ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol. 2, Issue 10, October 2016

A SURVEY ON QUESTION AND ANSWER SYSTEM BY RETRIEVING THE DESCRIPTIONS USING LOCAL MINING AND GLOBAL LEARNING

SAI KIRAN CHEBIYYAM[1], KANMANI S P[2]

PG Student, COMPUTER SCIENCE AND ENGINEERING, SRM UNIVERSITY KATTANKULATHUR [1]
Assistant Professor, COMPUTER SCIENCE AND ENGINEERING, SRM UNIVERSITY KATTANKULATHUR [2]

***Abstract*— Question answering is a modern type of data recovery described by data needs that are at any rate somewhat communicated as normal dialect articulations or addresses, and is a standout amongst the most regular types of human PC cooperation. This article gives an extensive and relative review of Question Answering Technology (QAT). Question retrieval in current community-based question answering (CQA) administrations does not, all in all, function admirably for long and complex inquiries. This paper introduces the quality question and answer (QA) sets amassed as thorough information bases of human knowledge. It helps clients to look for exact data by acquiring right answers straightforwardly, as opposed to skimming through substantial ranked arrangements of results. Hence to retrieve relevant questions and their corresponding answers becomes an important task for information acquisition. This paper discusses different focus of the QA task which is transformed from answer extraction, answer matching and answer ranking to searching for relevant questions with good ready answers.**

***Index Terms*— answer extraction, community based QA, question answering, question retrieval, matching and ranking.**

I. INTRODUCTION

Question retrieval in community based question answering (CQA) is not quite the same as general Web look [1]. Not at all like the Web internet searchers that arrives at a non-insignificant rundown of positioned archives, question recovery gives back a few applicable inquiries with conceivable answers specifically. While in customary question answering (QA), the principle undertakings are answer extraction [2]-[3], answer matching [4] and answer ranking [5], with CQA, the fundamental errand is to hunt down significant inquiries with great prepared answers [6].

One of the significant difficulties for inquiry recovery is the word bungle amongst inquiries and competitor questions.

For instance,

Question Q1: Why do individuals get colds all the more frequently in lower temperature?

Important Question: Q2: Why are you less inclined to come down with a bug or flu in spring summer and fall than winter months?

In the above inquiries are applicable to each other, yet the same importance is communicated with various word structures, for example, "get colds" and "contract a bug", "lower temperature" and "winter months". These make it non-unimportant for semantic level inquiry coordinating.

To handle the word crisscross issue, past work principally falls back on inquiry extension. Xu et.al [7] investigated neighborhood and worldwide elements to extend single terms in questions. Collins-Thompson et.al, [8] utilized equivalent words, sign words and foundation smoothing to decide inquiry affiliations. Nonetheless, the previous methodology neglects to dole out unequivocal weights to the extended angles and the later approach ignores phrase level proofs for inquiry extension. In the

ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol. 2, Issue 10, October 2016

interim, pseudo relevance feedback [9] and blending [10] are likewise two compelling ways to deal with tackle the word jumble amongst questions and the competitor reports in the term level. Zhou et.al, [11] used the Wikipedia as an outside learning base to improve the execution of inquiry recovery. In spite of their prosperity, writing that considers the idea level extension by abusing various outer information sources is still scanty.

II. SOME DEFINITIONS

Question is characterized as a characteristic dialect sentence, which more often than not begins with an interrogative word and communicates some data need of the client. Now and then an inquiry has a type of a basic build and begins with a verb. In such a case the data solicitation (information request) is called articulation (statement).

Question type is characterized as a specific semantic class of inquiries described by some regular properties. The real issue sorts are: tidbits, list, definition, and theoretical, causal, relationship, procedural, and affirmation questions. A tidbit inquiry is an inquiry, which more often than not begins with a Wh-examined word (What, When, Where, Who) and requires as an answer a reality communicated in the content body. A rundown inquiry is an inquiry, which requires as an answer a rundown of elements or realities. A rundown address generally begins as: List/Name [me] [all/at any rate NUMBER/some]. A definition inquiry is an inquiry, which requires finding the meaning of the term in the inquiry and ordinarily begins with what is. Identified with the last is the illustrative inquiry, which requests definitional data or for the depiction of an occasion, and the supposition address whose center is the feeling around an element or an occasion. A speculative inquiry is an inquiry, which requires data around a theoretical occasion and has the type of what might happen if. A causal inquiry is an inquiry, which requires clarification of an occasion or antique, similar to why. A relationship question gets some information about a connection between two elements. A procedural inquiry is an inquiry, which requires as an answer a rundown of directions for finishing the assignment specified in the inquiry. An affirmation inquiry is an inquiry, which requires a Yes or No as a response to an occasion communicated in the inquiry.

