a node represents a concept (word sense) and an edge represents any of the relation

between the concepts (Sinha et al. 2004, Dwivedi and Rastogi 2008, Das et al. 2010, Sharan et al. 2011, Jain et al. 2013, Jain et al. 2014).

In all these works, the researchers considered all the relations defined in Hindi WordNet as equally important. But in the real world, some relations can help in describing a given concept in a more reasonable way than the other. For instance, the words "वाहन" (vāhana, vehicle) and "कार" (kāra, car) share hypernymy/hyponymy relation. "कार" (kāra, car) is composed of an "इंजन" (iñjana, engine), "पहिया" (pahiyā, wheel) and "सीट" (sīṭa, seat) so the word pairs (कार (kāra, car), इंजन (iñjana, engine)), (कार (kāra, car), पहिया (pahiyā, wheel)), (कार (kāra, car), सीट (sīṭa, seat)) shares meronymy/holonymy relations. The word "कार" (kāra, car) can be described in a better way by using "वाहन" (vāhana, vehicle) rather than by using "इंजन" (iñjana, engine), "पहिया" (pahiyā, wheel) or "सीट" (sīṭa, seat).

Therefore, if hypernymy/hyponymy relations are given more importance than meronymy/holonymy relations, then the given concept may be understood in a better way. In a similar way the word pair (सामान (sāmāna, material), वस्तु (vastu, thing)) shares Hypernymy/Hyponymy relation and the word pair (सामान्य (sāmānya, ordinary), वस्तु (vastu, thing)) shares Modifies Noun relation (a cross parts of speech relation). The word "वस्तु" (vastu, thing) can be described by using the word "सामान" (sāmāna, material) in a more convenient way, than by using the word "सामान्य" (sāmānya, ordinary).

In this paper we address this relative importance of Hindi WordNet relations among words by assigning a strength measure $\mu \in [0, 1]$, to each relation. In our approach, initially, we build a graph for the given query using Hindi WordNet.

Here vertices represent words and edges represents relations (defined in Hindi WordNet) between words (synonym sets).

Now by assigning the strength to the graph edges according to the semantic importance of the relations we convert it into a fuzzy graph.

Our aim is to identify the important vertices in the fuzzy graph created for query other than those representing the query terms. To do so, we apply local connectivity measures (viz. Degree Centrality, Eigen Vector Centrality (PageRank, HITS (Hub, Authority)), Closeness Centrality, Betweenness Centrality) and global connectivity measures (viz. Entropy, compactness, edge density) on the fuzzy graph constructed (Jain and Lobiyal 2016) and find additional terms for the Hindi query.

The aforesaid method considers intended interpretation of the user while performing query expansion and it expands the query by adding unambiguous terms. Query expansion is often effective in increasing recall but in parallel it may also significantly decrease precision, particularly when the query contains ambiguous terms. The proposed method provides high recall, it also prevents decrease in precision problem. Since the semantic relations present in the Hindi WordNet are different from those present in the English WordNet, hence this paper proposes a method for Hindi query expansion only.

The rest of paper is organized as follows: In the next section we discuss related work. In the third section we analyse and compare the semantic importance of various Hindi WordNet relations. The proposed method i.e. Hindi query expansion using fuzzy graph connectivity measures is discussed in the next section after that. In the fifth section we present our Experimental Setup and Results. Conclusions and future work are presented in the last section.

2 Related Work

Fuzzy set theory (Zadeh 1965) is widely applied in information retrieval. Initially in 1975 Rosenfeld (Rosenfeld 1975) reviewed the basic properties of fuzzy relations and generalized to the case where the underlying set is a fuzzy set.

In 1987 Bhattacharya explored Fuzzy Graphs and indicated that results from (crisp) graph theory do not always have analogy for fuzzy graphs (Bhattacharya 1987).

Fuzzy analogues of several basic graph-theoretic concepts (e.g., bridges and trees) are also defined by him. In 2001, M.S. Sunitha

(Sunitha 2001) proposed the concepts of fuzzy bridges, fuzzy cut nodes, fuzzy trees, blocks and metric in fuzzy graphs defined by Rosenfeld. She also modified the definition of the complement of a fuzzy graph. The concept of connectivity plays an important role in both theory and applications of fuzzy graphs. Recently, fuzzy graphs are being used by R. Yager to represent the relations in social networks (Yager 2010). Very recently fuzzy graph connectivity measures are proposed and used for word sense disambiguation (Jain and Lobiyal 2016). Graph connectivity measures have been studied extensively in the social sciences, especially within the field of Social Network Analysis (Freeman 1979; Wasserman and Faust 1994).

Quantifying centrality and connectivity help in analyzing the characteristics of the structure and properties of large networks and is also used to make predictions about their behavior. Word sense disambiguation (WSD) is an important task in computational linguistics as it is essential for language processing applications, such as machine translation, information retrieval, question answering and text summarization etc. Navigli divided the methods of word sense disambiguation in his survey in three subcategories namely supervised, unsupervised and knowledge-based methods (Navigli 2009).

Through a series of experiments (Hai-Tao Zheng et al. 2009) show that the effectiveness of semantic relationships for clustering are (from highest to lowest): hypernymy, hyponymy, meronymy and holonymy. Many graph-based algorithms are proposed for WSD. Researchers proposed graph-based algorithms for large-scale WSD (Navigli and Lapata 2010). Authors used the measures of graph connectivity from social network analysis like degree centrality (Freeman 1979), key player problem (Borgatti 2003), HITS (Kleinberg 1998), PageRank (Litvak et al. 2006) and Betweenness centrality (Newman 2005) for Word Sense Disambiguation in Hindi language (Jain et al. 2013).

Voorhees used WordNet for query expansion by adding synonyms to the original query for expansion (Voorhees 1994). Rila et al. presented a co-occurrence-based method for making WordNet more useful in Information Retrieval (Rila et al. 1998). Kim et al. proposed a query term

expansion and reweighting method which considers the term co-occurrence within the feedback documents (Kim et al. 2001). In 2003 Christopher et al. found out that word sense disambiguation in information retrieval improves its performance (Christopher S et al. 2003). In 2004 initially Hindi WordNet developed at IIT Bombay is being used for Hindi word sense disambiguation (Sinha et al. 2004). Various approaches have been used for expanding queries using automatically derived thesaurus which were basically used in domain-specific search engines.

WordNet and Term Semantic Network were used in combination for query expansion (Gong et al. 2005). Shuang et al. determined the senses of words in queries by using WordNet (Shuang et al. 2005). In his approach, noun phrases present in a query were strengthened by including its synonyms, hyponyms and hypernyms. Cao et al. captured both direct and indirect term relationships for query expansion through external knowledge sources such as ontology and statistical processing of the document corpus respectively as independent usage of the sources showed minimal improvement in retrieval performance (Cao G. et al. 2005) (Collins-Thompson K. et al. 2005).

Cao et al. integrated two types of relationship viz. extracted from WordNet and co-occurrence relationships for information retrieval (Cao G. et al. 2005). The further categorization of query expansion approaches was given as extensional, intentional, or collaborative ones (Grootjen et al. 2006). The first one materializes information need in terms of documents, for instance relevance feedback and local analysis methods. The second i.e. intentional approach which take advantage of the semantics of keywords, is primarily thesauri/ontology based. Collaborative approaches are focused towards exploiting users' behavior, e.g., mining query logs, as a complement to previous approaches. Bhogal et al. has done a review of ontology-based query expansion (Bhogal et al. 2007).

The authors analyzed various query expansion approaches which include relevance feedback, corpus dependent knowledge models and corpus independent knowledge models.

The lack of test collections containing ambiguous queries is highlighted and a method for

creating collections from existing resources is described in (Mark 2008).

