Enhancing Technical Q&A Forums with CiteHistory

Adam Fourney[1,2] and Meredith Ringel Morris[2]

[1]University of Waterloo, Waterloo, ON, Canada, [2]Microsoft Research, Redmond, WA, USA

afourney@cs.uwaterloo.ca, merrie@microsoft.com

Abstract

Software engineers often use Q&A forums like Stack Overflow and MSDN to ask and answer technical questions. Through a survey study and web browser log analysis, we find that both askers and answerers of technical forum questions typically conduct extensive online research before composing their posts. The inclusion of links to these research materials is beneficial to the forum participants, though post authors do not always include such citations. Based on these findings, we developed CiteHistory, a browser plugin that simplifies the process of including relevant search queries and URLs as bibliographic supplements to forum posts, and supports information re-finding for post authors. We discuss the results of a two-week deployment of CiteHistory with professional software engineers, which demonstrated that CiteHistory increased reference inclusion in posts, and offered auxiliary benefits as a personal research tracker.

Introduction

Online forums, blogs, wikis, and other community-curated content embody a vast amount of technical knowledge and expertise. These resources have become invaluable tools to the modern knowledge worker. A recent trend in this trajectory is the emergence of technical question/answer (Q&A) sites, where participants can pose questions to a responsive audience of highly skilled peers. Such sites exist for a range of disciplines, and are particularly important in the software development community (Mamykina et al., 2011; Treude, Barzilay, and Storey, 2011). In this field, forums like Stack Overflow and the MSDN (Microsoft Developer Network) Forums are two of the most prominent, hosting over 3.6 million (Stack Exchange Inc., 2012) and 7.8 million questions (Microsoft, 2012) respectively.

In this paper we investigate, for the first time, the online web browsing contexts in which software development questions are asked and answered. In this vein, and forming the first major contribution of this paper, we present two large studies investigating the current use of online resources on two technical Q&A sites. The first is a survey study of 122 software professionals who either asked or answered a recent question on MSDN. The second is an analysis of the web browsing logs of 210 users who asked or answered questions on Stack Exchange Q&A forums (primarily Stack Overflow). From these two studies we find that conducting extensive online research before asking or answering a question is a common practice. Likewise, we find that answers receiving positive community feedback are more likely to include links to online references compared to their less successful counterparts. However, we find that such forum posts are in the minority: most forum posts fail to reference any of the consulted external materials, and authors often fail to provide any indication of information provenance.

Based on these findings, we developed CiteHistory – the second major contribution of this work. CiteHistory is a tool that facilitates sharing of research activities by Q&A forum participants. CiteHistory is intended to: (1) encourage participants to cite relevant source material, and (2) to simplify the work practices of valuable contributors who are already in the habit of including such citations. We report the findings of an evaluation of CiteHistory with professional software engineers, which indicate that CiteHistory increased reference inclusion in posts and offered auxiliary benefits as a personal research tracker.

The remainder of this paper is structured as follows: We first discuss related work, and then present the results of our survey and log studies. Next, we describe our CiteHistory tool, and then present its evaluation. Finally, we conclude with a discussion and directions for future work.

Related Work

Software Development and Online Resource Use

Software developers make extensive use of online resources in their day-to-day activities. For example, when Brandt et al. (2009) tasked experienced student-

programmers with developing and prototyping a web chat application, the researchers noted that participants spent 19% of their time conducting online research. Likewise, Ko, Aung, and Myers (2005) explored how Java developers allocate their time while maintaining code, and found that developers spent 11% of their time reading online API documentation. These findings are further supported by a small-scale study of web use and programming conducted by Goldman and Miller (2009), who reported that 23% of website visits temporally proximal to editing code were development related.

When faced with a difficult programming task, Q&A forums hosted by Stack Exchange are common destinations among software developers. Within only a few years, these forum websites have captured significant mindshare among developers, are a key resource used for resolving programming problems, and are often relied on as a substitute for official technical documentation for many technologies (Treude, Barzilay, and Store, 2011). Moreover, many forum participants consider their Stack Overflow achievements to be a valuable component of their professional resumes (Mamykina et al., 2011).

Given the propensity for technical details and sample code to appear in technical forum questions and answers, and the likelihood that experts frequent these forums, it seems plausible that technical Q&A participants review online material while authoring posts. Our work confirms this hypothesis, and contributes data about how software developers compose their posts on such sites.

Reference and Citation Use in General Q&A Sites

In contrast to technical Q&A forums, general community Q&A sites, such as Yahoo! Answers [http://answers.yahoo.com] are characterized by their broad appeal and support for a large number of topic areas. Such sites are significantly more active than their specialized technical counterparts, and have attracted the attention of many researchers. Gazan (2006) investigated the role of references in the Answerbag community, classifying answers as having been either written by a *synthesist* or by a *specialist*. Synthesist answers are those in which the author "make(s) explicit reference to other sources of information to support their answers." Conversely, specialist answers are those written by authors who "have proclaimed their expertise in a community, and who answer a given question without referring to any other source." Gazan reported that, across all Answerbag categories, synthesist answers tended to be rated more highly than specialist answers by Answerbag's users. Moreover, synthesist answers were especially highly regarded in technical categories such as "Drugs and Medicine" and "Science."

Harper et al. (2008) examined answer quality on general community Q&A sites, and found that the number of links included in answers was positively correlated with answer quality. The authors suggest that such links are indications of synthesist answers. Conversely, in a separate study investigating Yahoo! Answers, Shah and Pomerantz (2010) reported that the inclusion of references in posted answers was a poor predictor of quality. This discrepancy suggests the need for continued research.

