CQA system based on community structure and QAF algorithm with online forum support

* Amitesh Gadade, Abhijit Kad , Akshay Kumavat, Prof. Vikas Patil

[1] *College name.* Sinhgad Institute of Technology and Science

Abstract

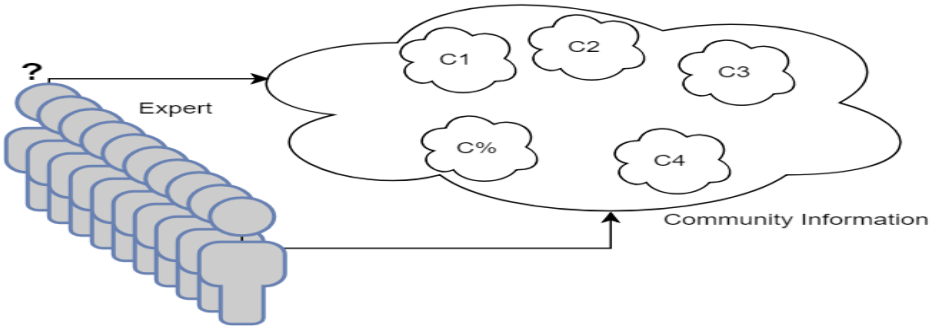
Web based CQA system come in to the focus when user searches question. Any web based user searches on the internet online QA system provides the answers using matching keywords and the matching concept. Because of that sometimes user do not get the proper answers of asked question. For that research on the QA system is going on which work on the social-based Q&A systems that rely on an asker's social friends to provide answers. However, this method cannot find answers for a question which does not belonging to the asker's interests. So, considering this problem, the new system CQA is proposed. This system improves the response latency and answer quality in both the social domain and global domain. It uses neural network based friend ranking method to identify answerer candidates by considering social closeness and Q&A activities. In existing works, we used weak tie assisted social based potential answerer location algorithm and an interest coefficient based uncategorized question forwarding algorithm. In this paper we are also having forum support when user do not get proper answer, and so user takes part in online discussion. Sometime user enters wrong question at time posting, so using the fuzzy dictionary we correct the words which help system to work properly.

Keywords - Distributed systems, Question and answer systems, Social networks, Information search.

Introduction

1. Introduction

CQA system has huge number of users where they have different type of questions. Too many CQA systems on web forums like Quora, Yahoo answer and Stack Overflow are more popularity Forums. In which user can post question and answers freely as per interest. User needs to sign up with interest and level of education where he will be in specific community of his interests. User asks questions in his community and rates the answers with rating so that answerer can act like expert. This expert can be of his community or outside of community and experts helps to find best answers and best answerer. User acknowledges the answerer when he gets best answer by rating that answer and likewise high rated answers act like experts and this system helps to find near to best answer w. r. to response rate and response delay condition. This has been seen on two sides a) a user can freely ask any question and can expect a best variety of answers based on the answers rating. b) It takes efforts to go through the provided answers of varying quality and to make sense of them.



Which have been used more options support to this all problem. The main approach is to propose a system which may help to automate the process of finding best answers of newly posed questions. So going forward to this paper, it introduced rank based QA pair and online forum support technique. The Community Question and Answer (CQA) system have large number of users where they have different types of questions. User asks questions in his community and rates the answers so that answerer can act like expert. User acknowledges the answerer when he gets best answer by rating same answer and likewise high rated answers act like experts and also helps to and near to best answer. It improves response rate and response delay of answer quality in both the social community and global community. Since, the interest coefficient based uncategorized QA forwarding algorithm and weak tie assisted social based potential answerer location algorithm.

.

Problem Definition:

The social community question answering pair work on centralized framework but not on different categories of subject. Search question gives more answers and create lots of confusion of analyzing the answer, to resolve problem CQA system doing separated subject and interest of user. The proposed system reducing response rate and response delay that occur in previous system.

Purpose

There have been many existing CQA systems that heavily depends web forums with many more features, such as question answer post, answer view by, multilingual CQA support, recommend on posted question. These features together with calculate response rate and response delay with more similar and semantic search technique, achieves excellent performance on best answer finding. This system aims to efficiently find best answer with the help of fine granted and QA forwarding technique based on answers rating.

