

ARTICLE

SOCIAL Q&A: AN ONLINE/OFFLINE SOCIAL NETWORK BASED QUESTION AND ANSWER SYSTEM

Prateeksha Chaurasia*, Shashank Joshi

Dept. of Computer Engineering, Bharti Vidyapeeth (Deemed to be University), College of Engineering, Pune, INDIA

ABSTRACT

Background: Question and Answer (Q&A) systems play a significant role in our standard of living for information and knowledge sharing. Users post questions and decide questions to answer within the system. As a result of the apace growing user population and therefore the variety of queries, it's unlikely for a user to come upon a question by chance that he will answer. Altruism doesn't encourage all users to provide answers, not to mention top quality answers with a brief answer wait time. **Methods:** The first objective of this paper is to enhance the performance of Q&A systems by actively forwarding questions to users who are capable and willing to answer the questions. To the present end, we've got designed and enforced Social Q&A, a web social network primarily based Q&A system. Social Q&A leverages the social network properties of common-interest and mutual-trust friend relationship to spot an verbalize through friendly relationship that are possibly to answer the question, and enhance the user security we tend to conjointly improve Social Q&A with security and efficiency by protecting user's privacy and also retrieving answers mechanically for recurrent queries. We tend to describe the design and algorithms, and conducted comprehensive large-scale simulation to gauge Social Q&A as compared with alternative strategies. **Results:** Our results recommend that social networks are often leveraged to enhance the solution quality and asker's waiting time. We tend to conjointly enforced a true 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. **Conclusions:** It removes the burden from answer providers by directly delivering them the questions they might be interested in, as opposed to requiring answer providers to search through a large collection of questions as in Yahoo! Answers or flooding a question to all of an asker's friends in an online social network.

INTRODUCTION

What is social media?

A social network platform could be a net platform that peoples are victimization these days on awfully high demand for making social networks that can even use a similar personal or Career familiarized interest, backgrounds and real time connections [1]. All social media services that is available on-line.

Loopholes in social media (Insecurity of Social Media)

The rapid growth and ubiquity of online social media services has given an impact to the way people interact with each other. Online social networking has become one of the most popular activities on the web. Social network analysis has been a key technique in modern sociology, geography, economics, and information science [2]. The data generated by social media services often referred to as the social network data. In many situations, the data needs to be published and shared with others. Social networks are online applications that allow their users to connect by means of various link types. As part of their professional network; because of users specify details which are related to their professional life. These sites gather extensive personal information, social network application providers have a rare opportunity direct use of this information could be useful to advertisers for direct marketing. Publish data for others to analyze, even though it may create severe privacy threats, or they can withhold data because of privacy concerns, even though that makes the analysis impossible [3]. .A privacy breach occurs when sensitive information about the user, the information that an individual wants to keep from public, is disclosed to an adversary. For examples, business companies are analyzing the social connections in social network data to uncover customer relationship that can benefit their services and product sales. The analysis result of social network data is believed to potentially provide an alternative view of real-world phenomena due to the strong connection between the actors behind the network data and real world entities. Social-network data makes commerce much more profitable. On the other hand, the request to use the data can also come from third party applications embedded in the social media application itself. For instance, Facebook has thousands of third -party applications and the number is growing exponentially [4]. Even though the process of data sharing in this case is implicit, the data is indeed passed over from the data owner (service provider) to different party (the application) The data given to these applications is usual not sanitized to protect users' privacy. Desired use of data and individual privacy presents an opportunity for privacy-preserving social network data mining. That is, the discovery of information and relationships from social network data without violating privacy.

Question soliciting is an imperative part from proficient learning. In any case, teachers are frequently overpowered with understudies' inquiries and in this way unfit to give opportune answers [5]. Data looking for is additionally rendered troublesome by the sheer measure of learning material accessible, particularly on the web. The utilization of cutting edge data recovery and common dialect handling procedures to answer students' inquiries and diminish the trouble of data looking is from this time forward especially encouraging. Question Answering (QA) frameworks appear to be appropriate for this assignment since they

KEY WORDS

Question and answer systems, Social networks, Information search

Received: 14 May 2018
Accepted: 9 June 2018
Published: 11 June 2018

*Corresponding Author

Prateekshachaurasia@gmail.com
Tel.: +918530089630

go for creating exact responses to normal dialect inquiries rather than just returning reports containing answers.

