CQAVis: Visual Text Analytics for Community Question Answering

Enamul Hoque[†‡], **Shafiq Joty**[†], **Lluís Màrquez**[†], **and Giuseppe Carenini**[‡]
[†]ALT group, Qatar Computing Research Institute, HBKU, Qatar Foundation
[‡]Department of Computer Science, University of British Columbia
[†] {sjoty,lmarquez}@qf.org.qa,[‡] {enamul,carenini}@cs.ubc.ca

ABSTRACT
Community question answering (CQA) forums can provide effective means for sharing information and addressing a user's information needs about particular topics. However, many such online forums are not moderated, resulting in many low quality and redundant comments, which makes it very challenging for users to find the appropriate answers to their questions. In this paper, we apply a user-centered design approach to develop a system, CQAVis, which supports users in identifying high quality comments and get their questions answered. Informed by the user's requirements, the system combines both text analytics and interactive visualization techniques together in a synergistic way. Given a new question posed by the user, the text analytic module automatically finds relevant answers by exploring existing related questions and the comments within their threads. Then the visualization module presents the search results to the user and supports the exploration of related comments. We have evaluated the system in the wild by deploying it within a CQA forum among thousands of real users. Through the online study, we gained deeper insights about the potential utility of the system, as well as learned generalizable lessons for designing visual text analytics systems for the domain of CQA forums.

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User Interfaces
I.2.7 Natural Language Processing:Text analysis

Author Keywords
Asynchronous conversation; community question answering; text visualization; computer-mediated communication;

INTRODUCTION
Community question answering (CQA) forums, such as StackExchange, Yahoo! Answers, and Quora are becoming more and more popular these days.[1] They represent effective means for communities of users around particular topics to share information and to collectively solve their information needs. CQA forums typically organize their content in the form of multiple topic-oriented *question–comment threads*, where a question posed by a user may be answered by a possibly long list of comments from other users.

Many such online forums are not moderated, which often results in very noisy and redundant content. Users tend to deviate from the original question and engage in discussions on completely irrelevant or only loosely related topics. At the same time, similar questions may be posted repeatedly with minor variations. This near-duplicity is difficult to track for users, who are usually offered only simple search capabilities by the forum interface. When relevant answers to user questions are scattered around multiple related conversations and buried among a large number of comments, the user is facing a challenging information processing task, which, without proper support, leads to information overload.

For example, consider John, who is an expatriate, arrived in Qatar and is seeking recommendations for a good bank. When he searches for 'Which is the best bank in Qatar?' in the Qatar Living forum[2], a very popular site in Qatar, it returns about a dozen of previously asked questions, such as ' What is the best bank to open an account?' or 'what is the best bank in Qatar for small business?' (see Figure 1). Each of these questions followed by a set of comments, resulting in hundreds of comments in total. Given the large number of comments from multiple related threads, it would be very difficult and time-consuming for John to identify and make sense of useful comments using a traditional interface.

While recently some natural language processing (NLP) techniques have been proposed to address this and similar problems [28, 29], there has been very little work to combine these techniques with information visualization to create intelligent interfaces to CQA forums. In this paper, we present CQAVis, an intelligent visual interface specifically tailored to help users to find comments that provide good answers to a new question (i.e., never asked in exactly this form before) in community-created forums. The system tightly integrates search and NLP technology (i.e., text analytic) with information visualization (InfoVis) to help users navigate the comments more effectively. CQAVis interface allows the user to start with a new question,

[1]stackexchange.com, answers.yahoo.com, quora.com

[2]http://www.qatarliving.com/forum

IUI 2017, March 13-16, 2017, Limassol, Cyprus
©2017 ACM. ISBN 978-1-4503-4348-0/17/03 $15.00
DOI: http://dx.doi.org/10.1145/3025171.3025210

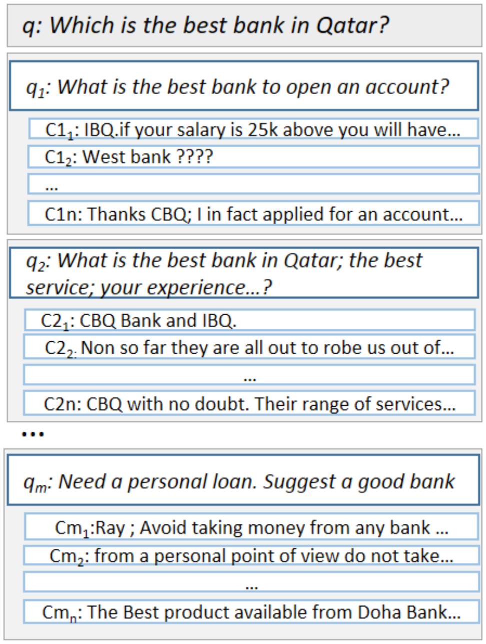

Figure 1: An example of a new question q asked by a user (shown on top) followed by a set of related thread questions $(q_1, \ldots, q_n)$ and their comments.

then to explore the related threads to find the ones that seem to be most relevant to her information needs, and eventually to navigate through the comments of a thread in search for relevant answers to her question. The underlying text analytic module dynamically ranks potential answers to a new question by combining two relevant measures: (*i*) how good or useful the comment is with respect to the thread question (e.g., q_1, q_2 in Figure 1), and (*ii*) how similar the thread question is with respect to the new question (q).

Our system was deployed in the Qatar Living forum site to evaluate our interface among hundreds of real users. Qatar Living forum was suitable for our study, because it represents the type of forums where the information overload problem (as described above) could be more prevalent due to unmoderated noisy content. Moreover, a large number of its users have limited expertise in using visual interfaces, which poses critical challenges to designing interfaces.

The primary contributions of our work include: 1) characterization of the CQA forums by identifying user tasks and some key design needs; 2) tightly integration of a novel set of NLP and InfoVis techniques to meet these needs; 3) the evaluation of the tool in the wild in an ecologically valid testing by deploying the system among real forum readers; and 4) generalizable lessons learned from the study that can be useful to design visual interfaces for online conversations in other domains (e.g., blogs, news comments), as well as to design for user population possibly having low visualization literacy.

RELATED WORK

Text Analytics for CQA

Several text analytics tasks have been explored for CQA. Three tasks are of special relevance when a user poses a new question to the website: (*i*) find existing related questions to the new question (i.e., *question-question similarity*), (*ii*) find relevant answers to the new question (i.e., *answer selection*), and (*iii*) find potential users who may provide good answers to the new question (i.e., *user selection*) [18, 19]. Another task that can help solving other tasks better is determining whether a comment within a question-comment thread is a good answer to the question of that thread (i.e., *answer goodness*).

Early attempts have mostly focused on bag-of-words models [20] and concept matching [45] to measure similarity between two texts (e.g., the question and a candidate answer). More recent work focus on ways to produce good syntactic and semantic representations to better capture the similarity between texts, mainly through the use of distributed representations and neural networks. Several neural models have been proposed for CQA tasks such as *answer goodness* [47], *question-question similarity* [46, 8], and *answer selection* [42, 32, 38, 10, 39]. Most of this work used advanced architectures based on convolutional neural network (CNN), long short-term memory (LSTM), attention mechanisms, etc. For example, Zhou et al. [47] treated the *answer goodness* task as a sequence labeling problem and used recurrent CNNs and LSTMs. Dos Santos et al. [8] combined CNN and bag-of-words representations for comparing questions.