The particular links included in a forum post may also implicitly reveal the author's topical expertise level, since domain experts visit different URLs and issue different search queries than novices (Bhavnani, 2002). Making experts' information-seeking strategies more transparent through revealing queries issued or URLs visited may be of benefit to more novice users (Moraveji et al., 2011). Q&A forum exchanges do not merely provide information to the original asker of a question, but are reused by subsequent searchers with similar information needs (Liu et al., 2011). This suggests that links included in forum answers serve another important role: they associate relevant reference material to user questions in cases where traditional search may fail, making such information more discoverable by novice users.

Another benefit to the inclusion of links to reference material is that such information can help readers assess an answer's credibility (Kapoun, 1998). Indeed, the lack of provenance information for much online content is a topic of sufficient concern that the W3C (World Wide Web Consortium) has established a working group on that topic (W3C Provenance Working Group, 2012). This paper adds to the body of knowledge regarding the current state of provenance information in technical forums, and introduces a system that encourages the practice of citation inclusion (i.e., synthesist answers) by simplifying and automating users' abilities to leverage their search and browsing history when composing forum posts.

Interacting with Search & Browser History

All mainstream web browsers allow users to view their complete browsing history as well as specially marked "bookmarks" or "favorite" URLs, though such functionality is rarely utilized in practice, with search engines being a more common way for users to re-find previously encountered information (Aula, Jhaveri, and Käki, 2005). Several research projects have sought to enhance these basic history and bookmarking capabilities. For example, Hunter Gatherer (Schraefel et al., 2002) helps users organize and make sense of content collected during web browsing sessions, and SearchBar helps users resume interrupted searching/browsing tasks (Morris, Morris, and Venolia, 2008). ActionShot (Li et al., 2010) allows users to record and share complex web interactions with friends (including

sub-page-level actions like clicking buttons or filling out forms).

Specific to our target demographic of software engineers, several research projects (Brandt et al., 2010; Goldman and Miller, 2009; Hartmann et al., 2011) have sought to directly support programmers' tendencies to refer to online material while writing code. For example, HyperSource (Hartmann et al., 2011) helps developers document and track the origin of code they copy from websites. Codetrail (Goldman and Miller, 2009) achieves a similar effect by comparing a developer's recently written code to pages in their web browsing history. When similarities are identified, an association is made between the code and the online document. Blueprint (Brandt et al., 2010) is a system that integrates web search into a development environment, and facilitates the process of adapting example code pulled from the web. Our system, CiteHistory, also enhances a user's ability to leverage their search and browser history, albeit in service of the specific task of adding citations to Q&A forum posts.

Survey Study of MSDN Forum Users

To better understand current practices, we conducted a survey study investigating the online resource utilization of technical Q&A forum participants. We identified MSDN forum participants who had stated in their profiles that they were software developers at an anonymized technology company with which we collaborated. We then identified 63 questions and 222 answers posted by these employees within the most recent two-month timeframe (May & June, 2012). The high number of answers as compared to questions likely reflects the relatively high expertise of this company's employees, and their propensity toward answering questions about their own company's software.

We sent each developer an email inviting him or her to participate in the survey. The email included a link to a question the participant had asked or answered in the two-month timeframe. Participants were instructed to consider only the referenced question when completing the survey. This design was intended to help participants avoid generalization and mitigate the inaccuracies that can be associated with retroactive self-report studies. A gift card drawing was offered as an incentive for survey completion.

107 of the 222 answerers (92% male) completed their surveys, as did 22 of the 63 askers (95% male) (such skewed gender ratios are typical among professional software engineers). 88% of answerers reported working in the field of software development for at least 5 years (63% reported having more than 10 years' experience). Askers were slightly less experienced, with 59% reporting working in the field for at least 5 years.

Results: Askers

21 of 22 askers (95%) reported using online resources to attempt to answer their own questions prior to submitting their posts to the MSDN forums. A variety of online resources were consulted: 55% visited official documentation, 59% consulted other technical Q&A forums, 73% visited blogs or other websites with journalistic-style content, and 91% consulted search engines.

Of the askers who reported using search engines, 85% considered themselves to be above-average searchers, with 50% reporting themselves as search experts. Only 25% of respondents reported difficulty in devising good search terms. This suggests that many participants would be able to accurately identify candidate answers, if they were to appear in search result pages. Despite this search expertise, when asked how long respondents searched prior to posting their questions, 70% reported spending *at least* 30 minutes on the task.

16 of the 22 respondents (73%) reported receiving helpful answers to their question. In 12 of these 16 cases (75%), the askers reported that helpful responses contained hyperlinks to related material. In 8 of these 12 cases (67%), respondents indicated that the hyperlinks alone were sufficient to answer their question. This suggests that an important role served by Q&A sites is the dissemination of links and resources relevant to user questions, especially in cases where traditional search fails.

Results: Answerers

56 of 107 answerers (52%) reported using online resources to help them answer the specified question. Again, a variety of online resources were used. Of those reporting using online resources, 73% visited official documentation, 18% consulted other technical Q&A forums, 21% visited blogs or other websites with journalistic-style content, and 82% consulted search engines. In cases where search was used, 70% of respondents reported searching for pages they already had in mind, while 30% reported searching for pages they had never visited before. This suggests that navigational queries (Broder, 2002) and/or instances of re-finding are especially likely when participants are answering questions.

Finally, while 52% of answerers reported conducting online research (as noted above), only 39% of all answers included hyperlinks or citations to relevant materials. This suggests that answerers fail to include potentially helpful references in a substantial portion their posts.