Privacy concerns in social networks can be mainly categorized into two types: inherent-data privacy and latent data privacy. Inherent-data privacy is related to sensitive data contained in the data profile submitted by users in order to receive data-related services.

Communication strategy on social media

While a great amount of literature has focused on the relationship between communication strategies and corporate reputation, there is no systematic research on the different kinds of social media communication strategies. Based on the corporate reputation and social media literature, this paper aims to contribute to this gap in the research in two main ways. First identifying which social media communication strategy is more effective with contrasting levels of reputations; second, analyzing the differences between high- and low-reputation companies with respect to their ability to use corporate communication [6].

Findings: Social media and communication between those media: egocentric, informal, selective, openness, secretive and supportive. The results additionally reveal distinct ways that within which high-, medium- and low-reputation companies' utilize the six complementary methods of communications [7].

Research limitations/implications: The study is predicated on one trade and on one single geographical market, and care should therefore be taken in generalizing the findings to alternative contexts. So emerges the chance to broaden this analysis to alternative similar service sector, like on-line examination, auditing to assess and generalize the results obtained. Additionally, this technique user is posting solely questions on the social media so this may prove the limitation for projected system. System should permit user to post the comments on social media so experts can comments on the precise post and additionally if stop words are there then system can take away that stop words.

Practical implications: From this study, users will add queries as many as they require so system can recommends the knowledgeable to users and consultants can answer to their queries so user can get precise or correct declare their queries.

Originality/value: This analysis extends between existing system and the proposed system will be prove that this proposed system will prove beneficial when we consider the concepts of security and privacy of user.

Social networking sites (SNS) supplement the system of connections display in the disconnected world by giving stage to dynamic correspondence amongst companions and more uninvolved perception through totaled surges of social news. Utilization of these locales has been related with more noteworthy levels of social capital, or advantages made conceivable by the presence of a social structure [8]. These advantages incorporate crossing over social capital, or access to new data through a different arrangement of associates, and holding social capital, or passionate help from dear companions [9].

Early investigations of the Internet analyzed the relationship between's chance online with results, for example, dejection[10], yet later examines separate between social exercises and unadulterated excitement, finding distinctive outcomes for various exercises

Literature survey on previous work

1. A Comparison of Information Seeking Using Search Engines and Social Networks. [1]

Author: M. R. Morris, J. Teevan, and K. Panovich [1] Published in: 2010

In this paper, they present a study in which 12 participants posted a question to Facebook while simultaneously investigating the same question via Web search. We compare the information participants found with these two methods and participants' satisfaction with each experience. They conclude by discussing the implications of our findings for the design of next-generation search tools.

Advantages

- 1) Accuracy is obtained.
- 2) NLP is used.
- 3) User satisfied with the answer.

Disadvantages:

- 1) Database should maintain.
- 2) User doesn't know the time to get answer.

2. What do People Ask Their Social Networks, and Why? A Survey Study of Status Message Q&A Behavior [2]

Author: M. R. Morris, J. Teevan, and K. Panovich Published in: 2010

In this paper we explore the phenomenon of using social network status messages to ask questions. We conducted a survey of 624 people, asking them to share the questions they have asked and answered of their online social networks. We present detailed data on the frequency of this type of question asking, the types of questions asked, and respondents' motivations for asking their social networks rather than using

more traditional search tools like Web search engines. We report on the perceived speed and quality of the answers received, as well as what motivates people to respond to questions seen in their friends' status messages.

Advantages

- 1) Trust friend maintain.
- 2) If answer not given by the system, friend and family who is trust worthy will give answer.

Disadvantages:

- 1) Time consuming.
- 2) Not efficiency given.
- 3) Problem occurred if trust worthy don't know the answer.

3. Questioning Yahoo! Answers [3].

Author: Z. Gyongyi, G. Koutrika, J. Pedersen, and H. Garcia-Molina Published in: 2008

In this paper, it seeks to understand YA's knowledge sharing activity. They analyze the forum categories and cluster them according to content characteristics and patterns of interaction among the users.

Advantages:

- 1) Less time required
- 2) Categories forms.
- 3) Assign question category wise

Disadvantages:

- 1) Accuracy not defined.
- 2) Not all questions answer is obtained.

