ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 4, April 2018

Retrieving Relevant Answer from Large Questions- Answer Dataset by Using Semantic Analysis and Natural Language Processing

Anusha T K[1], Hilda Jerlin C M[2], Jayashree M[3], S Muthumariammal M.Tech.,[4], Dr. R. Janarthanan [5]

B.E. Scholar, Department of CSE, T.J.S Engineering College, Chennai, India [1,2,3]

Assistant Professor, Department of CSE, T.J.S Engineering College, Chennai, India [4.]

Professor, Department of CSE, T.J.S Engineering College, Chennai, India[5.]

ABSTRACT: This paper focuses on addressing the lexical gap problem in question retrieval. Question retrieval in Community question answering(cQA) archives aims to find the existing questions that are semantically equivalent or relevant to the queried questions. However, the lexical gap problem brings a new challenge for question retrieval in cQA. In this paper, we propose to model and learn distributed word representations with metadata of category information within cQA pages for question retrieval using two novel category powered models. One is a basic category powered model called MB-NET and the other one is an enhanced category powered model called ME-NET which can better learn the distributed word representations and alleviate the lexical gap problem. To deal with the variable size of word representation vectors, we employ the framework of fisher kernel to transform them into the fixed-length vectors. Experimental results on large-scale English and Chinese cQA data sets show that our proposed approaches can significantly outperform state-of-the-art retrieval models for question retrieval in cQA. The evaluation results show that promising and significant performance improvements can be achieved.

KEYWORDS: Question and Answer Retrieval, Language Model, Local Mining, Global Learning

I. INTRODUCTION

The main aim of this project is to solving lexical gap problem in community question answering (Question answer forum) and find the existing questions that are semantically equivalent or relevant to the queried questions by using Natural Language Process (NLP).Over the past few years, a large amount of user-generated content has become an important information resource on the Web. These include the traditional Frequently Asked Questions (FAQ) archives and the merging community question answering (cQA) services, such as Yahoo! Answers, Live QnA, and Baidu Zhidao. The content in these web sites is usually organized as questions and lists of answers associated with metadata like user chosen categories to questions and asker's awards to the best answers. This data made cQA archives valuable resources for various tasks like question-answering and knowledge mining, etc.The best answers of similar questions in cQA will be used to answer the queried questions this is what known as question retrieval.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 4, April 2018

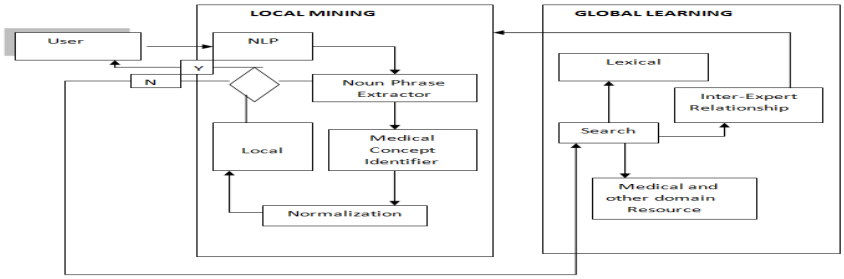

Fig.1Architecture Diagram

In many cases, the community generated content, however, may not be directly usable due to the vocabulary gap. Take Health-Tap as an example, which is a question answering site for participants to ask and answer health-related questions. The questions are written by patients in narrative language. The same question may be described in substantially different ways by two individual health seekers. On the other side, the answers provided by the well-trained experts may contain acronyms with multiple possible meanings, and no standardized terms. Recently, some sites have encouraged experts to annotate the medical records with medical concepts. However, the tags used often vary wildly and medical concepts may not be medical terminologies.

For example, “heart attack” and “myocardial disorder” are employed by different experts to refer to the same medical diagnosis. It was shown that the inconsistency of community generated health data greatly hindered data exchange, management and integrity. Even worse, it was reported that users had encountered big challenges in reusing the archived content due to the incompatibility between their search terms and those accumulated medical records. Therefore, automatically coding the medical records with standardized terminologies is highly desired. All the technical question solved by well know technical person where the user not asking question in same way that the will be ask in different way but the answer is same this is known as lexical gap.