Log Study of Stack Exchange Q&A Sessions

The asker and answerer surveys strongly suggest that Q&A participants make extensive use of online resources – especially search – when asking and answering questions. One

limitation of these surveys is that they relied on self-reported metrics, which may be inaccurate. Thus, to complement our survey findings and paint a more complete picture of technical Q&A forum behavior, we also conducted an empirical log study of Stack Exchange question and answer sessions.

We analyzed 6 months of anonymized web browsing data collected between February and July of 2012 by a popular consumer browser plugin. While anonymous, these data include a unique identifier for each user, allowing the browsing streams to be partitioned on a user-by-user basis. Within this dataset, we identified browsing sessions in which it appeared that a user had asked or answered a question on any of Stack Exchange's websites. These events can be detected through distinctive sequences of page visits. For both questions and answers, these sequences begin with a URL signaling a user's intent to submit a question or answer, followed by a redirection to the resultant post. Upon identifying candidate post events, we then compared the question or answer creation times, as listed on the Stack Exchange website, with the timestamps found in our logs. This step ensured that the posts were indeed accepted by the website. Performing a similar analysis on MSDN was not possible, due to that site's use of dynamic webpages (e.g., AJAX), which were not completely captured in the log files.

In the log data, we identified 918 questions which were asked and 120 answers which were posted. For each question and answer, we extracted a log excerpt ending at the moment the question or answer was posted, and extending backwards in time up to two hours. Such excerpts capture the website visits and searches that occurred prior to the Q&A transaction. For the 120 answers, the resultant dataset included 4,917 log entries, spanning 96 hours. To provide a comparable dataset of questions, we randomly sampled 90 questions from the 918 identified earlier. The resultant question dataset included 4,968 log entries, spanning 88 hours.

Figure 1 depicts a typical answer session. In this case, the individual answers a question relating to the exclusive use of the Windows operating system on a Macintosh computer. The entire sequence of events took 73 minutes to complete. Despite the extensive research conducted by the user, their answer contained no hyperlinks or reference material. We also note that this user discovered the Stack Exchange question after first issuing a search query. This early query demonstrates that the answerer had a preexisting information need.

Having extracted log excerpts surrounding each post, we manually labeled each log entry (either a URL visit or a search engine query) as either relevant or not relevant to the associated questions or answers. After independently coding 14 log excerpts, this paper's two authors disagreed on only 4 of 412 individual log entries. The resultant inter-

18:51:47 – Search: "run windows 7 only on mac"
18:51:49 – Stack Exchange question visited for the first time ("…can you remove OS X and run only Windows on new Mac Intel based hardware?")
18:53:17 – Search: "run windows 7 only on mac youtube"
18:53:27 – Page: "Mac Pro - Running Windows 7 Beta"
18:53:40 – Search: "use windows only on mac"
18:54:15 – Search: "uninstall mac os and keep window"
18:54:28 – Page: "how to uninstall mac os x windows 7"

… 4 searches and 5 page visits later …

19:58:11 – Page: "3 Ways to Run Windows on Your Mac"
20:05:38 – Stack Exchange answer submitted

Figure 1. An example log excerpt in which a user opportunistically answers a Q&A forum post. Bold entries represent interactions with the forum. Despite conducting 8 searches and visiting 8 topically relevant URLs while composing an answer, no links to external materials or discussion of information provenance were included in the answering post.

rater reliability of 0.90 (via. Cohen's **κ**) was deemed sufficiently high to allow one researcher to continue labeling the remaining 196 excerpts.

Results

Askers Conduct Research Prior to Posting
As expected, the browsing logs of people asking questions on the Stack Exchange Q&A forums show that most users do extensive research prior to posting a question. Specifically, log excerpts from 77 of the 90 questions (86%) show some evidence of research before posting. These research sessions last an average of 44 minutes (median: 39 minutes). Within these 77 sessions, an average of 9 relevant URLs are visited (median: 6). However, these relevant URLs are not typically cited in the forum posts: only 13 of the 90 questions included URLs (14%). As a result, the average number of URLs contained in questions is just 0.27 (median: 0).

Answerers Conduct Research Prior to Posting
The log data show that answerers also make extensive use of online resources when authoring answers to questions. Log excerpts from 67 of the 120 questions (56%) showed evidence of research prior to posting an answer.

Within the 67 answer sessions in which research was conducted, an average of 4 relevant URLs were visited (median: 3), over an average timespan of 20 minutes (median: 13 minutes). Again, many of these URLs are not cited in the forum answers: only 29 of 120 answers contained URLs (24%). This results in an average of 0.47 relevant URLs per post (median: 0), rising to 1.9 (median: 2) when we consider only posts containing at least one URL; this is lower than we might expect given the number of pages visited by the authors.

Interestingly, for 36 of the 67 researched questions (54%), it appears that the answerer was already researching the question's topic prior to discovering the Stack Exchange question thread. The excerpt in Figure 1 is an example of this behavior. We refer to these events as *opportunistic answers*, since the answerer's log entries suggest a pre-existing information need. In the remaining 31 cases, the answerers began research only after discovering the Stack Exchange question by other means (e.g., browsing the site for unanswered questions). We refer to these events as *deliberate answers*. As will later be shown, deliberate answers tend to be better received by the community than opportunistic answers.

Answers with Positive Feedback Contain More Links
On Stack Exchange websites, answers can be "up voted" by the community, and can be marked as "accepted." In either case, we consider the answer as having received positive feedback. Consistent with previous research (Gazan, 2006; Harper et al., 2008), we found that answers receiving positive feedback contain more hyperlinks on average (0.64 links) than answers not receiving positive feedback (0.30 links). This difference was found to be statistically significant by an independent samples t-test with $t(118) = 2.0$ and $p = 0.04$.