4. Routing Questions to Appropriate Answerers in Community Question Answering Services [5].

Author: Li and I. King Published in: 2006

The paper aims to route questions to the right answerers who have a top rank in accordance of their previous answering performance. In order to rank the answerers, we propose a framework called *Question Routing (QR)*.

Advantages:

- 1) QR technique used.
- 2) All questions can get answer.
- 3) Ranking is used.

Disadvantages:

- 1) Time consuming.
- 2) Accuracy not determined.

Our Objective

The primary aim of this work is to enhance the performance of Q&A systems by actively forwarding questions to users who are capable and willing to answer the questions. We additionally aim to enhance social question and answer with security and proficiency upgrades by ensuring users protection and recognizes and recovering answers naturally for intermittent inquiries. We depicted the architecture and calculations, and led far reaching substantial scale reenactment to assess social Q&A in examination with different strategies.

MATERIALS AND METHODS

The system consists of 2 modules in total.

Modular design of a system

1. User Module (social media):

A) Registration

In this module, the system let user to register into the system with general information like email id, age, contact number etc.

B) Login

In this module user will be asking their perspective question.

- Submit question on system

1. Answer finding and expert recommendation:

1. User's question is first processed to remove stop words like (Who, What, Where, How, Why, etc.)
2. Perform LDA for topic finding from user's question.
3. Search related questions and answers of similar topic on social network (user interest mapping) and online Google database (using Cosine and KNN algorithm), find experts for the same topic and send that question to that expert for answer.

2. Offline social network search for answer

1. Put question on user wall.

2. User's question is first processed to remove stop words like (Who, What, Where, How, Why, etc.)
3. Perform LDA for topic finding from user's question
4. Search related questions and answers of similar topic on social network (user interest mapping) (using Cosine and KNN algorithm), find experts for the same topic and send that question to that expert for answer.
5. Display the answer by finding accuracy of answer.

Combine result of online and offline together to show result to user and recommend experts to user

2. Admin

Login

1. Add question answers dataset of question answer system
2. Add some expert by default in the database.
3. Save the topic and user name.
4. Check the accuracy of answer using IF_IDF. And as per accuracy display answer to user.

Methodologies of problem solving and efficiency issues:

1. User Interest Analyzer

User Interest Analyzer uses every client's profile data in the social network and client collaborations (answers gave and addresses solicited) to decide the interests from the client in the predefined intrigue classes. This is on account of if a client asks or answers inquiries in an intrigue class, (s) he is probably going to be keen on this specific classification.

2. Question Categorizer

The essential errand of Question Categorizer is to order an inquiry into predefined intrigue classes in light of the topic(s) of the inquiry. We likewise enable clients to enter self characterized labels connect with questions, which are investigated being referred to parsing. Question Categorizer creates a vector of question "Qi's" interests, indicated by "VQi", utilizing a comparable calculation. While preparing an inquiry, Social Q&A utilizes Word Net to analyze the labels and content of the inquiry and creates a token string. The tokens are contrasted with Social Q&A's to decide the classifications where the inquiry has a place. We have ascertained the intrigue weight without standardization so as to foresee the client insight to answer an issue of Interest.

3. Question-User Mapper

Question-User Mapper distinguishes the proper answerers for a given inquiry. The potential answer suppliers are browsed the asker's companions in the online interpersonal organization. Note that the adjustments in a client's companions in the online interpersonal organization don't influence the execution of Social Q&A as it generally utilizes a client's present companions. To check the fittingness of a companion (U_k) as an answer supplier for an inquiry, two parameters are considered: 1) the intrigue closeness between the intrigue vectors of the companion and the inquiry. The previous speaks to the potential capacity of a companion to answer the inquiry, and the last speaks to the readiness of a companion to answer the inquiry.

Algorithm

1. TF_IDF

The TF-IDF value increases proportionally to the number of times a word appears in the document, but is often offset by the frequency of the word in the corpus, which helps to adjust for the fact that some words appear more frequently in general. Nowadays, TF-IDF is one of the most popular term-weighting schemes.

For example:

Document 1: The game of life is a game of everlasting learning

Document 2: The unexamined life is not worth living

Document 3: Never stop learning

Step 1: Term Frequency (TF):

Term Frequency also known as TF measures the number of times a term (word) occurs in a document. Given below are the terms and their frequency on each of the document.