Question answering is a data recovery errand obliged by an outflow of all or a part of the data need as an arrangement of regular dialect inquiries or articulations. Illustrations are: Who is the planner of the Hancock working in Boston? Or what is the stature of the Eiffel Tower? Questions in characteristic dialect, from one perspective indicate an all-around characterized data need, yet then again they convey more data than a basic rundown of pursuit terms, as they speak to syntactic and semantic connections between the inquiry terms.

Data (information) source is characterized as a gathering of data items (records, video, sound, content, documents or databases) accessible to the inquiry noting framework for separating answers.

Today's question answering is not constrained by the sort of report or information archive – it can address both customary databases and more propelled ones that contain content, pictures, sound and video. Organized and unstructured information accumulations can be considered as data sources being referred to noting [12]. Cases of organized information sources incorporate social databases (i), where the put away questions and their qualities have all around characterized semantics and connections; learning bases of master frameworks (ii), where notwithstanding the true information that is run of the mill of information stores, standards are included permitting thinking and critical thinking. Composed content, discourse, pictures, video and sound are all types of unstructured information. Unstructured does not suggest that the information is basically indistinguishable, in which case it would just be garbage, yet rather that its data is encoded in a structure that is troublesome for PCs to translate specifically. Unstructured information permits questioning of crude components (for instance, words in an assemblage of content), removing data with clear semantics connected (organized data), or a blend of both. Identified with this refinement amongst organized and unstructured information there is a customary qualification

between restricted domain question answering, (RDQA), and open domain question answering (ODQA). RDQA frameworks are intended to answer questions postured by clients in a particular space of skill, and as a rule depend on physically developed information or learning sources. They regularly focus on a class of clients who know and utilize the space particular wording in their inquiry detailing, as, in the therapeutic area [13-15]. ODQA concentrates on noting questions paying little respect to the subject area. Removing answers from an expansive corpus of printed reports is an average case of an ODQA framework [16]. As of late, we have seen a methodology of inquiry noting including semi-organized information. These information frequently contain content archives in which the structure of the record or certain removed data is communicated by a markup. Such markups can be credited physically (e.g., the structure of a record) and/or in a programmed way, e.g., markups for distinguished individual and organization names and their connections in daily paper articles.

A retrieval model is characterized by the structure used to speak to the data sources and the data need, and by the recovery capacity, which gauges the importance between the inquiry and the record. In this connection we can balance information recovery with data recovery. Information recovery, for instance, from a social database, expect an information recovery dialect, (for example, SQL) with a very much characterized language structure and semantics, and information that correspondingly show this sentence structure and semantics, permitting a deterministic match between a data need and the information. Data recovery, then again, ordinarily handles questions and archives where the structure and semantics are to an expansive degree vague, and diverse dubious understandings of the data need and record substance are basic. At the point when recovering answers from the report archive, frequently the match amongst inquiry and data is non-deterministic, yielding a positioning of the data as indicated by importance. Being referred to noting innovation, as no basic prerequisites on the information gathering and regular dialect inquiries or explanations are made, a non-deterministic positioning of the answers appears the most suitable recovery model.

III. QUESTION ANSWERING HISTORY-SHORT MANNER

In the writing the initially referred to address noting frameworks are BASEBALL [17], worked in 1961, and LUNAR [18], worked in 1972, both of which cross examined an organized database utilizing characteristic dialect questions. LUNAR gave an interface to information from investigations of rock tests amid the Apollo moon missions. The BASEBALL framework addressed inquiries regarding ball games played over a time of one year. Both frameworks investigated questions in light of an arrangement of regular dialect designs that were required to happen in the info. Since the subject area was limited, a thorough arrangement of examination examples implanted in a space particular vocabulary was assembled physically, so that the inquiries could be prepared and effectively deciphered into an organized inquiry structure expected to investigate the databases. BASEBALL and LUNAR were the principal cases of characteristic dialect interfaces to databases, NLIDB. Today's examination in NLIDB concentrates on non-specific ways to deal with the discovery of articles and their traits and connections in normal dialect questions, and on the interpretation of lexical things into the string tokens that are utilized to portray the database passages. An outline of NLIDB examination can be found in [19]-[20].