Another line of research in CQA uses tree kernels (TK) to measure similarity between two texts based on their syntactic representations. Nicosia et al. [30] achieved the best results on the answer goodness task in SemEval-2015 using Support Vector Machines (SVM) with TK and other numeric features. Barrón et al. [2] use thread-level features for the same task. In our work, the classifier for *answer goodness* uses comment- and thread-level features, and high-level feature representations learned from neural models in a TK-based SVM classification model. As we describe later, this classification is precomputed in our system. However, the *question-question similarity* is measured dynamically, and to make our system efficient for the users, our method directly uses the rank given by the search engine.[3] An initial demo of our text analytic system was presented in [17].

Visual interfaces for Online Conversations

Previous work on visualizing online conversations has often focused on supporting the users in exploring comments of interest by visualizing thread structures. Several techniques were developed to reveal the thread structure of a conversation using tree visualization techniques, such as using a mixed-model visualization to show both chronological sequence and reply relationships [40], thumbnail metaphor using a sequence of rectangles [43, 21], and radial tree layout [31]. However, such visualizations do not analyze the actual content (i.e., the text) of the conversations. In contrast, those that analyze the textual content of conversations primarily focused on summarizing and showing the major themes of conversations [34, 6], or visualizing the content evolution over time [44, 41]. More recently, a visual text analytic system named ConVis [13] is presented, which provides a visual overview of an online

[3]The best system participated in SemEval'16 *question-question similarity* task is slightly higher than the search engine results.

conversation and allows the user to navigate through the conversation based on topics, sentiments and different metadata. This system further supported the user in revising the topic model when the current results are not adequate to fulfill her information needs [14, 15]. Later, this interface was extended to allow the user to explore a collection of conversations [16]. However, the above work mainly focused on supporting exploratory tasks, whereas within the domain of CQA forums users have more specific questions in mind and the system needs to help users to find answers to those questions.

Unfortunately, only a few works are focused on supporting the tasks of finding high quality answers in the domain of CQA. Liu et al. [24] developed a location-aware real-time CQA interface, which routes a newly posted question to a ranked list of users through various recommendation algorithms. Through user studies they compared between different recommendation algorithms for user-ranking and question-ranking. However, this interface is limited for mobile devices and rely on the assumption that there will be active users at the same time in the same location of the asker, which often may not be the case for the CQA forum we are considering.

Finally, CommentIQ [7], a tool for supporting the task of news comments moderation was developed by combining text analytic and visualization techniques informed by a user centered design. In our work, we have followed a similar design approach, but we have studied a different domain (CQA vs. news comments) and more importantly our target users are different (i.e., novice users as opposed to expert moderators).

THE DESIGN PROCESS

Our design study process followed the nine-stage framework proposed by Sedlmair et al.[35]. In particular, we focused on four core phases of the design framework: **1) Discover**: In this stage, we analyzed the needs, problems, and requirements in the domain of CQA forums through literature review and conducting in-depth interviews with Qatar Living forum users and administrators. **2) Design**: After reaching a shared understanding of the CQA domain we explored the design space of text analytics and visualization to support users. We have applied an iterative design approach, starting with paper prototyping, followed by prototyping on a limited annotated dataset which led us to the final prototype on the whole forum datset. **3) Implement**: We developed both client and server side components in collaboration with the Qatar Living administrators. **4) Deploy**: Following several pilot studies and corresponding refinements of the prototype, we deployed the tool as a beta version in the Qatar Living website.

USER REQUIREMENTS ANALYSIS

Domain Characterization

To characterize the domain of question answering (QA) forums, we analyzed existing literature in the areas of human-computer interaction and computer supported collaborative work, focusing on what types of questions are asked [11, 25, 18], who answers and why [9, 25] and what are the predictors for answer quality [11].

Subjective nature of questions: Several research works found that there are more subjective and opinion-based questions than factual questions [25, 18, 12]. Morris et al. surveyed QA users and found that only 17% questions they asked were seeking factual information, while the most common categories of questions were requests for recommendation (29%) and opinions (22%) [25]. Similar results were found for a social-question-answering system, with 64.7% of the queries were found to be subjective [18]. Due to the nature of these questions, e.g., ‘best Italian restaurant in Doha’, often any particular subjective answer may not satisfy the information needs, therefore the user interface should effectively support browsing various answers from multiple related threads.

Variability in answer quality: Previous work also analyzed the characteristics of good answers. Harper et al. conducted a controlled field study to analyze different predictors of answer quality across several QA sites [11]. They found that while QA site like Yahoo! Answers provide lots of high-quality answers, users should also expect substantial variability in the quality of individual answers. To address this issue, it may be useful to apply an automatic approach of identifying high quality answers and help users to navigate through these answers.

Slower response: Some researchers have explored the factors affecting answer quality and response time on QA sites. Raban and Harper identified both intrinsic factors (e.g., perceived ownership of information, gratitude) and extrinsic factors (e.g., reputation systems) that motivate CQA users to answer questions [33]. However, even when motivated people are available to answer, their response times tend to be longer [25].

Interviews: In addition to analyzing existing literature, we also conducted two semi-structured interviews and a number of follow-up interviews with our collaborator at Qatar Living. The goal was to understand more specific needs and requirements for the type of forum that Qatar Living represents to directly inform our design process.

Many naive users: Qatar Living is one of the most popular sites in Qatar, with over 550,000 visitors per month and over 19 million page views a month from Qatar. Its forum is actively visited by hundreds of users everyday, who mainly try to fulfill their information needs in their topics of interest. However, a large portion of the forum users are naive and they are not proficient with sophisticated user interfaces. Therefore, an important design goal is to make the interface simple and intuitive. In addition to naive and non-expert users, there is a dedicated small group of forum moderators having higher level expertise about the topics. These users actively browse the new questions posted in the forum and try to answer them depending on their expertise. While we have mainly focused on supporting the former group of users, we argue that moderators can also benefit from our visual interface for their tasks.

Searching for previous questions rather than asking new ones: Our collaborator pointed out that usually the readers try to get their questions answered quickly. So, they often prefer to use the search feature within the forum to find similar questions to their current question, rather than posting their questions and waiting for answers. However, they have difficulty in exploring the similar question threads due to large volume of comments they need to read, which is time consuming and cumbersome using the existing search interface. This suggests the pressing need for improving the search interface for the forum to enhance the user's ability to find good answers.

Difficulty in finding good answers: Like many other CQA sites, Qatar Living forum contents are often noisy and redundant. Users tend to use very informal language, often writing very long stories with small pieces of relevant text only. Due to noisy and redundant content, the question threads can become longer with only a few relevant answers. As a result, searching for relevant answers often leads to the information overload problem. To make matter worse, although there is a upvoting/downvoting feature, most users either do not know how to use this feature or they do not bother to do it. Therefore, our collaborators agreed that an automatic comment classifier that is reasonably accurate can be effective in identifying good answers. More importantly, the interface should facilitate the user to find the good answers, which may be scattered among the large amount of comments from multiple different threads.

In summary, the system for supporting the information seeking tasks in the forums like Qatar Living (that are typically unmoderated, contain near duplicate questions and lot of noisy comments), should consider following user requirements: 1) The interface should effectively support the user in identifying multiple good answers from related question threads; 2) To address the variability in the answer quality, a classifier should be introduced to identify useful comments. 3) The interface should introduce interactive visualization components to enhance the user's ability to find good answers from large volume of comments. 4) To support users having lower visualization expertise, the interface should be simple and intuitive.

