Social question and answer sites: the story so far

Article · December 2016

CITATIONS
0

READS
1,197

1 author:

Aftab Alam
Kyung Hee University
30 PUBLICATIONS **85** CITATIONS

SEE PROFILE

Program

Social question and answer sites: the story so far

Shah Khusro, Aftab Alam, Shah Khalid,

Social question and answer sites: the story so far

Shah Khusro, Aftab Alam and Shah Khalid
Department of Computer Science, University of Peshawar, Peshawar, Pakistan

Received 12 November 2015
Revised 12 October 2016
Accepted 5 November 2016

Abstract

Purpose – Social question and answer (SQA) site is one of the factors that boosted up and popularized the vision of social web. It enables community members to post highly valued answers to globally asked questions and information seekers to grab intellectual information in a contextual, concise, and meaningful format at the cost of investing a few minutes. The purpose of this paper is to present a common architecture, history, and a comprehensive review of such sites.
Design/methodology/approach – A critical and analytical investigation of the state-of-the-art SQA sites and relevant literature has been carried out with the intention to explore the noticeable features of such sites.
Findings – By studying relevant literature, and analysing a number of existing systems, a number of research challenges are identified and a generic architecture of SQA sites is contributed.
Practical implications – The review contributes a comprehensive knowledge about SQA systems and aims to be helpful to new researchers who want to get a broad picture of SQA systems on a single platform. The domain is in its infancy and requires tremendous efforts from the research community to explore its salient aspects with respect to the human world.
Originality/value – The study inspects SQA sites on a large scale and makes an original contribution by presenting a comprehensive review, future research challenges, and a generic architecture of SQA sites.

Keywords Web 2.0, Social Q&A, Architecture of social Q&A, Collaborative information seeking, Community question and answers, Expert generation

Paper type General review

Introduction

Since the inception of the web, it has become the first place where people look for information. Information seekers use two alternatives most commonly on the web, namely keyword searching and asking questions which exploit human mental capabilities for acquiring information in natural language. Keyword-searching systems like web search engines and web directories use available resources to map user query for finding relevant resources. In many situations, retrieved results are not satisfactorily relevant and accurate. The results are retrieved in huge quantities, which results in cognitive overload where deciding the reliability of information becomes difficult. Ranking plays an important role in handing this situation by bringing the most relevant results on top of the search results' list; however, it has its own problems and limitations. What about the information that is in the human mind or not yet uploaded? Such information can neither be retrieved by web search engines nor browsed in the web directories.

Rather than searching a bulk of documents that may never provide satisfactory results, a person can find a context-aware, concise, and meaningful solution to a problem. The alternative is to find the solution to one's problem by asking others on the web. In this context, social question and answer (SQA) platforms have attracted millions of web users. At the time of writing this paper, there are a number of SQA platforms available and accessible on the web, e.g. Yahoo! Answers[1], Answer Bag[2], and much more. SQA platform can be described as a system that is mainly designed to allow users to ask questions as well as respond to questions on a vast array of topics (Farzan, Dimicco, Millen, Brownholtz, Geyer and Dugan, 2008). It is a way of getting help and helping others by giving answers which are otherwise difficult to retrieve on the web.

Online question and answer (Q&A) services can be classified as digital references, expert services, and SQA communities (Shah *et al.*, 2009). A digital reference is "where the services

Program
Vol. 51 No. 2, 2017
pp. 170-192

0033-0337
DOI 10.1108/PROG-11-2015-0073

are provided by library professionals in digital environment" (Lankes, 2004). An expert service is an online question answering system in which real-world subject experts answer questions. SQA sites allow users to submit problems to the community and get the response from community members having expertise or knowledge about the problem.

Scientists have used the terms SQA and community Q&A with almost equal frequency for such platforms where community members find the solution to a problem with collaborative efforts (Gazan, 2011; Shah *et al.*, 2009). The term "social" has been used because it supports the social interaction of community members and exemplifies the vision of social web. Throughout this paper, the word SQA has been used to scope the review.

The objective of this review paper is to discuss SQA sites from different perspectives and to provide a scholarly overview of these systems. The contribution of the paper can be summarized as follows:

- To present a common architecture of online SQA sites for the ease of explanation and understandability, and to create some common nomenclature. The functionalities of each layer and components' dependencies are spotlighted.
- Different aspects of SQA sites have been reviewed in an organized manner, which could help the new researchers in finding the relevant knowledge on a single platform.
- An up-to-date survey of standard SQA sites with relevant statistics while projecting their importance.
- Significant research challenges are discussed so that the need to develop efficient SQA sites could be addressed.

Remaining paper is structured as follows. First, we discuss the evolution of SQA sites and then present a common architecture, followed by a case study, and finally conclude discussion and put forward suggestions for future research.

Online SQA sites as evolution of the web

The ultimate consequence of web 2.0 applications is the exponential growth in the web size which resultantly made it difficult for users to memorize all URLs. In the mid-1990s, web search engines and directories got an overwhelming response as information retrieval tools and acted as gateway to online information.

A web search engine is the most common tool for searching information. The notion of a search engine is to retrieve information by exploiting the current semantic-free structure of the web (Zdravko and Daniel, 2007). A search engine process consists of three phases, i.e. web crawling, indexing, and searching. A search engine exploits available information and retrieves the required information using different information retrieval techniques. These corpus-based search engines use keywords to retrieve information. They utilize the available web resources and map users' keywords to find out relevant information. Web documents and resources like digital libraries, wikis, blogs, photo sharing, video sharing, forums, social networks, etc., are the soul of a search engine (Zdravko and Daniel, 2007).

However, there are certain limitations in using these search engines. A keyword-based search engine returns an enormous amount of information against the user's query and the user has to make further efforts in order to get the appropriate document. Search engines exploit available resources and if such information is not uploaded, it cannot be retrieved through information retrieval schemes. It is not always possible to ask questions in plain language, i.e. express feelings, give suggestions, recommendations, advice, or opinion. For example, "I have done intermediate with first division and now thinking to take admission in BS. What subject should I opt for?" Such kind of questions cannot be asked from search engines. Substitute solutions exist in the form of instant messages, e-mailing,

bulletin-board systems, and discussion forums. These platforms utilize human power rather than documents and enable a user to ask a question, to express feelings, give suggestions, recommendations, advice, or opinions in natural language.

The idea of online question answering is not new but has evolved since 1979 (Lueg and Fisher, 2012). In the broader sense, there are many platforms which support online question answering services and enable a user to ask questions in natural language from other users. E-mail is used to send electronic mail, but it is also a kind of platform through which one can ask a question. To use e-mail, one must know the exact address of the answerer. A bulletin-board system, such as Usenet (Lueg and Fisher, 2012), was a network application and was used for asking the question in a specific category and other users were able to answer or pass comments on the question. Web forums got popularity in the early days of the web and are still used. Forums are commonly used for discussion purpose and focus on a specific topic. The main purpose of forums is to have a discussion rather than question answering.

Digital reference services, expert's services, and SQA sites (Boyd and Ellison, 2007) are true question answer platforms and are solely designed for the purpose of online Q&A. A digital reference service is the evolution of conventional reference service where the services are provided by the library professionals in digital environment (Jenna, 2007), e.g. www.ipl.org. An expert service is a question answering service in which real-world subject experts provide answers to a problem. This is based on one-to-one interaction, which means that the asker gets the response from a specific subject expert for his/her question. An expert service can either be free or payment based. In some expert services, users are allowed to rate or comment on either experts or experts' answers. Examples are Allexperts[3], Justanswer[4], and many more. An SQA site is completely different in nature from the above services; it allows community members to post a question, which gets a response from other community members, as for example, Naver's Knowledge-iN[5] (KiN), Yahoo Answers, and Answerbag.

Shah *et al.* (2009) proposed a working definition of SQA site as a platform where asker(s) post question in natural language instead of keywords, other community members respond and suggest the solution to asker's question, and this volunteer participation builds a community.

Active members are the soul of SQA sites (Welser *et al.*, 2007). SQA sites map the question to the right person or to the community for answering (Horowitz and Kamvar, 2010). A user can ask the question in natural language (Shah *et al.*, 2008) from the community instead of keywords. In an SQA site, asker's question can receive personalized, concise, and contextualized answer(s) from the community members rather than a list of documents.

Almost all popular SQA sites, e.g. Naver's knowledge-iN and Yahoo! Answers, allow the user to explore existing knowledge without session management. However, to participate in SQA sites in the form of asking, answering, commenting, tagging, liking, disliking, and voting, it is mandatory to register. By tracking user activities, an SQA site finds user interests, expertise in a topic, and the quality of questions and answers. SQA sites exploit user profile information for question recommendation purpose and for finding a right person to answer the questions. A user profile is also helpful in predicting and evaluating content quality (Shah and Pomerantz, 2010).

Questioning and answering is not the sole purpose of SQA sites, but they also aim to create knowledge archives with collaborative efforts. In 2002, Naver started KiN with the aim to create a knowledge archive in South Korean (Nam *et al.*, 2009) because at that time most of the websites were in English. Yahoo! Answers is the second most visited educational reference site after Wikipedia (Jacob, 2007). After the creation of a knowledge archive, automatic SQA system can be built to return to the point and provide a brief answer to the information seeker rather than a list of documents (Wenyin *et al.*, 2009). Text Retrieval Conference[6] (TREC) encouraged automatic Q&A research. In TREC11, focus was directed towards question answer systems such that to return an exact answer to the question from the knowledge base (Page *et al.*, 1999).