In the 1980s and 1990s, information base frameworks, typically worked for limited application areas, turned out to be extremely prevalent. This innovation is appropriate to an inquiry noting structure, in which the client is defied with a specific issue and needs the answer. Access to the learning base is typically sorted out through menus or a characteristic dialect interface. The framework itself intuitively asks the client extra inquiries with a specific end goal to better comprehend the client's aim. To take care of the issue, the framework reasons utilizing the learning accessible as a part of the learning base and the extra data gave by the client. Such deductive inquiry noting concentrates on thinking and gives clarifications how a specific answer was acquired. Its initial roots do a reversal to early master framework innovation, for example, the MYCIN framework [21], which was intended to offer clarifications of

ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol. 2, Issue 10, October 2016

therapeutic ideas. Another case is the SHRDLU framework which as of now in 1972 offered an intuitive exchange interface to control robot communications with toy squares [22]. Today, characteristic dialect inquiries still draw in generous consideration as an approach to inquiry information bases.

The contemporary inquiry noting time began in 1999. The yearly Text Retrieval Conference (TREC) has included open area question replying as one of its opposition tracks [23]. The test is to give a compact response to a characteristic dialect question, given a vast gathering of printed reports. In late assessments two surely understood accumulations of reports were utilized: AQUAINT with more than 1 million records and AQUAINT2 with around 907 K archives and 2.5 GB in size. The scope of inquiry themes is unhindered, and frequently the examination of inquiry and of writings is somewhat shallow, and no propelled thinking methodologies are connected to discover the answers. TREC majorly affected enthusiasm for inquiry noting and on the improvement of assessment measures that think about the execution of distinction frameworks [24]. This brought about the principal Web-based QA framework, START, which was created at the Massachusetts Institute of Technology in 2004 [25]. From 2002 onwards, the Initiative for the assessment of XML recovery (INEX) [26] propelled various assignments concentrating on abusing the inner structure of stamped records or recovering pertinent data.

IV. THE ARCHITECTURE OF A TYPICAL QUESTION ANSWERING SYSTEM

The client composes a question by method for the client question interface. After that this question is utilized to concentrate all the conceivable responses for the information question. The engineering of Question-Answering framework is as appeared in Fig.1. The engineering which is given in Fig.1 works in 5 phases. The capacity of every stage is as per the following [36]:

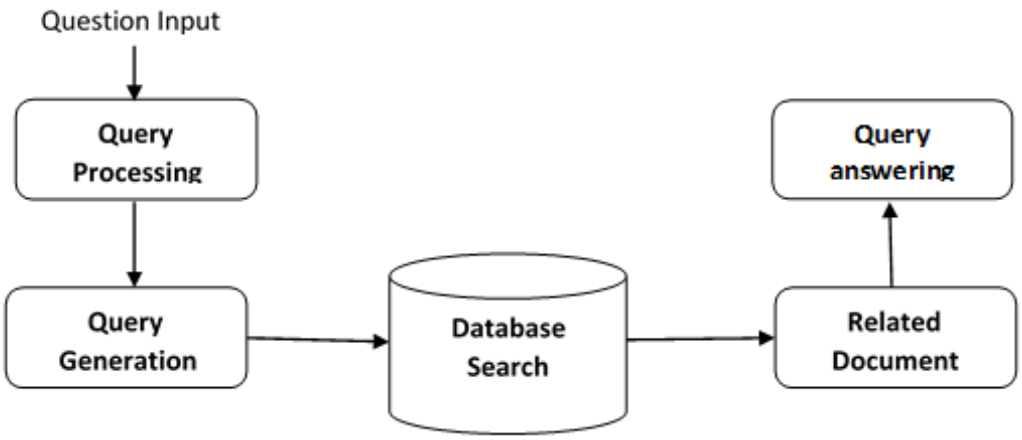

Fig. 1. Question-Answering System's Architecture

Query Processing

Query processing is used to retrieve the information from a database according to a set of retrieval criteria, the database itself remaining unchanged.

Given a natural language question as information, the general capacity of the inquiry preprocessing module is to prepare and break down the information question. This prompts the characterization of inquiry as having a place with any of the sorts bolstered by the framework.

Query Generation

In query generation we will utilize Query Logic Language (QLL) which is utilized to express the info question.

Database Search

Here the hunt of the conceivable results is done in the put away database, the related results that fulfill the given question with chose watchword and guidelines are sent to the following stage.

Related Document

The outcome which was produced by the past stage is put away as a report.