Project Scope

CQA is the system with social media behaviour and having more advantages related community question answer. We are going to form neural network of different community and bind with strong ties and weak ties so we can easily transfer question one community to another community. This system mainly uses those users who have more interest in particular topic and more confusion. The main aim is to calculate rating of answer that given by satisfied user (who agrees with answer) with feigned gained technique. The best answers get rewards with the same community as well as different community.

Overall Study

In this system we are using some techniques that are defined below: Feigned gained technique: it finds the best answers with the help of users rating. Social based potential answers: Find posted question answers in local community Global Based Potential answers: Find same question's answers in different answers. Reward System: Best answers get reward for best answers. In that feigned technique and reward system having high priority to

calculate or gives best answers. When user is not getting best answers, he/she enter that question in forum site In Fig 2 shows the flow of the whole system of CQA, in that main entity belongs with the feigned grained answers selection techniques. This form the best answers related to the previous system

SYSTEM ARCHITECTURE

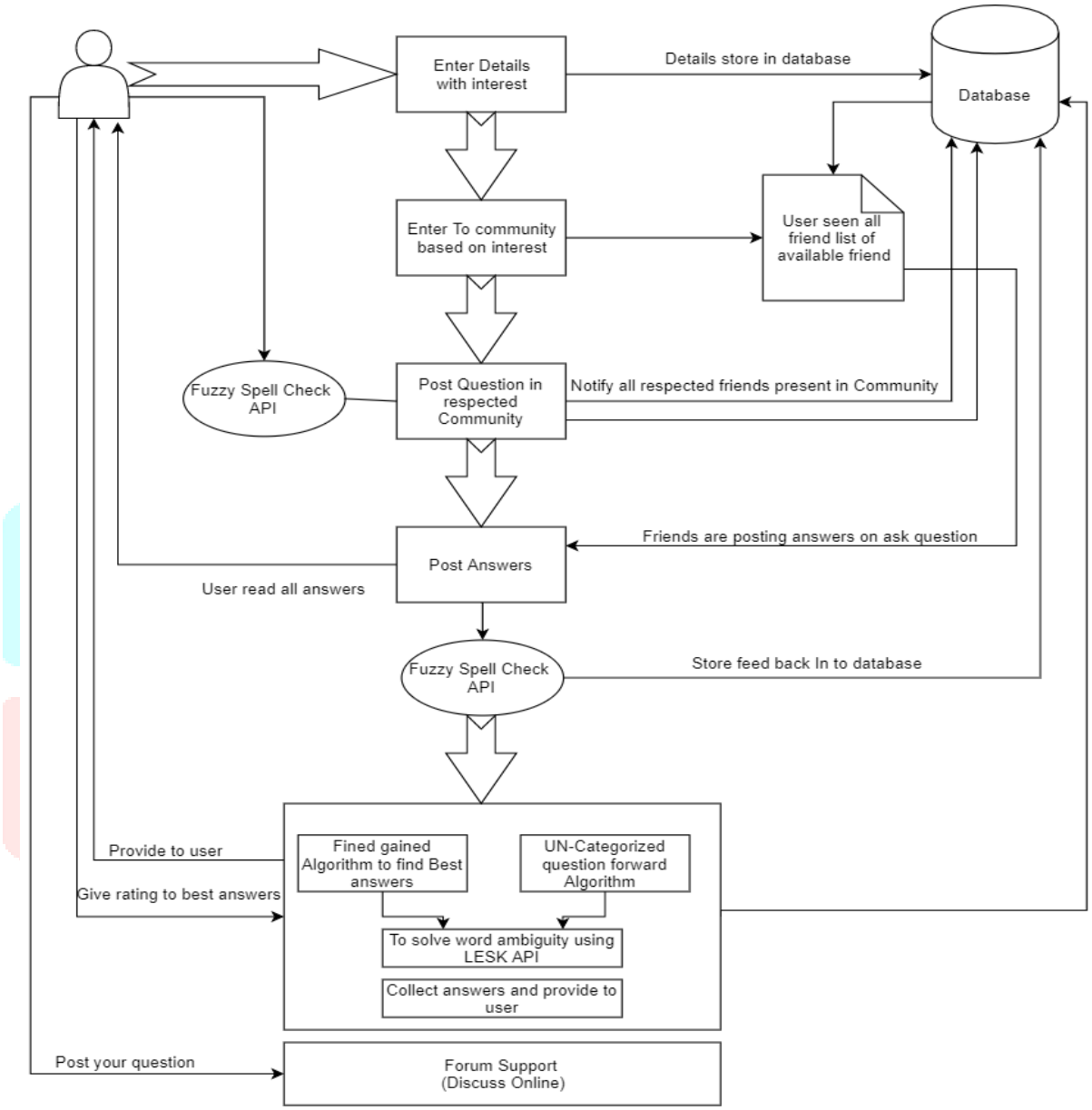


Fig.1. Architecture Diagram