Recently authors also proposed a new ontology Based Query Expansion (Barathi and Valli 2010). Wang et al. presented a term-reweighting method for query expansion (Wang et al. 2010). Researchers presented methods for improving English-Hindi cross-lingual information retrieval System (Das et al. 2010, Varshney and Bajpai 2013). Researchers proposed a part-of-speech tagging of multi-category words in Hindi language method using rough sets (Gupta et. al. 2011). Roi and Christina presented a principled graph-theoretic approach of computing term weights and to integrate discourse aspects into retrieval (Roi and Christina 2012).

Recently, Sanasam et al. presented a query expansion framework which explores user's real-time implicit feedback provided at the time of search to determine user's search context and identify relevant query expansion terms (Sanasam et al. 2013). Another technique for query expansion termed as automatic query expansion aimed at augmenting the original query terms with new features having similar meaning (Claudio & Giovanni 2012). Vaidyanathan et al. presented a query expansion method based on qui-width and equi-frequency partition (Vaidyanathan, et al. 2013). Researchers developed a semantic-based content mapping mechanism for an information retrieval system. This approach employs the semantic features and ontological structure of the content as the basis for constructing a content map (Pai et al. 2013). Query recommendation technology that suggests a list of related queries is presented to resolve short and ambiguous nature of queries (Song et al. 2014). A knowledge-based document representation approach was used to expand the terms in the document by using concepts and semantic relations between them (Franco et al. 2014). Authors have also used WordNet to automatically incorporate context meaning for English Query Expansion (Jain et al. 2014).

In this paper, authors have also considered the problem of ambiguity in the query, but they assigned equal importance to all the relations present in Hindi WordNet.

Very recently the concept of assigning significance to semantic relations hypernymy,

hyponymy, meronymy, holonymy and derivationally related forms has been discussed for word sense disambiguation (Vij et al. 2018). But this paper has a limitation that only a few semantic relations are being considered. Also, it is worth mentioning that the relations of Hindi and English WordNet are different. In the present paper we have assigned significance to all the 16 semantic relations in the Hindi WordNet. The assignment of strength is according to the semantic importance of relations which helps in improving the system performance.

3 Analyzing and Comparing the Semantic Importance of Various Relations in Hindi WordNet

In the literature, researchers have considered Hindi WordNet as graphs where nodes represent concepts/SynSets and edges represent any of the relations between concepts/SynSets. Here, the relation between any two concepts may be any of the relations defined in Hindi WordNet. In the literature most of the researchers have considered that the various relations existing between SynSets have equal semantic strength. In the present section we propose the notion that these relations are not equally semantically strong but different types of Hindi WordNet relations possess different degree of semantic strength. Here, we propose to assign the strength to a relation according to its semantic richness.

However, for query expansion we represent the user query as a subgraph of Hindi WordNet where edges represent any of the relations defined in Hindi WordNet (and not only the above four relations discussed by Hai-Tao Zheng et al.). Therefore, it is necessary to assess the semantic strength of all relations in Hindi WordNet. Now we compare the semantic strength of various relations so that the weight can be assigned to each relation depending upon the semantic strength.

This assigned weight $\mu \in [0,1]$ can be used to bring closer the words to the query terms which are connected by a semantically rich Hindi WordNet relation and to move farther away the words from the query terms which are connected by less semantically strong Hindi WordNet relations.

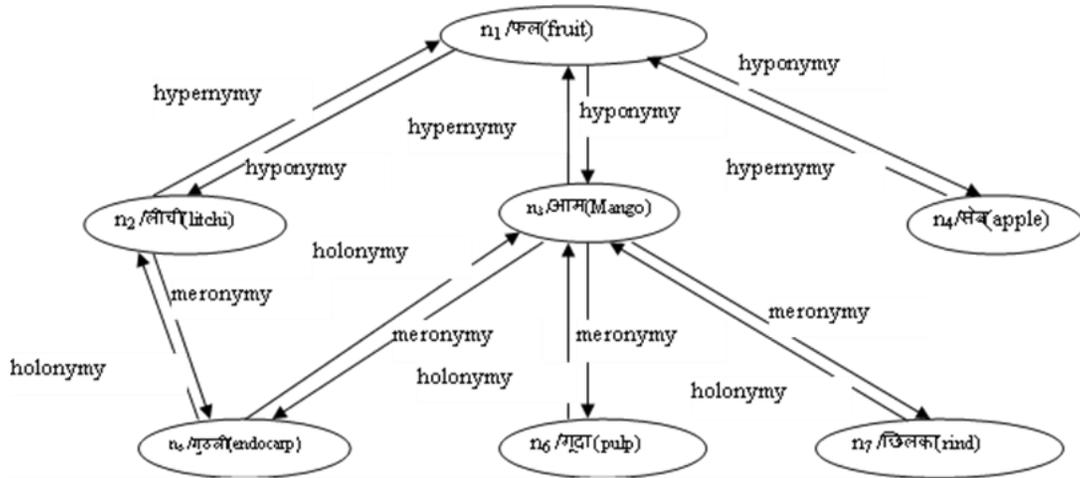


Fig. 1. Various relations for "फल" in Hindi WordNet

Let us now discuss an illustrative example to understand the variation in the semantic strengths of various relations existing in the Hindi WordNet.

Let $n_i(1 \leq i \leq 7)$ are words (SynSets) in Hindi WordNet such that:

n_1 has hyponymy relation with n_2, n_3, n_4 and n_2, n_3, n_4 have hypernymy relation with n_1 .

n_3 has meronymy relation with n_5, n_6, n_7 and n_5, n_6, n_7 has holonymy relation with n_3 .

n_2 has meronymy relation with n_5 and n_5 has holonymy relation with n_2 .

Fig.1 represents an instance set of $n_i(1 \leq i \leq 7)$. A noun usually has a single hypernym and hypernymy is a generalization relation (Fellbaum 1998) therefore there exist a strong semantic connection between a word and its hypernym. The hypernymy relation from n_2 to n_1 states that n_1 is a generalized concept of n_2 . In Fig.1 "फल" (phala, fruit) is a hypernym of "आम"(āma, mango).

Therefore, if a query has terms लीची (līcī, litchi), आम (āma, mango) etc., then its hypernym फल (phala, fruit) can be added to the query with the strong validity. Moreover, a noun usually has multiple hyponyms as hyponymy is a specialization relation (Fellbaum 1998). The hyponymy relation from n_1 to n_2 states that n_2 as a specialized concept of n_1 . In fig.1 "आम" (āma, mango) is a hyponym of

"फल" (phala, fruit). If a query has a term फल (phala, fruit) then फल (phala, fruit) is highly semantically related to लीची (līcī, litchi), आम (āma, mango) and सेब (sēba, apple) etc. Although, from the user query containing a term "फल" (phala, fruit) it can't always be decided judiciously whether the query is intended towards लीची (līcī, litchi), आम (āma, mango) or सेब (sēba, apple). Therefore, the semantic strength of hypernymy is always greater than the semantic strength of hyponymy.

Secondly, note that if a query contains a noun as a query term and this noun is composed of many parts then adding these parts to the query would lead to finer results.

It would help to describe the query term in more detail which may further improve the search results. For instance, if a query is having the term n_3 i.e. आम (āma, mango) we can expand the query by adding its parts n_5, n_6, n_7 i.e. गुठली (guṭhālī, endocarp), गूदा (gūdā, pulp), छिलका (chilakā, rind). This would help to strengthen the query so as to retrieve finer results. On the other side, Holonymy is a "part of" relation. The holonymy relation from n_5 to n_3 convey the semantics that n_5 is a part of n_3 .

In fig.1, आम (āma, mango) is a holonym of गुठली (guṭhalī, endocarp) and लीची (līcī, litchi) is also a holonym of गुठली (guṭhalī, endocarp). So, in a query having term गुठली (guṭhalī, endocarp), it can't always be decided judiciously whether the query is intended towards लीची (līcī, litchi) or आम (āma, mango). Therefore, the semantic strength of the holonymy relation is always lesser than the semantic strength of the meronymy relation.