Data and Task abstractions

Tasks: In our conversations with the Qatar Living admin, we learned several use-cases and tasks of the forum users. We analyzed these tasks according to a visualization task typology [27] in order to inform our design. At the high level, users are primarily interested in seeking information with a goal to *discover* new information or knowledge. At this level, the user may ask questions like "Which is the best bank in Qatar?" or " Where can I find a good Chinese restaurant in Qatar?". Once the user is presented with some related questions to her new questions, the next level task is to *search* for the most related questions of interest by *browsing* the list of questions presented to her. When they find the most related questions from the list, next they focus on *identifying*, *comparing* and *summarizing* the most useful answers to her original question.

Data: Based on our user requirements analysis, we derive how the data should be abstracted for visualizing to the user. As illustrated in Figure 1, an example dataset consists of a question asked by a user with the set of related questions found by the system. We encode the relatedness of a question to the new question by a rank value (ordinal). Each related question is also followed by a set of comments that tried to answer that question. For each of these comments, we derive the *goodness* score provided by a classifier with respect to its related question and represent as a normalized quantitative value between [0.0,1.0], by passing the score through a sigmoid function. We also assign each comment into one of six equally sized bins depending on its classification score to help the user understand how relevant a particular comment is. Based on this binning, we also compute the distribution of comments for each related question thread by counting how many comments fall into a particular bin. We compute this distribution, because it can effectively convey to the user how many comments are useful among all the comments of a question.

SYSTEM OVERVIEW

Figure 2 presents an overview of our system, which is organized in two parts. In the offline step (Figure 2a) we preprocess the datasets and we train a comment classifier. In the online regime (Figure 2b) the user enters a question as input, and the system performs three steps on the fly: retrieving the top n related question threads, ranking all the answers, and visualizing the results. We briefly discuss these steps below.

Offline Processing

To build the system, we used a recent dump of the Qatar Living forum (from March 2016), and we performed several preprocessing steps including the conversion of the XML dump to JSON format that our interface can process. This dump contains $202,304$ conversations and $2,043,022$ comments (On an average, each conversation consists of 10.21 comments).

We also used the datasets on CQA from SemEval-2016 Task 3 (subtask A), where the comments in the threads are manually annotated with `good` vs. `bad` labels, indicating how well the comments answer the question in the thread. Using this dataset, we extracted a collection of features and we trained an SVM-based comment classifier (see details in the next section) that scores each comment in a thread regarding its *goodness*.

Online Processing

When a user types a new question q, the system performs the following three steps on the fly: (*i*) *Retrieve related questions*, where Google local search is invoked to retrieve the top-n question threads in the Qatar Living forum that are most similar to q, $\{q_i\}_{i=1}^{n}$; (*ii*) *Rank the answers*, where all the comments from these top-n question threads are ranked based on their relevance with respect to q. (*iii*) *Visualize the results*, where the presentation module takes the related questions' threads together with the ranked lists of comments and the overall best selected answer, and presents them to the user.

TEXT ANALYTICS

The answer ranker module computes the relevance score of a comment c in a question thread q_i with respect to the new question q by combining two scores: (*i*) $\sigma(q,q_i)$, the similarity of q_i to q; and (*ii*) $\gamma(c,q_i)$, the goodness score for c with respect to q_i. Formally, the relevance score $\rho(c,q,q_i)$ is computed by:

$$\rho(c,q,q_i) = \sigma(q,q_i) \times \gamma(c,q_i) \qquad (1)$$

We use the inverse rank in the list returned by the Google search engine as $\sigma(q,q_i)$, and $\gamma(c,q_i)$ is computed by a comment classifier, indicating how well comment c answers q_i. The resulting score is used to rank all the comments from the retrieved question threads to obtain the best overall answer to the input question q. Intuitively, if a comment is a good comment with respect to the thread question, and the thread

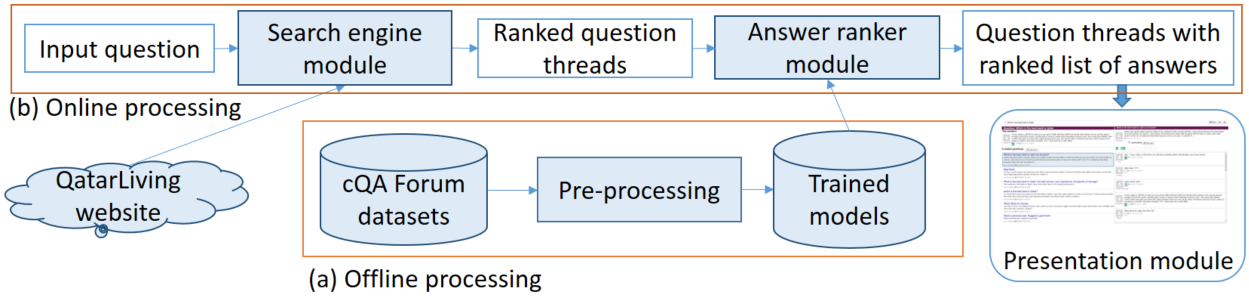

Figure 2: Overview of our interactive system for supporting community question answering.

question is related to the new question, then the comment is likely to be a relevant answer to the new question.[4] The core NLP component of this architecture is the comment classifier, which is briefly described below.

Comment Classifier

We use the comment classifier presented in our demo paper [17]. Given a question q and a list of comments associated with it $\{c_i\}_{i=1}^{m}$, the task of the classifier is to assign a relevance score to each of the comments according to their *goodness* at answering the question. This very problem was set at SemEval-2016 Task 3 [29], subtask A. We trained an SVM classifier on that dataset to distinguish between good and bad comments.[5] The dataset is split into training, development and test sets, with 2,669, 500, and 700 questions, and 17,900, 2,440, and 7,000 answers, respectively. The kernel function in our SVM is a linear combination of four functions: two linear kernels over numeric features and embeddings, and two tree kernels over shallow syntactic trees.

Numeric Features These features are inspired by [2, 30]. They include three types of information: (*i*) a variety of textual similarity measures computed between the question and the comment; (*ii*) several Boolean features capturing the presence of URLs, emails, positive/negative words, acknowledgments, forum categories, long words, etc.; (*iii*) a set of global features modeling dialogue and user interactions in the thread.

Embedding Features Higher level abstract features learned automatically by deep neural networks have proved to be quite beneficial for learning semantic similarity between two texts [37, 8, 32, 38]. We learn embeddings for questions and answers by training a convolutional neural network (CNN) on the comment classification task following the approach of [37]. Specifically, the input to the CNN is formed by two matrices containing word embeddings for the question and for the answer, respectively. The CNN performs a *convolution* and a *max-pooling* operations on the word embeddings and on the convoluted feature maps, respectively, to produce the question embedding q_E and the answer embedding c_E. These embeddings are then combined to produce a similarity value using a similarity matrix. The similarity and the embeddings along with other additional similarity features are then passed through a hidden layer and next to the output layer for classification. q_E and c_E are learned by backpropagating the (cross entropy) errors from the output layer. q_E and c_E vectors are finally concatenated and used as features in our SVM model.