Query Answering

The outcome is put away as a report which is in web design. At that point the outcome is changed over into required content which is required by the client and showed to the client.

V. QUESTION AND ANSWERING TASKS

Answer Extraction

Berger et.al, [27] introduced statistical approaches to bridging the lexical gap in FAQ retrieval. They reviewed an accumulation of addressed inquiries and portray the connection amongst inquiry and answer with a measurable model. Riezler et.al, [28] used a monolingual interpretation based recovery model for answer recovery. They acquainted sentence level summarizing system with catch lexical similitudes amongst inquiries and answers. Duan et.al, [29] first recognized inquiry theme and center by utilizing a tree cut technique. They then proposed another dialect model to catch the connection between inquiry point and center for inquiry recovery. Jeon et.al, [30] thought about four diverse recovery models, i.e., VSM, BM25, LM and interpretation model for inquiry recovery in CQA documents. Trial comes about uncover that the interpretation model beats alternate models.

Xue et.al, [1] joined the dialect model and interpretation model to an interpretation based dialect demonstrate and acquire better execution being referred to recovery. Taking after that, Wang et.al, [31] proposed a syntactic tree coordinating model to finding comparative inquiries, and exhibited that the model is powerful against linguistic blunders. Bernhanrd et.al, [32] utilized the monolingual parallel corpora, which are gathered from the WikiAnswer site, the definitions and shines of the same term in various lexical semantic assets, to prepare the interpretation model for inquiry recovery.

Cao et.al, [6] proposed the class smoothing based and address classification based ways to deal with upgrade the exhibitions of existing inquiry recovery models. Kumaran et.al, [33] addresses the issue of removing the ideal (as far as recovery effectiveness) sub-question from the first long inquiry. Their methodology includes removing a short rundown of applicant sub-questions utilizing the common data measure and displaying this rundown to the client, permitting her to supplant the first inquiry by one of the competitors from the rundown.

Cui et.al, [34] studied a fuzzy relation matching technique for factoid passage retrieval and also built up a framework which removes the developments terms in view of client's conduct which is put away in type of question logs. They kept up a rundown of the considerable number of archives went by for a specific question. Likelihood of archive being gone by when a specific question word is available in an inquiry is ascertained to find the pertinence of the report.

Answer Matching

The syntactic and semantic investigations of the inquiry and of record sentences in a literary database have two purposes. To begin with, they give extra limitations in the coordinating of data need and competitor answer sentences keeping in mind the end goal to conceivably build the accuracy of the subsequent answers. The syntactic structure distinguished in the sentences or the disambiguation of expressions by appointing them semantic marks are cases of such limitations. We likewise arrange expected answer sort finding under this class, as this additionally can possibly make the quest for the answer more exact. Second, the investigation may make an interpretation of the substance into a more broad structure, which may enhance the conceivable outcomes of coordinating and in this manner can enhance the recovery review. Regular dialect investigations in the end lead to an interpretation of the announcements into an organized shape, for example, an organized question or sentences, or into first-arrange rationale representations.

Answer Ranking

Jae-Hyun Park et.al, [35] proposed an inquiry term positioning methodologies. These are utilized to choose powerful terms from a verbose question by positioning terms. Highlights utilized for question term positioning and determination in past work don't consider syntactic connections between terms. To address this issue, we utilize syntactic components removed from reliance parsing consequences of verbose inquiries. We additionally adjust the technique for measuring the viability of question terms for inquiry term positioning.

VI. GLOBAL LEARNING AND LOCAL MINING APPROACHES

Local Mining Approach

Medical concepts are alluded to therapeutic space particular thing expressions and Medical wordings are insinuate as validated expressions by surely understood associations that are utilized to precisely portray the human body and related segments, conditions and procedures in a science-based way.

The local mining comprises of the thing expression mining, therapeutic thoughts location and standardization. The worldwide mining includes entomb master affiliation; bury phrasing relationship, probabilistic hyper diagram creation.

Global Learning Approach

Global learning is a critical technique, including neighborhood approach, and endeavored to delineate QA matches straightforwardly to the passages in outer lexicons with no pruning. This strategy by and large introduces issues since the outer word references actually cover moderately far reaching wordings and are a long ways past the vocabulary size of the given corpus. It might bring about the disintegration in coding execution in states of proficiency and viability. The issue is brought about by the over-turned extent of Vocabularies, which may take in erratic commotions and make the exact wording choice testing. Accordingly, a corpus mindful vocabulary phrasing is normally developed by nearby mining approach, which can be utilized as wording hole for further learning.