Further, note that if two nouns are associated with hypernymy/hyponymy relation it means the nouns share features of meaning (Fellbaum 1998), so these words mostly share common properties, common attributes, and similar characteristics and also belong to similar class in the real world. On the other hand, if two nouns share meronymy/holonymy relation it means one noun may be composed of different types of nouns. These different types of nouns generally belong to different types of semantic diversity.

For instance, a "कार" (cār, car) may be composed of a "इंजन" (īñjana, engine), "पहिया" (pahiyā, wheel) or "सीट" (sīṭa, seat). It may be noted that these components are quite semantically diversified and therefore the semantic strength of the relation (meronymy/holonymy) diminishes.

Therefore, it may be concluded that semantic strength of hypernymy/hyponymy relation is always greater than the semantic strength of meronymy/holonymy relation. In Hindi WordNet, troponymy relation (which exists between two verbs only) denotes a specific manner elaboration of another verb. It shows manner of an action, *i.e.*, X is a troponym of Y if *to X is to Y* in some manner, for example मुस्कुराना (muskurānā, to smile) is a troponym of हँसना (h̄ṣanā, to laugh).

According to Miller (Miller 1995), troponymy (manner-name) is for verbs what hyponymy is for nouns. Therefore, we assign equal strength to troponymy relation and hyponymy relation.

Further, in Hindi WordNet, entailment refers to a relationship between two verbs. Any verb A entails B, if the truth of B follows logically from the truth of A. The relation of entailment is unilateral, *i.e.*, it is one-way relation.

For example, खर्राटलेना (kharrāṭalēnā, to snore) entails सोना (sōnā, to sleep). The entailment relation between verbs resembles meronymy between nouns but meronymy is better suited to nouns than to verbs (Fellbaum 1998). Therefore, we assign equal strength to the entailment relation and meronymy relation.

Cross parts of speech linkages contain less information so these relations may be assigned minimum strengths as compared to all of the above-mentioned relations. Novischi (Novischi 2004) mentions that since WordNet does not contain gloss relation, it could be explicitly induced in a heuristic way. A pair of synsets S and S' is connected via a gloss relation if an unambiguous word $w \in S'$ occurs in the gloss of S.

Since a gloss relation is not defined in Hindi WordNet and can only be derived heuristically, therefore we assign minimal weight to it. Antonymy is a relation that holds between two words that (in a given context) express opposite meanings. Gradation is a lexical relation. It represents the intermediate concept between two opposite concepts. Antonymy and gradation relations are related to opposite meaning of the concepts so these relations may be assigned zero strength for query expansion.

Therefore, the comparative semantic strength (also called weight $w_r \in [0, 1]$) to each relation is assigned as follows:

$$W_{\text{hypernymy}} \geq W_{\text{hyponymy}} = W_{\text{troponymy}} \geq W_{\text{entailment}} \quad (1) \\ = W_{\text{meronymy}} \geq W_{\text{holonymy}} \geq W_{\text{cp}} \geq W_g.$$

w_{cp} and w_g represents the weight assigned to cross parts of speech linkage and gloss words respectively.

Once the strength to each relation is assigned, the Hindi WordNet could be viewed as a directed fuzzy graph, which can be used to extract the additional terms for the query. But to identify the additional and appropriate nodes/words out of all the candidate nodes to be added to the query, we need fuzzy graph connectivity measures.

4 Query Expansion using Hindi WordNet Relations and Fuzzy Graph Connectivity Measures

In this section we propose the algorithm for Hindi query expansion. We start by building the fuzzy graph which is called the query graph, $G = (V_i, E_i)$, corresponding to the given query Q using the reference lexicon, Hindi WordNet (please refer to step1 in Algorithm1). As discussed previously, the nodes in the graph are word senses and the edges represent the semantic relations. Our query expansion method uses Depth First Search (DFS) as edges are explored out of the most recently discovered vertex v which still has unexplored edges connecting it. When all edges connecting to node v have been explored, the search backtracks to explore edges leaving the vertex from which v was discovered. This practice continues until we have discovered all the vertices which are reachable from the original source vertex.

The use of depth-first search is encouraged by computational efficiency. However, there is nothing intrinsic in our formulation that restricts the method to DFS. For instance, we could have used breadth-first search (BFS) also. Now we assign the strength to each edge (relation) according to the semantic importance of that relation associated to it (discussed in Section 5). If we take the threshold value (δ) to be very less, the words/nodes in the query become disconnected when query graph is created, thus unable to identify the terms for expansion. Taking a high value for the threshold results in adding irrelevant nodes in the query graph thus degrades the performance. So, a mediocre value is taken. From the experimental study it is found that, taking its value as 6 provides better results.

On the constructed query graph, we apply all local connectivity measures to analyse the importance of each node (word sense) in the query graph. In this paper we considered the four local measures namely Degree Centrality, Eigen Vector Centrality (PageRank, HITS), Closeness and Betweenness. The weight of each node is defined by the average of the four measures. We have given equal weightage to all the local measures because each measure has its own pros and cons mentioned below.

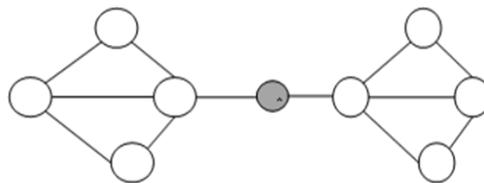


Fig. 2. Structure of graph where degree centrality fails to capture the ability to broker between groups

Degree Centrality is one of the simplest measures and widely being used in many NLP applications but in many situations, degree fail to capture the importance of a node. Consider the example in Fig. 2. The degree centrality of the shaded node fails to capture the ability to broker between groups and the likelihood that information originating anywhere in the network reaches this node.

The shaded node may be given good importance by using any of the Betweenness or Closeness Centrality. If the query graph is approximately fully connected, Betweenness centrality of all the nodes have approximately equal weightage, thus fail to capture the importance of the desired node. Closeness centrality may be helpful to capture the importance of the node in this case. Unlike other centrality measures, eigenvector centrality acknowledges that all connections are not equally important. This feature helps to measure the context of the query words and thus identifying the additional potential words for the query. Although in many cases senses of the same word are linked to each other, those senses recursively reinforce each other and thus both receive higher ranks (E. Agirre and A. Soroa 2009). Considering the pros and cons of all local measures viz. Degree centrality, Eigenvector Centrality, Betweenness, Closeness Centrality, we have given equal importance to all local measures.

In the original form of PageRank, the sum of PageRank over all pages was the total number of pages on the web at that time, so each page has an initial value of 1 (Brin and Page 1998).

However later versions of PageRank assume a probability distribution between 0 and 1. Researchers (Mihalcea et al. 2004) initiate each node in the graph with arbitrary values between 0 and 1 of PageRank.

PageRank is iterated until convergence below a given threshold is achieved. In (Esuli and Sebastiani 2007), authors have assigned initial value for PageRank as $1/(\text{total number of pages})$.

From the experimental study we observed that the converged value of resulting PageRank is dependent upon the initial selected value of PageRank though the ranking of pages/nodes remain same irrespective of the initial PageRank values assigned. We initialize the PageRank value as $1/(\text{total number of pages})$. Gupta et al. explained how HITS algorithm works and perform well. We initialize the HITS values as 1. The key details necessary for the interpretation of convergence of HITS can be explained as follows:

Let x represents a hub value and y represents authority value. The weights for hubs and authorities are modified according to the simple operation, $x = A^T y$ and $y = Ax$ therefore $x = A^T Ax$, similarly $y = AA^T y$.

The iteration therefore converges to the principal eigenvector of AA^T . We identify the nodes (except the nodes which represent the original query terms) which have high value of average centrality measures (please refer step2). The word senses exhibited by these nodes represent the additional words for the given query. It may be noted that additional words identified for the query are unambiguous (having only one sense). To disambiguate the original words, present in the query, we use global connectivity measures for the fuzzy graph.

