Seeking Information in Online Environments – Where, Who, and Why?

Chirag Shah, Rutgers University,
Jennifer Sonne, Rutgers University,

Abstract
Given the ubiquity of social media and Web 2.0 resources, there is a current lack of knowledge about the complementary aspects of individualized and social search strategies. This paper looks at the information resource preferences of users, focusing on online resources. Two exploratory studies were done to analyze the motivations behind online information seeking behavior, specifically looking at where people go for information and reasons behind that decision. The first study collected log data from users who used the Web for browsing and searching information, and also asked questions on Q&A sites. The second study used a survey with four different scenarios that asked respondents to rank different information resources. The findings from these studies provide a more comprehensive understanding of how and why people choose to use an information resource/method depending on their information needs.

Keywords: Information seeking, social media, Q&A websites, web search engines
Copyright: Copyright is held by the author(s).
Acknowledgements: This work was supported in part by the Institute of Museum and Library Services (IMLS) Early Career Development grant #RE-04-12-0105-12 awarded to Chirag Shah and in part by the National Science Foundation (NSF) BCC-SBE award #1244704. The authors are also grateful to Erik Choi, whose dissertation work led to Study-1 described here.
Contact: chirags@rutgers.edu, jennifer.sonne@rutgers.edu

1 Introduction
When people search for an answer to a question, they use all known available resources to find the answer most quickly. They satisfice (March & Simon, 1958) and crosscheck; comparing what they think they know with what they can find. Even if they are searching for information individually, they still “have a strong social inclination throughout the search process” and will utilize their social connections when possible (Evans & Chi, 2009). While there are several works in the literature that study individual information seeking (Kuhlthau, 1993; Bates, 2002) and social information seeking (Morris, Teevan & Panovich, 2010; Evans, Kairam, & Pirolli, 2010; Oh, Oh, & Shah, 2008), they are often studied exclusively, disregarding the larger context in which people seek information - some of which may be individual, other may be social, and some may be overlapping. To address this shortcoming, we took an exploratory research approach to investigate how different forms of information needs may lead to different methods of seeking information.

To find out more about how these social connections play a role in people’s overall information seeking strategies, we conducted two different studies – one based on log data collected from 18 participants over a period of a month, and the other using a survey completed by 105 responders. Analyzing the data from the first study yielded four interesting scenarios of information seeking, which were then used as the primary part of the survey for the second study. Study-2 participants were presented with four scenarios that asked them to pick their top three choices from five information resources: Person, Professional, Web Search, Social Media, and Q&A Site. By asking the participants to choose from an array of resources for a scenario, we hypothesized that they would hone in on the resource they personally found the most useful for similar scenarios. However, would there be one consistent resource chosen first for all four scenarios, or would participants value resources differently based on the type of scenario? How valuable are social connections to people’s overarching information seeking strategies? It turns out that when people go to seek information, who they choose to ask for information, and why they do so heavily depends on the kind of information need. While this may not come as a surprise, the work reported here locates all of these possibilities under one umbrella in order to directly compare individual-based and social-based information seeking methods.

In the end, by teasing out the implicit value of each resource, we hope to contribute to social information seeking research. This study’s preliminary results also supply implications for further research in the complementary aspects of individualized and social search strategies.

The rest of the paper is organized as follows. In Section 2, we provide an overview of several popular methods for seeking information online, divided into individual and social-based approaches. We then point out that there is a lack of research that connects these modalities by studying online...
information seeking in context. This leads to two studies we conducted, described in Section 3. The studies expose interesting insights that are reported in Section 4. We conclude the paper, listing contributions and limitations of this work, as well as giving pointers for future research.

2 Background
This section will provide a brief overview of different popular methods used for seeking information online and some of the recent works that study those methods.

