KNOWLEDGE ENHANCEMENT PLATFORM USING Q&A SOCIAL MEDIA SYSTEM

PUNEET PALAGI, ASSISTANT PROFESSOR, DEPT. OF CSE, NHCE.

Abstract: The Internet is an important source of information, where the amount of data is vast and constantly growing. Users rely on search engines to find specific information in this knowledge base. Search engines such as Google and Bing use keywords provided by the users to perform searches. Previous researches have indicated that search engines perform well in indexing web pages and providing users with relevant content to their search but are not suited for non-factual questions. In this paper, I propose Social Q&A, an online social network based Q&A system that actively forwards questions to those users with the highest likelihood (capability and willingness) of answering them with expertise and interest in the questions subjects.

1. INTRODUCTION:

The Internet is an important source of information, where the amount of data is vast and constantly growing. Users rely on search engines to find specific information in this knowledge base. Search engines such as Google and Bing use keywords provided by the users to perform searches. Recently, industrial research and development activities, such as Microsoft and Facebook's social-featured Bing search endeavor, try to combine search engines and online social networks for higher search performance.

As previous research has indicated, search engines perform well in indexing web pages and providing users with relevant content to their search but are not suited for non-factual questions such as "Which is the best local auto shop?". To address this particular class of non-factual questions, many Question and Answer (Q&A) systems such as Yahoo! Answers, Baidu Zhidao, Stack-Exchange, Quora and Ask have been developed. Since their inception, Q&A systems have proved to be a valuable resource for sharing expertise and consequently are used by a large number of Internet users. For example, Yahoo! Answers was launched at the end of the year 2005 and attracted more than 10 million users in February, and hit 200 million users in December of 2009. Q&A systems also preserve all questions and answers, thus acting as a repository for information retrieval. They are not only important for sharing technical knowledge, but also as a source for receiving advice and satisfying one's curiosity about a wide variety of subjects.

As a result, 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. This is confirmed by the study in. 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. For Baidu Zhidao, 22.7% of questions were successfully answered, and 42.8% of the unresolved questions were not answered at all. Thus, there is an increasing need for an advanced Q&A system that can decrease the number of unanswered questions, enhance the answer quality and decrease the response time.

2. PROBLEM DEFINITION:

The growing importance of Q&A systems demands an effort to better understand these systems and to improve them. The work studies the influence of different factors (e.g., users' profiles, messages prediction, system interactions and community size) in the social networks on Q&A performance. These study results lay the foundation of Social Q&A to leverage social network properties in the design. Note that the existing social network based on the asker-answerer relationship in current Q&A systems is different from online social network based on the social relationship, which is used in Social Q&A. The work is concentrated on locating experts and authoritative users. Instead, Social Q&A aims to find normal users that can answer questions including opinion-type questions. Some studies have been conducted to create reputation models in Q&A systems to increase the credibility of answers, and to determine the relationship between the reputation of the users and the quality of their provided answers. Social Q&A directly utilizes the social network property of mutual-trust friendship to motivate users to provide answers without relying on an additional reputation model. Social Q&A shares similarity with other peer-assistant systems such as in leveraging the collective power of peers for a certain goal.

3. LITERATURE SURVEY:

Survey From (G. Drosatos, P. Efraimidis, A. Arampatzis, G. Stamatelatos, and I. Athanasiadis.)

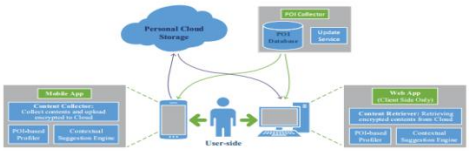


Fig 2.1 pythia system architecture.

Pythia is a new privacy-enhanced user-centric contextual suggestion system for tourism. The system exploits the user's digital trace to automatically generate and update a profile, which is safely kept at the user side. A contextual suggestion component, executed at the user's side, considers the profile and a location of interest, and generates POI suggestions. An overview of the Pythia system architecture is shown in Figure 1. The system comprises mobile, conventional (desktop or server-side), and cloud components. A personal cloud storage serves as intermediate storage for the application components and a POI collection framework manages the available POI data of the system.

Survey from (Ze Li, Haiying Shen Jin Li, Guoxin Liu.)

SOS incorporates an online social network, where nodes connect each other by their social links. As shown in Figure 1, 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.

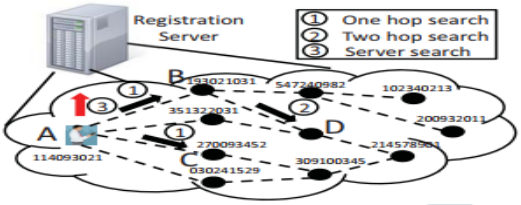
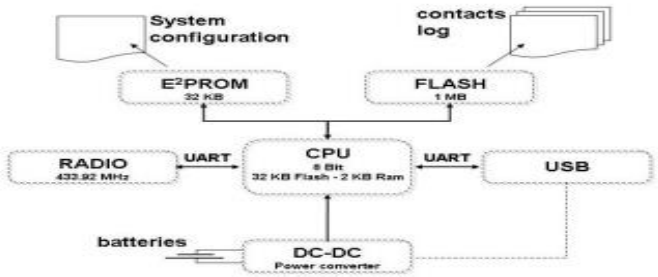


Fig 2.2 question routing process in SOS

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.

