A Comprehensive Survey and Classification of Approaches for Community Question Answering

IVAN SRBA and MARIA BIELIKOVA, Slovak University of Technology in Bratislava

Community question-answering (CQA) systems, such as Yahoo! Answers or Stack Overflow, belong to a prominent group of successful and popular Web 2.0 applications, which are used every day by millions of users to find an answer on complex, subjective, or context-dependent questions. In order to obtain answers effectively, CQA systems should optimally harness collective intelligence of the whole online community, which will be impossible without appropriate collaboration support provided by information technologies. Therefore, CQA became an interesting and promising subject of research in computer science and now we can gather the results of 10 years of research. Nevertheless, in spite of the increasing number of publications emerging each year, so far the research on CQA systems has missed a comprehensive state-of-the-art survey. We attempt to fill this gap by a review of 265 articles published between 2005 and 2014, which were selected from major conferences and journals. According to this evaluation, at first we propose a framework that defines descriptive attributes of CQA approaches. Second, we introduce a classification of all approaches with respect to problems they are aimed to solve. The classification is consequently employed in a review of a significant number of representative approaches, which are described by means of attributes from the descriptive framework. As a part of the survey, we also depict the current trends as well as highlight the areas that require further attention from the research community.

Categories and Subject Descriptors: H.3.2 [**Information Storage and Retrieval**]: Information Search and Retrieval; K.4.3 [**Computers and Society**]: Organizational Impacts

General Terms: Algorithms, Design, Measurement

Additional Key Words and Phrases: Community question answering, knowledge sharing, online communities, exploratory studies, user modeling, content modeling, adaptive collaboration support

ACM Reference Format:
Ivan Srba and Maria Bielikova. 2016. A comprehensive survey and classification of approaches for community question answering. ACM Trans. Web 10, 3, Article 18 (August 2016), 63 pages.
DOI: http://dx.doi.org/10.1145/2934687

1. INTRODUCTION

With the development of Web 2.0, popularity of systems based on user-generated content such as Wikipedia, YouTube, or Flickr is continuously increasing. Another example that has become quite prominent in the past few years is Community Question Answering (CQA). CQA is a web-based service where people can seek information by asking a question and share knowledge by providing answers on questions asked by the rest

This work was partially supported by Grants No. APVV-15-0508, No. VG 1/0646/15, No. VG 1/0752/14, and No. KEGA 009STU-4/2014 and is the partial result of the Research and Development Operational Programme for the project "University Science Park of STU Bratislava," ITMS 26240220084, co-funded by the ERDF.

Authors' addresses: I. Srba and M. Bielikova, Faculty of Informatics and Information Technologies, Slovak University of Technology in Bratislava, Ilkovičova 2, 842 16 Bratislava, Slovakia; emails: {ivan.srba, maria.bielikova}@stuba.sk.

DOI: http://dx.doi.org/10.1145/2934687

of the community. Some CQA systems allow users to ask questions without any topic restriction (e.g., Yahoo! Answers), while other CQA systems are dedicated to a specific domain (e.g., Stack Overflow focuses on programming-related questions).

A typical process of question answering in CQA systems consists of several steps. It is usually initiated by unsuccessful information seeking in traditional information retrieval systems, such as web search engines. An asker visits a CQA system and posts a new question by providing its title, description, and assignment to one or several predefined topics. In comparison with web search engines, the question description should be formulated in a natural language, and thus it does not have to be reduced to keywords or limited to some basic semantics. It means that the required information can be described more precisely and thus the asker can receive the appropriate answer more effectively. Afterwards, the question is distributed to potential answerers who are most likely to provide good answers. This phase is absolutely essential for the question-answering process because if the question is not available to suitable users, it may not be answered properly or in acceptable time. In addition, this phase covers activities, such as posting comments or voting on questions and answers. Finally, the asker can choose the best answer and optionally in some CQA systems also provide a rating how good the answer is. As soon as the best answer is selected, the question is marked as resolved and placed into the CQA archive.

The main goal of CQA systems is to provide the most suitable answers on the recently posted questions in the shortest possible time. Compared with traditional information retrieval systems, CQA systems are able to harness tacit knowledge (embedded in their diverse communities) or explicit knowledge (embedded in all resolved questions) in answering of an enormous number of new questions posted each day. Nevertheless, the growing number of new questions could make CQA systems without appropriate collaboration support become overloaded by users' requests. As a result, askers would not be able to receive satisfactory answers in acceptable time, and thus the main goal of CQA systems would not be achieved. For this reason, many approaches have been already proposed to support the process of question answering. In addition, a number of case studies and data analyses concerning with users, questions, and answers have been conducted so far.

In spite of the increasing number of research articles aimed at CQA systems in recent years, a comprehensive review of state-of-the-art approaches has not been published so far. As a result, the negative consequences of the missing survey are quite compelling. At first, the relation between CQA concepts and existing theories on users' collaborations has not been fully described yet. Second, due to the absence of systematical classification of problems solved in CQA, it is really difficult (particularly for newcomers) to get an overview of the current state of the research. Furthermore, many similar problems are termed differently in the articles (e.g., question routing, answerer recommendation, and expert finding quite often refers to the same concept of new question recommendation to potential answerers). These inconsistences together with missing classification make the orientation in the existing approaches chaotic and ambiguous. Moreover, the approaches themselves are really diverse in terms of employed algorithms, features, datasets, evaluation metrics, and so on, and thus it is quite difficult to identify the state-of-the-art solutions or the preferred evaluation methodologies.

In order to face these drawbacks, we attempted to perform the first comprehensive survey and classification of approaches employed in CQA systems. Our main contributions are as follows:

(1) *Summary of theories behind two perspectives on CQA systems (Section 2)*. At first, we suggest that CQA systems can be perceived besides the primary and commonly referred knowledge sharing perspective, also from an alternative and unique perspective of collaborative learning. Consequently, for each of these perspectives, we

situate CQA systems into modern theories that allow us to better understand the theoretical background behind the community question-answering process and its successfulness.

(2) *A proposal of a general descriptive framework (Section 3.4).* On the basis of an extensive study of 265 articles (our survey methodology is described in Section 3.2), we identified a set of attributes that accurately characterize the existing CQA approaches–problems they tackle with, proposed solutions, contributions, as well as evaluation techniques. By putting these attributes together, we introduce a framework that provides an abstract layer above the question-answering process as well as above the CQA approaches.
(3) *An introduction of a complex classification (Section 3.5)*. By utilization of the most distinguishing attributes from the proposed framework, we propose a complex three-level categorization of problems solved in CQA. As part of the classification proposal, we attempt to consolidate terminology to make orientation in the literature easier.
(4) *A review of a representative approaches (Section 4, 5, and 6).* We describe a significant number of 142 approaches assigned to three main categories from the proposed classification: exploratory studies, content and user modeling, and adaptive support methods, respectively. Moreover, a description of all reviewed articles according to the proposed framework is available as an electronic appendix of this article.

The main purpose of this survey is to provide existing as well as potential researchers in the domain of CQA with an overview of the state-of-the-art literature and related theories. More specifically, our survey can serve researchers as a comprehensive guide to:

(1) Recognize not only a knowledge sharing but also learning potential embedded in the question-answering process. In addition, researchers can take advantage of the provided hierarchy of theories, which describe collaboration in CQA systems and, consequently, build their solutions with respect to these theories.
(2) Get an overview of open problems present in CQA systems. Moreover, as we put emphasis in the summarizing discussions on trending problems that represent the possible pointers for future work, our survey can help beginning researchers to select an interesting area for their research.
(3) Identify the most successful features, methods, and techniques in order to propose novel state-of-the-art solutions addressing selected open problems.
(4) Identify the commonly employed datasets, ground-truth definitions, and evaluation metrics in order to conduct valid experiments. In addition, our survey can be helpful in selection of the existing methods that can be considered as appropriate baselines for results comparison.

2. TWO PERSPECTIVES ON COMMUNITY QUESTION ANSWERING

In order to understand principles and concepts of CQA systems, which stand behind their success, we recognized that the question-answering process can be perceived from two different perspectives. In the first perspective, CQA can be considered as an information system that is fundamentally based on *knowledge sharing*. Knowledge sharing refers to a process in which a knowledge is exchanged among members of a particular community. At the same time, searching for an answer to a question we are asking is actually a specific way of informal learning. And thus CQA systems can be perceived also from a perspective of *collaborative learning*.

The increasing importance of computer support in knowledge sharing and collaborative learning caused that these domains became an interesting and promising subject of research in computer science. This area of research is, however, highly multidisciplinary as it lies at the intersection of computer science with information,

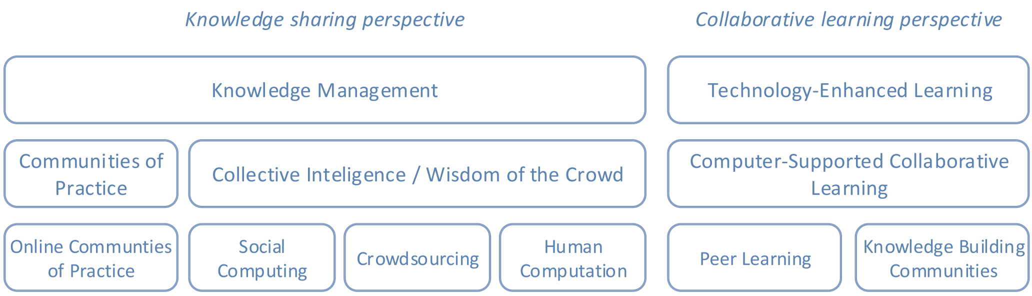

Fig. 1. Hierarchy of theories behind knowledge sharing and collaborative learning in CQA systems.

social, psychological, and pedagogical sciences. As a result, a number of theories and methodologies have been proposed. We emphasize that any research of information systems aimed to support users' interaction should be based on the existing theories and methodologies. Otherwise, the proposed research solutions are based only on naïve theories and, consequently, the achieved results are not usually significant enough.

Following this assumption, many authors in the current CQA-related literature refer to a number of various theories and concepts, such as knowledge management [Adamic et al. 2008], collective intelligence [Z. Li et al. 2012], or wisdom of the crowds [Mao et al. 2013]. In spite of that, the relations between these theories and CQA systems have not been systematically described so far, and thus they remained quite unclear and ambiguous. For this reason, we situate CQA systems into a hierarchy of theories that allows us to understand theoretical background behind the process of community question answering better (see Figure 1).

The primary perspective on CQA systems, which is commonly considered in the state-of-the-art literature, is perspective of knowledge sharing. Wasko and Faraj [2000] suggest that there are three main views of knowledge underlying the design of knowledge management systems: *knowledge as object*, *knowledge embedded in individuals*, and *knowledge embedded in communities*. CQA systems can be easily classified as knowledge management systems that are based on tacit knowledge embedded in communities of their users. Several theories have been proposed so far to analyze internal processes how these communities share knowledge. We recognized that users and their collaboration in CQA systems can be described by two of them: first, by a theory of *communities of practice* and, second, by a pair of related theories named *collective intelligence* and *wisdom of the crowd*.

Communities of practice are defined as groups of people who have a common interest in a subject and collaborate to share ideas or find solutions [Wenger 1998]. As users in CQA systems collaborate solely by means of the Internet, they can be characterized also as *online communities of practice*. Communities of practice are, however, sometimes criticized in that they excessively emphasize a community rather than a mutual cooperation of individuals [Lévy 1997]. It is especially true in online communities of practice, such as those in CQA systems, where mutual interconnections among individuals are quite week. We are inclined to this opinion and thus we emphasize that success of CQA concepts can be explained particularly by a pair of related theories of collective intelligence and wisdom of the crowd, which have been introduced to describe online knowledge sharing communities more adequately. Collective intelligence is built on an idea that nobody knows everything but everybody knows something, and thus we can harness intelligence of the whole community to solve tasks that would be hardly possible to be solved by an individual. The concept of wisdom of the crowd was described by Surowiecki [2005] as "the process of taking into account the collective opinion of a group of individuals rather than a single expert to answer a question." Wisdom of the crowd builds on an assumption that if we aggregate even imperfect data created by individuals, we are able to obtain a result that is better in comparison with the estimate created by the best expert. Collective intelligence and wisdom of the crowd represent a base for three subsequent well-known theories that are closely related to CQA systems: *human computation*, *crowdsourcing*, and *social computation*.

In contrast to the knowledge sharing perspective, the second perspective of collaborative learning provides a quite uncommon attitude to CQA systems. The question-answering process causes a flow of knowledge from more experienced users to less experienced ones. These users can gain new knowledge by reading, asking, and answering questions. In addition, users are able to elaborate solutions to solved problems by discussions attached to questions or answers. Thereby, it is very natural to speak about learning in CQA systems; nevertheless, there are only a very few studies concerned with the learning potential of CQA systems [Aritajati and Narayanan 2013].

This leaning potential present in the most of existing CQA systems can be characterized as *technology-enhanced informal learning* and, more specifically, *computer-supported collaborative learning*. Users in typical CQA systems lack a participation of an instructor or a teacher and thus we can recognize collaboration in CQA systems as a special case of collaborative learning termed *peer learning*. The key concept of peer learning is that learners are gaining new knowledge from each other instead of their instructor or teacher.

Besides peer learning, the second theory that can describe computer-supported collaborative learning in CQA systems is *knowledge building communities*. Establishment of a knowledge building approach was motivated by the current trend of the knowledge age in which sustained knowledge advancement is seen as essential for social progress [Scardamalia and Bereiter 2006].

In many cases, knowledge management and technology-enhanced learning are studied in isolation; however, the potential of their convergence has been already recognized [Chatti et al. 2012]. We suppose that CQA systems have a great potential to become an example of this convergence where knowledge sharing and collaborative learning is not only present side by side but can be utilized by a community of cooperating users to achieve more successful collaboration.

3. CLASSIFICATION AND DESCRIPTION OF CQA RESEARCH APPROACHES

3.1. Determination of Survey's Scope

The first Computer Supported Collaborative Work (CSCW) research articles tackling applications based on computer-mediated question answering date back to the mid-1990s, for example, Answer Garden 2 [Ackerman and McDonald 1996] and Lotus Notes [Whittaker 1996]. Since the mid-2000s, computer-mediated knowledge sharing has been significantly improved with the advancement of information and communication technologies, especially with emergence of Web 2.0. As a result, several variants of question-answering systems have been established. To explicitly determine the scope of our survey, we provide an overview of terminology and definitions of different kinds of question-answering systems with an emphasis how they differ from community question answering, which we are interested in.

Probably the most comprehensive hierarchical classification of computer-mediated question-answering systems was provided in Shah et al. [2014a] and Choi et al. [2012] together with analyses of how users' motivations, asking strategies, and types of questions differ between them.

(1) At first, it is possible to distinguish between automatic and human-driven question-answering services. In comparison with CQA systems, automatic question answering is concerned proposing methods that automatically answer questions asked by humans in natural language, while they do not involve any humans as answerers.

(2) Among human-driven services, there are two major categories: expert-based and peer-based services (commonly termed also as social Q&A services). In expert-based services (e.g., IM-an-Expert [Richardson and White 2011]), answers are provided by small groups of experts rather than an open community [Choi et al. 2012]. For this reason, they can easily employ effective real-time instant messaging, which is suitable especially if an asker desires a quick response (e.g., in enterprise or in virtual reference services). In addition, some expert-based services work on a pay-per-answer principle (e.g., former Google Answers, where a higher price was confirmed to increase likelihood of getting an answer, while it does not affect its quality [Jeon et al. 2010]) or market-based approach (e.g., mimir, where a mechanism of economic market managed to reduce non-serious and non-important questions [Hsieh and Counts 2009]).
(3) Finally, peer-based services can be further divided into three subgroups: besides community question answering, it is also collaborative question answering and social question answering (this denomination is proposed by Shah et al. [2014a] and Choi et al. [2012] despite possible confusion as the same term is commonly used also for a whole group of peer-based question-answering systems). Collaborative question-answering systems (e.g., Quora or Wiki Answers) use the same mechanisms as community question answering (e.g., to submit a question, to provide answers, to vote on them); however, they provide, in addition, an ability to edit and improve posts by collaboration with other users. Social question-answering systems utilize features of social networking sites as a means for knowledge sharing. Typical examples include system Aardvark [Horowitz and Kamvar 2010] or work by Nichols and Kang [2012], aimed at identifying possible answerers on Twitter.

On the one hand, all mentioned types of question-answering systems can be grouped together as particular cases of online QA systems. On the other hand, the differences between some of them are fundamental, and thus they are characterized by different research problems and solutions.

In our survey, we primarily focus on community question-answering systems. It is clear that automatic question answering represents a substantially different area, as it does not involve human answerers at all. Similarly, there are some significant differences also with regards to expert-based systems, especially in core principles they employ (e.g., a group of experts instead of an open community, a real-time chat instead of a community website with an archive of solved questions), and thus we decided to omit these systems from the survey. The differences between these variants of question-answering systems are obvious also with respect to underlying theories, especially in terms of knowledge type, which underlines their design (see Section 2). While automatic question answering relies on knowledge as object principle, expert-based systems utilize mainly knowledge embedded in individuals, and, finally, all peer-based systems take advantage of knowledge embedded in communities. However, as pointed out by Shah and Kitzie [2012], despite the difficulties in comparing them, some approaches from expert-based systems (e.g., expert finding [White and Richardson 2012] or community size analyses [White et al. 2011]) can be still interesting for CQA systems despite the fact that they cannot be applied directly.

From peer-based question-answering systems, collaborative question answering is very similar to community question answering in many aspects, and, thus, we cover them in the survey, too. On the other side, social question-answering systems differ significantly, as asking questions in standard social networking services poses a number of different challenges and, at the same time, many features characteristic for community question answering are not available here (e.g., best answer selection).

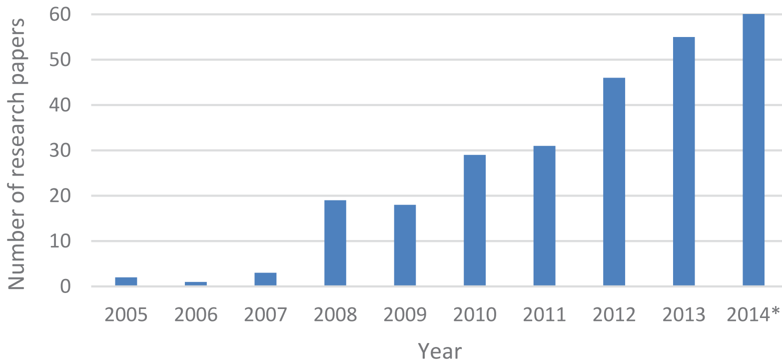

Fig. 2. Number of research articles concerned with CQA systems covered in the survey. *The last year 2014 contains the articles that were available in digital libraries in February 2015

3.2. Methodology

In order to ensure that the coverage of our survey is as good as possible, we paid significant attention to collection of research articles. During the initial search phase, we utilized search tools provided by major digital libraries that contain computer science articles, that is, ACM DL, IEEE Xplore, Springer Link, and ScienceDirect. More specifically, the following search queries were used: “Community Question Answering,” “CQA” and “Social Question Answering” (some researchers prefer to use this term, nevertheless it is usually dedicated to all peer-based question-answering systems). Consequently, from all returned search results, we manually selected those that explicitly concern with community and collaborative question-answering systems. Moreover, we enriched the list of relevant articles with additional publications identified from their related work. Finally, we obtained a list of 265 relevant articles authored in academia as well as in industry that have been published before the end of year 2014.

Deducing from the obtained list, the research on CQA systems is quite a new area that emerged with the first successful CQA systems, such as Yahoo! Answers, in 2005. From that time until present, we witness an increasing number of research articles that concern with various analyses and studies on the question-answering process (see Figure 2).

In addition, the obtained articles give us an interesting overview of conferences and journals, where the CQA approaches are published the most. A significant number of articles were published at major international conferences, such as the ACM International World Wide Web Conference (WWW), the ACM Conference on Research and Development in Information Retrieval (SIGIR), the ACM Conference on Computer Supported Cooperative Work (CSCW), the ACM Conference on Human Factors in Computing Systems (CHI), or the IEEE/ACM Conference on Advances in Social Networks Analysis and Mining (ASONAM). In comparison with conference articles, there is a smaller number of articles published in journals, such as in *ACM Transactions on Information Systems*, *Knowledge and Information Systems* (Springer), or *IEEE Transactions on Knowledge and Data Engineering*, to name a few.

All articles have been carefully reviewed and became the basis for a proposal of a descriptive framework, which was further used to propose a classification. Finally, the proposed framework was used to describe representatives of all problems covered by the classification. In general, each problem/group of approaches is described by three parts: (1) a definition; (2) an overview of proposed solutions (i.e., commonly used algorithms and inputs); and, finally, (3) evaluation techniques.

3.3. Previous Surveys

In spite of the big variance and increasing number of research articles, we found among the obtained publications only partial or outdated surveys that concern with CQA approaches.

The first survey on CQA approaches was conducted in 2009 by Shah et al. [2009]. This survey covers only a small number of studies due to the short previous history of CQA systems (see Figure 2). Authors classified all approaches into two categories: content-centered and user-centered studies. The content-centered studies had mainly focused on evaluation of answer quality. Topics covered by the user-centered studies were more diverse, such as roles of users during question answering, user information needs or identification of authoritative users. Later, Gazan [2011] conducted a broader survey aimed at all peer-based question-answering services. The author divided all approaches into four categories: (1) question classification and retrieval; (2) answer classification and quality evaluation; (3) user satisfaction; and (4) motivation, reputation, and perceived authority. A valuable contribution of this survey is also identification of future research directions. In comparison with these two surveys, the scope of current state-of-the-art approaches is, however, far wider.

In contrast to the previously mentioned studies, the next one by Furlan et al. [2013] focused specifically on question routing methods (i.e., recommendation of recently posted question to potential answerers—for more detailed description see Section 6.2). The survey provides a well-arranged generalization of approaches proposed to solve question routing so far.

The latest and probably the most complex survey on CQA approaches was published as a part of the dissertation thesis by Li [2014]. The author classified existing approaches into eight categories of which five are reviewed in more detail. In spite of the widest coverage among the previous surveys, this review still does not provide a comprehensive view of all major problems solved in the domain of CQA systems.

The described surveys are either quite outdated or they focus on some specific categories of problems solved in the domain of CQA systems, and thus they do not provide a comprehensive overview of all approaches proposed for CQA so far. In spite of that, they represent a valuable starting point for our review.

In contrast to the previous surveys on CQA approaches, there are several publications that provide comparative evaluation of CQA systems themselves (e.g., Fichman [2011], Chua and Balkunje [2012], Chua and Banerjee [2014]). Therefore, we decided to omit a review of the most popular CQA systems in this survey and focus specifically on approaches aimed to analyze and support collaboration in CQA systems.

3.4. General Framework to Describe CQA Approaches

In order to provide a comprehensive picture about the state-of-the-art research in the domain of CQA systems, we needed to lay the solid foundations for our further review. We achieved this by proposing a general framework that describes the question-answering process on a more abstract level and, moreover, defines descriptive attributes that are common for existing CQA approaches and are able to describe the proposed solutions and evaluations.

Question Lifecycle. In spite of heterogeneity of the question-answering process in the existing CQA systems, it is possible to generalize this process and identify four phases that describe the question lifecycle:

(1) *Question Creation.* At first, a user posts a question by formulating a title and a description of a problem that is the subject of the particular question. In addition,

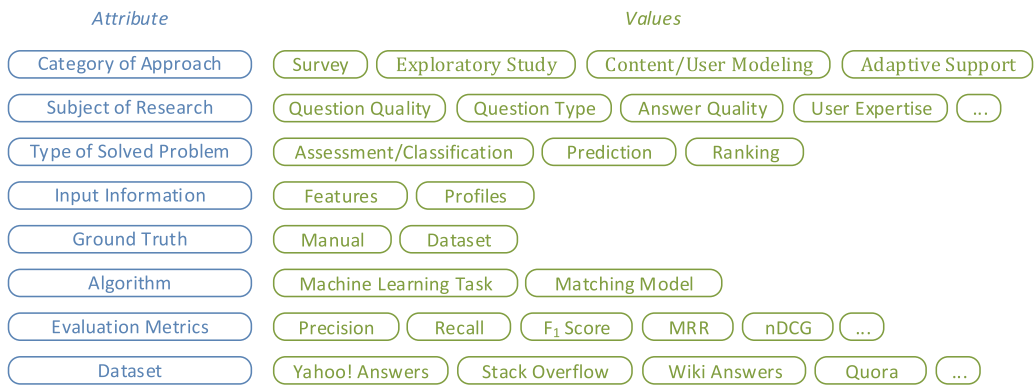

Fig. 3. Attributes and their possible values describing the existing CQA approaches.

it is usually necessary to select an appropriate question topic (a category or a set of related tags). We term the role of this user as an asker or a knowledge seeker.

(2) *Question Answering*. Afterwards, other users can collaborate and provide their answer-candidates on the posted question. These users play the role of answerers. Afterwards, all users can vote for the most appropriate answer-candidate and thus help the asker, the CQA system, and all other users who are involved in the question-answering process to identify answers with the highest quality.