2.1 Individual-based Approaches
Using a web search engine to find information has been studied extensively. Choo, Detlor, and Turnbull (2000) studied knowledge workers' use of the web to find information. They found that people who use the web daily “engage in a range of complementary modes of information seeking,” suggesting that the type of behavior depends on the type of need (Choo, Detlor, & Turnbull, 2000). A study conducted by Kellar, Watters, and Shepherd (2007) also found that different types of information seeking tasks had different characteristics. Additionally, almost all of the college students surveyed in a study by Head and Eisenberg (2011) “used Web search engines for gathering everyday life information,” and they cited the ease and familiarity of web search engines (specifically Google) as their primary reason to use web search. Given the ubiquity of web search use, we expect in our study to see most people use web search as their primary information resource.

Web search is just one of the many ways people can search for information individually. Other forms of individualized information resources include physical resources such as libraries, newspapers, and television. In this digital age, though, most individual information seeking is done online through the use of databases, online catalogues, and general web browsing. Robinson (2010) compared engineers’ information behavior between the percentage of time spent with human and nonhuman information sources, and found that people spent more time with nonhuman sources. This suggests that although people may use others to find information, they nonetheless spend a significant portion of time looking for information individually.

2.2 Social-based Approaches
Web search engines help people find information individually, but there are many online resources that people can use to help them find information socially. Shah, Kitzie & Choi (2014) defined social information seeking as a process that “involves the activities taking place within participatory social media sites, such as social networking (e.g., Facebook), social bookmarking (e.g., Delicious), social photo/video sharing (e.g., Pinterest), blogging (e.g., Tumblr), and micro-blogging (e.g., Twitter), in order to fulfill an individual’s desired information need.” Social information seeking encompasses two of our categories used in this study that are described in more detail below: social media and Q&A sites. It is important to note that the concept of social information seeking has been mentioned in recent literature, because our study is interested in comparing different resources found in social information seeking.

Social media has become a complementary information resource to web search engines. A survey conducted by Morris, Teevan, and Panovich (2010) evaluated the use of social networks as question and answer services. They found that social networks provide an avenue for information seeking that many people use in lieu of web search engines and Q&A sites. Depending on the type of need, social media can provide faster answers that are more trusted and personalized than those of Q&A sites or web search engines (Morris, Teevan & Panovich, 2010).

Evans, Kairam, and Pirolli (2010) also found that “social resources may be essential for answering a certain class of queries and, in partnership with powerful search engines, could greatly assist users on their quest for information.” Additionally, more than half of the college students surveyed by Kim, Sin, & Tsai (2014) were found to use social media as an information resource. As a rising resource, social media’s potential is still being explored. We would like to see how social media ranks as an information resource for users when they are confronted with a specific question.

Q&A site usage corresponds more to social media usage than web search engine, in that they are participatory rather than individualistic. A review of the literature on Q&A sites by Gazan (2011) reported that Q&A sites “are more than the sum of their questions and answers, and that people participate for reasons well beyond a quest for facts” (Gazan, 2011). Part of the appeal of Q&A sites is the ability to post intricate questions with greater anonymity than in social media (Morris, Teevan & Panovich, 2010). However, there are multiple motivations to use Q&A sites, including fulfilling cognitive needs and having fun (Choi, Kitzie, & Shah, 2014).
We can see that while there has been plenty of literature on individual and social information seeking, there is a lack of research that focuses on an individual’s information needs in the larger context of online information seeking, using multiple methods that involve different systems and people.

3 Method

3.1 Study-1: Log Analysis

We were initially interested in the information behavior of Q&A site users within web search sessions, so our first study collected log data from a total of 18 Q&A site users, 10 Yahoo! Answers users and 8 WikiAnswers users. These participants were selected from a pool of nearly 200 respondents to an online questionnaire, and chosen based on how they represent different groups of people. More details can be found in (Choi & Shah, In press).

To implement the log data collection, we used an extension toolbar for the Firefox and Chrome browser called Coagmento (http://coagmento.org) (see González-Ibáñez & Shah, 2012; Shah, 2010 for a detailed description of Coagmento). This extension toolbar automatically collected anonymized Web search information and also allowed the participants to manually write a diary entry about their questioning behavior each time they asked questions on Q&A sites.

