

## Question and Answer System Survey Based on Social media

<sup>1</sup>CHAMAKURA VINAYAK SAI, <sup>2</sup>Dr.V.SATHIYASUNTHARAM

<sup>1</sup>M.Tech, <sup>2</sup>Professor

Department of cse

CMR Engineering College, Kandlakoya (v), Medchal (D), Hyderabad, Telangana,

### ABSTRACT:

Question and Answer (Q&A) systems play a vital role in our daily life for information and knowledge sharing. Users post questions and pick questions to answer in the system. Due to the rapidly growing user population and the number of questions, it is unlikely for a user to stumble upon a question by chance that (s)he can answer. Also, altruism does not encourage all users to provide answers, not to mention high quality answers with a short answer wait time. The primary objective of this paper is to improve the performance of Q&A systems by actively forwarding questions to users who are capable and willing to answer the questions. To this end, we have designed and implemented Social Q&A, an online social network based Q&A system. Social Q&A leverages the social network properties of common-interest and mutual-trust friend relationship to identify an asker through friendship who are most likely to answer the question, and enhance the user security. We also improve Social Q&A with security and efficiency enhancements by protecting user privacy and identifies, and retrieving answers automatically for recurrent questions. We describe the architecture and algorithms, and conducted comprehensive large-scale simulation to evaluate Social Q&A in comparison with other methods. Our results suggest that social networks can be leveraged to improve the answer quality and asker's waiting time. We also implemented a real prototype of Social Q&A, and

analyze the Q&A behavior of real users and questions from a small-scale real-world Social Q&A system.

### I. INTRODUCTION

Traditional search engines like Google, Bing, and Yahoo provide information for factual queries. These do not provide any answers for nonfactual queries which are specific, imaginary, subjective and multi-dimensional. So in order to enhance the performance of search engines, social search engines are proposed. The social search engines gather and group the people of similar interest and initiate the search query to the relevant person in the group. Although the search engines answer factual queries that are already stored in centralized server hence this technique is not suitable for answering non factual queries. For example, "can anyone recommend me a Doctorate professor for doing my project in social network...?" If the valid information is not found in database then we need to forward these queries to the human, which are the most "intelligent machines". The persons who are expertise in that particular topic can give perfect answer.

The advent of Social Web is clearly a game-changer, on numerous fronts. The rush of social web to implement and focus on marketing specially Versus the business is more universal, "social media marketing" is just the latest technique in marketing, but the truth is it is a new way that can be used to market and promote your services and products. It also concentrates on interaction with colleagues, latent

customers as well as current clients and helps to spread message in a serene and conversational way. It also results in many costly disturbances to users by sending queries that cannot be answered and hence increasing the workload by looking on the queries through the pool of queries. It leads to high server bandwidth and high query congestion and maintenance costs. To enhance the askers satisfaction, researchers focused on systems in which the users post and answer in social network. SOS (Social based Q&A system) is such system in which the answerers in the network are connected socially. SOS also leverages a lightweight technique to transform the users closeness and their interests. So that the question is forwarded to the answerer who are close and have similar interests. By which, the node overload is reduced and a quick response is received within low cost. As the answerers have similar interest, the accuracy of answer is also increased. SOS transforms the question into ID's, so that the node can easily compare it with the social ID. Then the node forwards the question to the particular person with that social ID. After receiving the questions the answerer can answer if he knows the answer or if he doesn't, he can forward the question to his friends. As the answerers are socially related to askers, the willingness to answer the question is more compared to the strangers. By choosing the potential answerer from the friends list, the question is finally answered. SOS uses NLP and FOL techniques to calculate questions ID. NLP technique will divide the question into groups of related words (e.g., wh-type). First order logic will parse the question into tokens. In SOS, the question is forwarded to friends in the friend list. The question can be forwarded to TTL number of hops, so that if the askers' friends don't know the answer they can forward to their friends hop. By this the accuracy of getting an answer is improved.