These results can be extended and independently verified by accessing forum datasets made publicly available by Stack Exchange [http://data.stackexchange.com]. For example, we sampled 100,002 Stack Overflow answers at random and found that answers receiving positive feedback contain more links (mean of 0.72 links) than answers receiving no positive feedback (mean of 0.55 links). Again an independent samples t-test achieves statistical significance with $t(10^5) = 33.1$ and $p < 0.001$. Overall, only 34% of sampled Stack Overflow answers contained at least one link.

Deliberate Answers Fare Better than Opportunistic Answers
Independent samples t-tests also reveal that deliberate answerers spend less time conducting research on average than opportunistic answerers (14 minutes vs. 26 minutes, $t(65) = 2.27, p < .03$), yet their answers tend to score better (mean of 1.1 "up votes" vs. 0.47 "up votes", $t(65) = 2.50, p < .02$). Moreover, deliberate answers tended to include more links (mean of 1.00 links vs. 0.42 links, $t(65) = 2.08, p = 0.04$).

Discussion: Survey and Log Study Findings

The results of our survey and log studies indicate that both answerers and askers conduct extensive research and make heavy use of online resources before posting on technical Q&A forums. However, the resultant posts often fail to include citations or links to relevant – and potentially helpful

– material. In the case of askers, link inclusion may benefit answerers by listing resources that have already been investigated, and /or by providing an impression of the askers' level of expertise (Bhavnani, 2002; White, Dumais, and Teevan, 2009). For answerers, the dearth of links in posts is especially problematic: Respondents to our askers survey indicated that, when included in answers, links were often sufficient for addressing the askers' questions. Likewise, the log study revealed that answers containing links tended to be better received by the community.

In interpreting the relationship between links in answers and positive community feedback, it is important to note that there are several factors which, alone or in combination, may explain the observed outcomes. For example, it is possible that including relevant citations in forum answers improves the quality and utility of those posts. It is also possible that the authors of high-quality answers possess, as a common trait, the habit of including citations in their posts. In either case, these results motivate the need for a tool that facilitates the sharing of research in Q&A forum posts. Such a tool would encourage users to cite related material, and would optimize the workflow of the high-value forum members who are already including such references.

In the following sections, we describe the design and evaluation of CiteHistory, a tool we constructed to enhance Q&A forums through simplified research sharing.

CiteHistory: A Tool for Reference Sharing

CiteHistory is a web application paired with a browser plugin (for Internet Explorer, the primary browser used within the organization in which CiteHistory was evaluated). These components work together to meet our design and research objectives.

When installed, the CiteHistory plugin locally records a one-hour rolling log of the user's browsing history (based on our survey and log study findings that 30 – 60 minutes of research is typical). When the user visits the MSDN Forums or any of Stack Exchange's 88 websites, the plugin activates and performs two important actions:

First, the CiteHistory plugin alters the HTML of the Q&A forum so as to insert a new "Add CiteHistory" button into the page (Figure 2). Clicking this button displays a dialog that enables individuals to easily select and rate pages from their recent search and browsing history for inclusion in their posts (Figure 3).

Second, the plugin transmits a representation of the one-hour browsing log for our analysis. For privacy, transmitted log entries are irreversibly obscured using a 256-bit salt and cryptographic hashing with the SHA-256 message digest algorithm, thereby allowing only coarse analysis (e.g., detecting domain or page re-visitations, counting search

query terms, etc.). However, entries corresponding to the pages rated relevant by the user (in the first step) escape obfuscation, and are recorded in a readable format. In essence, the user's selections provide explicit permission for sharing and explicit relevance judgments for the pages in the user's browsing history.

In the next sections, we describe in detail the features of both the CiteHistory plugin and website.

CiteHistory Browser Plugin Features

Users interact with CiteHistory primarily through the "Add CiteHistory" button that is injected into Stack Exchange and MSDN Forum post submission pages. A typical CiteHistory session progresses as follows:

1. The user conducts online research as normal.

2. Prior to submitting a forum post, the user clicks the "Add CiteHistory" button (Figure 2), and the history selection dialog appears (Figure 3). Clicking this button, and interacting with CiteHistory, is voluntary.

3. The user is presented with a timeline representing their recent browsing history (Figure 3). Within this timeline, a smaller time range is highlighted. The highlighted time range marks the start and end of research conducted in service of writing the forum post. When selecting pages for citation (step 4), users can only interact with the entries included in this range. If the user is posting an answer, the time range is initially set to begin at the moment the question was visited for the first time. If the user is posting a question, the time range is initially set to begin 30 minutes before initiating the question-asking process. These choices of defaults are based on the time ranges observed in our earlier log study. Users can adjust the time range by dragging the selection area in the interface (Figure 3).

4. After adjusting the time range corresponding to research, authors then select entries to publicly include in their posts. This can be accomplished by clicking on the star icon next to each page visit or search query (Figure 3). By default, the most recent web search and the most recent page visit receive a star rating, since prior work suggests that the endpoints of "search trails" are likely to contain the sought-after information (Bilenko & White, 2008). Items in the research time range, but not receiving the star rating, are included in a private bibliographic record. This private record is visible only to the post's author.

