Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Automatic Query Expansion for Term Selection with BERT Score and WordNet Semantic Filtering

Rishabh Sachan[1], Mrs Sony Tripathi[2]

[1]M.Tech. Scholar, Department of Computer Science and Engineering, Naraina College of Eng. and Tech., Kanpur, Uttar Pradesh, India

[2]Assistant Professor, Department of Computer Science and Engineering, Naraina College of Eng. and Tech., Kanpur, Uttar Pradesh, India

ABSTRACT

In this modern era, with loads and tons of data, extracting a much relevant data has always been the user's desire. This whole process of retrieving the relevant document is known as information retrieval. Information retrieval system uses query expansion techniques for retrieving the most relevant document for satisfying the user's requirements. There have been several query expansions introduced few of which are knowledge based, corpus based, and pseudo relevance feedback. They make use of synonyms accordingly user's query with a purpose of extracting the relevancy and it also screens the documents that contains the search term but has a completely different meaning. A combination of genetic algorithms and fuzzy logic blocks approach is proposed in this work. We also presented a new method that attenuates the power of knowledge based or corpus based techniques.

Keywords : Information Retrieval, Query Extension, Pseudo Relevance Feedback, WordNet, Word2vec, Query Expansion, Theme Semantic Network

I. INTRODUCTION

Information Retrieval (IR) is an extensive-term. From book search to library management system for searching a document on the World Wide Web, search any kind of information is called as the retrieval of information. It stems from the search phone number in the phone bar to ask for chat details that require the agent to dial a specific person's phone number.

In the standard IR function, the user provides multiple title-specific terms to describe something as information required. The IR system is expected to retrieve texts that best fit the user's needs, so it starts with those documents which contain terms provided by the user.

Query Expansion

There is a huge amount of information available on the Internet, and it is growing exponentially. This unrestricted growth in knowledge has not been accompanied by concurrent technological advances in the production of relevant information. Often, web searches do not produce the right results. There are many reasons for this. First, user-generated keywords are unrelated to multiple topics; as a result, search results do not focus on the topic of interest. Second, the question can be so short that you can accurately capture what the user wants. This may

occur as a practice problem (e.g., average web search size of 2.4 words). Third, the user is often unsure of what he or she wants until he or she sees results.

Even if the user knows what he or she wants, he or she does not know how to make the right question (travel questions are not the same [12]. QE plays an important role in tracking relevant results in the above scenarios. Most web queries fall under these three basic categories:

- *Information Questions:* Questions covering a broad topic (e.g., India or journals) that can have thousands of relevant results.
- *Navigation Questions:* Questions that require a specific website or URL (e.g., ISRO).
- *Sales Questions:* Questions that show the user's intention to perform a particular task (e.g., downloading papers or purchasing books)

Currently, user queries are processed using indicators and ontologies, which work directly and are hidden from users. This leads to a time difference: user queries and search index are not based on the same set of terms. This is also known as vocabulary problem caused by a combination of synonymy and polysemy. Synonymy means many words that have the same meaning, e.g. \ Buy "and \ purchase". Polysemy refers to words that have multiple meanings, e.g., "mouse" (computer device or animal). Similar words and phrases are a barrier to obtaining relevant information; reduce levels of memory and accuracy.

However, there are also problems with QE strategies, e.g., There are accounting costs associated with the use of QE strategies. In the case of Internet search, where prompt response time is required, the calculation costs associated with the use of QE strategies restrict their use partially or completely. Another downside is that it can sometimes fail to establish relationships between the name in the corpus and those used in different communities, e.g., \ high citizen "and \ adults". Another issue is that QE can damage the recovery function of some queries. In terms of automation and end user engagement, QE strategies can be categorized as follows:

- Text Query Extension: Here, the user manually resizes the question.
- Default Query Extension: Here, the system automatically updates the question without user intervention.

Both the process of calculating the T0 set and the selection of D data sources are incorporated into system intelligence.

Interactive query extension: Here, query query occurs due to the interaction between the system and the user. It is a man-in-the-loop method where the system returns search results to an automated query, and users show visible results among themselves. Depending on the user's preferences, the system continues to modify the query and get results. The process continues until the user is satisfied with the search results.

