ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 19, April 2017

IMPROVED KEYWORD EXTRACTION USING SEMANTICS FOR QUERY RETRIEVAL

S.Vilma Veronica, S.Vincy, S.Padmavathi
UG Scholar, Kings Engineering College, Sriperumbudur, Chennai -602117.
Vil_95@yahoo.com,vbrs95@gmail.com, padmasree095@gmail.com

ABSTRACT

A huge amount of high quality question and answering (QA) pairs has been accumulated as comprehensive knowledge bases of human intelligence. It helps users to seek precise information by ranking[2] lists of results. Hence to retrieve relevant questions and their corresponding answers becomes an important task for information acquisition. Here we define question retrieval in CQA (Community Question and Answering) [7] services as a task in which new questions are used as queries to find relevant questions for which the answers are already available. For simplicity and consistency, we use the term "query" to denote new question posed by question retrieval in CQA[7] is different from general web search. Unlike the web search engines that return a long list of ranked documents, question retrieval can also be considered as a traditional Question Answering (QA) problem [1], but the focus of the QA task is transformed from answer extraction, answer matching and answer ranking [2] to search for relevant question with ready answers.

1. INTRODUCTION

The main aim of this paper is to capture the meaning of key phrases of the given question and retrieving the relevant descriptions by jointly utilizing local mining and global learning approaches.

This paper is completely based on information retrieval, which retrieves the information from a huge collections of data.

This in turn helps the user to retrieve the information and data which is more relevant for the user's query.

Supervised By Mrs. M.Vidhya Assistant Professor (Computer Science) Kings Engineering College, Chennai.

In this paper we post a query and get the result in our desired language. In this paper we have suggested an NLP (Natural Language Processor) technique which translates query to the desired language and also gets the answer from other users or experts finally the same answer is converted to the language in which we asked. The search is done efficiently, based on the key term of the query and the noun phrase and relevant results are acquired.

The queries can be answered by both other users and experts. If the user is not satisfied with the other users answers they can directly ask to the experts.

2. EXISTING SYSTEM

In the existing system, the major challenge is the word verboseness in the queries where important words may be surrounded by other additional words the word mismatch problem is even more common and severe than in general search.

Existing work mainly focused on core term discovery, query reformation, key concept identification on verbose queries. Despite the great success achieved these project mainly focused on distinguishing the key concepts from the non-key concepts and the importance among the key concept was not taken into consideration. In this paper, we propose a ranking based[2] method for key concept identification, which not only distinguishes the key concepts but also tackle the word mismatch problem, previous work mainly resorts to query expansion[3]. However, the former approach overlooks concept level evidences for query expansion[3] and fails to assign explicit weights to the expanded aspects and further reduce the performance of question retrieval. However, both of them are based on the term level expansion (Zhou et al) and the dependence between terms is not considered. However, the phrase based translation model makes little or no direct use of syntactic information, which leads to the limitation on the translation performance and further impact the question retrieval results.

2.1 PROBLEM DEFINITION

- Identification of Key Terms
- Word Mismatch Problem
- Lack of Artificial Intelligence

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 19, April 2017

- Lack of Machine Learning.
- Time Delay for answers

3.PROPOSED SYSTEM

We propose a novel scheme that is able to find the key concepts and also to give the ranking based relevant answers [2]. The main contributions of the proposed works are threefold:

To the best of our knowledge, this is the first thorough study of using multiple languages to bridge the semantic gaps in question retrieval task.

Second, we propose a ranking- based approach [2] to capture the important levels of key concepts in the target questions, which significantly outperforms the state-of-the-art binary classification approach.

Third, we demonstrate the usability of the paraphrase model to be compatible with existing question retrieval models, and show that it contributes additional semantic connection among the key concepts in the query and the retrieved questions.

So this paper helps its users to get the relevant data for the user's questions and the data is retrieved at high rate of speed since it works in offline.