(3) *Question Closing*. The asker can terminate the question-answering process by selecting the best answer that satisfies his or her information needs best. In situations when the asker never closes the question, the best answer can be selected by CQA system itself [Liu and Agichtein 2008] (e.g., in the case of Yahoo! Answers the best answer is selected according to the number of votes).

(4) *Question Search*. Consequently, the question is marked as answered and moved to the archive of solved questions. The existing CQA systems already contain a huge number of archived question-answer pairs, and thus besides the primary question-answering scenario, CQA archives offer an alternative way to find answers. The systems often provide facilities for discovering answered questions by full text search or navigation in taxonomy of questions' topics.

Domain Entities. We recognized that the question-answering process is characterized by the presence of three crucial domain entities: a question, an answer, and a user, who can play two main roles: an asker and an answerer. Afterwards, data coming from the question-answering process can be represented as a graph where nodes consist of these entities and edges represent various relationships between them (e.g., a user posts a question, a user votes on an answer). In our framework, we use a notation based on this graph representation inspired by Agichtein et al. [2008]: letters Q, A, and U stand for the particular question, answer, or user, respectively. Consequently, a sequence of letters, for example, Q.A.U, expresses a path in the question-answering graph, such as Q.U represents an asker of question Q, and, similarly, A.U represents an answerer who provided answer A.

In addition to the formalization of the question-answering process and domain entities, we define in the framework attributes that are mutual for all CQA approaches and describe their most important characteristics (see Figure 3).

Category of Approach. At first, the existing approaches can be distinguished by their fundamental character. Among the obtained articles, we identified surveys, exploratory studies, content/user modeling approaches, and approaches that tackle with providing adaptive support.

Subject of Research. Another quite significant attribute describing the existing approaches is a subject of research. In this attribute, we merged two closely connected characteristics: a domain entity, which an article focuses on (e.g., question), and its particular researched property (e.g., quality).

Type of Solved Problem. We recognized that the reviewed approaches tackle three different types of solved problems: (1) assessment or classification of a current domain entity's property (e.g., estimation of question complexity or question type classification), (2) prediction of a future value of domain entity's property (e.g., prediction whether an answer will be selected as the best one or not), and (3) ranking of domain entities according to a particular property (e.g., ranking answers according to their quality).

Algorithm. An additional important attribute, which distinguishes the analyzed approaches, is an employed algorithm. In content/user modeling, the problems are perceived as standard machine-learning tasks (e.g., classification, regression, learning to rank). In adaptive support, recommendations are usually provided on the basis of results from custom matching models (e.g., question routing approaches are based on similarity matching between a newly posted question and users' previously answered questions).

From machine-learning tasks, we omit a detail explanation of well-known classification and regression due to the article length limitation and we describe in particular the learning-to-rank (L2R) model. Learning-to-rank methods fall into three categories [Liu 2007]: pointwise, pairwise, and listwise. The pointwise strategy corresponds to a regression problem where each ranked document is characterized by a relevance score. In the pairwise strategy, the method learns a relationship between each pair of documents in order to compare their relevance to the query. Finally, the listwise strategy is based on complete ranking of all documents according to their relevance for the query.

Input Information. Besides the employed algorithm, each CQA approach is characterized by the scope of input information. In the case of content/user modeling, we recognized that the approaches employ a large set of low-level features describing domain entities (e.g., question length). In the case of adaptive support methods, high-level profiles are employed to describe domain entities (both the content and users). These profiles are commonly filled with characteristics calculated by content/user modeling approaches (e.g., user topical expertise).

To make the orientation in the large set of low-level features easier, we created their simple categorization (see Figure 4). Questions and answers have four groups of features:

(1) *Textual features* relate to the textual body of an analyzed question or answer itself: length (e.g., a number of words in the question title/description or answer body), structure (e.g., a presence of a code snippet, a number of URLs), and style/readability (e.g., a number of typos, a punctuation density, an averaged word length).
(2) *Non-textual features* capture important metadata about a question/answer: community feedback (e.g., a number of votes/favorites/abuse reports, a best answer selection) and temporal (e.g., time when the question was posted, time to the first answer).
(3) *Thread features* describe context of a question/answer: relevance/similarity (e.g., a number of words that overlap between a question and an answer) and thread statistics (e.g., a number of answers/comments, an answer position).

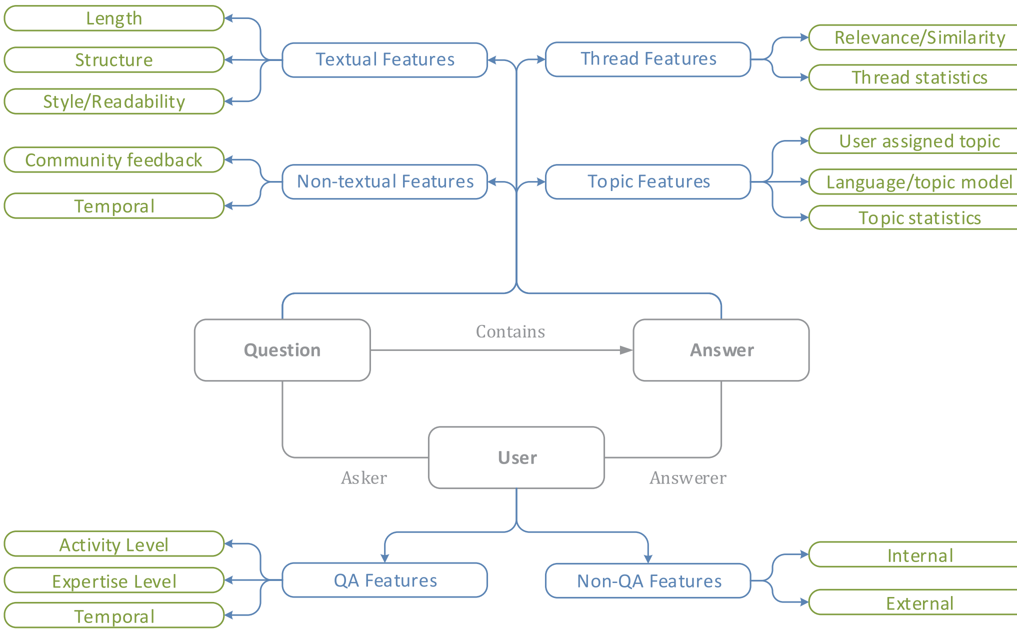

Fig. 4. Categorization of low-level features describing main domain entities.

(4) *Topic features* capture the meaning of the content as such: user assigned topic (e.g., tags or categories selected by an asker), language/topic model (e.g., bag of words, n-grams, Latent Dirichlet Allocation (LDA) topics), and topic statistics (e.g., a number of questions assigned to the same topic, an average score of questions with the same topic).

Users have two groups of features:

(1) *QA features* come from the question-answering process itself: activity level (e.g., a number of asked questions or posted answers/best answers), expertise level (e.g., a ratio of answers selected as the best), and temporal (e.g., time from the registration).
(2) *Non-QA features* describe a user on the basis of information that does not emerge directly from the question-answering process: internal (e.g., "about me" description, a number of followers) and external (e.g., connected accounts at social networking sites).

Ground Truth. In general, ground truth in the analyzed approaches is obtained by two main ways: manually (usually by Mechanical Turk workers or domain experts) or automatically from a dataset (e.g., in the best answer prediction, it is possible to utilize information whether an answer was actually selected as the best one or not). Manual evaluation is useful especially in cases when it is not possible to obtain ground truth directly from a dataset (e.g., in question subjectivity classification). However, manual evaluation can be really time consuming and thus it is not usually possible to apply it on great datasets.

Evaluation Metrics. CQA approaches used a number of standard evaluation metrics well known from information retrieval (e.g., precision, recall, F_1 score).

Dataset. Last but not least, CQA approaches can be distinguished by datasets used during evaluation. Datasets from many popular CQA systems are commonly utilized, such as from Yahoo! Answers, Stack Overflow, Wiki Answers, Quora, or Naver

Table I. Overview of Number of Articles Assigned to Top Two Levels of the Proposed Categorization with Information about Time Period of Their Publication

Category	Number of articles	The oldest article published in	The newest article published in
Exploratory Studies	**55**	**2008**	**2015**
System-wide Analyses	21	2008	2014
Content-related Analyses	12	2008	2014
User-related Analyses	22	2008	2015
Content and User Modeling	**91**	**2006**	**2015**
Question Quality	24	2008	2015
Question Type	12	2008	2014
Question Topic	12	2010	2014
Answer Quality	23	2006	2014
User Expertise	16	2007	2014
User Type	4	2012	2014
Adaptive Support	**90**	**2005**	**2015**
Question Retrieval	45	2005	2015
Question Routing	33	2005	2015
Question Suggestion	2	2013	2013
Answer Summarization	9	2008	2014
User Motivation	1	2013	2013

Knowledge-iN. In particular, there is a significant number of studies on Stack Overflow. There are two main reasons for this. The Stack Overflow dataset (as well as all datasets from Stack Exchange platform) is easily available, as it is regularly published under creative commons license. Furthermore, it was used twice as a data source during the mining challenges, which were held at the conference on Mining Software Repositories (MSR) in 2013 and 2015. On the other side, Quora is definitely an interesting and successful collaborative question-answering system; nevertheless, it is not so frequently used in research articles in comparison with Yahoo! Answers or systems built on top of the Stack Exchange platform. Probably, the main reason is that its dataset is not so easily available, and thus it has to be manually crawled (e.g., in Zhao et al. [2015]).

3.5. Classification of CQA Approaches

With regards to the introduced descriptive framework, we, consequently, proposed a complex three-level categorization of approaches that are specifically aimed to support the question-answering process (see Figure 5).

At the first level, we divided the CQA approaches into three categories according to the *category of approach* attribute from the introduced descriptive framework (see Figure 3). These categories are namely: exploratory studies, content and user modeling, and adaptive support. At the second level, the *subject of research* attribute was utilized to classify approaches. At the third level, we classify approaches according to the *type of solved problem*.

Quantity of articles at the first level of the proposed categorization is distributed quite evenly; nevertheless, at the second level, differences in the number of articles are rather significant (see Table I). Naturally, some of categories at the second level attract more research as well as provide more open problems to be investigated. In addition, we found out that categories at the first and the second levels are fairly stable in time, as almost all of them contain articles published from the very beginning of the CQA research until the present. However, it is not true when considering publication years at the third level of the categorization (because of the article length restrictions,

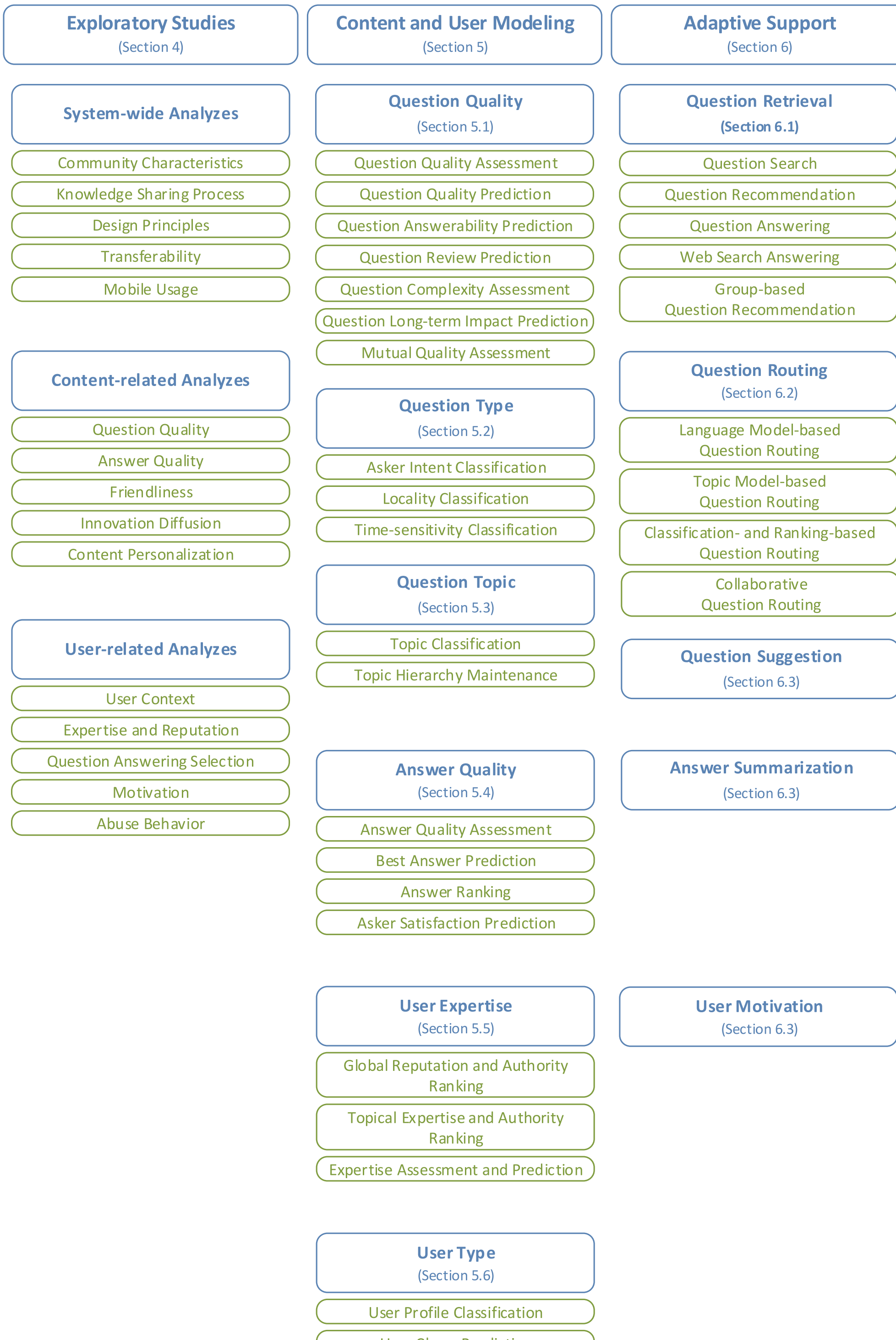

Fig. 5. Proposal of three-level classification of CQA approaches.

we do not provide a full overview of their quantity and publications years; however, this information can be easily found in the electronic appendix of our survey). The third-level categories, which are based on particular problems solved in the articles, are noticeably dynamic as they reflect actual needs, possibilities, and problems that constantly emerge as well as disappear in CQA systems.

4. EXPLORATORY STUDIES

In the proposed categorization, we focus particularly on content/user modeling and adaptive support approaches, which constitute the main body of CQA approaches. Nevertheless, exploratory studies still represent an important part of CQA research and thus at least the brief overview is provided.

Exploratory studies are concerned with analyses of data that are recorded during the question-answering process. Their outcomes represent an important contribution because they allow us to understand better how users interact during question answering, in particular, which circumstances make their collaboration successful or, on the contrary, prevent effective knowledge sharing. Afterwards, achieved findings can be utilized to propose more successful approaches to content/user modeling or providing adaptive support.

Exploratory studies can be widely divided into three subgroups according to the primary area they are concerned with: system-wide analyses, content-related analyses, and user-related analyses. In the following overview, we provide their short description together with references to selected representatives.

The first subgroup of *system-wide analyses* covers these topics:

—*Community characteristics* with a special emphasis on community evolution as CQA systems grow, for example, graph-based analyses of Quora community [Wang et al. 2013] or comparative analyses of the community in Yahoo! Answers and Baidu Zhidao [B. Li, Lyu, et al. 2012].
—K*nowledge sharing process* standing behind question answering, for example, Wang et al. [2014] proposed an analytical framework to describe thread-level communication between CQA users.
—*Design principles*, for example, Mamykina et al. [2011] analyzed the question-answering process at Stack Overflow in order to propose CQA design recommendations.
—*Transferability,* motivated by the high success of the CQA system on the open web, several studies concerning with CQA transferability to different domains have been conducted so far, for example, regarding how CQA concepts can be embedded into various software applications [Matejka et al. 2011], applied as a supplementary educational system [Aritajati and Narayanan 2013], or in a crowd-based support tool [Piccardi et al. 2014].
—*Mobile usage,* Lee et al. [2012] compared mobile CQA environments with traditional ones. They found out that in the mobile environment, users usually post questions/answers that are short in length and closely connected to their everyday life and thus highly dependent on spatial, temporal, and social context.

Another subgroup of exploratory studies concerns with *content-related analyses* and more specifically with these topics:

—Q*uestion and answer quality,* which is the aim of the largest part of content-related analyses, for example, Harper et al. [2008] compared answer quality in expert-based question-answering systems (Google Answers and virtual reference services) with CQA systems, Suzuki et al. [2011] investigated how various contextual information

included in a question can lead to better answers, and Yeniterzi and Callan [2014] analyzed how temporal and presentation bias affects community feedback.
—*Friendliness* of posts, which was explored by Cleary et al. [2013] in Stack Overflow community.
—*Innovation diffusion,* which was analyzed by Gomez et al. [2013] by means of link sharing on Stack Overflow.
—*Personalization* of posts, which was analyzed by Yardi and Poole [2009], who found that users commonly use various strategies to add their personal context into technologically oriented questions (e.g., by mentioning the reason why receiving an answer is important to them).

The last subgroup of exploratory studies relates to *user analyses* and covers these topics:

—*User context*, for example, Gardelli and Weber [2012] analyzed user pre-question behavior in order to identify situations in which users post questions to CQA systems after unsuccessful search in traditional search engines.
—*Expertise and reputation*, for example, Pal, Chang, et al. [2012] studied the evolution of experts, particularly how their behavior changes in time on a Stack Overflow dataset; similarly, Paul et al. [2012] studied authoritative users on Quora, Morrison and Murphy-Hill [2013] described how knowledge level relates to age of programmers at Stack Overflow, and Bosu et al. [2013] analyzed strategies of how a user can build his or her reputation at Stack Overflow.
—*Question-answering* s*election*, which refers to studies that try to describe how CQA users select questions to be answered, for example, Dearman and Truong [2010] surveyed users at Yahoo! Answers to determine situations when users decide to skip particular questions, and J. Yang, Tao, et al. [2014] identified two different groups of users at Stack Overflow: sparrows (i.e., users who answer a lot of easy questions) and owls (i.e., users who answer more difficult and popular questions).
—*Motivation*, which is essential for users' participation, for example, Grant and Betts [2013] examined how badges on Stack Overflow influence user behavior.
—*Abuse behavior*, which is a new emerging topic that arises from openness of CQA. Kayes et al. [2015] investigated a mechanism in Yahoo! Answers that allows users to mark inappropriate posts as abusive. The study revealed significant drawbacks of this mechanism as some users with many abuse reports were not actually harmful to the community, while many users, who violated community rules, were not reported at all.

From the proposed classification of exploratory studies, we would like to emphasize three groups of studies in particular. At first, abuse behavior has recently become a significant problem as we pointed out also in our exploratory study on Stack Overflow [Srba and Bieliková 2016]. By means of evaluation of community perception as well as data analyses, we showed that there are two emerging problems in Stack Overflow: an increasing failure rate (i.e., a proportion of questions that remain unanswered) and increasing churn rate (i.e., a proportion of users that leave the community). This negative development is highly related to the growing amount of low-quality content created by undesired groups of users (i.e., help vampires, noobs, and reputation collectors). In particular, help vampires purposefully abuse CQA systems to achieve their goals without returning the received help back to the community. Additional exploratory studies will be useful to describe the behavior of these users in more detail in order to propose suitable remedial actions (e.g., adaptive support methods) that will prevent abuse behavior in CQA.

Second, an interesting potential of CQA systems lies in their transferability to various additional environments and contexts. CQA systems have been recently successfully applied in enterprise environments (e.g., in crowd-based customer service [Piccardi et al. 2014]) or in educational systems (e.g., the CQA system Askalot [Srba and Bielikova 2015], which we specifically designed and developed for an educational and organizational environment), where it is possible to fully utilize the convergence between knowledge sharing and collaborative learning perspectives (see Section 2). As business or educational environments significantly differ from open CQA communities, additional exploratory studies, which will analyze possibilities how CQA concepts can be adapted in these environments, would be extremely helpful.

Finally, CQA systems in mobile environments represent another interesting challenge, as users can participate in question answering not only from their computers at home or work but also from anywhere as a meaningful way of spending their free time. Moreover, a lot of questions can arise as a result of everyday activities where temporal, spatial, and social context plays a significant role. This potential is present especially in general CQA systems (e.g., Yahoo! Answers) that contain mostly opinion-based questions that do not require further investigation of the problem (in comparison with domain-specific questions such as those asked on Stack Overflow).

5. CONTENT AND USER MODELING

The second category of approaches is concerned with modeling various characteristics of users, questions, and corresponding answers. Their purpose is to derive high-level attributes from low-level question-answering interactions and provide them as an important input for subsequent adaptation processes or core CQA features (e.g., the output of approaches to assess answer quality can be used to sort these answers).

5.1. Question Quality

In spite of a notable effort to guide users, posted questions in CQA systems are very diverse in their quality. The motivation to evaluate question quality lies in these four points [B. Li, Jin, et al. 2012]:

—at first, question quality directly affects answers quality;
—low-quality questions hamper an effective question-answering process; whereas, on the other hand,
—high-quality questions motivate users to contribute their knowledge; and, finally,
—question quality can be utilized in various adaptive approaches.

Question quality can be defined as effectiveness at attracting high-quality answers [Bian et al. 2009]. This homophily means that we can derive question quality directly from the quality of provided answers. A slightly different definition was provided by B. Li, Jin et al. [2012], in which authors emphasize an ability to attract answerers' attention and thus initiate answering attempts resulting in effectively obtaining the best answer. In another perspective, quality of questions can be measured by their complexity, popularity, or long-term term value. This diversity in question quality perception is caused not only by the absence of one established definition but also by its application as different forms of question quality are the most suitable in different scenarios.

Approaches aimed to evaluate various forms of question quality are described by means of attributes from the proposed framework in Table II, while Table III further depicts the employed low-level features (according to Figure 4).

5.1.1. Question Quality Assessment. Question quality assessment refers to estimation of question quality in order to identify/distinguish high- and low-quality questions. The

Table II. Description of Approaches Concerning with Question Quality

Problem	Machine-Learning Task	Ground Truth	Eval. Metrics	Dataset		Reference
Question Quality						
Q Quality Assessment	Classification; Stochastic Gradient Boosted Trees (2 classes)	Manual	P, R, F1	Yahoo! Answers	6.7K Q	[Agichtein et al. 2008]
	Regression + Classification; Linear + Discrimination (4 classes)	Evaluation rules	P	Stack Overflow	1.2M Q	[Ponzanelli et al. 2014]
	Classification; SVM (2 classes)	Manual	A, AUC	Yahoo! Answers	2.0K Q	[Shah et al. 2014b]
Q Quality Prediction	Regression; Linear	Score	-	Math Overflow	953 Q	[Tausczik and Pennebaker 2011]
	Classification; Linear/Non-linear (2 classes)	Score + # of views	A	Stack Overflow	66K Q	[Ravi et al. 2014]
Q Answerability Prediction	Classification; C4.5, Random Forest (3 classes)	Time to 1st A	P, R	Stack Overflow	N/A	[Asaduzzaman et al. 2013]
	Regression; Linear	# of answers	-	Weibo	13.8K Q	[Liu and Jansen 2013]
	Regression + Classification; Linear + Logistic regression (2 classes)	# of answers	RMSE, AUC	Yahoo! Answers	10M Q	[Dror, Maarek and Szpektor 2013]
Q Review Prediction	Classification; Stochastic Gradient Boosted Trees (2 classes)	Q closing	F1, A, AUC	Stack Overflow	203K Q	[Correa and Sureka 2013]
	Classification; AdaBoost (2 classes)	Q deletion	F1, A, AUC	Stack Overflow	470K Q	[Correa and Sureka 2014]
	Classification; Logistic regression (2 + 3 classes)	Q edit + Manual	P, R, F1	Stack Overflow	1K Q	[J. Yang, Hauff, et al. 2014]
Q Complexity Assessment	Classification; Logistic regression (2 classes)	Manual	P, R, F1, AUC	Server Fault	510 Q	[Burel and He 2013]
	Classification; Probability framework (2 classes)	Manual	P, R, F1	Yahoo! Answers	40K Q	[Lin et al. 2014]
Q Long-term Impact Prediction	Classification; Logistic regression (2 classes)	# of views	AUC	Stack Overflow	28,7K Q	[Anderson et al. 2012]
	Classification; Logistic regression (2 classes)	# of views	P	Tianya Wenda	20K Q	[Quan et al. 2012]
	Regression; Various	Score	RMSE	Stack Overflow, Math Overflow	2M Q	[Yao et al. 2014]
Mutual Quality Reinforcement	Classification; Logistic regression (2 classes)	Manual	P, R	Yahoo! Answers	250 Q	[Bian et al. 2009]
	Classification; Label propagation (2 classes)	Time to BA + # of A + # of favorites	A, Sensitivity, Specificity	Yahoo! Answers	8.5K Q	[B. Li, Jin, et al. 2012]
	Classification; Optimization framework (2 classes)	Score	P	Stack Overflow, Math Overflow	2M Q	[Yao et al. 2015]

Table III. Features Employed in Approaches Concerning with Question Quality

Problem	Question & Answer Features										User Features					Reference
	Textual			Non-textual		Thread		Topic			QA			Non-QA		
	Length	Structure	Style/Readability	Community feedback	Temporal	Relevance/similarity	Thread statistics	User assigned topic	Language/topic model	Topic statistics	Activity Level	Expertise Level	Temporal	Internal	External	
Question Quality																
Q Quality Assessment	Q		Q	Q				Q	Q	Q	Q.U	Q.U	Q.U			[Agichtein et al. 2008]
	Q	Q	Q													[Ponzanelli et al. 2014]
	Q	Q	Q	Q				Q			Q.U	Q.U				[Shah et al. 2014b]
Q Quality Prediction		Q			Q						Q.U	Q.U	Q.U		Q.U	[Tausczik and Pennebaker 2011]
	Q								Q	Q						[Ravi et al. 2014]
Q Answerability Prediction	Q	Q	Q		Q			Q	Q		Q.U					[Asaduzzaman et al. 2013]
	Q	Q	Q							Q	Q.U	Q.U				[Liu and Jansen 2013]
		Q	Q		Q			Q			Q.U			Q.U		[Dror, Maarek and Szpektor 2013]
Q Review Prediction	Q	Q	Q	Q				Q			Q.U	Q.U	Q.U			[Correa and Sureka 2013]
	Q	Q	Q	Q				Q			Q.U	Q.U	Q.U			[Correa and Sureka 2014]
									Q							[J. Yang, Hauff, et al. 2014]
Q Complexity Assessment	Q		Q	Q, Q.A	Q.A		Q				Q.U	Q.U	Q.U			[Burel and He 2013]
	Q.A									Q	Q.U, Q.A.U	Q.U, Q.A.U				[Lin et al. 2014]
Q Long-term Impact Prediction									Q	Q						[Anderson et al. 2012]
	Q.A			Q.A	Q, Q.A		Q				Q.U	Q.U				[Quan et al. 2012]
	Q, Q.A								Q		Q.U, Q.A.U	Q.U, Q.A.U				[Yao et al. 2014]
Mutual Quality Reinforcement	Q			Q, A	Q	Q-A	Q, A				Q.U	Q.U				[Bian et al. 2009]
	Q		Q								Q.U	Q.U				[B. Li, Jin, et al. 2012]
	Q, Q.A			Q, Q.A			Q				Q.U, Q.A.U	Q.U, Q.A.U				[Yao et al. 2015]

estimated quality can be consequently used as an input to ranking mechanisms (e.g., during displaying the results of question search).

