Designing Consumer Health Information Systems: What Do User-Generated Questions Tell Us?

Yan Zhang[1] and Wai-Tat Fu[2]

[1] School of Information, University of Texas at Austin, Austin, TX, 78701
[2] Department of Computer Science, University of Illinois, Urbana, IL 61801
yanz@ischool.utexas.edu, wfu@illinois.edu

Abstract. Searching for health information has become a prevalent activity on the web. The information found online has a significant impact on people's decisions on whether to seek medical care and what treatments to undergo. However, existing studies consistently suggest that general consumers have various difficulties in formulating search queries using existing search engines and the queries were often not effective in retrieving personal- and situational-relevant information. Understanding users' information needs is a gateway to designing effective information retrieval (IR) systems. In this study, we examined the types of information requested by users, the characteristics of consumers' expressions of their information needs, and their expectations for results by analyzing the questions that general users posted on Yahoo! Answers, a popular social Q&A site. Based on the results, we proposed design recommendations for facilitating users' ability to articulate their health information needs and recommendations for the presentation of information in health-related IR systems.

Keywords: Consumer health informatics, information retrieval, social Q&A, health information searching.

1 Introduction

Recent Pew studies show that more than 80% of the internet users in the U.S. seek health information online and this activity has become the third most popular activity across all age groups, after Email and search engine use [30]. Furthermore, the information found online significantly impacts people's decisions about their own health or the health of someone in their care [9].

In response to the high demand for health information and the significant impact of this information, a lot of efforts have been made, in the past decade, to make reliable and trustworthy health information sources available online. However, little attention has been paid to the usability of and the user's experience with such sources. Subsequently, there is still a significant lack of understanding of the factors that impact users' adoption of health information systems [14, 16]. Meanwhile, consumers consistently report difficulties in finding personally relevant information using existing sources [29]; and they frequently feel confused and frustrated by their searching experiences [1].

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 536–545, 2011.

In order to design not only useful but usable health information systems, it is necessary to understand how consumers conceptualize and articulate their health needs and how they use information systems to find relevant information [3, 14]. This study is an effort to improve the understanding of consumer health information needs. Specifically, we explore types of information requested by consumers, the ways in which they express their needs, and their expectations of the answers by analyzing health-related questions that general users posted on a social Q&A platform, Yahoo! Answers. The implications of the results to the design of health information systems are discussed as well.

2 Related Literature

The current research on consumer health information needs mainly takes two approaches. The first approach is from a clinical perspective using structured surveys to identify types of health information such as information about cure, spread of disease, and treatment, needed by patients or their caregivers. For example, several studies employed a structured information needs questionnaire (INQ) to examine the information needs of cancer patients or their care givers. The INQ includes nine items: information about cure, spread of disease, treatments, side effects of treatment, genetic risk, self care, sexual attractiveness, impact on social life, and impact on family [2, 7; 19]. Harrison, et al. [12] used the Toronto Informational Needs Questionnaire (TINQ) to survey women with breast cancer concerning their demand for specific information related to their disease, their treatment, and relevant investigative tests.

The second approach is analyzing transaction logs, particularly queries. In addition to identifying subject areas that people search for [22], query analyses can shed light on vocabularies that users use to search for health-related information and the difficulties that they have with the searching. These studies found that people submit short queries, with an average number of 2.2 to 3.3 terms per query, to search engines for health-related information [22, 27]. Short queries often lead to insufficient representation of health problems in search and unsatisfying results. At the same time, misspellings and abbreviations are common in queries, which tend to cause search failures [4, 20].

Query analyses also reveal conceptual difficulties that consumers have in searching for health information. First, their vocabularies do not match medical terminologies, such as those found in UMLA and used in the content or indexes of health information websites [15]. Second, they sometimes could not find proper terms to describe their actual needs. And third, they have different mental representations of their conditions from medical professionals and are likely to describe the conditions using simpler or more concrete terms [29].

Apparently, these two approaches provide valuable insight into what information is needed by general people and the difficulties that people have in retrieving health-related information. However, they fall short of revealing important elements of people's information search tasks that result from conscious information needs, such as how people conceptualize their problems, what their intentions are, and how they convey their conceptualizations to information systems [6, 21]. The research on

interactive information retrieval (IIR) consistently suggests that users' perceptions of tasks, such as intended goals or purposes, task complexity, and task familiarity, have a significant impact on their mental models of the system that they are interacting with, their behavior of using the systems, and their experience with the systems [e.g., 24]. Therefore, an enhanced understanding of consumer health information needs, which give rise to information search tasks, will be a rich source for informing the design of better information systems.

