Discovering Educational Potential Embedded in Community Question Answering

Ivan Srba and Mária Bieliková

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia

{name.surname}@stuba.sk

Abstract. Community Question Answering (CQA) systems, such as Yahoo! Answers or Quora, are mostly perceived and studied from their primary knowledge sharing perspective. In spite of that, CQA systems have also a potential to become an effective means for people to acquire new knowledge. In the present, we can witness initial efforts on taking advance of this secondary perspective as CQA concepts have been recently applied in several educational applications. However, these educational systems take quite different approaches to the transition of CQA concepts from the open web to organizational and educational environments. One of possible reasons is that it is not clear which aspects of the question answering process have the best potential for knowledge acquisition. Therefore, we performed a study with a dataset obtained from CQA system Stack Overflow, in which we identified three scenarios that positively leads to improvement of users' expertise. The results of the study provide us with a better understanding how CQA systems can be applied as nontraditional learning environments.

Keywords: Community Question Answering, Informal Learning, Knowledge Sharing, Stack Overflow

1 Introduction

Information retrieval systems (e.g. web search engines or digital libraries) provide users with powerful tools how to identify valuable information and knowledge in the great information space of the current web. However, these systems are effective only if required information has been already codified and made publicly available. Moreover, standard information retrieval systems are not always successful to answer subjective, non-factual and context-aware queries, such as recommendations, advices or complex domain-specific problems. Current possibilities of the web allow us to employ supplementary sources of information to overcome these problems. These non-traditional sources of knowledge are often based on collective intelligence. Concept of collective intelligence [4] refers to shared knowledge which emerges from common collaboration of a community of users that share common practice, interests or goals. Collective intelligence is present in many popular web systems, such as forums, social networking

sites or wikis (particularly Wikipedia is considered as one of the best examples of collective intelligence). In recent years, the new forms of web systems based on collective intelligence have appeared. One of them is Community Question Answering.

Community Question Answering (CQA) is a service where people can seek information by asking questions and share knowledge by providing answers on questions asked by other users. The well-known examples of CQA systems include Yahoo! Answers, Quora or domain-specific Stack Overflow where users concern with questions related to computer science and programming.

Millions of answered questions have already proved the successful concepts of CQA. Therefore, CQA systems became the interesting subject of many research studies. However, in spite of the increasing research effort in recent years, the beneficial effects of CQA systems has not been fully determined for intra-organizational environments yet, such as for educational or business organizations. Especially the educational domain, where students are quite often struggling with various problems related to a learning process, can benefit from the positive effects of the question answering process. Nevertheless, the full potential of CQA systems in the educational domain has not been discovered yet. In particular, the current state of research does not provide a clear answer which aspects of the question answering process contribute to successful knowledge acquisition.

The main goal of this paper is to describe an educational potential of CQA systems, specifically how CQA concepts can be successfully applied in educational and organization-wide environments, such as at universities. We determine this potential by means of an exploratory study with a dataset from CQA system Stack Overflow. The results from the study consist of an identification of three knowledge acquisition scenarios that lead to a positive improvement of users' expertise.

The paper is structured as follows: we describe a related work on CQA systems in Section 2; Section 3 contains a study aimed to discover a learning potential of CQA systems; finally, conclusions and future work are proposed in Section 4.

2 CQA Systems in an Educational Context

A typical process of question answering in CQA systems consists of several steps. At first, an asker posts a question by formulating a description of his or her problem. In addition, it is usually necessary to select an appropriate question's topic (a category or a set of related tags). Afterwards, answerers can provide their answer-candidates on the posted question, vote for the most appropriate answers and thus help all users, who are involved in the question answering process, to identify answers with the highest quality. The asker can finish the answering process by selecting the best answer, which satisfies his or her information needs best, and consequently the question is marked as answered and moved to the archive of solved questions.

The significant part of state-of-the-art research on CQA systems studies the question answering process from a perspective of knowledge sharing. In this perspective, the goal of a CQA system is to harness knowledge of a whole community to provide the

most suitable answers on recently posted questions in the shortest possible time. Besides this primary view, we stress that there is also another interesting perspective how CQA systems can be perceived. People can gain new knowledge by reading, asking and also by answering questions. In addition, they are able to perceive different perspectives on a problem by discussions attached to a question or related answers. Thereby, it is very natural to speak about this knowledge acquisition as a special kind of informal learning in CQA systems.