Agichtein et al. [2008] considered question quality assessment as a classification problem. A number of features describing an asker and a question (textual, non-textual, and topical) have been considered. The obtained results indicated a strong influence of community feedback on classification precision. It was, however, necessary to normalize community feedback by average values obtained by questions in the same category, proveing the significant variance of categories. Shah et al. [2014b] report a significant improvement in question quality assessment with a two-stage approach when questions are classified according to their type, and, consequently, question quality is estimated according to textual features for each question type separately. To save time for Stack Overflow moderators, Ponzanelli et al. [2014] proposed a hybrid linear regression and classification approach to identify good-quality misclassified questions that have been reported and added to the review queue (a list of questions that have to be manually reviewed by moderators in order to decide whether to close or keep them). Readability and asker-related features were recognized as the most useful ones in this scenario.

There is no definite way to obtain objective ground truth for question quality assessment. Shah et al. [2014b] pointed out that answering features and community feedback might be unreliable ground truth because there is no indication that users had understood correctly the asker's information need. Therefore, Agichtein et al. [2008] and Shah et al. [2014b] decided to asses question quality manually by human assessors. On the other side, as manual evaluation can be subjective and is not applicable to label a larger number of questions, Ponzanelli et al. [2014] relied rather on the judgement of the community.

5.1.2. Question Quality Prediction. In contrast to the previous group of approaches, question quality prediction attempts to estimate question quality at or only shortly after creation time, that is, with no or only minimal answer information and community feedback.

The study by Tausczik and Pennebaker [2011] on CQA system MathOverflow showed that users' online reputation (calculated from previous question-answering activity) and offline reputation (derived from external non-QA data) can be effective to predict quality of new questions. Ravi et al. [2014] focused rather on question textual content, particularly on the length of a question's title and body together with latent topics and their global popularity. These features served as an input to several binary classification models, from which the best one achieved the prediction accuracy at 72%.

In comparison with question quality assessment, approaches to question quality prediction can take advantage that CQA datasets provide clear information about the future state of questions. Thereby obtaining ground truth is quite straightforward, and there is no necessity to employ manual labelling. Tausczik and Pennebaker [2011] attempted to predict question quality represented by the number of votes as it is perceived by the community. However, this score can be strongly influenced by topic popularity and thus Ravi et al. [2014] proposed to normalize this score with a view count.

5.1.3. Question Answerability Prediction. Question answerability prediction is closely related to question quality prediction as it predicts the future number of answers. The predicted value can be used during question formulation to give a preemptive warning to an asker that the question can remain unanswered and recommend him/her to rephrase the question [Dror, Maarek, and Szpektor 2013].

Dror, Maarek, and Szpektor [2013] introduced a series of scalable classification and regression models to predict whether questions at Yahoo! Answers will be answered and how many answer they will receive, respectively. Asaduzzaman et al. [2013] manually analyzed 400 randomly selected unanswered questions from the Stack Overflow system and created a taxonomy of the most common reasons of unanswered questions (e.g., too short, unclear, or vague questions or a program specific questions without a code snippet). The identified reasons of unanswered questions lead to a selection of various textual and user-related features considered in a classification model, in which authors goes beyond the previous work and attempt to predict a time span (divided into three categories) how long a question will remain unanswered. In Liu and Jansen [2013], social features, which are neglected in the previous studies, are taken into consideration in a regression model. The results showed that number of asker's followers or presence of emoticons can positively contribute to the number of answers, whereas, on the other side, questions with a higher number of hashtags as well as user mentions receive fewer answers, which contradicts the stated hypotheses.

5.1.4. Question Review Prediction. Question review prediction aims to predict a review process (i.e., whether a question will be closed, deleted, or edited), which is an indication of question (low) quality.

Correa and Sureka [2013] focused specifically on closed questions on Stack Overflow. Questions can be closed for many reasons, for example, when a question is duplicate, off topic, or subjective. The proposed classification model is able to predict closed questions with accuracy of 70.3%. The same authors [Correa and Sureka 2014] complemented the previous research with a binary classification task aimed at predicting whether a question will be deleted or not. Deleted questions can be predicted with accuracy of 66%. J. Yang, Hauff, et al. [2014] randomly selected 600 questions from Stack Overflow, which were edited by an asker or by a different user, and manually annotated the type of edit (e.g., source code refinement, supplementing a question with an example or with HW/SW details). Consequently, a binary classifier was trained to predict whether a question will be edited or not and, moreover, three additional classifiers attempted to predict the three most common edit types, although the achieved accuracy was not satisactory.

5.1.5. Question Complexity Assessment. Question complexity assessment aims to estimate a value that represents the difficulty and level of user expertise required for answering a question [Burel and He 2013].

Burel and He [2013] trained logistic regression models to identify complex questions in the dataset from the CQA system Server Fault. Despite modest accuracy, authors discovered that complex questions depend on five factors: asker topical focus, asker ratio of successfully answered questions, asker community age, question existing value, and a number of question views. Lin et al. [2014] addressed the question complexity assessment problem as a probability model that utilizes a phenomenon called a *knowledge gap* (i.e., questions asked by users with high expertise are answered by other expert users, while the inverse is true for users with a low level of expertise).

As CQA datasets do not contain any objective information about questions' complexity, in both approaches human annotators were asked to manually evaluate randomly selected questions.

5.1.6. Question Long-Term Impact Prediction. Question long-term impact prediction problem refers to a prediction of size of community that can benefit from the question and its answers (i.e., whether a question will continue to draw an attention from a community) [Yao et al. 2014]. The same problem was termed alternatively also as a *prediction of question popularity* [Quan et al. 2012] or a *prediction of long-lasting value* [Anderson et al. 2012].

Quan et al. [2012] formulated the problem of long-term impact prediction as a binary classification task. A logistic regression model was built with just three advanced features available immediately after question creation: an average number of views achieved by top similar questions and a relevance frequency of popular and unpopular terms in the body of the question. The achieved precision 83.7% of the proposed model outperform two classification models (k-NN (k-Nearest Neighbors), SVM (Support Vector Machines)) selected as a baseline. Similarly, Anderson et al. [2012] employed a logistic regression. Features derived a short time (1, 3, 24, and 72 hours) after question creation were examined with the result that even features available after only 1 hour provide enough information to the prediction. Yao et al. [2014] proposed a family of various comprehensive, flexible, and adaptive algorithms to predict questions' long-term impact. The prediction was performed after 24 hours from question creation.

There are several options of how to measure long-term impact of questions: a number of favorites, a number of views, and question's score. As the number of favorites is usually quite sparse, the remaining two measures were selected to automatically determine ground truth in the described articles.

5.1.7. Mutual Quality Assessment. Following an assumption that question quality is closely related to answer quality and asker expertise, the last group of approaches aims to assess their quality/expertise by their mutual reinforcement.

Bian et al. [2009] proposed a semi-supervised coupled mutual reinforcement framework in which question/answer quality and user expertise are calculated simultaneously. In contrast to the standard question quality assessment approaches, it requires only a small number of manually labeled data to initialize the training process. In B. Li, Jin, et al. [2012], question quality was assessed together with asker's expertise only. Finally, a homophily between question and answer quality was utilized in Yao et al. [2015] to jointly detect high-quality questions and answers.

5.2. Question Type

In the domain of traditional web search, the classification of search queries was proved as a valuable input for enhancing adaptive search methods. In particular, the taxonomy proposed by Broder [2002] is a well-known approach that puts search queries into three categories: informational, navigational, and transactional. Therefore, it is natural to examine the similar positive effect also in CQA systems and classify questions according to various taxonomies (see Tables IV and V for an overview of approaches and employed features).

5.2.1. Asker Intent Classification. The first research problem related to question type is classification of questions according to taxonomies that describe asker intent (e.g., subjective and objective questions). Consequently, assigned question type can be utilized to improve other approaches to domain modeling or adaptation support, such as assessment of question quality [Shah et al. 2014b] or answer quality [Toba et al. 2014].

Various taxonomies describing asker intent have been proposed so far. At first, according to questions' subjectivity, the questions can be characterized as objective or subjective [Li and Agichtein 2008]. Harper et al. [2009] divided questions into two groups: informational (i.e., fact- or advice-oriented questions asked with intent of getting information) and conversational (i.e., question asked with the intent to stimulate a discussion). The same categories of questions were used by Mendes Rodrigues and Milic-Frayling [2009], but they are termed differently as non-social and social questions. L. Chen, Zhang, and Mark [2012] mixed the previous taxonomies and categorized questions to three types: subjective, objective, and social. A more complex function-based taxonomy was used by Bu et al. [2010] in which questions are categorized with respect to an expected type of answer: fact, list, reason, solution, definition,

Table IV. Description of Approaches Concerning with Question Type and Topic

Problem	Machine-Learning Task	Ground Truth	Eval. Metrics	Dataset		Reference
Question Type						
Asker Intent Classification	Classification; SVM (2 classes)	Manual + co-training	F1	Yahoo! Answers	1K Q	[Li and Agichtein 2008]
	Classification; SVM (2 classes)	Manual	P, R, F1	Yahoo! Answers, MSN QnA	4K Q	[Mendes Rodrigues and Milic-Frayling 2009]
	Classification; Naive Bayes (2 classes)	Manual	A	Yahoo! Answers, AnswerBag, Ask Metafilter	490 Q	[Bian et al. 2008]
	Classification; Markov Logic Network (6 classes)	Manual	P, R, F1, A	Baidu Zhidao, Sina iAsk	5.8K Q	[Bu et al. 2010]
	Classification; SVM (2 classes)	Manual + co-training	F1	Yahoo! Answers	1.5K Q	[L. Chen, Zhang and Mark 2012]
	Classification; Logistic regression (6 classes)	Manual	A	Yahoo! Answers	5.8K Q	[Toba et al. 2014]
	Classification; SVM (3 classes)	Manual	A, AUC	Yahoo! Answers	3K Q	[Shah et al. 2014b]
Locality Classification	Classification; SVM (2 classes)	Yahoo! Placemaker Tags	F1	Yahoo! Answers, Wiki Answers	5.3M Q	[L. Chen, Zhang and Levene 2012]
Time-sensitivity Classification	Classification; C4.5, SVM (2 classes)	Manual	F1, A	Yahoo! Answers	1.7K Q	[Liu et al. 2009]
Question Topic						
Topic Classification	Hierarchical classification + Keyword extraction (389 classes)	User-assigned	P, R	Oshiete! goo	787K Q	[Nishida and Fujimura 2010]
	Classification; k-NN (1065 classes)	User-assigned	F1	Yahoo! Answers	1.1M Q	[Singh and Visweswariah 2011]
	Classification; Maximum Entropy (262 classes)	User-assigned	A	Yahoo! Answers	2M Q	[Cai et al. 2011]
	Classification; SVM (10 classes)	User-assigned	F1	Naver KiN	15K Q	[Bae and Ko 2012]
	Hierarchical kernelized categorization (6 classes)	User-assigned	F1	Yahoo! Answers	11K Q	[Chan et al. 2013]
	Probabilistic hypergraph learning	User-assigned	P@n, S@n	Zhihu	218K Q	[Nie et al. 2014]
Topic Hierarchy Maintenance	Clustering; EM	Manual	Hit Number, Weighted Precision	Yahoo! Answers	6K Q	[Miao et al. 2010]
	Ranking	User-assigned	Match@K	Yahoo! Answers	-	[G. Zhou, Cai, et al. 2012]

Table V. Features Employed in Approaches Concerning with Question Type and Topic

Problem	Question & Answer Features										User Features					Reference
	Textual			Non-textual		Thread		Topic			QA			Non-QA		
	Length	Structure	Style/Readability	Community feedback	Temporal	Relevance/similarity	Thread statistics	User assigned topic	Language/topic model	Topic statistics	Activity Level	Expertise Level	Temporal	Internal	External	
Question Type																
Asker Intent Classification									Q							[Li and Agichtein 2008]
	Q	Q					Q	Q	Q	Q	Q.U	Q.U				[Mendes Rodrigues and Milic-Frayling 2009]
								Q	Q		Q.U					[Bian et al. 2008]
									Q							[Bu et al. 2010]
					Q			Q	Q		Q.U					[L. Chen, Zhang and Mark 2012]
			Q						Q							[Toba et al. 2014]
									Q							[Shah et al. 2014b]
Locality Classif.		Q							Q							[L. Chen, Zhang and Levene 2012]
Time-sensitivity Classification					Q.A				Q, Q.A	Q						[Liu et al. 2009]
Question Topic																
Topic Classification									Q							[Nishida and Fujimura 2010]
									Q							[Singh and Visweswariah 2011]
									Q							[Cai et al. 2011]
									Q	Q						[Bae and Ko 2012]
									Q, Q.A							[Chan et al. 2013]
									Q		Q.U			Q.U		[Nie et al. 2014]
Topic Hierarchy Maintenance									Q							[Miao et al. 2010]
									Q							[G. Zhou, Cai, et al. 2012]

and navigation. Harper et al. [2010] proposed more formal taxonomy of questions grounded in Aristotelian rhetorical theory: advice, identification, (dis)approval, quality, prescriptive, and factual. Finally, Toba et al. [2014] proposed taxonomy with six question types: definition, factoid, opinion, procedure, reason, and yes/no question.

For the purpose of classification, authors employed different supervised and semi-supervised learning algorithms, mostly SVM, but also decision trees or Markov logic networks. Data describing a question, an asker, and even related answers are used as the classification features, such as question topic, question or answer creation time, or asker reputation.

As CQA systems do not contain any information about asker intent, it is possible to employ only manual labelling to obtain ground truth. The valuable time of manual annotators (domain experts as well as MTurk workers) can be saved by utilizing semi-supervised technique of co-training that have been successfully applied in approaches by Li and Agichtein [2008] and L. Chen et al. [2013]. In co-training, two independent feature sets are used to train two classifiers. Unlabeled data can be classified according to the first classifier and, consequently, used as labelled data for the second classifier and vice versa [Li and Agichtein 2008]. It means that two classifiers iteratively learn from each other.

5.2.2. Locality Classification. The problem of locality classification points to identification of questions' geographical dependency. This kind of information can be utilized, for example, in question routing as questions concerned with a particular location can be usually answered best by local answerers.

We are aware of only one approach belonging to this category. L. Chen, Zhang, and Levene [2012] proposed classifying questions as local (i.e., questions with strong spatial context) and global (i.e., the required information is independent on specific geographical location). Authors demonstrated that a SVM classifier was able to achieve the successful results and even classify correctly local questions that do not contain any specific place name.

5.2.3. Time-Sensitivity Classification. Questions can be classified not only by their geographical dependency but also by time-sensitivity. Similarly, as in the previous problem, the result of classification can be utilized in adaptive support methods, for example, urgent questions should be served preferentially as for such questions, late answers may not be useful.

Liu et al. [2009] classify questions according to their time sensitivity as urgent, non-urgent, and seemingly urgent. Question text provided sufficient information to achieve a high accuracy, which can be further improved by consideration of question category. On the other side, answer features did not improve classification performance, which means that the classification can be successfully performed even at question creation time.

5.3. Question Topic

CQA systems contain questions related to many different topics and thus it is not possible to put all these questions on the same pile. For this reason, a topic organization provides users with a simple and comprehensible structure of topics and questions assigned to them. When adding a new question, users are usually asked to select one or several topics related to the particular question. The purpose of topic organization is multiple. At first, assigned topics are commonly used to identify similar questions or to find users with similar interests. Second, it provides users with a navigation to the required questions more easily (e.g., when finding a question to answer or when searching archives of already solved questions).

The most common topic organization is based on a hierarchy of categories (e.g., three-level hierarchy of topics employed in Yahoo! Answers). In this solution, it is usually possible to assign a question only to one category that is a leaf of the whole category hierarchy. However, many questions can be related to more topics simultaneously [Nishida and Fujimura 2010]. This problem is usually solved by another topic organization based on tags where a question can be related to as many tags as necessary. On the other side, tags have only a plain structure without hierarchical relations (for example, as it is in Stack Overflow).

Tables IV and V provide an overview of question topic-related approaches and employed input low-level features.

5.3.1. Topic Classification. The hierarchy of topics can be quite broad, particularly in non-focused CQA systems. Therefore, assignment of a question to an appropriate topic can be a difficult task, especially for less-experienced askers [Singh and Visweswariah 2011]. In addition, the significant proportion of questions has usually at most one manually assigned tag, what finally hinders all tag-based mechanisms, such as following tags users are interested in [Nie et al. 2014]. These problems can be solved by approaches that automatically assign topics to existing or newly posted questions. Question classification into the predefined topics can be, however, a difficult task because features available before question posting (i.e., question title and body) are usually very limited in terms of available topical information and, moreover, there is a large number of categories in which questions could be assigned to (e.g., the hierarchy on Yahoo! Answers contains more than 1,000 leaf categories).

At first, approaches to topic classification differ in question representation. The majority of approaches employ only question content as an input feature, which can be modeled by a bag-of-words (BoW) model [Nishida and Fujimura 2010], by different variants of language models, such as by a translation-based language model [Singh and Visweswariah 2011], a translation-based language model enriched with semantic knowledge obtained from Wikipedia [Cai et al. 2011], or a category-based language model [Bae and Ko 2012]. The latest approaches employ more advanced representations, such as kernels based on keywords/n-grams/Part-of-speech (POS) tagging/syntactic trees [Chan et al. 2013] or an adaptive probabilistic hypergraph [Nie et al. 2014], in which hyperedges can be constructed not only from content but also from the question-answering history of asker and his/her followees.

Second, approaches employ various machine-learning algorithms to identify appropriate topics: a k-nearest neighbor classifier [Singh and Visweswariah 2011], a maximum entropy classifier [Cai et al. 2011], or SVM [Bae and Ko 2012]. Nie et al. [2014] proposed probabilistic hypergraph learning to identify semantically similar questions and, subsequently, a heuristic approach to further filter out the obtained tag candidates. In contrast to all the above-mentioned approaches, the following approaches assign questions into hierarchy of topics. Nishida and Fujimura [2010] proposed a hierarchical classification method, in which hierarchy of tags consists of three abstraction levels: category, theme, and keyword. Chan et al. [2013] proposed a sparse hierarchical classification method that was able to integrate information from several kernels in order to eliminate sparseness of data.

5.3.2. Topic Hierarchy Maintenance. Questions posted in CQA systems reflect actual information needs of web users. Therefore, their topics change dynamically according to actual trending problems. In contrast to dynamic character of users' questions, topic models in CQA systems do not change very often. As a result of this discrepancy, many CQA systems suffer with the problem of increasing size of categories dedicated to "other" questions that cannot be assigned to any more specific category. The purpose of topic hierarchy maintenance is to identify new categories that properly supplement the existing topic structure.

Miao et al. [2010] proposed a method to deal with new category identification problem based on probabilistic latent semantic analysis (PLSA), which was already used several times in the topic modelling problem in other domains with promising results. The proposed method was applied on Yahoo! Answers, more specifically on questions that were posted in an Internet category and were not assigned to any specific subcategory. The results of the experiment showed that the method was able to identify groups of questions that relate to similar topics, such as Twitter or eBay. G. Zhou, Cai, et al. [2012] directly extended the previous approach. Similarly, as it is important to identify new categories, it is also necessary to assign them appropriate labels. Traditional labelling approaches are not sufficient for this purpose because it is necessary to consider not only the content of questions in the newly identified category but also the existing hierarchy of topics. The new label has to be consistent with existing siblings: It should be on the same level of abstraction and, in addition, it cannot overlap with other categories. Authors decided to employ Wikipedia as a source of concepts related to the group of questions that should be named. These concepts are consequently filtered to obtain only suitable candidates that complement existing categories.

5.4. Answer Quality

Question quality and answer quality are an important attribute in CQA. Despite the significant effort to ensure only high-quality content in CQA systems, similarly to questions, the provided answers are not always appropriate. Actually, their quality can be very diverse, and thus high-quality answers can be mixed up with inappropriate answers or even with spam and abusive content. More specifically, an answer can be considered irrelevant if it does not relate to the question, incomplete if only partial information is provided, or incorrect if the provided information or part of it is not true or is no longer valid, and, finally, an answer can be biased when it considers only one person's view and does not reflect different perspectives [Sakai et al. 2011].

All these kinds of unsatisfactory answers make the effective knowledge sharing in CQA systems difficult and thus it is useful to classify the answers according to their quality. Afterwards, it is possible to filter out low-quality answers and, on the other side, highlight the most useful ones. Moreover, answer quality can be utilized to estimate user expertise or to evaluate successfulness of adaptive collaboration support approaches.

The importance of answer quality estimation was recognized in much research. However, they are not united with regard to what answer quality is and how ground truth should be set. Answer quality is usually defined by attributes that describe its content and relation to a corresponding question, such as responsiveness, accuracy, and comprehensiveness to the question [Bian et al. 2009]; satisfaction of information needs captured by the question [Sakai et al. 2011]; or relevance, completeness, objectivity, and originality [Zhu et al. 2009].

The overview of approaches concerned with answer quality and their features is provided in Tables VI and VII, respectively.

5.4.1. Answer Quality Assessment. Answer quality assessment refers to approaches that are proposed to evaluate quality of an individual answer, while quality can be represented by means of ordinal classes (e.g., low- and high-quality answers) or by a value from a predefined interval.

