Contextualizing Consumer Health Information Searching: an Analysis of Questions in a Social Q&A Community

Yan Zhang
School of Information
University of Texas at Austin
Austin, TX, 78701
(1)512-471-9448

yanz@ischool.utexas.edu

ABSTRACT

The existing research on consumer health information searching has focused on types of information that users need concerning a particular disease and their querying behavior in general search engines. The context of health information searching, such as people's goals, motivations, and emotions, has received comparatively little attention. Given the significant impact of context on the result of information searching, this study sets out to explore contextual factors of consumer health information searching by analyzing health-related questions that people posted on Yahoo Answers, a social Q&A site. Particularly, we looked at the following factors: linguistic features of the questions that users formulated, their motivations for asking the questions, the time when the questions were asked, and their cognitive representations of the problem space. The results could improve our understanding of consumer health information searching and provide implications for the design of general-purpose health information search engines and consumer health information systems.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: User/Machine Systems – *human factors, human information processing.*

General Terms

Design, Human Factors

Keywords

Consumer health information, information searching in context, mental models, Q&A community, information retrieval

1. INTRODUCTION

A Pew study reported that 83% of internet users in the U.S. look on the web for health information. The information found online showed significant impact on consumers' decisions about how to treat an illness or condition and approaches to maintain their health or the health of someone they help take care of [10]. As one of the most popular and increasingly influential activities online, consumer health information searching has become an important research area and calls for a more systematic exploration of patterns of the searching behavior [15].

The existing research on consumers' health information searching and navigation behavior often relies on transaction logs [27]. The research results consistently suggested that users formulated short queries and the queries were not effective in retrieving relevant information [38, 48]. Queries are a simplistic or minimalist expression of consumer health information needs. They can provide valuable insight into the subject of the information needed by people, but they have limited power in revealing contextual information such as user goals, motivations, and background knowledge that give rise to the need [28, 32]. As has been suggested by researchers in information retrieval (IR), a more fruitful approach to IR is to focus on structured representations of information needs that capture pertinent contextual information of the task at hand [2, 22]. Thus, understanding contextual factors of consumer health information searching is of vital importance for designing effective consumer health systems.

When faced with a problem, the most natural way for a person to seek information is to ask questions of someone with more expertise or share similar experience [44]. Web-based technologies, from newsgroups, listservs, and bulletin boards, to the more recent blogs, wikis, social networks sites, and social Q&A sites (where users can ask questions and respond to others' questions), allow people to ask questions, in full sentences, to a broad audience and elicit answers from them. Unlike queries submitted to search engines, the messages posted on these platforms are expressed in natural language, unconstrained by query syntax or prescribed metadata. They can be a good source for authentic user needs and contextual factors (e.g., motivations and emotion) of user information searching [8, 39]. In this study, we analyze health-related questions that users posted on a major social Q&A website, Yahoo Answers, with the intention to explore contextual elements that characterize general consumers' health information needs and information searching. The study is explorative in nature and the results could have implications for the design of information retrieval (IR) and consumer health information systems at both algorithm and interface levels.

2. RELATED LITERATURE

One main approach for consumers to seek health-related information is to search. It was consistently reported that queries that general users submitted to search for health information were simple and short. Spink et al. [38] reported an average of 2.2 terms per health-related query in general-purpose web search engines. Zeng et al. [45] found that about 90% of the queries that users submitted to a hospital website and MedlinePlus, a consumer health information portal created by the National Library of Medicine, contained 1-3 terms. A more recent analysis of queries for medical and health information from general search engines reported an average number of 3.3 terms a query [43].

Medicine is a highly specialized area. When formulating queries, general consumers, including patients, their caregivers, and people concerned about health, without medical training often have difficulties in formulating their requests in a few terms. The difficulties are at three levels. At the mental model level, laypeople often have drastically different mental representations of diseases or medical specialty from medical professionals. They are more likely to describe the diseases or conditions using simpler or more concrete terms [45]. At the semantic level, laypeople's vocabulary often does not match with the medical terminologies, such as those found in UMLS and used in the content of health information websites or index services of medical information collections [19]. Sometimes, general users can not even find proper terms to describe their actual intentions [47]. At the lexical level, misspellings, partial words, and use of abbreviations, are common in health information searching, which often cause search failures [27, 36].

Queries can provide valuable insight into the subject of information requested by users and offer guidance for better query recommendation and processing mechanisms for medical IR systems [24, 43, 26, 48]. Due to the succinct nature and users' conceptual and lexical difficulties involved in formulating queries, however, they are limited in informing researchers other aspects of people's information needs, particularly the goals and motivations of their health information seeking and the emotions involved in it. Thus, queries are not a sufficient representation of consumers' health information needs [28].

In the recent decade, online health-related communities (or online support groups) gradually became widespread on the web. The emergence of these communities was considered as one of the most promising aspects of the rise of ehealth [9]. These communities often consist of patients with a particular disease (e.g., breast cancer, knee injuries, or Lou Gehrig's disease), their caregivers, and sometimes, a few healthcare professionals. They have appeared in increasingly interactive forms, from the earlier newsgroups, listservs [21, 41], and bulletin boards [7, 30] to the later web forums [35], blogs [20, 1], wikis, social networks [11], and social Q&A sites. On these user-centered collaborative platforms, users seek information and emotional support, respond to questions, raise another person's spirit, post announcement, thank others for the support, or express advocacy or political viewpoints [12, 21, 30, 41].

As a recent development, social Q&A sites are gradually gaining popularity [34]. The sites are a form of web 2.0 platform designed specifically for people to ask questions and answer questions from their peers. Different from other aforementioned collaborative platforms, where users-posted messages serve a wide range of purposes and not all contain questions [7, 11, 33], the questioning messages posted on Q&A sites tended to be more homogeneous in terms of their purposes: they are often an expression of users' information needs motivated by a recognition of a knowledge gaps or an anonymous state of knowledge (ASK) [2, 42, 44].