One of possible ways, how to provide students with innovative learning environments, is to adopt concepts of Web 2.0 knowledge sharing applications, such as wikis, forums, social networking sites or content creation tools [8]. Due to the learning potential of CQA systems, it is natural that they have been also already applied as a model in proposals of educational systems. However, we are aware of only a few studies concerned with an employment of CQA systems within the educational domain.

At first, OpenStudy [5] is a large-scale open social learning environment which promotes knowledge sharing through Web 2.0 technologies. It adapts concepts of many social applications, such as CQA systems, online forums, real-time chats and social networking sites. In the present, OpenStudy (http://openstudy.com) has more than 1 million users that come from 160 countries.

While OpenStudy involves a great open community of students, remaining educational CQA systems are focused on smaller groups of students who enrolled for the same class. Piazza (http://piazza.com) is a learning system that is directly inspired by CQA. It is an online platform which offers a refined question answering process along with key features for effective course collaboration. It supports student-to-student collaboration as well as student-to-teacher discussions.

Green Dolphin is another social question answering board designed to support collaborative learning of programming [1]. Green Dolphin automatically identifies students who are experts on a particular topic. Afterwards, students can ask other students or directly experts identified by the system. This recommendation ensures high-quality answers while minimizing teachers overload. Students are awarded by points for asking and answering questions. Afterwards, they can use the earned points to direct their own questions to the recommended experts or teachers. The important concept of Green Dolphin is that new questions are postponed and hidden from teachers for some time, so students have enough time to provide answers by themselves. Only if a question cannot be answered in the given time, a teacher is notified and asked to take a participation on students' collaboration.

Authors in [2] investigated how to meet the needs of students and instructors while providing them with possibilities of social tools. Classroom Salon was proposed in which collaboration takes place in small groups termed Salons. Each Salon can be open to the entire community or only to a particular group. Students can use these Salons to post various documents, such as a piece of text, a program or a series of questions. Additionally, it is possible to annotate or vote on these documents. Authors in the set of experiments confirmed that their system based on principles of social tools, CQA being one of them, can successfully replace and outperform traditional online forums.

On the basis of performed analyzes, we identified an open problem that directs our further research. The analyzed educational systems are quite diverse in an application

of CQA concepts for the purpose of learning. One of the main reasons is that their potential to elicit students' participation and collaborative learning is only to be discovered and thus it has not been well-documented yet [1]. The existing systems assume that an active participation on the question answering process (e.g. asking a question, providing an answer or a comment, searching for solved questions) leads somehow to an improvement of users' knowledge about particular topics. However, it is not very clear which specific situations in the question answering process have a potential for participating users to acquire new knowledge.

We suppose that especially collaborative scenarios in which users participate on question answering with other more experienced users can lead to improvement in their knowledge. With tangible and well-described identification of this kind of scenarios, it will be possible to optimally utilize the learning potential embedded in the question answering process. Naturally, the identification of these scenarios plays even more important role in the educational domain. It will suggest how to propose more effective learning environments as well as methods for adaptive collaboration support that will guide students towards these scenarios. There are many ways how to achieve it, e.g. we can provide students with a recommendation to take a participation on questions in which some of the identified collaborative scenarios will occur with a high probability.

3 Determination of CQA Learning Potential

In order to answer the identified open problem, we conducted a study aimed to discover in which scenarios and how users improve their knowledge in CQA systems. For this purpose, we analyzed a dataset from Stack Overflow and investigated relations between users' interactions and their level of expertise. Following the standard process in CQA systems, we identified three scenarios which occur during the question answering process and which can lead to improvement in expertise of participating users. Consequently, we stated a hypothesis associated with each of these scenarios. More specifically, we suppose that users improve their expertise:

- When an asker receives a high quality answer (Scenario 1).

  H1: The number of asked questions containing high quality answers (with a quality that exceeds asker's current expertise) is positively associated with improvement in asker's expertise.

- When an answerer utilizes additional external sources, which supplement his or her own expertise, to provide better answers (Scenario 2).

  H2: The number of provided high quality answers (with a quality that exceeds current answer's expertise) is positively associated with answerer's expertise.

- When a user participates (i.e. provides an additional answer-candidate while the best answer has not been selected yet) on a question answering in which other high quality answers has been previously posted (Scenario 3).

  H3: The number of provided answers to questions that has previously received other high quality answers (with a quality that exceeds current answerer expertise) is positively associated with answerer's expertise.