The expanded query is ambiguous, to disambiguate the expanded query, we construct the interpretation graph corresponding to each interpretation of the expanded query (please refer step 3). Let each $t_i \in Q$ ($1 \leq i \leq n$) has m_i senses defined in Hindi WordNet. The given query has $I = \prod_{i=1}^n m_i$ interpretations. For each interpretation we construct the interpretation graph $G_i = (V_i, E_i)$, $G_i \subseteq G$ ($1 \leq i \leq I$). Interpretation graph having highest average value of global connectivity measures would represent the intended interpretation of the given query. Thus, the set $(V_Q \cap V_{final}) \cup V_{lcm}$ obtained in step4 represents the disambiguated terms for the expanded Hindi query (please refer algorithm1 for the formal procedure). Query expansion methods provide high recall, but at the cost of low precision (Manning et al. 2008).

The reason behind low precision is only the ambiguous nature of the query (Manning et al. 2008).

The approach proposed in the paper focuses on ambiguity resolution thus eradicating the cause of low precision. The basis of the proposed query expansion method is to investigate the query graph (which is being created for the query) by using the various fuzzy graph centrality measures. Based on various centrality measures the most related nodes/words (representing word senses) to the query words are identified and chosen as additional terms to the query. Graph structural properties (various global connectivity measures) are used to disambiguate the expanded query.

Let us illustrate the proposed method by considering different queries having polysemous word "फल" (phala, Fruit/Blade/solution/result/consequence/board/shield/interest/nutmeg) which has 9 different senses defined in Hindi WordNet.

Let us consider the following two user queries (preprocessed for stemming, stop word removal etc.):

Query1: "प्रौद्योगिकी (prōdyōgakī, Technology) उत्तीर्ण (uttīrṇa, Pass) फल (phala, result)"

Query2: "दशहरी (daśaharī, a type of mango) चौसा (causā, a type of mango) फल (phala, fruit)"
Fig.3 shows the excerpt of Hindi WordNet graph for the query. For the proposed query expansion method, we consider the following strength associated to the relations in Hindi WordNet (please refer to equation (1)):

- Hypernymy = 1,
- Hyponymy=Troponymy=0.9,
- Meronymy=Entailment=0.8,
- Holonymy=0.7,
- Cross parts of linkage=0.6,
- Gloss=0.5.

We explain our query expansion method using the query "प्रौद्योगिकी (prōdyōgakī, technology) उत्तीर्ण (uttīrṇa, pass) फल (phala, result)" first. Fig. 4 shows the fuzzy graph created for the query by following step 1 in algorithm 1. In step 2, we compute all the local connectivity measures (various centrality measures) for the constructed fuzzy graph. Table 1 shows the computed value for the same.

Algorithm 1. Algorithm For Hindi Query Expansion

Given a Query $Q = (t_1, t_2, t_3, \dots, t_n)$ that consists of one or more than one ambiguous terms, where $t_i (1 \leq i \leq n)$ represents the i^{th} query term after the task of pre-processing is performed (stemming and stop words removal etc.).

Step1: Construct the directed fuzzy graph $G_f = (V_f, E_f)$ named as query graph corresponding to the query Q by applying the following steps:

- i. Let $V_Q = \cup_{i=1}^n \text{senses}(t_i)$ represents all possible word senses in Q where $\text{sense}(t_i)$ represents set of all possible senses of t_i in

Hindi WordNet. Initially we set: $V_f = V_Q, E_f = \emptyset$

- ii. For each node $v \in V_Q$, perform Depth First Search (DFS) of the Hindi WordNet Graph considering v as the source node up to length $\leq \delta$, where δ is the threshold value. If a node $v' \in V_Q$ is found in the path, such that $v \neq v'$ and $v' \& v$ are not polysemous, then add all the nodes and edges which are encountered in the path $(v, v_1, v_2, \dots, v_k, v')$ in the graph G_f .

$$V_f = V_f \cup \{v_1, v_2, \dots, v_k\}$$

$$E_f = E_f \cup \{(v, v_1), (v_1, v_2), \dots, (v_k, v')\};$$

- iii. In the constructed graph G_f , assign the weight $w_r \in [0, 1]$ (where $r \in R$, and R is either the set of relations defined in Hindi WordNet or the relation between word and word of its gloss) corresponding to the edge representing relation r in the following manner:

$$W_{\text{hyponymy}} \geq W_{\text{hyponymy}} = W_{\text{troponymy}} \geq W_{\text{entailment}} = W_{\text{meronymy}} \geq W_{\text{holonymy}} \geq W_{\text{cp}} \geq W_g$$

where g represents gloss and cp represents cross parts of speech linkage.

// Note: Causative relation (represents the base form of the word) is already considered at the time of pre-processing (stemming)

of query. Antonymy, Gradation relations are related to opposite meaning of the word so these relations are not considered here.

Step2:

- i. Compute the following local connectivity measures of the fuzzy graph G_f
 - a. Degree Centrality
 - b. Eigen Vector Centrality (PageRank, HITS(HUBS,AUTHORITY))
 - c. Closeness Centrality
 - d. Betweenness Centrality
- ii. For each $v \in V_f$, compute $\text{avg}(v) =$ average value of all local connectivity measures. Identify all the nodes v , such that $v \notin V_Q$ and $\text{avg}(v) \geq \alpha$ where α is the threshold value. Let V_{lcm} represents the set of such nodes. All the terms belongs to V_{lcm} represents the additional terms for the user's query.

Step3:

- i. Let V_i represent the set of words in the i^{th} interpretation of the query. $V_i \cup V_{lcm}$ represents the i^{th} interpretation of the expanded query. For each interpretation, create the fuzzy graph (named as interpretation graph) $G_i = (V_i, E_i), G_i \subseteq G_f$ where $V_i = V_i \cup V_{lcm} \cup V_{int}$ and V_{int} is the set of intermediate nodes in the path which connects any pair of the nodes in $V_i \cup V_{lcm}$, $E_i \subseteq E_f$ is the set of edges in the path which connects any pair of nodes in $V_i \cup V_{lcm} \cup V_{int}$.
- ii. Discard all the interpretation graphs amongst all G_i which contain at least one disconnected node.
 - // A graph having disconnected node is not able to convey any valid interpretation therefore it is discarded.
- iii. For each remaining G_i compute the following global connectivity measures.

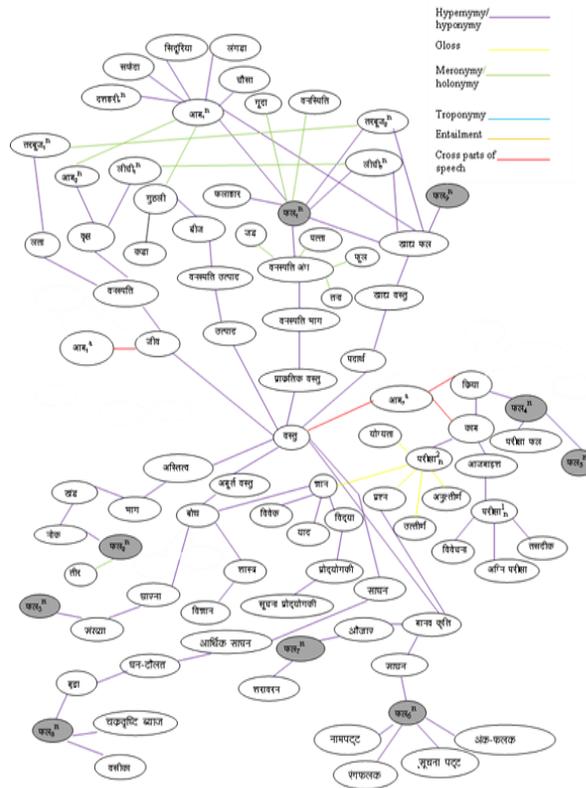


Fig. 3. An excerpt of Hindi WordNet graph around the words in queries “प्रौद्योगिकी(prōdyōgakī, Technology)दशहरी(daśaharī) चौसा(causā) फल(phala)”

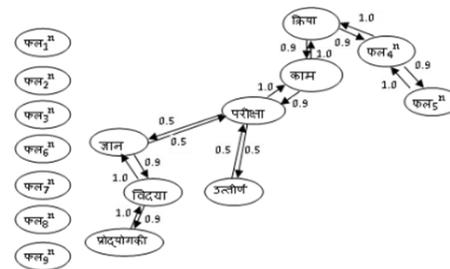


Fig. 4. Fuzzy graph for the query “प्रौद्योगिकी (prōdyōgakī, Technology) उत्तीर्ण(uttīrṇa, Pass)फल(phala, result)”

Now we identify the nodes (other than those representing query terms) which have high average value of local connectivity measures. Taking the threshold value $\alpha=0.2$, we find that the word/node परीक्षा₂ⁿ(parīkṣā₂ⁿ, examination) (here ₂ⁿ represents superscript *n* as noun and subscript 2 represents second sense in Hindi WordNet) is eligible to be used as the term for query expansion,

as it has average value of local connectivity measures = 0.22.