Past research on online support groups mainly focused on identifying demographic characteristics of the participants (e.g., gender and age), their use of the communities (e.g., number of active participants, frequency of usage, and themes or categories of the messages) [1, 12, 21, 25, 30], linguistic features of consumer health vocabularies [37], and subsequent behavioral changes of the participants [20]. However, few treated user-generated questions as a manifestation of consumer health information needs and as a source for informing IR and consumer health information systems [41]. Understanding users' information needs is fundamental for the design of information systems [2]. Designing effective consumer health information systems also requires in-depth understanding of users' health information needs [19]. Since questions expressed in natural language are a comparatively rich reflection of people's information needs, this study intends to analyze characteristics of user-generated health-related questions posted on a major social Q&A website, Yahoo Answers, to increase the current understanding of consumer health information needs and information searching.

3. RESEARCH METHOD

3.1 Dataset

Yahoo Answers (http://answers.yahoo.com) is a major social Q&A site, having about 74% of the market share of U.S. visits among Q&A websites [17]. The site has about 26 top level categories, such as arts, entertainment, music, travel, and health. The data was crawled in early 2008. The initial broad categories to start the crawl were "Health", "Education and Reference", "Sports", "Science and Mathematics", and "Arts" [23]. Because we are interested in health related questions, we extracted the questions in the Health categories, which include 77,903 questions across 23 sub-categories, such as cancer, heart disease, diabetes, and women's health. In order to identify the characteristics of general consumer health information needs and the ways in which users express them, we randomly selected 276 questioning messages across all the 23 sub-categories for a content analysis.

3.2 Data analysis

The randomly selected 276 questioning messages were imported into QSR N6, a content analysis software. Each message is a coding unit. We first coded the number of questions contained in each message. A question is loosely defined as a request that an asker send to one or more respondents to solicit knowledge on a certain subject, which the asker lacks but sincerely wants [41]. Identifying questions was based on meaning, not form. A declarative sentence beginning with "I would like to know" and soliciting input from respondents, was treated the same as questions beginning with interrogatory words such as "what" and "where".

Then, open coding method was employed to code the data. First we read questioning messages several times to gain an overview of the overall content. Then codes were derived inductively by closely reviewing themes appeared in each message. Whenever a contextual description appears, such as for whom a question was asked, the motivations of asking the question, emotions involved, and difficulties in expressing their needs, it was coded into a category. When coding a new text to a category, the text was compared to those already assigned to that category [13]. The constant comparison method allowed us to fully understand the theoretical properties of the category. We then examined the categories resulted from the open coding process to make sense of the properties and dimensions of the categories, identify relationships between them, and uncover patterns [4]. To ensure the validity of the results, a second coder coded 20% of the questions. The inter-coder agreement reached 87.9%.

4. RESULTS

The context within which a person seeks information is multidimensional and interactive in nature. Although no consensus on what constitutes a context has been reached, researchers have generally agreed that the elements of context include cognitive, affective, social, linguistic, technical, organizational, or cultural factors related to a person and his/her tasks and goals [6, 16, 18]. In this study, we explore the context of consumer health information searching in terms of the following contextual elements: basic linguistic features of askers' expressions of their health information needs, their cognitive, social, and emotional motivations for asking questions on a Q&A site, time when the questions were asked, and askers' cognitive structure of their problem spaces.

4.1 Basic linguistic features of consumer health questions

As suggested by its name, Yahoo Q&A is a site designed mainly as a platform for users to ask questions. In the study, questions were loosely defined as messages that an asker posted with the intention to elicit useful information from other users to fill his/her knowledge gap. An analysis of the selected questioning messages found that 9 (3.3%) messages did not meet the definition. Among the 9 messages, 3 were advertisements, 3 tried to elicit comments from people, 1 asked for prayers, 1 was a survey, 1 urged for early check for colon cancer, and 1 was a rhetorical message condemning drug companies. These messages were excluded from the subsequent analysis. For linguistics features of the questions in Yahoo Q&A, we looked at the length of the messages containing questions and the density of questions.

4.1.1 Length of the messages containing questions

Unlike in general web search engines where users are used to typing in short queries, users of Q&A sites have the freedom of typing questions as long as they want. The length of the messages that contain questions varied, ranging from 3 to 364 terms (*Mean* = 47.5, *SD* = 47.9; *Median* = 32.5). An example of a 3-term question is "*What is prehypertension?*" Longer messages often provide more contextual information, such as age, gender, symptoms, and medical history.

4.1.2 Density of questions

Density is defined as the average number of sub-questions per message. In Yahoo Q&A, a questioning message could contain multiple sub-questions. The sub-questions were also coded based on meaning rather than form. In one message, if the same thing were asked in two different forms or using two different expressions, it was counted as one sub-question instead of two. For messages containing questions, the number of sub-questions ranged from 1 to 5 (*Mean* = 1.34; *SD* = 0.70), with the majority (74.8%) containing only one sub-question. Forty-nine messages (18.4%) contained two sub-questions and the rest (6.8%) contained 3 to 5 sub-questions.

4.2 Motivations for asking questions

People ask questions when faced with problems [44]. Similarly, users came to Yahoo Q&A to ask health questions often do so due to a recognition of their lack of knowledge about certain health-related concepts or conditions. As one asker expressed in his/her question "[…] *Dr. suggested MCT diet but I know nothing about it. Does anyone have specific information?*" Recognition of the knowledge gap might be a necessary condition for people to ask questions, but it is not the only reason. Other cognitive, social, or emotional motivations could work together to encourage users to ask questions on a Q&A site. The analysis of the questioning messages identified the following motivations. The number in the parentheses is the number of incidents.