In recent years, supported by the fast development of social media technologies, community-based online health forums, such as bulletin boards, blogs, wikis, social networks, and social Q&A sites, became widespread on the web. The emergence of these communities was considered critical in empowering patients to make decisions concerning their own health and promoting patient-centered healthcare [8]. These communities often consist of patients with a particular disease, such as breast cancer, or people who share a particular topic of interest, such as running, patients' caregivers, and sometimes, some health professionals. 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 who share similar experiences [28]. Therefore, the most popular activity on these social platforms is asking health-related questions or answering questions posted by peer users [10, 17]. Unlike in web search engines where users often type in short queries, these platforms allow the users to ask questions in natural language and in full sentences. Therefore, these questions can be a good source for learning a range of factors involved in consumers' health information needs and their related information search tasks, such as background of the askers, the askers' perception of the tasks, the nature and goals of the tasks, and the constraints around the tasks [23, 24].

This study presents a preliminary effort to explore the characteristics of consumer health information needs that have direct impact on information searching by analyzing health-related questions that general consumers posted on Yahoo! Answers, a major social Q&A site. Implications of the results for designing effective consumer health information systems are discussed, as well.

3 Research Method

A dataset including 77,903 questioning messages across 23 health-related categories, such as diabetes, heart disease, cancer, and diet and fitness were crawled from Yahoo! Answers [18]. At the time the data was crawled, this site had about 74% of the market share of U.S. visits among Q&A websites [13]. We randomly selected 276 questioning messages from the dataset and analyzed the messages using the qualitative content analysis method. A content analysis software, QSR, was employed to assist the data analysis. Each questioning message was a unit of analysis.

We first coded the number of questions contained in each message. A question is loosely defined as a request sent by an asker to one or more respondents to solicit knowledge on a certain subject, which the asker lacks but sincerely wants [26]. Then, the open coding method was employed to code the data. At the beginning of the coding, we read the text several times to gain an overview of the overall content. Then codes were derived inductively by closely reviewing themes that appeared in the

messages. Whenever a theme related to health information needs appeared, such as for whom a question was asked, what information is needed, difficulties in expressing their needs, and expectations for answers, 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 [11]. 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 [5]. To improve the validity of the results, a second coder coded 20% of the questions. The inter-coder agreement reached 87.9%.

4 Results

Among the 276 questioning messages, nine were advertisements or messages eliciting comments and prayer from others. These messages were not intent to solicit knowledge that the asker lacked, thus did not meet our definition of questions. They were excluded from the subsequent analysis. In this section, we report three important aspects of consumer health information needs that are directly related to information searching and information system design: types of information needed, characteristics of users' expressions of the information needs, and users' expectations for results.

4.1 Types of Information Needed

Information systems, particular IR systems, are designed to help people find relevant information to solve particular tasks. Therefore, knowledge about types of information requested by users is important for system designers. The analysis of the questioning messages revealed the following major types of information needs. The number in the parentheses is the number of messages.

Information about a particular disorder/disease (123). These questions were asking about a specific aspect of a particular disease, mainly symptoms, causes, diagnoses, treatments, and prognoses. When asking about symptoms, users' questions followed two patterns: what are the symptoms of a disease or condition and whether it is normal for people with a particular disease or condition to have certain symptoms. When asking about causes, users were concerned about what factors or behaviors cause a disease or a condition. An example is: "*What raises people's blood pressure? Please give me list of stuff that raises your blood pressure and how it raises.*"

Questions concerning diagnoses had three different patterns. The first was that askers described certain symptoms, conditions, or behaviors and asked for possible diagnoses. For example, "*I'm really desperate to find out what the lump in the back of my throat is. I also have an inflamed gland on the side of my throat (neck). […] Do you think this sound like throat cancer?*" The second pattern was that askers were concerned about themselves having a particular condition and asked whether a symptom suggests that particular condition. The third pattern was that askers wanted to double check a doctor's diagnoses. For example, "*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.*"

Questions concerning treatments were asking about effective treatments to certain conditions, the cost of a treatment, how long it takes, whether it is hard to get through, and whether there are other options. Askers asking for prognoses often had a diagnosis and wanted to find the prognosis for that condition. For example, "*Hubby has dilated cardiomyopathy, stage 4, what is his prognosis?*"

Some askers knew about diagnoses, but wanted to find out whether the disease or condition would affect their other life plans, such as carrying babies and traveling. An example is: "*What kind of risk [is] involved if I carry baby with heart disease?*" Some askers were asking about attributes of a disease or a condition, such as whether it is genetic, whether it is contagious, whether it is fatal, how serious it could be, how fast it spreads, whether it is rare, whether it will recur, and the age of getting the disease. A few askers also requested general information about a disease, for example, "*What is cancer?*" and "*What is pink eyes?*" A few people asked questions concerning the recovery of a disease, such as how long it will take to recover.