It may be noted that the additional terms for the given query are unambiguous here. Although, the word परीक्षा (parīkṣā, examination) has two senses in Hindi WordNet, our method automatically identifies the correct one. Now the word sense परीक्षा₂ⁿ (parīkṣā₂ⁿ, examination) is the additional

फल...
Noun(3)

- (R)(E)(A)(Be)(Bo)(G)(K)(Ka)(Ko)(M)(Ma)(Mi)(N)(O)(P)(S)(T)(Te)(U) फल, फर, प्रसून - वनस्पति में होने वाला गूदे या बीज से भरपूर बीजकोश जो किसी विशिष्ट ऋतु में फूल आने के बाद उत्पन्न होता है *उसने फल की दुकान से एक किलो आम खरीदा*
- (R)(E)(A)(Be)(Bo)(G)(K)(Ka)(Ko)(M)(Ma)(Mi)(N)(O)(P)(S)(T)(Te)(U) फल, गाँस, गाँसी, गाँस, गासी, गाँसी, अंकुड़ा, अंकुड़ा, अंकड़ा, अंकड़ा, आँकुड़ा - तीर या बरछी आदि के आगे का धारदार भाग *इस तीर का फल बहुत नुकीला है*
- (R)(E)(A)(Be)(Bo)(G)(K)(Ka)(Ko)(M)(Ma)(Mi)(N)(O)(P)(S)(T)(Te)(U) परिणाम, अंजाम, अन्जाम, अंत, अन्त, नतीजा, प्रतिफल, फल, हथ, ताबीर, खामियाजा, खामियाजा, खामियाजा, खामियाजा, परिणति, खुट्टि, विपाक, रिजल्ट - किसी कार्य के अंत में उसके फलस्वरूप होनेवाला कार्य या कोई बात *उसके कार्य का परिणाम बहुत ही बुरा निकला*



Fig. 5. Snapshot of Hindi WordNet for the word “फल” परीक्षा2n (parīkṣā2n)

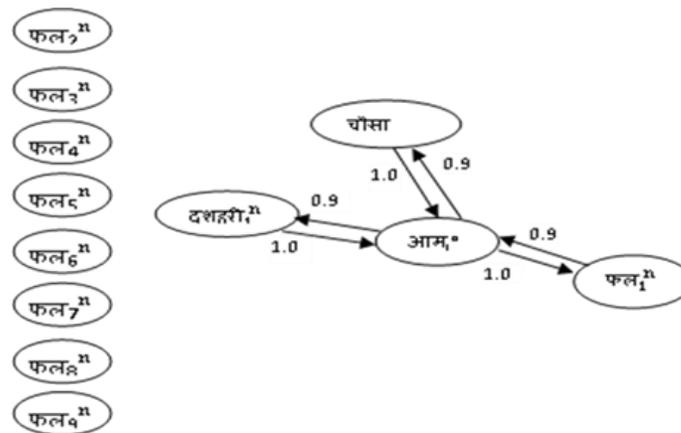


Fig. 6. Fuzzy graph for query "दशहरी (daśaharī, a type of mango) चौसा (causā, a type of mango) फल (phala,fruit)"

term for the query "प्रौद्योगकी (prōdyōgakī, technology) उत्तीर्ण (uttīrṇa, pass) फल (phala, result)". The intermediate query now becomes "प्रौद्योगकी (prōdyōgakī, technology) उत्तीर्ण (uttīrṇa, pass) फल (phala, result) परीक्षा₂ⁿ (parīkṣā₂ⁿ, examination)".

Now we disambiguate the original terms of the query using global connectivity measures. In the query फल (phala, Fruit/Blade/solution/result/consequence/board/shield/interest/nutmeg) is the ambiguous term which has 9 different senses. Therefore, the resulting query has 9 different interpretations.

For each interpretation we create the fuzzy subgraph (please refer fig.5) using step3. We notice that the graphs for 7 different interpretations

viz. "प्रौद्योगकी (prōdyōgakī, technology) उत्तीर्ण (uttīrṇa,pass) फल₁ⁿ(phala₁ⁿ,fruit) परीक्षा₂ⁿ (parīkṣā₂ⁿ, examination)", "प्रौद्योगकी (prōdyōgakī, technology) उत्तीर्ण (uttīrṇa,pass) फल₂ⁿ (phala₂ⁿ, result) परीक्षा₂ⁿ (parīkṣā₂ⁿ)", "प्रौद्योगकी (prōdyōgakī, technology) उत्तीर्ण (uttīrṇa, pass) फल₃ⁿ (phala₃ⁿ, nutmeg) परीक्षा₂ⁿ (parīkṣā₂ⁿ)", "प्रौद्योगकी (prōdyōgakī, technology) उत्तीर्ण (uttīrṇa, pass) फल₄ⁿ(phala₄ⁿ, interest) परीक्षा₂ⁿ (parīkṣā₂ⁿ)", "प्रौद्योगकी (prōdyōgakī, technology) उत्तीर्ण (uttīrṇa, pass) फल₅ⁿ(phala₅ⁿ,board) परीक्षा₂ⁿ (parīkṣā₂ⁿ)", "प्रौद्योगकी (prōdyōgakī, technology) उत्तीर्ण (uttīrṇa, pass) फल₆ⁿ (phala₆ⁿ, consequence) परीक्षा₂ⁿ (parīkṣā₂ⁿ)" and "प्रौद्योगकी (prōdyōgakī, technology) उत्तीर्ण

Table 1. Local connectivity measures for the query "प्रौद्योगिकी(prōdyōgakī) उत्तीर्ण(uttīrṇa) फल(phala)"

	क्रिया kiryā	काम kām	परीक्षा ² _n parīkṣā ² _n	ज्ञान jñāna	उत्तीर्ण uttīrṇa	विद्या vidyā	प्रौद्योगिकी prōdyōg akī	फल ¹ _n phala ¹ _n	फल ² n phal a ² _n	फल ³ _n phala ³ _n	फल ⁴ _n phala ⁴ _n	फल ⁵ _n phala ⁵ _n	फल ⁶ _n phala ⁶ _n	फल ⁷ _n phala ⁷ _n	फल ⁸ _n phala ⁸ _n	फल ⁹ _n phala ⁹ _n
Degree	0.12	0.12	0.13	0.09	0.03	0.12	0.06	0.00	0.00	0.00	0.1	0.06	0.00	0.00	0.00	0.00
PageRank	0.07	0.11	0.14	0.13	0.08	0.08	0.53	0.05	0.05	0.05	0.71	0.04	0.05	0.05	0.05	0.05
HITS: Authority	0.25	0.32	0.32	0.32	0.19	0.19	0.12	0.00	0.00	0.00	0.07	0.12	0.00	0.00	0.00	0.00
HITS: Hub	0.22	0.29	0.42	0.26	0.16	0.22	0.09	0.00	0.00	0.00	0.19	0.03	0.00	0.00	0.00	0.00
Closeness	0.23	0.24	0.24	0.20	0.13	0.21	0.16	0.00	0.00	0.00	0.22	0.17	0.00	0.00	0.00	0.00
Betweenne ss	0.05	0.06	0.08	0.05	0.0	0.02	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
Average Value	0.15	0.19	0.22	0.17	0.09	0.14	0.16	0.00	0.00	0.00	0.22	0.07	0.00	0.00	0.00	0.00