4.2.1 Cognitive motivations

Recognizing the lack of knowledge on a particular subject does not necessarily motivate people to ask questions on a social Q&A site. In the study, we identified two cognitive motivations that might drive people to such collaborative platforms. One was dissatisfaction with the amount or extent of information from medical professionals and the other was inability to complete an assignment.

Askers felt that they did not get enough information from healthcare professionals (5). Some askers reported that MDs or nurses refused to discuss with them issues of interest, had not told them much, could not seem to find what it was, or simply shrugged their questions off. Here is an example:

> […] About open heart bypass surgery. How much is the post [operation] pain&suffering? How long till all stitches healed? What about removal of wires which they use to reset the ribcage in place? What are risks of this procedure? MDs refuse to discuss these issues with me. I insist to know the reality of what might happen.

For assignments (1 explicitly stated). Users also use Yahoo Q&A to seek help with assignments. One asker stated explicitly this purpose in the question that he/she posted:

> Most of the information on the EKG represents electrical activity of_____of the myocardium? Can anyone help me with my assignment? I can't find the answers in my book...I know it's there but I can't seem to find it!

Several other questions were highly likely to be assignments from classes, but they were not counted here due to the inferences involved.

4.2.2 Social motivations

Some askers cannot go to doctors because of a lack of social support, particularly insurance and income (2). An example is:

> What is the cause of High Blood Pressure? I have never had High blood pressure before.I developed Diabetes 7 years [ago], and have had no insurance to get to a Doctor.

4.2.3 Emotional motivations

Three kinds of emotions were identified to motivate people to ask questions: suspicions concerning the current information that they have, concern about family and friends, and uneasy and disturbing feelings about conditions of themselves or of people who they care about.

Askers reported suspicions about doctors' diagnoses or information heard from other people (9). Some askers expressed doubts concerning the judgment of their doctors. They were worried about whether the doctor's diagnosis was right or asked for second opinion about the diagnosis. An example is:

> There is a hyper density 2mm in my liver. Can anybody tell me if it's something to worry about? My doctor said it's nothing but I need to be sure and a second opinion never hurts.

Sometimes, they tried to validate information from other people, for example:

> […] I've heard that if someone does the same exercises everyday instead of doing 1 type of exercise 1 day the second type the 2nd day and so on his muscles will stop gaining mass ? is this true?

Askers cared about family and friends (25 explicitly stated). In Yahoo Q&A, people ask questions both for themselves and for others they care about or care for. Among the questioning messages, 25 explicitly identified themselves as asking for someone other than themselves, including for sisters/brothers, mothers/fathers, husbands, daughters/sons, grandfathers/grandmothers, partners, nieces, and friends). An example is:

> Can a child of 11years have acidity, i.e., burning sensation in chest? My brother has this problem. Sometimes he feels burning sensation after he takes his meal.

One asker asked a question for his/her dog.

Askers felt disturbed (29). Research suggested that emotions have a determining effect on people's information behavior [29]. Askers who came to the Yahoo Q&A site for health-related questions were mostly driven by negative feelings. Some of them explicitly expressed such feelings in the questions. The intensity of the feelings varied on a wide range of spectrum. Some askers were curious and wondering about something; some felt embarrassed, troublesome, nervous, worried, upset, and anxious; while others felt miserable, desperate, going crazy, freaking out, and scared to death. An example is:

> I am 17 and recently I feel pain in my lower abdomen and checked with my doctor last week and he told me not to worry because there were no lumps, but now for the past 2 days I have been feeling pain in my lower back body with experience please help. I am going crazy and extremely worried with this.

4.3 When to ask questions

Health information needs of individuals occur at various stages of health and illness and the nature and extent of the needs maybe different at different stages [6]. The analysis of the questions identified 8 stages at which people asked health-related questions on the Yahoo Q&A site.

Stage 1: when healthy. At this stage, people ask questions mostly for preventative purpose, for health promotion, or for consequences of an unhealthy life style, such as smoking. Here is an example:

> I've smoked from age 19 - 23, but I've quit now, what are my chances of getting lung cancer?

Stage 2: when asker thinks he/she might be ill. At this stage, people had some symptoms of health problems. They were speculating about what the problems could be and they had not received any medical tests.

> Last week I had heart strain or pain then my heart pounded maybe irregular. I had to take a couple deep breaths then I felt light headed. This week under my left rib cage it feels like my stomach is constantly digesting. Should I be concerned or could it be indigestion. I am 29 years old. I haven't eaten any differently.

Stage 3: right before getting a medical test or checkup. At this stage, askers asked questions concerning various aspects of a medical test, such as what they need to do to prepare for the test, what is looked for in a test, and how do they pass a test. An example is:

> How long do you fast before a cholesterol test? It's at 9am... I just don't eat in the morning, right? […] [They] never said HOW LONG to fast for. 10 or 11 tonight should be fine, right?

Stage 4: when diagnosed or self-diagnosed as ill. At this stage, the askers (or the people they care for) have been diagnosed or self-diagnosed as ill. They asked about what they could do, how serious is the situation, second opinions about their diagnoses, possible treatments, or whether the disease was genetic. For example, one asker asked:

> I have recently been diagnosed with cardiomyopathy (hypertrophic) and am wondering if others have experienced this, and what medical interventions have proved useful in treating the problems associated with this.