II. RELATED WORK

Burke et al. combine lexical similarity and semantic similarity between questions to rank FAQs. Berger et al. study several statistical approaches to bridge the lexical gap for FAQ retrieval, including a translation-based approach anda query expansion approach using a word mapping learned from a collection of question-answer pairs. Jijkoun and Rijke use a vector space model for question-answer pair retrieval. Riezler et al. provide a translation model for question retrieval; their translation model is trained on a large amount of data extracted from FAQ pages on the Web. Soricut and Brill build an answer passage retrieval system. Question retrieval has recently been revisited for Community Question Answering (CQA) data. Duan et al. propose a solution that makes use of question structures for retrieval by building a structure tree for questions in a category of Yahoo! Answers, meaning that important phrases in questions are given higher weight in question matching. Wang et al. employ a parser to build syntactic trees of questions, and questions are ranked based on the similarity between their syntactic trees and the syntactic tree of the query question.

As observed by Wang et al., current parsers are not well-trained to parse reallife questions, especially informally stated questions. They report that they outperform a Bag-of-Word baseline (that matches words between query and question) by 11.99% in terms of MAP. Bian et al. propose an interesting learning framework for question retrieval However, this approach needs training data (which would be difficult to get for general questions) and experiments are conducted on factoid questions.Recently, the research of question retrieval hasbeen further extended to the cQA data.

Zhouet al. (2011) proposed a phrase-based translationmodel for question retrieval and obtained the stateof-the-art performance.However, all the existing methods ignore the latenttopics information in calculating the semanticsimilarity between questions. In this paper,we present a new approach to discover the latenttopic of questions for improving the performanceof translation-based language models for questionretrieval. Moreover, we introduce the category informationinto the process of discovering the latenttopics. To the best of our knowledge, none ofthe existing

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 4, April 2018

studies addressed question retrieval incQA by learning the latent topics.Chekuri and Raghavan introduce the idea of utilizing classificationfor document retrieval. However, their focus is on the automaticclassification of Web documents (no category informationavailable) into high-level categories of the Yahoo! taxonomy; theydo not investigate how to leverage the classification results for documentretrieval. Similarly, Lam et al. also focus on developingan automatic text categorization approach.Finally, we note that other works exists on CQA services thatconsider category information, e.g., [1, 13]. However, these focuson other aspects of CQA, not on question retrieval.

III. EXISTING SYSTEM

Medical blogs are restricted to general user answers as it might lead to negative results on human lives. It is worth mentioning that there already exist several efforts dedicated to research on automatically mapping medical records to terminologies. Most of these efforts, however, focused on hospital generated health data or health provider released sources by utilizing either isolated or loosely coupled rule-based and machine learning approaches. Compared to these kinds of data, the emerging community generated health data is more colloquial, in terms of inconsistency, complexity and ambiguity, which pose challenges for data access and analytics.

There are two different models first translation models, which leverage the question-answer pairs to learn the semantically related words for improving traditional IR models. The basic assumption is that question answer pairs are “parallel texts” and relationship of words can be established through the word-to-word translation probabilitiesThe **Disadvantages** as follows**:**Time Delay for answers, Ambiguity in key terms**,** Lexical gap, Lack of Artificial Intelligence and Lack of Machine Learning.

IV. PROPOSED SYSTEM

In cQA, the systems can directly return answers to thequeried questions instead of a list of relevant documents,thus providing an effective alternative to the traditional adhocinformation retrieval.Unlike in Web search, opinionbased queries are also answered here by expertsand users based on their personal experiences. Thequestion and answers are also enhanced with richmetadata like categories, subcategories, user expertlevel, user votes to answers etc. To make full use of the large scalearchives of question-answer pairs, it is critical to have functionalityhelping users to retrieve historical answers. Therefore, it is a meaningful task to retrieve thequestions that are semantically equivalent or relevant to thequeried questions.