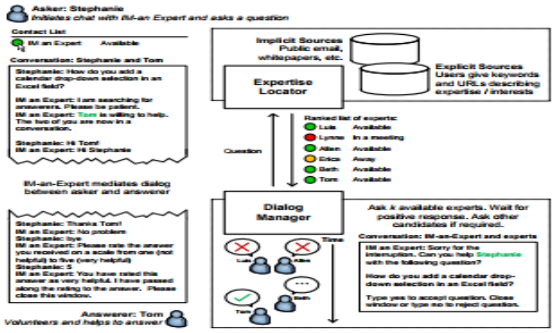
Survey from (Sabrina Gaito, Elena Pagani , Gian Paolo Rossi.)



The design of a specific device for trace recording is mainly motivated by the need of observing and recording very short contact periods, few seconds, that arise from random mobility in a dense area. As a consequence, PMTRs have been designed to operate with beaconing times ranging from 1 sec. to some

configurable value which depends on the mobility environment we wish to observe. Secondly, the devices have to enable unmanned experiments lasting 3-4 weeks without batteries substitution. Layer 2 beacons are the unique frames a PMTR broadcasts to its neighbors. The first time a beacon is received from a given encounter, the current time is recorded for the contact start, together with the encounter ID; a contact ends when beacons from a certain encounter have been missing for more than t seconds, with t = 60 in our experiments. The local memory size should be dimensioned to store the contacts of the experiments. Our test beds have generated on average 2000 contacts per device with beaconing time set to 1 sec. No specific bandwidth and processing requirements have been envisaged for PMTRs.

Survey from (Matthew Richardson and Ryen W. White)



IM-an-Expert is an automated service that receives questions via IM, locates and contacts potential answerers with expertise or interest in the question topic, and mediates the dialog between asker and answerer. Figure 1 illustrates the interaction flow in the IM-an-Expert service. The asker initiates an IM conversation with IM-an-Expert and poses the question. The question is used to retrieve a candidate set of experts based on profile information (described later). A small group of those experts who are currently available (not busy, away, in a meeting, in a call, etc. available via presence information from the IM client) are contacted via IM three at a time, in descending order of their expertise, to determine whether they are willing to help answer the question. If and when an answerer accepts, other requests are cancelled. If a candidate answerer does not respond in time or rejects the question, the service asks others. Once an answerer accepts, IM-an-Expert mediates the conversation between the asker and answerer. Once the conversation ends, the asker is asked to optionally rate the quality of the answer they received on a scale from one (not helpful) to five (very helpful). To support the functionality described in this paragraph, the system has two components: (i) expertise locator: selects users who are most likely able to answer a question, and (ii) dialog manager: handles question processing and communication management throughout the question lifecycle.

4. EXISTING SYSTEM

It is found that Yahoo! answers, only 17.6% of questions satisfactorily; for the remaining 82.4%, one fifth of the questions remained unanswered. For Baidu Zhidao, 22.7% of questions were successfully answered, and 42.8% of the unresolved questions were not answered at all. Thus, there is an increasing need for an advanced Q&A system that can decrease the number of unanswered questions, enhance the answer quality and decrease the response time. Some research categorizes questions into predefined categories, making it easier for users to locate previously asked questions and for experts to find questions they can answer.

Quan et al. proposed three new supervised term weighting schemes for question categorization, and evaluated each scheme using a trace from Yahoo! Answers.

Song et al. proposed a sequential process including topic-wise word identification and weighting, semantic mapping, and similarity calculation.

Disadvantages:

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.

Since Social Q&A is built upon social networks, the asker and answerer are socially close to each other. Therefore, protecting the privacy is important and challenge.

5. PROPOSED SYSTEM

We propose Social Q&A, an online social network based Q&A system that actively forwards questions to those users with the highest likelihood (capability and willingness) of answering them with expertise and interest in the questions' subjects. The design of Social Q&A is based on two social network properties. First, social friends tend to share similar interests (e.g., lab members majoring in computer systems). Second, social friends tend to be trustworthy and altruistic due to the property of "friendship fosters cooperation".

The design of Social Q&A 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.

Advantages:
Different from the existing Q&A systems, due to the importance of users privacy, we future introduce security and efficiency enhancement to protect users privacy while users using social network answering questions.
Social Q&A incorporates three methods to enhance its security and efficiency performance. The bloom filter based personal information exchange method protects users' privacy including friendship and interest information.

The onion routing based answer forwarding method protects the identities of the asker and the answerer from being exposed. The answer retrieval for recurrent questions automatically finds the answers for recurrent questions. We conducted comprehensive large-scale simulation to evaluate Social Q&A in comparison with other methods. Our results suggest that Social Q&A improves the quality of answers and reduces the wait time for answers.