Tree kernels Tree kernels provide effective ways to learn by comparing syntactic structures of two texts in the SVM framework, which has been shown to give state-of-the-art results in CQA [30]. First, we produce shallow syntactic trees for the question and for the comment using the Stanford parser. Following [36], we link the two trees by connecting nodes such as NP, PP, VP, when there is at least one lexical overlap between the corresponding phrases of the trees, and we mark those links using a specific tag. The kernel function K is defined as: $K((q_1,q_2),(c_1,c_2)) = TK(q_1,c_1)+TK(q_2,c_2)$, where $TK(q,c)$ is a tree kernel function operating over a pair of question (q) and comment (c) trees.[6]

Classification Performance We evaluated our comment classifier on the SemEval-2016 test set with the official scorer, obtaining the following results: MAP=77.66, AvgRec=88.05, MRR=84.93, F_1=66.16, Acc=75.54. Compared to the participant systems at SemEval-2016, our system scores in second position regarding the official MAP evaluation metric (-1.5 points below the best). In contrast, we achieve better F_1 (+1.8) and better Accuracy (+0.4) than the top system. For a full description of the results from SemEval-2016, see [29].

CQAVIS DESIGN AND IMPLEMENTATION

In order to explore a large number of design choices, we carried out an iterative design process, starting from early mockups and prototypes using paper and Powerpoint. We then developed a mid-level prototype which works on a small CQA annotated corpus [29], where the comments are annotated with good vs. bad labels by human experts. Finally, we developed a fully functional system and deployed within a real CQA site. Throughout the design process, we performed formative evaluations [23] to identify potential usability issues and to iteratively refine the prototype. We now present the final design of the CQAVis interface[7], along with justifications for the key design decisions based on our user requirements analysis and the InfoVis literature.

[4] As discussed in the SemEval-2016 Task 3 description paper [29], this is a very simple way to obtain good results for the general task of ranking answers for new questions.

[5] The conversations in the SemEval dataset was written in the same language as the QatarLiving forum site.

[6] We use Partial Tree Kernel and Syntactic Tree Kernel [26, 5] to instantiate TK.

[7] A live demo of CQAVis is available at iyas.qcri.org

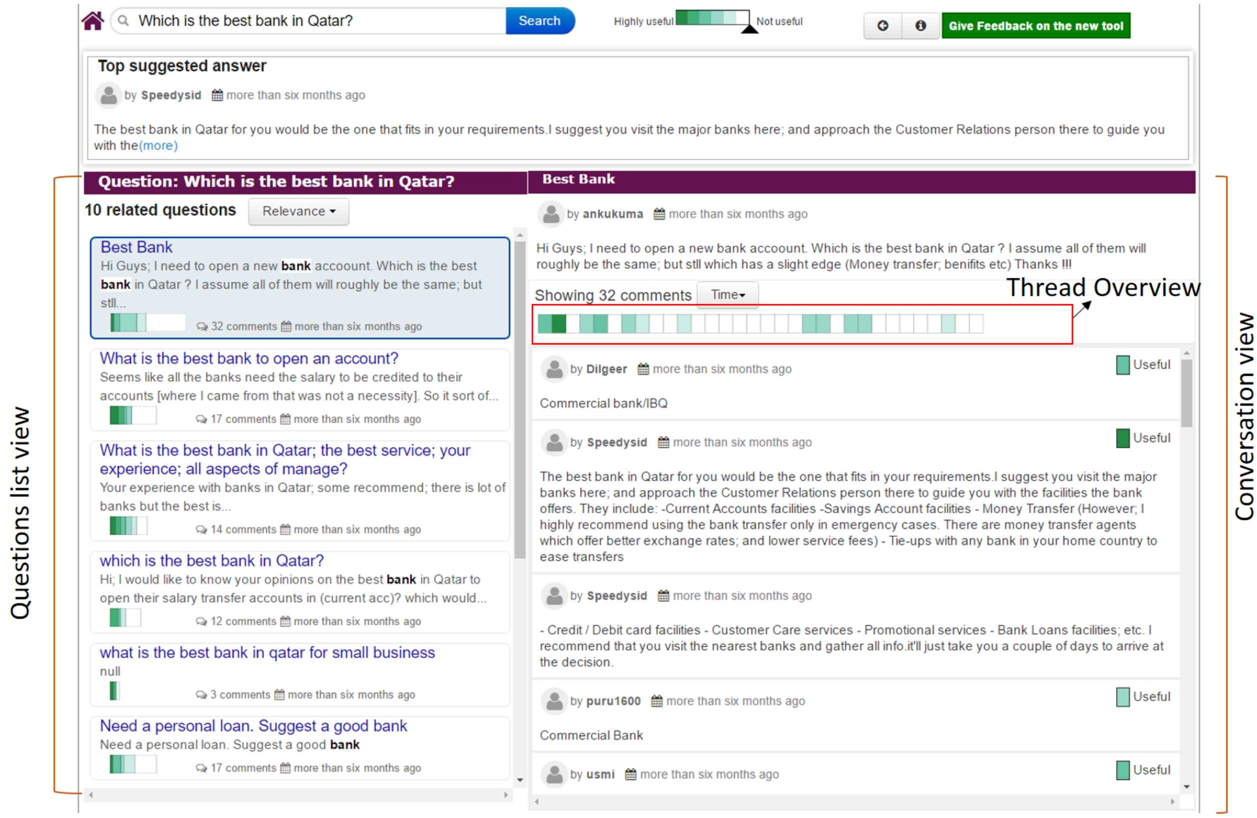

Figure 3: A screenshot of the interface showing the top answer and related questions for a user's question. As the user selects a related question (marked by the blue rectangular boundary), the interface shows the corresponding thread in the conversation view.

A high level design decision for the interface was to follow an overview+detail approach, where the overview represents the *question list view* showing the top-most relevant questions to the user's question; and the detail view (i.e., *conversation view*) showing the question followed by the answers for a particular question thread (see Figure 3). We made this choice because this allows users to browse comments of a specific question, while still having the context of the other related questions, and also because this approach has been found to be more effective for text comprehension tasks than other approaches such as zooming and focus+context [4].

Questions list view

After the system finds the related questions to the user's question, it presents an overview of the the ranked list of relevant questions in a scrollable list view (see Figure 3, left). Each item within the *questions list view* represents a question thread, showing a set of metadata i.e., the original question, the posting date, the total number of comments, as well as a stacked bar with the distribution of useful comments. Since we are representing an ordered sequence of values, we used a set of six sequential colors by varying monotonically on the green color channel ranging from dark green (highly useful) to white (not useful) chosen from ColorBrewer [1]. In this way, the user can quickly get a sense of which threads seem to be more relevant and which threads may contain the most useful answers.

The questions are ordered by their relevance rank by default, but the user can change this order by selecting criteria from the popup menu 'Order by'. For instance, she can order the question threads based on the number of useful answers within each of these threads.

Another important feature of the interface is that at any time the user can filter out comments with low usefulness score by using the slider of the widget (containing sequence of colored rectangles) at the top (see Figure 3). In this way, the user can quickly narrow down the set of less useful comments of different question threads and focus on the ones that are potentially good answers to her question.

Note that at the top of the question list view, the interface also shows the comment that has received the best score with respect to the new question ("Top suggested answer"). This feature was designed to support the user in finding a very good answer to her question immediately, without having to open any question thread and then navigating to answers within that thread. This was motivated by the user requirements analysis, from which we learned that users would like to find some very good answers quickly, therefore showing the top ranked answer right away could be very useful.

