However, the previous broadcasting and centralized methods are not suitable to the mobile environment, where each mobile node has limited resources. Broadcasting questions to a large number of friends cannot guarantee the quality of the answers.

II.SYSTEM DESIGN

A. Question Processing

Social Network based Question Answer system(SNQA) incorporates an online social network, where nodes connect each other by their social links. A registration server is responsible for user registration. Each user has an interest ID, which represents his/her interest. Users who have been willing to answer questions and provided high quality answers to node *i*'s questions previously are more likely to be willing to answer node *i*'s questions and provide high quality answers. Thus, SNQA has a metric best answerer $(BA_{(qi,j)})$ that measures likelihood of node *j* to be able and willing to answer node *i*'s question *qi* with a high quality answer. It is determined by the interest similarity $(S_{(qi,j)})$ between the question q_i's interest and node *j*'s interest as well as the answer quality ($Q_{(i,j)}$) of node *j* to node i's previous questions.

B.Question/User Interest Mapping

When a user first uses the SNQA system, s(he) is required to complete his/her social profile such as interests, professional background and so on. Based on the social information, the registration server recommends friends to the user, and the user then adds friends into his/her friend list. Each user locally stores his/her own profile and interest ID, and friend list and their interest IDs and answer quality values. Each user calculates his/her own interest ID on his/her soci**al** information and sends it to their friends. To calculate interest ID, a node first drives the first-order logic representation (FOL) from its social information, then conducts first-order logic inference to infer its interests, from which it decides the interest ID.

To parse a question, the node first processes the question using natural language processing (NLP), and then represents the question in the FOL format and uses the FOL inference to infer the question's interests. Finally, it transforms the question to a question ID in the form of a numerical string. After a node i parses its initiated question *qi* to a question ID, it calculates the interest similarity $S(qi\,j)$ for each of its friends $j \in F_i$, where F_i denotes the set of node i's friends. It then calculates the best answerer value ($BA_{(qi,j)}$) for each friend *j* by combining $S_{(qi,j)}$ and answer quality from friend $j(Q_{(i,j)})$.Finally, node i choose top K friends that have the highest $BA_{(qi,j)}$ values to send the question. By comparing the similarity between a question's ID and its friend's interest ID, a node can identify its friends that are able to answer questions.

C.First-order Logic Inference

The FOL[12] inference component consists of three parts: (1) fuzzy database, (2) rules and axioms, (3) inference engine. The goal of the inference is to identify node interests represented by a numerical string that can accurately represent the capability of a node to answer questions. The fuzzy database is used to store words that have relationships, including subset, alias(x), related, with the information in profiles. For example, related (cinema) =movie, subset (computer science, algorithm), alias (USA) =US.The rule and axioms provide basic formulas for the inference.The inference engine checks the rules and finds related but not obvious information. It sets each interest as an inference goal and builds lattice inference structure, as shown in Figure 1, to connect All the FOL symbols with the goals. Each node in the lattice is an FOL syntax symbol and the arrows represent the connective symbols that connect the symbols.

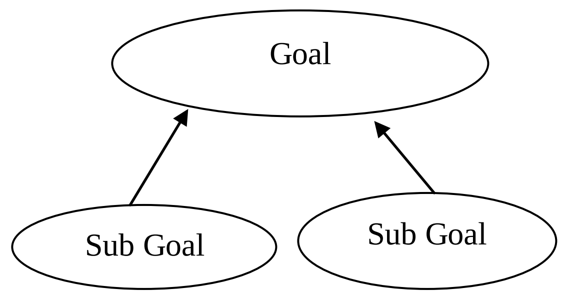

Fig.1 Lattice in inference engine

III. SIMILARITY VALUE CALCULATION

After user's social information and questions are transformed into numerical strings, the similarity between a user and a question can be calculated based on two parts: Interest similarity between the user and question, and answer quality between the question sender and receiver.