Table 2. Global connectivity measures for the expanded query "प्रौद्योगिकी (prōdyōgakī) उत्तीर्ण (uttīrṇa) फल (phala) परीक्षा²_n (parīkṣā²_n)"

Global Measures	Interpretation	
	प्रौद्योगिकी (prōdyōgakī) उत्तीर्ण (uttīrṇa) फल ⁴ _n (phala ⁴ _n) परीक्षा ² _n (parīkṣā ² _n)	प्रौद्योगिकी (prōdyōgakī) उत्तीर्ण (uttīrṇa) फल ⁵ _n (phala ⁵ _n) परीक्षा ² _n (parīkṣā ² _n)
Compactness	0.837	0.830
Graph entropy	0.620	0.580
Edge density	0.410	0.310
Average Value	0.622	0.573

(uttīrṇa, pass) फल⁹_n (phala⁹_n, solution) परीक्षा²_n (parīkṣā²_n)" contain disconnected nodes, therefore these interpretations are discarded. Now we compute all global connectivity measures (viz.

Graph entropy, compactness, edge density) for rest 2 interpretations "प्रौद्योगिकी (prōdyōgakī, technology) उत्तीर्ण (uttīrṇa, pass) फल⁴_n (phala⁴_n, result) परीक्षा²_n (parīkṣā²_n, examination)", "प्रौद्योगिकी (prōdyōgakī, technology) उत्तीर्ण (uttīrṇa, pass) फल⁵_n (phala⁵_n, consequence) परीक्षा²_n (parīkṣā²_n, examination)" (please refer to table 2). Then we compute the average value of these global connectivity measures. The interpretation having the highest average value for global connectivity measures is identified as the final expanded query, which in our case is "प्रौद्योगिकी (prōdyōgakī,

technology) उत्तीर्ण (uttīrṇa, pass) फल⁴_n (phala⁴_n, result) परीक्षा²_n (parīkṣā²_n, examination)".

Similar procedure is followed for the second user query "दशहरी (daśaharī, a type of mango) चौसा (causā, a type of mango) फल (phala, fruit)". Fig. 6 shows the fuzzy graph created for this query. We compute all local connectivity measures for the same, which is being shown in Table 3.

The term having high average value of local connectivity measures is आम¹_n (āma¹_n, mango) so the intermediate query is "दशहरी (daśaharī, a type of mango) चौसा (causā, a type of mango) फल (phala, fruit) आम¹_n (āma¹_n, a type of mango)". Now we disambiguate the intermediate query (please refer to Table 4). Therefore the final expanded query is "दशहरी (daśaharī, a type of mango) चौसा

Table 3. Local connectivity measures for the query "दशहरी (daśaharī) चौसा (causā) फल (phala)"

	दशहरी ⁿ daśaharī	आम ⁿ āma ⁿ	चौसा causā	फल ⁿ phala ⁿ								
Degree	0.09	0.25	0.09	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PageRank	0.07	0.10	0.07	0.07	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
HITS (Authority)	0.46	0.55	0.46	0.45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HITS(Hub)	0.51	0.53	0.51	0.45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Closeness	0.21	0.25	0.17	0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Betweenness	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Average Value	0.22	0.23	0.21	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 4. Global connectivity measures for the expanded query "दशहरी(daśaharī)चौसा(causā)फल(phala)आमⁿ(āmaⁿ)"

Global Measures	Interpretation
	"दशहरीचौसाफल ⁿ आम ⁿ "
Compactness	0.96
Graph entropy	0.63
Edge density	0.93

(causā, a type of mango) फलⁿ (phalaⁿ, fruit) आमⁿ (āmaⁿ, a type of mango)".

Figure 5 highlights the results obtained after searching for the word "फल" on Hindi WordNet.

The query expansion process expands the query according to intended interpretation by the user. In case of first query, the word "फल (phala, fruit/Blade/solution/result/consequence/board/shield/interest/nutmeg)" possesses 9 word senses as defined in Hindi WordNet. Our method expands the query according to the intended interpretation by the user as the user intended to use the term "फल (phala, fruit/Blade/solution/result/consequence/board/shield/interest/nutmeg)" to refer to "result". Similarly, in the second query, the user intended to use the term फल (phala, Fruit/Blade/solution/result/ consequence/ board/shield/interest/nutmeg) to refer "fruit". Moreover, in the second query, the user is not certain about which fruit he/she is interested in.

But the query formulated by him/her is biased towards the fruit mango (आम (āma, mango)

because दशहरी (daśaharī, a type of mango) & चौसा (causā, a type of mango) are two varieties of fruit mango आम (āma, mango).

Therefore, by expanding this query, our method adds the unambiguous word आमⁿ (āmaⁿ). Therefore, the updated query is now capable of searching for other varieties of the fruit आम (āma, a type of mango) also, which the user might be interested in, as his query is oriented towards the varieties of fruit mango (आम (āma, mango)). Hence, the proposed method identifies the desired word sense of the query, expands it and thus enhances the retrieved results.

5 Experimental Setup and Results

We used Hindi WordNet 1.4 for identification of the additional terms for the query and for disambiguation of query terms. The Forum for Information Retrieval Evaluation (FIRE) creates evaluation frameworks like TREC, CLEF and

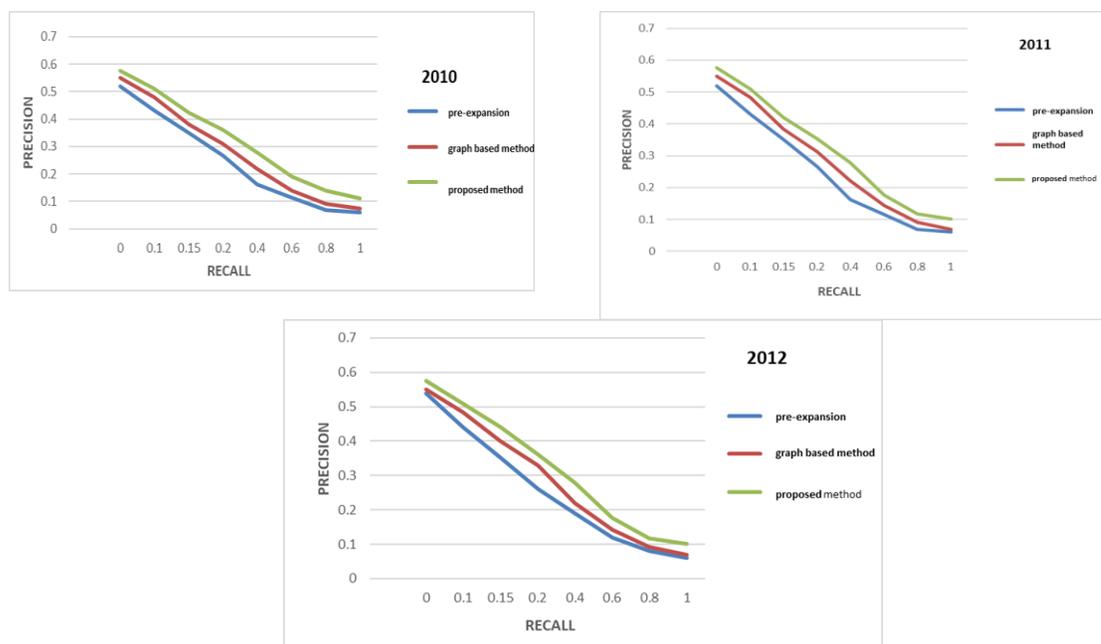


Fig. 7. Precision-Recall graph for proposed query expansion for FIRE dataset for 2010, 2011 and 2012 respectively

NTCIR for the South Asian languages. In this paper, we have used their corpuses and its query sets for 3 consecutive years (2010, 2011 and 2012) in Hindi language. The corpus consists of around 95,215 Hindi documents and 178861 distinct words. We tested the proposed method on the given queries and have compared the performance of the proposed method with the results obtained from original query (without expansion) for all the 3 years. We have also compared the performance of proposed query expansion method with the graph-based query expansion method (Jain et. al. 2014) which suffered from a major limitation that all the semantic relations were given equal importance while generating the WordNet graph.