3.1 Dataset Description

We employed a dataset from Stack Overflow in our analyses. Stack Overflow is a domain-specific open CQA system in which users concern with various questions about programming. Stack Overflow is not applied directly in the educational domain, nevertheless our goal is to determine the learning potential embedded in the question answering process, which appears independently on the particular domain where CQA system is applied in. In addition, its large dataset is publicly available and thus it represents the best option for the purpose of our study.

Stack Overflow was founded in July 2008 and so far it contains more than 8.1M of questions and more than 14M of answers. It is considered as the fastest CQA system with median time to the first answer lower than 10 minutes. Several voting mechanisms are available: each question and answer can be voted up or down; each comment can be marked as a useful one. In addition, a question can be starred as a favorite one.

The anonymized dataset from Stack Overflow as well as from all other CQA systems built on the top of Stack Exchange infrastructure is published regularly under Creative Commons license and contains all publicly available data (http://blog.stackexchange.com/category/cc-wiki-dump/). The main part of dataset consists of users' posts (i.e. questions, answers and comments), their revision history and metadata (i.e. tags, votes, received badges).

At first, we analyzed the evolution of Stack Overflow during its history. More specifically, we focused on the number of new questions. Figure 1 shows that the amount of new questions is growing from the very beginning and is quite stable recently. Therefore, we decided to limit our further analyses at the content posted between January and December 2013. During this one year-long interval, users asked about 2.3M of questions and provided more than 3.4M of answers. About 91.9% of all questions received at least one answer and in addition, asker selected the best answer in the case of 48.7% of all questions. Quite impressive is also the speed of the question answering process as 67.5% of questions received the first answer in 10 minutes after being posted.

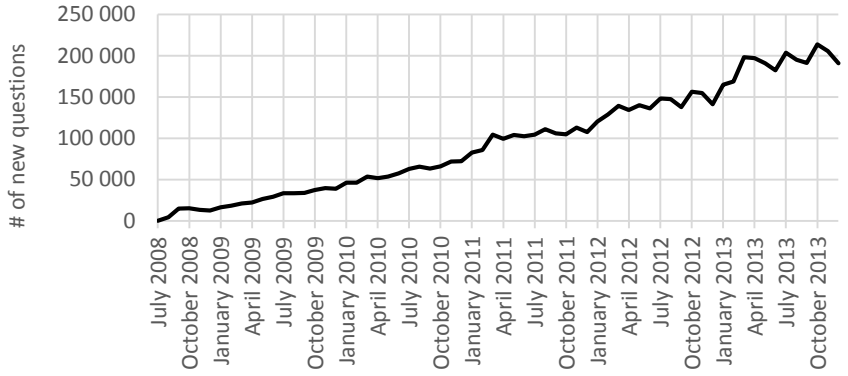

Fig. 1. Number of new questions posted each month during Stack Overflow history.

3.2 Estimation of Answer Quality and User Topical Expertise

Stack Overflow dataset provides information only about low-level interactions (e.g. creation of a new question). In order to confirm the stated hypotheses, it was necessary to process users' actions to more abstract variables.

Topic. To model question and answer topic, we decided to employ tags assigned by users at the time of the question creation. Corresponding answers inherit topics from the question they are related to. Some other studies supplement user-assigned tags by latent topics identified by methods such as Latent Dirichlet Allocation (LDA), e.g. [7]. However, this approach is important especially when questions are assigned only to one category (e.g. in Yahoo! Answer) and thus a little information is known about a question's topic. On the other side, in Stack Overflow a question can have unlimited number of tags (each question has 2.95 tags in average in the selected dataset).

Answer Quality. Determination of answer quality in CQA systems is quite a difficult and challenging task because most of questions can be subjectively oriented and the context of each question and corresponding asker is unique and sometimes not well known. Several different techniques have been used to determine answer quality in CQA systems so far. Some authors relied on an assumption than an answer is of high quality if it was selected by an asker as the best answer. Authors in [6] pointed out a problem of this 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. Another possibility how to achieve more precise answers' quality estimation is manual evaluation what is, however, really time consuming and thus it is not possible to apply it on great datasets. Finally, Stack Overflow, as well as other CQA systems, provides users with a voting mechanism which determines answers' quality by means of score. Score refers to a sum of positive and negative votes on questions as well as answers. Positive votes on an answer determine its correctness and giving a negative vote means that the answer is incomplete, incorrect or biased. More than 70% of all votes are created within 48 hours after the corresponding post was created. Therefore, we can consider score as a quite precise estimation of answer quality which is mostly independent on time when the answer was posted.