A common architecture of SQA

An SQA is an online website that:

- acts as a hub for community members to post their problems under a wide variety of categories ranging from daily life to school subjects and to find solutions with collaborative efforts;
- allows users to control the quality of the contents with collaborative efforts and enables the users to rate, support, edit, or register complaint either about user(s) or content(s);
- allows users to define an online profile;
- allows the user to set access permissions, privacy settings, and preferences; and
- builds user reputation on the basis of user activities and other users' feedback.

However, variations exist technologically, algorithmically, and in the way the community members are motivated for long-term participation. The life cycle of a question in SQA sites goes through different algorithmic, economic, and social processing strategies such as answers, supports, and comments (Gazan, 2011), as shown in Figure 1.

The above-mentioned attributes can be mapped to the following architectural components:

- users profile;
- user activity;
- social feedback;
- user reputation; and
- access control.

From the above facts, a common layered architecture has been developed based on the client-server model as shown in Figure 2. On the client side, user agents are used for

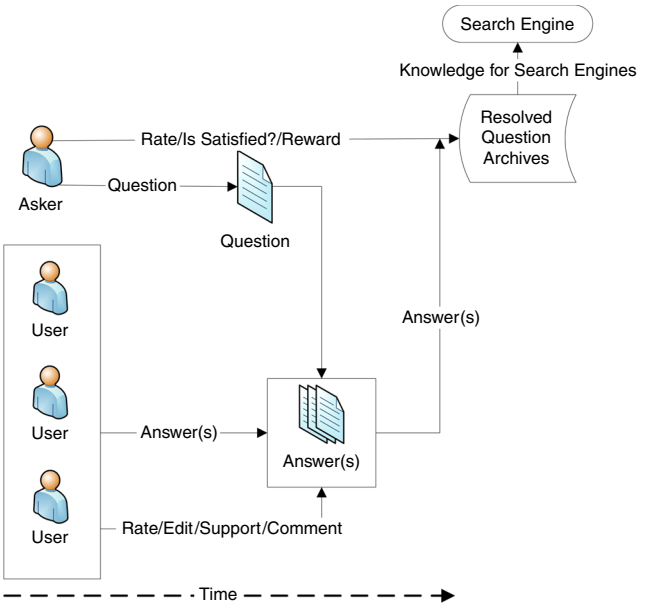

Figure 1.
Question life cycle in SQA sites

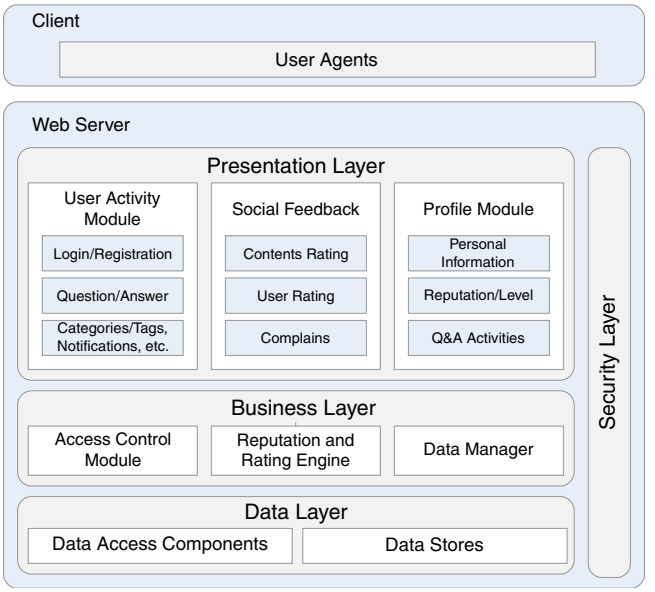

Figure 2.
A common architecture of SQA site

accessing the data where the core SQA application resides on the server and comprises of three layers, i.e. presentation layer, business layer, and data layer. The security layer is shown as cross-cutting because security issues are mandatory and common to all layers. The security cross-cutting layer of SQA system supports operations like authorization, authentication, exception management, and validation.

Presentation layer

Presentation layer facilitates user interaction and consists of user interface components and presentation logic. This layer consists of three main components, i.e. user activity module, social feedback, and profile module. user activity module provides basic facilities like user registration, login, answering, and asking a question under a specific category or tag. User activity module uses the services of data manager and access control module.

SQA sites are based on user-generated contents which are produced with collaborative efforts. Social feedback module allows community members to rate contents and users. This module also facilitates users to complain abusive material. The social feedback module rating works as an input for reputation and rating engine. The third component of this layer is profile module which uses the services of data manager to display user activities, reputation, and information.

Business layer

Business layer works as a mediator between presentation layer and data layer and implements the core functionality of SQA system. SQA business logic layer commonly consists of components like access control module, reputation and rating engine, and data manager. Access control module is responsible for controlling the access of community members according to their level, position, and authority by using predefined access control schemes. Reputation and rating engine is used to distinguish quality contents from bad ones and also to scale the reputation of community members accordingly. Data manager component facilitates the maintenance and retrieval of contents.

Data layer

Data layer consists of two components, namely, data access components and data stores. The SQA data such as profile information, questions, answers, reputation data, and other necessary information are stored in data stores. Business layer accesses this data by exploiting the services of data access components.

Important aspects of SQA sites

The key aspects of SQA sites are users, questions, and answers. Users play a vital role in SQA sites because these sites are developed to serve them whereas questions and answers are the nutrients of such sites. Researchers are investigating to diagnose these aspects in the quest of finding solutions to problems like:

- How to categorize questions and bring accuracy in question retrieval?
- How to classify answers and evaluate its quality?
- How to motivate community members for long-term participation?

The fourth derived aspect is to build user reputation so that it can be trusted in future. In this section, these aspects of SQA sites are under discussion.

Questions

An SQA site builds a large archive of questions and answers. This archive can be exploited to automatically find semantically similar questions and recommend them to the asker as a quick solution. In this context Jeon *et al.* (2005) used Naver's KiN-based data set to identify questions asked for the same problem but variations exist in the way of expression. Answer text for similar questions were analysed to create a probabilistic transition model and to find semantic similar words. This model was exploited to retrieve similar questions when a new question was asked. In terms of finding the semantically similar question, they found that their system outperformed.

Cao *et al.* (2009) proposed an effective framework for question retrieval by exploiting the category structure of Yahoo! Answers. They used local smoothing technique and hierarchical clustering to compute the probability of new questions belonging to existing category. This research work was further modified and improved by Cao *et al.*(2010) by considering the local and global relevance, i.e. query relevance to a question within a category and query relevance to a category, respectively. Similarly, Bernhard and Gurevych (2008) used the combination of similarity matrix and query processing strategies for similar question retrieval. They came out with the result that question answering accuracy exceeds 80 per cent for questions paraphrased from Wiki Answers collection.

People ask thousands of questions on a regular basis using SQA sites. There exist various types of questions. Roughly, these can be categorized as factual questions, procedural questions, opinion-oriented questions, task-oriented questions, and advice-seeking questions. In order to understand user behaviour and community outcomes, scientists have tried to explore different types of questions which will help the research community and practitioners to automate, and understand the nature and reason behind those questions being asked in online SQA sites.

Kateryna *et al.* (2009) tried to investigate the types of questions being asked on SQA sites such as Yahoo! Answers. They used the modified question annotation scheme developed by Graesser *et al.* (1994), i.e. concept compilation, definition, procedural, comparison, causal, disjunctive, verification, quantification, and general information need. For experimental purpose, they used the compiled data set of Yahoo! Answers consisting of 755 questions in the field of Data Mining, Natural Language Processing, and eLearning. The annotators were allowed to label a question with multiple tags in case of ambiguity. They came out with the

result that the frequency of Conceptual Completion is on peak consisting of 46.3 per cent of the questions where only 0.5 per cent of the questions are categorized as General Information Need and about one over five of the questions were lexically, syntactically, or semantically bad formatted.

Harper *et al.* (2009) tried to distinguish informational questions (user's aim is to learn rather than discuss) from conversational questions (user's aim is either to start discussion or get opinions) in terms of archive value, writing quality, and structural difference by exploiting different classifiers, i.e. human coding, machine learning algorithms, and statistical analysis. For research purpose, a data set was compiled based on the data being extracted from three popular SQA sites, i.e. Yahoo! Answers, Ask Metafilter, and Answerbag. They found that informational question carries dominancy over the conversational question in terms of archive value where the classification accuracy achieved by human coding and machine learning algorithms was 91.4 and 89.7 per cent, respectively. The same data set was used by Harper *et al.* (2010) to operationalize the questions in online SQA with the intention to produce a taxonomy of questions and then to study the properties and qualities of the question types being identified. A taxonomy of six types of questions were designed under Aristotle's three species of rhetoric (Aristotle, 2007), i.e. deliberative, epideictic, and forensic questions were further sub-categorized as advice and identification, approval/disapproval, quality, factual, and prescriptive, respectively. The experts agreed on 283 questions as the primary type out of 300 and disagreed on the remaining 17 questions. The archived value of advice, quality, prescriptive, and factual was found high as compared to (dis)approval and identification type of questions where advice questions were labelled as most personalized and the factual questions as most generic.