Fig. 7 shows the precision and recall obtained. For each query, the first 100 documents were extracted, and the precision of the system was calculated at different points i.e. at the point where 10, 20, 30 documents were retrieved.

The outcomes of the experimentation as illustrated in Fig. 7 show that the proposed method performs better than these two. It can be observed

that as far as Hindi query expansion is being concerned, the proposed method is able to achieve better results as compared to the ones obtained by the state-of-art (Jain et. al. 2014).

This is majorly since the assignment of semantic significance to the relations in Hindi WordNet, now the related words are appearing nearby to each other in the WordNet graph. Hence our proposal seems to be logically significant too.

This work can be extended for other languages too provided their WordNet is available and all the major semantic relations are taken care of for the assignment. In Table 5 we present the glossary of the used Hindi words.

6 Conclusion

In this paper we have used local and global measures of graph connectivity and found them well suited for the task of query expansion. The Hindi query expansion method was proposed to consider the context of the user query and to inculcate the concept of semantic importance of the relations in Hindi WordNet. Fuzzy graph

Table 5. Glossary

Hindi Word	Corresponding English Meaning	Hindi Word	Corresponding English Meaning
मिलना(milana)	Visit/meeting/reception/get/blend/occur/place/go_steady/combine/join	हँसना (hँsanā)	To laugh
खाना(khānā)	Meal/box/eating/diet/drawer/square/food/eat/eat/-/-/destroy/consume/waste/-/trouble/deplete/pocket	खरटालेना (kharrāṭālēnā)	To snore
पाना(pānā)	Wrench/duel/get/get/pick_up/equal/derive/-/experience/-/place/ accept/achieve	सोना (sōnā)	To sleep
दल(dala)	Team/side/set/force/-/panel/group/petal/party/herd/leaf/battalion	प्रोद्योगकी (prōdyōgakī)	Technology
बैठना (baiṭhanā)	Visit/-/retreat/-/-/settle/-/ensconce/hop_on/sink/-/devolve	उत्तीर्ण (uttīrṇa)	Pass
बांधना (bāndhanā)	Rope/bunch/-/-/sew/hire/capture/construct/situate/knot/pack/subscribe	दशहरी (daśaharī)	A type of mango
अच्छा (acchā)	Good/all_right/well/faultless/glorious/proper/-/healthy/delectable/well/advantageously	चौसा (causā)	A type of mango
वाहन (vāhana)	Vehicle	परीक्षा (parīkṣā)	Examination
इंजन(iñjana)	Engine	काम(kāma)	Deed
पहिया(pahiyā)	Wheel	ज्ञान (jñāna)	Cognition
सीट(sīṭa)	Seat	विद्या (vidyā)	Skill
सामान (sāmāna)	Material	सेब (sēba)	Apple
वस्तु(vastu)	Thing	गुठली (guṭhalī)	Endocarp
सामान्य (sāmāny)	Ordinary	गूदा (gūdā)	Pulp
फल (phala)	Fruit/Blade/solution /result/ consequence/board/shield/interest/nutmeg	छिलका (chilakā)	Rind
लीची(līcī)	Litchi	कार (kāra)	Car
आम(āma)	Mango	मुस्कराना (muskurānā)	To smile

connectivity measures were used in investigation of fuzzy graph structure which further helped in identification of additional terms for the query.

We analyzed that some Hindi WordNet relations play more important role than the other in identification of the context of the user query and thus helped in improving query expansion.

Also, our method was able to add the right unambiguous terms to the user query.

We found that the proposed method identifies the desired word senses of the query terms and thus helped in enhancing the retrieval results. The proposed method is evaluated on the Forum for Information Retrieval (FIRE) dataset for 3 consecutive years and shows optimum results.

Since WordNet has been used for a number of different purposes in information systems, including document classification, automatic text

summarization, to determine the similarity between words etc., hence, in future, this idea of assigning semantic importance to relations in Hindi WordNet can be used in these applications which would help in improving their performance.

References

1. **Agirre, E. & Soroa, A. (2009).** Personalizing Pagerank for Word Sense Disambiguation. *Proceedings of EACL '09*, pp. 33–41.
2. **Barathi, M. & Valli, S. (2010).** Ontology Based Query Expansion Using Word Sense Disambiguation. *International Journal of Computer Science and Information Security*, Vol. 7, No. 2, pp. 022–027.
3. **Bhattacharya, P. (1987).** Some Remarks on Fuzzy Graphs. *Pattern Recognition letters*, Vol. 6, No. 5, pp. 297–302. DOI: 10.1016/0167-8655(87)90012-2.
4. **Bhogal, J., Macfarlane, A., & Smith, P. (2007).** A review of ontology based query expansion. *Information Processing and Management*. Vol. 43, No. 4. pp. 866–886. DOI: 10.1016/j.ipm.2006.09.003.
5. **Borgatti, S.P. (2003).** Identifying Sets of Key Players in a Network. *Proceedings IEMC '03 Managing Technologically Driven Organizations*, pp. 127–131. DOI: 10.1109/KIMAS.2003.1245034.
6. **Cao, G., Nie, Y.J., & Bai, J. (2005).** Integrating word relationships into language models. *Proceedings of the 28th annual international ACM SIGIR conference on research and development in information retrieval*, pp. 298–305. DOI: 10.1145/1076034.1076086.
7. **Collins-Thompson, K. & Callan, J. (2005).** Query expansion using random walk models. *Proceedings of the 14th ACM Intl conference on information and knowledge management*, pp. 704–711. DOI: 10.1145/1099554.1099727.
8. **Stokoe, C., Oakes, M.J., & Tait, J.I. (2003).** Word Sense Disambiguation in Information Retrieval revisited. *SIGIR '03 Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 159–166. DOI: 10.1145/860435.860466.
9. **Carpineto, C. & Romano, G. (2012).** A survey of automatic query expansion in information retrieval. *Journal ACM Computing Surveys*. Vol. 44, No. 1, pp. 1–50. DOI: 10.1145/2071389.2071390.
10. **Grossman, D.A. & Frieder, O. (2004).** Information Retrieval Algorithms and Heuristics. *Springer*.
11. **Das, S., Seetha, A., Kumar, M., & Rana, J.L. (2010).** Post Translation Query Expansion using Hindi WordNet for English-Hindi CLIR System. *Forum for Information Retrieval Evaluation (FIRE)*.
12. **Dwivedi, S.K. & Rastogi, P. (2008).** An Entropy Based Method for Removing Web Query Ambiguity in Hindi Language. *Journal of Computer Science*, Vol. 4, No. 9, pp. 762–767. DOI: 10.3844/jcssp.2008.762.767.
13. **Esuli, A. & Sebastiani, F. (2007).** PageRanking WordNet Synsets: An Application to Opinion Mining. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pp. 424–431.
14. **Freeman, L.C., Boragatti, S.P., & White, D.R. (1991).** Centrality in Valued Graph: A Measure of Betweenness Based on Network Flow. *Proceedings of Social Networks*, Vol. 13, No. 2, pp. 141–154. DOI: 10.1016/0378-8733(91)90017-N.
15. **Fellbaum, C. (1998).** WordNet: An Electronic Lexical Database, MIT Press.
16. **Franco-Salvador, M., Rosso, P., & Navigli, R. (2014).** A Knowledge-based Representation for Cross-Language Document Retrieval and Categorization. *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 414–423.
17. **Miller, G.A. (1995).** WordNet: A lexical database for English. *International Journal of Lexicography*. pp. 31–41.
18. **Gong, Z., Cheang, C., & Leong-Hou, U. (2005).** Web query expansion by WordNet. In **Andersen, K., Debenham, J., and Wagner, R., editors**, Database and Expert Systems Applications, Lecture Notes in Computer Science, pp.166–175.
19. **Grootjen, F. & van Der-Weide, T. (2006).** Conceptual query expansion. *Data and Knowledge Engineering*, Vol. 56, No. 2, pp.174–193.
20. **Gupta, G.K., (2006).** *Introduction to Data Mining With Case Studies*. PHI, pp. 238-240.
21. **Gupta, J.P., Tayal, D.K., & Gupta, A. (2011).** A TENGGRAM method based part-of-speech tagging of multi-category words in Hindi language. *Expert Systems with Applications*, Vol. 38, No. 12, pp. 15084–15093.
22. **Hai-Tao, Z., Bo-Yeong, K., & Hang-Gee, K. (2009).** Exploiting noun phrases and semantic relationships for text document clustering. *Journal on Information Sciences*, Vol. 179, pp. 2249–2262.
23. **Hindi WordNet (2014).** Center for Indian Language Technology Solutions, IIT Bombay,