Stage 5: before a treatment, surgery, or taking certain medications. At this stage, askers or the people they cared about are going to receive surgeries or treatments. Their questions often included whether the surgery would cure the disease, how much is the pain and suffering, what are the risks of a particular procedure, and how long does it take to recover. For example,

> I will have surgery on the 4th. of January. […] I would like to know if anybody has had surgery on their kidney for the removal of cancer? The Drs. are very up and up on my recovery, but I am wondering if it will take a lot of time to recover? […]

At this stage, some askers also sought good wishes and prayers from their peers.

Stage 6: when receiving or taking treatments, medications, or exercise routines. At this stage, askers wanted to know whether the treatment, medication, or exercise routine would be effective for them or not. They also looked for suggestions on coping with effects of a treatment, such as hair loss during a chemotherapy or side effects of medications. An example is:

> My father was diagnosed with cancer he is going through a very strong form of chemo. This morning he went to his normal daily doctor visit to have his blood cell count checked. The doctor said his white cells are very low and if they drop below 1000 he will put him in the hospital, i do not know what his exact count was but it sounds like it could be very low like below 2000. Now my question is there anything that he could take to help bring his count up and maintain it (I.E. vitamin, fruit,vegetable, over the counter drug). Any help you could give me would be helpful I live 3 hours away, dad doesn't tell mom a whole lot and mom's upset.thank you.

Stage 7: after surgery. At this stage, askers asked questions about the precautionary steps for recovery, the possibility of the reoccurrence of a disease, causes for infections after a surgery, possible complications, and what to watch for. An example is:

> What are the precautionary body movements after an angio seal surgery? Two days ago, my Dad's doctors found that a couple of his arteries were blocked. They performed a Stent on his right femoral artery. What are the precautionary steps my Dad need to take regarding movements like working around the house now, and in the near future? Thank you all so much.

Stage 8: when chronically ill/living with illness. At this stage, the asker or the person he/she cared about lives with a particular disease, such as high blood pressure or heart disease, for a long time. The questions asked often included how they can cope with or manage the situation, whether there are any alternative treatment options, whether it could be cured, and what they can do to treat the disease without medications. An example is,

> I genetically have high blood pressure, how can I lower it without medication? I was diagnosed with hypertension at age 10, doctors said it was inherited from my father. Now I am 24, I don't take medication because it makes me even worse. IS THERE AN ORGANIC WAY TO LOWER MY B.P. OTHER THAN MEDS?

People with chronic disease, particularly heart disease, were also concerned about risks of certain activities, such as losing weight, carrying baby, or traveling.

4.4 Askers' cognitive representations of their problem spaces

Questions and the ways in which they are expressed not only represent people's information needs, but also their cognitive understanding of the problem space. This section analyzes askers' cognitive representations of their problems in relation to four aspects: types of information requested (or goals of questioning), types of information that they provided in the questions, difficulties in phrasing the questions, and expectations for answers of the questions.

4.4.1 Types of information that askers requested (or goals of questioning)

Askers needed various types of information to fulfill their information needs. The following is an overview of major types of information that askers looked for on the Yahoo Q&A site. The number in the parentheses is the number of instances.

- Information about a particular disorder/disease
  - Symptoms (13);
  - Causes (25);
  - Diagnoses (33);
  - Treatments (30);
  - Prognoses (4);
  - Various attributes (23);
  - General information (5)
- Information about drugs or supplements (23)
- Information about lifestyle, including diet and exercise (10)
- People with similar conditions or people they can talk to (23)
- Others
  - Information sources (12);
  - Information about medical professions (6);
  - How to cope or help others cope with bad life situations (5);
  - Interpretations of readings from medical tests (4);
  - Health insurance and policies (2)

The majority of the sample questions were asking about a specific aspect of a particular condition, such as symptoms, causes, diagnoses, treatments, and prognoses. Questions asking about different aspects often follow certain patterns. Due to the space limit, the results will not be reported here. Some askers were concerned about other attributes of a condition, such as whether it is genetic, whether it is contagious, curable, how serious it could be, how fast does it spread, whether it will recur, and age of getting certain disease. A few askers also requested general information about a disease, for example, "*What is kavachi disease?*"

Questions concerning drugs or supplements mainly asked for what drugs to take, whether a particular drug works, how much to take, and when to take. Some askers also requested information associated with drug use, such as interference between different drugs, side effects, safety issues, effects of stopping using a drug, and the storage of a particular drug. An example is:

> What are the best supplements out there for building muscle and losing fat? Right now I am taking Creatine, Arginine, Glutamine, and Protein, and a product called Novadex for suppositive testosterone increase....whats the best stuff out there?

Questions in the lifestyle (diet and exercise) category often asked for recommendations on diets or exercise routines to help lose weight, build muscle, stay healthy, or recover from a disease or a surgery. An example is:

> I wonder if its healthy to maintain a diet of no more than 1000calories/day and to workout everyday pushing myself to the limit....

Yahoo Q&A is a web 2.0 application that builds on user-generated content and the interaction between users. Therefore, it is natural for askers to use it to look for people who have undergone similar conditions for information on treatments and medications or to seek successful stories and emotional support. An example is:

> I recently had a heart virus that has left me with a low ejection fraction of 15-20%. Has anyone had this same problem, and does the ejection fraction increase with time? I discuss specifics with my cardiologist, but I wanted some real world experience from someone

who deals with the same issue. Thanks! I believe it was myocarditis, and please tell me how old you are, or the person is that you are explaining. That way I can compare to myself.I was in the hospital for 26 days, and inpatient P.T. for another week. I have a defibrillator, and am taking Enalapril and Metoprolol for my heart (a beta blocker and an ace inhibitor).

Some askers came to Yahoo Q&A for recommendations of information sources, such as a reliable website to buy medication online, animated diagrams of heart attack, or a best yoga DVD for pregnant women. Some asked questions about the medical specialty, such as "*What is an electrophysiologist?*", or what they need to do to become medical personnel. A few askers asked for help with interpreting readings from medical tests or measurements, particularly blood pressure and BMI, for example, "*How bad is my blood pressure 148 over 114?*' Several askers asked questions concerning how to cope or help others cope with their life conditions and questions about health insurance.