Information about drugs or supplements (23)**.** Questions about drugs or supplements were mainly about the following themes: whether a particular drug is effective, what drugs to take, how much to take and when, whether there are side effects, what the ingredients are, whether it is safe, whether it interferes with other medical tests, whether the way in which a drug had been stored has impact on its effect. An exemplar question asking about the effectiveness of a drug is: "*Does relacore really work? I am exercising and I have been eating pretty well. I was thinking of taking relacore to help me lost weight, has anyone used it? Does it work?*"

Information about lifestyle, mainly diet and exercise (10). Questions concerning lifestyle often asked for recommendations for a healthy diet or exercise routine, given a specific weight, height, age, or health condition or checked with others whether their current exercise routine is reasonable, for example, "*Hi, I am 33 yrs old… My heart rate drops as I started exercising everyday (for about 2 months). I run 3-4 miles, sit-ups and play ping pong EVERYDAY.....I have a blood pressure monitor with which I measure my BP. and it seems to me it dropped down and heart rate dropped down from 72BPM to 55 BPM [...], I am also on Lexapro, (medicine for depression)… I am afraid I am over-exercising? I don't smoke or consume alcohol...*" Some of the questions in this category also asked for more information on a particular diet, such as calorie shifting diet, and a lifestyle to adopt after experiencing a disease. For example, "*What lifestyle to adopt after gastric ulcer disease? [..].*"

People with similar conditions (23). There were a number of askers looking for people with similar conditions on Yahoo! Answers, for general information or advice, success stories, treatment information, and drug information. An example is: "*Has anyone tried ortho-k lenses coupled with eye exercises for permanent vision improvement? How well does it work?*"

Information sources (12). Some askers asked about sources for information, such as quotes to cheer up friends, reliable websites to buy medications, animated diagrams of a heart attack, statistics about the causality of lung cancer and percentage of people smoking worldwide, or a yoga DVD for pregnant women.

Others (20). Askers also asked questions concerning medical professions, for example, "*I have a niece interested in becoming a Physical Therapy Assistant. What is the pay? We are in NC,*" body working mechanisms, for example, "W*hat would happen if the right side of heart pumped faster than the left side?*", interpretations of medical tests readings, such as blood pressure, information about coping with bad life situations and managing stress, and health insurance and policies. An example about insurance and policies is, "*Can I bring mum over to Australia for treatment for lung cancer from UK? Does NHS subsidize any costs (or medicare)? […]*"

It is worth noting that many askers were asking questions not for themselves, but for related others, such as sisters, brothers, parents, grandparents, and friends. Among the 267 messages, 25 explicitly stated that the questions asked were for someone other than themselves.

4.2 Characteristics of Users' Expressions of Health Information Needs

Characteristics of users' expressions of health information needs refer to linguistic features of the questioning messages. Queries or questions are a major means through which users interact with an information system. Therefore, features of users' expressions impact the quality of the results of user-system interaction. In this study, users' expressions of health information needs were analyzed at both the term level and the question level.

At the term level, inappropriate terms, particularly, misspellings (e.g., canser, diagnoised, suppliment, and gentically) and run-together phrases (forcancer, to'prevent'whatever, medication"zimvastatin 10mg", and cardiologygrizzln) were prevalent in the sample questioning messages. Sometimes, the misspellings were due to the lack of auto-correction mechanisms in Yahoo! Answer's question submit form. Sometimes, it was due to the fact that users had difficulties in spelling a medical term. For example, one asker stated that, "*I'm looking for a website on a rare skin disease, it sounds like dariase DARE E AZE? Uncertain of the correct spelling […].*" The second term level linguistic feature that could cause communication difficulties between user-system or user-user was the use of acronyms, such as B.P., AML, EKG, ECHO, CKMB, CK, and NHL. In this study, about 5% of the sample questioning messages contained such acronyms. The third term level difficulty was conceptual, specifically, users could not find terms to describe their conditions. One asker stated: *[…] 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. […]*"