Fig shows the details system of CQA, System will feature aesthetically pleasing and easy-to-use, complete with decision integrity controls to ensure accurate information. The layout of the system must be clear and clean, and utilize fast-loading additional tool like Gmail API and Fuzzy Search. The functional areas of system will be quick loading and agile in response time, to produce result. To ensure system protection, we will provide trusted and secure dataset. Additionally, system must provide clear navigation and transparency. In that user is the main person related to all system, in that he/she registered with the all information with his interest, education and knowledge. This may reflect with the particular community, in that community all related users are present in entered interest, education and knowledge checks with data server. The user of system has authority of post the new question into search box. **Fuzzy Spell: fig 2**There are find all words at distance from the misspelled word i.e. words having one character changed, deleted or removed from the misspelled word. If found, suggest any

one of them (which is not a very good suggestion maybe). Although suggesting one of the several valid words seems very nice approach, one modification could be to use a corpus of text along with the dictionary and for each word in the corpus which is present in the dictionary, find the probability of occurrence of that word in the corpus i.e. Num of time word occurs/Total number of words in corpus. **Forum site: fig 3** A Web forum is a website or section of a website that allows visitors to communicate with each other by posting messages. Most forums allow anonymous visitors to view forum postings, but require you to create an account in order to post messages in the forum. When posting in a forum, you can create new topics or post replies within existing message.