A. *Interest similarity calculation*

To evaluate the interest similarity of a question of user i (*qi*) and a user *j*, we use a method proposed in [25]. We use ID_{qi} and ID_j to denote the interest strings of question *qi* and user *j*, respectively. We use $n_{(qi,j)}$ to denote the number of interests owned by ID_{qi} but not by ID_j; use $l_{(qi,j)}$ to denote the number of categories of interest elements owned both by ID_{qi} and ID_j , and $m_{(qi,j)}$ the number of categories of interest elements owned by ID_j but not by ID_{qi}. Then the interest similarity of question *qi* and user *j* is defined as:

$$S_{(qi,j)}=\frac{l(qi,j)+1}{2}\left(\frac{1}{l(qi,j)+n(qi,j)+2}+\frac{1}{l(qi,j)+m(qi,j)+2}\right). \quad (1)$$

ISSN:2249-5789
Bagwan Ruksar et al , International Journal of Computer Science & Communication Networks,Vol 4(6),193-196

The value of (qi, j) ranges in the classical spectrum [0, 1], and it represents the level of likelihood that two strings under comparison are actually similar. If two strings have complete overlapping (n=m=0), $S(qi, j)$ approaches 1 as the number of common features grows. The underlying idea of Equation (1) is that two strings with longer complete overlapping should have higher similarity than the two strings with less complete overlapping. In the case of no overlapping ($l=0$), the function approaches to 0 as long as the number of non-shared entries grows. It indicates that two strings with a larger number of entries and share no common entries are more likely to have smaller similarity than the two strings with a smaller number of entries and share no common entries.

B. Answer quality calculation

Social closeness value calculation mechanisms are based on the whole social network topology, which are energy consuming. It is even worse when the social network dynamically changes. Therefore, the topology base social closeness calculation methods are not suitable for energy stringent mobile devices in SNQA. Performance of the SNQA largely depends on the activeness and the knowledge base of the users, user *i* considers the number of received answers from user *j* and their associated quality ratings when calculating the answer quality of user *j*. we call it as feedback mechanism. For each received answer, an asker can rate the quality of the answer within rating scale R=[1,5]. The answer quality value is updated based on the number of answers received from friend *j* during each period T and the associated quality rating ($r \in [1, 5]$). For the *kth* question sent from node *I* to node *j*, if node *I* receives an answer from node *j* during *T*, $xk=1$; otherwise, $xk=0$. The parameter xk is used to represent the willingness of node *j* to answer questions from node *i*. Then, the answer quality (i,j) is calculated by:

$$Q(i,j) = \alpha . (i,j) + 1 - \alpha . k(xk . rk / R)\ (xk = 0,1) \quad .(2)$$

Where $\alpha \in [0, 1]$ is a damping factor, *rk* is node *i*'s quality rating for the *kth* answer received from node *j*. A larger $(i,)$ implies that user *j* is willing and able to provide high-quality answers to user *i*. Considering the high dynamism of the social networks, in which the willingness of users to answer questions and the quality of answers from a user to another user may change over time, we add damping factor α into the answer quality calculation.

A. Best answer metric calculation

Based on above sections, for its generated or received question *qi* that it cannot answer, node *i* calculates the best answer metric of each of its friends. That is

$$B(qi, j) = \beta S\ qi, j + (1 - \beta) Q(i, j) \quad (3)$$

Where $\beta \in [0,1]$ is a parameter used to adjust the weight of the similarity and answer quality. Node *i* then selects the top *K* friends that have the highest (qi, j) values and forwards the question to them. Social trust between two nodes decrease exponentially with distance. This relationship has been confirmed by other studies [28, 29]. A reduction in social distance between two persons significantly increases the trust between them.
Algorithm 1 shows the pseudo code of the process for the best answerer metric calculation and best answerer selection conducted by node *i*. If node *i* does not receive answers for its created question during the time corresponding to TTL, it resorts to the centralized server for the answers, where all users conduct Q&A activities in online Q&A sites.