- Mumbai, India. <http://www.cfilt.iitb.ac.in/wordnet/wordnet/webhwn/Acce>.
24. Jain, A., Sudesh, Y., & Devendra, K.T. (2013). Measuring Context Meaning for Open Class words in Hindi Language. *6th International Conference on Contemporary Computing*, pp. 118–123.
 25. Jain, A., Mittal, K., & Tayal, D.K. (2014). Automatically Incorporating Context Meaning for Query Expansion using Graph Connectivity Measures. *Progress in Artificial Intelligence*, Vol. 2, No. 2, pp.129–139.
 26. Jain, A. & Lobiyal, D.K. (2016). Fuzzy Hindi WordNet and Word Sense Disambiguation using Fuzzy Graph Connectivity Measures. *ACM Transactions on Asian and Low-Resource Language Information Processing*, Vol. 15, No. 2.
 27. Kim, B.M., Kim, J.Y., & Kim, J. (2001). Query term expansion and reweighting using term co-occurrence similarity and fuzzy inference. *Proceedings of the Joint 9th IFSA World Congress and 20th NAFIPS International Conference*, Vol. 2, pp. 715–720.
 28. Kleinberg, J.M. (1998). Authoritative Sources in a Hyperlinked Environment. *Proceeding Ninth Symp. Discrete Algorithms*, pp. 668–677.
 29. Kumar, S., & Mansotra, V. (2012). Query Optimization: A Solution for Low Recall Problem in Hindi Language Information Retrieval. *International Journal of Computer Applications*, Vol. 55, No. 17, pp. 6–17.
 30. Lee, K.H. (2005). *First Course on Fuzzy Theory and Applications*. Springer, Berlin-Heidelberg.
 31. Litvak, N., Scheinhardt, W., & Volkovich, Y. (2006). *In-Degree and Page-Rank of Web Pages: Why Do They Follow Similar Power Laws?*
 32. Manning, C.D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*, Cambridge University Press.
 33. Mark, S. (2008). Ambiguous queries: test collections need more sense. *Proc. of SIGIR '08*, pp. 499–506.
 34. Mihalcea, R., Tarau, P., & Figa, E. (2004). PageRank on semantic networks, with application to word sense disambiguation. *Proceedings of the 20th international conference on Computational Linguistics*. Association for Computational Linguistics.
 35. Navigli, R. (2009). Word Sense Disambiguation: a Survey. *ACM Computing Surveys*, Vol. 41, No. 2, pp.1–69.
 36. Navigli, R. & Lapata, M. (2010). An Experimental Study of Graph Connectivity for Unsupervised Word Sense Disambiguation. *IEEE transaction on pattern analysis and machine learning*, Vol. 32, No. 4, pp. 678–692.
 37. Newman, M.E.J. (2005). *A Measure of Betweenness Centrality Based on Random Walk*. Social Network, Elsevier.
 38. Novischi, A. (2004). Combining Methods for Word Sense Disambiguation of WordNet Glosses. *Proc. 17th Florida Artificial Intelligence Research Soc.*
 39. Mao-Yuan, P., Ming-Yen, Ch., Hui-Chuan, Ch., & Yuh-Min, Ch. (2013). Development of a semantic-based content mapping mechanism for information retrieval. *Expert Systems with applications*, Vol. 40, No. 7, pp. 2447–2461.
 40. Rila, M., Takenobu, T., & Hozumi, T. (1998). The use of WordNet in Information Retrieval. *Proc. of the COLING-ACL workshop on Usage of WordNet in Natural Language Processing*, pp. 31–37.
 41. Blanco, R. & Lioma, C. (2012). Graph-based term weighing for information retrieval. *Information retrieval*, Vol. 15, No. 1, pp. 54–92.
 42. Salton, G. & McGill, M. (1988). *Introduction to Modern Information Retrieval*. McGraw-Hill.
 43. Sanasam, R., Singh, A., Murthy, T., & Gonsalves, A. (2013). Inference based Query Expansion Using User's Real Time Implicit Feedback. *Knowledge Engineering and Knowledge Management, Communications in Computer and Information Science*, Vol. 272, pp.158–172.
 44. Shuang, L., Clement, Y., & Weiyi, M. (2005). Word Sense Disambiguation in queries. *Proc. of CIKM'05*, Bremen, Germany, pp. 525–532.
 45. Sinha, M., Reddy, M.K., Bhattacharya, R.P., Pandey, P., & Kashyap, L. (2008). Hindi Word Sense Disambiguation. *International Symposium on Machine Translation, Natural Language Processing and Translation Support Systems, Delhi, India*.
 46. Song Wei, Jiu Zhen Liang, Xiao Long Cao, & Soon Cheol Park (2014). An effective query recommendation approach using semantic strategies for intelligent information retrieval. *Expert Systems with applications*, Vol. 41, No. 2, pp. 2366–372.
 47. Sunitha, M.S. (2001). *Studies on Fuzzy Graph*. Ph.D. Dissertation, Cochin University of Science and Technology, Cochin, India.
 48. Tanveer, S. & Tiwari, U.S. (2008). *Natural Language Processing and Information Retrieval*. Oxford University Press.
 49. Vaidyanathan, R., Das, S., & Srivastava, N. (2013). *Query Expansion based on Equi-Width and Equi-Frequency Partition*. LNCS, 7536:13–22.

50. **Varshney, S. & Bajpai, J. (2013).** Improving performance of English-Hindi cross language information retrieval using transliteration of query terms. *International Journal on Natural Language Computing*, Vol. 2, No. 6.
51. **Vij, S., Jain, A., Tayal, D., & Castillo, O. (2018).** Fuzzy Logic for Inculcating Significance of Semantic Relations in Word Sense Disambiguation Using a WordNet Graph. *International Journal of Fuzzy Systems*, Vol. 20, No. 2, pp. 1–16.
52. **Voorhees, E.M. (1994).** Query Expansion using Lexical-Semantic Relations. *SIGIR '94: Proceedings of the 17th Annual International ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 61–69.
53. **Wang, C., Yajun, D.U., Zhang, P., & Han, B. (2010).** A Term-Reweighting Method for Query Expansion. *Journal of Computational Information Systems*, Vol. 6, No. 11, pp. 3779–3785.
54. **Wasserman, S. & Faust, K. (1994).** *Social Network Analysis: Methods and Applications*. Cambridge Univ. Press.
55. **Yager, R.R. (2010).** Concept Representation and Database Structure in Fuzzy Social Relational Networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, Vol. 40, No. 2, pp. 413–419.

*Article received on 17/07/2018; accepted on 19/06/2019.
Corresponding author is Oscar Castillo.*