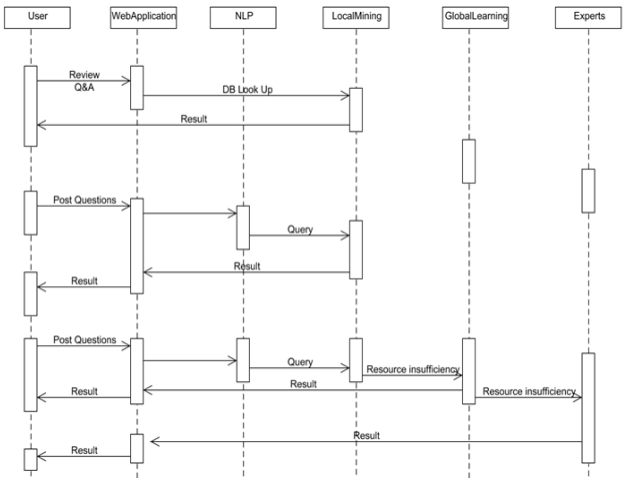

Fig.2 Sequence Diagram

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 4, April 2018

As illustrated in Figure 2, the proposed scheme contains multiple domain QA (question answer) forums.To the best of our knowledge, this is the first work on automatically coding the community generated question answer data, which is more complex, inconsistent and ambiguous compared to the single domain QA forum.It proposes the concept entropy impurity approach to comparatively detect and normalize the medical concepts locally, which naturally construct a corpus-aware terminology vocabulary with the help of external knowledge. It builds a novel global learning model to collaboratively enhance the local coding results. This model seamlessly integrates various heterogeneous information cues.

V. METHODOLOGY

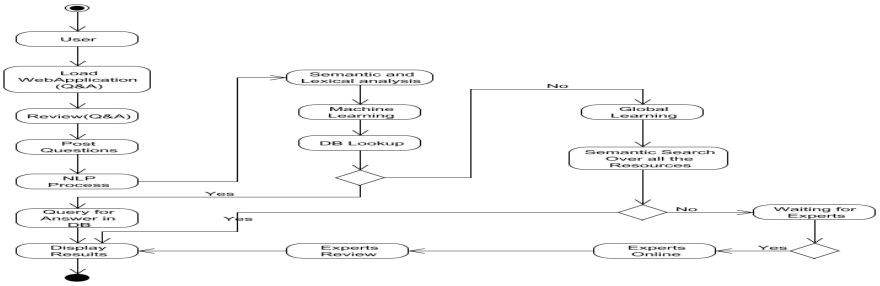

Fig.3 Activity Diagram

We propose a novel scheme that is able to code the medical records with corpus-aware terminologies. The main contributions of this work are as follows

1. Q and A Application
2. Information Retrieval
3. NLP Process
4. Bridging lexical Gap
5. Machine Learning

1. Q AND A APPLICATION:

Question Answering(QA) application is designed based on the technique from informationretrieval and natural language processing (NLP), which arerelated to developed systems that automatically answerquestions posed by humans in a natural language rather thanthe keyword based retrieval mechanism.This is similar to Web search. There are many disadvantages of Websearch. It gives lot of web pages for searching a single word. It is time consuming. Automated Question Answering (QA), the technology that locates,extracts, and represents a specific answer to a user question expressed in natural language, is now being studied veryactively by researchers and practitioners. The overall aim of this QA track was to retrieve small pieces of text that contain the actual answer to the questionrather than the list of documents traditionally returned by retrieval engines.Generally, we can avoid redundancy and user un-believability especially for medical related doubts, clarifications and questions. In medical sites a medical Experts who can give believable answers should be available all the time which is practically not possible and time Consuming .So we build a Efficient Q and A Scheme which could give Instant Answers Analyzing the users Objective behind the Question.