Various classification or regression approaches have been applied to solve this task so far. The first approaches employed sets of automatically extracted features, either non-textual [Jeon et al. 2006] or a combination of textual and non-textual [Blooma

Table VI. Description of Approaches Concerning with Answer Quality

Problem	Machine-Learning Task	Ground Truth	Eval. Metrics	Dataset	Reference
Answer Quality					
Answer Quality Assessment	Classification; Maximum entropy (2 classes)	Manual	P, R	Naver 2.6K A	[Jeon et al. 2006]
	Regression; Linear	Manual	-	Yahoo! Answers 2.1K A	[Blooma et al. 2008]
	Classification; Stochastic Gradient Boosted Trees (2 classes)	Manual	P, R, F1	Yahoo! Answers 8.3K A	[Agichtein et al. 2008]
	Regression + Classification; Linear + Discrimination (2 classes)	Manual	A	Answerbag 50 A	[Zhu et al. 2009]
	Classification; Logistic regression (2 classes)	Manual	P, R, F1, A	Baidu Zhidao 5.0K A	[C. Chen et al. 2013]
	Classification; Logistic regression (6 + 2 classes)	Manual	A	Yahoo! Answers 46.8K A	[Toba et al. 2014]
Best Answer Prediction	Brouwer Fixed Point	BA selection	A	MSN QnA 1.3M A	[Lee et al. 2009]
	Classification; Logistic regression (2 classes)	BA selection	A	Yahoo! Answers 600 A	[Shah and Pomerantz 2010]
	Classification; Logistic regression (2 classes)	BA selection	A	Yahoo! Answers 400 Q	[John et al. 2011]
	Classification; Decision Tree (2 classes)	BA selection	P, R, F1, AUC	SCN, SF, Cooking 213K A	[Burel et al. 2012]
	Classification; Decision Tree (2 classes)	BA selection	P, R, F1, AUC	Stack Exchange 12M A	[Gkotsis et al. 2014]
Answer Ranking	L2R; SVM Rank (pairwise)	BA selection	MRR, P@n	Yahoo! Answers 143K Q	[Hieber and Riezler 2011]
	L2R; SVM Rank (pairwise), ListNet (listwise)	BA selection	MRR, P@n	Yahoo! Answers 150K Q	[Z.-M. Zhou et al. 2012]
	Regression; Linear	Manual	MAP, NDCG, P@n	Yahoo! Answers 21.5K A	[B.-C. Chen et al. 2012]
	L2R; Random Forest (pointwise)	Answer score	NDCG	Stack Overflow 53K A	[Dalip et al. 2013]
	L2R; SVM Rank (pairwise)	Answer score	MRR, NDCG	Stack Overflow 100K Q	[Ginsca and Popescu 2013]
Asker Satisfaction Prediction	Classification; C4.5, Random Forest, SVM, AdaBoost, Naive Bayes (2 classes)	BA selection & rating	P, R, F1, A	Yahoo! Answers 5K Q	[Yandong Liu et al. 2008]
	Classification; Logistic regression (2 classes)	Bounty offered	AUC	Stack Overflow 28.7K Q	[Anderson et al. 2012]

Table VII. Features Employed in Approaches Concerning with Answer Quality

Problem	Question & Answer Features										User Features					Reference
	Textual			Non-textual		Thread		Topic			QA			Non-QA		
	Length	Structure	Style/Readability	Community feedback	Temporal	Relevance/similarity	Thread statistics	User assigned topic	Language/topic model	Topic statistics	Activity Level	Expertise Level	Temporal	Internal	External	
Answer Quality																
Answer Quality Assessment	A			A			A.Q				A.U	A.U				[Jeon et al. 2006]
	A		A			Q-A		A.Q			A.U, A.Q.U	A.U, A.Q.U				[Blooma et al. 2008]
	A, A.Q		A			Q-A			A			A.U				[Agichtein et al. 2008]
			A													[Zhu et al. 2009]
									A, A.Q		A.U, A.Q.U					[C. Chen et al. 2013]
	A, A.Q	A	A, A.Q			Q-A										[Toba et al. 2014]
Best Answer Prediction				A							A.U					[Lee et al. 2009]
	A, A.Q	A		A.Q			A.Q				A.U, A.Q.U	A.U, A.Q.U				[Shah and Pomerantz 2010]
	A, A.Q		A	A		Q-A					A.U, A.Q.U	A.U, A.Q.U				[John et al. 2011]
	A		A	A	A, A.Q		A, A.Q				A.U	A.U		A.U		[Burel et al. 2012]
	A		A	A	A, A.Q		A.Q		A		A.U					[Gkotsis et al. 2014]
Answer Ranking	A, A.Q	A	A, A.Q			Q-A			A, A.Q							[Hieber and Riezler 2011]
											A.U	A.U		A.U		[Z.-M. Zhou et al. 2012]
				A							A.U					[B.-C. Chen et al. 2012]
	A	A	A	A, A.Q	A, A.Q	Q-A	A, A.Q				A.U	A.U	A.U			[Dalip et al. 2013]
														A.U		[Ginsca and Popescu 2013]
Asker Satisfaction Prediction	A.Q		A.Q	A.Q	A.Q		A.Q			A.Q	A.U, A.Q.U	A.U, A.Q.U	A.U, A.Q.U			[Yandong Liu et al. 2008]
	A			A, A.Q	A						Q.U	A.U, A.Q.U				[Anderson et al. 2012]

et al. 2008; Agichtein et al. 2008]. In contrast to the previous studies, Zhu et al. [2009] identified 13 dimensions of quality (e.g., politeness, completeness), manually extracted them for a set of questions, and used them to construct a linear regression model. Later, more advanced approaches emerged, such as online detection of spam answers posted as a part of commercial campaigns [C. Chen et al. 2013] or hierarchical categorization that at first identifies a type of a corresponding question (e.g., factoid or opinion question; see Section 5.2 for more approaches concerning question type classification), and, consequently, the second question-type-specific classifier labels an answer as a good/bad one [Toba et al. 2014].

Evaluation of answer quality assessment is a difficult and challenging task because there is no clear information about true answer quality. All analyzed approaches were decided for manual evaluation; nevertheless, this has several significant drawbacks. As most of the questions are subjectively oriented and the context of each question/asker is unique and sometimes not well known, manual evaluation of answer quality can be inconsistent. In addition, manual evaluation of larger datasets (such as in Toba et al. [2014]) is really time consuming. Another possibility, regarding how to evaluate answer quality, is to employ an answer score. It can be significantly biased (e.g., by topic popularity or spam voting), but there are possibilities how this bias can be eliminated (e.g., as proposed in B.-C. Chen et al. [2012]).

5.4.2. Best Answer Prediction. The best answer prediction problem refers to the ability to predict whether an individual answer will be selected by an asker as the best one or not.

At first, Lee et al. [2009] proposed best answer prediction based on community votes weighted by a voting score that captures how often the previous votes of the particular user agreed with the best answer selection. This assumption leads to a circular definition that can be solved by an iterative computational process. The next approaches consider best answer identification as a classification problem with two classes that correspond to a best answer flag. Shah and Pomerantz [2010] compared manually and automatically extracted features. The authors created Human Intelligence Task (HIT) in the Amazon Mechanical Turk (MTurk) service to evaluate the answer's quality on a 5-point scale according to 13 quality criteria previously defined in Zhu et al. [2009]. The obtained manual evaluations as well as the automatically extracted features were consequently employed in the construction of classification models based on a logistic regression. The results pointed out that the manual evaluation without further context (e.g., asker or answerer history) is not sufficient to predict best answers. On the other side, Blooma et al. [2010] considered manually as well as automatically obtained features and recognized a significant prediction value of manually obtained ones. Besides content and user features, Burel et al. [2012] employed a set of thread features and found out that a thread-based feature score ratio (a proportion of score given to an answer in a total score given to all answers) represents a successful predictor for best answer identification. This feature, however, cannot be obtained near answer posting time. Gkotsis et al. [2014] showed that it is possible to achieve similar prediction accuracy also when considering only textual features at answer creation time.

Approaches that tackle best answer prediction can take advantage of easily accessible ground truth, as best answer selection is directly included in all employed datasets. We should emphasize that this kind of ground truth is appropriate for the best answer prediction as described above, although asker-selected answers do not have to be necessarily of high quality. Sakai et al. [2011] employed four assessors, who manually evaluated answers' quality on Yahoo! Answers, to point out a problem of best-answer discrete taxonomy. The best answer selected by an asker can be chosen subjectively, and thus it can be biased while there can be also other high-quality answers. A similar observation was confirmed by B.-C. Chen et al. [2012], who showed that a significant part (70%) of all answers that are marked as the best answers by an asker or by a community were manually evaluated by experts as fair or even bad.

5.4.3. Answer Ranking. Answer ranking approaches aim to obtain a relative rank of all answers provided for a particular question according to their quality.

In order to rank answers, learning-to-rank (L2R) or regression models can be applied. Hieber and Riezler [2011] treated answer ranking problem from information retrieval perspective. A question was expanded with snippets from web search in order to calculate more precise question-answer similarity features. Consequently, pairwise SVM Rank was employed to learn how to rank answers. B.-C. Chen et al. [2012] proposed a vote calibration model that predicts the potential bias in users' behavior. Achieved results on a manually annotated dataset proved that eliminating voting bias is able to significantly improve ranking precision. Dalip et al. [2013] ranked answers by means of a large set of 186 features describing a question, an answer, and an answerer. Z.-M. Zhou et al. [2012] demonstrated that it is possible to rank answers with the answerer-related QA features (e.g., total count of answer, best answer rate) and internal non-QA features (e.g., presence of a picture). Moreover, Ginsca and Popescu [2013] managed to successfully rank answers only with the answerer internal non-QA features (e.g., self-description, answerer age, or links to external platforms).

In order to obtain ground truth for answer ranking, various methods have been utilized so far. At first, it is possible to obtain ground truth manually as in the previous approaches. Second, the position of the best answer (selected by an asker or a community) in the calculated ranking can be evaluated. In this place, we would like to repetitively point out the already-described possible bias in the asker's best answer selection, which means that the best answer does not have to be necessarily the answer with the highest quality. Finally, a score provided by a community, which can be more objective, can be considered.

5.4.4. Asker Satisfaction Prediction. In the cases when an asker did not choose the best answer personally and the best answer was selected by a community or by a CQA system itself, we do not have any information whether the information need of the asker was fulfilled. The asker satisfaction problem solves the prediction of whether the asker would be satisfied with the answers provided by a community or not.

Agichtein et al. [2009] employed various classification algorithms (e.g., decision trees, SVM or Naïve Bayes) to predict whether the provided answers have satisfied the asker information needs or not. Experiment results showed that the classification outperformed not only the baseline (random selection) but also the selection derived from human judgments (Mechanical Turk workers). The same classification problem was addressed in Anderson et al. [2012] on the dataset from Stack Overflow.

To obtain ground truth, it is possible to take advantage of specific features implemented in CQA systems. In Yahoo! Answers, an asker can not only select the best answer but also express his/her satisfaction by explicit feedback at 5-point scale. Consequently, Agichtein et al. [2009] defined the asker's satisfaction as a situation in which the asker chose the best answer and provided a rating at least 3. Stack Overflow provides askers with a possibility to offer a bounty for answering their questions. The amount of bounty corresponds to points that are subtracted from the asker's reputation and thus Anderson et al. [2012] considered the act of providing bounty as an expression of insufficient satisfaction with the answers provided so far.

5.5. User Expertise

In research publications on CQA systems, user expertise is closely associated with several different terms such as user authority, user topical authority, user reputation, and user topical expertise. The common characteristic of all these terms is that they refer to a user-related measure that captures an amount of user knowledge and his/her potential to provide high-quality answers. As a result, they are sometimes used interchangeably, and thus the differences between them are neglected. To understand differences between them more precisely, we recognized that user expertise measures can be characterized by two dimensions:

(1) At first, user expertise measures can refer to expertise at a global level or at a particular topic (i.e., a user assigned tag/category or an automatically extracted topic). Global measures are usually represented by a single value that provides simple comprehensive information about a user, and thus it can be easily displayed in the user interface or utilized to rank users. On the other side, topical sensitive measures are represented by a rather more complex variable that naturally depends on particular topics. It can be used in situations when identification of experts on a certain topic is important, for example, in recommendation of recently posted questions to potential answerers (so-called question routing).
(2) Second, user expertise measures can be broadly divided according to the employed calculation method into graph-based and feature-based approaches. The graph-based approaches work with a social graph underlying users' interactions in CQA systems (mainly between askers and answerers). Various graph-based algorithms (e.g., algorithms developed to rank websites, such as PageRank and HITS) are then applied on these graphs in order to identify expert users in the community. The second group of feature-based approaches is based on historical question-answering records about users as well as about content created by them. Consequently, various mostly numerical methods are employed to derive user expertise.

In our categorization, we distinguish among various user expertise measures as follows. If user expertise is estimated at a global level, then it can be represented by *user authority* or *user global expertise*, which is often termed *user reputation*. While user authority is a graph-based measure, user reputation can be characterized as a feature-based approach. User reputation can be calculated either by reputation schemas (rule-based mechanisms commonly employed in the existing CQA systems) or numerically derived from users' question-answering history. Similarly, at a topical level, we can estimate user expertise by a graph-based measure, *user topical authority,* or by a feature-based measure, *user topical expertise*. Various particular methods aimed to calculate user expertise measures and their features are described in Tables VIII and IX, respectively.

There are two major approaches how to evaluate user expertise estimation [J. Liu et al. 2011]: employing traditional information retrieval evaluation metrics to compare the calculated estimation with ground truth (e.g., precision, recall, rank correlation) or evaluating user expertise estimation rather indirectly by evaluating the quality of answers provided by identified experts. In the first approach, it is necessary to obtain ground truth. However, both an automatically as well as a manually obtained ground truth has some significant disadvantages. Jurczyk and Agichtein [2007] created ground truth automatically from user features, such as best answer ratio. Unfortunately, such ground truth is obtained according to a certain heuristic method that itself can be considered as an approach to estimate user expertise [J. Liu et al. 2011]. Zhang et al. [2007] decided to employ two human evaluators to manually rank users according to their expertise. As it was necessary to read hundreds of answers posted by each user, human evaluators managed to evaluate only 135 users. Therefore, manual evaluation is absolutely inappropriate to create ground truth for robust datasets that are currently available from CQA systems. For this reason, the automatic evaluation derived from user features remains the only suitable technique to build usable ground truth.

Table VIII. Description of Approaches Concerning with User Expertise and Type

Problem	Machine-Learning Task	Ground Truth	Eval. Metrics	Dataset		Reference
User Expertise						
Global Reputation and Authority Ranking	Ranking, HITS (Asker-Replier)	Score, BA	Pearson	Yahoo! Answers	458K U	[Jurczyk and Agichtein 2007]
	Ranking; Z-score, ExpertiseRank, HITS (Asker-Replier)	Manual	Spearman, Kendall's τ	Java Forum	135 U	[Zhang et al. 2007]
	Ranking; TrueSkill, SVM (Competition-based)	BA, Manual	nDCG, P@n, MRR	Yahoo! Answers	6.5K U	[J. Liu et al. 2011]
	Ranking; PageRank, HITS, InDegree, Harmonic centrality (Competition-based)	BA selection	A	Yahoo! Answers	20M Q	[Aslay et al. 2013]
Topical Expertise and Authority Ranking	Ranking; Topical Random Surfer model (Asker-Replier)	Manual	P@n, MRR, MAP	Yahoo! Answers, Tianya Wenda	446K U	[Zhu, Cao, et al. 2011]
	Ranking; ExpertRank (Asker-Replier weighted by A Quality)	Manual	nDCG	Stack Overflow, Turbo Tax	90 U	[Cai and Chakravarthy 2013]
	Ranking; Topical Random Surfer model (Asker-Replier)	Manual	P@n, MRR, MAP	Yahoo! Answers	286K U	[Zhou et al. 2014]
Expertise Assessment and Prediction	Ranking + Classification; InDegree (Asker-Best Answerer) + Discrimization (2 classes)	Manual	-	Yahoo! Answers	-	[Bouguessa et al. 2008]
	Classification; SVM, C4.5 (2 classes)	Manual	P, R, F1	Turbo Tax	605K U	[Pal et al. 2011]
	Classification; Bagging metaclassifier (2 classes)	Z-score + manual	P, R, F1	Stack Overflow, Turbo Tax	287K U	[Pal, Harper, et al. 2012]
	Classification; Random Forest (2 classes)	SO reputation	P, R, F1	Stack Overflow	1.3M U	[Movshovitz-Attias et al. 2013]
User Type						
User Profile Classification	Ranking; QP-Borda count ranking model	Manual	P	Quora	N/A	[Song et al. 2013]
	Classification; Artificial neural network (10 classes)	Derived from cluster analysis	P	Stack Exchange (5x)	102K U	[Furtado et al. 2013]
User Churn Prediction	Classification; Naive Bayes, Logistic Regression, SVM, Decision Tree, Random Forest, k-NN (2 classes)	User history	F1, AUC	Yahoo! Answers	20K U	[Dror et al. 2012]
	Classification; Logistic Regression, SVM, Decision Tree (2 classes)	User history	A	Stack Overflow	1M U	[Pudipeddi et al. 2014]

Table IX. Features Employed in Approaches Concerning with User Expertise and Type

	Question & Answer Features										User Features					
	Textual			**Non-textual**		**Thread**		**Topic**			**QA**			**Non-QA**		
Problem	Length	Structure	Style/Readability	Community feedback	Temporal	Relevance/similarity	Thread statistics	User assigned topic	Language/topic model	Topic statistics	Activity Level	Expertise Level	Temporal	Internal	External	**Reference**
User Reputation																
Expertise Assessment and Prediction			U.A	U.A	U.A						U	U	U			[Pal et al. 2011]
											U	U				[Pal, Harper, et al. 2012]
											U	U				[Movshovitz-Attias et al. 2013]
User Type																
User Profile Classification																[Song et al. 2013]
				U.Q			U.A				U	U	U			[Furtado et al. 2013]
User Churn Prediction	U.Q/A	U.A		U.Q	U.A		U.A.Q	U.Q			U	U	U	U		[Dror et al. 2012]
	U.Q/A			U.Q/A	U.A		U.A.Q				U	U	U			[Pudipeddi et al. 2014]

5.5.1. Global Reputation and Authority Ranking. The first group of approaches tackles with user expertise at a global level and perceives its estimation as a ranking problem.

At first, we introduce how popular CQA systems rank users according to their global user expertise. CQA systems utilize estimation of user reputation as a part of their gamification systems in order to provide users with motivation to actively participate on question answering. This reputation is calculated according to various reputation schemas based on simple rules in order to be transparent for a community. In addition, system administrators can simply influence the community behavior by gamification in order to promote insufficient actions in the system (e.g., by giving them more reputation points). Users in CQA system Yahoo! Answers are divided into eight categories based on their reputation score. Each level has limitations in a number of questions and answers a user can contribute each day. Users gain and lose reputation based on their actions in the system. Similarly, the reputation schema of CQA systems on a Stack Exchange platform also works on point-based reputation rules.[1]

Besides rule-based reputation schemas applied in the existing popular CQA systems, it is possible to find several more or less simple measures of user global expertise also in the research articles concerned with CQA systems. In these approaches, mostly graph-based techniques are used to rank users by means of so-called *community expertise networks*—social networks in which nodes represent users and edges represent the flow of knowledge [Aslay et al. 2013]. There are two basic types of community expertise networks: *asker-replier network* and *asker-best answerer network*. While the first one contains the edges weighted by a number of all provided answers and ignores best answers, the second one considers only best answers and ignores other non-best answers. As these ranking-based approaches do not employ features, they are not included in Table IX, and, instead of that, the type of employed graph is depicted in an overview of methods in Table VIII.

At first, asker-replier networks were employed by link analysis approaches that are well known from tasks related to measure a web page centrality. Jurczyk and Agichtein [2007] adapted the Hyperlink-Induced Topic Search (HITS) algorithm for the purpose of user expertise estimation. Zhang et al. [2007] proposed a different algorithm, named ExpertiseRank, which is inspired by PageRank. These link analysis approaches are, however, quite computationally inefficient and achieve a level of performance very similar to much simpler metrics such as Z-score, proposed by Zhang et al. [2007]. Z-score describes how many answers and questions a user previously posted in the CQA system: $Z_{score} = \frac{a-q}{\sqrt{a+q}}$, where a represents a number of posted answers and q is a number of asked questions. The assumption is that true experts only provide answers and do not ask any questions.

Each of two basic expertise networks, used in the above-described approaches, ignores very important information about the selection of the best answer or other non-best answers, respectively. Therefore, J. Liu et al. [2011] and Aslay et al. [2013] proposed to create *competition-based expertise networks,* which combine all available information into one community expertise network. They are based on pairwise comparison of the users. The basic idea is that user expertise can be expressed relatively according to the following two assumptions: (1) the answerer who provided the best answer has higher expertise in comparison with all other non-best answerers, and (2) the best answerer has higher expertise as the asker. J. Liu et al. [2011] applied this kind of network to calculate the relative expertise levels of users by competition-based models, such as TrueSkill and the SVM model. Experiments on the dataset obtained from Yahoo! Answers confirmed that the competition-based models are able to significantly outperform

[1] http://stackoverflow.com/help/whats-reputation.

standard graph-based baseline methods, that is, HITS and PageRank. Similarly, Aslay et al. [2013] compared the performance of various link analysis methods (e.g., PageRank, HITS, InDegree) on all three kinds of expertise networks. Methods utilizing the competition-based expertise network consistently outperformed the basic expertise networks.

5.5.2. Topical Expertise and Authority Ranking. In comparison with the previous group of approaches, topical expertise and authority ranking approaches rank users according to their expertise on particular topics (rather than overall expertise). Moreover, some of these approaches consider also community feedback, which was neglected in the previous approaches. While also some global authority ranking approaches can be utilized to estimate topical expertise (e.g., by building expertise networks only with questions from the same category), these approaches consider topics explicitly.

In order to rank users according to their expertise in a particular category, Zhu, Cao, et al. [2011] exploited information not only from the target category but also from other relevant categories, which are identified by a similarity measure based on an LDA topic model. Consequently, the Topical Random Surfer (TSR) model was applied to rank users in these extended category link graphs. Cai and Chakravarthy [2013] proposed an ExpertRank framework which is able to consider besides graph structure also domain-specific information. More specifically, authors enhanced a simple asker-replier graph with edges weighted by answer quality. Experimental results revealed positive influence on ranking performance. Zhou et al. [2014] combined a graph-based PageRank with a LDA semantic model in order to take into account not only link structure but also topical similarity between askers and answerers. The proposed method achieved better performance in comparison with traditional and even the most successful competition-based models proposed by J. Liu et al. [2011].

5.5.3. Expertise Assessment and Prediction. Expertise assessment and prediction refers to an estimation of current or future value of user expertise in order to identify users with high expertise (experts) and users with low expertise.

One of the drawbacks of global authority ranking is the determination of how many users should be chosen as authoritative from the obtained ranked list. Bouguessa et al. [2008] used a simple metric named InDegree to rank users (i.e., a number of nodes that link to the node that represents the analyzed user). In the asker-best answerer network, it actually represents the number of best answers provided by the particular user. Consequently, they proposed a probabilistic approach based on a mixture model to automatically discriminate authoritative users from non-authoritative ones.

While the approach proposed by Bouguessa et al. [2008] still belongs to the graph-based approaches, subsequent approaches are rather feature based. Pal et al. [2011] explored the possibility to predict whether a user will develop to an expert user or not according to his/her behavior during the first 2 weeks in the CQA system TurboTax. Similarly, Movshovitz-Attias et al. [2013] recognized differences in initial activity of users that were used as predictors of their long-term contribution. Pal, Harper, et al. [2012] explored question selection bias to identify expert users. The authors proposed the so-called existing value of a question that represents the overall value of already provided answers on the particular question. An assumption was that experts intentionally select those questions that have only a very low existing value (with none or only a few low-quality answers) and thus they can provide a valuable contribution by answering that question. A probabilistic model that captured answerers' question selection preferences was used as a new source of features for a set of classification algorithms. Among them, bagging metaclassifier consistently provided better results in comparison with other classifiers (e.g., SVM, decision trees).

5.6. User Type

Besides user expertise, user behavior in CQA systems can be significantly diverse in terms of type, quantity, as well as quality of carried out contributions. User behavioral patterns thus represent another subject of user analyses.

5.6.1. User Profile Classification. User profile classification approaches investigate user behavioral patterns in order to assign users into various user profiles.

We are aware of only two approaches that are concerned specifically with user classification in CQA systems. Song et al. [2013] attempted to discover leading users on Quora. Authors introduced a leading capacity model, which considered three user characteristics: authority, activity, and influence. Consequently, a QP-Borda count model is employed to make a collective decision from user positions in three rankings derived for each characteristic. Furtado et al. [2013] used clustering above user behavior metrics (e.g., number of asked questions, number of days when a user was active) to classify the random sample of users into groups with similar contributions profiles. Ten contribution profiles were identified and manually labelled, for example, occasional, unskilled or expert answerer, answer activist, or hyperactivist. Consequently, a neural network was trained to classify the remaining users to the identified profiles. This approach allowed authors not only to compare composition of five various systems in Stack Exchange platform but also to study the dynamics of user behavior over time.

5.6.2. User Churn Prediction. The churn prediction problem refers to identification of users who are about to quit a CQA system. The motive for the identification of this kind of user is that acquiring new expert users is more complicated, such as by taking appropriate remedial steps to provide churning users with an extrinsic motivation.

Dror et al. [2012] as well as Pudipeddi et al. [2014] tried to predict user churn for new users (i.e., with limited previous user activity). A classifier proposed by Dror et al. [2012] revealed a number of provided answers and community feedback (e.g., a number of positive votes on provided answers) as significant signals for churn prediction. On the other side, Pudipeddi et al. [2014] identified a time gap between user subsequent posts as a significant predictor.

In the case of churn prediction, it is possible to derive ground truth directly from user history (e.g., a user can be considered as churned if he/she had no activity for more than 6 months).

5.7. Discussion

In this section, we provided the extensive overview as well as the in-depth insight into approaches on content and user modeling. In spite of a significant number of existing research articles, all defined and described groups of approaches still provide new opportunities for further research, as shown by the presence of many articles published in the past few years. Naturally, questions and answers became the interesting subject of research and thus the majority of approaches have focused exactly on their attributes so far. On the other side, approaches aimed at modeling users' attributes and behavior are still rarer while we suppose that they deserve more attention from researchers.

Among question- and answer-oriented approaches, we emphasize especially approaches that build on various specifics and phenomena present in CQA systems, such as the *homophily* between question and answer quality (e.g., Yao et al. [2015]) or the *knowledge gap* (e.g., Lin et al. [2014]). They perceive CQA systems from the wider angle and conveniently model characteristics of more than just one domain entity at the same time. In addition, we see new possible directions of research in combining of various modeling tasks (e.g., questions type and quality as in Toba et al. [2014]). These approaches have a potential not only to overcome content diversity more successfully but also to conveniently take advantage of it.