4.MODULES

4.1 Q and A Application
4.2 Key Concept Detection
4.3 Bridging Gap
4.4 Machine Learning

4.1 Q AND A APPLICATION

Generally, In Existing web applications the questions posted by the users are answered by the other user which might result in redundancy and user un- believability especially for medical related doubts, clarifications and questions. So a Medical Experts who can give believable answers and will be available all the time but which is practically not possible and time consuming .So we build an efficient Q and A Scheme[5] which could give instant answers by analysing the users objective behind the question. In this, questions will be asked in many languages. it will be translated to the user with the respective language.The user can also see the review question and answers.

4.2 KEY CONCEPT DETECTION

The user questions is processed by a Natural Language Processing (NLP) technique for instant answers. So that the proper meaning is retrieved. The NLP process comprises a several steps. Of which Parts Of Speech Tagging (POST) results in Phrases and Nouns Extraction[7]. The Keywords thus extracted is subject to Stemming Process which eliminates the stop words in the sentence and also trims the keyword for Base Word.

4.1 BRIDIGING THE ANSWERS

The Proper meanings will be analysed with an English Dictionary and the Medical Terms will be Normalized based on Domain Specific Knowledge. Medical Terminologies were collected and grouped so that the checking with the synonyms of keywords[7] could result in Normalization. The Normalized words will be checked for contradictions with medical terminologies and the related answers will be queried from Local Mining Database.

For simplicity and consistency, we use the term "query" to denote new question posed by question retrieval in CQA[7] is different from general web search. Unlike the web search engines that return a long list of ranked documents, question retrieval can also considered as a traditional question.

4.4 MACHINE AND LANGUAGE TRANSLATOR

`Machine Learning in our approach is achieved by the use of Local Mining and Global Learning techniques. Local Mining database gets updated by the Global Learning data's once user posts a new kind of query to the answering system. The Global learning comprises a large collection of Medical related resources in its backend which helps to retrieve a related resource to the query based on terminology keywords.

This Search is completely indexed a thus the retrieval time is faster. In case of resource insufficiency the Query and the Question will be left in pending state till an expert arrives.

We propose a novel scheme that is able to find the key concepts and also to give the ranking based relevant answers [2].

Once Experts reviewed the query and the response are reported to the Medical Seekers and also update the Local Mining Database for future instant retrieval to the related Query from other users. [4] discussed about a method, This scheme investigates a traffic-light-based intelligent routing strategy for the satellite network, which can adjust the pre-calculated route according to the real-time congestion status of the satellite constellation. In a satellite, a traffic light is deployed at each direction to indicate the congestion

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 19, April 2017

situation, and is set to a relevant color, by considering both the queue occupancy rate at a direction and the total queue occupancy rate of the next hop.

5.ALGORITHMS

- Clustering
- Normalization
- NLP (Using POS Tagger ,Word Net and Spell Checker)
- Indexing (Using Lucene Library)
- Bubble Sort (For Prioritizing Results)
- Machine Learning

5.1 BUBBLE SORT

Bubble sort is a simple and well-known sorting algorithm. It is used in practice once in a blue moon and its main application is to make an introduction to the sorting algorithms. Bubble sort belongs to $O(n^2)$ sorting algorithms, which makes it quite inefficient for sorting large data volumes. Bubble sort is stable and adaptive**.**

5.1.1 ALGORITHM

1. Compare each pair of adjacent elements from the beginning of an array and, if they are in reversed order, swap them.
2. If at least one swap has been done, repeat step 1.

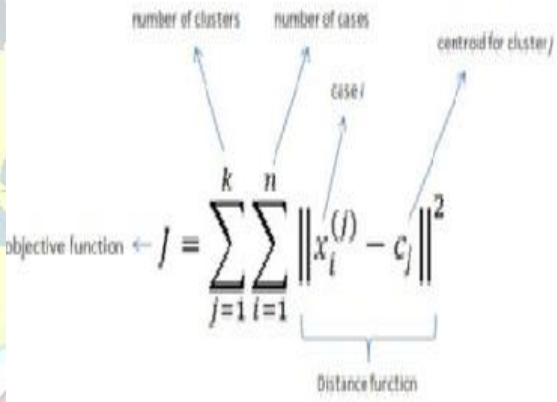
Java

```
public void bubble sort(int[] arr)
{
boolean swapped = true;
int j = 0;
int tmp;
while(swapped)
{
swapped = false;
j++;
for(int i=0;i<arr.length-j;i++)
{
if(arr[i]>arr[i+1])
{
tmp-arr[i];
arr[i]=arr[i+1];
arr[i+1]=tmp;
swapped = true;
}
}
}
}
```