Working of Query Expansion

The process of Query Expansion consists of these four steps:

1. Pre-processing of data
2. Extraction of terms
3. Weighting and ranking of terms
4. Expansion of terms

The process which undergoes the following steps is shown in the figure below

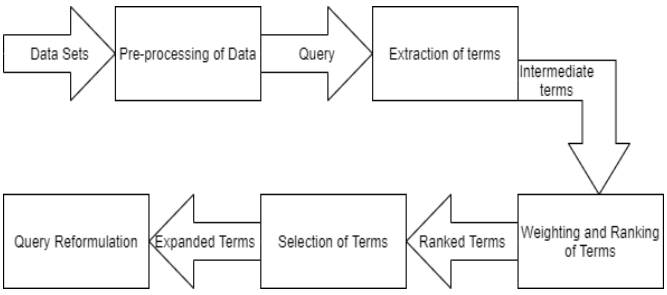

Fig 1: Working of Query Expansion Process

Classification of QE approaches

On the basis of data sources used in QE, several approaches have been proposed. All these approaches can be classified into two main groups:

(1) Global analysis and (2) Local analysis.

Global and Local analyses can be further split into four and two subclasses respectively as shown in Fig.. This section discusses the QE approaches based on the properties of various data sources used in QE as shown in Fig.

1. ***Global Analysis*** In the global analysis, QE strategies completely select the terms of the extension from the hand-crafted information resources or from a large corpora to change the first question. Terms for each question only it is considered to amplify the first question. Extensions are exactly the same as first words. Each term is given weight; extension terms can be assigned a minimum weight to compare with the terms of the original question. Global analysis can be divided into four categories in basis for terms of inquiries and data sources: (i) Language-Based, (ii) Copy-Based, (iii) Login-Based, and (iv) Web-Based.
2. ***Local Analysis:*** Local analysis includes QE techniques that select extension words from the text collection retrieve and answer the user's first question for the user. The belief that documented the returns obtained in response to the user's first question are valid, which is why the terms exist in these texts should also be accompanied by the original question. Using location analysis, there are two ways to do it extend the user's original question: (1) The Relevance Feedback and (2) the pseudo relevance feedback.

Since several years, Query Expansion (QE) methods have emerged in an effective manner to abode ambiguity of the information contained in the documents in the process of Information Retrieval (IR). The purpose is to reinforce the query by summing up the words associated to the concept, especially using similar words. QE methods can be categorized by global or local methods. In contrary, global methods extend the primal question without any repercussions. In general, WordNet is an established tool for opting new terms associated with the primal query [1]. Whereas, local methods use a consistent response, in which they perform initial retrieval of the result used to select the most promising terms [2].

Typically, Pseudo Relevance Feedback (PRF) has been endorsed to reduce computer costs, as this method inevitably removes new words from the top set of documents marked “k”, found in the retrieval process [3][4].The above approaches faced drawbacks. Often, global approaches require ambiguous vocabulary, as words retrieved on the basis of ontological knowledge based methods (such as WordNet) are polysomic. Thus local context methods are involved, using words from the top documents that may be relevant and found in the first return. Considering the concurrence of those documents can add audio words to the extended question. To prevail over these drawbacks, a recent study in QE has been reported where word embedding is used as a semantic model [5] [6]. Word embedding (WE) is still distributed by word presentations, usually found in the form of a neural network which reflects the shared distribution of thesaurus [7].