4.4.2 Types of information that askers provided in the questions

The information that askers provided in the questions is not only a representation of their understanding of the current conditions, but also is expected to effectively communicate their intentions to peers. Table 1 shows types of information that askers provided in their questions. The frequency is the number of messages that contain each type of information.

Table 1. Types of information that askers provided in the questions

	User provided information	Frequency
Demo-graphic info.	Age	33
	Gender	4
	Ethnicity	4
	Height	7
	Weight	10
	Geographic location	5
	Professions	6
	Family composition	1
Medical info.	Body parts (e.g., heart, knees)	20
	Conditions/diseases	32
	Stage (cancer)	6
	Symptoms	33
	Medical tests	9
	Treatments	12
	Time for surgery or treatments	2
	Lifestyle (diet, exercise)	11
	Drugs currently using	16
	Cholesterol	3
	Blood pressure	4
	Personal medical history	19
	Family medical history	4
	Health insurance	1
	Time in hospital	1
	Picture of a bone condition	1

As shown in the table, askers provided mainly two types of information, demographic and medical information, to contextualize their requests. Major demographic information included age, gender, ethnicity, height, weight (it should be noted

that the count for gender only included messages that explicitly provided this information; messages that implicitly indicated the information were not counted). Geographic information was provided when the requested information was location dependent, such as the salary of a specialist and insurance policies. A few askers mentioned their professions, such as trainee nurse, high school student, and singer, to better indicate the purpose of the questions.

Many different types of medical information were provided by askers in their questions. Some explicitly pointed out specific body parts that were of concern, such as heart, knees, and spleen. Some listed names of the conditions or diseases that have been diagnosed or self-diagnosed, such as hypertension, leukemia, and colon cancer. Among askers who asked questions about cancer, six indicated the stage of the cancer. In many cases, askers provided information about their symptoms, for example,

> I have been having these headaches accompanied with this constant pain in my heart with occasional sharp pains. I have been feeling short of breath the entire week and i seem to only be feeling worse not better and and i get dizzy very easily...

Some askers provided information about medical tests that they had taken, such as echocardiogram, blood test, x-ray, and CT scan. Some mentioned treatments, such as chemotherapy, bone marrow transplant, and Lasik surgery, that they were going through, had experienced, or would receive in the near future. A dozen askers also listed information about their lifestyle, including diet, exercise routines, drink, and smoke.

To help others interpret their problems, some askers listed drugs or supplements that they are currently using or provided readings from medical tests, mostly blood pressure and cholesterol. Some went further to provide information about their personal medical history or medical history that runs through their families. A couple of them also mentioned health insurance (e.g., Medicare) and time in hospital. It is worth nothing that, to facilitate others to make judgment, an asker posted a link to a picture illustrating his/her bone condition.

4.4.3 Difficulties in expressing health information needs

Difficulties involved in phrasing questions indicate deficiencies in askers' cognitive representations of their problems and their ability to articulate the problems. Two difficulties directly related to effective IR were observed in the questions. One is spelling. Misspellings, including misspellings, run-together phrases, word/phrase split inappropriately, appeared with high frequency in the questions. Most askers did not realize or were not bother to recognize misspellings, except that one stated in his/her question:

> Ways to lose weight with a heart disease? I have an irregular heartbeat it beats too fast. I think it is called "Tacacardia". Don't hold me to the spelling. I don't know if it is the right ways to spell it or not but anyway. […]

The other problem is the lack of proper medical terms to describe the condition of concern, as one asker stated in his/her question:

> Genetic problem? What's it called? Apparently I've inherited a condition which gives me more of a chance of getting lung and liver cancer if I smoke and drink. My question is, what is the condition called? I want to know more about its history. […]

Another observation that could possibly affect IR is the frequent use of acronyms, such as AML, EKG, ECHO, CK, CKMB, ASD, and IV, by askers in their questions.

4.4.4 Expectations for the answers

Some askers' representations of their problem spaces also included expectations for answers. Some cared about the quality, credibility, reliability, and specificity of the answers. For example, one asker indicated that he/she wishes to see the source of information.

> Is Down's Syndrome hereditary, and what are the increased risks of it if having a child after 30? I am aware that 1 in 733 live births is the average statistic. Please include your source of information. Thanks!

The other asked for answers from professionals only:

> What about the side effects for the hoodia pills for losing weight? (for professionals only).

Some askers wanted the answers to be genuine, candid, kind, and/or serious. For example,

> I want to hear from anyone who had or knows about open heart bypass surgery. Kindly reply only.

5. DISCUSSION

Searching for health information online has become more and more popular. In recent years, a large number of web-based consumer health information sources and services have been created. However, just making information available does not ensure the use of it. The design of effective consumer health information systems requires an understanding of the context of consumer health information searching [2]. This study is an initial effort to explore factors that contextualize consumer health information searching. Specifically, we investigated some basic linguistic features of consumer health questions as expressed in natural language, askers' motivations for asking the questions, time when they asked the questions, and their cognitive representations of the problem spaces.

In the study, a question was defined as a manifestation of a person's information need arising from an intention to fill his/her knowledge gaps. A sample of 276 questioning messages were randomly selected from the Yahoo Q&A site. Over 95% of the selected messages contained questions. The median number of terms in the messages that contain questions was 33. Compared to short queries that users submitted to general web search engines, most messages on Yahoo Q&A provided more contextual information, revealing not only different aspects of consumer health information needs and the ways in which the needs were represented, but also the goals, motivations, and emotions behind the needs. Therefore, the questions are a rich source for exploring consumer health needs to inform the design of IR and consumer health information systems.