5. Optionally, users can redact private history items. This is achieved by clicking an "X" icon next to each entry (Figure 3). When redacted, a black bar replaces the entry. Redacted items are never transmitted, and do not appear in any bibliographic record.
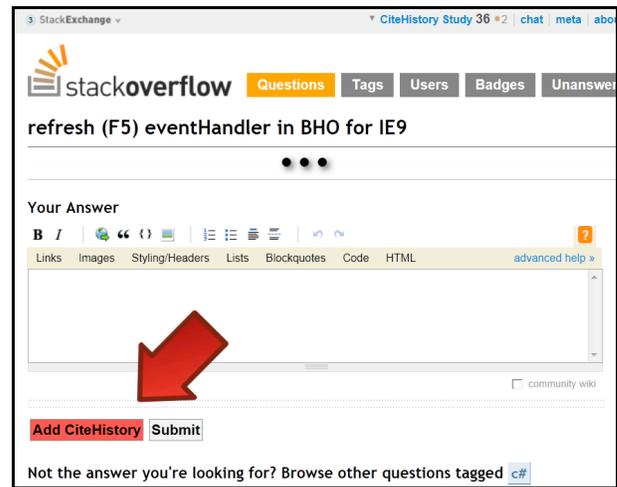

Figure 2. A Stack Overflow page augmented with the "Add CiteHistory" button (arrow superimposed).
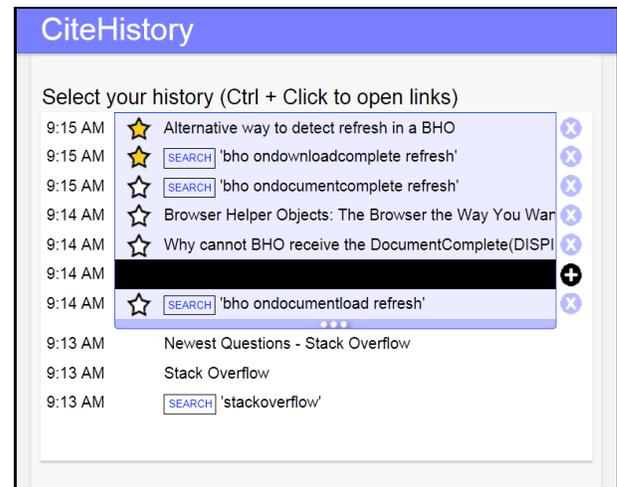

Figure 3. CiteHistory's history selection dialog. From this interface, users can adjust the time range representing their research, select (star) items for inclusion in the public bibliography, and redact any sensitive items (the black bar corresponds to a redacted item).

6. Finally, the user submits their selection. The user is then presented with a choice of a compact or long-form bibliographic record they can copy into the forum post (Figure 4). This bibliographic record is automatically formatted in a manner that is appropriate for the forum in which the post is submitted.

CiteHistory Website Features

While users primarily interact with CiteHistory as described above, the CiteHistory website provides some additional functionality. In particular, the CiteHistory website aggregates all public bibliographies posted by CiteHistory's users (Figure 5). These public records link to their as-

sociated forum posts, and list additional aggregate statistics such as the total time spent, searches performed and pages visited while the author was researching the post.

Additionally, the CiteHistory website provides a "My-Cites" page where users can retrieve their private bibliographies (Figure 6). As noted earlier, private bibliographies include all pages indicated by the author as being part of the research period, including those receiving star ratings and those not receiving stars.

Evaluating CiteHistory

In order to encourage and facilitate the sharing of research resources among Q&A forum participants, CiteHistory makes it easy for users to annotate their search and browsing histories, and to include those histories in their posts. In evaluating CiteHistory, we seek to determine how responsive the Q&A community is to this general strategy, and to evaluate CiteHistory's features.

Method

We deployed CiteHistory for a two week evaluation within the same large software company as with the survey study described earlier – in fact, all survey participants received invitations to participate in the study, along with many additional employees, all of whom were software engineers. As such, the demographics for this study were largely the same as in the survey described earlier. Invitations were sent by email, which included a link to a video demonstrating CiteHistory's features, as well as instructions for installation. A gift card drawing was offered as an incentive for installing CiteHistory for a two week period. To receive the incentive, participants were required only to install the plugin. Participants were neither required to post questions or answers, nor to use any of CiteHistory's features when authoring such posts – we wanted to evaluate natural use.