In SQA sites, a document is generated by the collaborative efforts of the community members and the first step is posting a question. If the quality of the posted question is high so is the final document, otherwise it is junk. In this background, Fong *et al.* (2015) came out with an idea that the question quality can be improved by providing guidelines on how to post a good quality question and to avoid the bad one. They stated that if the question is good (questions with accepted answer were called good questions) then there are high chances that it will get a positive response from the community members compared to a bad question (questions ignored by the community were called bad questions). To accomplish this task and to predict the quality of the question, they proposed a classifier which exploits the component of attributes being extracted by using factor analysis. The research objectives of Burel *et al.* (2015) were to extract patterns from the question selection behaviour of answerers and then to apply random forests-based learning to rank models to predict relevant questions for the answerer. During experimental analysis, they found that random forests provides promising results with mean reciprocal rank of 0.44. Jiarpakdee *et al.* (2016) built prediction models that calculate the chance of a question of getting no reply. They concluded that community-based features (answer badge count, question badge count, tag badge count, etc.) and affective features (politeness, positive and negative sentiments) are the most significant features in this regard. Similarly, Ahasanuzzaman *et al.* (2016) tried to predict duplicate questions in Stack Overflow[7] at the time of posting a new question. A discriminative model classifier along with BM25 scoring function was used and claimed that this proposed model outperformed DupPredictor in terms of the recall rate while using the data of Stack Overflow.

Answers

Submitted questions receive answers from the community members and enrich the document with each contribution. Different SQA sites use different social processing techniques to encourage the answerer and to enhance the quality of the received answers.

Such sites generate enormous amount of data anonymously and give birth to different research questions, i.e.:

RQ1. How to judge the quality of the answers received?

RQ2. How to measure the correctness and usefulness of answers?

RQ3. Content quality comparison between SQA sites and other mediums generated by experts.

RQ4. Effects of financial incentives, authority, response time, answer length and citations on the quality of answers.

RQ5. What can be the criteria for the best answer selection?

A number of scientific studies were conducted by scientists to answer these research questions. In the quality comparison context, researchers found that (Harper *et al.*, 2008; Shachaf, 2009; Shachaf and Rosenbaum, 2009) the answer quality of SQA sites can compete or even surpass the library reference service and experts services. During a comparative study Harper *et al.* (2008) reported that the quality of answers are proportional to financial incentives, i.e. the answer quality of fee-based Q&A sites (e.g. Google Answers[8]) are longer and are of good quality as compared to free SQA sites (Yahoo! Answers). Furthermore, they stated that the success of Yahoo! Answers lies in its huge community, responsiveness, and high answer diversity. Later on Jeon *et al.* (2010) re-examined the Harper *et al.* (2008) research and found that the financial incentive may affect the response rate and answer size but not necessarily the answer quality; however, the reputation of the answerer is influential in this regard. Some other studies have also shown that the reputation, answer length, and response time have positive effects on the answer quality (Li, He, Jeng, Goodwin and Zhang, 2015; John *et al.*, 2011; Chua and Banerjee, 2013). In another study Gazan (2006) studied the Answerbag community members' rating behaviour and categorized the community members as specialists and synthesists. Specialists are those who consider themselves as specialists and answer questions without references whereas synthesists are those who answer with references. He showed that the answers of synthesists are rated more highly as compared to answers of specialists. Similarly, Harper (Kateryna *et al.*, 2009) showed that the answers with references receive a high rating as compared to answer with no references.

Kim *et al.* (2007) made an attempt to find the best answer selection criteria mentioned in comments by the askers while choosing an answer as best answer on Yahoo! Answers site. Only 465 comments justified the best answer selection criteria out of 1,200 comments being collected for research purpose. They designed a framework of best selection criteria consisting of seven categories, i.e. content, cognitive, socio-emotional, extrinsic, information source, utility, and General statement. Adaji and Vassileva (2016) studied the commenting and editing behaviour of users in Stack Overflow and noticed that the best answer receives more comments compared to other answers. They argued that such behaviour needs to be encouraged through rewards as it can have a positive effect on the quality of the answer. Jeon and Rieh (2014) conducted a quasi-field study while using Yahoo! Answers and observed that attitude, trustworthiness, and expertise tend to be considered as positive indications in credibility assessment of the answer.

A multi-dimension model has been developed, analysed, and evaluated by Zhu *et al.* (2009) to assess answer quality of SQA sites for eLearning. Natural Language Processing techniques were used to automate the process. In their quality model, they considered 13 dimensions including informativeness, politeness, completeness, readability, relevance, conciseness, truthfulness, level of detail, originality, objectivity, novelty, usefulness, and expertise. Additionally, Li, Zhu, Lu, Ding and Gu, (2015) examined the wiki

functionality of Stack Overflow with the intention to find its effects on the content quality and users' motivation. They found the collaborative editing feature beneficial and suggested that it needs to be incorporated in other SQA sites as well. Song *et al.* (2014b) presented a theoretical model of relationship between the users' motivation and user-generated contents while considering social exchange and achievement need theories. For research purpose, they used an SQA site Baidu iKnow and interviewed 12 active participants. During their study, they claimed that achievement need, power need, affiliation need, and extrinsic reward are in positive correlation where input cost is in negative correlation with SQA contents quality and quantity. Deng *et al.* (2015) researched that poor answer quality and low participation affect users' satisfaction.

User motivation, trust, and reputation

People are the soul of SQA sites and active community members play a vital role in the success of SQA sites (Welser *et al.*, 2007; Harper *et al.*, 2008; Shah *et al.*, 2008). Users in SQA sites share information for free. A cost in terms of time and effort is associated with such information. To reward long-term and frequent users for participation, each SQA site has its own incentive model. Research showed that the information quality of SQA sites is proportional to the incentive model (Jain *et al.*, 2009; Li, He, Jeng, Goodwin and Zhang, 2015; Wei *et al.*, 2015). In a broader sense, there are two types of motivation that motivate the user to participate in SQA sites, i.e. intrinsic and extrinsic (Raban and Harper, 2008).

Intrinsic motivation comes from within an individual, e.g. self-satisfaction and pleasure. People invest their time and make efforts to answer others questions to gain self-satisfaction, pleasure, sense of reciprocity for having returned a favour, and feelings of gratitude and respect (Raban and Harper, 2008). Beyond these motivations, SQA systems provide extrinsic motivation. Extrinsic motivation is derived from outside world, e.g. money, grades, marks, points, social position, etc. SQA sites provide extrinsic motivation to bring content quality and attract community members. Different approaches are used to extrinsically motivate community members for long-term participation, such as awarding points, social comparison (Harper *et al.*, 2007), level titles, money, and granting site privileges. Some SQA sites provide extrinsic motivation in the form of money, e.g. Mahalo answer (Hsieh *et al.*, 2010b) and buy answer (Wenyin *et al.*, 2009), whereas others provide incentives in the shape of points like a game, e.g. Yahoo answers, Naver's KiN, etc. Previous studies showed that rewarding financial incentives improved answer quality (Harper *et al.*, 2008). Answers on pay-based Q&A sites are longer, have higher archive value, and are of good quality as compared to free SQA sites (Harper *et al.*, 2008; Janes *et al.*, 2001; Hsieh and Counts, 2009). The researcher stated that financial incentive model works well for the factual questions but is not feasible for discussion or conversational type of questions (Hsieh *et al.*, 2010a). Farzan *et al.* (2008) showed that point-based incentive system is the reason behind consistent contribution from community members. A number of factors which motivate community members to participate in online virtual communities are altruism, belonging, collaboration, egoism, egotism, emotional support, empathy, knowledge, power, reciprocity, reputation, self-esteem, self-expression, and wisdom (Moore and Serva, 2007).

Search engines use authority for trust whereas SQA sites use intimacy, reputation, and expertise. Reputation can be defined as the beliefs or opinions about a person or thing (Chen *et al.*, 2007). In the context of SQA sites, trusting someone is a commitment to an action based on a belief that the future action of a community member will lead to a good result (Golbeck and Hendler, 2006). Reputation and expertise building is a significant indication for participants' decision-making in activities on SQA and social networking

sites. For example, 12 per cent employers in the USA use popular social networks to investigate potential employees (Zhou *et al.*, 2008). Offline we trust because of real-world reputation but in SQA sites interaction take place anonymously. So the question arises whether the solution from a stranger is reliable or not? In SQA sites, systems are emerging that respect anonymity and build community members' reputation. Long-term participants build reputation and create an incentive for good behaviours. Resnick *et al.* (2000) stated three properties of reputation system, i.e. a long-term interaction of the entity, current interaction must be rated and should be visible in the future, and current interaction with this entity is dependent on its past reputation. The main theme of reputation is to increase trust in SQA environment because these systems are built on user-generated contents. It has been studied that the reputation is relative to the contents' quality (Baltadzhieva and Chrupała, 2015) and also positively correlated to the contribution of a diverse number of tags (MacLeod, 2014). Approaches to decide the level of user expertise in SQA sites can be classified into two types: link analysis and point-based approach.