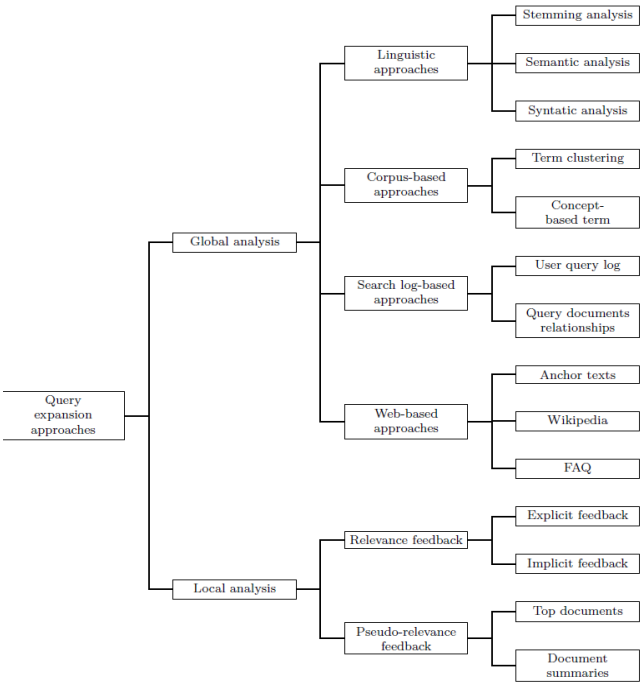

Fig 2: Classification of Query Expansion

These vector declarations have recently been used in a variety of natural language processing activities [8]. Word2vec [9] is the perfect way to accomplish this map these days. These methods are actually based on the assumption that each word of the question can select the best candidates with respect to the semantic approach. This leads to the utilization of extension techniques such as combining N-words closely or the exploration for relevant areas of terms [10]. Under this perspective, the semantics of the query are made up of atoms and is taken locally as the expected words are chosen by the query words at the same time. So to enrich QE performance, we need to measure the semantics of the whole query from the conception of possible words. Accordingly, we aim to enhance the quality of candidate terms related to query and add new linked terms according to the most relevant terms.

An automatic query expansion approach for term selection methods has been proposed in this paper. This approach uses the combination of genetic algorithms and fuzzy logic blocks. We used the concept of pseudo relevance feedback and ranked the documents with the help of Okapi-BM25. We have used the BIM co-occurrence method and corpus based methods to pass through the fuzzy logic blocks. The combination of these term selection methods gives the most unique terms. These fuzzy logic blocks give the final termed weight as output. Then we used those final weighted terms to calculate the BERT score. Later, we processed those terms in WordNet based semantic filter. These semantic filters sometime filters few irrelevant terms too. So to avoid this problem we passed those weighted terms through mutation based genetic algorithm. Thus the proposed approach ascertains the most relevant terms. It reweights the most favorable candidate terms using fuzzy rules.

This thesis is categorized into following chapters: chapter 2 gives the related work of automatic query expansions. It highlights the work of few authors. Chapter 3 discussed the proposed work and its preliminaries. Chapter 4 shows the result work and analysis. Chapter 5 concludes the paper and gives the idea for the future work.

II. Related Work

Semantic language vector space models points each and every word by a value which represents the real vector and it can also be used as a purchase vector for various applications. Word acquisition statistics in the corpus are the main source of information for reading those presentations. These genres incorporate standard mathematical complexity and integrity of the natural language, thereby denoting a structure in a coherent manner; e.g. shared presentation [11]. In variance to local presentations, distributed interpretation is an example of a space where names under similar conditions may have similar carriers. Thus the distributed presentation fully encapsulates a wide range of similarities [12]. In view the analogy “king” means “male” as “queen” means “female” tends to be encoded in the same vector space as the equality relationship felt in the equation such as king-queen = male-female. Most word processing methods anticipates the distance or midnight voice vectors as the foremost means of assessing the quality of such similarities.

The research in [13] is an example of a local QE strategy named after the Local Content System [2]. In this way, the extended question is made on the basis of retrieved search documents and original query. Therefore, the effectiveness of the method depends on accuracy of high results.

Research in [14] used real vectors found in the global embedding of the data warehouse. In contrary, they were tested with high N values from 1 to 10 in each term of the question. Although they were not able to enter the correct value in all the tested categories, the authors reduced the distance between 4 and 10 terms.

Similar to the last research, authors in [15] chose words that are very similar to vocabulary in each QE. Both of the above cases show that embedding gains significant mathematical improvement over two older questioning methods such as Pseudo Relevance Feedback and Mutual Information.

In [16] authors calculated the scores of similarities between inclusion and a small collection of vocabulary words. The authors suggest three ways to select candidates' policies based on neighborhoods close to K (KNN). In each case they form a set Q × N e, where Q = {q 1, · · Mum, q Q} are the terms of the question and N e = {t 1, · · · · · · ·, T, e N e} a set that contains all the corresponding neighbors of each K. After that, they calculated the cosine equilibrium rate between the elements of the set Q × N e. Set N is usually made up of vocabulary words selected using one of the three proposed combinations. The final list of student names is obtained by ordering set N e in proportion to the combined cosine similarity and keeping its top properties as terms of expansion. In order to incorporate new concepts into the first question, they proposed a consistent language model based on the variation of text possibilities. A positive aspect of their approach is that they use unigrams and bigrams from the K-neighbors' first computer query words, indicating that the word formation of the queries increases the effectiveness of their method.