2. INFORMATION RETRIEVAL:

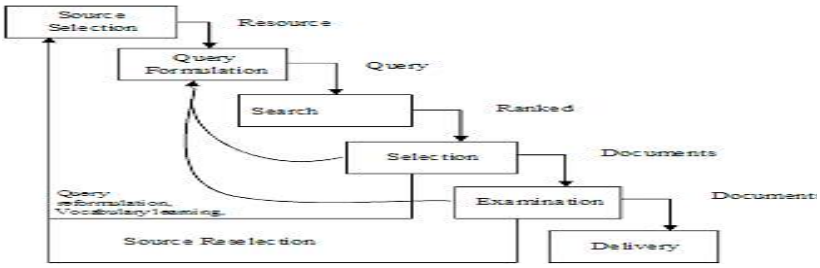

Fig.4 Information Retrieval Cycle

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 4, April 2018

Information retrieval is an essential component in Community QuestionAnswer (CQA) services. The retrieval task in a Q&A archive is to find relevant question-answer pairs for new questions posed by the user. To solve the word mismatch problem, we focus on translation-based approaches since the relationships between words can be explicitly modeled through word-to-word translation probabilities.Besides designing the translation based retrieval model, another important problem is how to learn good word-to-wordtranslation probabilities. In a Q&A archive, since theasker and the answerer may express similar meanings withdifferent words, it is natural to use the question-answer pairsas the"parallel corpus". Since the question part and answer part arewritten in the same language, the word-to-word translationprobabilities can be learned with either part as the sourcelanguage and the other part as the target. Intuitively, thesame word will be related to different sets of words whicheverpart it appears in. Thus, combining the word-to-word translationprobabilities learned with different source and targetconfigurations is beneficial.

InformationRetrieval deals with the representation, storage, organizationand access to information items. The basic architecture of Question- Answering systems isbased on an information retrieval system that considers theword of the questions as queries to retrieve the appropriateanswers

Translation Models: Learning wordor phrase level translation models fromquestion-answer pairs in parallel corpora ofsame language. Retrieving semantically similar questions can bethought of as a classification problem with largenumber of categories. Here, each category containsa set of related questions and the numberof questions per category is small. It is possiblethat given a test question, we find that there are noquestions semantically related to it in the archives,it will belong to a entirely new unseen category.Thus, only a subset of the categories is known duringthe time of training.

For a practical Q&A retrieval system, the search resultsshould be presented to the user in a hierarchical way, wherea list of questions is first presented, and after the user selectsa specific question the corresponding answers are presented.This hierarchical strategy is more effective for the user thandirectly presenting a list of question-answer pairs, since thisresult list could easily be overwhelmed a single question withmany answers. Since the relevance of the answer to its correspondingquestion is usually guaranteed (with the exception of "spam" answers), the retrieval performance of a systemcan be measured by the rank of relevant questions it returns.Thus, in our experiments, relevance judgments are based onquestions. This observation can be explained as follows. Sometimes,different wording to the concept of the question canbe directly observed in the corresponding answer, thus thequery likelihood language model for the answer can be consideredas a good complement to the translation-based languagemodel for the question.

VI. NLP PROCESS

Natural Language Processing (NLP) is the computerized approach to analyzing text that is based on both a set oftheories and a set of technologies. Its goal is "to accomplish human-like language processing". The choice of the "word processing" is verydeliberate, and should not be replaced with "understanding". NLP techniques are used in applications that make queries to databases, extract information from text, retrieverelevant documents from a collection, translate from one language to another, generate text responses, or recognizespoken words converting them into text. NLP is a very difficult task because human being has good common sense andreasoning mechanism which they use to answer the question.The most commonapplications utilizing NLP include the following: systems ofmachine translation, information retrieval (IR), etc.

Here, the User posts Questions for instant answers is processed by a natural language processing technique so that the proper meaning would be revealed. The NLP Process comprises a several steps of which Parts Of Speech Tagging (POST) results in Phrases and Nouns Extraction. 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.

VII. BRIDGING LEXICAL GAP

The Proper meanings will be analyzed with a English Dictionary and the Medical Terms will be Normalized based on Domain Specific Knowledge. Different domain Terminologies were Collected and grouped so that the checking with the synonyms of keywords 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.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 4, April 2018

VIII. MACHINE LEARNING

Machine Learning in Our Approach is achieved by the use of Local Mining and Global Learning techniques in medical domain.To overcome the vocabulary gap, a new scheme is used which combines two approaches namely local mining and global learning.