A community member's reputation can be boosted or inferred by the reputation of other connected users. Link analysis employs algorithms like PageRank (Page *et al.*, 1999) and Hyperlink-Induced Topic Search (HITS) (Kleinberg, 1999) or their modified versions for user reputation and experts' ranking. Reputation computation by using linked analysis approach is also referred as flow or propagation model (Lee *et al.*, 2013; Josang *et al.*, 2007) because reputation is computed by transitive iteration and the rank value flow through looped/arbitrarily long chain. The HITS algorithm (Chen and Nayak, 2008; Jurczyk and Agichtein, 2007a, b) divides users into hub and authority user groups that ask questions and respond/answer, respectively. Hub and authority values are recursively calculated which converge after several iterations. PageRank ranks a user based on the number of users that (s)he helps (Jiao *et al.*, 2009; Zhang *et al.*, 2007). Link analysis works well if users in a Q&A site behave properly but actions like answering as many questions as one can without worrying about the quality of the answers, are common. It is reported that 1/3 of the given answers have some sort of quality problems and 1/10 of these given answers are bad (Jeon *et al.*, 2006).

Jurczyk and Agichtein (2007a, b) used HITS algorithm for finding the reputation of the user by exploiting the linked structure of Yahoo! Answers community. They performed some analysis and found that HITS works well for some categories but not for all. Zhang *et al.* (2007) used a set of network-based algorithms such as PageRank and HITS for user reputation to analyse the interaction between asker and answerer in Java community. They reported that 54.9 per cent are askers, 13 per cent are answerers, and 12.3 per cent are both. Jiao *et al.* (2009) used the modified version of PageRank algorithm called ExpertRank for user reputation and used Microsoft discussion group for evaluation. Expert profiles were used to calculate the relevance score for the reflection of dynamic relevance between candidate experts and user input query while considering the spam experts. They calculated user reputation and were successful in filtering spam experts. Zhou *et al.* (2012) used the modified version of linked analysis algorithms for finding experts while considering the topic link structure and topic similarity among community members. They performed experiments on Yahoo! Answers data set and found their proposed algorithm more effective. Yang *et al.* (2013) proposed a multipurpose general model called Topic Expertise Model for an expert by combining topic interest with expertise evaluation. Li *et al.* proposed topic-level expert learning model by combining graph-based link analysis and content semantic analysis for expert modelling in community Q&A. Liu *et al.* (2013) considered user subject relevance, user reputation, and authority of a category in finding experts. Chen *et al.* (2011) proposed a bias-smoothed tensor model for user reputation in comment rating environment. During their study, they showed that the proposed model outperformed. Furthermore, Hong *et al.* (2009) did a

comparative study on PageRank- and HITS-like schemes for modelling user reputation on different topics in Yahoo! Answers. Their results showed that the PageRank-based approaches dominate HITS schemes. Song *et al.* (2014a) tried to find the leading user in user-centric community Q&A and proposed a leading ability detecting model. They used multiple kernel learning algorithms on the basis of leading capacity and showed the effectiveness of the proposed method. Yang and Manandhar (2014) also tried to explore experts and proposed a probabilistic model while considering the topic-specific abilities of the community members and the social links structure within SQA sites. For experimental purpose, they considered the data set of Stack Overflow and turned their proposed model as a candidate solution for investigating experts in SQA sites.

The points-based user reputation model allows users to ask and answer questions and rate each other. A reputation computation engine computes user reputation (reputation score) based on user ratings and earned points. Naver's KiN, Yahoo! Answers, Answerbag, and Stack Overflow are some of the best examples of an accumulative reputation display pattern that measures the reputation of community members using points. Reputation scores increase or decrease either monotonically or arbitrarily according to points. Naver's KiN uses a points-based approach where ten points are awarded for an answer, 25 points for the best answer, one point for voting, and three points for logging in every day (Nam *et al.*, 2009). In a similar fashion, Yahoo! Answers grants 100 points to users for creating an account, one point for each login, five points for asking a question, ten points for selecting an answer as the best answer, and three points for selecting the best answer for his/her question (Gyongyi *et al.*, 2007). Stack Overflow is another points-based SQA system for computer programmers, where reputation is achieved by convincing the community members that they have subject matter expertise. Stack Overflow grants privileges as the user's reputation increases.

There are variations of the user reputation model. Chen *et al.*(2007) developed a user reputation model for CuteAid, which relies on a combination of social network analysis (SNA) and user ratings while considering different relationships including asking and answering questions, identifying correct answers, supporting and complaining about answers, and commenting on questions. McNally *et al.* (2013) proposed a collaborative user reputation model for social web and considered the aggregate vote greater than zero by using three different SQA systems on stock exchange network[9]. Similarly, an incentive model for the IBM social networking site has been developed to encourage community member participation (Farzan, Dimicco, Millen, Dugan, Geyer and Brownholtz, 2008; Farzan, Dimicco, Millen, Brownholtz, Geyer and Dugan, 2008). It is noteworthy that points-based SQA systems are very popular with users because they encourage prompt responses (Vasilescu *et al.*, 2014). If the answerer earns badges or rewards for answering questions, he/she is consequently more active while answering more questions (Cavusoglu *et al.*, 2015). Points-based user reputation models are more prevalent than link analysis approaches because they are accurate, bring quality to the contents, and differentiate novice from experts, because of their collaborative filtering approach, and because they attract community members for long-term active participation.

State of the art

Figure 3 shows popular SQA sites and their launching dates. Jimmy Wales and Larry Sanger launched Wikipedia on 15 January 2001. Wikipedia provides two types of services: Wikipedia free encyclopaedia and Wikipedia reference desk (SQA services launched in 2001). Wikipedia reference[10] desk provides an online Q&A services. Users ask the question and Wikipedia volunteers provide the solution. Wikipedia is organized in categories having seven top-level categories, i.e. computer and IT, entertainment, humanities, languages, mathematics, science, and miscellaneous.

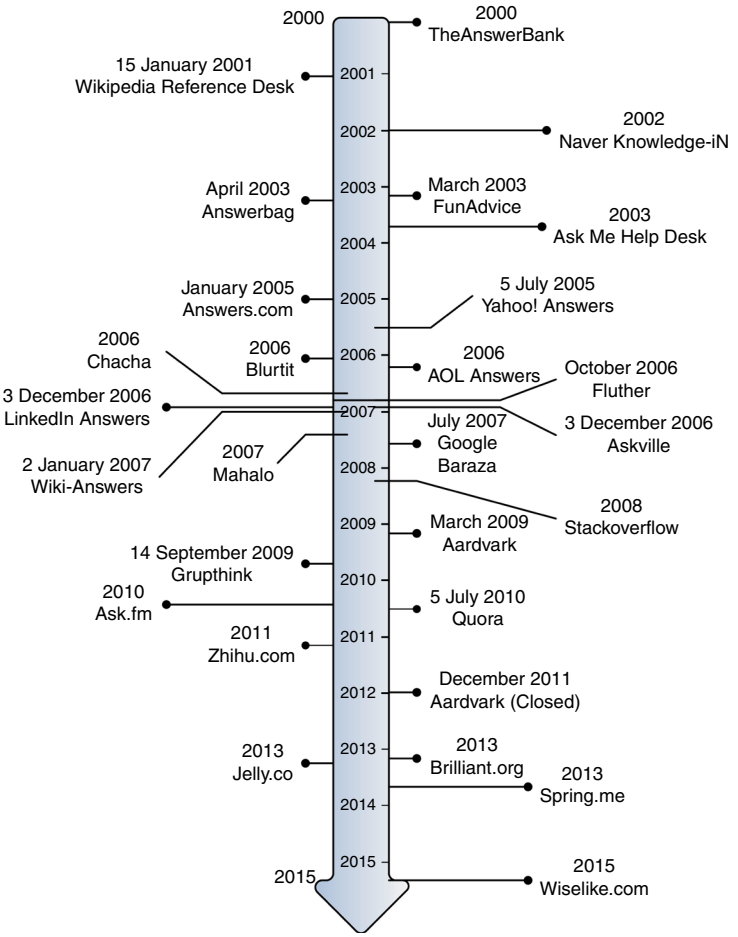

Figure 3.
Launch dates of popular SQA sites

The reliability level of Wikipedia reference desk in term of accuracy and verifiability is better than that of Wiki Answers, Yahoo! Answers, and Askville. The response rate of Wikipedia help desk is 96 per cent whereas the response rates of Wiki Answers, Yahoo! Answers, and Askville are 16, 85 and 74 per cent, respectively (Shachaf and Rosenbaum, 2009). Even the quality of Wikipedia reference desk competes with traditional reference service which is run by professionals and experts (Shachaf and Rosenbaum, 2009).

Naver's KiN is a large general-purpose question answering community, started in South Korea. Naver[11] is the most popular search engine and question answer community site in South Korea, launched by NHN Corporation in 2002 by ex-Samsung employees (Jeon *et al.*, 2005). At that time, most of the documents, web pages, and information were available in English. It was difficult to build a Korean search engine. KiN started a question answer community where the user can ask and answer many questions, thus enabling users to research the KiN database (Nam *et al.*, 2009). For user attraction and motivation, KiN displays the most active and top answerers on the main page. In KiN, a registered user is allowed either to ask question or answer. Unlike discussion

forums, only one answer can be given to a question. KiN's community users can interact with other members directly via e-mail or messages. KiN's community members post 44,000 questions per day and get 110,000 answers from about 4.5 million regular visitors on average (Nam *et al.*, 2009). KiN consist of 14 top-level categories and over 3,500 subcategories. KiN extrinsically motivates the community members, i.e. based on point system.