Among user-oriented approaches, we described a significant effort to model user expertise. This group of approaches is well characterized by employing many algorithms adapted from other domains (e.g., HITS and PageRank in Aslay et al. [2013] or TrueSkill in J. Liu et al. [2011]). At the same time, J. Yang, Tao, et al. [2014] pointed out a problem that is present in standard methods aimed to estimate user expertise at both a global level as well as a topical level. These methods very often misclassify very active users (denoted by authors as "sparrows") for experts (denoted as "owls"). While sparrows generate most of the content, owls provide valuable answers to questions that are perceived as important by the community. The existing expert identification methods, however, targeted mainly sparrows, as they focused mainly on the amount of users' activity in the system rather than on the quality of their contributions. As a result, these methods suffer a serious issue: The calculated estimation of user expertise does not usually reflect real users' knowledge level. The similar problem is present also in reputation schemas employed in the existing CQA systems. The negative consequences of these reputation schemas, which also favor user activity, lie in reputation abuse. As we pointed out in Section 4, exploratory studies have already confirmed increasing population of several kinds of undesired types of users. Among them, reputation collectors intentionally abuse the reputation system in order to collect reputation by answering as many questions as possible. To address these drawbacks, it is necessary to propose novel methods that balance the influence of user activity and quality of contributions. In our previous work [Huna et al. 2016], we proposed such novel reputation mechanism that focus primarily on quality and difficulty of users' contributions, and thus it can prevent undesired user behavior. However, this kind of reputation mechanism does not provide good transparency for the end users as does simple rule-based reputation schemas. Finding an optimal balance between precision (including also violence robustness) and transparency of methods for user reputation calculation provides an interesting direction for further research.

Finally, in contrast to question type, which is broadly examined, we recognized just a few articles aiming at user type classification. Nevertheless, we suppose that automatic identifying of various stereotypes of users (e.g., newcomers, users who are new to areas their questions are about) have the potential to become interesting asset to adaptive support.

6. ADAPTIVE SUPPORT

Adaptive support approaches build on results from exploratory and content/user modeling studies in order to directly influence users' collaboration. In general, CQA systems provide two main sources of knowledge: archives of already-solved questions (i.e., explicit knowledge) and community of users (i.e., tacit knowledge). Two major categories of adaptive approaches can be distinguished according to these sources of knowledge:

(1) *Question retrieval* refers to a recommendation of archived question-answer pairs that provide a user with the same information as is required to answer his/her original question.
(2) *Question routing* approaches aim to recommend the best potential answerers while taking estimations of user expertise, user activity, and user motivation into consideration.

These two categories of approaches have probably the best chance to improve the success and effectiveness of collaboration during the question-answering process. Nevertheless, adaptive support can be provided also by other additional approaches, such as *question suggestion*, *answer summarization,* or *user motivation*.

6.1. Question Retrieval

CQA systems became the repositories of very unique knowledge related to many various topics, and thus they can benefit not only from community-embedded knowledge but also from huge archives of previously answered questions. The fundamental task for reusing the content in CQA systems is to retrieve similar questions for various forms of queries [Zhang et al. 2014]. Question retrieval approaches (see Table X) can be characterized by presence of:

(1) a query profile, which represents information needs of a knowledge seeker;
(2) a question/answer profile, which describes existing questions or answers; and
(3) a matching model, which calculates similarity between these profiles.

6.1.1. Question Search. The problem of question search refers to the following task: Given a query (i.e., a question or a set of keywords), find the questions that are semantically similar to the query [Cao et al. 2008]. The question search problem was well known in FAQ data even before the emergence of CQA systems. However, CQA systems differ significantly (e.g., in the number of archived questions), and thus new methods have been proposed specifically for CQA. Question search approaches can mainly serve users in two situations: in searching a CQA archive to determine whether their question has been already asked and before asking a new question, when a CQA system can display similar questions in order to prevent posting a duplicated question.

The big challenge of question search (and also of other question retrieval tasks) is that users tend to express their equivalent information needs with different words, which finally creates a *lexical gap* during matching queries and existing posts [Zhang et al. 2014]. For this reason, the traditional IR approaches (i.e., the Vector Space Model (VSM), the BM25 model, and the Query Likelihood Language Model (QLLM)) achieve only poor performance. According to different solutions to overcome this issue, existing works can be broadly grouped under the following three topics.

At first, various advanced models are applied on question content to prevent the lexical gap. Translation Models (TM) and Translation-Based Language Models (TBLM) were confirmed to improve successfulness of question search (e.g., Cao et al. [2012], and Wu et al. [2014]). In K. Wang et al. [2009], authors identified similar questions by finding similarity in syntactic trees created for queries and questions. Similarly, J. Wang et al. [2010] proposed to utilize POS tagging and syntactic tree kernels. A completely different approach was proposed by Muthmann and Petrova [2014], who addressed identification of question duplicates as a classification task in order to decide whether two questions are topical near-duplicates or not.

Second, approaches can analyze questions' content to extent matching models with information that captures required information needs more precisely. K. Wang et al. [2010] improved question search performance by segmentation of multi-sentence questions in order to identify their fragments that can possible capture users' different information needs. Singh [2012] extended a translation model with semantic concepts (entities). In addition, L. Chen et al. [2013] and Wu et al. [2014] created intent-based models by combining translation-based models and user intent classification (see Section 5.2.1).

Finally, approaches can exploit a question context. Some models exploit topical "parallelism" between questions and their answers [Xue et al. 2008]. Jeon et al. [2005] used only a simple query likelihood language model; nevertheless, this model was able to find similarity between questions according their corresponding answers. Zhang et al. [2014] pointed out the drawback of this approach, whereby answers can be of very diverse quality and, consequently, low-quality answers can hinder a question search model, and thus the authors proposed to leverage answer quality. Besides answers,

Table X. Description of Approaches Concerning with Question Retrieval

Problem	Query Profile	Question Profile	Matching Model	Ground Truth	Eval. Metrics	Dataset	Reference
Question Retrieval							
Question Search	BoW of answers	BoW of answers	QLLM	Manual	S@n	NHN Corp. CQA 5.2K Q	[Jeon et al. 2005]
	Syntactic tree	Syntactic tree	Syntactic tree matching (STM)	Manual	P@n, MAP	Yahoo! Answers 302K Q	[K. Wang et al. 2009]
	Lexical semantics, POS tagging, Syntactic tree	Lexical semantics, POS tagging, Syntactic tree	Composite kernel	Manual	P@n, MAP	Yahoo! Answers 50K Q	[J. Wang et al. 2010]
	BoW/Syntactic tree + Sentence segmentation	BoW/Syntactic tree + Sentence segmentation	VSM/Syntactic tree matching (STM)	Manual	P, R, F1	Yahoo! Answers 0.8M Q	[K. Wang et al. 2010]
	BoW + Category classification	BoW + Category information	Category-based VSM, BM25, QLLM, TM, TBLM	Manual	P@n, MAP, MRR	Yahoo! Answers 3.1M Q	[Cao et al. 2012]
	BoW + Entity annotation	BoW + Entity annotation	Entity-based TM	Manual	R, P , P@n, MAP, MRR	Yahoo! Answers 5M Q	[Singh 2012]
	BoW + User intent classification	BoW + User intent classification	QLLM, TBLM, Intent-based LM	Manual	P@n, MAP	Yahoo! Answers, Wiki Answers 5.3M Q	[L. Chen et al. 2013]
	BoW + Category classification	BoW + Category hierarchy	BM25, QLLM, TM, TBLM + Reranking	Manual	P@n, MAP	Yahoo! Answers 4.1K Q	[Chan et al. 2014]
	Features (textual)	Features (textual)	Classification	Duplicates	P, R, F1	Stack Exchange 402K Q	[Muthmann and Petrova 2014]
	BoW + User intent classification	BoW + User intent classification	Intent-based LM	Manual	nDCG	Yahoo! Answers, Quora 128M Q	[Wu et al. 2014]
	BoW	BoW + Answer quality	Topic-based LM	Manual	P@n, R-Prec, MAP, MRR	Yahoo! Answers, Baidu Zhidao 2M Q	[X.-J. Wang et al. 2009]
Question Recommendation	Base noun phrases + WH-ngrams	Base noun phrases + WH-ngrams	MDL-based tree cut model	Manual	R-Precision, MAP, MRR	Yahoo! Answers 0.5M Q	[Cao et al. 2008]
	BoW	BoW + Question popularity	QLLM	Manual	P@n, MAP, MRR	Yahoo! Answers 17.6K Q	[Wang et al. 2011]
	BoW	BoW + LDA	Topic-enhanced TBLM	Manual	P@n, MAP, MRR, Bpref	Yahoo! Answers 4.5M Q	[Zhou et al. 2015]

(Continued)

Table X. Continued

Problem	Query/Community Profile	Question/Answer Profile	Matching Model	Ground Truth	Eval. Metrics	Dataset	Reference
Question Answering	Features (textual, statistical, user)	Features (textual, statistical, user)	L2R; GBrank (pairwise)	TREC patterns, manual	P@n, MAP, MRR	Yahoo! Answers 90K QA	[Bian et al. 2008]
	BoW	BoW	TBLM (Q) + QLLM (A)	Manual	P@n, MAP	Wondir 1M QA	[Xue et al. 2008]
	BoW	BoW + Answer quality	QLLM + Quality	Manual	P@n	Yahoo! Answers 537K QA	[Suryanto et al. 2009]
	BoW	BoW + Answer length	QA Refinement (QAR)	TREC patterns, manual	P@n, A, MAP, MRR	Yahoo! Answers 4.5M Q	[Pera and Ng 2011]
	BoW + LDA + Features (textual, similarity)	BoW + LDA + Features (textual, similarity)	VSM + Random forest	Online experiment + Manual	BAR, P, R	Yahoo! Answers 600 Q	[Shtok et al. 2012]
Web Search Answering	Features (query clarity)	Features (query-question match, answer quality)	Logistic Regression; Composite model	Manual	NDCG, Kendall's τ	Yahoo! Answers 171K QA	[Q. Liu et al. 2011]
	Features (statistical, POS tagging, syntactic roles)	Features (statistical, POS tagging, syntactic roles)	BM25/LM + POS and syntactic role scoring + SVM Rank	Web search navigation + Manual	MRR, NDCG, MAP, P@n	Yahoo! Answers 100K + 1K QA	[Carmel et al. 2014]
Group-based Question Recommendation	User ratings	Features (textual, temporal, asker)	Majority-based perceptron algorithm	Question score	Error rate of preference pairs	Yahoo! Answers 298K Q	[Sun et al. 2009]
	Topic preference	BoW	Topic-based group recommendation	Community feedback	P, R, F1	Yahoo! Answers 15K Q	[Liu et al. 2014]

category information [Cao et al. 2012] and category hierarchy [Chan et al. 2014] can be harnessed.

The majority of question search approaches were evaluated with a dataset obtained from Yahoo! Answers. Unfortunately, this dataset does not contain any information about questions' mutual similarity, and thus human assessors were asked to evaluate the correctness of the results. The only exception is approach by Muthmann and Petrova [2014], which was evaluated at datasets from a Stack Exchange platform that directly contained duplicated questions identified by the community.

6.1.2. Question Recommendation. Question recommendation (also alternatively termed as *question suggestion*) is closely related to question search. Instead of searching for semantically similar questions, it provides a recommendation of semantically related questions that reflect different aspects of the user query and provide supplementary information [Cao et al. 2008]. This kind of recommendation can be displayed besides each question to direct users to additional (but not the same) information about a topic of the currently displayed question.

Similarly, as in question search and in question recommendation, various types of language and tree-based models enhanced by context information can be employed. Cao et al. [2008] addressed question recommendation as a tree-cutting problem in which Minimum Description Length (MDL) was employed to identify the best cuts. Wang et al. [2011] extended a language model with a question popularity prediction to provide better question recommendations. Zhou et al. [2015] proposed a topic-enhanced translation-based language model that also incorporates answer information.

6.1.3. Question Answering. In comparison with question search, question answering goes one step further. Given a query submitted by a user, the goal of question answering is to return answers that provide the same information as captured by the query. This group of approaches is also commonly termed *learning-to-rank answers*, which can be, however, easily confused with answer ranking, which tackles a completely different problem of answer quality evaluation (see Section 5.4.3). The purpose of question answering is to reuse the existing answers to solve questions that remain unanswered for a long time or provide askers with immediate answers after question creation.

Bian et al. [2008] proposed GBrank, a supervised learning-to-rank method, to retrieve factual answers with textual, statistical, and user features. Xue et al. [2008] combined a translation-based language model to calculate similarity between the query and the question part, and, consequently, a query likelihood language model to calculate similarity with the answer part. Suryanto et al. [2009] introduced a framework that incorporates answer quality and answer relevance (determined by a query likelihood language model between the query and the question part only). Pera and Ng [2011] proposed a QA Refinement (QAR) system that identifies answers in two phases: question matching, which identifies existing questions and their answers with the same or similar information needs, and answer ranking, which reranks the obtained answers according to their length, similarity with the query, and their own corresponding question. Shtok et al. [2012] likewise proposed a two-stage approach. In the first step, answer candidates are identified by similarity calculation between the query and existing questions (represented by VSM). In the second step, a random forest classifier above the textual and similarity features was used to determine whether an answer candidate is suitable to answer the query or not.

Besides standard manual evaluation, other interesting solutions were proposed to obtain ground truth for question answering. Datasets containing sample questions and corresponding answer patterns from Text Retrieval Conference (TREC) were employed in Bian et al. [2008] and Pera and Ng [2011]. Moreover, Shtok et al. [2012] created three robots that automatically answer new questions in Yahoo! Answers with existing answers. Consequently, the best-answer feedback received by automatically posted answers were compared with average feedback received by answers in the same category provided by real users. The results revealed a significantly higher success rate of automatically posted answers. However, it is necessary to mention that robots provided answers only in situations where the model achieved a high degree of certainty that the existing answer is similar/appropriate to resolve a new question.

6.1.4. Web Search Answering. Web search answering can be perceived as an alternative to question answering (see Section 6.1.3), where queries are created by users in external web search tools instead of CQA systems themselves. The motivation behind this group of approaches is that existing answers in CQA systems have a potential to efficiently satisfy also information needs of users who submitted their queries to standard search engines [Q. Liu et al. 2011]. In other words, CQA sites can be used as complementary sources (or verticals) for web search systems.

This kind of web search answering scenario poses new challenges for both search engines as well as for CQA systems [Q. Liu et al. 2011]. The first attempt to evaluate the usefulness of the CQA archives for external researchers was conducted by Q. Liu et al. [2011]. The authors proposed and evaluated several approaches aimed at predicting whether a web searcher will be satisfied with existing answers from Yahoo! Answers. Three groups of features were considered (query clarity, query-question match, and answer quality) in a single regression model as well as in a composite method, which utilizes three separate regressions for each group of features individually and which outperforms a single regression model. In Carmel et al. [2014], authors analyzed a large log of web search sessions that finally landed on the Yahoo! Answers website. In order to effectively face the short length of search queries, authors proposed the term weighting method, which utilizes syntactic information for each word from the query. The relative importance of each feature was learned by the SVM Rank algorithm. Automatic evaluation on a dataset containing the actual web search queries together with visited resolved questions as well as manual evaluation by editors revealed that consideration of syntactic information led to consistent improvement.

In the cases where web searcher information need is not satisfied, web searchers can change their role to askers. In this cases, CQA systems can take advantage of the previous search session, for example, by automatic question suggestion (see Section 6.3)

6.1.5. Group-Based Question Recommendation. The last group of question retrieval approaches has a slightly different aim—to recommend answered questions to subcommunities of users. It means that a community profile, which represents users and their interests, is employed instead of a query profile. Due the massive number of questions and answers created in CQA systems every day, information overload can become a significant problem, while this kind of recommendation can help users to identify interesting content [Liu et al. 2014]. The output from recommendation can be utilized in many ways, from displaying the recommended questions in the system's dashboard to sending them as a part of newsletter.

The first approach in this category [Sun et al. 2009] attempts to recommend questions by predicting how likely a question is to be recommended by most users (i.e., question popularity). To predict the popularity of this question, the authors proposed a new algorithm, called the majority-based perceptron algorithm, which is able to learn users' preferences from their previous ratings. In the prediction, features available at question creation time are used (i.e., textual, temporal and asker-related). Liu et al. [2014] proposed a group-based recommendation. This approach at first generates community profiles by aggregating previously posted questions and answers with correspondence to the weight of each community member (e.g., his/her reputation). Afterwards, the recommended questions are selected by considering relevance as well as complementarity between knowledge captured in the recommended questions and in the community profile. In both approaches, community feedback was considered as ground truth.

6.2. Question Routing

One of the most important goals of CQA systems is to provide an asker with a suitable answer in the shortest possible time. However, there are still questions that remain unanswered for a long time. The problem of a high rate of unanswered questions is increasing in CQA systems because of the growing number of newly posted questions each day. Question routing tries to solve this problem by recommending questions to potential answerers who are most likely to provide a satisfying answer. This type of recommendation in CQA systems is sometimes alternatively termed *answerer recommendation*, *expert finding*, or even *question recommendation*, which causes ambiguity with question retrieval (see Section 6.1.2).

In contrast to other categories of CQA approaches, question routing has been already described in a survey by Furlan et al. [2013]. Authors introduced a presentation paradigm that provides a detailed view on implementation of several selected approaches. In our survey, we provide a more abstract overview of various question routing solutions also within the context of other CQA approaches and including also the most recent works.

The problem of question routing can be formalized as follows: given a newly posted question q, we need to create an ordered list of top k users $u_1, u_2, \ldots, u_k$ who are the most suitable to answer question q. This list is usually ordered by a probability that user u would answer given question q. To obtain the list of suitable answerers, it is necessary to solve three sub-problems [Guo et al. 2008] that also characterize question routing approaches (see Table XI):

(1) construction of a question profile, which represents question's topics;
(2) construction of a user profile, which represents user expertise/interest and optionally also additional characteristics (e.g., authority or availability);
(3) matching between profile of a new question and all relevant user profiles.

Question routing differs from user global/topical expertise ranking and expertise assessment (see Section 5.5) because of its explicit orientation to identify possible answerer-candidates for a particular new question instead for a whole system or a topic. The suitable level of user expertise/interest on the question topic definitely represents an important condition to provide a suitable answer, but also additional aspects that influence answer suitability and users' willingness to provide an answer can be considered.

(1) At first, some approaches (e.g., Liu et al. [2010] and Yang et al. [2013]) consider user authority as a good predictor, as authoritative authors will probably give more authoritative answers.
(2) Users may also differ considerably in the degree of their overall activity [Liu et al. 2010]. Some users can be very active at the beginning and, afterwards, they can be silent for a long time. User interest can die away due to different reasons and circumstances, such as a user can lose interest in the particular topic or he/she does not have time to participate in question answering any more. Therefore, users with a high level of (recent) activity are more likely to provide answers on new questions.
(3) Finally, sometimes potential answerers have necessary knowledge and are even active enough to answer the routed question, but they are not motivated to do that. Therefore, a few of the existing approaches (e.g., Luo et al. [2014]) consider also

Table XI. Description of Approaches Concerning with Question Routing

Problem	Question Profile	User Profile	Matching Model	Ground Truth	Eval. Metrics	Dataset		Reference
Question Routing								
Language Model-based Question Routing	BoW	Expertise (BoW)	QLLM, RM, Cluster-based LM	Actual answerers	MRR	Wondir	852K A	[Liu et al. 2005]
	BoW	Expertise (BoW + A quality) + Availability	QLLM + Jelinek-Mercer smoothing	Actual answerers	MRR	Yahoo! Answers	19K Q	[Li and King 2010]
	BoW + Category information	Expertise (BoW + Category)	Category-sensitive LM	Actual answerers	P@n, MAP, MRR	Yahoo! Answers	433K Q	[Li et al. 2011]
	BoW + Category information	Expertise (BoW + Category + A Quality)	TM	Actual answerers	P@n, MRR	Yahoo! Answers	359K Q	[G. Zhou, Liu, et al. 2012]
	BoW + Category information	Expertise (BoW + Category + A Quality + Temporal) + Authority	Hybrid method	Manual	P@n, MAP, MRR	Yahoo! Answers	200 Q	[Liu et al. 2013]
Topic Model-based Question Routing	BoW + UQA topic model	Expertise (BoW + UQA topic model)	Term (BM25) + topic (UQA) similarity	Actual answerers	P@n	Yahoo! Answers	119K Q	[Guo et al. 2008]
	BoW	Expertise (BoW + LDA) + Authority + Activity	Term (QLLM) + topic (LDA) similarity	BA answerer	P@n	Iask	963K Q	[Liu et al. 2010]
	BoW	Expertise (PLSA)	Topic similarity	BA answerer	A	Yahoo! Answers	246K Q	[Xu et al. 2012]
	BoW	Expertise (BoW, LDA, STM)	Cosine similarity, QLLM, topic similarity	BA answerer	S@n	Stack Overflow	119K Q	[Riahi et al. 2012]
	BoW + LDA topic model + Category information	Expertise (BoW/LDA + Category + Temporal)	Vector similarity (dot-product) + Diversification	Online experiment (A/B test)	Activity level	Yahoo! Answers	119K U	[Szpektor et al. 2013]
	BoW	Expertise (TEM topic model) + Authority	Topic similarity	Actual answerers	nDCG, Pearson, Kendall's τ	Stack Overflow	8.9 K Q	[Yang et al. 2013]
	BoW	Expertise + Descriptive ability (UTAM) + Social links (USTA)	Topic similarity	Actual answerers + A score	nDCG, Spearman, Kendall's τ	Stack Overflow	9.1 K Q	[Yang 2014]

(Continued)

Table XI. Continued

Problem	Question Profile	User Profile	Matching Model	Ground Truth	Eval. Metrics	Dataset	Reference
Classification- and Ranking- based Question Routing	Features (textual, category, user, bias)	Features (question-driven, explicit relations, bias)	Classification; Multi-Channel Recommender (2 classes)	BA answerer	A, AUC	Yahoo! Answers 1.3M Q	[Dror et al. 2011]
	Features (textual, Q-U relationship)	Features (activity, expertise)	Classification + Ranking; SVM (2 classes) + SVM Rank (pairwise)	Actual answerers	P@n, MAP, MRR	Stack Overflow 92 K Q	[Ji and Wang 2013]
	Features (topic - sLDA model)	Features (expertise - sLDA topic model)	Ranking; RankSLDA (pairwise)	Actual answerers + A score	P@n, nDCG, MAP, MRR	Cross Validated (Stack Exchange) 22 K Q	[San Pedro and Karatzoglou 2014]
	Features (Q type, topic - BoW)	Features (expertise, motivation, availability)	Classification; Linear regression + Diversification	Actual answerers	A	IBM Connections 2.9 K Q	[Luo et al. 2014]
	Features (topic - BoW)	Features (expertise, social relationships)	Graph Regularized Matrix Completion	Actual answerers + BA answerer	P@n, nDCG, MRR, A	Quora 444K Q	[Zhao et al. 2015]
Collaborative Question Routing	Features (textual, topic - LDA, LSA)	Features (activity, expertise, non-QA)	Clustering; k-NN	Actual community	P@n, MRR	IBM Connections 60K Q	[Pal et al. 2013]
	Features (temporal, topic - LDA, keywords)	Features (expertise, availability, user compatibility)	Greedy algorithm	Actual answerers	P@n, R@n, MSC	Stack Overflow 162K Q	[Chang and Pal 2013]

user motivation or users' mutual familiarity as another aspect, which affects the question routing process.

To obtain the ground truth of appropriate answerers for a particular question, the majority of previous studies (e.g., Liu et al. [2010], Guo et al. [2008], and Zhu, Chen, et al. [2011]) utilized either the best answerer or the list of all users who actually provided an answer on this question. We would like to point out that this approach does not completely correspond to the real interest from users. For example, as soon as a question receives at least one high-quality answer, other suitable candidates can express their expertise just by providing a positive vote or simply by skipping the routed question and attempting to answer another one (unfortunately, as voting and question views are usually anonymous, CQA datasets do not provide this kind of information). In spite of this drawback, the list of actual answerers can be still considered as fair precise ground truth for question routing. Other less-common evaluation methods include online experiments (e.g., in Szpektor et al. [2013]) or manual assessment (e.g., in Liu et al. [2013]).

We divide analyzed question routing approaches according to various types of matching models following the classification proposed in Li [2014]. In addition, we recognized a problem of collaborative question routing, which we address separately.

6.2.1. Language Model-Based Question Routing. The first group of approaches to question routing is based on language models. These approaches represent both question and user profiles as a bag of words (the user profile is created from all questions the corresponding user previously answered or asked). Consequently, user profiles are ranked according to various kinds of language models that calculate a probability that user profiles will generate terms of the routed question.