Step 2: Inverse Document Frequency (IDF):

The main purpose of doing a search is to find out relevant documents matching the query. In the first step all terms are considered equally important. In fact certain terms that occur too frequently have little power in determining the relevance. We need a way to weigh down the effects of too frequently occurring terms. Also the terms that occur less in the document can be more relevant. We need a way to weigh up the effects of less frequently occurring terms.

Let us compute IDF for the term game:
 $idf(t, D) = \log |D| + 1 - \log |\{d \in D: t \in d\}|$

Step 3: TF * IDF

Remember we are trying to find out relevant documents for the query: life learning For each term in the query multiply its normalized term frequency with its IDF on each document. In Document1 for the term “life” the normalized term frequency is 0.1 and its IDF is 1.405507153. Multiplying them together we get 0.140550715 (0.1 * 1.405507153).

Given below is TF * IDF calculations for life and learning in all the documents.

Step 4: Vector Space Model – Cosine Similarity

From each document we derive a vector. If you need some refresher on vector refer here. The set of documents in a collection then is viewed as a set of vectors in a vector space. Each term will have its own axis. Using the formula given below we can find out the similarity between any two documents.

$$\cos(\theta) = \frac{v \cdot w}{\|v\| \|w\|} = \frac{\sum_{i=1}^n v_i w_i}{\sqrt{\sum_{i=1}^n v_i^2} \sqrt{\sum_{i=1}^n w_i^2}}$$

Fig. 1: Calculation formula to find the similarity of the documents.

In [Fig. 1], the calculation formula is used to find the similarity between the documents. Term Frequency also known as TF measures the number of times a term (word) occurs the document. The TF-IDF esteem builds relatively to the circumstances a word shows up in the record, however is regularly balanced by the recurrence of the word in the corpus, which helps to adjust for the fact that some words appear more frequently in general.

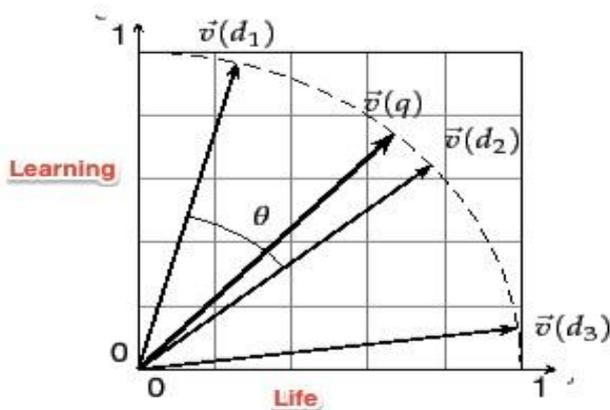


Fig. 2: Vector Space Model of Life and learning in documents.

In [Fig. 2], the vector space model has been implemented for calculations for life and learning in all the documents using the similarity formula. From each document we derive a vector. The set of documents in a collection then is viewed as a set of vectors in a vector space. Each term will have its own axis.

2. KNN algorithm :

Steps of the Algorithm are:

- 1 .Determine parameter k = number of nearest neighbour.
2. Calculate the distance between the query instance and all the training samples.
3. Sort the distance and determine nearest neighbour based on the k th minimum distance.
4. Gather the category y of the nearest neighbour.
5. Use simple majority of the category of nearest neighbour as the prediction value of the query instance

Architectural Diagram

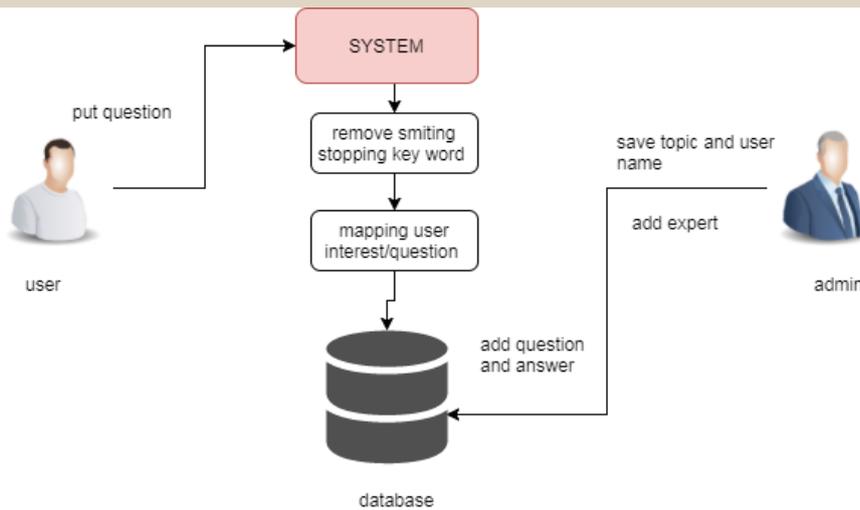


Fig. 3: Architectural diagram of proposed system.