A total of 112,476 Web pages were viewed by the 18 participants during 4 weeks. An average of page visits among the participants was 6,140.33 (S.D=8,557.23), ranging from 111 to 30,856 page visits. This signifies that some participants in the log data collection may be more likely active Web users than other participants, or particular information search strategies in the online environments may cause more page visits depending on the topics or contexts of information seeking. In addition, 205 incidents of asking a question in Yahoo! Answers and WikiAnswers were collected through the extension toolbar during a 4-week period of period of log data collection. The participants were also required to choose a specific motivation and expectation, and to provide a short narrative of their motivation and expectation each time they wrote their diary about the question.

3.1.1 Information Seeking Strategies within Web Search Session

We examined how online Q&A services (i.e., Yahoo! Answers, WikiAnswers) were utilized to seek information within each web search session. Since the participants’ interactions within the browser were targeted this analysis investigated the participants’ information seeking strategies that consisted of not only search engines (e.g., Google, Yahoo, etc.), but also online Q&A services, in order to find relevant information that helps satisfy their needs. However, it was found that the participants exclusively use online Q&A to seek information without employing any other Websites (e.g., search engines, etc.) when they ask a question in online Q&A (N=161, 78.54%). This suggests that even though people have access to multiple websites sequentially or simultaneously when they ask a question, it may not necessarily indicate that they search a similar topic using these multiple Websites, including online Q&A. In other words, people tend to search information on a specific topic through the question-answering processes within the online Q&A environment, while they search dissimilar topics via other sites or use the Internet for different reasons (e.g., access to their social networking sites such as Facebook, etc.).

A small number of incidents show that the participants utilize different information sources before and/or after asking a question in online Q&A for a variety of reasons, including verifying information, receiving diverse information, etc. One participant (#YA7), for example, first used Yahoo! Answers to ask a question, “what is the name of this plant?” and attached a specific photo of the plant to provide a better understanding of the plant. Once they received answers to their question, they used a web search engine to search for the specific name of the plant, “scented geranium.” Additionally, it was found that another participant (#WA6) attempted to search for specific information about filing taxes jointly for a same-sex married couple using different websites through a search engine. However, after visiting a variety of websites, they visited WikiAnswers to ask, “Can a married same-sex couple use online Turbo Tax to file state taxes in a state (VA) that doesn't recognize their marriage?” It can be assumed in this information search session that the users were either verifying information found from the websites they visited before WikiAnswers, or that they failed to find satisfactory information and then used WikiAnswers to receive an answer for their specific need.

Finally, it was found that 127 (61.95%) incidents of asking a question were done solely to seek information, and that the participants tended to ask more than one question in order to satisfy their need (N=78, 38.05%). However, 64 out of 78 incidents where multiple questions were asked were not related to one another, suggesting that people may sometimes ask multiple unrelated questions to seek
information on different topics when they are still in online Q&A sites. Table 1 presents a summary of how the participants used online Q&A within their information search sessions in the online environment.

3.1.2 Four Information Seeking Scenarios

We examined the questions that the 18 users from study-1 posted to Q&A sites and chose four questions that covered different topics but that also demonstrated a mix of motivations and expectations along with different behaviors before, during, and after asking their questions.

User 19’s sample question, “Can a human bladder be stretched through exercise to reduce bathroom trips?” was asked after doing a Google search for “bladder stretch.” They then went to health.com and looked at similar questions while they waited for their WikiAnswer question to be answered. User 19 spent a total of 23 minutes on this question, and never revisited it. This session was very targeted, as User 19 did not visit any other websites that did not pertain to this question.

Alternatively, User 27 was not targeted at all in their question, “where can you audit classes for free as a senior citizens?” They asked the question directly in Yahoo! Answers, but did not do any related web searches and did not look at any other websites related to their question. In fact, User 27 did not even revisit their question after posting it to Yahoo! Answers.