In [17] authors expand the query using topical-guided local embedding. Local embedding is found in a subset of datasets found by firing a certain query. They used a different language model for Kullback-Leibler Divergence (KLD) to score goals. It is noteworthy that following the insertion of the local entries for each query, they reschedule the documents instead of retrieving another one. These outcomes show that local embedding is better in this way.

In summary, vector word presentations points out to be as a good initial point for modeling the basic similarities between terms and QE. However, we view fundamental thinking in current approaches as a major issue. Basically, considering in these Question-Guided Ideas that each query term can select the best candidates for QE. Therefore, in the search for scopes, current methods take one query at a time to find a set of terms in the semantic vector space. Thus each query word is "responsible" for the choice of vocabulary words as extension words: the semantics of the query are made up of atoms and held in place. Furthermore, the presentation of a vocabulary word demonstrates its most common concept in the corpus, as all content of the word combination leads to a term. When looking for candidates for the process of QE, the context of the term is ignored, so the term of which context can interpret the closest meaning is vitally important [18]. Local QE methods use published documents generated by the original query. Basically refers to the success of compliance with a fake relationship response (Buckley et al. 1994) ways to change the question. These methods use high-quality restored documents with the first query. However, the high-quality documents obtained may not always provide it good word extensions, especially difficult or short questions with few appropriate conditions texts in the collection that do not share relevant principles. These methods lead to the topic erosion and negative impact on outcomes (Macdonald and Ounis 2007).The authors in Cao et al. (2008) also examined the assumptions that provide PRF. It is assumed that the most common terms in duplicate texts - the answer is useful to retaliation do not apply to reality. In Chen and Lu (2010), the authors have shown that extension words cannot be distinguished only from their distribution in response documents and throughout the collection. Suggest a word combination of the separation process to predict the usefulness of the extension terms. Recently, in Colaceet al. (2015) authors have demonstrated the effectiveness of a new method of expanding that they issued two pairs of words based on relevant or fake

related documents. They have them to include a study to evaluate ways to select useful terms from a set of candidate expansion policies within the PRF framework. The results obtained showed that the QE approach depending on their new structure it exceeds the original foundation (Colace et al. 2015). They used the name of the embedded introduction, in Almasri et al. (2016), the authors examined the use of vector-extracted relationships in in-depth QE studies. They have shown that embedding is a promising source of an additional query by comparison with PRF and in the extension method, the same data is used.

Global QE methods in contrast to local QE, in international methods, candidate guidelines are derived from the compilation of all the texts contains the appropriate (fake) documents. In Xu et al. (1996), authors have argued that the application of global analytical strategies produces real results in both it is more effective and more predictable than a simple local response. Such QE methods are usually based on the exclusion of terms between terms of the whole text collection and based on their events when the size of the window used is a document.

In Javelin et al. (2001), authors have developed a drag-and-drop model data model extension of the question. Based on three levels of output: concept, language, and line levels. Concepts and relationships between them must be at the level of the mind.

The language level provides natural language expressions for concepts. Each sentence has another or many similar patterns at the level of the thread. In Gong et al. (2006), authors have used WordNet and Theme Semantic Network (TSN) for the development of Tsauri based on word combinations. TSN was used as a filter and provided a WordNet extension. However, it was realized that Thesauri's construction strategy was complex and tedious.