At the question level, in the sample questioning messages, about 75% of the messages contained only one question, 18% contained two questions, and the rest 7% contained 3 to 5 questions. Multiple questions in a questioning message were often asking about different aspects of a disease, drug, surgery or procedure. For example, an asker asked two related questions about a condition: "*Is clogging of the heart arteries reversible? If so how?*" Similarly, another asked "*What is a spleenic lesion and how is it treated?*" In many cases, one question in such messages was asking about whether there was someone with similar experience. For example, "*Who is taking hydrochlorothiazide 12.5 mg and lisnopril 10mg and does it help to lower your blood pressure? What is your blood pressure after taking these two medicines?*"

4.3 Expectations for the Answers

In asking questions, askers also expressed their expectations for the answers. The expectations had three dimensions, one concerning the quality of the information, one concerning the personal or situational relevance of the information, and the other concerning the social attitudes involved in the answers. For the quality of information, some pointed out in the messages that they wanted reliable information, for example, one asker asked, "*Where is a reliable site to buy medication online?*" Some askers indicated that they wanted answers from medical professionals, and some indicated that they wanted to see the source of the information.

As mentioned above, many users of Yahoo! Answers tried to seek information, advice, or personal experience from people with similar conditions. One reason for this phenomenon seems to be that users favor personally and contextually relevant information. For example, one asker tried to elicit personal stories "*Do you have a child with congenital heart disease? [Has he/she] undergone an open heart surgery? […] How's the operation? Please tell something about after the surgery.*"

Users of Yahoo! Answers also cared about the attitudes of the people who answered their questions. Askers wanted the answers to be genuine, candid, kind, and serious. For example, "*What's the best yoga/workout post natal DVD? I'm pregnant but due fairly soon, and what to start a collection of some workout dvds and need some input on the best for a post-natal body. ** Serious answers only … kids DON'T respond saying some dumb remarks and NO ignorant remarks […]*"

5 Discussion and Implications

Understanding the characteristics of consumer health information needs is essential to designing and constructing effective consumer health information systems [3]. In this study, we explored the types of information requested by users, the characteristics of their expressions of information needs, and their expectations for results. It is apparent that consumers need different types of health-related information, with the majority of the request focusing on different aspects of categories including diseases, drug and supplements, lifestyles, people with similar conditions, and information sources. This observation suggests that consumer health information systems should construct effective information architecture to support users' access to these different types of information. The first three categories were also identified in several other studies [e.g., 25, 26] and were made available in many consumer health websites, such as MedlinePlus, however, little attention has been paid to providing effective access to people with similar conditions and other relevant information sources (like websites and DVDs). Future design of consumer health information systems should make an effort to improve access to these two types of information.

In this study, it was also found that users sometimes have several different questions concerning a particular disease, medication, or surgery. This suggests that systems can provide cognitive assistance to help users find relevant information. Such assistance could be implemented by providing access to related topics. The selection of related topics could be based on data mining of user-generated questions.

Consistent with the existing literature [4, 20, 22], this study found that users have various linguistic and conceptual difficulties in expressing their health information needs. Misspellings and run-together phrases were common and these errors sometimes prevented systems or other users from understanding the messages correctly. Most of the misspellings and run-together phrases were due to users' cognitive slips and could be solved by providing a spelling check function in the question submit form. Sometimes, the misspellings were due to the fact that users do not know how to spell a medical term or they do not know what terms to use to describe their conditions. These difficulties suggest that the system should provide a function to assist users in finding appropriate medical terms to describe their thoughts. This help could significantly improve users' experience with health information searching.

Looking for health-related information online is a highly social activity. Many people asked questions for family and friends who they care about. People wanted information, advice, personal stories, and emotional support from those with similar conditions; and people wished the answers to be genuine, candid, and serious. These observations suggest that users want personally and situationally relevant information and they value the information from peers. Most health information systems provide systematic scientific knowledge on consumer health topics, however, due to the static nature, they often fall short in supporting people's highly idiosyncratic health information needs. One way to improve users' experience with health information seeking is to integrate social networks with traditional information retrieval system and leverage information searching using collective wisdom. Future research could investigate how to integrate social systems with traditional IR systems to best augment general users' cognition and achieve more effective health information searching.

This study also found that users are concerned about the quality and credibility of the results. They wanted to see the source for answers and they preferred answers from medical professionals. They also wanted the answers to be sincere and candid. This suggests that a consumer health information system should provide mechanisms to help users judge the quality of information. If a social network component is involved, it will be desirable to have means to prevent, block, or penalize malicious and ignorant remarks. In future studies, it will be worthwhile to investigate what criteria users employ to judge the quality and relevance of results.