<table>
<thead>
<tr>
<th>Information seeking strategies</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Related search*</td>
<td></td>
</tr>
<tr>
<td>Search engines → Online Q&amp;A</td>
<td>13 (6.34%)</td>
</tr>
<tr>
<td>Other sites → Online Q&amp;A</td>
<td>8 (3.90%)</td>
</tr>
<tr>
<td>Online Q&amp;A → Search engines</td>
<td>9 (4.39%)</td>
</tr>
<tr>
<td>Online Q&amp;A → Other sites</td>
<td>14 (6.83%)</td>
</tr>
<tr>
<td>Unrelated search**</td>
<td></td>
</tr>
<tr>
<td>Unrelated search → Online Q&amp;A → Unrelated search</td>
<td>161 (78.54%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>205</td>
</tr>
</tbody>
</table>

Table 1. A summary of the participants’ information seeking strategies.

*Related search means that the participants use different online information resources in order to search a similar topic after/before posing a question in online QA

** Unrelated search indicates that the participants use those resources to seek information for different topics after/before posing a question in online QA.

User 7 exhibited behavior that corresponded the closest to our predicted structure when they asked their question on Yahoo! Answers, “What are some good party games for college aged people?” They asked their question directly without browsing the web beforehand. After they asked their question the user did a lot of web browsing of unrelated websites. User 7 then checked their question in the same session, 6 minutes after first asking the question, and then re-checked their question again 83 minutes later.

In the middle of a very long web session, User 13 asked their question on Yahoo! Answers, “What advice do you have for dealing with a new college roommate?,” but they never revisited their question after asking it. User 13 viewed many websites both related and unrelated to their question before and after asking their question.

3.2 Study-2: Survey

After analyzing the data and the four sample questions from our preliminary work, we found the behavior of the Q&A site users intriguing. We had no explanation for why users would go to an online Q&A site in the midst of doing unrelated web searches. People are obviously utilizing multiple resources online while searching for information, but what drives them to each resource? To explore this phenomenon further, we created a survey based on both the literature (Evans and Chi, 2009) and the findings from our first study to gain insight into why and how people use different information resources to find information.
Particularly, we were interested in user explanations for their rankings of different information resources based on the question they were given.

We used Amazon Mechanical Turk (MTurk - https://www.mturk.com/) to conduct our survey. Since we were primarily interested in exploration, the readily available users made MTurk a preferable tool. We received 110 responses, but removed 5 due to nonsensical answers.

Most of the respondents were between 25-44 years old and had received a Bachelor’s degree. There was a fairly even split between Male and Female respondents. See Table 2 for a complete breakdown of the demographics.

### 3.2.1 Survey Content

The survey had two sections: the first section asked respondents to answer questions on how often and why they used each information resource category. The second section included four scenarios that gave each respondent a question and asked them to pick their top three choices from the five information resource categories and explain how and why they chose each resource. The former was taken from Evans and Chi (2009), and the latter was constructed using the results from study-1 described earlier. The four scenarios presented in the second part are as follows:

- **Scenario 1**: “Where can you audit classes for free as a senior citizen?”
- **Scenario 2**: “Can a human bladder be stretched through exercise to reduce bathroom trips?”
- **Scenario 3**: “What are some good party games for college aged people?”
- **Scenario 4**: “What advice do you have for dealing with a new college roommate?”

Scenarios 1 and 3 are everyday questions that fit into the Everyday Life Information Seeking (ELIS) model (Savolainen, 1995), Scenario 2 is a health related question, and Scenario 4 is a social question. We used these questions because we wanted to see if different question types elicited different responses.

### Analysis and Findings

#### 4.1 Information Resource Preferences (Where and Who)

For Scenarios 1 and 3, the majority of the participants chose to search on the web first (see Tables 4a and 4b). These questions deal with everyday life and health issues, so beginning with a web search is very common (Head & Eisenberg, 2011). Most participants chose to consult a person or professional second, giving the reason that it would not be as convenient or quick (“I think it would take longer for valuable answers”). They did believe consulting a person or professional would be useful, but the timeliness of using a web search was more desirable than getting the opinions of others.