In addition to the global approach based on the construction of Thesaurus, we focus on the organization manages the mines aimed at obtaining consistent patterns (Agrawal 1994) from the collection of documents. Organizational law binds two sets of principles namely a foundation and conclusion. This means that the conclusion occurs whenever there is a basis seen in a set of texts. In each relationship rule, a certain amount of confidence is given measuring opportunities for associations. In the letter, it is proved that the use of such a QE dependence can significantly increase recovery performance (Wei et al. 2000). (e.g.) Organizational rules show a clear and solid connection between goals. Use this combination of expanding questions to enrich the presentation of the question by adding a set of related goals and as a result improve the performance of the return with similarity to other texts. Therefore, authors in Tangpong and Rungsawang (2000) make small improvements when using the APRIORI algorithm (Agrawal and Skirant 1994) with a high level of confidence (over 50%) that produced a small number of organizational rules. Using the lowest confidence limit (10%), the authors performed better results (Tangpong and Rungsawang 2000). In Hadad et al. (2000), authors suggest that the same way to improve when using APRIORI algorithm to extract organizational rules. Good progress is made with low confidence rates. The method in Martin-Bautista et al. (2004) refined the question in terms of organizational rules.

Given the first set of texts found on the web, text transactions are being created and organizational rules were made. These rules are used by the user to add additional in terms of the question of improving the accuracy of the return. More converted mines a textual algorithm that protects the unwanted organization rules in mines is proposed in Latiri et al. (2012). Unwanted organization rules between terms are used to enhance the user the question considers all the principles from the conclusions of these laws which are the basis of it contained in the first question. Experimental tests of this method show

improvement of IR function. Near our work, in Song et al. (2007), the authors proposed a semantic expansion question process that combines the rules of integration with ontologies and Strategies for Natural Language Processing. This method uses explicit semantics and other non-structural corpus language structures. It includes status structures of important principles found in organizational rules, as well as ontology entries, which are added to the question by separating the meanings of the words.

Word embedding uses Word2Vec and GloVe to measure similarities between queries and texts [31]. Process the database document using Word2Vec to search term candidates using the same semantic term. Calculate the approximate term of the term candidates by question using cosine similarity and is defined as a bipartite graph. Calculate the proximity of the questions to the candidate's objectives using the higher cosine similarity. Calculate the approximate term of the candidate with the question using the cosine-like similarity [32].

Some previous studies using semantic-based expansion questions were performed by adding candidates' questions using Harman, croft, okapi, and algorithm lesk statistics [30]; using combinations from the response text above; uses WordNet to understand semantic questions; integrate the process of compiling document integration and its semantics using WordNet; using place name embedding (to process the correct response according to the document database), and counts its semantic relations. In the meantime, retrieving semantic data is becoming an integral part of anything data repair engine. Semantic adjectives are often defined by ontology, which is clearly defined by the thinking of things. It plays an important role in explaining semantic details. Several projects aim to build ontology, (Suresh & Zayara, 2014) has used synthetic synthetic and semantic-based semantics to subtract divisive Bayes to extract the conceptual relationship from the informal text of the automatic construction of the ontology, in that the list of symptoms and the combination of a given seed concept is automatically extracted. In reference (Bentricia et al., 2017) suggested a method based on merger patterns for extraction semantic relations from the Quranic Arabic corpus to enrich the automatic structure of the Quran ontology. The role of building such an ontology is to provide direct and complete knowledge in the world, with the aim of reducing the role of expert knowledge in structured ontology. (Denis & Wasito, 2017) suggested a completely automatic method that combines two methods (learning ontology from texts and the structure of the ontology structure) building an ontology for a bilingual domain precisely in the knowledge of Alzheimer's domain.

In general, ontologies are used in many applications such as the expansion of the query. Question extensions apply to the IRS where new terms can be added to the user's query to improve and increase the effectiveness of the recovery process. Investigators are examining the questionnaire by extension of the question; tax class questionnaires provided in (Dipasree et al., 2015). Related working on the expansion of the question can be broadly divided into three groups: global, domestic and foreign.