By analyses of Stack Overflow dataset, we found out that values of answer score follow a typical long-tail distribution and thus any calculation with them can be significantly skewed. For this reason, we performed two transformations on the score:

1. A logarithmic transformation to eliminate the undesired power law distribution.
2. A min-max normalization which transforms positive values to interval $\langle 0,1 \rangle$ and negative values to interval $\langle -1,0 \rangle$. This normalization is important because negative votes in the selected dataset represent only 11% of all votes and thus we emphasized their weight.

Consequently, we propose to calculate quality of an answer A_i as a normalized score (S') increased in the case when the answer was selected as the best answer (BA). We empirically set the influence of asker's best answer vote/selection as a value 0.1 that corresponds approximately to 5 standard up-votes by users other than the asker.

$$quality(A_i) = S'(A_i) + BA(A_i) \quad (1)$$

$$BA(A_i) = \begin{cases} 0.1 & if\ A_i\ was\ selected\ as\ the\ best\ answer \\ 0 & otherwise \end{cases} \quad (2)$$

User topical expertise. In the most of research works, an estimation of users' expertise is based on their previous contributions. Zhang et al. [10] proposed a very simple metric named z-score which describes how many answers and questions a user previously posted in the CQA system: $Z\_score = (a - q)/\sqrt{(a + q)}$, 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. Bouguessa et al. [3] proposed a probabilistic model that is based on another simple metric named InDegree. This metric represents only a number of best answers provided by the particular user.

In our approach, we similarly utilize the previous user activity. However in order to derive user expertise more precisely, we estimate it from quality of previously provided answers. More specifically, user expertise is calculated as a cumulative average of answers' quality. In addition, it is calculated separately for each topic (tag). It means that users can have different values of expertise for various topics and thus we are able to model real user expertise even more precisely. Similarly as all approaches based on previous user activities, also our approach has a drawback that we are not able to calculate user expertise for users with no or minimal activity in the CQA system. It means, that we have an estimation of user expertise with a high degree of uncertainty at the beginning and with the following answers, we are able to refine the expertise level and estimate it with a significantly higher confidence.

3.3 Evaluation of Knowledge Acquisition Scenarios

We employed the estimation of answer quality and user topical expertise to evaluate the stated hypotheses. At first, we identified all occurrences of three analyzed scenarios for each user and separately for each tag he or she provided answers on. Consequently, we calculated two numbers each time when user expertise has changed (i.e. when the particular user provided a new answer):

1. Number of scenario occurrences in the time interval from the analyzed point in time until user's last activity in the selected dataset.
2. Relative change of user's expertise between the analyzed point in time and user's last activity in the dataset.

Time intervals with a low-confidence estimation of user expertise were omitted because they can bring an undesired distortion to the evaluation. Finally, the calculated relative change of expertise was averaged across all users and tags with the respect to the number of scenario occurrences (see Figure 2).

The obtained results pointed out that all three scenarios are positively related with user expertise. We can see a logarithmic distribution for scenario H2 (when an answerer provides a high quality answer) and H3 (when an answerer provides an answer besides other high quality answers). It means that these two scenarios provide the best potential to boost knowledge acquisition. On the other hand, influence of scenario H1 (when an asker receives a high quality answer) follows rather a linear trend.

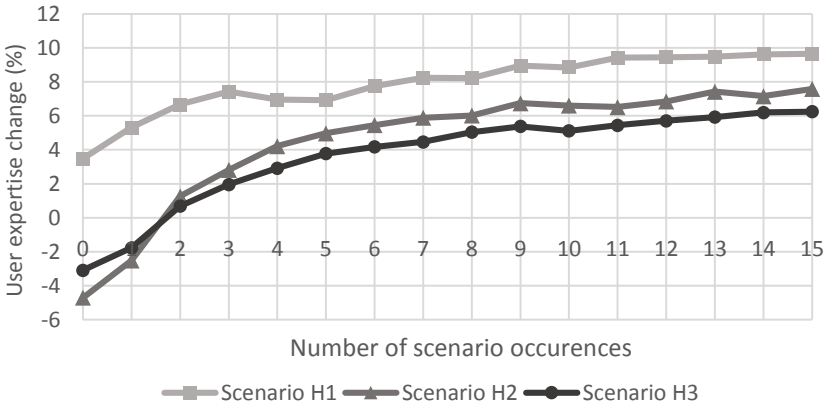

Fig. 2. Relation between average change of user topical expertise and the number of scenario occurrences in the observed time intervals.