Wiki Answers[12] is a platform where registered members can post questions and can answer other questions. Wiki Answers was launched as FAQ Farm in 2002 by Chris Whitten. Answer Corporation bought FAQ Farm in November 2006 for two million US dollars. In January 2007, it was renamed as Wiki Answers. Unlike Yahoo! Answers and other SQA sites where answers are voted and there is no way to improve existing contents, Wiki Answers allows its user to edit existing Q&A to improve the quality of contacts with the passage of time by using wiki technology. Under the category supervisor, duplicated questions are identified, merged, and generalized with collaborative efforts (Seth, 2009). Wiki Answers provides extrinsic motivation in the form of points and power to control the site. Trusted and providing "super power" tool are almost similar to the webmaster. Wiki Answers has over 7,000 categories where user can post and get the answer from the community.

Answerbag is a free SQA site launched in April 2003 by Joel Downs as a collaborative FAQ online database based on one-question-one-answer model. Answerbag was acquired by Infosearch Media in early 2006. Registered users can ask a question from the Answerbag community and can get multiple answers in more than 4,000 categories. Unlike Yahoo! Answers, there is no method for selecting a best answer in Answerbag. Another difference is that one can attach a picture or a video along with Q&A. Answerbag users are able to rate answers and questions. Answers are listed in descending order by rating. Answerbag had more than 7 million unique visitors/month, 40,000 of them were registered users, in June 2008 (Gazan, 2009). Answerbag uses point-based system and users are awarded points; as the points increase the user levels up.

Yahoo started Yahoo! Answers at the end of 2005. Total number of Yahoo! Answers registered users are 200 million (Harper *et al.*, 2010) and 15 million users visit per day (Harper *et al.*, 2008). Yahoo is also an SQA site. In order to participate, a user has to create an account or can use existing Yahoo or Facebook[13] account to login. Users are allowed to post a question in detail along with question heading under a relative category to which the question belongs. Other community volunteers are allowed to answer. The asked question is open for 4 days and is extensible for another 4 days. During this period, other community members can answer along with optional references. After getting satisfactory answers from the community members, the asker is allowed to choose the best answer. Resolved questions are stored permanently in Yahoo! Answers archive, and otherwise deleted. After the question is resolved other community members are allowed either to comment or vote. One can search or browse Yahoo! Answers categorywise (25 top-level categories). Yahoo! Answers has organized topics in hierarchal order having two to three level depth (Gyongyi *et al.*, 2007). Yahoo! Answers' incentive model is based on points. Users with maximum points make it into the "leader board", possibly gaining respect within the community. Yahoo! Answers is a good platform for posting questions and getting quick, satisfactory, and effective answers from the community (Shah, 2011). Kateryna *et al.* (2009) showed that in Yahoo! Answers, 1/5 of the questions are poorly formatted. Shen *et al.* (2015) studied the hidden features of Yahoo! Answers and found that about ten per cent of the community members are posting good quality answers.

SQA communities like Ask Me Help Desk, Help, AOL Answer, Quara[14], Google Barasza[15], TheAnswerBank[16], Baidu Zhidao[17], Stackexchange, and many more are trying to contribute to this domain.

Analysis and comparison of the state-of-the-art SQA sites

After analysing five popular SQA sites, i.e. Naver's KiN, Wiki Answers, Answerbag, Yahoo! Answers, and Stack Overflow, we identified 24 design features. These design features have been categorized under architectural components as shown in Figure 2. In Table I, ✓ and × mark means the presence and absence of a particular design feature, respectively, in the corresponding SQA.

Conclusion and future recommendations

In this paper, the existing SQA systems were extensively surveyed and their importance was explored from the perspective of research and practitioners. The review contributes a comprehensive knowledge about SQA systems and aims to be helpful to new researchers who want to get a broad picture of SQA systems on a single platform. SQA systems are in their infancy and require a huge effort to be invested by the researchers, academia, and organizations. As they mature, SQA sites will release online information seekers from keyword constraints and will empower them to grab relevant material simply by asking questions in natural language. Undoubtedly, the area is growing and attracting an increasing number of researchers day by day. By analysing a number of existing systems,

Architectural components	Design features	SQA site KiN	WA	AB	Y!A	QR	SO
User activities	Registration required for participation	✓	✓	✓	✓	✓	✓
	Question can be edited by other community members	×	✓	×	×	✓	✓
	Answer can be edited by other community members	×	✓	×	×	✓	✓
	Category-based management system	✓	✓	✓	✓	×	×
	Tag-based management system	×	×	×	×	✓	✓
	Linking profile with other social networks (FB, etc.)	✓	✓	✓	✓	✓	✓
	Contents sharing	×	✓	✓	✓	✓	×
	Community tagging	✓		×	×	✓	✓
	Answers restriction per question for a community member	×	✓	×	✓	✓	×
	Notification system	✓	✓	✓	✓	✓	✓
	Question time restriction	×		×	✓	×	×
	Private Messaging	✓	✓	✓	✓	✓	×
Social feedback	Questions can be rated by other community members	✓	×	×	×	✓	✓
	Badges	×	✓	✓	×	×	✓
	Answers can be rated by other community members	✓	✓	✓	✓	✓	✓
	Points on asking question	✓	×	×	✓	×	×
	Points on answering	✓	✓	✓	✓	×	✓
	Report abusive behaviour	✓	✓	✓	✓	✓	✓
Profile	User personal information	✓	✓	✓	✓	✓	✓
	User current level of expertise/reputation	✓	✓	✓	✓	×	✓
	History of question and answer activities	✓	✓	✓	✓	✓	✓
	Points achieved	✓	✓	✓	✓	×	✓
	Privileges granted on contribution	×	✓	×	×	×	✓
	Leaderboard	✓	×	✓	✓	×	✓

Notes: KiN, Naver's Knowledge-iN; WA, Wiki Answers; AB, Answerbag; Y!A, Yahoo! Answers; SO, Stack Overflow; QR, Quora

Table I. Popular SQA sites with its respective features

a number of challenges were identified which need to be addressed immediately by the research community. Some of the promising future research directions are as follows:

- Qualitative and quantitative research methods: the fact that participants provide their highly valued knowledge on SQA sites results in the development of a collective wisdom highlighting the new avenues of research. So, to determine the increasing reliability and credibility of these systems as well as of the participants, qualitative and quantitative research methods encompassing the area are sharply needed to be researched. These methods should find answers and increase scholars' understandability about a number of pertinent questions which include the following: what attracts folks to participate? Why do they choose SQA sites to get suggestions from anonymous and unverified experts? Why do people invest their time to share information for free? How to produce high quality contents and to attract more folks for participation? Qualitative and quantitative research is required to find the answers to these questions.
- User reputation models for SQA: current state-of-the-art SQA sites use different aggregative and propagative models for user reputation calculation and content quality evaluation. The aim is to attract the users for long-term active participation and to differentiate novices from experts. Further refinement and investigation are required to improve this area and to calculate the reputation of users with more accuracy. The more refined version of the virtual user reputation model can be combined with real-life academic or professional achievements to benefit users online as well as offline.
- Development tool kits: the success and popularity of an area depend on the rate of software development in that domain. Developing software from scratch normally requires a greater amount of effort, potential, cost, time, etc., which would not be usually available. Therefore, highly specialized development tool kits are needed to be researched to help developers in developing highly attractive, reliable, and accurate SQA systems in a short span of time.
- Software agents and simulators: scientists have tried to study user behaviour and the reason why the community members are attracted to SQA sites. However, an enormous amount of research is still required to study the hidden dynamics of SQA sites, such as the effects of incentive models while considering different factors, the patterns of interaction, studying community members' behaviour, social influence locality, and much more. The scientific community has borrowed software agents and simulators from other domains for the testing their models and evaluations. The kind of accuracy and validity of such agents and simulators in the SQA environment is not clear. However, an attempt has been made by Aumayr and Hayes (2014) who have proposed an agent-based model, but still further research and refinement is the demand of this domain.
- Application of semantic web technologies in SQA systems: in the state-of-the-art SQA systems, people suffer from a significant amount of cognitive overload in evaluating contents quality and determining users' reputation, etc. Semantic web technologies are deemed as operational tools for converting information into machine-processable format, in order to reduce cognitive overload. Thus, applying already matured semantic web technologies in designing advanced SQA systems can significantly reduce the cognitive overload of people. Similarly, domain ontologies can be built either from the scratch or by using upper-level ontologies.
- Pluggable architecture for SQA sites: today most of the valuable systems (e.g. Facebook, etc.) use pluggable architectures, enhancing/increasing their functionalities by allowing

third-party developers to easily develop and integrate plug-ins. Therefore, for complete and widely applicable SQA systems, generous and pluggable architectures are needed to be researched. Thus, it may allow a wide range of developers and researchers belonging to diverse fields to add functionalities regarding their fields and may result in the increasing spectrum of these systems.