In the expansion of the global questionnaire, the corporation as a whole is considered to be choosing the principles of expansion. An international technology proposed in (Jing & Croft, 1994) based on co-occurrence data in corpus, they choose extension words that are very similar to the question. In local QE strategies, terms are selected from the first set of texts returned in response to the original question for example is the coherence answer (Bilel et al., 2011; Picariello et al., 2007). In the program lack of user feedback, assuming that a few high-quality texts are appropriate for a cohesive response. Pragati et al. (2014) developed a fake qualification limit the expansion of the questionnaire in response by suggesting the integration of corporate-based

information with a fuzzy genetic approach and the concept of semantic similarity. Colace et al. (2015) suggested new way to extend a question based on a weighty two-word process. This structure has been removed from a set of documents received with compliance response and added to the original question. In external QE techniques, researchers incorporated the concept of semantics through foreign language knowledge such as WordNet ontology. Abbache et al. (2014) used Arabic WordNet extensions Arabic query by adding similar words to the original queries; the method does not give better results compared to the method of interaction and therefore in (Abbache et al., 2016) use a method that selects automatic synonyms extracted from Arabic WordNet in accordance with the merging rules, this method improves the return results in relation to means standard specification (MAS). Tests show that in a good way of similar words to choose from Arabic WordNet as a source of language knowledge in an automatic query expansion improves the efficiency of acquisition of Arabic knowledge. General ontology was presented in (Audeh et al., 2014) for the development of named businesses using the "Yago" ontology.

III. Performance Evaluation

We have used the benchmark dataset i.e. Trec-3 and implemented up to fifty queries on it and compared it to few similar approaches. We calculated the F-measure and precision parameters and compared them to few other similar approaches.

Query Wise Retrieval effectiveness:

In this paper, we've got computed F-Measure to examine the overall performance of every question of proposed technique and as compared the outcomes with different approaches. We have decided F-degree fee at 3 cut-offs. These cut-offs are top ten, twenty, fifty retrieved documents for TREC-3 dataset. Better F-measure value is obtained in proposed approach as compared to other similar approaches.

Fig.2 demonstrates the F-measure for top ten queries on TREc-3 data set. Fig. 3 shows the comparison for TREC-3 at top twenty retrieved documents cut-off. This figure shows that better F-measure are obtained by proposed approach approach over FLTSBAQE, Tomiye et al. approach, Parapar et al. approach and original query. F-measure values of FLTSBAQE approach are equal to Parapar et al. approach for one query only. Fig. 4 demonstrates the results for TREC-3 at top fifty cut-off.

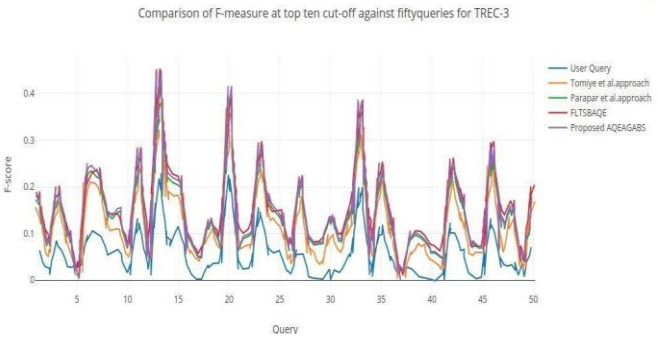

Fig 4: Top Ten cut-off queries

Fig 4. Demonstrates the F-measure comparison for top 10-cutoff against fifty queries comparison with original user query, tomiye et.al, parapar et al approach with proposed approach at TREC-3 dataset.

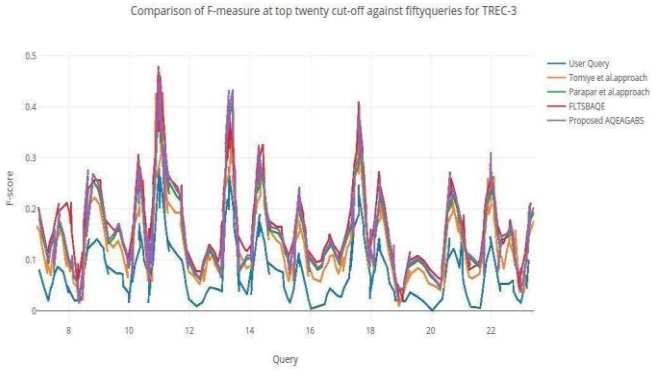

Fig 5: Top twenty cut-off queries

Fig 5. Demonstrates the F-measure comparison for top 20-cutoff against fifty queries comparison with original user query, tomiye et.al, parapar et al approach with proposed approach at TREC-3 dataset.