A total of 44 people participated in the two-week deployment study of CiteHistory. During this timeframe, we collected usage logs describing the activities of these 44 individuals, and we conducted 13 user interviews to directly observe their use of CiteHistory. In each of these 13 interviews, we asked participants to attempt to answer a Stack Exchange or MSDN forum question. Each interview lasted at most one hour. In 4 cases, participants could not find an unanswered question they felt comfortable answering in this timeframe. In these cases we altered the protocol to allow users to answer questions which had already received accepted replies. In these cases, we asked participants to go through the motions of writing an answer for the purpose of evaluating CiteHistory's features, but not to submit their answer (since submitting answers to already-answered questions is discouraged by forum moderators).
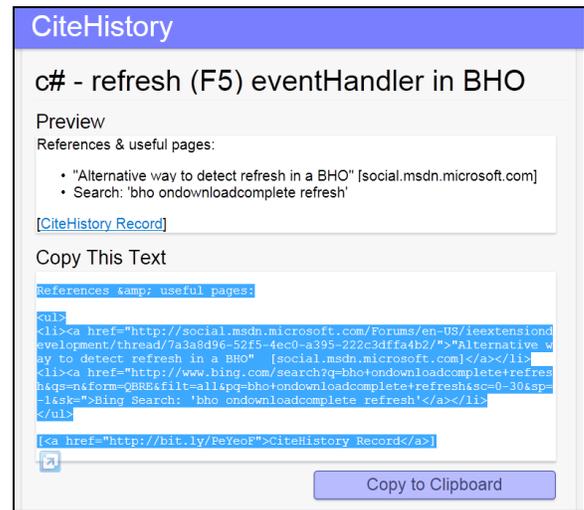
Figure 4. Based on the selected history items, CiteHistory automatically formats bibliographies for inclusion into posts.
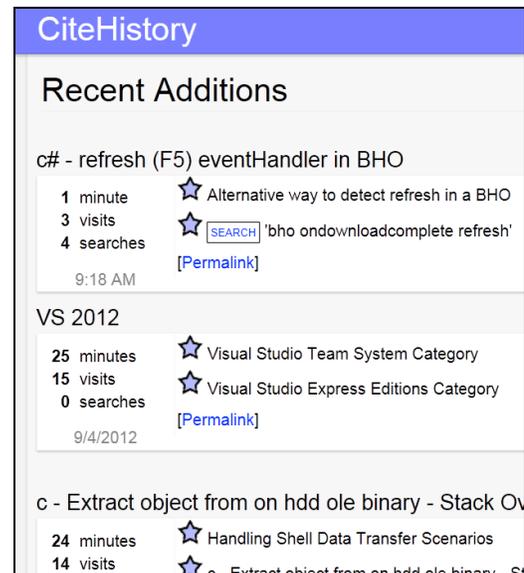
Figure 5. CiteHistory's "Recent Additions" page lists recent bibliographies created by CiteHistory users.
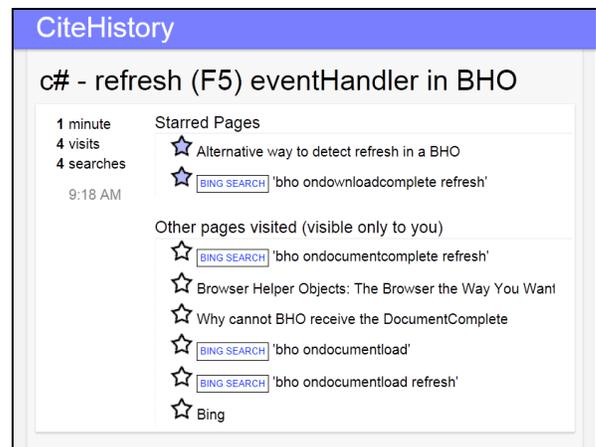

Figure 6. CiteHistory's "My Cites" feature allows users to access their set of private bibliographies.

Results

We first present an empirical analysis of the logs generated by CiteHistory's instrumentation. This analysis describes how CiteHistory was used on a day-to-day basis. We then describe qualitative and quantitative feedback provided by its users, as revealed by the interviews.

Empirical Findings – Instrumentation

As noted above, 44 users participated in CiteHisory's deployment study, contributing a total of 15,063 log entries. In this two-week period, participants answered 53 questions and created 29 distinct bibliographic records (we would not expect users to add bibliographies to all answers, since our earlier survey and log studies indicated that only about half of answers involve online research). Despite CiteHistory supporting the workflows of both asking and answering questions, all 29 records correspond to users answering; this propensity toward answering rather than asking is likely explained by the high expertise levels of our participants, who were all professional software engineers, most of whom had at least 5 to 10 years of on-the-job experience.

In the 29 bibliographic records, users' private bibliographies indicate that they visited an average of 6.03 pages (median: 5), conducted an average of 1.17 searches (median: 1), and spent an average of 4.8 minutes conducting research. The publicly posted portions of the bibliographies contained an average of 1.17 linked pages (median: 1) and 0.17 searches (median: 0). Only 5 of the 29 bibliographic records contained a public inclusion of a search; while this is an increase compared to status quo behavior (such as that in our Stack Exchange log study, in which search terms were included in posts in a negligible number of cases), the low use of this capability suggests that post authors do not appear to value including searches in the bibliographic record, even with the introduction of tools that facilitate doing so.

As noted above, bibliographies were created for 29 of 53 answers (55%). This proportion is considerably higher than the 24% proportion reported in our earlier log study of Stack Exchange (statistically significant by an independent samples t-test with $t(171) = 4.09$ and $p < 0.001$). While we acknowledge that it is difficult to directly compare log data to self-report data, we note that this 55% proportion is also greater than the 39% of MSDN survey respondents who reported including links in their answers. Regarding the number of links per answer, bibliographies publicly listed an average of 1.17 links (noted earlier). This difference in means is significantly lower than the comparable average of 1.93 links for the subset of Stack Exchange answers containing at least one URL ($t(56) = 3.335$, $p < 0.002$). As such, we find that, while CiteHistory increases the proportion of answers containing hyperlinks, the benefits may concentrate at the low end. In other words, CiteHistory may be encouraging more users to include links, but does not appear to increase the number of hyperlinks included by authors who were already in the habit of adding such reference material.

Of the 29 CiteHistory-annotated answers in our dataset, six received positive feedback. Specifically, 2 answers were "up voted," 3 answers were marked as an "accepted" solution, and one answers was both up voted and marked as accepted. While a direct comparison to our Stack Exchange log study is difficult (CiteHistory answers were analyzed after only a few weeks of being posted, while the answers in the Stack Exchange study had up to 6 months to collect votes), we note that we received no overtly negative feedback – the remaining 23 CiteHistory answers simply received no feedback at all. We hope to revisit this comparison in future work.