For messages containing questions, the average number of questions per message (density of questions) was 1.25. Over 25% of the messages encompassed more than one question. In most interaction log analysis studies, the unit of analysis was a query. Although it has been pointed out that some users input multiple search queries in a health information search session [38, 43],

little is known about why users modify their queries over time, whether it is a rephrase of the same question resulted from unsatisfying results or a natural switch/progression to search for another pre-conceived question. The density of the questions indicated that some users have multiple different but related questions when they come to search for health information. IR systems could capitalize on this pattern of user needs and provide mechanisms or functions to identify and support this type of request. In future research, analysis should be performed to identify the characteristics of the transition from one question to the next in one questioning message.

Motivations for seeking information influence people's judgment of the relevance of information. Past research, mainly through surveys or interviews, identified specific reasons for people to search for health information. It has been reported, for example, that people seek health information because they have been diagnosed with a medical condition, they have been prescribed a new medication or course of treatment, or they are preparing for doctors' appointments [10]. Using a factor analysis, Rice [31] identified three categories of motivations for health information searching: problematic health conditions, lack of access to physicians, and care for significant others. In this study, we found that there was a common cognitive motivation for people to use the Q&A site for health-related information: they have recognized a knowledge gap of their own and strove to find information to fill the gap. Meanwhile, other cognitive, emotional, and social reasons work together to foster the questioning behavior. At the cognitive level, some askers asked questions because they felt that they did not get enough information on the issue of concern from doctors or they were unable to complete an assignment. At the social level, askers turned to the site because they lacked income and health insurance. At the emotional level, askers sought health information because of their suspicion about information that they have received, concern about family and friends, or disturbance resulted from negative feelings, such as worried, anxious, and desperate, about their health conditions. The emotions that askers demonstrated in the questions were mostly negative but with various degrees of intensity. Designers for future health IR systems could utilize the intensity of emotions as an element to make the systems more emotion-responsive, therefore, more personalized.

Time matters in health information seeking. At different stages of health and illness, consumers have different levels of needs and require different types of information [5]. The study identified eight stages on which general users could search for health information: when healthy, when they think they might be ill, before getting a medical test, when diagnosed as ill, before treatment, when receiving the treatment, after treatment/surgery, and when being chronically ill. At different stages, people had different foci of concern and asked for different types of information. For example, before receiving a medical test, askers would ask for how to prepare for the test and what they will go through during the test. After surgery, askers asked for suggestions about precautionary steps for recovery and information about possible implications. Information of the stage which a patient is experiencing during his/her illness trajectory could shed light on the kinds of information that might be more relevant to him/her. For system design, temporal organization of information materials or of user actions could be experimented

with to see whether they can provide more effective service to consumers.

People's mental representations of a problem space have direct impact on the ways in which they search for information and the effectiveness of search results. Past research has shown that laypeople used different sources and query terms, and had different sequences of search goals in a search session, from people with medical training [3, 43]. In this study, we explored askers' mental representations of their problem spaces in relation to four aspects: types of information that askers requested (or the goals of questioning), types of information that askers provided in the questions, their difficulties in expressing their needs, and expectations for the answers.

Past research has identified that, for specific diseases, consumers often seek information for diagnoses, treatments, symptoms, medications, and prognoses [14, 41]. For general health and wellness concerns, they seek information such as nutrition, disease prevention, and fitness [40]. Consistent with previous findings, on Yahoo Q&A, askers mainly sought information about diagnoses, treatments, causes, and symptoms of a particular disorder/disease. When questioning about drugs or supplements, askers mainly wanted to know what drugs to take and the effectiveness of the drug for a particular purpose. In addition, they were interested in a range of attributes of a disease, such as whether it is genetic and contagious and whether it recurs.

It is worth noting that many askers valued personal experience and specifically sought for people whom they can talk to. This finding suggests that a social network approach that connects consumers with peers with similar conditions could have a great potential to keep them informed and to encourage them to be more actively engaged in their own healthcare. Another finding was that some askers needed help with interpreting test readings, such as blood pressure and cholesterol, which suggests that applications could be developed to facilitate people to read common medical test results. On Yahoo Q&A, askers also sought other types of information, including suggestions or advices for lifestyle, credible health information resources, emotional support for people in difficult life situations, and information about medical professions and health policies.

In comparison to the types of information requested by consumers, little attention has been paid to what information that consumers provided in order to elicit relevant answers. Information provided by consumers reflects their current status of knowledge of their conditions. In the study, askers provided mainly two types of information, demographic (e.g., age, gender, and weight) and medical information, in their questions. When providing medical information, many askers described symptoms that they were experiencing, identified specific body parts that they felt uncomfortable with, or listed specific diseases that they had been diagnosed or self-diagnosed, name of the drugs that they were using, and treatments that they went through. The information that askers provided in questions reflects consumers' ways of representing and expressing their requests and could be used to support customization or personalization of health information searching. In future research, more efforts should be dedicated to design IR interfaces or information architecture that have both the structure and flexibility to accommodate to users' representations of their problem spaces. It is worth noting that one asker provided a picture to help illustrate his/her current condition. This phenomenon suggests that IR systems or community-based health forums, such as social Q&A sites, should also provide mechanisms to allow people to express their needs through different media formats.

Health-related search queries submitted by general consumers, in many cases, were misspelled or did not match with standard medical terminologies [27, 46]. This study confirmed the observations. Askers expressed, in the questions, their difficulty with spellings and with coming up with proper terms to describe a condition. Such knowledge deficiencies could prevent users from getting optimal results. In health IR systems, both conceptual and lexical tools, such as spelling check, term recommendation, auto-completion of medical terms, or auto-expansion of medical acronyms, should be provided to help users articulate their needs. The same functions should also be implemented on the question form on social Q&A sites to facilitate the asking of better questions.