5.2 CLUSTRING ALGORITHM (K- MEANS CLUSTERING)

5.2.1 Algorithm

1. Clusters the data into *k* groups where *k* is predefined.
2. Select *k* points at random as cluster centers.
3. Assign objects to their closest cluster center according to *Euclidean distance* function.
4. Calculate the centroid or mean of all objects in each cluster.
5. Repeat steps 2, 3 and 4 until the same points are assigned to each cluster in consecutive rounds



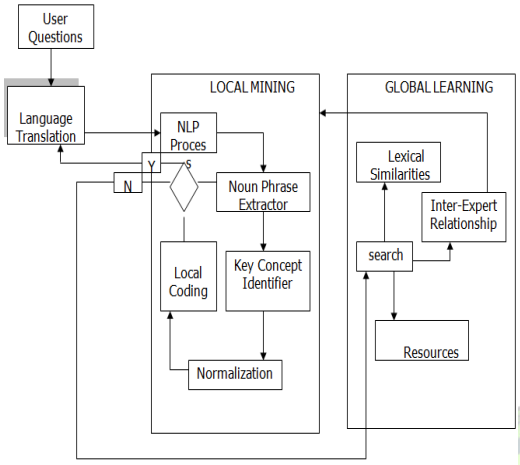
6.ARCHITECTURE DIAGRAM

This figure shows how, the local mining and Global mining process is working .When the query is given it is processed by NLP Processor and then translates the query to the necessary language and the key term is extracted from the query using the noun phrase .The Search is made accordingly that the user can ask answer from other user using Local Mining technique.Else the user can ask their query to the experts especially the Doctors through the technique called Global Learning .Each block of this paper has a unique feature and retrives the results as quick as possible.

The search results are more relevant to the user's queries because the search is made based on the extracted keyword from the query and based on the query the noun phrase is extracted using the noun phrase extractor and the search results are produced on the bases of keyterm and noun phrase.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 19, April 2017



7.ENHANCEMENT

- ✓ Separate Server Implementations for Local Mining and Global Learning
- ✓ Medi Net Library
- ✓ NLP Techniques
- ✓ PDF Searching using Indexing
- ✓ Artificial Intelligence
- ✓ Multiple language Translation

8. CONCLUSION

In this paper, we proposed a key concept paraphrasing based approach to effectively tackle the major problems of word verboseness and word mismatch in question retrieval by exploring the translations of pivot languages. Further, we expanded queries with the generated paraphrases for question retrieval.

9.RESULT

This is the final output of this paper which gets the answers from the doctors and the the result is transulated into the desired language of the user.

9. REFERENCES

[1] Xiaobing xue, "*Retrieval Models for Question and Answer Achieve*" USAxueub@cs.umass.edu.

[2] Jae-Hyun Park and W. Bruce Croft, *"Query Term Ranking based on Dependency Parsing of Verbose"* Center for Intelligent Information Retrival,USA.

[3] Ashish Kankaria, *"Query Expansion Technicals"*, Indian Institute of Technology Bombay, Mumbai.

[4] Christo Ananth , P.Ebenezer Benjamin, S.Abishek, "Traffic Light Based Intelligent Routing Strategy for Satellite Network", International Journal of Advanced Research in Biology, Ecology, Science and Technology (IJARBEST), Volume 1,Special Issue 2 - November 2015, pp.24-27

[5] Jiwoon Jeon, W. Bruce Croft and Joon Ho Lee, "*Finding Similar Questions in Large Question and Answer Archives"*, Center for Intelligent Information.