Finally, the time windows we selected for logging (based on our earlier study findings) appear to be good default choices that achieve the goal of capturing research but minimizing capture of off-topic browsing – log entries were redacted from only 4 bibliographic entries, which suggests that our tight logging window combined with our data encryption appropriately balanced the goals of supporting users' privacy and supporting research sharing.

User Feedback – Interviews & Observations

In addition to the log study of CiteHistory usage, we visited 13 CiteHistory users for one-hour sessions at their place of work, and observed them answering a forum question. Interview participants were primarily positive about CiteHistory. Two themes present in positive feedback were praise for CiteHistory's facilitation of reference sharing, as well as praise for the ability of CiteHistory to track browsing history for one's own personal use. A sampling of positive feedback is listed below:

P5: "This means I don't have to keep track (and save) what I find while researching a problem."

P6: "It standardizes/formalizes the process for (in my case anyway) citing references when I answer."

P10: "The [MyCites] feature is very useful, and can be a personal Q&A notebook"

P11: "I like that it stores my search history for programming sites in one place."

User feedback also revealed areas for improvement. Here, one theme dominated: when selecting pages to include in bibliographies, the page title and URL are often insufficient for discriminating websites on a common topic. 4 of 13 participants commented on this issue. For example, P2 stated the problem best:

P2: *"The short webpage description wasn't sufficient to distinguish between links since all the links in my query were similar. In fact, I accidentally selected the incorrect link."*

A straightforward solution to this issue would be to either provide thumbnail previews of pages during the selection process, or to allow users to pre-emptively flag informative pages as they encounter them online.

Other features requested by interviewees included adding functionality to annotate and categorize one's own bibliographies, and affording users the capability of specifying the locations within pages where helpful information can be found.

User Feedback – Questionnaire

In addition to collecting open-ended feedback, we administered a questionnaire at the end of each interview. In the first portion of the questionnaire, participants were asked about the value of including various types of reference material in answers and questions by rating their agreement to various statements on a 6-point forced-choice Likert-scale. This scale ranged from 1 (strongly disagree) to 6 (strongly agree), and provided no neutral option. This design was intended to partially address the central tendency bias.

All 13 participants responded positively to a statement asserting the value of including reference material when answering forum questions. The average rating for this question was 5.62 on a scale from 1 - 6.

Participants were also asked about the need for sharing reference material when asking questions. Here 12 of 13 responded positively, giving an overall average rating of 4.85 points.

Regarding material that should be listed in posts' bibliographies, all participants agreed that relevant websites should be included (average rating of 5.85), while 12 of 13 felt that relevant search queries should also be included (average rating of 5.00). This is in stark contrast to the results reported earlier, which noted that only 17% of Cite-History bibliographies included listings for searches. Future research would be necessary to reconcile this discrepancy between user's stated preferences versus their actual use of the tool.

The questionnaire employed the same mechanism (a 6-point Likert scale) to assess some of CiteHistory's other features, namely the ability for users to: (1) revisit their own public bibliographies aggregated on CiteHistory's website, (2) revisit their own private browsing histories associated with those bibliographies, and (3) use the Cite-History website to browse the public bibliographies of other users. In all three cases, responses were primarily positive (Respectively: 100% positive, mean rating 5.46; 92% positive, mean rating 5.08; and 92% positive, mean rating 4.77).

Discussion: CiteHistory

The CiteHistory system was designed to facilitate the process of including citations to online resources when composing technical forum posts. Our evaluation showed that the system successfully increased the practice of linking to online resources.

Longer term deployment is necessary to see if this will result in increased ratings of posts by the forum communities (as "up-votes" tend to accumulate slowly over time), but our Stack Exchange log study as well as prior work on more general-purpose Q&A forums (Harper et al., 2008) suggest that link inclusion is likely to enhance post value. Longer term use is also necessary to reveal whether users learn to achieve a good balance of original explanatory text and external pointers in their posts; community feedback on CiteHistory-annotated answers may help authors learn to optimize their answer style.

Measuring clickthrough rates on links included in these bibliographies would be another valuable metric of success; our initial version of CiteHistory used tracked links, but pilot testing revealed that such tracking was not well-received by the highly technical user base of these forums, who deleted such posts. Hence, link-tracking within the forum posts was disabled in our final system.

When we initially designed CiteHistory, our intent was to allow forum participants to visit the CiteHistory web page to view the entire research process associated with a given forum post (via a link embedded in the post). This presentation would include both helpful and unhelpful aspects of the research process, distinguishing between the two via the "star" interface. We hypothesized that this information might have pedagogical value to members of the technical community (e.g., by warning users away from unhelpful routes of inquiry). However, early pilot testing revealed that users redacted unhelpful items rather than including them in their research histories (perhaps due to a desire to manage their image and reputation on the forums); this redaction had the unintended consequences of giving inaccurate impressions of the amount of research conducted, and in removal of these items from users' private MyCites entries. Hence, we modified the tool's design so that only starred items are revealed publicly (along with summary statistics including the total time spent researching, the total number of pages visited, and the total number of searches performed, as in Figure 5).

Finally, although our main design goal in developing CiteHistory was to enrich technical Q&A forums with valuable content about information provenance, our evaluation found that the auxiliary benefit of coalescing a user's various technically-related research efforts into a single repository was valued as well. This "selfish" goal in using the tool could help incentivize the more altruistic benefit of sharing one's sources with the larger community. Addi-

tional self-reflection features, perhaps modeled after tools like the Search Dashboard (Bateman, Teevan, and White, 2012), could be added to enhance this aspect of the system.