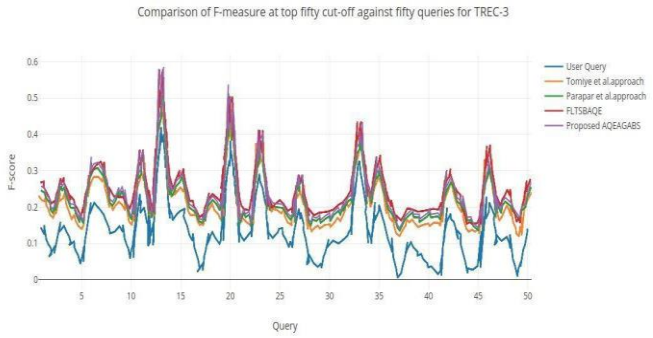

Fig 6: Top fifty cut-off queries

Fig 6. Demonstrates the F-measure comparison for top 50-cutoff against fifty queries comparison with original user query, tomiye et.al, parapar et al approach with proposed approach at TREC-3 dataset.

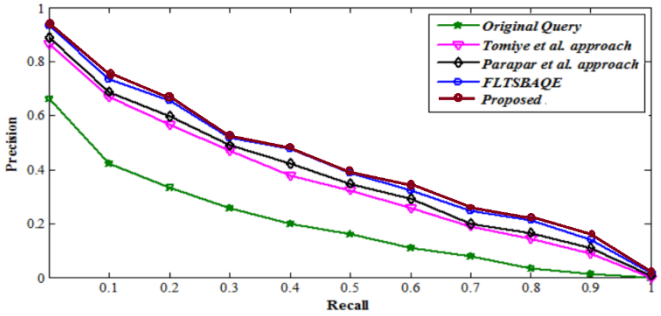

Fig 7. Precision-Recall curves of all approaches for TREC-3

Figure 7 demonstrates the Precision –Recall curve for original query, Tomiye et al. approach, Parapar et al.approach, FLTSBAQE and proposed approach. The proposed approach has the highest precision recall curve.

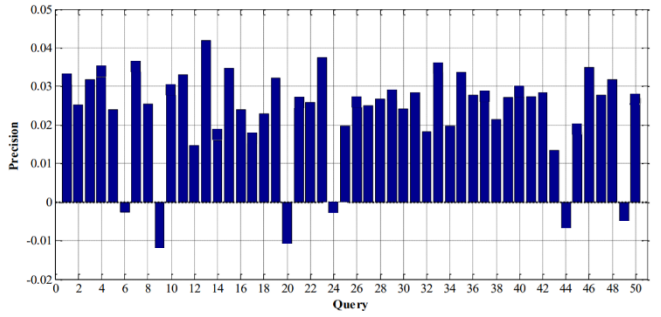

Fig. 8.Query-wise precision variation of proposed model against Parapar et al. approach for TREC-3

Figure 8 depicts the query wise precision variation of proposed model against Parapar er al. approach for TREC-3 dataset.

IV. Conclusion

With loads and heaps of data to draw from, users have always wanted to draw upon the most relevant facts. Information retrieval is the entire process of finding relevant documents. A system designed to retrieve information use query expansion methods in order to provide the most relevant result for the user. While many query expansion techniques have been used, a number of them rely on specific data sources (for example, a corpus, a knowledge base, or pseudo-relevance feedback), these types of techniques have just recently been presented. In terms of user queries, they use synonyms in accordance to its meaning to extract the relevance and they also look for papers containing the search word but with a different meaning. This paper suggests a mix of evolutionary algorithms and fuzzy logic to get results. We also demonstrated a novel strategy that suppresses the impact of corpora- or knowledge-based approaches. Computed F-Measure has been used to evaluate every question and method and to assess how they performed in comparison. F-degree charge will be implemented at 3 cut-offs. These papers have been selected as the ten most valuable, twenty most valuable, and fifty most valuable for the TREC-3 dataset. Using the suggested method, a higher F-measure value is achieved.

V. References

[1]. G. W. Furnas, T. Landauer, L. Gomez, and S. Dumais, “The vocabulary problem in human-system communication,” Communications of the ACM, vol. 30, no. 11, pp. 964-971, 1987.

[2]. H. Chen, J. Yu, K. Furuse, and N. Ohbo, “Support IR query refinementby partial keyword set,” Proceedings of the second international conference on web information systems engineering, Singapore, pp. 245–253, 2001.