## II. Literature survey:

### **SOS: A Distributed Mobile Q&A SystemBased on Social Networks**

Some ideas are taken from SOS implemented in Subjective Q&A system:

SOS incorporates an online social network, where nodes connect each other by their social links. Here, a registration server is responsible for node registration. Each user has an interest ID, which represents his/her interest. The closeness of two user's interest IDs means the similarity between the two users' interests. Users sharing more common interests with an asker are more likely to be able to answer the asker's questions. Also, users having shorter social distances with an asker are more likely to be willing to answer the asker's questions. SOS has a metric similarity ( $S$ ) that measures the likelihood of a node to be able and willing to answer another node's question. It is determined by the interest similarity between the question's interest and the receiver's interest as well as the social closeness between the question receiver and sender. SOS defines a constant  $K$ , which is the largest number of friends that a node can send/forward a question in its friend list. SOS allows each node to define TTL, which is the maximal number of hops that a question can be forwarded. A node determines TTL depending on how urgent the question is. Figure shows the question routing process in SOS. After asker A initiates a question, it forwards the question to the top  $K$  friends (nodes B and C) who have the highest  $S$  in its friend list with the question.

## III. Social Q&A

SQA sites exemplify the Web 2.0 model of user-generated and user-rated content. Anyone can ask and answer questions, rate content submitted by others, and view the community's aggregate assessment of

which questions, answers, and users are best. Sites featuring answers by experts are not strictly social Q&A, though many SQA sites offer a function to identify contributors as peer experts, given users' assessments of their past contributions in particular topic areas. In their proposal for a social Q&A research agenda, Shah et al. (2009) propose a working definition of social Q&A as a site or service requiring:

A method for users to present an information need in the form of a natural language question (as opposed to a keyword query)

- A forum for public response
- A community, based on participation level, in which the above transactions are embedded

They identify three primary research areas in social Q&A: user motivation and behavior, information quality assessment, and design and technological factors impacting participation.

In 2002, South Korean company NHN launched what is generally credited as the first SQA site, Knowledge-iN, as a component of its popular Naver search engine. The first SQA site in the United States was Answerbag, launched in April 2003, but when Yahoo! Answers launched in December 2005 (after a 6-month beta test), with its installed base of information-hungry searchers, SQA became popularized and institutionalized (Table 1). From its inception, Yahoo! Answers was by far the most widely used SQA site, with a reported peak of 62 million unique visitors per month in the United States alone in 2010. Only recently have competing SQA sites begun to challenge its popularity; after Answers.com merged several other properties into its overall database, its traffic approached Yahoo! Answers' level at approximately

50 million unique visitors per month in the United States.

The predominance of Yahoo! Answers as a data source in the literature is primarily because of the combination of its dominant market share and the ready availability of a subset of its data through a public application programming interface (API) (<http://developer.yahoo.com/answers/>). Other SQA sites are less forthcoming with their data, and researchers are limited to the data and tools available through the sites' public interfaces

However, with the amount of data sharing and crossover traffic between sites, unique visitors per month is a less and less meaningful metric. SQA sites have created widgets or Facebook apps allowing users to access relevant topic-focused content from other sites, and there are differences in the rules and mechanics of site interaction that also have an effect on traffic. For example, Yahoo! Answers limits the length of time a question is open, from as little as four hours to a maximum of 8 days, and restricts the number of questions, answers or comments less experienced members can submit. Some systems allow askers to declare an answer the best, while others aggregate ratings and responses from other users indefinitely, and present the highest-rated answers first.

#### IV. IMPLEMENTATION

##### Base 64 Algorithm used for data encoding

- Base 64 is an encoding scheme that converts binary data into text format so that encoded textual data can be easily transported over network un-corrupted and without any data loss. Base64 is used commonly in a number of

applications including email via MIME, and storing complex data in XML.

**Algorithm:**

Lets take an example. We have to encode string “MENON” into base64 format. Lets call “MENON” as **input\_str**, base64 character set (“ABC..+”) as **char\_set** and resultant encoded string as **res\_str**.