Conclusion

When software developers ask and answer questions in online technical discussions, they make extensive use of online resources. In this paper, we explored these behaviors and developed CiteHistory, a tool for sharing references. Our primary contributions included (1) a survey of 129 professional software developers' use of online reference material when posting to forums; (2) a log analysis of web browsing activity proximal to 210 Stack Exchange forum posts; (3) the creation of CiteHistory, a system that facilitates research sharing in forum posts, and (4) an evaluation of CiteHistory through logs from a two-week deployment to 44 software engineers, plus 13 interview/observation sessions.

To conclude, our survey and log studies found that users conduct online research when composing posts for technical Q&A forums, that users generally do not include links to their online research in the posts themselves, and that those posts that do include such links are received more positively by users. Consequently, our CiteHistory tool facilitates and encourages the sharing of reference material, and helps authors keep track of their own research and forum activities. Based on the value we found in CiteHistory from these initial studies, we have released CiteHistory for public use at: http://research.microsoft.com/en-us/um/redmond/projects/citehistory/

References

Aula, A., Jhaveri, N., and Käki, M. 2005. Information search and re-access strategies of experienced web users. In Proc. WWW '05. New York, NY: ACM.

Bateman, S., Teevan, J., and White, R. W. 2012. The search dashboard: how reflection and comparison impact search behavior. In Proc. CHI '12. New York, NY: ACM.

Bhavnani, S. K. 2002. Domain-specific search strategies for the effective retrieval of healthcare and shopping information. In CHI '02 Extended Abstracts. New York, NY: ACM.

Bilenko, M., and White, R. W. 2008. Mining the search trails of surfing crowds: Identifying relevant websites from user activity. In Proc. WWW '08. New York, NY: ACM.

Brandt, J., Dontcheva, M., Weskamp, M., and Klemmer, S. R. 2010. Example-centric programming: Integrating web search into the development environment. In Proc. CHI '10. New York, NY: ACM.

Brandt, J., Guo, P. J., Lewenstein, J., Dontcheva, M., and Klemmer, S. R. 2009. Two studies of opportunistic programming: Interleaving web foraging, learning, and writing code. In Proc. CHI '09. New York, NY: ACM.

Broder. 2002. A Taxonomy of web search. SIGIR Forum, 36(2): 3–10

Gazan, R. 2006. Specialists and Synthesists in a Question Answering Community. In Proceedings of the 69th Annual Meeting of the American Society for Information Science and Technology. Austin, TX: Richard B. Hill

Goldman, M., and Miller, R. C. 2009. Codetrail: Connecting source code and web resources. J. Vis. Lang. Comput., 20(4): 223–235

Harper, F. M., Raban, D., Rafaeli, S., and Konstan, J. A. 2008. Predictors of answer quality in online Q&A sites. In Proc. CHI '08. New York, NY: ACM

Hartmann, B., Dhillon, M., and Chan, M. K. 2011. HyperSource: bridging the gap between source and code-related web sites. In Proc. CHI '11. New York, NY: ACM.

Kapoun, J. 1998. Teaching undergrads WEB evaluation: A guide for library instruction. C&RL News, July/August 1998, 59(7).

Ko, A. J., Aung, H., and Myers, B. A. 2005. Eliciting design requirements for maintenance-oriented IDEs: A detailed study of corrective and perfective maintenance tasks. In Proc. ICSE '05. New York, NY: ACM.

Li, I., Nichols, J., Lau, T., Drews, C., and Cypher, A. 2010. Here's what I did: Sharing and reusing web activity with ActionShot. In Proc. CHI '10. New York, NY: ACM.

Liu, Q., Agichtein, E., Dror, G., Gabrilovich, E., Maarek, Y., Pelleg, D., and Szpektor, I. 2011. Predicting web searcher satisfaction with existing community-based answers. In Proc. SIGIR '11. New York, NY, USA: ACM.

Mamykina, L., Manoim, B., Mittal, M., Hripcsak, G., and Hartmann, B. 2011. Design lessons from the fastest Q&A site in the west. In Proc. CHI '11. New York, NY: ACM.

Microsoft. 2012. MSDN Site Statistics. Retrieved September 12th, 2012, from http://social.msdn.microsoft.com/forums/en-US/categories/

Moraveji, N., Morris, M.R., Morris, D., Czerwinski, M., and Henry Riche, N. 2011. ClassSearch: Facilitating the development of web search skills through social learning. In Proc. CHI '11. New York, NY: ACM.

Morris, D., Morris, M.R., and Venolia, G. 2008. SearchBar: A search-centric web history for task resumption and information re-finding. In Proc. CHI '08. New York, NY: ACM.

Schraefel, M. C., Zhu, Y., Modjeska, D., Wigdor, D., and Zhao, S. 2002. Hunter Gatherer: Interaction support for the creation and management of within-web-page collections. In Proc. WWW '02. New York, NY: ACM

Shah, C., and Pomerantz, J. 2010. Evaluating and predicting answer quality in community QA. In Proc. SIGIR '10. New York, NY: ACM.

Stack Exchange Inc. 2012. Stack Overflow. Retrieved September 12th, 2012, from http://stackoverflow.com/questions

Treude, C., Barzilay, O., and Storey, M.-A. 2011. How do programmers ask and answer questions on the web?: NIER track. In Proc. ICSE '11. New York, NY: ACM/IEEE.

W3C Provenance Working Group. 2012. Provenance WG Wiki. Retrieved September 14th, 2012, from http://www.w3.org/2011/prov/wiki/Main_Page

White, R. W., Dumais, S. T., and Teevan, J. 2009. Characterizing the influence of domain expertise on web search behavior. In Proc. WSDM '09. New York, NY, USA: ACM.