The respondents’ preferences are slightly different between the second and third choices for Scenarios 1, 2, and 3. This is more interesting to us, because there is a large jump between the first choice and the last two choices in those scenarios. Perhaps the other information resources are viewed as more interchangeable due to the direct and indirect social connections, whereas web search is the only resource that someone would use that has no direct social connections. Social interactions can be nuanced and vary greatly, so respondents could have chosen information resources that they have had success with in the past. Our survey does not ask the respondents to actually answer the questions, but rather asks them to hypothesize what resources they would use to answer the questions.
For Scenario 4, a question that is very clearly socially driven, the majority of respondents chose to consult a person first. The second choice for this scenario was more evenly split between consulting a person, searching on a web search site, or asking on social media. Many participants said web search would not be as useful because of the social nature of this question, but others believed that they could find a good answer in the large amount of information available. In general, respondents were looking for opinions and advice of people they trust.

Respondents did not choose one resource first for all four scenarios (as can be seen clearly in Table 4b). Instead, they chose the resource depending on the type of question. Also, given that people tend to be creatures of habit, our results found that the participants chose resources based on familiarity and ease rather than on how well the resource fit the question.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Person</th>
<th>Professional</th>
<th>Web Search</th>
<th>Social Media</th>
<th>Q&amp;A Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 1- 1st Choice</td>
<td>10.48%</td>
<td>8.57%</td>
<td>78.10%</td>
<td>2.86%</td>
<td>0%</td>
</tr>
<tr>
<td>Scenario 1- 2nd Choice</td>
<td>6.67%</td>
<td>16.19%</td>
<td>72.38%</td>
<td>0.95%</td>
<td>3.81%</td>
</tr>
<tr>
<td>Scenario 1- 3rd Choice</td>
<td>27.62%</td>
<td>0.95%</td>
<td>43.81%</td>
<td>22.86%</td>
<td>4.76%</td>
</tr>
<tr>
<td>Scenario 2- 1st Choice</td>
<td>47.62%</td>
<td>4.76%</td>
<td>32.38%</td>
<td>11.43%</td>
<td>3.81%</td>
</tr>
<tr>
<td>Scenario 2- 2nd Choice</td>
<td>26.67%</td>
<td>20.95%</td>
<td>16.19%</td>
<td>12.38%</td>
<td>23.81%</td>
</tr>
<tr>
<td>Scenario 2- 3rd Choice</td>
<td>8.57%</td>
<td>33.33%</td>
<td>20.95%</td>
<td>6.67%</td>
<td>30.48%</td>
</tr>
<tr>
<td>Scenario 3- 1st Choice</td>
<td>23.81%</td>
<td>3.81%</td>
<td>35.24%</td>
<td>20.95%</td>
<td>16.19%</td>
</tr>
<tr>
<td>Scenario 3- 2nd Choice</td>
<td>21.90%</td>
<td>13.33%</td>
<td>26.67%</td>
<td>24.76%</td>
<td>13.33%</td>
</tr>
<tr>
<td>Scenario 3- 3rd Choice</td>
<td>32.38%</td>
<td>26.67%</td>
<td>9.52%</td>
<td>19.05%</td>
<td>12.38%</td>
</tr>
<tr>
<td>Scenario 4- 1st Choice</td>
<td>31.43%</td>
<td>26.67%</td>
<td>20%</td>
<td>14.29%</td>
<td>7.62%</td>
</tr>
<tr>
<td>Scenario 4- 2nd Choice</td>
<td>21.90%</td>
<td>7.62%</td>
<td>15.24%</td>
<td>26.67%</td>
<td>28.57%</td>
</tr>
<tr>
<td>Scenario 4- 3rd Choice</td>
<td>17.14%</td>
<td>11.43%</td>
<td>21.90%</td>
<td>20.95%</td>
<td>28.57%</td>
</tr>
</tbody>
</table>

Table 4a. Percentage of Respondents that Chose Each Information Resource.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Person</th>
<th>Professional</th>
<th>Web Search</th>
<th>Social Media</th>
<th>Q&amp;A Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 (.346)</td>
<td>3 (.279)</td>
<td>1 (.894)</td>
<td>5 (.154)</td>
<td>4 (.160)</td>
</tr>
<tr>
<td>2</td>
<td>4 (.214)</td>
<td>2 (.417)</td>
<td>1 (.895)</td>
<td>5 (.090)</td>
<td>3 (.216)</td>
</tr>
<tr>
<td>3</td>
<td>2 (.468)</td>
<td>5 (.054)</td>
<td>1 (.665)</td>
<td>3 (.422)</td>
<td>4 (.224)</td>
</tr>
<tr>
<td>4</td>
<td>1 (.643)</td>
<td>5 (.152)</td>
<td>2 (.530)</td>
<td>3 (.308)</td>
<td>4 (.200)</td>
</tr>
</tbody>
</table>

Table 4b. Ranking of Information Resource Scenarios according to Mean Reciprocal Ranks (MRR) (MRR calculation in parentheses).