At first, three traditional language models well known from the information retrieval domain were employed to rank the user profiles [Liu et al. 2005]: QLLM, RM (Relevance Model), and a cluster-based language model. Similarly, a query likelihood language model was used in a question routing framework proposed by Li and King [2010]. It considers not only user expertise (derived with answer quality information) but also user activity (a prediction whether an answerer will respond to the request in short time). An experiment on a dataset obtained from a Yahoo! Answer portal confirmed that taking answer quality and user activity into consideration can significantly improve the recommendation. Later, Li et al. [2011] proposed to include the routed question category in user expertise estimation. In the basic category-sensitive language model, authors considered two questions related only if they are assigned to the same category. In a more advanced version termed the transferred category-sensitive language model, a more sophisticated approach was used to calculate the similarity between categories: If there are many answerers who contribute to two categories frequently, then these two categories should be similar to each other.

In the above-described traditional language models, data sparseness can lead to word mismatch between the routed question and the user profiles, which can be caused by co-occurrence of random words in user profiles or questions [G. Zhou, Liu, et al. 2012]. This problem is solved by translation models (TM) that employ statistical machine translation to overcome data sparseness and that are able to differentiate between exact matched words and translated semantically related ones. Finally, Liu et al. [2013] proposed a hybrid method that combines all kinds of information employed in the previous approaches: user expertise (calculated with answer, category, and temporal information), reputation, and authority score.

6.2.2. Topic Model-Based Question Routing. Traditional language models are based on exact word matching, and thus they are not able to capture more advanced semantics and solve the problem of the lexical gap between the posted question and user profiles [Liu et al. 2010]. As a solution to this issue, latent topic models, such as Probabilistic Latent Semantic Analysis (PLSA) or LDA, are employed to consider not only syntactic but also semantic similarities.

The first approaches can be considered a transition between traditional language models and more advanced semantic models. Guo et al. [2008] combined the topic and the term levels to represent questions' focus and answerers' expertise. To discover latent topics, a novel generative model named UQA was proposed. Consequently, matching was performed by a BM25 language model at the term level and by a topical matching method at the topic level. In a similar way, besides a QLLM model, an LDA model was employed in Liu et al. [2010].

Guo et al. [2008] created user profiles by concatenation of all asked and answered questions; similarly, Liu et al. [2010] concatenated all previously answered questions. This solution, however, has two significant drawbacks. At first, Xu et al. [2012] pointed out that users in CQA systems play two different roles simultaneously: an asker and an answerer. While answering a question can be perceived as an expression of expertise on question topics, asking a question, on the other side, can be perceived as a lack of this expertise. Experimental evaluation revealed that considering only answerer role leads to better question routing performance, while mixing these two roles can even impair recommendation results. Second, when only answered questions are concatenated to one document, the LDA model cannot take advantage of the internal structure of user profiles, as each answered question can relate to a different topic. Riahi et al. [2012] proposed a Segmented Topic Model (STM), which can discover the hierarchical structure of topics, and thus, instead of grouping all user's questions under one topic, it allows each question to have a different topical distribution.

In the above-mentioned works, latent topics and user expertise were modelled separately. Yang et al. [2013] introduced a Topic Expertise Model (TEM) to jointly model topics and expertise (besides content, also tags and answer quality was considered). Authors consequently proposed a CQArank framework that combines user topical expertise estimation from the TEM model and user authority derived from link analysis above a QA graph. The TEM model, however, did not consider the two-role perspective introduced by Xu et al. [2012] as it derived user expertise from questions and answers simultaneously. Yang [2014] addressed this drawback and proposed a User Topical Ability Model (UTAM) that models separately users' descriptive ability (i.e., ability to ask good questions) and user expertise. Consequently, authors integrated the results from the UTAM model with social link analysis and created probably the most comprehensive topic model named User Social Topic Ability (USTA).

A different approach was proposed in Szpektor et al. [2013], who represented question topics with three vectors based on an LDA model, a lexical model, and a category model. In contrast to the previous approaches, vector similarity (calculated by a dot-product) was employed in the matching phase and, in addition, a diversification was employed to prevent a well-known recommendation problem: filter-bubble.

6.2.3. Classification- and Ranking-Based Question Routing. The third group of approaches tackles question routing in comparison with previous language and topic models as a classification problem, a ranking problem, or a missing value estimation. Question and user profiles are represented as a set of features that can provide important information that is not fully utilized in the language and topic models as they consider primarily only text similarities [Ji and Wang 2013].

Dror et al. [2011] addressed question routing as a classification task whether a particular question will be interesting for a user or not. A set of different question- and answerer-related features supplemented with bias features were considered in a multi-channel recommender system (e.g., a question with several answers is less attractive as completely unanswered question). Ji and Wang [2013] proposed a ranking-based approach consisting of two methods, SVM and RankSVM, which utilized a set of features describing a question, a user, and a question-user relationship (i.e., similarities calculated by QLLM and LDA models). As a typical SVM classifier provides only a binary classification, authors decided to use the probability estimation function, which is able to produce the probability of a user having enough expertise to provide a suitable answer, which was finally used to rank potential answerers. San Pedro and Karatzoglou [2014] proposed an algorithm named RankSLDA—a learning-to-rank extension to supervised LDA applied specifically to solve the question routing problem. Zhao et al. [2015] considered question routing from the perspective of missing value estimation. For estimation of missing values in the rating matrix, authors employed a method called *graph regularized matrix completion*. Besides user expertise, social relations between users were considered to obtain recommendations.

The approach proposed by Luo et al. [2014] differs from the above-described ones in several aspects. At first, authors focused on the enterprise CQA system IBM Connections rather on an open CQA system. The proposed question routing method considered employees' expertise, willingness (e.g., amount of previous activity) and readiness (e.g., recent activity, current load) determined from the CQA system itself as well as from non-QA data sources. In addition, the method purposefully diversifies recommended answerers with less-active employees to grow the pool of answerers.

6.2.4. Collaborative Question Routing. In contrast to the previous question routing approaches, collaborative question routing refers to a recommendation of recently posted questions to a group of compatible users, who would collaborate together to create a content with lasting value, instead of to a top-k most desirable experts [Chang and Pal 2013].

To solve the collaborative question routing problem, k-NN clustering [Pal et al. 2013] and a greedy algorithm [Chang and Pal 2013] were employed. In both cases, features describing a question, users, and their mutual similarity/compatibility were considered.

6.3. Other Adaptive Support Approaches

Besides question retrieval and question routing, additional adaptive approaches proposed for CQA systems exist. In this section, we provide their brief overview.

Question Suggestion. Internet users naturally direct their information needs to search engines. In spite of developed mechanisms, which have been proposed already to support web search, there are still many cases in which web search ends with failure. However, many of these unsatisfied searches could be addressed by CQA systems [Liu et al. 2012]. Asking a question in a CQA system after an unsuccessful web search is a process in which users change their role from searchers to askers. Liu et al. [2012] performed an exploratory study of users' behavior to understand better this transition. The results point out that users coming from search engines behave quite differently in CQA systems. For example, they are not so patient while waiting for the best answer in comparison with other CQA users, and thus CQA systems should promote even more questions asked by searchers.

The main focus of adaptive approaches to support the transition from search engines to CQA systems is how to map search queries to appropriate questions in natural language. At first, it is possible to automatically identify the most suitable questions that have been already posted in a CQA system and thus navigate web-searchers easily to their answers [Gao et al. 2013]. Another question suggestion approach suggests synthetic questions based on keywords provided in the original search query [Dror, Maarek, Mejer, et al. 2013].

Answer Summarization. Answer summarization takes question answering (see Section 6.1.3) one step further. Especially in the case of multi-sentence questions (i.e., questions that comprise several sub questions), even the best answers can be incomplete while the remaining answers can provide additional aspects [Chan et al. 2012]. Answer summarization aims to summarize the best answer together with other answers to provide users with one coherent answer.

Approaches that are concerned with the answer summarization problem in CQA systems take different metadata into consideration. As open and opinion-based questions usually have multiple good answers, Yuanjie Liu et al. [2008] considered a question type as a guide to answer summarization based on the clustering algorithm. Tomasoni and Huang [2010] combined quality, coverage, relevance, and novelty to generate trustful, complete, and relevant answers. X. Liu et al. [2011] extracted sentences from answers according to their salience score, which is assigned by a graph-based random walk model incorporating user social features and answers' content. Chan et al. [2012] introduced a Conditional Random Fields (CRF) framework with group L_1-regularization, which addressed answer summarization as a sequence labeling process: Each answer sentence is labeled as a summary or a non-summary sentence and, consequently, the summarization is created by concatenation of sentences with the summary label. Pande et al. [2013] proposed to create summaries by creating graphs using various answer-level, sentence-level, and similarity features in which nodes represent particular answers. Consequently, summarization is created by finding a path in this graph that covers the most important information.

User Motivation. Finally, adaptive approaches can contribute to keeping users motivated and dedicated to CQA communities and even steer users towards certain types of actions. One typical example of how to motivate people in online communities is gamification. CQA systems employ several kinds of gamification techniques but mainly badges (e.g., Stack Exchange) and reputation points (e.g., Yahoo! Answers).

We are aware of only one article by Anderson and Huttenlocher [2013], which addressed steering user behavior directly in CQA systems. At first, authors studied how badges can influence user behavior and created a formal model that can predict how badge assignments affect users' contributions in Stack Overflow system. Second, the authors examined the badge placement problem (i.e., identification of optimal amounts of activities when badges should be assigned to users) and proposed several high-level design principles that are supposed to help system designers create good gamification mechanisms.

6.4. Discussion

Similarly to content and user modeling approaches, adaptive support is also characterized by several topics that are popular among researchers and some that are rather neglected. While question retrieval and question routing represent a large body of the existing research, some emerging problems have not been addressed in these approaches so far. In our previous study [Srba and Bieliková 2016], we pointed out two emerging problems that hamper a long-term sustainability of Stack Overflow: an increasing failure and churn rate, which are caused especially by an increasing amount of low-quality content created by undesired group of users—help vampires, noobs, and reputation collectors. The majority of existing approaches focused on providing adaptive support in CQA systems even unintentionally support these undesired groups of users [Srba and Bieliková 2016]. The main reason for this is that the existing approaches can be characterized as asker oriented (i.e., they primarily focus on askers while preferences and expectations of answerers are suppressed) and expert oriented (i.e., only a small proportion of highly active experts are involved). Question routing can be provided as a typical example. Most existing methods recommend new questions to users with a high level of expertise regardless of real question difficulty. Consequently, experts are easily overloaded while capacity of other users is left unutilized. This attitude to question routing is really successful in achieving askers' goals (to receive a high-quality answer), although it completely overlooks those experts who prefer to answer more difficult and challenging questions within their limited time capacities. In addition, this attitude ignores other users who are involved only very rarely; nevertheless, they might be interested in getting recommendations.

Just a few of recent existing articles have already reflected these problems (e.g., exploratory analyses of content abusers [Kayes et al. 2015] or diversification applied to question routing [Szpektor et al. 2013]). We suggest that it will be necessary to propose a new family of adaptive support methods that are more answer oriented and in addition involve the whole community in the question-answering process. We have already illustrated this kind of approach on the novel question routing method [Srba et al. 2015]. It can consider non-QA data sources about answerers in order to estimate users' expertise early and more precisely and thus recommend questions also to newcomers and lurkers (users with a low level of QA activity).

Many other areas are also worth deeper examination, such as personalized recommendation of solved questions to individuals (so far only group-based recommendation approaches have been proposed, see Section 6.1.5) or better preservation of motivation, for example, with gamification mechanisms such as reputation score or badges.

As we stated in the conclusion of exploratory studies (see Section 3), another interesting potential of CQA systems lies in their application in various environments and contexts. These exploratory studies pointed out that these environments differ significantly from open CQA communities and thus the first domain-specific adaptive support approaches have already appeared: question recommendation in the enterprise CQA system IBM Connections [Luo et al. 2014] or in Massive Open Online Courses (MOOC) systems [D. Yang et al. 2014]. We suppose that educational CQA systems in particular present an interesting topic for additional research. The existing approaches are not applicable here because additional constraints should be satisfied in this kind of environment [D. Yang et al. 2014] (e.g., a student cannot be overloaded with too many questions or difficulty of recommended questions should match student's knowledge level). Therefore, proposing novel methods for appropriate adaptive support (mainly question routing and question retrieval) represents a promising direction for future work.

In addition, we recognized the potential of CQA systems to be used more intensively in mobile environments, where user spatial, temporal, and social context plays a significant role. In many cases, a user wants to use his/her smartphone or tablet to ask a question that is specific for particular location and time. As an illustrative example, we can provide a question asked at Yahoo! Answers[2] with a title: "*Good coffee house in Houston (preferably with live music)*?" It is clear that this kind of question needs to be treated in CQA systems differently from the majority of questions. During question routing, it should be recommended only to a specific subgroup of people who are familiar with a particular locality, and, in addition, it should be answered fast as the answer is useful only if delivered in an acceptable time limit. Similarly, during question retrieval, the provided answers on this kind of questions are not usually suitable to answer new questions that are related to different location or time. Above all these specifics, mobile-based question answering requires also novel user interfaces that are specifically designed with respect to limitations as well as advantages of mobile touch screens. As a result, CQA systems can be definitely used to answer highly contextual

[2]https://answers.yahoo.com/question/index?qid=20151123183553AAg1XfM.

dependent questions in mobile environments, but it is necessary to provide users with appropriate adaptive collaboration support that will take these contextual dependencies into consideration. The first approaches following this direction has just appeared (e.g., a real-time question-answering system *RealQA* proposed by Liu et al. [2015]).

7. CONCLUSION

Community and Collaborative Question Answering (CQA) systems, such as Yahoo! Answers, Quora or Stack Overflow, represent a substantial source of knowledge in the current web. They conveniently supplement traditional search engines in answering complex, subjective, or conversational questions. CQA systems have become a valuable part of the current web and have also become an interesting subject of many research articles that tackle with a variety of different aspects. In spite of their heterogeneity, the research on CQA systems have missed a comprehensive survey and categorization. As a result, orientation in this domain and determination of state-of-the-art approaches was quite challenging. In addition, the existing articles use ambiguous and inconsistent terminology, as the same problems are commonly termed differently, and, on the other hand, some terms refer to several different problems.

To address these issues, we performed an extensive survey and review of 265 research studies in order to precisely evaluate the state of the art in this domain. The main purpose of the survey was to make the orientation in the research on CQA systems more convenient for researchers who only begin with the research in this area as well as who already concern with CQA systems. The outcomes from our survey are, therefore, multiple. We proposed the general descriptive framework, which depicts the distinguishing attributes of CQA approaches, and the complex three-level classification. Both the framework and the categorization can be easily adapted to the future evolution of CQA research. For example, if a new group of approaches will emerge (as a respond to a new task/problem in CQA systems), it will be possible to easily integrate it to the proposed classification. Finally, by means of the proposed descriptive framework and categorization, we also provide the extensive description of the existing approaches.

Besides current trends in CQA research, we emphasized in the survey also new possibilities for future research. In order to summarize these possibilities, we can group them according to two major phenomena, which can provide directions for additional research in the existing tasks and open problems solved in CQA systems: *preservation of a long-term sustainability* and *exploiting transferability of CQA systems*.

Current CQA systems have witnessed an emergence of new problems that have recently started to appear as a result of the increasing number of users and questions posted every day. The most serious problems, which can significantly hamper the long-term sustainability of CQA systems, include a constantly growing amount of low-quality content, passive users, or even users who purposefully abuse the question-answering process. In order to prevent this undesired evolution, we stress that additional exploratory studies, content, and user modeling as well as adaptive support approaches should be conducted. In our previous work, we already started to examine this phenomenon in more detail as we conducted the exploratory case study on recent evolution of community and content on Stack Overflow [Srba and Bieliková 2016]. Community perception as well as data analyses showed that the emerging problems are highly related to the growing amount of low-quality content created by undesired groups of users (i.e., help vampires, noobs, and reputation collectors). In order to face the emerging problems, we proposed the following: (1) to improve the existing motivation mechanisms based on various gamification principles in order to reflect the real value of users' contributions more precisely (as illustrated in Huna et al. [2016]) and (2) to provide users with novel answerer-oriented adaptive support that, in addition, involves a whole community in question answering (as illustrated in Srba et al. [2015]). These approaches represent an eminent attitude change in the existing question-answering support methods with the aim to preserve the long-term sustainability of CQA ecosystems.

Second, we are witnessing growing interest in the application of CQA concepts in various additional environments, such as in the educational domain, in business (e.g., in crowd-based support tools or inside software applications), or on mobile devices. In our previous work [Srba and Bielikova 2015], we investigated mainly transferability of CQA systems into the education domain and proposed the novel concept of an organization-wide educational CQA system and its implementation by means of the system Askalot. Nevertheless, the research possibilities of these untraditional environments are much wider.

Besides the open problems and tasks that already have been addressed in the existing research articles, CQA systems certainly have the potential to be utilized in many new scenarios that will naturally make novel open problems appear.

In the beginning, it is possible to utilize the tremendous collective intelligence of large CQA communities for many additional crowdsourcing tasks. Stack Overflow can be taken as an example. In September 2015, an expansion of the system to incorporate software documentation was introduced.[3] The incentive for this idea is that official documentation of many software products is commonly infrequently updated and incomplete. In these cases, users on Stack Overflow can mutually build alternative documentation artifacts by employing a bottom-up approach. Consequently, they can benefit from documentation accompanied by a number of related questions, and, on the other hand, it will be possible to get rid of many repeatedly asked duplicated questions concerning with the same issues resulting from insufficient official documentation. These kinds of various extensions of the standard question-answering process are applicable not only in Stack Overflow but also in other domain-specific as well as general CQA systems. As a result, they will provide new and interesting possibilities on how to scaffold users' collaboration.

Second, large CQA datasets containing millions of answered questions can be employed also in additional research domains. Especially in natural language processing, it would be possible to improve automatic question answering by utilizing knowledge embedded in the archived question-answer pairs or to just use text of questions and answers as a dataset for name entity recognition, word sense disambiguation, sentiment analyses, and so on.

ELECTRONIC APPENDIX

The electronic appendix for this article, which contains a complete overview of all analyzed approaches (i.e., also with approaches that are not included in the survey due to space restrictions), can be accessed in the ACM Digital Library.

REFERENCES

Mark S. Ackerman and David W. McDonald. 1996. Answer Garden 2: Merging Organizational Memory with Collaborative Help. In *Proc. of the 1996 ACM conference on Computer Supported Cooperative Work-CSCW'96.* ACM Press, New York, NY, 97–105. DOI:http://dx.doi.org/10.1145/240080.240203

Lada A. Adamic, Jun Zhang, Eytan Bakshy, and Mark S. Ackerman. 2008. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In *Proceeding of the 17th International Conference on World Wide Web-WWW'08.* ACM Press, New York, NY, 665–674. DOI:http://dx.doi.org/10.1145/1367497.1367587

Eugene Agichtein, Carlos Castillo, Debora Donato, Aristides Gionis, and Gilad Mishne. 2008. Finding High-Quality Content in Social Media. In *Proc. of the International Conference on Web Search and Web Data Mining-WSDM'08.* ACM Press, New York, NY, 183–194. DOI:http://dx.doi.org/ 10.1145/1341531.1341557

[3]http://meta.stackoverflow.com/questions/303865.

Eugene Agichtein, Yandong Liu, and Jiang Bian. 2009. Modeling Information-Seeker Satisfaction in Community Question Answering. *ACM Trans. Knowl. Discov. Data* 3, 2 (April 2009), 1–27. DOI:http://dx.doi.org/10.1145/1514888.1514893

Ashton Anderson and Daniel Huttenlocher. 2013. Steering User Behavior with Badges. In *Proc. of the 22nd International Conference on World Wide Web-WWW'13.* International World Wide Web Conferences Steering Committee, 95–105.

Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2012. Discovering Value from Community Activity on Focused Question Answering Sites: A Case Study of Stack Overflow. In *Proc. of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD'12.* ACM Press, New York, NY, 850–858. DOI:http://dx.doi.org/10.1145/2339530.2339665

Chulakorn Aritajati and N. Hari Narayanan. 2013. Facilitating Students' Collaboration and Learning in a Question and Answer System. In *Proc. of the 2013 Conference on Computer Supported Cooperative Work Companion-CSCW'13.* ACM Press, New York, NY, 101–106. DOI:http://dx.doi.org/10.1145/2441955.2441983

Muhammad Asaduzzaman, Ahmed Shah Mashiyat, Chanchal K. Roy, and Kevin A. Schneider. 2013. Answering Questions about Unanswered Questions of Stack Overflow. In *Proc. of 10th Working Conference on Mining Software Repositories-MSR'13.* IEEE, 97–100. DOI:http://dx.doi.org/10.1109/MSR.2013.6624015

Çiğdem Aslay, Neil O'Hare, Luca Maria Aiello, and Alejandro Jaimes. 2013. Competition-Based Networks for Expert Finding. In *Proc. of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval-SIGIR'13.* ACM Press, New York, NY, 1033–1036. DOI:http://dx.doi.org/10.1145/2484028.2484183

Kyoungman Bae and Youngjoong Ko. 2012. An Effective Category Classification Method Based on a Language Model for Question Category Recommendation on a CQA service. In *Proc. of the 21st ACM International Conference on Information and Knowledge Management-CIKM'12.* ACM Press, New York, NY, 2255–2258. DOI:http://dx.doi.org/10.1145/2396761.2398614

Jiang Bian, Yandong Liu, Eugene Agichtein, and Hongyuan Zha. 2008. Finding the Right Facts in the Crowd: Factoid Question Answering over Social Media. In *Proceeding of the 17th International Conference on World Wide Web-WWW'08*. ACM Press, New York, NY, 467. DOI:http://dx.doi.org/10.1145/1367497.1367561

Jiang Bian, Yandong Liu, Ding Zhou, Eugene Agichtein, and Hongyuan Zha. 2009. Learning to Recognize Reliable Users and Content in Social Media with Coupled Mutual Reinforcement. In *Proc. of the 18th International Conference on World Wide Web-WWW'09.* ACM Press, New York, NY, 51–60. DOI:http://dx.doi.org/10.1145/1526709.1526717

Mohan John Blooma, Alton Y. K. Chua, and Dion Hoe-Lian Goh. 2008. A Predictive Framework for Retrieving the Best Answer. In *Proc. of the 2008 ACM Symposium on Applied Computing-SAC'08.* ACM Press, New York, NY, 1107–1111. DOI:http://dx.doi.org/10.1145/1363686.1363944

Mohan John Blooma, Alton Yeow-Kuan Chua, and Dion Hoe-Lian Goh. 2010. Selection of the Best Answer in CQA Services. In *Proc. of 7h International Conference on Information Technology: New Generations.* IEEE, 534–539. DOI:http://dx.doi.org/10.1109/ITNG.2010.127

Amiangshu Bosu, Christopher S. Corley, Dustin Heaton, Debarshi Chatterji, Jeffrey C. Carver, and Nicholas A. Kraft. 2013. Building Reputation in StackOverflow: An Empirical Investigation. In *Proc. of the 10th Working Conference on Mining Software Repositories-MSR'13.* Piscataway, NJ, USA: IEEE Press, 89–92. DOI:http://dx.doi.org/10.1109/MSR.2013.6624013

Mohamed Bouguessa, Benoît Dumoulin, and Shengrui Wang. 2008. Identifying Authoritative Actors in Question-Answering Forums-The Case of Yahoo! Answers. In *Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD'08.* ACM Press, New York, NY, 866–874. DOI:http://dx.doi.org/10.1145/1401890.1401994

Andrei Broder. 2002. A taxonomy of web search. *ACM SIGIR Forum* 36, 2 (September 2002), 3. DOI:http://dx.doi.org/10.1145/792550.792552

Fan Bu, Xingwei Zhu, Yu Hao, and Xiaoyan Zhu. 2010. Function-based question classification for general QA. In *Proc. of the 2010 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA, USA: Association for Computational Linguistics, 1119–1128.