Sometimes, askers' mental representations of their problems also encompass expectations for the answers. Askers wanted answers with good quality, specific, and reliable. Due to the social nature of the platform, askers also cared about the attitudes of people who provided answers. They wished to receive genuine, candid, kind, and serious answers from their peers. In future research, more research could be conducted to investigate how consumers judge the relevance of health information in general and health information generated by peer users on collaborative platforms in particular. In some cases, askers preferred answers from medical professionals. A possible design implication for question forms on social Q&A sites is to provide a checkbox for askers to indicate their preference.

It should be pointed out that, in the study, the identification of contextual elements were based on user self-generated descriptions. Not all the users were equally likely to provide detailed contextual information of their problems. For example, many askers identified themselves as asking questions on behalf of others, but some might not explicitly identify this purpose in the questions. The study is qualitative in nature. The number of instances was used only to tally questions that explicitly provided a particular type of information and could not be used to make statistical inferences. In future studies, semi-structured interview methods could be employed to examine these contextual factors to gain a more in-depth understanding of consumer health information searching. Future studies could also examine whether different types or contexts of questions result in different patterns of responses. Furthermore, it will be worthwhile to have users perform real-life tasks to observe how they make use of contextual factors in actual search sessions.

6. CONCLUSION
In summary, this study is an initial effort to contextualize consumer health information searching. The results suggested that Q&A sites are a fruitful resource for identifying authentic consumer health information needs and the ways in which the needs were expressed. Consumers' questioning behavior was motivated by the recognition of a knowledge gap, disturbing feelings, or the lack of social resources. Time is an important factor affecting health information seeking. People tended to have different concerns at different stages of their health and illness. Overall, on Yahoo Q&A, askers requested many different types of

information. They were particularly interested in finding people with similar experiences. In order to get relevant answers, they provided various types of demographic and medical information to communicate their problems to others. Askers had both conceptual and lexical difficulties in formulating questions. Sometimes, they had a high expectation for the quality of the answers. Future consumer health IR systems should be able to accommodate to the ways in which consumers represent their problems and facilitate their expressions of the needs in a more effective way.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Adams, S.A. 2010. Blog-based applications and health information: two case studies that illustrate important questions for consumer health informatics (CHI) research. *International Journal of Medical Informatics, 79*(6), 89-96.

[2] Belkin, N., J., Oddy, R. N., and Brooks, H. M. 1982. Ask for information retrieval: Part I. background and theory. *Journal of Documentation, 38*, 61-71.

[3] Bhavnani, S. K. 2002. Domain-specific search strategies for the effective retrieval of healthcare and shopping information. In the *Proceedings of CHI 2002*, 610-611.

[4] Bradley, J. 1993. Methodological issues and practices in qualitative research. *Library Quarterly, 63*(4), 431-449.

[5] Buckland, S. 1994. Unmet needs for health information: a literature review. *Health Libraries Review, 11*, 82-95.

[6] Cool, C., and Spink, A. 2002). Issues of context in information retrieval (IR): an introduction to the special issue. *Information Processing and Management, 38*, 605-611.

[7] Culver, J.D., Gerr, F., and Frumkin, H. 1997. Medical information on the Internet: A study of an electronic bulletin board. *Journal of General Internal Medicine, 12*, 466-470.

[8] Downie, J.S., and Cunningham, S.J. 2002. Toward a theory of music information retrieval queries: system design implications. In *Proceedings of the 3rd International Conference on Music Information Retrieval*, 299-300.

[9] Eysenbach, G., Powell, J., Englesakis, M., Rizo, C., and Stern, A. 2004. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. *BMJ, 328*(7449): 1166-1171.

[10] Fox, S., and Jones, S. 2009. *The Social Life of Health Information*. Pew Internet & American Life project. Retrieved March 3, 2010, from http://www.pewinternet.org/~/media//Files/Reports/2009/PIP_Health_2009.pdf

[11] Frost, J. H., and Massagli, M. P. 2008. Social use of personal health information within PatientsLikeMe, an online patient community: What can happen when patients have access to one another's data. *Journal of Medical Internet Research, 10*(3): e15.

[12] Ginossar, T. 2008. Online participation: A content analysis of differences in utilization of two online cancer communities by men and women, patients and family members. *Health Communication, 23*, 1-12.

[13] Glaser, B.G., and Strauss, A.L. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine.

[14] Graesser, A.C., Lang, K., and Horgan, D. 1998. A taxonomy for question generation. *Questioning Exchange, 2*, 3-15.

[15] Graham, L., Tse, T., and Keselman, A. 2006. Exploring user navigation during online health information seeking. *AMIA 2006 Symposium Proceedings*, 299-303.

[16] Ingwersen, P., and Jarvelin, K. 2005. *The Turn: Integration of Information Seeking and Retrieval in Context.* Springer.

[17] Jasra, M. March 20, 2008. Yahoo Answers down but dominating Q&A websites. Accessed on May 3, 2010 at: http://www.webanalyticsworld.net/2008/03/yahoo-answers-dominates-q-websites.html.

[18] Kelly, D. 2006. Measuring online information-seeking context, part 1. Background and method. *Journal of the American Society for Information Science & Technology, 57*(13), 1729-1739.

[19] Keselman, A., Smith, G.A., Divita, G., Kim, H., Browne, A.C., Leroy, G., and Zeng-Treitler, Q. 2008. Consumer health concepts that do not map to the UMLS: Where do they fit? *Journal of American Informatics Association, 15*, 496-505.