Algorithm 1 [11]
Pseudo code of the best answerer identification executed by node *i*.
1: **Input:** *IDi*, *IDj* , *Q*(*i*,*j*) (*j* ∈ *Fi*)
2: **Output:** top-*K* best answerers
3: //*Periodically update Q*(*i*,*j*) (*j* ∈ *Fi*)
4: **for** each friend *j* in friend list *Fi* **do**
5: Update *Q*(*i*,*j*) based on Equation (2)
6: **end for**
7: **if** create a question or receive a question it cannot answer **then**
8: **if** TTL>0 **then**
9: **for** each friend *j* in friend list *Fi* **do**
10: Calculate *S*(*qi*,*j*) using *IDqi* and *IDj* based on Equation (1)
11: Calculate *BA*(*i*,*j*) using *Q*(*i*,*j*) and *S*(*qi*,*j*) based on Equation (3)
12: Add *BA*(*i*,*j*) to a list *List*
13: **end for**
14: QuickSort partition around the *Kth* largest element in *List*
15: Find the top-*K* friends having the highest *BA*(*i*,*j*)
16: TTL-=1
17: Send the question to the *ID*entified *K* friends
18: **end if**
19: **end if**
20: **if** does not receive answers for its created question during the time corresponding to TTL **then**
21: Resort to the centralized server for the answers
22: **end if**

Line4-Line6 are used to periodically update answer quality of each of its friends. Line8-Line13 calculates each friend's best answerer metric and generates a list including all metric values.

Line14-Line17 identifies the top-K friends with the highest best answerer metric values and send question to them. Answer quality (i,j) is pre-processed and only interest similarity $S\ qi\ ,j$ need to be calculated at run time. The $(qi\ ,j)$ calculation has a time complexity of $O\ Fi$. As the number of keywords in a question is generally very small, the calculation of $S(qi\ ,j)$ should take a short time and costs little computation resources of the mobile devices. This top-K friend selection algorithm has a time complexity of$O$ (| F_i |).

IV.CONCLUSION

SNQA systems are used by a large group of people for purposes such as information retrieval, academic assistance, and discussion. The growing importance of SNQA systems has led to numerous research developments that are directed toward making SNQA systems more effective. The motivation for this survey is to study design of SNQA system. Also it describe algorithm that used for best answer selection. Finally,SNQA system can be viewed as alternative solution to search engines.

REFERENCES

[1] Google. http://www.google.com

[2] B. M. Evans and E. H. Chi. An elaborated model of social search Information Processing & Management, 2009.

[3] E. Amitay, D. Carmel, N. Har'El, S. Ofek-Koifman, A. Soffer, S. Yogev, and N. Golbandi. Social search and discovery using a unified approach. In *Proc. of HT*, 2009.

[4] X. Yan. Smallblue: Social network analysis for expertise search and collective intelligence. In *Proc. of ICDE*, 2009.

[5] Harper, D. Raban, S. Rafaeli, and *J*. Konstan. Predictors of answer quality in online Q&A sites. In *Proc. of SIGCHI*, 2008.

[6] 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.

[7] J. Teevan, M. R. Morris, and K. Panovich. Factors affecting response quantity, quality, and speed for questions asked via social network status messages. In *Proc. of AAAI*, 2011.

[8] M. Richardson and R. W. White. Supporting synchronous social q&a throughout the question lifecycle. In *Proc. ofWWW*, 2011.

[9] J. Raacke and *J*. Bonds-Raacke. MySpace and Facebook: applying the uses and gratifications theory to exploring friend-networking sites.CyberPsychology & Behavior, 2008.

[10] M. R. Morris, *J*. Teevan, , and K. Panovich. A comparison of information seeking using search engines and social networks. In *Proc. of ICSWM*,2010.

[11] Z. Li, H. Shen, G. Liu, and *J*. Li. SNQA: A Distributed Context-AwareQuestion Answering System Based on Social Networks. In *Proc. ofICDCS*, 2012.

[12] M. Kirsten and S. Wrobel. Extending K-Means Clustering to First-Order Representations. In *Proc. of ICILP*, 2000.

[13] Z. Li and H. Shen. SOAP: A social network aided personalized and effective spam filter to clean your E-mail box. In Proc. of INFOCOM, 2011.

[14] E-commerce. http://en.wikipedia.org/wiki/E-commerce.

[15] C. Lampe, *J*. Vitak, R. Gray, and N. B. Ellison. Perceptions of facebook's value as an information source. In *Proc. Of CHI*, 2012