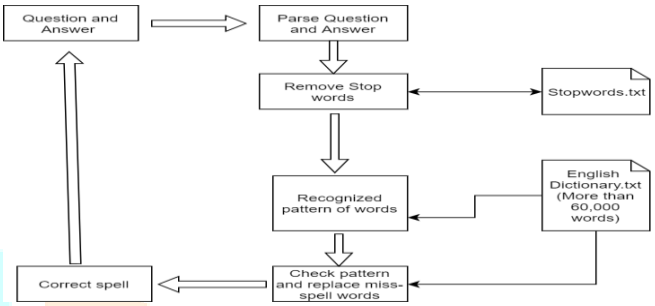


Fig 2

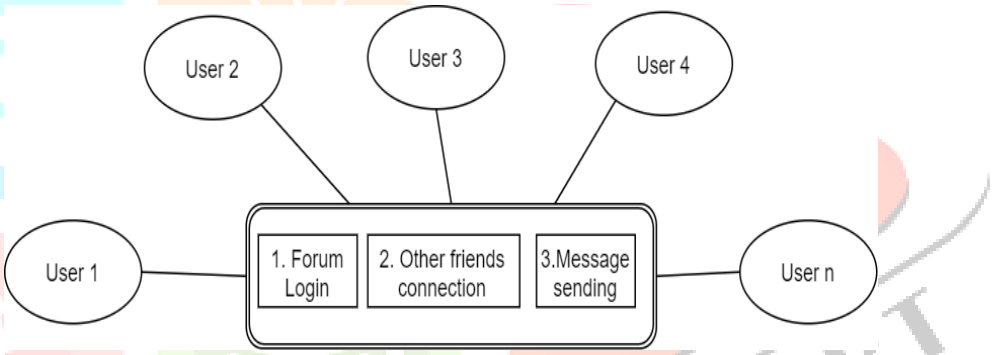


Fig 3

Related Works

Recently, research in social-based QA systems increased rapidly [2], [3], [4], [5], [6], [8], [9], [10], [11]. Most of social-based QA systems use distributed approach to identify credible answerer. Ze Li and Haiying Shen [2] proposed a distributed Social-based mobile QA System (SOS), which enables mobile users to send questions to potential answerer in their friend lists in a decentralized manner. It depends on lightweight knowledge engineering techniques to accurately find out friends who are able to and willing to answer questions, thus reducing the search and computation costs of mobile nodes [2]. Even though recent commercial search engines are mostly based on information related with QA, but it is still difficult to collect an appropriate related content from numerous user-specified answers in QA websites. And in order to get the users which could help people to find relevant answers, GunWoo [3] proposed a ranking algorithm called Influence Rank, which works as the base for analyzing the relationship between user's activities and their mutual understanding. QA system in [4] helps to find best answer based on up votes and down votes using users rating on answers. But, there is very much less knowledge about properties of experts and non-experts and on what basis experts need to be decided in general topics or a specific topic. Social networks have been used for efficient and cooperative file sharing and distribution in peer-to-peer (P2P) Networks [5], [6]. Cheng et al. [5] suggested NetTube, a good associated video forwarding framework that looks for the clustering in social networks for short video sharing. NetTube solved a long queue of key-design issues to understand the system, which included bi-layer overlay, an effective indexing method, and a pre-fetching methodology using social networks. The works in [6] focuses on locating experts and authoritative users as potential answers for QA systems. To recognize reliable users and content in social media, Jiang Bain [6] developed a semi-supervised coupled common boosting framework. This framework concurrently calculates elements quality and user position. This framework needed relatively less

marked exam-ples to start the training process of the system. David et al [8] re-ranked the search results by calculat-ing their relevance with individuals in the requester's social network. S. Bao [9] optimizes a web search by using social annotations from the following two aspects: similarity ranking and Static ranking. Guangyou Zhou et al [10] Question retrieval in CQA can automatically find the most relevant and recent questions that have been solved by other users. However, the word ambiguity and word mismatch problems bring about new challenges for question retrieval in CQA. State-of-the-art approaches address these issues by implicitly expanding the queried questions with additional words or phrases using monolingual translation models. Rui Zhao et al [11] BoW-based vector representation of a document, each element denotes the normalized number of occurrence of a basis term in the document. To count the number of occurrence of a basis term, BoW conducts exact word matching, which can be regarded as a hard mapping from words to the basis term. BoW representation suffers from its intrinsic extreme sparsely, high dimensionality, and inability to capture high-level semantic meanings behind text data. To address the above issues, proposed a new document representation method named Fuzzy Bag-of-Words (FBoW). In addition, propose to use word clusters instead of individual words as basis terms and develop Fuzzy Bag-of-Word Clusters (FBoWC) models. Document representations learned by the proposed FBoW and FBoWC are dense and able to encode high-level semantics. The results on seven real word document classification datasets in comparison with six document representation learning methods have shown that our methods FBoW and FBoWC achieve the highest classification accuracies.

CQA Module List:

Implementation of CQA system is divided into 5 main modules.

(a) Login and Registration with interest - In this module, the system receive information related to user and stored his/her subject of interest.

(b) Post Question and Answer- In this module, the system uses Gmail API to send posted question and answer that are come newly.

(c) Fine grained - In this module, the system uses a rating of answer given by satisfied system user and calculates the average rating of every answer. Based on rating find the best answerers.

(d) QA forwarding – Question in other community are come in different community transfer into correct community using weak and strong ties.

(e) Forum site – Question are not found in system user can post on online forum system.

TOOLS UDED

Software Requirement: A social networking service is an online platform which people use to build social relation with other people who share similar personal or career interests, activities, backgrounds or real-life connections. Social networking services are Internet-based applications.

- Operating System : windows 8 and above..
- Application Server : Tomcat5.0/6.X
- Language : Java
- Front End : HTML, JSP
- Database : MySQL

- **Hardware Requirement:** The hardware design of the system includes designing the hardware units and the interface between those units.