In the [Fig. 3], User will register to the system with normal information. User put their question into the system, the user's question is first processed to remove stop words like (who, What, Where, How, Why etc). The system will perform LDA for topic finding from users question and search related question and answers of similar topic on social network using user interest mapping and by using Cosine and KNN algorithm. It will find expert for the same topic and send that question to that expert for answer. Finally system will display the answer by finding accuracy of answer.

Admin

Administrator can likewise include inquiries into the database.

1. Admin can add question and answer dataset of question answer system.
2. It can add some expert by default in database.
3. It can save the topic and user name.
4. It can check accuracy of answers using IF_IDF. And according to the precision show answers to users.

Advantages of proposed system

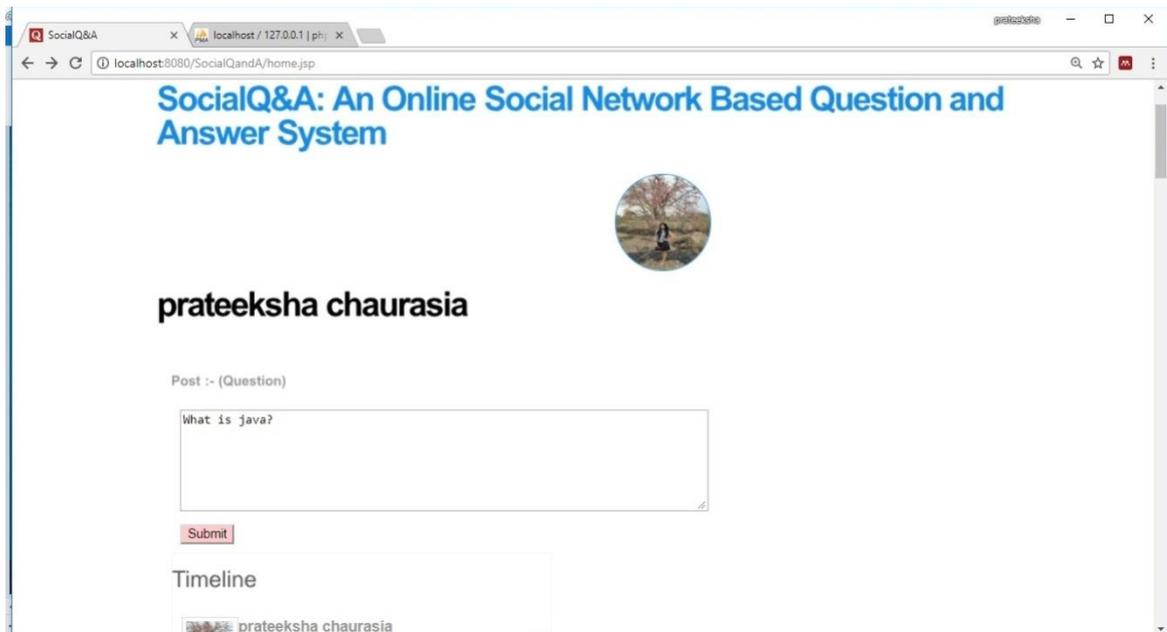
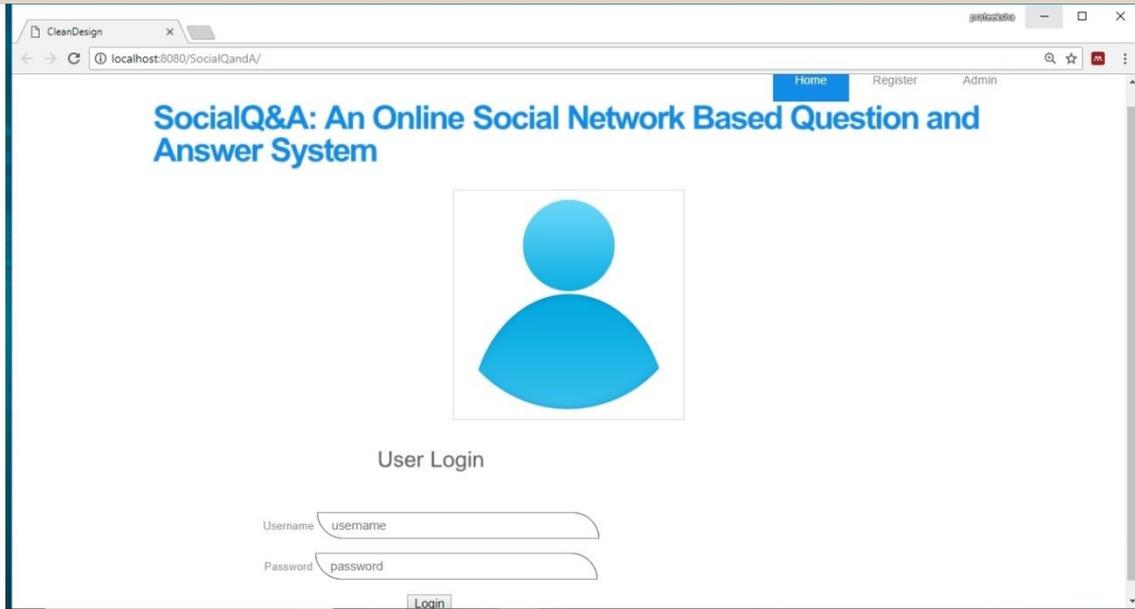
- With the help of proposed system user will get answer as soon as possible.
- Proposed system can work reasonably to balance privacy and security.
- Proposed system will recommend the expert to the user as per topic so that user will get exact and proper answer.