- Promoting community-based information seeking in educational and organizations environment: state-of-the-art SQA sites are suitable for factual/exploratory questions and are not appropriate for the educational and organizational environment. Liu *et al.* showed that the integration of SQA sites with learning management system can be advantageous and will help the students to share and manage knowledge (Yang *et al.*, 2013). Similarly, Ortbach *et al.* (2014) identified additional design principles for SQA sites that account for their application in the organizational background. We suggest that further research is required to investigate the design principals of SQA sites in the context of educational environment and organizations and to answer research questions that include the following: what kind of motivation model is more beneficial to motivate the community members for high-quality content creations (e.g., virtual point-based user reputation model or real-world reward system), what can be the effects of anonymity of users?
- Exploiting SQA archive for automatically answering questions: users' participation in SQA sites is increasingly data-driven. Exploiting such data will provide better services to the web community in general. Automatic Q&A are instantly beneficial to a wide spectrum of end users. However, further research is required to analyse the questions' text, find relevant and suitable answer(s), and present it to the end user.
- Social networks and social information seeking: key intentions of SQA sites are to use the efforts of the active community members to find the solutions to questions with collaborative efforts. Too many question recommendations or searching through a gigantic pool of questions can increase cognitive overload on the answerers, e.g. in Yahoo! Answers. Online social networks such as Facebook, LinkedIn, and many more, have active members in millions. These social networks can be used as a base layer for social information seeking, as shown in Figure 4, where relevant reputation model can be developed to encourage the users for active participation and to build an online reputation. In such a case, deferent factors, i.e. social ties, intimate relationship, real-world expertise and trust, active participation, SNA, and user ratings can be considered to build an effective user reputation model. Furthermore, algorithms can be designed to recommend a question to relevant expert(s).

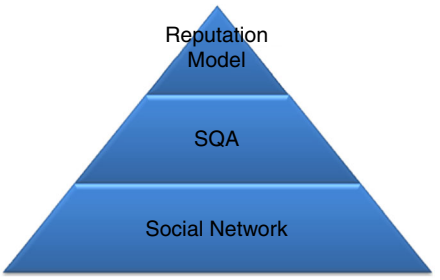

Figure 4.
Social networks can be used as a base layer for social information seeking

Notes

1. www.answers.yahoo.com
2. www.answerbag.com
3. www.allexperts.com
4. www.justanswer.com
5. http://kin.naver.com
6. http://trec.nist.gov/overview.html
7. www.stackoverflow.com
8. www.answers.google.com – Fee-based Q&A site lunched in 2002 and discontinued in 2006.
9. www.stackexchange.com/
10. www.en.wikipedia.org/wiki/Wikipedia:Reference_desk
11. www.naver.com
12. www.wiki.answers.com
13. www.facebook.com
14. www.quora.com
15. www.google.com/baraza/en
16. www.theanswerbank.co.uk
17. http://zhidao.baidu.com

References

Adaji, I. and Vassileva, J. (2016), "Modelling user collaboration in social networks using edits and comments", in Cena, F., Desmarais, M., Dicheva, D. and Zhang, J. (Eds), *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, *Halifax, Nova Scotia, Canada*, *13-17 July*, ACM, New York, NY, pp. 111-114.

Ahasanuzzaman, M., Asaduzzaman, M., Roy, C.K. and Schneider, K.A. (2016), "Mining duplicate questions in Stack Overflow", in Kim, M., Robbes, R. and Bird, C. (Eds), *Proceedings of the 13th International Conference on Mining Software Repositories*, New York, NY, pp. 402-412.

Aumayr, E. and Hayes, C. (2014), "Modelling user behaviour in online Q&A communities for customer support", in Hepp, M. and Hoffner, Y. (Eds), *International Conference on Electronic Commerce and Web Technologies, Vol. 188, Munich, 1-4 September*, Springer, Cham, pp. 179-191, availbale at: www.insight-centre.org/content/modelling-user-behaviour-online-qa-communities-customer-support

Baltadzhieva, A. and Chrupała, G. (2015), "Predicting the quality of questions on stackoverflow", Recent Advances in Natural Language Processing, Hissar.

Bernhard, D. and Gurevych, I. (2008), "Answering learners' questions by retrieving question paraphrases from social Q&A sites", *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications, Association for Computational Linguistics*, *Columbus, OH*, pp. 44-52.

Boyd, D.M. and Ellison, N.B. (2007), "Social network sites: definition, history, and scholarship", *Journal of Computer-Mediated Communication*, Vol. 13 No. 1, pp. 210-230.

Burel, G., Mulholland, P., He, Y. and Alani, H. (2015), "Modelling question selection behaviour in online communities", in Gangemi, A., Leonardi, S. and Panconesi, A. (Eds), *Proceedings of the 24th International Conference on World Wide Web*, *Florence*, ACM, New York, NY, pp. 357-358.

Cao, X., Cong, G., Cui, B. and Jensen, C.S. (2009), "The use of categorization information in language models for question retrieval", in Cheung, D.W.-L., Song, I.-Y., Chu, W.W., Hu, X. and Lin, J.J. (Eds), *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, *Hong Kong*, ACM, New York, NY, pp. 265-274.

Cao, X., Cong, G., Cui, B. and Jensen, C.S. (2010), "A generalized framework of exploring category information for question retrieval in community question answer archives", in Rappa, M., Jones, P., Freire, J. and Chakrabarti, S. (Eds) *Proceedings of the 19th International Conference on World Wide Web*, *Raleigh, NC*, ACM, New York, NY, pp. 201-210.

Cavusoglu, H., Li, Z. and Huang, K.-W. (2015), "Can gamification motivate voluntary contributions?: The case of stackoverflow Q&A community", *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*, ACM, Vancouver.

Chen, B.-C., Guo, J., Tseng, B. and Yang, J. (2011), "User reputation in a comment rating environment", in Apté, C. Ghosh, J. and Smyth, P. (Eds), *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *San Diego, CA*, *21-24 August*, ACM, New York, NY, pp. 159-167.

Chen, L. and Nayak, R. (2008), "Expertise analysis in a question answer portal for author ranking", *Web Intelligence and Intelligent Agent Technology: IEEE Computer Society*, *Sydney*.

Chen, W., Zeng, Q. and Wenyin, L. (2007), "A user reputation model for a user-interactive question answering system", *Concurrency and Computation: Practice and Experience*, Vol. 19 No. 15, pp. 2091-2103.

Chua, A.Y. and Banerjee, S. (2013), "So fast so good: an analysis of answer quality and answer speed in community question-answering sites", *Journal of the American Society for Information Science and Technology*, Vol. 64 No. 10, pp. 2058-2068.

Deng, S., Fang, Y., Liu, Y. and Li, H. (2015), "Understanding the factors influencing user experience of social question and answer services", *Information Research: An International Electronic Journal*, Vol. 20 No. 4, available at: www.informationr.net/ir/20-4/paper694.html#.WS2dLmiGOUl (accessed 30 May 2017).

Farzan, R., Dimicco, J.M., Millen, D.R., Brownholtz, B., Geyer, W. and Dugan, C. (2008), "When the experiment is over: deploying an incentive system to all the users", in Masthoff, J., Reed, C. and Grasso, F. (Eds), *Proceedings of the Symposium on Persuasive Technology, in Conjunction with the AISB*, *Aberdeen*, *April*.

Farzan, R., Dimicco, J.M., Millen, D.R., Dugan, C., Geyer, W. and Brownholtz, E.A. (2008), "Results from deploying a participation incentive mechanism within the enterprise", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Florence, pp. 563-572.

Fong, S., Zhuang, Y., Liu, K. and Zhou, S. (2015), "Classifying forum questions using PCA and machine learning for improving online CQA", in Berry, M.W., Mohamed, A. and Yap, B.W. (Eds), *International Conference on Soft Computing in Data Science*, *Putrajaya*, *2-3 September*, Springer, pp. 13-22.

Gazan, R. (2006), "Specialists and synthesists in a question answering community", in Grove, A. (Ed.), *Proceedings of the American Society for Information Science and Technology*, Wiley Subscription Services, Inc., Austin, TX, 3-8 November, pp. 1-10.

Gazan, R. (2009), "When online communities become self-aware", *System Sciences, HICSS'09, 42nd Hawaii International Conference on System Sciences (HICSS-42)*, *Waikoloa, HI*, *5-8 January*, IEEE Computer Society, Washington, DC, pp. 1-10.

Gazan, R. (2011), "Social Q&A", *Journal of the American Society for Information Science and Technology*, Vol. 62 No. 12, pp. 2301-2312.

Golbeck, J. and Hendler, J. (2006), "Inferring binary trust relationships in web-based social networks", *ACM Transactions on Internet Technology*, Vol. 6 No. 4, pp. 497-529.

Graesser, A.C., Mcmahen, C.L. and Johnson, B.K. (1994), "Question asking and answering", in Ann, G.M. (Ed.), *Handbook of Psycholinguistics*, Academic Press, San Diego, CA, pp. 517-538.

Gyongyi, Z., Koutrika, G., Pedersen, J. and Garcia-Molina, H. (2007), "Questioning Yahoo! Answers", technical report, Stanford InfoLab, available at: http://ilpubs.stanford.edu:8090/819/ (accessed 30 May 2017).

Harper, F.M., Li, S.X., Chen, Y. and Konstan, J.A. (2007), "Social comparisons to motivate contributions to an online community", in de Kort, Y., Ijsselsteijn, W., Midden, C., Eggen, B. and Fogg, B.J. (Eds), *Proceedings of the 2nd International Conference on Persuasive Technology*, Springer-Verlag, Palo Alto, CA, pp. 148-159.