Global learning model is manufactured to cooperatively enlarge the nearby coding results. This model impeccably acclimatizes various heterogeneous data signals.

VII. CONCLUSION

This paper displays a medicinal phrasing task plan to connect the vocabulary hole between wellbeing seekers and social insurance learning. The plan contains two segments, nearby mining and worldwide learning. The previous sets up a tri-stage structure to locally code every medicinal record. However, the nearby mining methodology may experience the ill effects of data misfortune and low accuracy, which are

brought about by the nonattendance of key therapeutic ideas and the nearness of the insignificant medicinal ideas. This propels us to propose a worldwide learning way to deal with make up for the inadequacy of nearby coding approach. The second part cooperatively learns and spreads wordings diverse hidden associated restorative records. It empowers the mix of heterogeneous data. Broad assessments on a genuine dataset exhibit that our plan can create promising execution when contrasted with the overarching coding strategies. All the more significantly, the entire procedure of our methodology is unsupervised and holds potential to handle extensive scale information. Later on, we will explore how to adaptably arrange the unstructured restorative substance into client needs-mindful philosophy by utilizing the prescribed therapeutic wordings.

In the future, we plan to generate the concept paraphrases to jointly estimating their probabilities on the multiple linguistic resources. Meanwhile, we will consider to adopt the word or phrase embedding approach to explore the phrasal paraphrases due to its power on measuring words or phrases similarities using the context of monolingual resource.

REFERENCES

[1] Xue, X, Jeon, J, and Croft, W.B. (2008). "Retrieval models for question and answer archives." In Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR'08, pp. 475–482.

[2] Kwok, C, Etzioni, O, and Weld, D. S. (2001). "Scaling question answering to the web." ACM Trans. Inf. Syst. Vol. 19, issue 3, pp. 242–262.

[3] Moldovan, D, Pas¸ca, M, Harabagiu, S, and Surdeanu, M. (2003). "Performance issues and error analysis in an open-domain question answering system." ACM Trans. Inf. Syst. Vol. 21, issue 2, pp. 133–154.

[4] Cui, H, Kan, M.-Y, and Chua, T.-S. (2007). "Soft pattern matching models for definitional question answering." ACM Trans. Inf. Syst. Vol. 25, issue 2.

[5] Ko, J, Si, L, Nyberg, E, and Mitamura, T. (2010). "Probabilistic models for answer-ranking in multilingual question-answering." ACM Trans. Inf. Syst. Vol. 28, issue 3, pp.16:1–16:37.

[6] Cao, X, Cong, G, Cui, B, Jensen, C.S, and Yuan, Q. (2012). "Approaches to exploring category information for question retrieval in community question-answer archives." ACM Trans. Inf. Syst. Vol. 30, issue 2.

[7] Xu, J., and Croft, W. B. (1996). "Query expansion using local and global document analysis". In Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR'96, pp. 4–11.

[8] Collins-Thompson, K, and Callan, J. (2005). "Query expansion using random walk models." In Proceedings of the 14th ACM international conference on Information and knowledge management, CIKM '05, pp. 704–711.

[9] Baeza-Yates, R. A., and Ribeiro-Neto, B. A. (2011). "Modern Information Retrieval - the concepts and technology behind search," Second edition. Pearson Education Ltd., Harlow, England.

[10] Belkin, N.J, Cool, C, Croft, W.B, and Callan, J.P. (1993). "The effect multiple query representations on information retrieval system performance." In ACM SIGIR conference on Research and development in information retrieval, SIGIR'93, pp. 339–346.

[11] Zhou, G, Liu, Y, Liu, F, Zeng, D, and Zhao, J. (2013). "Improving question retrieval in community question answering using world knowledge." In IJCAI, pp. 2239–2245. AAAI Press.

[12] Marie-Francine Moens, (2006). "Information Extraction: Algorithms and Prospects in a Retrieval Context (The Information Retrieval Series 21)," Springer Verlag, New York, Inc., Secaucus, NJ, USA, 2006.

[13] Pierre Zweigenbaum, (2003). "Question answering in biomedicine," in: Proceedings of the EACL2003 Workshop on NLP for Question Answering, Budapest, 2003, pp. 1–4.

[14] Erik Tjong Kim Sang, Gosse Bouma, Maarten de Rijke, (2005). "Developing offline strategies for answering medical questions," in: Proceedings of the AAAI-05 Workshop on Question Answering in Restricted Domains, Pittsburgh, PA, USA, 2005, pp. 41–45.