Conversation view

When the user selects a particular question thread from the list, the system presents the corresponding thread in the conversation view (see Figure 3). Again, we followed an overview+detail approach, where at the top we show a visual overview of the entire thread along with the question, followed

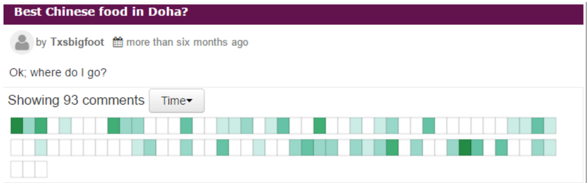

Figure 4: An example of a thread overview that splits a large number of comments into multiple rows to deal with horizontal space constraints.

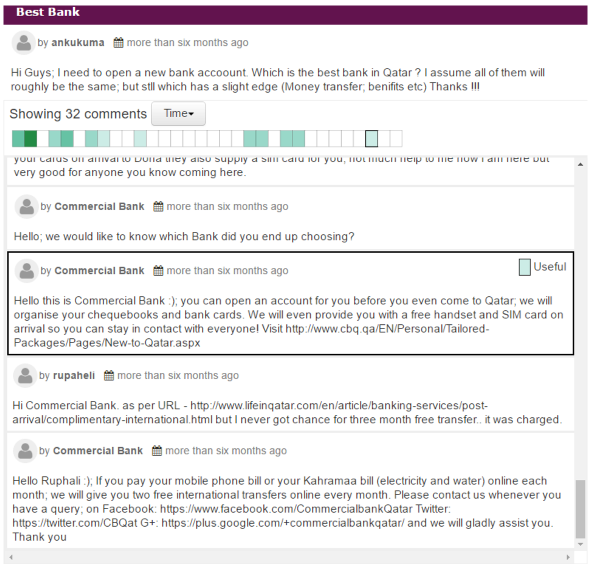

Figure 5: When the user clicks on a rectangle in the thread overview representing a comment, the interface scrolls to that comment (marked by black color) in the conversation view.

by a detail view containing the list of comments. Here, the thread overview visually encodes the comments using a sequence of rectangles from left to right, where each rectangle represents a comment. The color within each rectangle encodes the classification score of the comment represented by that rectangle. If the horizontal space is not sufficient for showing all the comments, then it shows the rectangles in multiple rows as shown in Figure 4. In this way, the thread overview visualization can scale with hundreds of comments, which is sufficient for a typical CQA forum conversation.

From the thread overview, the user can quickly notice which comments are more useful and then immediately navigate to a particular comment by clicking on the rectangle representing that comment (see Figure 5). Note that the two views are coordinated, i.e., hovering on a rectangle in the thread overview highlights the corresponding comment in the detailed view (by scrolling if needed) and vice-versa. Finally, the user can reorder the comments of a thread based on their classification score to quickly go through the most useful answers.

Throughout the design of CQAVis, an important goal was to make the interface simple and intuitive for the naive users, who constitute a large portion of users of Qatar Living and similar forums. To achieve this goal we focused on using visualization metaphors that are common and easily understood (e.g., bar graph based visualization and sequence of rectangles) and a small set of simple, low cost interactions [22] that can be easily triggered and reversed.

Implementation

The system is implemented as a Java Web application and runs on an Apache Tomcat Server. The back-end of the system is developed using Java. The presentation module, on the other hand, is implemented in Javascript (using the D3 and JQuery libraries). The system is sufficiently fast to respond in real time to the user's actions. A key factor for the efficiency is the fact that we precomputed and stored the goodness scores for all the comments in all the question-threads from the static snapshot of the Qatar Living database. In this way, at running time there is no need to classify the comments of the already stored question-comment threads.

WEB-BASED USER STUDY

To better understand the potential utility of our approach in real world scenarios we undertook a large-scale, Web-based study. The primary aim of our study was to empirically examine how real users would use CQAVis and what their reaction would be to such an InfoVis-enabled search interface. The main research questions were: 1) What are the possible benefits and limitations of the CQAVis interface in supporting the task of information seeking? 2) When we compare CQAVis with a typical interface for forum search (as instantiated by Qatar Living forum), is there any difference in subjective reactions?

Methodology

While a Lab-based user study would allow us to have more control over the users and tasks, realism would be largely lost [3]. Therefore, we decided to run the study in Web-based environments to enhance the ecological validity, since participants can then work in their own settings performing their own tasks [23]. It also gives us the advantage of collecting interaction logs from a large number of users to get deeper insights that are arguably more generalizable than a lab-study.

Study setup and procedure

In order to run the user study, we discussed with our collaborators at Qatar Living, who agreed to incorporate our web-based tool as a beta version of the forum site. Our system was deployed at a server and then a Web-link of the system was made available on the forum search page for the real users of the Qatar Living forum. To avoid compatibility issues, we tested our interface on the Web browser versions of Mozilla Firefox, Apple Safari, and Google Chrome to ensure that we can support a wide range of participants.

Participants were guided through three main steps of the study: 1) *Introduction*: In the home page, some background information and example queries were provided to get started, along with an invitation to use the interface. The page also contained a short video to demonstrate the main features of the interface. 2) *Interaction*: The main part of the study was the interaction with CQAVis. Here, users were not asked to complete any specific task; instead they could perform their own set of information seeking tasks. 3) *Feedback*: Participants were free to fill up a post-study questionnaire at any time

	Min	Median	Mean	St. Dev	Max
Length of session (seconds)	1.79	142	416	666.45	3,327
Queries per session	0	1	1.47	1.54	16
Query length (characters)	1	20	22.81	14.4	200

Table 1: Overview of user study sessions and queries.

Question Type	Percent	Example
Recommendation	21.82	Where can I find italian restaurants in Doha ?
Opinion	18.51	Is QnB a good bank?
Factual knowledge	31.21	When does Ramadan start in 2016?
Rhetorical	0.55	How it is to live in Qatar?
Invitation	0.83	need tennis partner
Others	27.07	razor racing car

Table 2: Breakdown of query types along with examples

during their interaction by clicking on the 'Give feedback on the new tool' button. The form also allowed them to provide free-form comments and suggestions. Finally, the questionnaire sought voluntary information about the age, gender, and Web experience of participants. Throughout the sessions, we logged interface actions along with their timestamps to better understand the usage patterns of the CQAVis tool.

Pilot study

Before making the beta version publicly available and running the online study, all study aspects, including instructions and setup, went through several iterations of a pilot study. We ran this pilot study in a lab-based setting with six participants, where we collected the data in the form of questionnaires, interviews, and observations.

The pilot study helped us in refining both the study procedure and the prototype. For example, the pilot study suggested that background questions should be asked at the end of the study instead of at the beginning, because participants wanted to immediately explore the system without requiring to fill up the questionnaire. We also modified the types of questions being asked (e.g., we provided less open-ended questions). The pilot study also led us to simplify the interface, by eliminating the encoding of less useful data to be shown, such as encoding the comment length using the width of rectangles. A more detailed description of the lessons learned from the pilot study is provided in the next section.

Participants

Our online study attracted 768 participants over a period of 18 weeks. The users were recruited through the beta version link of Qatar Living, as well as through publicizing in online social networks (i.e., Facebook and twitter) and mailing lists.