RESULTS

Here, Whole System taken many more attribute for the input purpose but here author mainly focuses on the Time and performance of system. Based some few attributes we will getting following analytical result for our proposed system.

The following modules has been obtained:

1. Users registration window.
2. Users put their question into the system.
3. Offline results of the Question asked by user with expert recommendation.
4. Online results of the question asked by users.



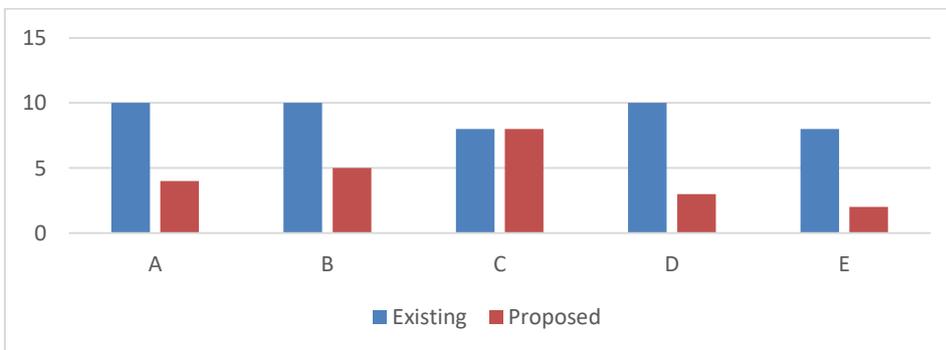
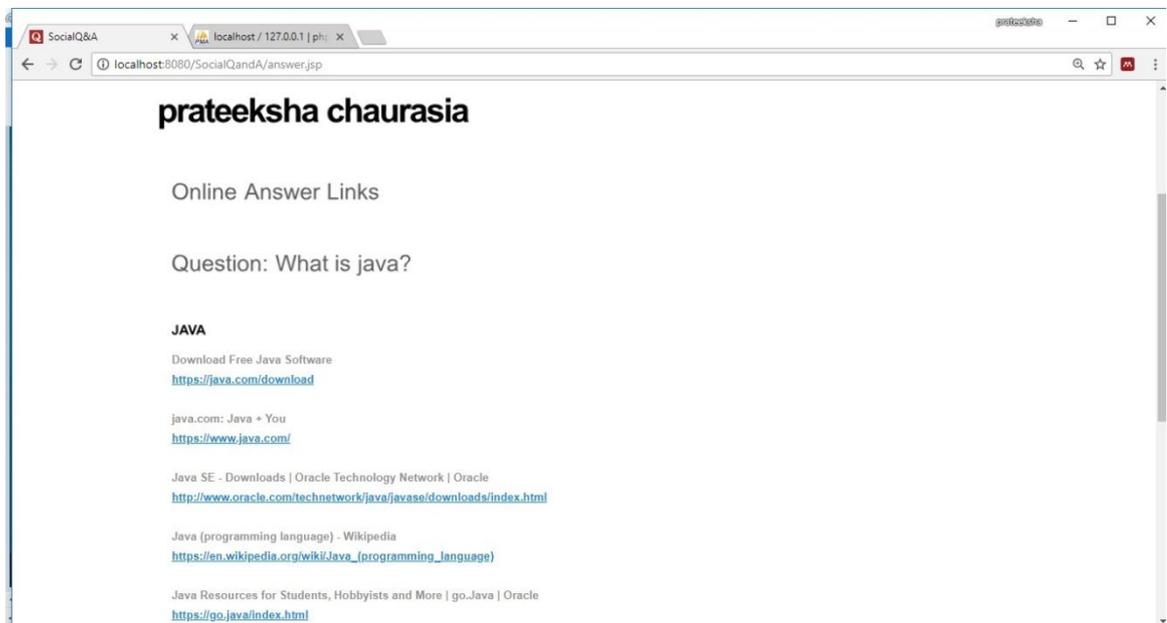
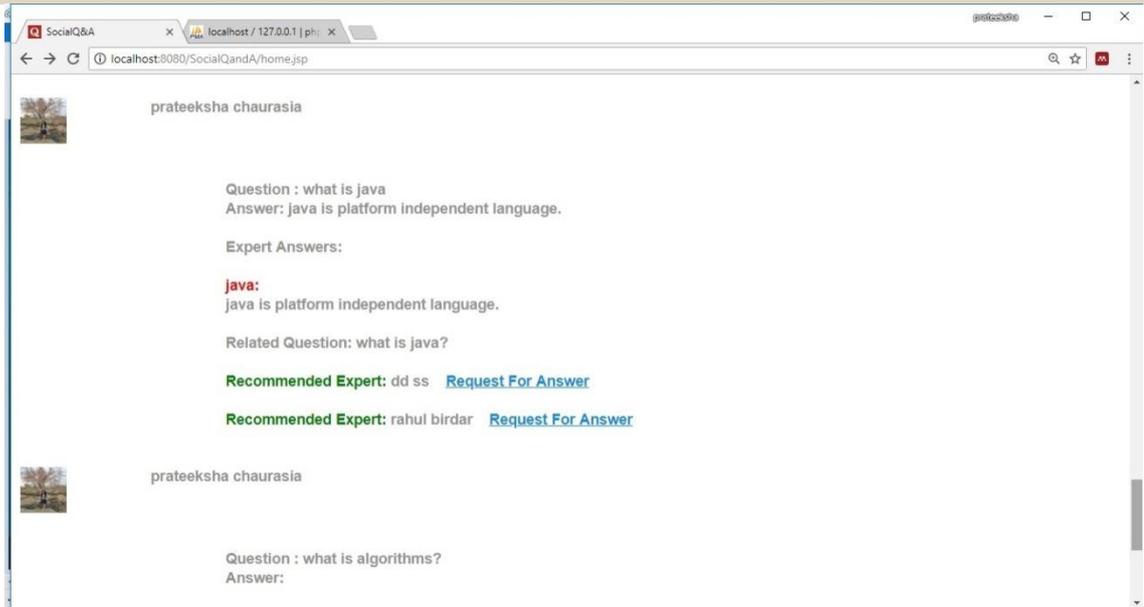


Fig. 4: Result analysis

In [Fig. 4], the analysis of the results is showing in the form of graph visualization. The chart shows the comparison between existing system and proposed system according to various parameters in the timeline.

Where,

A = Computation Cost.

B = Time Consumption.
 C = Scalable.
 D = Waiting Time.
 E = User Friendly

DISCUSSION

In the old system, it was very time taking to give the exact answers to their users, and correctness of the answers was also missing. We have overcome from this by applying TF_IDF algorithm and creating experts of particular fields.