[15] Diego Mollá, José Luis Vicedo, (2007). "Question answering in restricted domains: An overview," Computational Linguistics Vol. 33, issue 1, pp. 41–61.

[16] Mark T. Maybury (Ed.), (2004). "New Directions in Question Answering," AAAI Press, 2004.

[17] B.F. Green, A.K. Wolf, C. Chomsky, K. Laughery. (1961). "BASEBALL: An automatic question answerer," in: Proceedings of Western Computing Conference, vol. 19, 1961, pp. 219–224.

[18] W.A. Woods, R.A. Kaplan, B. Nash-Webber, (1972). "The lunar sciences natural language information system:" Final report: BBN Report #2378. Technical report, Bolt Beranek and Newman Inc., Cambridge, MA. June 1972.

[19] Ann Copestake, Karen Sparck Jones, (1990). "Natural language interfaces to databases," Knowledge Engineering Review, Vol. 5, pp. 225–249.
[20] Ion Androutsopoulos, (1995). "Natural language interfaces to databases – An introduction," Natural Language Engineering, Vol. 1, pp. 29–81.
[21] Edward H. Shortliffe, (1976). "Computer Based Medical Consultations: MYCIN," American Elsevier, 1976.
[22] Terry Winograd, (1972). "Procedures as a representation for data in a computer program for understanding natural language," Cognitive Psychology, Vol. 3, issue 1, pp. 1–191.
[23] Hoa Trang Dang, Diane Kelly, Jimmy J. Lin. (2007). "Overview of the TREC 2007 question answering track," in: Proceedings of The Sixteenth Text Retrieval Conference (TREC 2007), 2007.
[24] Ellen Voorhees, Donna Harman, (2005). "TREC: Experiment and Evaluation in Information Retrieval," The MIT Press, Cambridge, MA, 2005.
[25] Boris Katz, Sue Felshin, Jimmy J. Lin, Gregory Marton, (2004). "Viewing the Web as a virtual database for question answering," in: Mark T. Maybury (Ed.), New Directions in Question Answering, AAAI Press, 2004, pp. 215–226.
[26] Norbert Fuhr, Jaap Kamps, Mounia Lalmas, Andrew Trotman, (Eds.) (2008). "Focused Access to XML Documents," 6th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2007, Selected Papers, Lecture Notes in Computer Science, Springer, vol. 4862, 2008.
[27] A. L. Berger, R. Caruana, D. Cohn, D. Freitag, and V. O. Mittal, (2000). "Bridging the lexical chasm: statistical approaches to answer finding," in Proc. 23rd Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2000, pp. 192–199.
[28] S. Riezler, A. Vasserman, I. Tsochantaridis, V. O. Mittal, and Y. Liu, (2007). "Statistical machine translation for query expansion in answer retrieval," in Proc. 45th Annual Meeting Association of Computer Linguistics, 2007, pp. 464–471.
[29] H. Duan, Y. Cao, C.-Y. Lin, and Y. Yu, (2008). "Searching questions by identifying question topic and question focus," in Proc. ACM International Conference Information Knowledge. Management, pp. 156–164.
[30] J. Jeon, W. B. Croft, and J. H. Lee, (2005). "Finding similar questions in large question and answer archives," in Proc. 14th ACM International Conference Information Knowledge. Management, pp. 84–90.
[31] K. Wang, Z. Ming, and T.-S. Chua, (2009). "A syntactic tree matching approach to finding similar questions in community-based QA services," in Proc. 32nd Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2009, pp. 187–194.
[32] D. Bernhard and I. Gurevych, (2009). "Combining lexical semantic resources with question & answer archives for translation-based answer finding," in Proc. ACL, 2009, pp. 728–736.
[33] Giridhar Kumaran and James Allan. A case for shorter queries, and helping user create them. In Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics, pages 220–227, 2006.
[34] H. Cui, J.-R. Wen, J.-Y. Nie, and W.-Y. Ma. Probabilistic query expansion using query logs. In Proceedings of the 11th international conference on World Wide Web, pages 325–332. ACM, 2002.
[35] Jae-Hyun Park and W. Bruce Croft, "Query Term Ranking based on Dependency Parsing of Verbose Queries".
[36] Sahu, Shriya, Vasnik, Nandkishor, and Roy, Devshri. (2012). "Proshanttor: A Hindi Question Answering System", International Journal of Computer Science & Information Technology (IJCSIT) Vol 4, No 2, 2012, 149-158.