Harper, F.M., Weinberg, J., Logie, J. and Konstan, J.A. (2010), "Question types in social Q&A sites", *First Monday*, Vol. 15 No. 7, available at: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2913/2571 (accessed 20 May 2017).

Harper, M., Moy, D. and Konstan, J. (2009), "Facts or friends? Distinguishing informational and conversational questions in social Q&A sites", in Olsen, D.R., Arthur, R.B., Hinckley, K., Morris, M.R., Hudson, S.E. and Greenberg, S. (Eds), *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, *Boston, MA*, *4-9 April*, ACM, New York, NY, pp. 759-768.

Harper, M., Raban, D., Rafaeli, S. and Konstan, J. (2008), "Predictors of answer quality in online Q&A sites", in Czerwinski, M., Lund, A.M. and Tan, D.S. (Eds), *Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, ACM, Florence, pp. 865-874.

Hong, L., Yang, Z. and Davison, B.D. (2009), "Incorporating participant reputation in community-driven question answering systems", *International Conference on Computational Science and Engineering CSE'09*, *IEEE Vancouver, IEEE Computer Society, Washington, DC*, *29-31 August*, pp. 475-480, available at: http://ieeexplore.ieee.org/document/5284222/

Horowitz, D. and Kamvar, S. (2010), "The anatomy of a large-scale social search engine", in Rappa, M., Jones, P., Freire, J. and Chakrabarti, S. (Eds), *Proceedings of the 19th International Conference on World Wide Web*, *Raleigh, NC*, *26-30 April*, ACM, New York, NY, pp. 431-440.

Hsieh, G. and Counts, S. (2009), "A market-based real-time question and answer service", in Olsen, D.R., Arthur, R.B. Jr, Hinckley, K., Morris, M.R., Hudson, S.E. and Greenberg, S. (Eds), *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, *Boston, MA*, *4-9 April*, ACM, New York, NY, pp. 769-778.

Hsieh, G., Kraut, R. and Hudson, S. (2010a), "Why pay? Exploring how financial incentives are used for question & answer", in Mynatt, E.D., Schoner, D., Fitzpatrick, G., Hudos, S.E., Edwards, W.K. and Rodden, T. (Eds), *Proceedings of the 28th International Conference on Human Factors in Computing Systems*, ACM, New York, NY and Atlanta, GA, pp. 305-314.

Hsieh, G., Kraut, R.E. and Hudson, S.E. (2010b), "Why pay?: Exploring how financial incentives are used for question & answer", in Mynatt, E.D., Schoner, D., Fitzpatrick, G., Hudos, S.E., Edwards, W.K. and Rodden, T. (Eds), *Proceedings of the 28th International Conference on Human Factors in Computing Systems*, *Atlanta, GA*, *10-15 April*, ACM, New York, NY, pp. 305-314.

Jacob, L. (2007), "A librarian's worst nightmare, Yahoo! Answers, where 120 million users can be wrong", *Slate Magazine*, December, available at: www.slate.com/articles/technology/technology/2007/12/a_librarians_worst_nightmare.html (accessed 30 May 2007).

Jain, S., Chen, Y. and Parkes, D. (2009), "Designing incentives for online question and answer forums", in Chuang, J., Fortnow, L. and Pu, P. (Eds), *Proceedings of the 10th ACM Conference on Electronic Commerce*, *Stanford, CA*, *6-10 July*, ACM, New York, NY, pp. 129-138.

Janes, J., Hill, C. and Rolfe, A. (2001), "Ask-an-expert services analysis", *Journal of the American Society for Information Science and Technology*, Vol. 52 No. 13, pp. 1106-1121.

Jenna, W. (2007), "After 10 years of blogs, the future's brighter than ever", *Wired Magazine*, December, available at: www.wired.com/2007/12/after-10-years-of-blogs-the-futures-brighter-than-ever/ (accessed 30 May 2017).

Jeon, G.Y. and Rieh, S.Y. (2014), "Answers from the crowd: how credible are strangers in social Q&A?", *iConference 2014 Proceedings, iSchools*.

Jeon, G.Y., Kim, Y.-M. and Chen, Y. (2010), "Re-examining price as a predictor of answer quality in an online Q&A site", in Mynatt, E.D., Schoner, D., Fitzpatrick, G., Hudson, S.E., Edwards, W.K. and Rodden, T. (Eds) *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, *Atlanta, GA*, *10-15 April*, ACM, New York, NY, pp. 325-328.

Jeon, J., Croft, W.B. and Lee, J.H. (2005), "Finding similar questions in large question and answer archives", in Herzog, O., Schek, H.-J., Fuhr, N., Chowdhury, A. and Teiken, W. (Eds), *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, *Bremen, Germany*, *31 October-5 November*, ACM, New York, NY, pp. 84-90.

Jeon, J., Croft, W.B., Lee, J.H. and Park, S. (2006), "A framework to predict the quality of answers with non-textual features", in Efthimiadis, E.N., Dumais, S.T., Hawking, D. and Järvelin, K. (Eds), *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, *Seattle, WA*, ACM, New York, NY, pp. 228-235.

Jiao, J., Yan, J., Zhao, H. and Fan, W. (2009), "ExpertRank: an expert user ranking algorithm in online communities", *International Conference on New Trends in Information and Service Science, NISS '09*, *Beijing*, *30-2 July*, pp. 674-679.

Jiarpakdee, J., Ihara, A. and Matsumoto, K.-I. (2016), "Understanding question quality through affective aspect in Q&A site", *Proceedings of the 1st International Workshop on Emotion Awareness in Software Engineering*, ACM, *Austin, TX*, pp. 12-17, available at: http://dl.acm.org/citation.cfm?id=2897000

John, B.M., Chua, A.Y.-K. and Goh, D.H.-L. (2011), "What makes a high-quality user-generated answer?", *IEEE Internet Computing*, Vol. 15 No. 1, pp. 66-71.

Josang, A., Ismail, R. and Boyd, C. (2007), "A survey of trust and reputation systems for online service provision", *Decision Support Systems*, Vol. 43 No. 2, pp. 618-644.

Jurczyk, P. and Agichtein, E. (2007a), "Discovering authorities in question answer communities by using link analysis", in Silva, M.J., Laender, A.H.F., McGuinness, D.L., Olstad, B., Olsen, Ø.H. and Falco, A.O. (Eds), *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management*, *Lisbon*, *6-10 November*, ACM, New York, NY, pp. 919-922.

Jurczyk, P. and Agichtein, E. (2007b), "Hits on question answer portals: exploration of link analysis for author ranking", in Kraaij, W., de Vries, A.P., Clarke, C.L.A., Fuhr, N. and Kando, N. (Eds), *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, *Amsterdam, The Netherlands*, *23-27 July*, ACM, New York, NY, pp. 845-846.

Kateryna, I., Cigdem, T., Delphine, B. and Iryna, G. (2009), "Annotating question types in social Q&A sites", Tagungsband des GSCL Symposiums 'Sprachtechnologie und eHumanities, Essen, Abteilung für Informatik und Angewandte Kognitionswissenschaft Fakultät für Ingenieurwissenschaften Universität Duisburg-Essen, pp. 44-49.

Kennedy, G.A. (2007), *Aristotle, On Rhetoric: A Theory of Civic Discourse*, Oxford University Press, New York, NY.

Kim, S., Oh, J. and Oh, S. (2007), "Best-answer selection criteria in a social Q&A site from the user-oriented relevance perspective", *Proceedings of the Association for Information Science and Technology*, Vol. 44 No. 1, pp. 1-15.

Kleinberg, J.M. (1999), "Authoritative sources in a hyperlinked environment", *Journal of the ACM (JACM)*, Vol. 46 No. 5, pp. 604-632.

Lankes, D. (2004), "The digital reference research agenda", *Journal of the American Society for Information Science and Technology*, Vol. 55 No. 4, pp. 301-311.

Lee, U., Yi, E. and Ko, M. (2013), "Mobile Q&A: beyond text-only Q&A and privacy concerns", *Proceedings of CHI 2013*, ACM, San Antonio, TX, pp. 533-542.

Li, G., Zhu, H., Lu, T., Ding, X. and Gu, N. (2015), "Is it good to be like Wikipedia? Exploring the trade-offs of introducing collaborative editing model to Q&A sites", in Cosley, D., Forte, A., Ciolfi, L. and McDonald, D. (Eds), *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, *Vancouver, BC, Canada*, ACM, New York, NY, pp. 1080-1091.

Li, L., He, D., Jeng, W., Goodwin, S. and Zhang, C. (2015b), "Answer quality characteristics and prediction on an academic Q&A site: a case study on ResearchGate", *Proceedings of the 24th International Conference on World Wide Web*, ACM, pp. 1453-1458.

Liu, D.-R., Chen, Y.-H., Kao, W.-C. and Wang, H.-W. (2013), "Integrating expert profile, reputation and link analysis for expert finding in question-answering websites", *Information Processing & Management*, Vol. 49 No. 1, pp. 312-329.

Lueg, C. and Fisher, D. (2012), *From Usenet to CoWebs: Interacting with Social Information Spaces*, Springer Science & Business Media, London.

McNally, K., O'Mahony, M.P. and Smyth, B. (2013), "A model of collaboration-based reputation for the social web", *Seventh International AAAI Conference on Weblogs and Social Media*, *Boston, MA*, *8-10 July*.