- Processor - Pentium –III
- RAM - 1 GB (min)

Hard Disk - 20 GB

ALGORITHM

DESIGN OF THE STUDY

Algorithm:

CQA Search Algorithm:

1) Input: Question Q=(q1,q2,.....qn)
2) Output: Best Answers
3) Find best answers with feigned grained techniques (no of question and answers are present in community). ***Algorithm 1***
4) The question is not correlated with current community then we are using the question forwarding algorithm. ***Algorithm 2***
5) Best answerer gives rating related to given answers.
6) This answers gives to users A=(a1,a2,....an)
7) Best answerer takes reward to best answers. ***Algorithm 3***

Algorithm 1:

Fine Grained Reputation Algorithm

1. Input: reputation of users
2. Output: Representation of users (expert).
3. Step 1: Root server calculate global reputation of users fan.
4. Step2: Use Buj to calculate % of best answers from the answers.
5. Step 3: Virtual server calculation.
6. Step 4: Reputation of uj
7. Step 5: Find top answers and if not found.
8. Step 6: Question is posted to forum.

Algorithm 2:

Uncategorized question forwarding

1. Input: Uncategorized question by user
2. output: Question is forwarded to community
3. Step 1: find interest which is far away from askers interest
4. step 2: calculate interest coefficient
5. step 3: use asker interest and friend interest

6. step 4: for each interest we calculate coefficient interest
7. step 5: use maximum interest coefficient for asker and answerer
8. step 6: we calculate avg interest similarity
9. Step 7: question is forwarded to similar community.

Algorithm 3

Reward based system

1. Input: Question to friends
2. Output: reward reputed expert
3. Step 1: get the virtual currency
4. Step 2: set threshold currency for best answers
5. step 3: send currency for best answers
6. Step 4: add higher reward with highest reputation with all best answerer currency.

Proposed System Algorithm:

Forum Support

1. User U= {u1, u2, u3.....un); complete the registration(User)
2. getUniqueId() <-0;
3. message<-null , subject<-null , language<-L= (l1, l2,l3....ln);
4. for each(i in U) do
5. if (User register)then

   //generate_unique_userId
6. getUniqueId<-i;
7. else

   //not_register
8. registration(User)
9. End if
10. includeCurrentDiscussionTopic()

    //previous user are talking on particular subject
11. getSubject()
12. getFamiliarLanguage(language l)
13. discussForum()

//send message in the group

14. message<-sent message;
15. return message;