1. Take 3 characters from **input\_str** i.e “MEN” since each character size is 8 bits we will have  $(8 * 3) 24$  bits with us.
2. Group them in a block of 6 bits each  $(24 / 6 = 4$  blocks). (why 6?) because  $2^6 = 64$  characters, with 6 bits we can represent each character in **char\_set**.
3. Convert each block of 6 bits to its corresponding decimal value. Decimal value obtained is the index of resultant encoded character in **char\_set**.
4. So for each 3 characters from **input\_str** we will receive 4 characters in **res\_str**.
5. What if we have less than 3 characters in **input\_str** left i.e “ON”. We have 16 bits and blocks will be  $16 / 6 = 2$  blocks. Rightmost 4 bits will not make a proper block (1 block = 6 bits) so we append zeros to right side of block to make it a proper block i.e 2 zeros will be appended to right. Now we have 3 proper blocks, find corresponding decimal value of each block to get index.
6. Since There were less than 3 characters (“ON”) in **input\_str** we will append “=” in **res\_str**. e.g “ON” here  $3 - 2 = 1$  padding of “=” in **res\_str**.

**Database Schema**

For storing the details of registered users a database with the name users is created. It contains a table named “users” to store all the form details. The default values of some fields is put null. The schema is as follows:

**Schema of users table**

The above table contains six fields to store the uid, name, pwd, email, city and profilepic. Once the user enters the details into the registration page the details are transferred to the above table. Whenever the registered user enters the email and password in the login page the field entries are compared with the corresponding table entries. If there is no match a message is displayed to the user indicating login failure and asking the user to enter the valid details.

In a similar way other tables like calc, frequest, friends, interaction, query, synset and weights are created in database to store processed data from the system.

**Administrator Module**

**Maintaining Category Synset**

After admin is logged into the system under Category Synset tab he can add he can maintain Category Synset. This has category, sub category and all the keywords which are synonyms to the category. While processing a question, Q&A system uses WordNet to examine tags and text of question and generates token string. These tokens are compared to words in Category Synset to determine categories where questions belongs to.

**Question-User Mapping:**

This module identifies appropriate answerers for a given question. 2 main parameters are considered here: 1. the interest similarity between interest vectors of friend and question (represents capability of friend to answer the

question) and 2. Social closeness between friend and asker (represents willingness of friend of answer question). Questions forwarded on this basis are available under Queries tab when that particular user logs in

### **User Module**

#### **Registration**

As soon as the user clicks on the “signup” button on the main page a signup page is displayed. If the user is not a registered use, he has to register by clicking the link. The user has to enter first name, last name, user name, password address and upload users profile pic. Once the user enters the details into the registration page, the details are transferred to the database table.

#### **Login**

user has to login in order to access the system. He has to enter the email and password which were used during registration phase. These credentials are checked with the database entries, if there is a mismatch a message is displayed to the user indicating login failure and is prompted to enter valid credentials. Upon successful login the user can enter queries and find out the most promising location to the user provided clues.

#### **User Interest Analyser**

After user completes process of registration Subjective Q&A system collects the interests of user and stores them into database in encoded formats. This encoding is done using Base64 encoding algorithm. These interests collected are used for admin module for question forwarding procedure

### **Building Social Network**

Subjective Q&A system provides users with facilities of send friend requests for other users of the system by searching for them. Users can also accept or decline the friend request they got from other users. Using these 2 features any registered user can build a social network with altruist friends.

### **Viewing queries and answering them**

Admin module is responsible for question-user mapping, so based on user interests and answering frequency admin forwards questions to the users. User can post a question into the system under Post Query tab. Users can answer these forwarded questions which will be available under Queries tab. Users can also view questions answered which are in users interest zone under Your Queries tab.

## **V. SYSTEM ARCHITECTURE:**

### **Existing System**

Q&A systems on social search engines provide a way to get information from the people who are socially connected. But they are not providing satisfactory results as the friends reply with unnecessary answers to the questions which lead to high server bandwidth. Traditional search engines like Google, Bing and Yahoo provide information for factual queries. These do not provide any answers for nonfactual queries which are specific, imaginary, subjective and multi-dimensional. Current Q&A systems may not meet the requirement of providing high quality answer with a short answer wait time, though users wish to receive satisfactory answers quickly. It found that for Yahoo! Answers, only 17.6% of questions were answered satisfactorily; for the

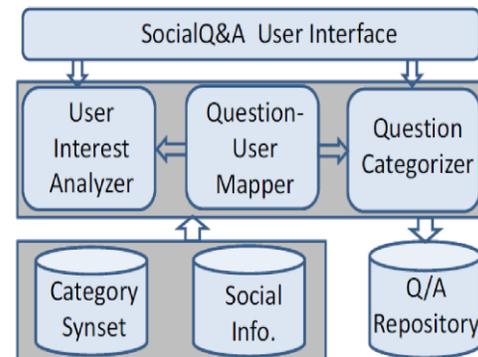
remaining 82.4%, one fifth of the questions remained unanswered.