Existing system

Q&A play a very important role in our everyday life for information and knowledge sharing however in our existing system quality of answer weren't that much correct or not satisfied to user and also the waiting time was more; to beat by this situation and user satisfaction we are adding new options within the system. [1]

Disadvantages of Existing System

1. Wait time – waiting time more, user need to wait to get the solution.
2. Quality data – user not satisfied by the solution.

Proposed system

Q&A play an important role in our daily life for information and knowledge sharing but traditional they were not that much accurate or not satisfied to user and the waiting time was more, to overcome by this problem we are enhancing the system by improving 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 that 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.

How insecurity reduced through proposed system

Social question and answer system with security and privacy enhancements protects user and retrieve answers automatically for questions post by users. This system will forward the questions to experts who can give the correct and expected answer as soon as possible.

Advantages of proposed system

1. With the help of proposed system user will get answer as soon as possible.
2. Proposed system can work reasonably to balance privacy and security.
3. Proposed system will recommend the expert to the user as per topic so that user will get exact and proper answer.

CONCLUSION

In our undertaking we are furnishing quality answer with less holding up time to number of clients. For quality answer and less hold up time we have created and prototyped an online informal community based Q&A framework, called Social Q&A.

It uses the properties of an interpersonal organization to forward an inquiry to potential answer suppliers, guaranteeing that a given inquiry gets a superb answer in a brief timeframe. It expels the weight from answer suppliers by straightforwardly conveying them the inquiries they may be occupied with, rather than requiring answer suppliers to seek through an extensive accumulation of inquiries as in Yahoo! Answers or flooding an inquiry to the majority of an asker's companions in an online informal community. The blossom channel based improvement strategies scramble the intrigue and Companionship data traded between clients to ensure client protection, and record all n-grams of addressed inquiries to naturally recover answers for intermittent question. Since same inquiries might be displayed distinctively and a similar inquiry might be addressed contrastingly in various circumstance.

CONFLICT OF INTEREST

There is no conflict of interest

ACKNOWLEDGEMENTS

The wastage of time of the users will be preserved. So they can get answers as soon as possible. Stop words are removed from the topic so that system will recommend the experts based on the topic only.

FINANCIAL DISCLOSURE

None

REFERENCES

- [1] MR Morris, J Teevan, K Panovich. [2010] What do People Ask Their Social Networks, and Why? A Survey Study of Status Message Q&A Behavior. In Proc. of CHI.
- [2] MR Morris, J Teevan, K Panovich. [2013] A Comparison of Information Seeking Using Search Engines and Social Networks. In In Proc of ICWSM,
- [3] MD Pandit. English translation of Sanskrit Grantha Lilavati. Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.
- [4] Z Gyongyi, G Koutrika, J Pedersen, H Garcia-Molina. [2008] Questioning Yahoo! Answers. In Proc. of QAWeb, 2008. Pawan Goyal, Gérard Huet, Amba Kulkarni, Peter Scharf3 Ralph Bunker, A Distributed Platform for Sanskrit Processing.
- [5] B Li, I King. [2006] Routing Questions to Appropriate Answerers in Community Question Answering Services. In Proc. of CIKM, 2010. fusion, SIGIR'06.
- [6] Yahoo! Answers Team. Yahoo! Answers BLOG. <http://yahoanswers.tumblr.com/>, [Accessed on 10/20/2014]. Konrad Rieck, Christian Wressnegger; 17(9):1-5, 2016, Journal of Machine Learning Research Homepage Harry: A Tool for Measuring String Similarity.
- [7] LA Adamic, J Zhang, E Bakshy, MS Ackerman. [2008] Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In Proc. of WWW.
- [8] D Bernhard, I Gurevych, [2008] Answering Learners' Questions by Retrieving Question Paraphrases from Social Q&A Sites, in Proceedings of the 3rd Workshop on Innovative Use of NLP for Building Educational Applications in conjunction with ACL 2008, (Columbus, Ohio, USA), pp. 44-52.
- [9] JS Coleman. [1988] Social capital in the creation of human capital. American journal of sociology 94, S1 95.
- [10] Burke M, Marlow C, Lento T. [2010] Social Network Activity and Social Well-Being. Proceedings of CHI, 1909-1912.
- [11] Kraut R, Patterson M, Lundmark V, Kiesler S. Mukopadhyay, T, Scherlis W. [1998] Internet paradox: A social technology that reduces social involvement and psychological well-being? American Psychologist, 53 (9):1017-1031.