Local mining extracts medical concepts from medical records and then map them to normalized terminologies based on standardized dictionary. Local mining suffer from problem of missing key concepts. Global learning overcome the issues in local mining by finding the missing key concepts.Local Mining database gets updated by the Global learning data's once user posts a newer 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 an thus the retrieval time is faster. In case of resource insufficiency the Query and the Question will be left in pending state till a expert arrives. Once Experts reviewed the query the answers not only dispatches to the Medical Seekers and also updates the Local Mining Database for future instant retrieval to the related Query from other Users, other domain answers retrieved from dataset.

Medical sites are among the most popular internet sites today through which people can get more knowledge about their health conditions. Most of the medical sites such as mayo clinic, Medscape are consumer oriented and provide their sound advice about general medical topics. Due to tremendous number of records have been accumulated in their repositories and in most circumstances user may directly locate good answers by searching rather than waiting for experts to answer. However users with diverse background do not necessary share same vocabulary, the same question may be written in different native languages which is difficult for other health seekers to understand, so to bridge vocabulary gap corpus aware terminology is used.

A. Local Mining

Local Mining involves a three stage framework. First stage is the noun phrase extraction in which the noun phrase are extracted. In the second stage the medical concepts are detected using concept entropy impurity(CEI). CEI also measures the specificity of that concept in the particular domain. Finally,normalization is performed. Normalizes the medical concepts based on the authenticated vocabulary.

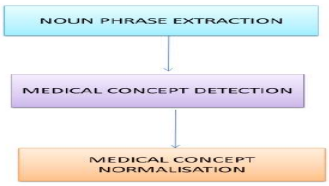

Fig.5 Three Stage Framework In local Mining

(i) Noun Phrase Extraction: As text is an unstructured source of information, to make it a suitable input to an automatic method of information extraction it is usually transformed into a structured format. Part of Speech Tagging is one of the preprocessing steps which perform semantic analysis by assigning one of the parts of speech to the given word. Part-of-speech taggingis the process of marking up a word in a text. It is the process of assigning a part-of speech marker to each word in an input text. In noun phase extraction, it takes the speech types part into account. In this process many unwanted words are stopped because that words are uninterested meaning. To extract the noun phrases, speech tags are assigned by Stanford POS tagger to every word of medical record given by the user. The noun phrases should contain zero or more adjectives or nouns, followed by an optional group of a noun and a preposition, followed all over again by zero or more adjectives or nouns, followed by a single noun. To make up a noun phrase, sequences of tags are matched in a pattern.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 4, April 2018

Fig.6 Parts of Speech Tagger

(ii) Medical Concept Detection: It detects the medical concepts and differentiates it from other phrases. Concept Entropy Impurity is used to analyses the specificity of the medical concept. Larger CEI value indicates more relevant the concept in that domain.

(iii) Medical Concept Normalization: Medical concepts may not be standard terminologies so it is necessary to normalize the concepts based on a authenticated vocabulary. Local mining suffer from incompleteness due to the missing key concepts. Second problem is the lower precision this is due to the irrelevant medical concepts in the records.

B. Global Learning

An enhanced and novel approach of global learning isbeing built for enhancing the result of local coding.

(i) Relationship Identification: The Inter-terminologyand Inter-expert relationships are analyzed from the medical records.

(ii)Inter-Terminology Relationship: The terminologiesin SNOMED-CT are arranged in hierarchies. For example,viral pneumonia is-an infectious pneumonia is-pneumonia is-a lung disease. Terminologies may also have multiple parents. For example, infectious pneumonia is also a child of infectious disease[1]. This hierarchial representation improves the coding.

(iii)Inter-Expert Relationship: Analyses the historical data of experts and checks whether the experts are in the same or related area. Jaccard coefficient is used to analyze the experts relationship .