To evaluate the relation between variables numerically, we used Kendall tau rank correlation coefficient as an evaluation metric. Kendall tau correlation was selected due to the non-linear character of evaluated relations as well as its better robustness in comparison with other standard correlation coefficients (i.e. Spearman rank correlation and Pearson correlation). All scenarios achieved strong correlations with a high significance (see Table 1).

Table 1. Overview of the achieved Kendall tau correlations for the evaluated hypotheses.

Hypothesis	*Correlation*	*P-value*
H1 (when an asker receives a high quality answer)	0.9167	< 0.001
H2 (when an answerer provides a high quality answer)	0.9333	< 0.001
H3 (when an answerer provides an answer besides other high quality answers)	0.9830	< 0.001

4 Conclusion and Future Work

Without any doubt, fast and reliable availability of information is the crucial step to success in business as well as in academia. We consider Community Question Answering (CQA) systems as promising knowledge management systems that appeared only recently. While the primary goal of CQA systems is to provide high quality answers on new questions in the shortest possible time, we recognized their potential to become novel learning systems that supplement a formal educational process.

On the basis of the dataset from Stack Overflow, which is one of the most popular CQA systems, we identified three scenarios that positively contribute to development of users' knowledge. The results of our study give us a deeper insight into the learning potential embedded in the question answering process. More specifically, the identified

knowledge acquisitions scenarios can be utilized by researchers to propose more effective educational environments or methods for collaboration support. For example, during a personalized recommendation of questions to students, we can promote those recommendations that will direct students into the identified scenarios.

In our additional work, we took into consideration the knowledge acquisitions scenarios together with organizational and educational specifics to propose an educational organization-wide CQA system named Askalot. Askalot demonstrates the learning potential of CQA system as a complementary tool dedicated to knowledge sharing besides other educational systems ALEF [9] and PopCorm [8], which were developed and are used at our faculty.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11 and by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. 009STU-4/2014.

References

1. Aritajati, C., Narayanan, N.H.: Facilitating students' collaboration and learning in a question and answer system. Proc. of the 2013 conf. on Computer supported cooperative work companion - CSCW '13. pp. 101–106. ACM Press, New York, USA (2013).
2. Barr, J., Gunawardena, A.: Classroom salon: a tool for social collaboration. Proc. of the 43rd ACM technical symposium on Computer Science Education - SIGCSE '12. pp. 197–202. ACM Press, New York, USA (2012).
3. Bouguessa, M., Dumoulin, B., Wang, S.: Identifying authoritative actors in question-answering forums. Proc. of the 14th ACM SIGKDD international conf. on Knowledge discovery and data mining - KDD 08. pp. 866–874. ACM Press, New York, USA (2008).
4. Lévy, P.: Collective Intelligence: Mankind's Emerging World in Cyberspace. Perseus Books, Cambridge, MA, USA (1997).
5. Ram, A., Ai, H., Ram, P., Sahay, S.: Open social learning communities. Proc. of the International Conf. on Web Intelligence, Mining and Semantics - WIMS'11. ACM Press, New York, USA (2011).
6. Sakai, T., Ishikawa, D., Kando, N., Seki, Y., Kuriyama, K., Lin, C.-Y.: Using graded-relevance metrics for evaluating community QA answer selection. Proc. of the fourth ACM international conf. on Web search and data mining - WSDM '11. pp. 187–196. ACM Press, New York, USA (2011).
7. Szpektor, I., Maarek, Y., Pelleg, D.: When Relevance is not Enough : Promoting Diversity and Freshness in Personalized Question Recommendation. Proc. of the 22nd international conf. on World Wide Web. pp. 1249–1259 (2013).
8. Srba I., Bieliková, M.: Designing Learning Environments Based on Collaborative Content Creation. In Proceeding of Workshop on Collaborative Technologies for Working and Learning, pp. 49–53, CEUR, (2013).
9. Šimko, M., Barla, M., Bieliková, M.: ALEF: A Framework for Adaptive Web-Based Learning 2.0. In Proc. of IFIP Advances in Information and Communication Technology, Vol. 324, pp. 367-378. Springer, Heidelberg (2010).
10. Zhang, J., Ackerman, M.S., Adamic, L.: Expertise networks in online communities. Proc. of the 16th international conf. on World Wide Web - WWW '07. pp. 221–230. ACM Press, New York, USA (2007).