We have prototyped the Social Q&A system with user interfaces, and conducted a real-world small-scale test with real users from India, the United Kingdom, and the United States for a period of approximately one month. We have analysed the features of the questions posted, the questioning and answering activities of users, the quality of answers, and the wait time for answers. Analytical results show the benefits of Social Q&A in enhancing answer quality and wait time.

6. SYSTEM ARCHITECTURE

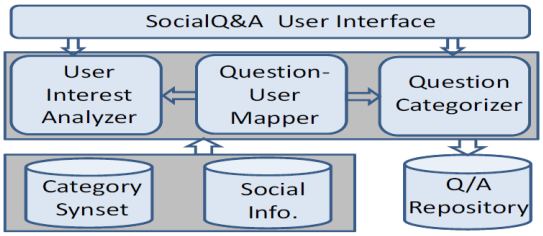


Fig 4.1 System Architecture

The high-level architecture of Social Q&A and the interaction between the core components: User Interest Analyzer, Question Categorizer, and Question-User of the Mapper. User Interest Analyzer analyzes data associated with each user in the social network to derive user interests. Question Categorizer categorizes the user question into interest categories based on the Category Synsets, which stores the synonyms of all categories' keywords from WordNet. Question-User Mapper connects these two components by identifying potential answerers who are most likely to be willing to and be able to provide satisfactory answers. The data from user questions and answers is stored on *Q/A* Repository to serve subsequent similar questions. Below, we present each component and user interface.

7. SYSTEM DESIGN

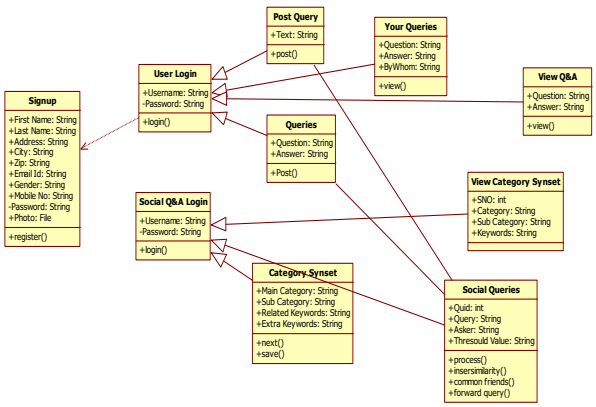


Fig 4.2 Project class

8. CONCLUSION

Q&A systems are used by many people for purposes such as information retrieval, academic assistance, and discussion. To increase the quality of answers received and decrease the wait time for answers, I am developing and prototyping an online social network based Q&A system, called Social Q&A. It utilizes the properties of a social network to forward a question to potential answer providers, ensuring that a given question receives a high-quality answer in a short period of time. 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.

The bloom filter based enhancement methods encrypt the interest and friendship information exchanged between users to protect user privacy, and record all n-grams of answered questions to automatically retrieve answers for recurrent question. The onion routing based answer forwarding protects the identities of askers and answers. Our comprehensive trace driven experiments and analysis results on the real-world Q&A activities from the Social Q&A prototype show the promises of Social Q&A to enhance answer quality and reduce answer wait time in current Q&A systems, and demonstrate the secure and efficiency improvement achieved by the enhancements. Since same questions may be presented very differently and the same question may be answered differently in different situation. In the future, will cooperate with other techniques into Social Q&A to find the redundant question with a large scale user set. Due to the dynamic of user behavior, Social Q&A can cooperate a machine learning method to adjust three parameters appropriately, which needs a large user base and much more usage. We will conduct tests on a large user base in the real-world experiment.

9. REFERENCES

[1] M. R. Morris, J. Teevan, and K. Panovich. A Comparison of Information Seeking Using Search Engines and Social Networks. In In Proc. of ICWSM, 2010.

[2] 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.

[3] Z. Gyongyi, G. Koutrika, J. Pedersen, and H. Garcia-Molina. Questioning Yahoo! Answers. In Proc. of QAWeb, 2008.

[4] Yahoo!Answers Team. Yahoo! Answers BLOG. http://yahooanswers.tumblr.com/, [Accessed on 10/20/2014].

[5] B. Li and I. King. Routing Questions to Appropriate Answerers in Community Question Answering Services. In Proc. of CIKM, 2010.

[6] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In Proc. of WWW, 2008.

[7] G. Drosatos, P. Efraimidis, A. Arampatzis, G. Stamatelatos, and I. Athanasiadis. Pythia: A privacy-enhanced personalized contextual suggestion system for tourism. In COMPSAC, 2015.

[8] S. Li, Q. Jin, X. Jiang, and J. Park. Frontier and Future Development of Information Technology in Medicine and Education: ITME 2013. Springer Science & Business Media, 2013.

[9] A. Mtibaa, M. May, C. Diot, and M. Ammar. Peoplerank: Social Opportunistic Forwarding. In Proc. of Infocom, 2010.