MacLeod, L. (2014), "Reputation on stack exchange: tag, you're it!", in Barolli, L., LI, K.F., Enokido, T., Xhafa, F. and Takizawa, M. (Eds), 28th *International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, *13-16 May*, IEEE Computer Society, Washington, DC, pp. 670-674.

Moore, T. and Serva, M. (2007), "Understanding member motivation for contributing to different types of virtual communities: a proposed framework", in Niederman, F., Summer, M., Lending, D. and Vician, C. (Eds), *Proceedings of the 2007 ACM SIGMIS CPR Conference on Computer Personnel Research: The Global Information Technology Workforce*, *St Louis, MO*, *19-21 April*, ACM, New York, NY, pp. 153-158.

Nam, K.K., Ackerman, M.S. and Adamic, L.A. (2009), "Questions in, knowledge in? A study of naver's question answering community", in Olsen, D.R. Jr, Arthur, R.B., HInkley, K., Morris, M.R., Hudson, S.E. and GreenBerg, S. (Eds), *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, *Boston, MA*, ACM, New York, NY, pp. 779-788.

Ortbach, K., Gaß, O., Köffer, S., Schacht, S., Walter, N., Maedche, A. and Niehaves, B. (2014), "Design principles for a social question and answers site: enabling user-to-user support in organizations", in Tremblay, M.C., Vandermeer, D., Rothenberger, M., Gupta, A. and Yoon, V. (Eds), *Advancing the Impact of Design Science: Moving from Theory to Practice: 9th International Conference, DESRIST, Miami, FL, 22-24 May Proceedings*, Springer International Publishing, Cham, pp. 54-68.

Page, L., Brin, S., Motwani, R. and Winograd, T. (1999), "The PageRank citation ranking: bringing order to the web", Stanford InfoLab, Stanford University, available at: http://ilpubs.stanford.edu:8090/422/ (accessed 30 May 2017).

Raban, D. and Harper, M. (2008), "Motivations for answering questions online", in Caspi, D. and Azran, T. (Eds), *New Media and Innovative Technologies*, Ben Gurion University, Negev Press, Tzivonim Publications, Beer Sheva, pp. 73-97, available at: https://pdfs.semanticscholar.org/6efa/bba4114be8aef270ca926d19515a78b0c102.pdf (accessed 30 May 2017).

Resnick, P., Kuwabara, K., Zeckhauser, R. and Friedman, E. (2000), "Reputation systems", *Communications of the ACM*, Vol. 43 No. 12, pp. 45-48.

Seth, F. (2009), "What's in a name? Everything, when you're talking wiki value", available: www.guardian.co.uk/technology/2009/feb/12/wiki-answers-wikia

Shachaf, P. (2009), "The paradox of expertise: is the wikipedia reference desk as good as your library?", *Journal of Documentation*, Vol. 65 No. 6, pp. 977-996.

Shachaf, P. and Rosenbaum, H. (2009), "Online social reference: a research agenda through a STIN framework", *Proceedings of the iConference, University of North Carolina*, *Chapel Hill, NC*.

Shah, C. (2011), "Effectiveness and user satisfaction in Yahoo! Answers", *First Monday*, Vol. 16 No. 6, available at: http://firstmonday.org/ojs/index.php/fm/article/vie%20w/3092/2769 (accessed 30 May 2017).

Shah, C. and Pomerantz, J. (2010), "Evaluating and predicting answer quality in community QA", in Crestani, F., Marchand-Maillet, S., Chen, H.H., Efthimiadis, E.N. and Savoy, J. (Eds), *Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, *Geneva*, *19-23 July*, ACM, New York, NY, pp. 411-418.

Shah, C., Oh, J.S. and Oh, S. (2008), "Exploring characteristics and effects of user participation in online social Q&A sites", *First Monday*, Vol. 13 No. 9, available at: www.firstmonday.dk/ojs/index.php/fm/article/view/2182/2028 (accessed 30 May 2017).

Shah, C., Oh, S. and Oh, J. (2009), "Research agenda for social Q&A", *Library & Information Science Research*, Vol. 31 No. 4, pp. 205-209.

Shen, H., Li, Z., Liu, J. and Grant, J.E. (2015), "Knowledge sharing in the online social network of Yahoo! Answers and its implications", *IEEE Transactions on Computers*, Vol. 64 No. 6, pp. 1715-1728.

Song, S., Wang, W., Que, X., Tian, Y. and Han, W. (2014a), "Modeling leading users in professional social-network based community question answering services", *Journal of Computational Information Systems*.

Song, Y., Jin, M.-Z. and Ren, P.-Y. (2014b), "Empirical study on personal motivation of UGC in Q&A community", *Pakistan Journal of Statistics*, Vol. 30 No. 5, pp. 651-666.

Vasilescu, B., Serebrenik, A., Devanbu, P. and Filkov, V. (2014), "How social Q&A sites are changing knowledge sharing in open source software communities", in Fussell, S.R., Lutters, W.G., Morris, M.R. and Reddy, M. (Eds), *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, Baltimore, MD*, ACM, New York, NY, pp. 342-354.

Wei, X., Chen, W. and Zhu, K. (2015), "Motivating user contributions in online knowledge communities: virtual rewards and reputation", in Bui, T.X. and Sprague, R.H. Jr (Eds), *International Conference on System Sciences (HICSS), 2015 48th Hawaii, 5-8 January*, IEEE Computer Society, Washington, DC, pp. 3760-3769.

Welser, W., Gleave, E., Fisher, D. and Smith, M. (2007), "Visualizing the signatures of social roles in online discussion groups", *Journal of Social Structure*, Vol. 8 No. 2, pp. 1-32.

Wenyin, L., Hao, T., Chen, W. and Feng, M. (2009), "A web-based platform for user-interactive question-answering", *World Wide Web*, Vol. 12 No. 2, pp. 107-124.

Yang, B. and Manandhar, S. (2014), "Exploring user expertise and descriptive ability in community question answering", in Wu, X., Ester, M. and Xu, G. (Eds), *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, *Beijing, China*, IEEE Computer Society, Washington, DC, pp. 320-327.

Yang, L., Qiu, M., Gottipati, S., Zhu, F., Jiang, J., Sun, H. and Chen, Z. (2013), "CQArank: jointly model topics and expertise in community question answering", in He, Q., Iyengar, A., Nejdl, W., Pei, J. and Rastogi, R. (Eds), *Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management*, *San Francisco, CA*, *27 October-1 November*, ACM, New York, NY, pp. 99-108.

Zdravko, M. and Daniel, T.L. (2007), *Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage*, ISBN: 978-0-471-66655-4, Wiley-Interscience, New York, NY.

Zhang, J., Ackerman, M. and Adamic, L. (2007), "Expertise networks in online communities: structure and algorithms", in Williamson, C.L., Zurko, M.E., Patel-Schneider, P.F. and Shenoy, P.J. (Eds), *Proceedings of the 16th International Conference on World Wide Web*, *Banff, Alberta*, ACM, pp. 221-230.

Zhou, B., Pei, J. and Luk, W. (2008), "A brief survey on anonymization techniques for privacy preserving publishing of social network data", *SIGKDD Explorations Newsletter*, Vol. 10 No. 2, pp. 12-22.

Zhou, G., Lai, S., Liu, K. and Zhao, J. (2012), "Topic-sensitive probabilistic model for expert finding in question answer communities", in Xue-wen, C., Guy, L., Haixun, W. and Mohammed, J.Z. (Eds), *Proceedings of the 21st ACM International Conference on Information and Knowledge Management, Maui, HI*, ACM, New York, NY, pp. 1662-1666.

Zhu, Z., Delphine, B. and Iryna, G. (2009), "A multi-dimensional model for assessing the quality of answers in social Q&A sites", in Bowen, P.L., Elmangarmid, A.K., Österle, H. and Sattler, K.U. (Eds), *Proceedings of the 14th International Conference on Information Quality*, *Hasso Plattner Institute, University of Potsdam*, *7-8 November*, pp. 264-265.

Further reading

Li, M., Zhao, T., Li, C. and Bian, N. (2013), "Topic-level expert modeling in community question answering", *Proceedings of the 2013 SIAM International Conference on Data Mining*, *Austin, TX, 2-4 May*, pp. 776-784.

About the authors

Dr Shah Khusro is working as a Professor at the Department of Computer Science, University of Peshawar, Pakistan. He received his PhD Degree from the Vienna University of Technology, Vienna, Austria. His research interests include web semantics, web engineering, information retrieval, web and mobile-based systems, and augmented reality. He is currently working on several research projects in these areas. He regularly publishes his work in international journals and conferences. Dr Shah Khusro is the corresponding author and can be contacted at: khusro@upesh.edu.pk

Aftab Alam is currently a PhD Scholar at the Department of Computer Science and Engineering, Kyung Hee University, South Korea. He received his MS Degree in Computer Science with specialization in Web Engineering from the Department of Computer Science, University of Peshawar, Pakistan. His research interests include social information seeking, social networks, social question and answers sites, web engineering and digital libraries. He is working as an IT Manager at the Quaid-i-Azam University, Islamabad, Pakistan.

Shah Khalid received his MS Degree in Computer Science from the Department of Computer Science, University of Peshawar, Pakistan in the year 2014. His research interests include semantic web, linked open data, information retrieval, web engineering and digital libraries. He is currently pursuing PhD at the School of Computer Science and Communication Engineering, Jiangsu University.