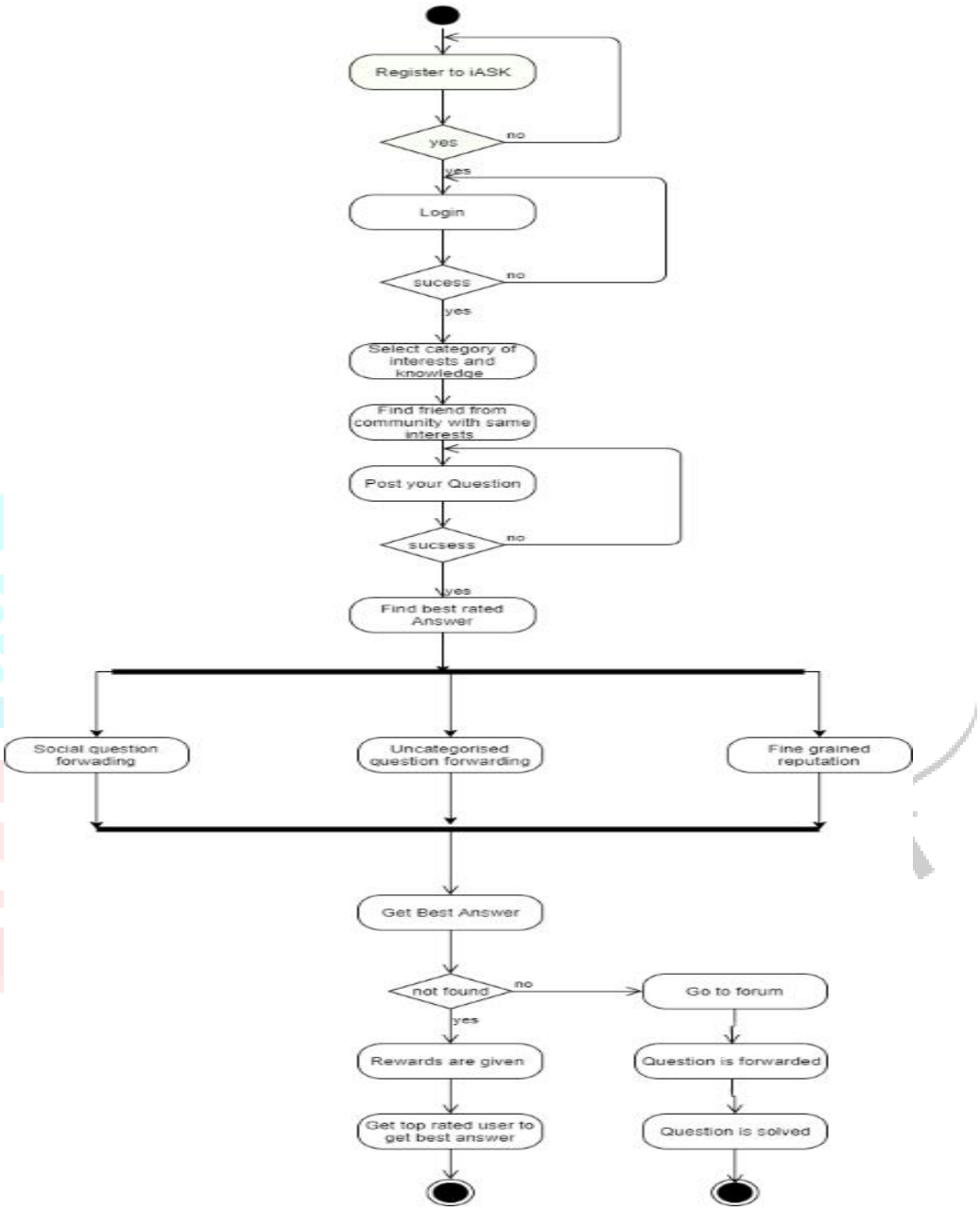


Fig 2: Activity Diagram

OUR APPROACH

To accurately find a friend's current interest and education to be an answerer, for each of user friends, CQA periodically calculates the following social and Q&A activities: response rate, mutual interaction frequency, response delay and precision rate.

Symbol	Description
$_{Ij}$ u_a	different degrees of knowledge in different interests User present in system
f_i R_{fi}	User's all friends list Response rate
Qf_i $ACKf_i$ Mf_i	question sent and forwarded to f_i response from f_i and Qf_i Is denoted by mutual interaction frequency
T	time of prediction
Df_i P^{Ij}_{fi}	Response delay of question sending precision rate of question and anwers
G G^{Ij}_{fi}	upper bound precision score of an answer in the system is the average precision score of all answers

Table 1. Notation Table

A) Response rate:

$$Rfi = \frac{ACKfi}{Qfi}$$

It is measured by the percentage of questions of answered or forwarded by friend's list because forwarding a question is also considered as a responding behavior. This metric reflects the cooperativeness of a friend.

B) **Mutual interaction frequency:**

$$Mfi = \frac{ACKfi}{T}$$

It is measured by the number of interactions between f_i and u_a in a unit time period. This metric reflects the social closeness of the two users.

C)**Response delay:**

$$Df_i = \Sigma_{j\in[1;\ ACKfi]}\ Djfi/ACK_{fi}$$

This metric reflects the responsiveness of interactions and Q&A activities between the two users.

D)**Precision rate:**

$$P^{Ij}_{fi} = G^{Ij}_{fi} / G$$

The response rate and mutual interaction frequency represent the willingness of friends to answer or forward a question. The response delay represents the timeliness of a friend's response. The precision rate reflects the degree that a friend's answer can precisely answer the user's question.

LESK API:

We propose an approach where all the words in the context window are simultaneously disambiguated in a bid to get the best combination of senses for all the words in the window instead of only the target word. We call this the global approach. As opposed to the local approach discussed above, the sense chosen for the target word depends on those chosen for the words around it, and vice versa. In this approach a score is computed for every possible combination of senses, where each such combination contains exactly one sense for each word in the context window, and is distinct from every other combination.