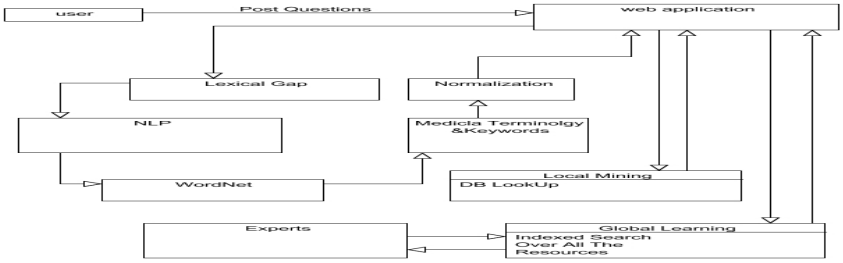

Fig.7Collaboration Diagram

IX. TECHNOLOGIES USED

a) JAVA

It is Platform Independent and object-oriented programming language intended to replace C++, although the feature set better resembles that of Objective C.

Java consists of two things:

- Programming language
- Platform

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 4, April 2018

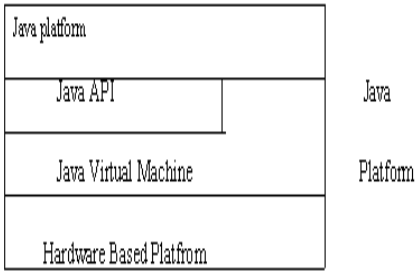

Fig.8JAVA Platform

With a compiler, you translate a Java program into an intermediate language called **Java byte codes** – the platform independent codes interpreted by the Java interpreter. With an interpreter, each Java byte code instruction is parsed and run on the computer. Compilation happens just once; interpretation occurs each time the program is executed. This figure illustrates how it works :

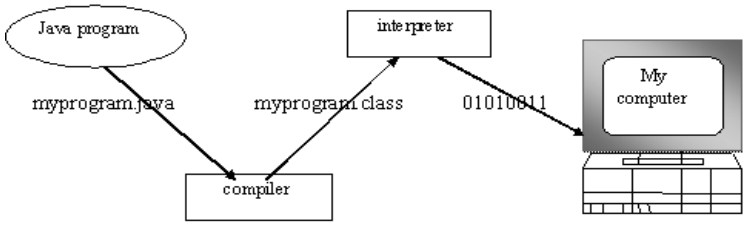

Fig.9 Implementation of JVM

b) Cloud Computing

Cloud computing is a style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet. Cloud computing customers do not generally own the physical infrastructure serving as host to the software platform in question. Instead, they avoid capital expenditure by renting usage from a third-party provider.They consume resources as a service and pay only for resources that they use.

c) Apache Tomcat Server

Apache Tomcatis a web container developed at the Apache Software Foundation. Tomcat implements the servlet and the JavaServer Pages (JSP) specifications from Sun Microsystems, providing an environment for Java code to run in cooperation with a web server. Tomcat includes its own HTTP server internally, it is also considered a standalone web server. The Tomcat servlet engine is often used in combination with an Apache web server or other web servers. Tomcat can also function as an independent web server.

Tomcat is increasingly used as a standalone web server in high-traffic, high-availability environments.Since its developers wrote Tomcat in Java, it runs on any operating system that has a JAVA Virtual Machine.

X. RESULT AND CONCLUSION

We proposed the overview of community question answer retrieval by using two enhanced model such as basic category and enhanced category model. Whereas user posted question retrieved based multiple domain category and we concentrate all the domain answer especially on medical sites. In our approach, discovery of the latent topic of questions for improvingthe performance of question retrieval and phrase-based machine translationmodels have shown superior performance. Our scheme is able to produce promising performance as compared to the prevailing coding

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 4, April 2018

methods. More importantly, the whole process of our approach is unsupervised and holds potential to handle large-scale data.