4.2 Reasons for Choosing Information Resources (Why)

Respondents generally provided similar reasons in the freeform responses for why they chose each information resource. The specific wording they chose was not consistent, however, and so to capture the similarities and differences between reasons, we used the framework proposed by Shah and Kitzie (2012) to categorize each information resource ranking. Although there are many frameworks we could have used, this is a reasonable basis for our analysis because their article delves deeply into the literature to find characteristics of answers that users consider good answers (Shah and Kitzie, 2012). We ultimately chose this framework because it mirrors what we asked our users in study-2: what makes their chosen resource a good resource?
In this framework, the three main concepts of relevance, quality, and satisfaction are further divided into cognitive, social-situational, social-affective, and social-organizational/collaborative factors (Shah and Kitzie, 2012). The 12 factors function “as distinct, high-level categories,” and using these factors allows us to evaluate the strengths and weaknesses of our five information resources (Shah and Kitzie, 2012).

**High-Level Concepts:**

- Relevance
- Quality
- Satisfaction

**Sub-Categories of Each Concept:**

- Cognitive
- Social-Situational
- Social-Affective
- Social-Organizational/Collaborative

The reasons we observed that respondents gave most frequently are listed in Table 3 (examples of respondents’ wording is provided in italics). The reasons we found in our survey closely match those in the literature (Evans & Chi, 2009; Borgatti & Cross, 2003; Morris, Teevan, & Panovich, 2010). For questions that had a social component to what is considered a ‘good’ response, the participants often chose social information resources such as social media or consulting a person or professional. For questions that users wanted cognitive answers to, they often chose to use web search.

Motivation for information seeking has been extensively researched and is not the primary interest in our exploration, but validating that our respondents were answering our survey to the best of their ability was important. We can therefore learn about their resource preferences from their responses.

5 Conclusion

People love to use a web search engine to find information. However, we found that people do have other resource preferences depending on the type of question. This typically depends on the kind of information need. From the survey we conducted, we also found that people tend to value web search the most when looking for information. Web search, while chosen the most, was not chosen exclusively as the first choice. Additionally, people’s social connections play a large role in their information seeking strategies, as they chose other people, professionals, and social media frequently in their survey answers. An interesting realization is that while we thought at least some of the sources of information were mutually exclusive, the findings suggest that often people use multiple modes of seeking information to complement the information they find. For instance, we found (mainly from study-1) that people do not necessarily use social Q&A services when their web search fails; they may use both these methods because each of them provides different value.
There are limitations to the work we conducted. Many of our findings are based on an online survey, which limits the interactions we could have with the respondents. Their responses are also limited to short questions that we asked, so the questions were externally prompted rather than self-motivated. The participant pool was limited to 105 people, and did not necessarily cover people with all kind of backgrounds and experiences. This limits the generalizability of our findings. Due to the nature of this method and the way it was executed (survey with Mechanical Turk), we also do not know the real identity of our respondents. We did, however, check for any erratic responses and removed those that were found to be problematic in any way.

Overall, our findings validate that people use different information resources based on the type of question they are trying to answer. This is not a new concept, but it is important to realize that in the current environment where there are a multitude of resource choices, people rely on familiar tools and resources. They also tend to use multiple sources of information depending on the type and intensity of need. And because of this, it would be ideal to combine current technologies to take advantage of the strengths of each resource. Although people mostly begin their search process by utilizing a web search, they often also receive affirmation and other opinions from their social networks at different points before, during, and after their search. It will be important to continue to research the connections between information resources, especially social media and Q&A sites.

References