Conclusion:

In this system proposing concept CQA which will use more than one communities to answer the question asked by user. To find best answerer candidates in a user's social network, CQA uses a QA forwarding which consider multiple factors in evaluating the answer QoS of the user's friends. If answers are not obtained in social then answer is forwarded to global community. CQA builds central servers where information is stored to efficiently locate answerer candidates in the interest of the question. CQA has a fine grained reputation system to find experts, and which depends on a reputation-based reward system that adaptively rewards question answerers based on their reputations, in order to provide some value in answering questions we also use weak tie assisted social based potential answerer location algorithm and the interest coefficient based uncategorized question forwarding algorithm to further improve its performance. So, this system helps user to find best answer based on experts based in same community and other too.

Reference:

[1] Ask, http://www.ask.com, [Accessed in May 2015].

[2] Answers, http://www.answers.com, [Accessed in May 2015].

[3] Yahoo! Answers, http://answers.yahoo.com, [Accessed in May 2015].

[4] stack overflow, http://stackoverflow.com/, [Accessed in May 2015].

[5] Quora, http://www.quora.com, [Accessed in May 2015].

[6] J. Jeon, W. B. Croft, and J. H. Lee, "Finding similar questions in large question and answer archives," in CIKM, 2005, pp. 84–90.

[7] M. R. Morris, J. Teevan, and K. Panovich. What Do People Ask Their Social Networks, and Why? A Survey Study of Status Message Q&A Behavior. In Proc. of CHI, 2010.

[8] X. Cheng and J. Liu. NetTube: Exploring Social Networks for Peerto

Peer Short Video Sharing. In Proc. of INFOCOM, 2009.

[9] F. Harper, D. Raban, S. Rafaeli, and J. Konstan. Predictors of Answer Quality in Online Q&A Sites. In Proc. of SIGCHI, 2008.

[10] R. W. White, M. Richardson, and Y. Liu. Effects of Community Size and Contact Rate in Synchronous Social Q&A. In Proc. Of CHI, 2010.

[11] L. Cai, G. Zhou, K. Liu, and J. Zhao, "Learning the latent topics for question retrieval in community qa," in IJCNLP, 2011, pp. 273–281.

[13] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman, "Knowledge sharing and yahoo answers: Everyone knows something," in WWW, 2008, pp. 665–674.

[14] P. GunWoo, Y. SoungWoung, L. SooJin, and L. SangHoon. Credible user identification using social network analysis in a q&a site. 2011.

[15] Z. Li, H. Shen, G. Liu, and J. Li. SOS: A Distributed Mobile Q&A System Based on Social Networks. In Proc. of ICDCS, 2012.

[16] J. Bian, Y. Liu, D. Zhou, E. Agichtein, and H. Zha. Learning to Recognize Reliable Users and Content in Social Media with Coupled Mutual Reinforcement. In Proc. of WWW, 2009.

[17] W. Zhang, Z. Ming, Y. Zhang, T. Liu, and T. Chua, "Exploring key concept paraphrasing based on pivot language translation for question retrieval," in AAAI, 2015, pp. 410–416.

[18] H. Zhang, T. N. Dinh, and M. T. Thai. Maximizing the Spread of Positive Influence in Online Social Network. In Proc. of ICDCS, 2013.

[19] D. Bernhard and I. Gurevych, "Combining lexical semantic resources with question & answer archives for translation-based answer finding," in ACL, 2009, pp. 728–736.

[20] Z. Li and H. Shen. Collective Intelligence in the Online Social Network of Yahoo! Answers and Its Implications. In Proc. of CIKM, 2012.

[21] L. Zhang, X. Li, Y. Liu, Q. Huang, and S. Tang. Mechanism Design for Finding Experts Using Locally Constructed Social Referral Web. In Proc. of INFOCOM, 2012.

[22] E. Bakshy, I. Rosenn, C. Marlow, and L. A. Adamic. The Role of Social Networks in Information Diffusion. CoRR, 2012.