Those participants who chose to provide their background information held a variety of occupation, including students, expatriates working as engineers, architects and consultants, researchers and professors in universities etc. The majority of participants were young (85% of them were below 45). Among those who indicated their gender information, 65% were male participants. In general, most of the participants were quite familiar with using the Web, with 72% of them indicating that they visit the Web quite frequently (several times a day). However, when it comes to uses of online forums, the responses were mixed, ranging from rarely to very frequently.

Analysis of results

We now present both our quantitative and qualitative analysis, as well as the results based on the data collected from the user logs and questionnaires.

Sessions and queries

During the study, we captured quantitative data regarding 1,122 queries from 768 users. A summary of the queries and sessions is provided in Table 1. From the table, we can see that based on the medians, a typical participant spent 142 seconds with the system and issued just 1 query per visit. The average session lengths are considerably larger, as some participants engaged with the system for much longer time periods.

We categorized the questions asked by the participants by following [25] to understand the nature of information needs that were prevalent among our target users. The distribution of question types is shown in Table 2. Here, both *opinion* and *recommendation* questions are subjective in nature; *opinion* questions ask for a rating of a specific item whereas *recommendation* questions ask for open-ended requests and suggestions. In contrast, *factual* questions expect objective answers. *Rhetorical* questions are intended to promote discussions as opposed to eliciting specific answers. An *invitation* asks for attending an event. Finally, the *other* category consists of queries, that do not fall into any of the previous categories.

The distribution of questions was similar to what has been found in the existing literature [25], with subjective questions (i.e., *opinion* and *recommendation*) being strongly prevalent among participants (41%). This justifies the rationale for tailoring our interface to deal with subjective questions which may require the user to read many useful comments to get the answers from various perspectives.

Subjective ratings

After interacting with the interface the user could chose to provide feedback by clicking on the Feedback button. 56 users chose to provide feedback on the tool. In the feedback form, participants rated four different measures on a standard 5 point Likert scale. The results of these questionnaires are presented in Figure 6. From the Figure, we can readily see that the majority of the responses were dominated by positive ratings. In particular, most users agreed that the tool is useful and it enabled them to find answers relevant to their questions.

Preference

In the questionnaire, participants were also asked if they would prefer this tool over their regular forum search tool. 68.75% of participants indicated a preference for CQAVis, with only a

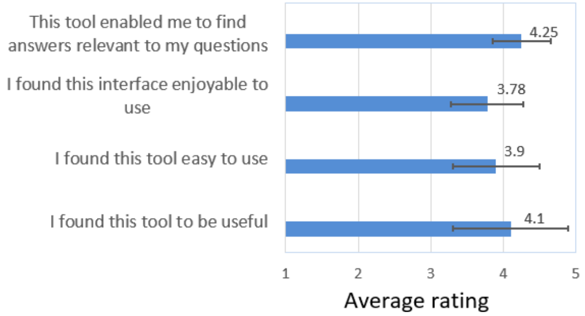

Figure 6: Average rating of interfaces by the participants on four different measures. Longer bars indicate higher rating.

small fraction of them (6.25%) choosing the regular one. 25% indicated that they were indifferent between the two interfaces.

Interaction patterns

In addition to questionnaires, we analyzed the log data to get insights into the interaction patterns of users Figure 7 shows the percentage of users who used each interactive features of the interface at least once. As expected, almost all the participants typed at least one query during the interaction. Similarly, most of them hovered and clicked on conversations in the question list view. When interacting with the conversation view, over 54% of the users hovered on the thread view and 39.4% clicked on rectangles in the thread view representing comments. This result is rather encouraging, because despite being completely new visualization features, they were used by a large portion of users. Finally, sorting and filtering comments were used by a smaller number of users (12% and 9% respectively). A possible explanation is that many participants did not notice these features, while interacting with the interface. Another reason could be that users were able to fulfill their information needs with other interactive features.

Qualitative Data Analysis

We analyzed the free-from text provided by 39 participants to gain insight into the users' experience with the interface. In order to make sense of these comments and suggestions, we carried out a bottom-up coding approach: first, read all the free-form texts to gain an overview of the participant's feedback; second, find common themes and topics and associate codes accordingly; and finally, categorize the themes into the main types of feedback. The resulting themes are described below:

General Feedback: From this analysis, we found that the feedback towards CQAVis was positive (68%), but there was also some negative (24%) and neutral (8%) feedback. More specifically, those who were positive towards the CQAVis interface expressed that the interface was simple and easy to use, which was an important design goal. According to participant P20, *"The design of this tool is very simple and easy to use. I am impressed with the tool's accessibility and how intuitive it was..."*. Also, some participants' perceived speed of task completion was enhanced by the interface as pointed out by P29 *"Quick and reliable"*.

A number of participants thought that the system was able to satisfy their information needs effectively. For instance, P13 mentioned that *"It gave me the answer I was looking for in a straightforward way, which is what you want from a*

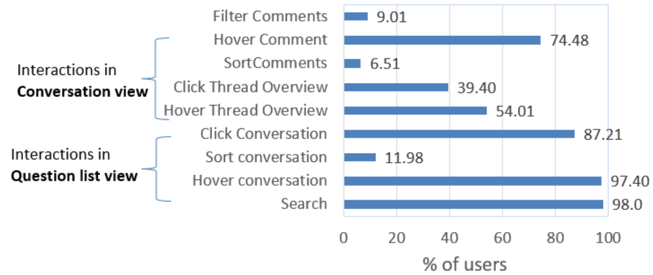

Figure 7: Interface features used by the participants.

search tool. No need to scroll through lots and lots of Google pages...". Similarly, P1 liked the idea of finding high quality comments from similar question, *"I like that you can get similar questions and their corresponding high quality answers immediately, without having to read all the comments"*. Some people also compared their positive search experience with the traditional Qatar Living search tool: *"Qatar Living is difficult to search but with this tool it gets much easier"* [P22].

Those who were critical about the interface mentioned that the text analytics techniques need to be more accurate '*"Need more accuracy for the result"* [P6]. Some people also questioned the reliability of the comments and suggested a way to filter out spam comments: *"It could be made better by filtering out spam comments. Some of the information has no actual basis..."* [P21]. Also, one participant suggested that for time sensitive questions, the system should consider the timestamp of answers for ranking, *"I asked: when does Ramadan start? But the top answer was actually posted few years ago"* [P18].

Reactions to interface features: We also found recurring comments on particular features of the interface. For instance, some participants liked the idea of having the question view and the conversation view side-by-side: *"It is nice to have the questions and answers load quickly side to side without having to open many tabs in the browser"* [P22].

Several people were impressed by the visual thread overview and the color coding to represent the usefulness of a comment. *"I liked the color coding idea of the comments in the tool. It is very useful"* P[24]. However, learning this feature requires sometimes for one participant *"At the beginning it was not clear what the colored squares are..."* [P8]. One possible explanation is that very few participants (2%) actually watched the video tutorial provided on the introduction page.

Suggestions for improvement: A few participants felt that the user interface needs some improvements in general. There were also few specific suggestions about the components, for instance, the size of the slider at the top needs to be increased, so that it can be easily noticed [P10] and the interface should show the textual label 'not useful' explicitly for the comments that fall into the least useful bin [P34].

DISCUSSION

Summary of findings

Based on our analysis of the results, we now revisit our research questions mentioned at the beginning of the previous section. The first question was what are the possible benefits and limitations of CQAVis in supporting information seeking tasks. From the feedback data, the majority of participants who filled up the questionnaires found the interface to be useful and felt that it enabled them to find the relevant answers to their questions. Also, the qualitative feedback from participants suggests that their overall impression was quite positive. With regards to the second research question, when the participants were asked to indicate a preference, the majority of them chose CQAVis over the traditional forum search tool.