### Proposed system

Subjective Question and Answer system, is an online social network based Q&A system, that actively forwards questions to those users with the highest likelihood of answering them with expertise and interest in the questions subjects. The design of this system is based on two social network properties. First, social friends tend to share similar interests. Second, social friends tend to be trustworthy and altruistic. Subjective Question and Answer system is composed of three components: User Interest Analyzer, Question Categorizer, and Question-User Mapper. User Interest Analyzer associates each user with a vector of interest categories. Question Categorizer associates a vector of interest categories to each question. Then, based on user interest and social closeness, Question-User Mapper identifies potential answerers for each question.

When a question is posted by the user, the node processes the question using NLP (Natural Language Processing) and then represent the in first order logic format by dividing question into tokens keywords then apply inference rules on the tokens to infer the questions interest. Finally a question id is created based on the interest. This question id is compared with the friends Interest id and if the id's match, question is forwarded to those friends. When a question is posted by the user, the node processes the question using NLP (Natural Language Processing) and then represent the in first order logic format by dividing question into tokens keywords then apply inference rules on the tokens to infer the questions interest. Finally a question id is created based on the interest. This question id is compared with the friends Interest id and if the id's match, question is forwarded

to those friends. After finding the friends with similar interest, it will select k number of best answerers among those friends. It will select the k answerers based on the feedback or performance of the users. If they have a good response feedback then they are selected among k friends



### VI. CONCLUSION

In this paper, we present how these Q&A system can accurately identify the best answerers who are expertise in that area. SOS leverages lightweight techniques to identify the related friends. This system also provides answers to non-factual queries which are specific and multi-dimensional. It uses FOL for the retrieval of interest of the user and the question. It also earns high user satisfaction ratings for the accurate answers. SOS generates very less overhead with limited question forwarding. Since each user is connected to several social groups, it selects most probable answerers and forward to an answerer that can provide an answer We can also get response from any location by accessing through internet by means of cloud server which provides storage of large resources. All the question and answers are stored in the cloud. The future of Q&A system in cloud is demandable and scope full

## REFERENCES

- [1] Haiying Shen, Ze Li, Guoxin Liu, and Jin “SOS: A Distributed Mobile Q&A System Based on Social Networks”, IEEE Transactions on Parallel and Distributed Systems Vol. 25, No. 4, April 014.
- [2] D. Horowitz and S.D. Kamvar, “The Anatomy of a LargeScale Social Search Engine,” Proc. 19th Int’l Conf. World Wide Web(WWW ’10), 2010
- [3] Nikhil Vitlani “SOCIALQ&A: A NOVEL APPROACH TO NOTIFYING THE CORRECT USERS IN QUESTION AND ANSWERING SYSTEMS”, (2012). All Theses. Paper 1423.
- [4] A. Mtibaa, M. May, C. Diot, and M. Ammar, “Peoplerank: Social Opportunistic Forwarding,” Proc. IEEE INFOCOM, 2010.
- [5] Z. Li, H. Shen, and K. Sapra, “Leveraging Social Networks to Combat Collusion in Reputation Systems for Peer-toPeer Networks,” Proc. IEEE Int’l Parallel and Distributed Processing Symp. (IPDPS ’11), 2011.
- [6] K. Toutanova and C.D. Manning, “Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger,” Proc. Joint SIGDAT Conf. Empirical Methods in Natural Language Processing and Very Large Corpora (SIGDAT ’00), 2000.
- [7] H. Shen and K. Hwang, “Locality-Preserving Clustering and Discovery of Resources in Wide-Area Distributed Computational Grids,” IEEE Trans. Computers, vol. 61, no. 4, pp. 458-473, Apr. 2012.
- [8] H. Shen and C. Xu, “Locality-Aware and Churn-Resilient Load Balancing Algorithms in Structured Peer-to-Peer Networks,” IEEE Transactions on Parallel and Distributed Systems, Vol.18, No. 6, June 2007.