Grégoire Burel and Yulan He. 2013. A Question of Complexity – Measuring the Maturity of Online Enquiry Communities. In *Proc. of the 24th ACM Conference on Hypertext and Social Media-HT'13.* ACM Press, New York, NY, 1–10. DOI:http://dx.doi.org/10.1145/2481492.2481493

Grégoire Burel, Yulan He, and Harith Alani. 2012. Automatic Identification of Best Answers in Online Enquiry Communities. In *Proc. of 9th Extended Semantic Web Conference-ESWC'12.* 514–529. DOI:http://dx.doi.org/10.1007/978-3-642-30284-8_41

Li Cai, Guangyou Zhou, Kang Liu, and Jun Zhao. 2011. Large-scale question classification in CQA by leveraging Wikipedia semantic knowledge. In *Proc. of the 20th ACM International Conference on Information and Knowledge Management-CIKM'11*. ACM Press, New York, NY, 1321–1330. DOI:http://dx.doi.org/10.1145/2063576.2063768

Yuanzhe Cai and Sharma Chakravarthy. 2013. Expertise Ranking of Users in QA Community. In *Proc. of 18th International Conference on Database Systems for Advanced Applications-DASFAA'13*. 25–40. DOI:http://dx.doi.org/10.1007/978-3-642-37487-6_5

Xin Cao, Gao Cong, Bin Cui, Christian S. Jensen, and Quan Yuan. 2012. Approaches to Exploring Category Information for Question Retrieval in Community Question-Answer Archives. *ACM Trans. Inf. Syst.* 30, 2 (May 2012), 1–38. DOI:http://dx.doi.org/10.1145/2180868.2180869

Yunbo Cao, Huizhong Duan, Chin-Yew Lin, Yong Yu, and Hsiao-Wuen Hon. 2008. Recommending questions using the mdl-based tree cut model. In *Proceeding of the 17th International Conference on World Wide Web-WWW'08*. ACM Press, New York, NY, 81–90. DOI:http://dx.doi.org/10.1145/1367497.1367509

David Carmel, Avihai Mejer, Yuval Pinter, and Idan Szpektor. 2014. Improving Term Weighting for Community Question Answering Search Using Syntactic Analysis. In *Proc. of the 23rd ACM International Conference on Conference on Information and Knowledge Management-CIKM'14*. ACM Press, New York, NY, 351–360. DOI:http://dx.doi.org/10.1145/2661829.2661901

Brendan Cleary, Carlos Gomez, Margaret-Anne Storey, Leif Singer, and Christoph Treude. 2013. Analyzing the Friendliness of Exchanges in an Online Software Developer Community. In *Proc. of 6th International Workshop on Cooperative and Human Aspects of Software Engineering-CHASE'13*. IEEE, 159–160. DOI:http://dx.doi.org/10.1109/CHASE.2013.6614756

Denzil Correa and Ashish Sureka. 2013. Fit or unfit: Analysis and Prediction of "Closed Questions" on Stack Overflow. In *Proc. of the first ACM Conference on Online Social Networks-COSN'13*. ACM Press, New York, NY, 201–212. DOI:http://dx.doi.org/10.1145/2512938.2512954

Denzil Correa and Ashish Sureka. 2014. Chaff from the Wheat: Characterization and Modeling of Deleted Questions on Stack Overflow. In *Proc. of the 23rd International Conference on World Wide Web-WWW'14*. ACM Press, New York, NY, 631–642. DOI:http://dx.doi.org/10.1145/2566486.2568036

Daniel Hasan Dalip, Marcos André Gonçalves, Marco Cristo, and Pavel Calado. 2013. Exploiting User Feedback to Learn to Rank Answers in Q&A Forums: a Case Study with Stack Overflow. In *Proc. of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval-SIGIR'13*. ACM Press, New York, NY, 543–552. DOI:http://dx.doi.org/10.1145/2484028.2484072

David Dearman and Khai N. Truong. 2010. Why Users of Yahoo! Answers Do Not Answer Questions. In *Proc. of the 28th International Conference on Human Factors in Computing Systems-CHI'10*. ACM Press, New York, NY, 329–332. DOI:http://dx.doi.org/10.1145/1753326.1753376

Gideon Dror, Yehuda Koren, Yoelle Maarek, and Idan Szpektor. 2011. I Want to Answer, Who Has a Question? Yahoo! Answers Recommender System. In *Proc. of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD'11*. ACM Press, New York, NY, 1109–1117. DOI:http://dx.doi.org/10.1145/2020408.2020582

Gideon Dror, Yoelle Maarek, Avihai Mejer, and Idan Szpektor. 2013. From Query to Question in One Click: Suggesting Synthetic Questions to Searchers. In *Proc. of the 22nd International Conference on World Wide Web-WWW'13*. International World Wide Web Conferences Steering Committee, 391–401.

Gideon Dror, Yoelle Maarek, and Idan Szpektor. 2013. Will My Question Be Answered? Predicting "Question Answerability" in Community Question-Answering Sites. In *Proc. of European Conference on Machine Learning and Knowledge Discovery in Databases-ECML PKDD'13*. Springer, Berlin, 499–514. DOI:http://dx.doi.org/10.1007/978-3-642-40994-3_32

Gideon Dror, Dan Pelleg, Oleg Rokhlenko, and Idan Szpektor. 2012. Churn prediction in new users of Yahoo! answers. In *Proc. of the 21st International Conference Companion on World Wide Web-WWW'12 Companion*. ACM Press, New York, NY, 829–834. DOI:http://dx.doi.org/10.1145/2187980.2188207

Pnina Fichman. 2011. A comparative assessment of answer quality on four question answering sites. *J. Inf. Sci.* 37, 5 (October 2011), 476–486. DOI:http://dx.doi.org/10.1177/0165551511415584

Bojan Furlan, Bosko Nikolic, and Veljko Milutinovic. 2013. A survey and evaluation of state-of-the-art intelligent question routing systems. *Int. J. Intell. Syst.* 28 (2013), 686–708. DOI:http://dx.doi.org/10.1002/int.21597

Adabriand Furtado, Nazareno Andrade, Nigini Oliveira, and Francisco Brasileiro. 2013. Contributor Profiles, their Dynamics, and their Importance in Five Q&A Sites. In *Proc. of the 2013 Conference on Computer Supported Cooperative Work-CSCW'13*. ACM Press, New York, NY, USA: 1237–1252. DOI:http://dx.doi.org/10.1145/2441776.2441916

Yunjun Gao, Lu Chen, Rui Li, and Gang Chen. 2013. Mapping Queries to Questions: Towards Understanding Users' Information Needs. In *Proc. of the 36th International ACM SIGIR Conference on*

Research and Development in Information Retrieval-SIGIR'13. ACM Press, New York, NY, 977–980. DOI:http://dx.doi.org/10.1145/2484028.2484138

Giovanni Gardelli and Ingmar Weber. 2012. Why do you ask this? Using Toolbar Data to Identify Common Patterns of Q&A Users. In *Proc. of the 21st International Conference Companion on World Wide Web-WWW'12 Companion.* ACM Press, New York, NY, 815–822. DOI:http://dx.doi.org/10.1145/2187980.2188205

Rich Gazan. 2011. Social Q&A. *J. Am. Soc. Inf. Sci. Technol.* 62, 12 (December 2011), 2301–2312. DOI:http://dx.doi.org/10.1002/asi.21562

Alexandru Lucian Ginsca and Adrian Popescu. 2013. User Profiling for Answer Quality Assessment in Q&A Communities. In *Proc. of the 2103 Workshop on Data-Driven user Behavioral Modelling and Mining from Social Media-DUBMOD'13.* ACM Press, New York, NY, 25–28. DOI:http://dx.doi.org/10.1145/2513577.2513579

George Gkotsis, Karen Stepanyan, Carlos Pedrinaci, John Domingue, and Maria Liakata. 2014. It's all in the Content: State of the art Best Answer Prediction based on Discretisation of Shallow Linguistic Features. In *Proc. of the 2014 ACM Conference on Web Science-WebSci'14.* ACM Press, New York, NY, 202–210. DOI:http://dx.doi.org/10.1145/2615569.2615681

Carlos Gomez, Brendan Cleary, and Leif Singer. 2013. A Study of Innovation Diffusion through Link Sharing on Stack Overflow. In *Proc. of 10th Working Conference on Mining Software Repositories-MSR'13.* IEEE, 81–84. DOI:http://dx.doi.org/10.1109/MSR.2013.6624011

Scott Grant and Buddy Betts. 2013. Encouraging User Behaviour with Achievements: An Empirical Study. In *Proc. of 10th Working Conference on Mining Software Repositories-MSR'13.* IEEE, 65–68. DOI:http://dx.doi.org/10.1109/MSR.2013.6624007

Jinwen Guo, Shengliang Xu, Shenghua Bao, and Yong Yu. 2008. Tapping on the Potential of Q&A Community by Recommending Answer Providers. In *Proceeding of the 17th ACM Conference on Information and Knowledge Mining-CIKM'08*. ACM Press, New York, NY, 921–930. DOI:http://dx.doi.org/10.1145/1458082.1458204

F. Maxwell Harper, Daniel Moy, and Joseph a Konstan. 2009. Facts or Friends? Distinguishing Informational and Conversational Questions in Social Q&A Sites. In *Proc. of the 27th International Conference on Human Factors in Computing Systems-CHI 09.* ACM Press, New York, NY, 759. DOI:http://dx.doi.org/10.1145/1518701.1518819

F. Maxwell Harper, Daphne Raban, Sheizaf Rafaeli, and Joseph A. Konstan. 2008. Predictors of Answer Quality in Online Q&A Sites. In *Proceeding of the 26th Annual CHI Conference on Human Factors in Computing Systems-CHI'08*. ACM Press, New York, NY, 865–874. DOI:http://dx.doi.org/10.1145/1357054.1357191

F. Maxwell Harper, Joseph Weinberg, John Logie, and Joseph A. Konstan. 2010. Question types in social Q&A sites. *First Monday* 15, 7 (2010).

Felix Hieber and Stefan Riezler. 2011. Improved Answer Ranking in Social Question-Answering Portals. In *Proc. of the 3rd International Workshop on Search and Mining user-Generated Contents-SMUC'11.* ACM Press, New York, NY, 19–26. DOI:http://dx.doi.org/10.1145/2065023.2065030

Damon Horowitz and Sepandar D. Kamvar. 2010. The anatomy of a large-scale social search engine. In *Proc. of the 19th International Conference on World Wide Web-WWW'10.* New York, New York, USA: ACM Press, 431–440. DOI:http://dx.doi.org/10.1145/1772690.1772735

Gary Hsieh and Scott Counts. 2009. mimir: A Market-Based Real-Time Question and Answer Service. In *Proc. of the 27th International Conference on Human Factors in Computing Systems-CHI 09.* ACM Press, New York, NY, 769–778. DOI:http://dx.doi.org/10.1145/1518701.1518820

Adrian Huna, Ivan Srba, and Maria Bielikova. 2016. Exploiting Content Quality and Question Difficulty in CQA Reputation Systems. In *Proc. of International Conference on Nework Science-NetSciX'16*. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, Berlin, 68–81. DOI:http://dx.doi.org/10.1007/978-3-319-28361-6_6

Wen Chan, Jintao Du, Weidong Yang, Jinhui Tang, and Xiangdong Zhou. 2014. Term Selection and Result Reranking for Question Retrieval by Exploiting Hierarchical Classification. In *Proc. of the 23rd ACM International Conference on Conference on Information and Knowledge Management-CIKM'14.* ACM Press, New York, NY, 141–150. DOI:http://dx.doi.org/10.1145/2661829.2661938

Wen Chan, Weidong Yang, Jinhui Tang, Jintao Du, Xiangdong Zhou, and Wei Wang. 2013. Community Question Topic Categorization via Hierarchical Kernelized Classification. In *Proc. of the 22nd ACM International Conference on Conference on Information & Knowledge Management-CIKM'13.* ACM Press, New York, NY, 959–968. DOI:http://dx.doi.org/10.1145/2505515.2505676

Wen Chan, Xiangdong Zhou, Wei Wang, and Tat-Seng Chua. 2012. Community Answer Summarization for Multi-Sentence Question with Group L1 Regularization. In *Proc. of the 50th Annual Meeting of the*

Association for Computational Linguistics-ACL'12. Stroudsburg, PA, USA: Association for Computational Linguistics, 582–591.

Shuo Chang and Aditya Pal. 2013. Routing Questions for Collaborative Answering in Community Question Answering. In *Proc. of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining-ASONAM'13*. ACM Press, New York, NY, 494–501. DOI:http://dx.doi.org/10.1145/2492517.2492559

M. A. Chatti, U. Schroeder, and M. Jarke. 2012. LaaN: Convergence of Knowledge Management and Technology-Enhanced Learning. *IEEE Trans. Learn. Technol.* 5, 2 (April 2012), 177–189. DOI:http://dx.doi.org/10.1109/TLT.2011.33

Bee-Chung Chen, Anirban Dasgupta, Xuanhui Wang, and Jie Yang. 2012. Vote Calibration in Community Question-Answering Systems. In *Proc. of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval-SIGIR'12*. ACM Press, New York, NY, 781–790. DOI:http://dx.doi.org/10.1145/2348283.2348388

Cheng Chen, Kui Wu, Venkatesh Srinivasan, and R. Kesav Bharadwaj. 2013. The Best Answers? Think Twice: Online Detection of Commercial Campaigns in the CQA Forums. In *Proc. of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining-ASONAM'13*. ACM Press, New York, NY, 458–465. DOI:http://dx.doi.org/10.1145/2492517.2492553

Long Chen, Dell Zhang, and Mark Levene. 2012. Identifying Local Questions in Community Question Answering. In *Proc. of 8th Asia Information Retrieval Societies Conference-AIRS'12*. 306–315. DOI:http://dx.doi.org/10.1007/978-3-642-35341-3_26

Long Chen, Dell Zhang, and Mark Levene. 2013. Question Retrieval with User Intent. In *Proc. of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval-SIGIR'13*. ACM Press, New York, NY, 973–976. DOI:http://dx.doi.org/10.1145/2484028.2484129

Long Chen, Dell Zhang, and Levene Mark. 2012. Understanding User Intent in Community Question Answering. In *Proc. of the 21st International Conference Companion on World Wide Web-WWW'12 Companion*. ACM Press, New York, NY, 823–828. DOI:http://dx.doi.org/10.1145/2187980.2188206

Erik Choi, Vanessa Kitzie, and Chirag Shah. 2012. Developing a typology of online Q&A models and recommending the right model for each question type. *Proc. Am. Soc. Inf. Sci. Technol.* 49, 1 (2012), 1–4. DOI:http://dx.doi.org/10.1002/meet.14504901302

Alton Y. K. Chua and Radhika Shenoy Balkunje. 2012. Comparative Evaluation of Community Question Answering Websites. In *Proc. of 14th International Conference on Asia-Pacific Digital Libraries-ICADL'12*. Springer, Berlin, 209–218. DOI:http://dx.doi.org/10.1007/978-3-642-34752-8_27

Alton Y. K. Chua and Snehasish Banerjee. 2014. Where to Ask and How to Ask? The Case of Community Question Answering Sites. In *Proc. of Internation Conference on Science and Information Conference-SAI'14*. IEEE, 888–895. DOI:http://dx.doi.org/10.1109/SAI.2014.6918291

Grace YoungJoo Jeon, Yong-Mi Kim, and Yan Chen. 2010. Re-examining Price as a Predictor of Answer Quality in an Online Q&A Site. In *Proc. of the 28th International Conference on Human Factors in Computing Systems-CHI'10*. ACM Press, New York, NY, 325–328. DOI:http://dx.doi.org/10.1145/1753326.1753375

Jiwoon Jeon, W. Bruce Croft, and Joon Ho Lee. 2005. Finding Semantically Similar Questions Based on Their Answers. In *Proc. of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval-SIGIR'05*. ACM Press, New York, NY, 617–618. DOI:http://dx.doi.org/10.1145/1076034.1076156

Jiwoon Jeon, W. Bruce Croft, Joon Ho Lee, and Soyeon Park. 2006. A Framework to Predict the Quality of Answers with Non-Textual Features. In *Proc. of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval-SIGIR'06*. ACM Press, New York, NY, 228–235. DOI:http://dx.doi.org/10.1145/1148170.1148212

Zongcheng Ji and Bin Wang. 2013. Learning to Rank for Question Routing in Community Question Answering. In *Proc. of the 22nd ACM International Conference on Conference on Information & Knowledge Management-CIKM'13*. ACM Press, New York, NY, 2363–2368. DOI:http://dx.doi.org/10.1145/2505515.2505670

Blooma John, Chua Alton Yeow-Kuan, and Goh Dion Hoe-Lian. 2011. What Makes a High-Quality User-Generated Answer? *IEEE Internet Comput.* 15, 1 (2011), 66–71. DOI:http://dx.doi.org/10.1109/MIC.2010.82

Pawel Jurczyk and Eugene Agichtein. 2007. Discovering Authorities in Question Answer Communities by Using Link Analysis. In *Proc. of the Sixteenth ACM Conference on Conference on Information and Knowledge Management-CIKM'07*. ACM Press, New York, NY, 919–922. DOI:http://dx.doi.org/10.1145/1321440.1321575

Imrul Kayes, Nicolas Kourtellis, Daniele Quercia, Adriana Iamnitchi, and Francesco Bonchi. 2015. The Social World of Content Abusers in Community Question Answering. In *Proc. of the 24th International*

Conference on World Wide Web-WWW'15. International World Wide Web Conferences Steering Committee, 570–580.

Chong Tong Lee, Eduarda Mendes Rodrigues, Gabriella Kazai, Nataša Milic-Frayling, and Aleksandar Ignjatovic. 2009. Model for Voter Scoring and Best Answer Selection in Community Q&A Services. In *Proc. of IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology-WI-IAT'09*. IEEE, 116–123. DOI:http://dx.doi.org/10.1109/WI-IAT.2009.23

Uichin Lee, Hyanghong Kang, Eunhee Yi, Mun Yi, and Jussi Kantola. 2012. Understanding Mobile Q&A Usage: An Exploratory Study. In *Proc. of the 2012 ACM Annual Conference on Human Factors in Computing Systems-CHI'12*. ACM Press, New York, NY, 3215–3224. DOI:http://dx.doi.org/10.1145/2207676.2208741

Pierre Lévy. 1997. *Collective Intelligence: Mankind's Emerging World in Cyberspace*, Perseus Books, Cambridge, MA.

Baichuan Li. 2014. *A Computational Framework for Question Processing in Community Question Answering Services*. The Chinese University of Hong Kong.

Baichuan Li, Tan Jin, Michael R. Lyu, Irwin King, and Barley Mak. 2012. Analyzing and Predicting Question Quality in Community Question Answering Services. In *Proc. of the 21st International Conference Companion on World Wide Web-WWW'12 Companion*. ACM Press, New York, NY, 775–782. DOI:http://dx.doi.org/10.1145/2187980.2188200

Baichuan Li and Irwin King. 2010. Routing Questions to Appropriate Answerers in Community Question Answering Services. In *Proc. of the 19th ACM International Conference on Information and Knowledge Management-CIKM'10*. ACM Press, New York, NY, 1585–1588. DOI:http://dx.doi.org/10.1145/1871437.1871678

Baichuan Li, Irwin King, and Michael R. Lyu. 2011. Question Routing in Community Question Answering: Putting Category in Its Place. In *Proc. of the 20th ACM International Conference on Information and Knowledge Management-CIKM'11*. ACM Press, New York, NY, 2041–2044. DOI:http://dx.doi.org/10.1145/2063576.2063885

Baichuan Li, Michael R. Lyu, and Irwin King. 2012. Communities of Yahoo! Answers and Baidu Zhidao: Complementing or Competing? In *Proc. of International Joint Conference on Neural Networks-IJCNN'12*. IEEE, 1–8. DOI:http://dx.doi.org/10.1109/IJCNN.2012.6252435

Baoli Li and Eugene Agichtein. 2008. CoCQA: Co-Training Over Questions and Answers with an Application to Predicting Question Subjectivity Orientation. In *Proc. of the Conference on Empirical Methods in Natural Language Processing-EMNLP'08*. Stroudsburg, PA, USA: Association for Computational Linguistics, 937–946.

Ze Li, Haiying Shen, and Joseph Edward Grant. 2012. Collective Intelligence in the Online Social Network of Yahoo!Answers and Its Implications. In *Proc. of the 21st ACM International Conference on Information and Knowledge Management-CIKM'12*. ACM Press, New York, NY, 455–464. DOI:http://dx.doi.org/10.1145/2396761.2396821

Chih-Lu Lin, Ying-Liang Chen, and Hung-Yu Kao. 2014. Question Difficulty Evaluation by Knowledge Gap Analysis in Question Answer Communities. In *Proc. of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining-ASONAM'14*. IEEE, 336–339. DOI:http://dx.doi.org/10.1109/ASONAM.2014.6921606

Duen Ren Liu, Yu Hsuan Chen, and Chun Kai Huang. 2014. QA document recommendations for communities of question-answering websites. *Knowledge-Based Syst.* 57 (2014), 146–160. DOI:http://dx.doi.org/10.1016/j.knosys.2013.12.017

Duen Ren Liu, Yu Hsuan Chen, Wei Chen Kao, and Hsiu Wen Wang. 2013. Integrating expert profile, reputation and link analysis for expert finding in question-answering websites. *Inf. Process. Manag.* 49, 1 (2013), 312–329. DOI:http://dx.doi.org/10.1016/j.ipm.2012.07.002

Jing Liu, Young-in Song, and Chin-yew Lin. 2011. Competition-based User Expertise Score Estimation. In *Proc. of the 34th International ACM SIGIR Conference on Research and Development in Information-SIGIR'11*. ACM Press, New York, NY, 425–434. DOI:http://dx.doi.org/10.1145/2009916.2009975

Mingrong Liu, Yicen Liu, and Qing Yang. 2010. Predicting Best Answerers for New Questions in Community Question Answering. In Lei Chen, Changjie Tang, Jun Yang, & Yunjun Gao, eds. *Proc. of the 11th International Conference on Web-Age Information Management-WAIM'10*. Springer, Berlin, 127–138. DOI:http://dx.doi.org/10.1007/978-3-642-14246-8_15

Qiaoling Liu, et al. 2011. Predicting Web Searcher Satisfaction with Existing Community-based Answers. In *Proc. of the 34th International ACM SIGIR Conference on Research and Development in Information-SIGIR'11*. ACM Press, New York, NY, 415–424. DOI:http://dx.doi.org/10.1145/2009916.2009974

Qiaoling Liu, Eugene Agichtein, Gideon Dror, Yoelle Maarek, and Idan Szpektor. 2012. When Web Search Fails, Searchers Become Askers: Understanding the Transition. In *Proc. of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval-SIGIR'12*. ACM Press, New York, NY, 801–810. DOI:http://dx.doi.org/10.1145/2348283.2348390

Qiaoling Liu, Tomasz Jurczyk, Jinho Choi, and Eugene Agichtein. 2015. Real-Time Community Question Answering: Exploring Content Recommendation and User Notification Strategies Qiaoling. In *Proc. of the 20th International Conference on Intelligent User Interfaces-IUI'15.* ACM Press, New York, NY, 50–61. DOI:http://dx.doi.org/10.1145/2678025.2701392

Tie-Yan Liu. 2007. Learning to Rank for Information Retrieval. *Found. Trends® Inf. Retr.* 3, 3 (2007), 225–331. DOI:http://dx.doi.org/10.1561/1500000016

Xiaoying Liu, Zhoujun Li, Xiaojian Zhao, and Zhenggan Zhou. 2011. Using Concept-Level Random Walk Model and Global Inference Algorithm for Answer Summarization. In *Proc. of 7th Asia Information Retrieval Societies Conference-AIRS'11*. Springer, Berlin, 434–445. DOI:http://dx.doi.org/10.1007/978-3-642-25631-8_39

Xiaoyong Liu, W. Bruce Croft, and Matthew Koll. 2005. Finding Experts in Community-Based Question-Answering Services. In *Proc. of the 14th ACM International Conference on Information and Knowledge Management-CIKM'05.* ACM Press, New York, NY, 315–316. DOI:http://dx.doi.org/10.1145/1099554.1099644

Yandong Liu and Eugene Agichtein. 2008. You've Got Answers: Towards Personalized Models for Predicting Success in Community Question Answering. In *Proc. of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies-HLT'08.* Stroudsburg, PA, USA: Association for Computational Linguistics, 97–100.

Yandong Liu, Jiang Bian, and Eugene Agichtein. 2008. Predicting Information Seeker Satisfaction in Community Question Answering. In *Proc. of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval-SIGIR'08.* ACM Press, New York, NY, 483–490. DOI:http://dx.doi.org/10.1145/1390334.1390417

Yandong Liu, Nitya Narasimhan, Venu Vasudevan, and Eugene Agichtein. 2009. Is this urgent? Exploring Time-Sensitive Information Needs in Collaborative Question Answering. In *Proc. of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval-SIGIR'09.* ACM Press, New York, NY, 712–713. DOI:http://dx.doi.org/10.1145/1571941.1572091

Yuanjie Liu, Shasha Li, Yunbo Cao, Chin-yew Lin, Dingyi Han, and Yong Yu. 2008. Understanding and Summarizing Answers in Community-Based Question Answering Services. In *Proc. of the 22nd International Conference on Computational Linguistics-COLING'08.* Stroudsburg, PA, USA: Association for Computational Linguistics, 497–504.