XI. ADVANTAGES

- Enables high quality answers from the archive to be retrieved.
- Users can directly obtain answers.
- We reduce the computational complexity bydirectly using character-level representations of question-answer pairs
- This model willgive better performance in datasets with good mixof questions that are lexically and semantically

XII. CONSTRAINTS IN IMPLEMENTATION

A hierarchical structuring of relations may result in more classes and a more complicated structure to implement. Therefore it is advisable to transform the hierarchical relation structure to a simpler structure such as a classical flat one. It is rather straightforward to transform the developed hierarchical model into a bipartite, flat model, consisting of classes on the one hand and flat relations on the other. Flat relations are preferred at the design level for reasons of simplicity and implementation ease. There is no identity or functionality associated with a flat relation. A flat relation corresponds with the relation concept of entity-relationship modeling and many object oriented methods.

XIII. ACKNOWLEDGEMENT

I take this opportunity to convey my deep and sincere thanks to our Head Of the Department **Mr. Dr. R. Janarthanan**. I also extend my deep gratitude to the project coordinator **Mrs.T.Karpagam** and also to my guide **Mrs. S. Muthumariammal** for their valuable help and support in presenting the project. I express my sincere gratitude to all the staffs of Computer Science & Engineering Department and my beloved family members and friends who helped me with their timely suggestions and support.

REFERENCES

[1]. J. Jeon, W. B. Croft, and J. H. Lee, “Finding similar questions in largequestion and answer archives,” in Proceedings of the CIKM, 2005, pp.84–90.
[2]. X. Xue, J. Jeon, and W. B. Croft, “Retrieval models for question andanswer archives,” in Proceedings of the SIGIR, 2008, pp. 475–482.
[3]. L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman, “Knowledgesharing and yahoo answers: Everyone knows something,” in Proceedingsof the WWW, 2008, pp. 665–674.
[4]. H. Duan, Y. Cao, C.-Y. Lin, and Y. Yu, “Searching questions byidentifying question topic and question focus,” in Proceedings of ACL,2008, pp. 156–164.
[5]. J.-T. Lee, S.-B. Kim, Y.-I. Song, and H.-C. Rim, “Bridging lexical gapsbetween queries and questions on large online q&a collections withcompact translation models,” in Proceedings of the EMNLP, 2008, pp.410–418.
[6]. D. Bernhard and I. Gurevych, “Combining lexical semantic resourceswith question & answer archives for translation-based answer finding,”in Proceedings of the ACL, 2009, pp. 728–736.
[7]. X. Cao, G. Cong, B. Cui, and C. S. Jensen, “A generalized frameworkof exploring category information for question retrieval in communityquestion answer archives,” in Proceedings of the WWW, 2010, pp. 201–210.
[8]. G. Zhou, L. Cai, J. Zhao, and K. Liu, “Phrase-based translation modelfor question retrieval in community question answer archives,” in Proceedingsof the ACL, 2011, pp. 653–662.
[9]. A. Singh, “Entity based q&a retrieval,” in Proceedings of the EMNLP,2012, pp. 1266–1277.
[10]. K. Zhang, W. Wu, H. Wu, Z. Li, and M. Zhou, “Question retrieval withhigh quality answers in community question answering,” in Proceedingsof the CIKM, 2014, pp. 371–380.
[11]. S. Robertson, S.Walker, S. Jones, M. Hancock-Beaulieu, and M. Gatford,“Okapi at TREC-3,” in Proceedings of TREC, 1994, pp. 109–126.
[12]. L. Cai, G. Zhou, K. Liu, and J. Zhao, “Learning the latent topics forquestion retrieval in community qa,” in Proceedings of the IJCNLP, 2011,pp. 273–281.
[13]. [1] E. Agichtein, C. Castillo, D. Donato, A. Gionis, andG. Mishne. Finding high-quality content in social media. In*WSDM*, pp. 183–194, 2008.
[14]. A. Berger, R. Caruana, D. Cohn, D. Freitag, and V. Mittal.Bridging the lexical chasm: statistical approaches toanswer-finding. In *SIGIR*, pp. 192–199, 2000.
[15]. J. Bian, Y. Liu, E. Agichtein, and H. Zha. Finding the rightfacts in the crowd: factoid question answering over socialmedia. In *WWW*, pp. 467–476, 2008.