However, recall that the questionnaire data was filled up only by a fraction of participants. While this prevents us from making strong claims from the questionnaire data alone, we complement the analysis with qualitative observation based on the free-form comments as well as from the interaction log data to get a deeper understanding about both positive and negative aspects of the interface features and their usage.

Note that the log data was analyzed over all the participants, thus arguably reflects overall usage patterns. In particular, the log data reveals that not all the interface features were equally used. While some of the new interface features such as the thread overview were used by a fair number of participants, still some participants did not use them. A possible explanation is that some participants might prefer the traditional way of scrolling through the comments of the thread, while still having a situational awareness by looking at the thread view.

Lessons learned

We now reflect upon on our design and evaluation of the CQAVis interface to summarize the lessons learned that can arguably be generalizable to other conversational domains.

Design

Most target users in our domain did not have enough familiarity with complex interactive visualization. To support such users, we have focused on following design principles which can be applicable to other domains where users have similar expertise level.

Less is more: In our early prototypes, we considered some advanced features, such as visually encoding additional data (e.g., comment length) and more complex interactions (e.g., navigate through the related-question-graph) with an aim to better support users. However, the feedback from users throughout the pilot studies led us to simplify the interface iteratively, eliminating such kinds of interactive visualization features.

Based on our experience, we suggest that when designing for similar populations in the domain of conversations, the designer should simplify interface features iteratively to retain features that are not only useful but also simple and intuitive.

Enhance learnability: We found that in our study users do not tend to spend time reading the instructions or watch the video tutorial to learn the new interface. Therefore, the interface should enhance the learnability by providing self-explanatory components by adding more textual labels and tooltips.

Introduce familiar visualizations: During the prototyping stage, we realized that novice users in the Qatar Living forums often find it difficult to understand complex visualizations. Therefore, the interface should use the visualization components that are easily understood by most people.

Evaluation

While we argue that the web-based online study enhanced the ecological validity by evaluating with real forum readers performing their real tasks, it also posed several challenges. For instance, it was difficult to collect sufficient amount of quantitative and qualitative feedback from a large number of participants. While it is common to collect users' background and demographic information in the form of a pre-study questionnaire, in a pilot study we found that participants were reluctant to fill-up the questionnaire. Therefore, the questionnaire was included in the feedback form that the user could fill out, after they had interacted with the interface.

Even then the challenge was how to get feedback from a large amount of participants who have interacted with the interface. While a button for providing feedback was available at the top of the interface, some participants did not even notice it. To further enhance the likelihood of obtaining some feedback, we introduced a pop-up screen that would appear reminding the user to submit the feedback when they move their cursor at the top of the screen. To provide further incentives to the user, the message mentioned that the participant will be entered into a lottery of winning 50 QAR gift cards. While all of the above techniques helped us to receive more feedback, we call for more research on how to get a rich amount of feedback from a large amount of participants in a Web-based study.

CONCLUSIONS AND FUTURE WORK

We have presented CQAVis, an interactive system that supports users to find good answers to a newly-posed question, by combining a novel set of NLP and InfoVis features, informed by an understanding of the user requirements in the domain of CQA. The underlying NLP techniques automatically retrieve and rank a set of comments with respect to the new question, (*i*) by selecting a set of question threads that are relevant to the user question, (*i*) by assigning a goodness score to the comments within these threads, and (*iii*) by measuring the similarity between the new question and the thread questions. The visual interface combines two coordinated views (i.e., question view and conversation view) that help users in rapidly navigating through the useful comments, even if they are scattered around multiple different threads.

Our large-scale Web study underlines the potential for tightly integrating NLP and InfoVis, offering the users a new way of information seeking in CQA forums. It also reveals important lessons for designing and studying such systems for real users with varying levels of expertise, which can be generalizable for the design and evaluation of other conversational domains.

In the future, we plan to further improve our comment classifier. We would also like to apply our visual interface for supporting information seeking tasks using other forum datasets.

Acknowledgments

We thank Alberto Barrón-Cedeño, Giovanni Da San Martino, Alessandro Moschitti, Preslav Nakov, Salvatore Romeo, Scott Cyphers and Mitra Mohtarami for their help in the NLP system. Many thanks to Qatar Living, and especially Ash Bashir, for their help in deploying the demo in their website.

REFERENCES

1. 2016. Colorbrewer. (2016). http://colorbrewer2.org/.
2. Alberto Barrón-Cedeño, Simone Filice, Giovanni Da San Martino, Shafiq Joty, Lluís Màrquez, Preslav Nakov, and Alessandro Moschitti. 2015. Thread-Level Information for Comment Classification in Community Question Answering. In *Proc. ACL-IJCNLP*.
3. Sheelagh Carpendale. 2008. Evaluating information visualizations. In *Information Visualization: Human-Centered Issues and Perspectives*. Springer, 19–45.
4. Andy Cockburn, Amy Karlson, and Benjamin B Bederson. 2008. A review of overview+ detail, zooming, and focus+ context interfaces. *ACM Computing Surveys (CSUR)* 41, 1 (2008), 2.
5. Michael Collins and Nigel Duffy. 2002. Convolution Kernels for Natural Language. In *Advances in NIPS*, T. G. Dietterich, S. Becker, and Z. Ghahramani (Eds.). MIT Press, 625–632.
6. Kushal Dave, Martin Wattenberg, and Michael Muller. 2004. Flash forums and forumReader: navigating a new kind of large-scale online discussion. In *Proc. ACM Conf. on CSCW*. 232–241.
7. Park Deokgun, Sachar Simranjit, Diakopoulos Nicholas, , and Elmqvist Niklas. 2016. Supporting Comment Moderators in Identifying High Quality Online News Comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 769–778.
8. Cicero dos Santos, Luciano Barbosa, Dasha Bogdanova, and Bianca Zadrozny. 2015. Learning Hybrid Representations to Retrieve Semantically Equivalent Questions. In *Proc. 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. 694–699.
9. Nicole B Ellison, Rebecca Gray, Jessica Vitak, Cliff Lampe, and Andrew T Fiore. 2013. Calling All Facebook Friends: Exploring Requests for Help on Facebook.. In *Proceedings of the ICWSM*.
10. Minwei Feng, Bing Xiang, Michael R Glass, Lidan Wang, and Bowen Zhou. 2015. Applying deep learning to answer selection: A study and an open task. *arXiv preprint arXiv:1508.01585* (2015).
11. F. Maxwell Harper, Daphne Raban, Sheizaf Rafaeli, and Joseph A. Konstan. 2008. Predictors of Answer Quality in Online Q&Amp;A Sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. 865–874.
12. Marti A. Hearst. 2011. Natural Search User Interfaces. *Commun. ACM* 54, 11 (Nov. 2011), 60–67.
13. Enamul Hoque and Giuseppe Carenini. 2014. ConVis: A Visual Text Analytic System for Exploring Blog Conversations. *Computer Graphics Forum (Proc. EuroVis)* 33, 3 (2014), 221–230.
14. Enamul Hoque and Giuseppe Carenini. 2015. ConVisIT: Interactive Topic Modeling for Exploring Asynchronous Online Conversations. In *Proc. ACM conf. on Intelligent User Interfaces*.
15. Enamul Hoque and Giuseppe Carenini. 2016a. Interactive Topic Modeling for Exploring Asynchronous Online Conversations: Design and Evaluation of ConVisIT. *ACM Transactions on Interactive Intelligent Systems* 6, 1 (Feb. 2016), 7:1–7:24.
16. Enamul Hoque and Giuseppe Carenini. 2016b. MultiConVis: A Visual Text Analytics System for Exploring a Collection of Online Conversations. In *Proc. ACM IUI*. 96–107.
17. Enamul Hoque, Shafiq Joty, Lluís Màrquez, Alberto Barrón-Cedeño, Giovanni Da San Martino, Alessandro Moschitti, Preslav Nakov, Salvatore Romeo, and Giuseppe Carenini. 2016. An Interactive System for Exploring Community Question Answering Forums. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*. Osaka, Japan, 1–5. http://aclweb.org/anthology/C16-2001
18. Damon Horowitz and Sepandar D Kamvar. 2010. The anatomy of a large-scale social search engine. In *Proceedings of the 19th international conference on World wide web*. ACM, 431–440.
19. Gary Hsieh and Scott Counts. 2009. mimir: A market-based real-time question and answer service. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 769–778.
20. Jiwoon Jeon, W. Bruce Croft, and Joon Ho Lee. 2005. Finding Similar Questions in Large Question and Answer Archives. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management (CIKM '05)*. 84–90.
21. Bernard Kerr. 2003. Thread arcs: An email thread visualization. In *IEEE Symposium on Information Visualization*. 211–218.
22. H. Lam. 2008. A Framework of Interaction Costs in Information Visualization. *IEEE Trans. Visualization & Comp. Graphics* 14, 6 (2008), 1149–1156.
23. H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. 2012. Empirical Studies in Information Visualization: Seven Scenarios. *IEEE Trans. Visualization & Comp. Graphics* 18, 9 (2012), 1520–1536.
24. Qiaoling Liu, Tomasz Jurczyk, Jinho Choi, and Eugene Agichtein. 2015. Real-Time Community Question Answering: Exploring Content Recommendation and User Notification Strategies. In *Proceedings of the 20th International Conference on Intelligent User Interfaces (IUI '15)*. 50–61.