[3]. B. Kim, J. Kim, and J. Kim, "Query term expansion and reweighting using term co-occurrence similarity and fuzzy inference," Proceedings of the joint ninth IFSA world congress and 20th NAFIPS international conference, Vancouver, Canada, pp. 715–720, 2001.

[4]. Chang, S. Chen and C. Liau, "A new query expansion method based on fuzzy rules," Proceedings of the seventh joint conference on AI, Fuzzy system, and Grey system, Taipei, Taiwan, Republic of China, 2003.

[5]. B. Yates, and R. Neto, "Modern Information Retrieval," Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 1999.

[6]. Y. Bade, R. Bhat, and P. Borate, "Optimization techniques for improving the performance of information retrieval system," International Journal of research on advanced technology, vol. 2, no. 2, pp. 263–267, 2014.

[7]. K. C. Thompson, "Reducing the risk of query expansion via robust constrained optimization," Proceeding of the 18th ACM conference on Information and knowledge management, NY, USA, pp. 837–846, 2009.

[8]. K. Raman, R. Udupa, P. Bhattacharyya, and A. Bhole, "On improving pseudo-relevance feedback using pseudo-irrelevant documents," Proceedings of ECIR, pp. 573–576, 2010.

[9]. R. White, and G. Marchionini, "Examining the effectiveness of real-time query expansion," Information Processing and Management, vol. 43, no. 3, pp. 685–704, 2007.

[10]. Z. Ye, J. Huang, and H. Lin, "Finding a good query-related topic for boosting pseudo-relevance feedback," Journal of the association of Information Science and Technology, vol. 62, no. 4, pp. 748–760, 2011.

[11]. J. Singh, and A. Sharan, "Context window based co-occurrence approach for improving feedback based query expansion in information retrieval," International Journal of Information Retrieval, vol. 5, no. 4, pp. 31–45, 2015.

[12]. Y. Li, S. Chung, and J. Holt, "Text document clustering based on frequent word meaning sequences," Data and Knowledge Engineering, vol. 64, no. 1, pp. 381–404, 2008.

[13]. J. Aguera, and L. Araujo, "Comparing and combining methods for automatic query expansion," Advances in natural language processing and applications, research in computing science, vol. 33, pp. 177–188, 2008.

[14]. M. Valdivia, M. Galiano, A. Raez, and L. Lopez, "Using information gain to improve multi-modal information retrieval systems," International Journal on Process Management, vol. 44, no. 3, pp. 1146–1158, 2008.

[15]. C. Carpineto, and G. Romano, "A survey of automatic query expansion in information retrieval," ACM Computer Survey, vol. 44, no. 1, pp. 1–50, 2012.

[16]. A. Tomiye, A. Samuel, A. Ijesunor, and I. Udo, "A fuzzy-ontology based information retrieval system for relevance feedback," International Journal of Computer Science, vol. 18, no. 1, pp. 382-389, 2011.

[17]. J. Parapar, M. Quindimil, and A. Barreiro, "Score Distributions for Pseudo Relevance Feedback," Information Sciences, vol. 273, pp. 171-181, 2014.

[18]. S. Robertson, "On term selection for query expansion," Journal of documentation, vol. 46, no. 4, pp. 359-364, 1990.

[19]. J. Swets, "Information retrieval systems," Journal of Science, vol. 141, pp. 245-250, 1963.

[20]. C. Lee, "Fuzzy logic in control systems: Fuzzy logic controller, Parts I and II," IEEE Transaction on System, Man and Cybernetics, vol. 20, pp. 404–435, 1990.

[21]. Y. Gupta, and A. Saini, "A novel Fuzzy-PSO term weighting automatic query expansion

approach using semantic filtering," Knowledge Based System, vol. 136, pp. 97-120, 2017.

[22]. Gupta, Yogesh, and Ashish Saini. "A novel term selection based automatic query expansion approach using PRF and semantic filtering." International Journal of Engineering and Advanced Technology 8.C (2019): 130-137.

[23]. Singh, Saharan," Rank fusion and semantic genetic notion based automatic query expansion model." Swarm and Evolutionary Computation. Vol. 1(38),pp-295-308,Feb 2018.