Zhe Liu and Bernard J. Jansen. 2013. Factors Influencing the Response Rate in Social Question & Answering Behavior. In *Proc. of the 2013 Conference on Computer Supported Cooperative Work-CSCW'13.* ACM Press, New York, NY, 1263–1274. DOI:http://dx.doi.org/10.1145/2441776.2441918

Lin Luo, Fei Wang, Michelle X. Zhou, Yingxin Pan, and Hang Chen. 2014. Who Have Got Answers? Growing the Pool of Answerers in a Smart Enterprise Social QA System. In *Proc. of the 19th International Conference on Intelligent User Interfaces-IUI'14.* ACM Press, New York, NY, 7–16. DOI:http://dx.doi.org/10.1145/2557500.2557531

Lena Mamykina, Bella Manoim, Manas Mittal, George Hripcsak, and Björn Hartmann. 2011. Design Lessons from the Fastest Q&A Site in the West. In *Proc. of the 2011 Annual Conference on Human Factors in Computing Systems-CHI'11.* ACM Press, New York, NY, 2857–2866. DOI:http://dx.doi.org/10.1145/1978942.1979366

Yuqing Mao, Haifeng Shen, and Chengzheng Sun. 2013. Online Silk Road: Nurturing Social Search through Knowledge Bartering. In *Proc. of the 2013 Conference on Computer Supported Cooperative Work-CSCW'13.* ACM Press, New York, NY, 1193–1202. DOI:http://dx.doi.org/10.1145/2441776.2441911

Justin Matejka, Tovi Grossman, and George Fitzmaurice. 2011. IP-QAT: In-Product Questions, Answers & Tips. In *Proc. of the 24th Annual ACM Symposium on User Interface Software and Technology-UIST'11.* ACM Press, New York, NY, 175–184. DOI:http://dx.doi.org/10.1145/2047196.2047218

Eduarda Mendes Rodrigues and Natasa Milic-Frayling. 2009. Socializing or knowledge sharing?: characterizing social intent in community question answering. In *Proceeding of the 18th ACM Conference on Information and Knowledge Management-CIKM'09*. ACM Press, New York, NY, 1127. DOI:http://dx.doi.org/10.1145/1645953.1646096

Yajie Miao, Chunping Li, Jie Tang, and Lili Zhao. 2010. Identifying New Categories in Community Question Answering Archives: A Topic Modeling Approach. In *Proc. of the 19th ACM International Conference on Information and Knowledge Management-CIKM'10.* ACM Press, New York, NY, 1673–1676. DOI:http://dx.doi.org/10.1145/1871437.1871701

Patrick Morrison and E. Murphy-Hill. 2013. Is Programming Knowledge Related To Age? In *Proc. of the 10th Working Conference on Mining Software Repositories-MSR'13.* 69–72.

Dana Movshovitz-Attias, Yair Movshovitz-Attias, Peter Steenkiste, and Christos Faloutsos. 2013. Analysis of the Reputation System and User Contributions on a Question Answering Website: StackOverflow.

In *Proc. of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining-ASONAM'13.* ACM Press, New York, NY, 886–893. DOI:http://dx.doi.org/10.1145/2492517.2500242

Klemens Muthmann and Alina Petrova. 2014. An Automatic Approach for Identifying Topical Near-Duplicate Relations between Questions from Social Media Q/A Sites. In *Proceeding of WSDM 2014 Workshop: "Web-Scale Classification: Classifying Big Data from the Web."*

Liqiang Nie, Yi-Liang Zhao, Xiangyu Wang, Jialie Shen, and Tat-seng Chua. 2014. Learning to Recommend Descriptive Tags for Questions in Social Forums. *ACM Trans. Inf. Syst.* 32, 1 (January 2014), 1–23. DOI:http://dx.doi.org/10.1145/2559157

Jeffrey Nichols and Jeon-Hyung Kang. 2012. Asking Questions of Targeted Strangers on Social Networks. *Proc. ACM 2012 Conf. Comput. Support. Coop. Work-CSCW'12* (2012), 999–1002. DOI:http://dx.doi.org/10.1145/2145204.2145352

Kyosuke Nishida and Ko Fujimura. 2010. Hierarchical Auto-Tagging: Organizing Q&A Knowledge for Everyone. In *Proc. of the 19th ACM International Conference on Information and Knowledge Management-CIKM'10.* ACM Press, New York, NY, 1657–1660. DOI:http://dx.doi.org/10.1145/1871437.1871697

Aditya Pal et al. 2011. Early Detection of Potential Experts in Question Answering Communities. In *Proc. of the 19th International Conference on User Modeling, Adaption, and Personalization-UMAP'11.* ACM Press, New York, NY, 231–242.

Aditya Pal, F. Maxwell Harper, and Joseph A. Konstan. 2012. Exploring Question Selection Bias to Identify Experts and Potential Experts in Community Question Answering. *ACM Trans. Inf. Syst.* 30, 2 (May 2012), 1–28. DOI:http://dx.doi.org/10.1145/2180868.2180872

Aditya Pal, Shuo Chang, and Joseph A. Konstan. 2012. Evolution of Experts in Question Answering Communities. In *Proc. of the Sixth International AAAI Conference on Weblogs and Social Media-ICWSM'12.* 1–8.

Aditya Pal, Fei Wang, Michelle X. Zhou, Jeffrey Nichols, and Barton A. Smith. 2013. Question Routing to User Communities. In *Proc. of the 22nd ACM International Conference on Conference on Information & Knowledge Management-CIKM'13.* ACM Press, New York, NY, 2357–2362. DOI:http://dx.doi.org/10.1145/2505515.2505669

Vinay Pande, Tanmoy Mukherjee, and Vasudeva Varma. 2013. Summarizing Answers for Community Question Answer Services. In *Proc. of 25th International Conference on Language Processing and Knowledge in the Web-GSCL'13*. Springer, Berlin, 151–161. DOI:http://dx.doi.org/10.1007/978-3-642-40722-2_16

Sharoda A. Paul, Lichan Hong, and Ed. H. Chi. 2012. Who is Authoritative? Understanding Reputation Mechanisms in Quora. In *Proc. of Conference on Collective Intelligence-CI'12.*

Maria Soledad Pera and Yiu-Kai Ng. 2011. A Community Question-Answering Refinement System. In *Proc. of the 22nd ACM Conference on Hypertext and Hypermedia-HT'11.* ACM Press, New York, NY, 251–260. DOI:http://dx.doi.org/10.1145/1995966.1995999

Tiziano Piccardi, Gregorio Convertino, Massimo Zancanaro, Ji Wang, and Cedric Archambeau. 2014. Towards Crowd-based Customer Service: A Mixed-Initiative Tool for Managing Q&A Sites. In *Proc. of the 32nd Annual ACM Conference on Human Factors in Computing Systems-CHI'14.* ACM Press, New York, NY, 2725–2734. DOI:http://dx.doi.org/ 10.1145/2556288.2557202

Luca Ponzanelli, Andrea Mocci, Alberto Bacchelli, Michele Lanza, and David Fullerton. 2014. Improving Low Quality Stack Overflow Post Detection. In *Proc. of IEEE International Conference on Software Maintenance and Evolution-ICSME'14.* 541–544. DOI:http://dx.doi.org/10.1109/ICSME.2014.90

J. S. Pudipeddi, L. Akoglu, and H. Tong. 2014. User Churn in Focused Question Answering Sites: Characterizations and Prediction. In *Proc. of the 23st International Conference Companion on World Wide Web-WWW'14 Companion.* International World Wide Web Conferences Steering Committee, 469–474. DOI:http://dx.doi.org/10.1145/2567948.2576965

Xiaojun Quan, Yao Lu, and Wenyin Liu. 2012. Towards Modeling Question Popularity in Community Question Answering. In *Proc. of IEEE 11th International Conference on Cognitive Informatics and Cognitive Computing-ICCI*CC'12.* IEEE, 109–114. DOI:http://dx.doi.org/10.1109/ICCI-CC.2012.6311134

Sujith Ravi, Bo Pang, and Ravi Kumar. 2014. Great Question! Question Quality in Community Q&A. In *Proc. of AAAI International Conference on Weblogs and Social Media-ICWSM'14.* 426–435.

Fatemeh Riahi, Zainab Zolaktaf, Mahdi Shafiei, and Evangelos Milios. 2012. Finding Expert Users in Community Question Answering. In *Proc. of the 21st International Conference Companion on World Wide Web-WWW'12 Companion.* ACM Press, New York, NY, 791–798. DOI:http://dx.doi.org/10.1145/2187980.2188202

Matthew Richardson and Ryen W. White. 2011. Supporting Synchronous Social Q&A Throughout the Question Lifecycle. In *Proc. of the 20th International Conference on World Wide web-WWW'11.* ACM Press, New York, NY, 755–764. DOI:http://dx.doi.org/10.1145/1963405.1963511

Tetsuya Sakai, Daisuke Ishikawa, Noriko Kando, Yohei Seki, Kazuko Kuriyama, and Chin-Yew Lin. 2011. Using Graded-Relevance Metrics for Evaluating Community QA Answer Selection. In *Proc. of the Fourth ACM International Conference on Web Search and Data Mining-WSDM'11.* ACM Press, New York, NY, 187–196. DOI:http://dx.doi.org/10.1145/1935826.1935864

Jose San Pedro and Alexandros Karatzoglou. 2014. Question Recommendation for Collaborative Question Answering Systems with RankSLDA. In *Proc. of the 8th ACM Conference on Recommender Systems-RecSys'14.* ACM Press, New York, NY, 193–200. DOI:http://dx.doi.org/10.1145/2645710.2645736

Marlene Scardamalia and Carl Bereiter. 2006. Knowledge building: Theory, pedagogy, and technology. In K. Sawyer, ed., *Cambridge handbook of the Learning Sciences*. Cambridge University Press, New York, NY, 97–118.

Chirag Shah, Vanessa Kitzie, and Erik Choi. 2014a. Modalities, motivations, and materials-investigating traditional and social online Q&A services. *J. Inf. Sci.* 40, 5 (October 2014), 669–687. DOI:http://dx.doi.org/10.1177/0165551514534140

Chirag Shah, Vanessa Kitzie, and Erik Choi. 2014b. Questioning the Question-Addressing the Answerability of Questions in Community Question-Answering. In *Proc. of 47th Hawaii International Conference on System Sciences-HICSS'14.* IEEE, 1386–1395. DOI:http://dx.doi.org/10.1109/HICSS.2014.180

Chirag Shah, Sanghee Oh, and Jung Sun Oh. 2009. Research agenda for social Q&A. *Libr. Inf. Sci. Res.* 31, 4 (2009), 205–209. DOI:http://dx.doi.org/10.1016/j.lisr.2009.07.006

Chirag Shah and Jefferey Pomerantz. 2010. Evaluating and Predicting Answer Quality in Community QA. In *Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval-SIGIR'10*. ACM Press, New York, NY, 411–418. DOI:http://dx.doi.org/10.1145/1835449.1835518

Anna Shtok, Gideon Dror, Yoelle Maarek, and Idan Szpektor. 2012. Learning from the Past: Answering New Questions with Past Answers. In *Proc. of the 21st International Conference on World Wide Web-WWW'12.* ACM Press, New York, NY, 759–768. DOI:http://dx.doi.org/10.1145/2187836.2187939

Amit Singh. 2012. Entity based Q&A retrieval. In *Proc. of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language-EMNLP-CoNLL'12.* Stroudsburg, PA, USA: Association for Computational Linguistics, 1266–1277.

Amit Singh and Karthik Visweswariah. 2011. CQC: Classifying Questions in CQA Websites. In *Proc. of the 20th ACM International Conference on Information and Knowledge Management-CIKM'11.* ACM Press, New York, NY, 2033–2036. DOI:http://dx.doi.org/10.1145/2063576.2063883

Siqi Song, Ye Tian, Wenwen Han, Xirong Que, and Wendong Wang. 2013. Leading Users Detecting Model in Professional Community Question Answering Services. In *Proc. of IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing.* IEEE, 1302–1307. DOI:http://dx.doi.org/10.1109/GreenCom-iThings-CPSCom.2013.226

Ivan Srba and Maria Bielikova. 2015. Askalot: Community Question Answering as a Means for Knowledge Sharing in an Educational Organization. In *Proc. of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing-CSCW'15 Companion.* ACM Press, New York, NY, 179–182. DOI:http://dx.doi.org/10.1145/2685553.2699001

Ivan Srba and Mária Bieliková. 2016. Why Is Stack Overflow Failing? Preserving Sustainability in Community Question Answering. *IEEE Softw.* 33, 4 (2016).

Ivan Srba, Marek Grznar, and Maria Bielikova. 2015. Utilizing Non-QA Data to Improve Questions Routing for Users with Low QA Activity in CQA. In *Proc. of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015-ASONAM'15.* ACM Press, New York, NY, 129–136. DOI:http://dx.doi.org/10.1145/2808797.2809331

Ke Sun, Yunbo Cao, Xinying Song, Young-In Song, Xiaolong Wang, and Chin-Yew Lin. 2009. Learning to Recommend Questions Based on User Ratings. In *Proceeding of the 18th ACM Conference on Information and Knowledge Management-CIKM'09*. ACM Press, New York, NY, 751–758. DOI:http://dx.doi.org/10.1145/1645953.1646049

James Surowiecki. 2005. *The Wisdom of Crowds*, Anchor.

Maggy Anastasia Suryanto, Ee Peng Lim, Aixin Sun, and Roger H. L. Chiang. 2009. Quality-Aware Collaborative Question Answering: Methods and Evaluation. In *Proc. of the Second ACM International Conference on Web Search and Data Mining-WSDM'09.* ACM Press, New York, NY, 142–151. DOI:http://dx.doi.org/10.1145/1498759.1498820

Saori Suzuki, Shin'ichi Nakayama, and Hideo Joho. 2011. Formulating Effective Questions for Community-based Question Answering. In *Proc. of the 34th International ACM SIGIR Conference on Research and Development in Information-SIGIR'11.* ACM Press, New York, NY, USA: 1261–1262. DOI:http://dx.doi.org/10.1145/2009916.2010149

Idan Szpektor, Yoelle Maarek, and Dan Pelleg. 2013. When Relevance is not Enough: Promoting Diversity and Freshness in Personalized Question Recommendation. In *Proc. of the 22nd International Conference on World Wide Web-WWW'13.* 1249–1259.

Yla R. Tausczik and James W. Pennebaker. 2011. Predicting the Perceived Quality of Online Mathematics Contributions from Users' Reputations. In *Proc. of the 2011 Annual Conference on Human Factors in Computing Systems-CHI'11.* ACM Press, New York, NY, 1885–1888. DOI:http://dx.doi.org/10.1145/1978942.1979215

Hapnes Toba, Zhao Yan Ming, Mirna Adriani, and Tat Seng Chua. 2014. Discovering high quality answers in community question answering archives using a hierarchy of classifiers. *Inf. Sci. (Ny).* 261, March (2014), 101–115. DOI:http://dx.doi.org/10.1016/j.ins.2013.10.030

Mattia Tomasoni and Minlie Huang. 2010. Metadata-Aware Measures for Answer Summarization in Community Question Answering. In *Proc. of the 48th Annual Meeting of the Association for Computational Linguistics-ACL'10.* Stroudsburg, PA, USA: Association for Computational Linguistics, 760–769.

G. Alan Wang, Harry Jiannan Wang, Jiexun Li, Alan S. Abrahams, and Weiguo Fan. 2014. An Analytical Framework for Understanding Knowledge-Sharing Processes in Online Q&A Communities. *ACM Trans. Manag. Inf. Syst.* 5, 4 (December 2014), 1–31. DOI:http://dx.doi.org/10.1145/2629445

Gang Wang et al. 2013. Wisdom in the Social Crowd: an Analysis of Quora. In *Proc. of the 22nd International Conference on World Wide Web-WWW'13.* International World Wide Web Conferences Steering Committee, 1341–1351.

Jun Wang, Xia Hu, Zhoujun Li, Wenhan Chao, and Biyun Hu. 2011. Learning to Recommend Questions Based on Public Interest. In *Proc. of the 20th ACM International Conference on Information and Knowledge Management-CIKM'11.* ACM Press, New York, NY, 2029–2032. DOI:http://dx.doi.org/10.1145/2063576.2063882

Jun Wang, Zhoujun Li, Xia Hu, and Biyun Hu. 2010. A Novel Composite Kernel for Finding Similar Questions in CQA Services. In *Proc. of 11th International Conference on Web-Age Information Management-WAIM'10*. Springer, Berlin, 608–619. DOI:http://dx.doi.org/10.1007/978-3-642-14246-8_59

Kai Wang, Zhao-yan Ming, Xia Hu, and Tat-seng Chua. 2010. Segmentation of Multi-Sentence Questions: Towards Effective Question Retrieval in CQA Services. In *Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval-SIGIR'10*. ACM Press, New York, NY, 387–394. DOI:http://dx.doi.org/10.1145/1835449.1835515

Kai Wang, Zhaoyan Ming, and Tat-Seng Chua. 2009. A Syntactic Tree Matching Approach to Finding Similar Questions in Community-based QA Services. In *Proc. of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval-SIGIR'09.* ACM Press, New York, NY, 187–194. DOI:http://dx.doi.org/10.1145/1571941.1571975

Xin-Jing Wang, Xudong Tu, Dan Feng, and Lei Zhang. 2009. Ranking Community Answers by Modeling Question-Answer Relationships via Analogical Reasoning. In *Proc. of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval-SIGIR'09.* ACM Press, New York, NY, 179–186. DOI:http://dx.doi.org/10.1145/1571941.1571974

Molly Mclure Wasko and Samer Faraj. 2000. "It is what one does": why people participate and help others in electronic communities of practice. *J. Strateg. Inf. Syst.* 9, 2–3 (September 2000), 155–173. DOI:http://dx.doi.org/10.1016/S0963-8687(00)00045-7

Etienne Wenger. 1998. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press.

Ryen W. White and Matthew Richardson. 2012. Effects of Expertise Differences in Synchronous Social Q&A. In *Proc. of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval-SIGIR'12.* 1055–1056. DOI:http://dx.doi.org/10.1145/2348283.2348466

Ryen W. White, Matthew Richardson, and Yandong Liu. 2011. Effects of Community Size and Contact Rate in Synchronous Social Q&A. In *Proc. of the 2011 Annual Conference on Human Factors in Computing Systems-CHI'11.* ACM Press, New York, NY, 2837–2846. DOI:http://dx.doi.org/10.1145/1978942.1979364

Steve Whittaker. 1996. Talking to strangers: An Evaluation of the Factors Affecting Electronic Collaboration. In *Proc. of the 1996 ACM Conference on Computer Supported Cooperative Work-CSCW'96.* ACM Press, New York, NY, 409–418. DOI:http://dx.doi.org/10.1145/240080.240352

Haocheng Wu, Wei Wu, Ming Zhou, Enhong Chen, Lei Duan, and Heung-Yeung Shum. 2014. Improving Search Relevance for Short Queries in Community Question Answering. In *Proc. of the 7th ACM International Conference on Web Search and Data Mining-WSDM'14.* ACM Press, New York, NY, 43–52. DOI:http://dx.doi.org/10.1145/2556195.2556239

Fei Xu, Zongcheng Ji, and Bin Wang. 2012. Dual Role Model for Question Recommendation in Community Question Answering. In *Proc. of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval-SIGIR'12.* ACM Press, New York, NY, 771–779. DOI:http://dx.doi.org/10.1145/2348283.2348387

Xiaobing Xue, Jiwoon Jeon, and W. Bruce Croft. 2008. Retrieval Models for Question and Answer Archives. In *Proc. of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval-SIGIR'08*. ACM Press, New York, NY, 475–482. DOI:http://dx.doi.org/10.1145/1390334.1390416

Baoguo Yang. 2014. Exploring User Expertise and Descriptive Ability in Community Question Answering. In *Proc. of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining-ASONAM'14*. IEEE, 320–327. DOI:http://dx.doi.org/10.1109/ASONAM.2014.6921604

Diyi Yang, David Adamson, and Carolyn Penstein Rosé. 2014. Question Recommendation with Constraints for Massive Open Online Courses. In *Proc. of the 8th ACM Conference on Recommender Systems-RecSys'14*. ACM Press, New York, NY, 49–56. DOI:http://dx.doi.org/10.1145/2645710.2645748

Jie Yang, Claudia Hauff, Alessandro Bozzon, and Geert-jan Houben. 2014. Asking the Right Question in Collaborative Q&A systems. In *Proc. of the 25th ACM Conference on Hypertext and Social Media-HT'14*. ACM Press, New York, NY, 179–189. DOI:http://dx.doi.org/10.1145/2631775.2631809

Jie Yang, Ke Tao, Alessandro Bozzon, and Geert-Jan Houben. 2014. Sparrows and Owls: Characterisation of Expert Behaviour in StackOverflow. In *Proc. of the 22th International Conference on User Modeling, Adaption, and Personalization-UMAP'14*. Vania Dimitrova, Tsvi Kuflik, David Chin, Francesco Ricci, Peter Dolog, and Geert-Jan Houben, eds. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 37–48. DOI:http://dx.doi.org/10.1007/978-3-319-08786-3

Liu Yang, et al. 2013. CQArank: Jointly Model Topics and Expertise in Community Question Answering. In *Proc. of the 22nd ACM International Conference on Conference on Information & Knowledge Management-CIKM'13*. ACM Press, New York, NY, 99–108. DOI:http://dx.doi.org/10.1145/2505515.2505720

Yuan Yao, Hanghang Tong, Tao Xie, Leman Akoglu, Feng Xu, and Jian Lu. 2015. Detecting High-quality Posts in Community Question Answering Sites. *Inf. Sci. (Ny)*. 302, December (May 2015), 70–82. DOI:http://dx.doi.org/10.1016/j.ins.2014.12.038

Yuan Yao, Hanghang Tong, Feng Xu, and Jian Lu. 2014. Predicting Long-Term Impact of CQA Posts: A Comprehensive Viewpoint. In *Proc. of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD'14*. ACM Press, New York, NY, 1496–1505. DOI:http://dx.doi.org/10.1145/2623330.2623649

Sarita Yardi and Erika Shehan Poole. 2009. Please Help! Patterns of Personalization in an Online Tech Support Board. In *Proc. of the Fourth International Conference on Communities and Technologies-C&T'09*. ACM Press, New York, NY, 285–294. DOI:http://dx.doi.org/10.1145/1556460.1556501

Reyyan Yeniterzi and Jamie Callan. 2014. Analyzing Bias in CQA-based Expert Finding Test Sets. In *Proc. of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval-SIGIR'14*. ACM Press, New York, NY, 967–970. DOI:http://dx.doi.org/10.1145/2600428.2609486

Jun Zhang, Mark S. Ackerman, and Lada Adamic. 2007. Expertise Networks in Online Communities: Structure and Algorithms. In *Proc. of the 16th International Conference on World Wide Web-WWW'07*. ACM Press, New York, NY, 221–230. DOI:http://dx.doi.org/10.1145/1242572.1242603

Kai Zhang, Wei Wu, Haocheng Wu, Zhoujun Li, and Ming Zhou. 2014. Question Retrieval with High Quality Answers in Community Question Answering. In *Proc. of the 23rd ACM International Conference on Conference on Information and Knowledge Management-CIKM'14*. ACM Press, New York, NY, 371–380. DOI:http://dx.doi.org/10.1145/2661829.2661908

Zhou Zhao, Lijun Zhang, Xiaofei He, and Wilfred Ng. 2015. Expert Finding for Question Answering via Graph Regularized Matrix Completion. *IEEE Trans. Knowl. Data Eng.* 27, 4 (April 2015), 993–1004. DOI:http://dx.doi.org/10.1109/TKDE.2014.2356461

Guangyou Zhou, Li Cai, Kang Liu, and Jun Zhao. 2012. Exploring the Existing Category Hierarchy to Automatically Label the Newly-arising Topics in CQA. In *Proc. of the 21st ACM International Conference on Information and Knowledge Management-CIKM'12*. ACM Press, New York, NY, 1647–1651. DOI:http://dx.doi.org/10.1145/2396761.2398490

Guangyou Zhou, Kang Liu, and Jun Zhao. 2012. Joint Relevance and Answer Quality Learning for Question Routing in Community QA. In *Proc. of the 21st ACM International Conference on Information and Knowledge Management-CIKM'12*. ACM Press, New York, NYork, 1492–1496. DOI:http://dx.doi.org/10.1145/2396761.2398459

Guangyou Zhou, Jun Zhao, Tingting He, and Wensheng Wu. 2014. An empirical study of topic-sensitive probabilistic model for expert finding in question answer communities. *Knowl.-Based Syst.* 66, August (2014), 136–145. DOI:http://dx.doi.org/10.1016/j.knosys.2014.04.032

Tom Chao Zhou, Michael Rung-Tsong Lyu, Irwin King, and Jie Lou. 2015. Learning to suggest questions in social media. *Knowl. Inf. Syst.* 43, 2 (May 2015), 389–416. DOI:http://dx.doi.org/10.1007/s10115-014-0737-z

Zhi-Min Zhou, Man Lan, Zheng-Yu Niu, and Yue Lu. 2012. Exploiting User Profile Information for Answer Ranking in CQA. In *Proc. of the 21st International Conference Companion on World Wide Web-WWW'12 Companion.* ACM Press, New York, NY, 767–774. DOI:http://dx.doi.org/10.1145/2187980.2188199

Hengshu Zhu, Huanhuan Cao, Hui Xiong, Enhong Chen, and Jilei Tian. 2011. Towards Expert Finding by Leveraging Relevant Categories in Authority Ranking. In *Proc. of the 20th ACM International Conference on Information and Knowledge Management-CIKM'11.* 2221–2224. DOI:http://dx.doi.org/10.1145/2063576.2063931

Hengshu Zhu, Enhong Chen, and Huanhuan Cao. 2011. Finding Experts in Tag Based Knowledge Sharing Communities. In *Proc. of 5th International Conference on Knowledge Science, Engineering and Management-KSEM'11*. Springer, Berlin, 183–195. DOI:http://dx.doi.org/10.1007/978-3-642-25975-3_17

Z. Zhu, D. Bernhard, and I. Gurevych. 2009. *A Multi-Dimensional Model for Assessing the Quality of Answers in Social Q&A Sites*, Technical Report TUD-CS-2009-0158.

Received August 2015; revised December 2015; accepted May 2016