[20] Kim, D., and Chang, H. 2007. Key functional characteristics in designing and operating health information websites for user satisfaction: An application of the extended technology acceptance model. *International Journal of Medical Informatics, 76*, 790-800.

[21] Klemm, P., Reppert, K., and Visich, L. 1998. A nontraditional cancer support group: the Internet. *Computer in Nursing, 16*(1), 31-26.

[22] Lin, J., and Demner-Fushman, D. 2005. Representation of information needs and the elements of context: a case study in the domain of clinical medicine. *Proceedings of the SIGIR 2005 Workshop on Information Retrieval in Context (IRiX 2005)*, 51-53.

[23] Liu, Y., Bian, J., and Agichtein, E. 2008. Predicting information seeker satisfaction in community question answering. In *Proceedings of SIGIR'08.*

[24] Luo, G., Tang, C., Yang, H., and Wei, X. 2008. MedSearch : A specialized search engine for medical information retrieval. In *Proceedings of CIKM'08*, 143-152.

[25] Maloney-Krichmar, D., and Preece, J. 2002. A multilevel analysis of sociability, usability, and community dynamics in a online health community. *ACM Transactions on Computer-Human Interaction, 12*(2), 201-232.

[26] McCray, A.T., Dorfman, E., Ripple, A., Id, N.C., Jha, M., Katz, D., Loane, R.F., and Tse, T. 2000. Usability issues in

developing a web-based consumer health site. In *Proceedings of AMIA 2000*, 556-560.

[27] McCray, and Tse, T. 2003. Understanding search failures in consumer health information systems. In *Proceedings of AMIA Symposium 2003*, 430-434.

[28] McCray, A.T., Loane, R.F., Browne, A.C., and Bangalore, A.K. 1999. Terminology issues in user access to web-based medical information. In *Proceedings of AMIA Symposium 1999*, 107-111.

[29] Nahl, D. 2007. Introduction. In Diane, N., & Dania, B. (Eds.), Information and Emotion: *The Emergent Affective Paradigm in Information Behavior Research and Theory* (pp. xvii-xxix). Information Today, Inc.

[30] Preece, J. 1998. Empathic communities: researching out across the web. *Interactions, March + April*, 32-43.

[31] Rice, R.E. 2005. Influences, usage, and outcomes of Internet health information searching: Multivariate results from the Pew survey. *International Journal of Medical Informatics, 75*, 8-28.

[32] Rose, D.E., & Levinson, D. 2004. Understanding user goals in web search. In *Proceedings of WWW 2004*, 13-19.

[33] Schoch, N. A., and White, M. D. 1997. A study of the communication patterns of participants in consumer health electronic discussion groups. In *Proceedings of the Sixtieth Annual Meeting of the American Society for Information Science*.

[34] Shah, C., Oh, S., and Oh, J.S. 2009. Research agenda for social Q&A. *Library & Information Science Research, 31*, 205-209.

[35] Slaughter L.A, Soergel D, and Rindflesch T. C. 2006. Semantic representation of consumer questions and physician answers. *International Journal of Medical Informatics, 75*(7), 513-529.

[36] Smith, C.A., Stavri, P.Z., and Chapman, W.W. 2002. In their own words? A terminological analysis of E-mail to a cancer information service. In *Proceedings of AMIA Symposium 2002*, 697-701.

[37] Smith, C. A., and Wicks, P. J. 2008. PatientsLikeMe: Consumer health vocabulary as a folksonomy. In *Proceedings of AMIA Symposium 2008*, 682-686.

[38] Spink, A., Yang, Y., Jansen, J., Nykanen, P., Lorence, D. P., Ozmutlu, S., and Ozmutlu, H.C. 2004. A study of medical and health queries to web search engines. *Health Information and Libraries Journal, 21*, 44-51.

[39] Till, J.E. 1995. Discussion groups on the Internet: Journaling. *Canadian Journal of Oncology, 5*(3), 379-380.

[40] Warner, D., and Procaccino, J.D. 2004. Toward wellness: Women seeking health information. *Journal of American Society for Information Science and Technology, 55*(8), 709-730.

[41] White, M.D. 2000. Questioning behavior on a consumer health electronic list. *Library Quarterly, 70*(3), 302-334.

[42] White, M.D., and Iivonen, M. 2001. Questions as a factor in web search strategy. *Information Processing and Management, 37*, 721-740.

[43] White, R.W., Dumais, S., and Teevan, J. 2008. How medical expertise influences web search interaction. In *Proceedings of SIGIR'08*, 791-792.

[44] Wildemuth, B.M., Bliek, R., Friedman, C.P., and Miya, T.S. 1994. Information-seeking behaviors of medical students: a classification of questions asked of librarians and physicians. *Bulletin of Medical Library Association, 82*(3), 295-304.

[45] Zeng, Q.T., Kogan, S., Ash, N., Greenes, R.A., and Boxwala, A.A. 2002. Characteristics of consumer terminology for health information retrieval. *Methods of Information in Medicine, 41*, 289-298.

[46] Zeng, Q.T., Kogan, S., Plovnick, R.M., Crowell, J., Lacroix, E.M., and Greenes, R.A. 2004. Positive attitudes and failed queries: an exploration of the conundrums of health information retrieval. *International Journal of Medical Informatics, 73*(1), 45-55.

[47] Zeng, Q.T., and Tse, T. 2005. Exploring and developing consumer health vocabularies. *Journal of the American Medical Informatics Association, 13*(1), 24-29.

[48] Zeng, Q.T., Crowell, J., Plovnick, R.M., Kim, E., Ngo, L., and Dibble, E. 2006. Assisting consumer health information retrieval with query recommendations. *Journal of American Medical Informatics Association, 13*, 80-90.