25. Meredith Ringel Morris, Jaime Teevan, and Katrina Panovich. 2010. What Do People Ask Their Social Networks, and Why?: A Survey Study of Status Message Q&A Behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. 1739–1748.

26. Alessandro Moschitti. 2006. Efficient Convolution Kernels for Dependency and Constituent Syntactic Trees. In *Machine Learning: ECML 2006*, Johannes Fürnkranz, Tobias Scheffer, and Myra Spiliopoulou (Eds.). Lecture Notes in Computer Science, Vol. 4212. 318–329.

27. Tamara Munzner. 2014. *Visualization Analysis and Design*. CRC Press.

28. Preslav Nakov, Lluís Màrquez, Walid Magdy, Alessandro Moschitti, Jim Glass, and Bilal Randeree. 2015. SemEval-2015 Task 3: Answer Selection in Community Question Answering. In *Proc. Workshop SemEval*. 269–281.

29. Preslav Nakov, Lluís Màrquez, Alessandro Moschitti, Walid Magdy, Hamdy Mubarak, Abed Alhakim Freihat, Jim Glass, and Bilal Randeree. 2016. SemEval-2016 Task 3: Community Question Answering. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval '16)*. Association for Computational Linguistics, San Diego, California.

30. Massimo Nicosia, Simone Filice, Alberto Barrón-Cedeño, Iman Saleh, Hamdy Mubarak, Wei Gao, Preslav Nakov, Giovanni Da San Martino, Alessandro Moschitti, Kareem Darwish, Lluís Màrquez, Shafiq Joty, and Walid Magdy. 2015. QCRI: Answer Selection for Community Question Answering - Experiments for Arabic and English. In *Proc. Workshop SemEval*.

31. Vıctor Pascual-Cid and Andreas Kaltenbrunner. 2009. Exploring asynchronous online discussions through hierarchical visualisation. In *IEEE Conf. on Information Visualization*. 191–196.

32. Xipeng Qiu and Xuanjing Huang. 2015. Convolutional Neural Tensor Network Architecture for Community-based Question Answering. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI 2015)*.

33. D Raban and F Harper. 2008. Motivations for answering questions online. *New media and innovative technologies* 73 (2008).

34. Warren Sack. 2000. Conversation map: an interface for very-large-scale conversations. *Journal of Management Information Systems* 17, 3 (2000), 73–92.

35. Michael Sedlmair, Miriah Meyer, and Tamara Munzner. 2012. Design study methodology: reflections from the trenches and the stacks. *IEEE Trans. Visualization & Comp. Graphics* 18, 12 (2012), 2431–2440.

36. Aliaksei Severyn and Alessandro Moschitti. 2012. Structural Relationships for Large-scale Learning of Answer Re-ranking. In *Proc. SIGIR*. 741–750.

37. Aliaksei Severyn and Alessandro Moschitti. 2015. Learning to rank short text pairs with convolutional deep neural networks. In *Proc. SIGIR*. 373–382.

38. Yikang Shen, Wenge Rong, Zhiwei Sun, Yuanxin Ouyang, and Zhang Xiong. 2015. Question/Answer Matching for CQA System via Combining Lexical and Sequential Information.. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI 2015)*. 275–281.

39. Ming Tan, Bing Xiang, and Bowen Zhou. 2015. LSTM-based Deep Learning Models for non-factoid answer selection. *arXiv preprint arXiv:1511.04108* (2015).

40. Gina Danielle Venolia and Carman Neustaedter. 2003. Understanding sequence and reply relationships within email conversations: a mixed-model visualization. In *Proc. Conf. CHI*. 361–368.

41. Fernanda B Viégas, Scott Golder, and Judith Donath. 2006. Visualizing email content: portraying relationships from conversational histories. In *Proc. Conf. CHI*. 979–988.

42. Di Wang and Eric Nyberg. 2015. A Long Short-Term Memory Model for Answer Sentence Selection in Question Answering. In *ACL*. 707–712.

43. Martin Wattenberg and David Millen. 2003. Conversation thumbnails for large-scale discussions. In *extended abstracts on CHI*. 742–743.

44. Furu Wei, Shixia Liu, Yangqiu Song, Shimei Pan, Michelle X Zhou, Weihong Qian, Lei Shi, Li Tan, and Qiang Zhang. 2010. Tiara: a visual exploratory text analytic system. In *Proc. ACM Conf. on Knowledge Discovery and Data Mining*. 153–162.

45. Jian Zhao, C. Collins, F. Chevalier, and R. Balakrishnan. 2013. Interactive Exploration of Implicit and Explicit Relations in Faceted Datasets. *IEEE Trans. Visualization & Comp. Graphics* 19, 12 (2013), 2080–2089.

46. Guangyou Zhou, Tingting He, Jun Zhao, and Po Hu. 2015a. Learning Continuous Word Embedding with Metadata for Question Retrieval in Community Question Answering. In *Proc. 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 250–259.

47. Xiaoqiang Zhou, Baotian Hu, Qingcai Chen, Buzhou Tang, and Xiaolong Wang. 2015b. Answer Sequence Learning with Neural Networks for Answer Selection in Community Question Answering. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. 713–718.