Acknowledgments. The authors acknowledge the support of the Research Grant from the Office of the Vice President for Research at the University of Texas at Austin.

References

1. Arora, N.K., Hesse, B.W., Rimer, B.K., Viswanath, K., Clayman, M.L., Croyle, R.T.: Frustrated and confused: the American public rates its cancer-related information-seeking experiences. Journal of General Internal Medicine 23(3), 223–228 (2007)
2. Beaver, K., Witham, G.: Information needs of the informal carers of women treated for breast cancer. European Journal of Oncology Nursing 11, 16–25 (2007)
3. Belkin, N.J., Oddy, R.N., Brooks, H.M.: Ask for information retrieval: Part I. background and theory. Journal of Documentation 38, 61–71 (1982)

4. Boden, C.: Overcoming the linguistic divide: a barrier to consumer health information. Journal of the Canadian Health Libraries Association 30, 75–80 (2009)
5. Bradley, J.: Methodological issues and practices in qualitative research. Library Quarterly 63(4), 431–449 (1993)
6. Bystrom, K., Hansen, P.: Conceptual framework for tasks in information studies. Journal of the American Society for Information Science and Technology 56(10), 1050–1061 (2005)
7. Degner, L.F., Kristjanson, L.J., Bowman, D., et al.: Information needs and decisional preferences in women with breast cancer. Journal of the American Medical Association 277, 1485–1492 (1997)
8. Eysenbach, G., Powell, J., Englesakis, M., Rizo, C., Stern, A.: Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 328(7449), 1166–1171 (2004)
9. Fox, S., Jones, S.: The Social Life of Health Information. Pew Internet & American Life project (2009) http://www.pewinternet.org/~/media//Files/Reports/2009/PIP_Health_2009.pdf (retrieved March 3, 2010)
10. Ginossar, T.: 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 (2008)
11. Glaser, B.G., Strauss, A.L.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine, New York (1967)
12. Harrison, D.E., Galloway, S., Graydon, J.E., Palmer-Wickham, S., van der Bij, L.: Information needs and preference for information of women with breast cancer over a first course of radiation therapy. Patient Education and Counseling 38, 217–225 (1999)
13. Jasra, M.: Yahoo Answers down but dominating Q&A websites (March 20, 2008), http://www.webanalyticsworld.net/2008/03/yahoo-answers-dominates-q-websites.html(accessed on May 3, 2010)
14. Keselman, A., Logan, R., Smith, C.A., Leroy, G., Zeng, Q.: Developing informatics tools and strategies for consumer-centered health communication. Journal of the American Medical Informatics Association 15(4), 473–483 (2008)
15. Keselman, A., Smith, G.A., Divita, G., Kim, H., Browne, A.C., Leroy, G., Zeng-Treitler, Q.: Consumer health concepts that do not map to the UMLS: Where do they fit? Journal of American Informatics Association 15, 496–505 (2008)
16. Kim, D., Chang, H.: 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 (2007)
17. Klemm, P., Reppert, K., Visich, L.: A nontraditional cancer support group: the Internet. Computer in Nursing 16(1), 31–36 (1998)
18. Liu, Y., Bian, J., Agichtein, E.: Predicting information seeker satisfaction in community question answering. In: Proceedings of SIGIR (2008)
19. Luker, K.A., Beaver, K., Leinster, S.J., Owens, R.G.: The information needs of women with breast cancer: a follow up study. Journal of Advanced Nursing 23, 487–495 (1996)
20. McCray, A.T., Tse, T.: Understanding search failures in consumer health information systems. Proceedings of AMIA Symposium, 430–434 (2003)
21. Rose, D.E., Levinson, D.: Understanding user goals in web search. In: Proceedings of WWW 2004, pp. 13–19 (2004)

22. Spink, A., Yang, Y., Jansen, J., Nykanen, P., Lorence, D.P., Ozmutlu, S., Ozmutlu, H.C.: A study of medical and health queries to web search engines. Health Information and Libraries Journal 21, 44–51 (2004)
23. Taylor, R.S.: Question-negotiation and information seeking in libraries. College & Research Libraries 29, 178–194 (1968)
24. Toms, E.G., O'Brien, H., Mackenzie, T., Jordan, C., Freund, L., Toze, S., Dawe, E., MacNutt, A.: Task effects on interactive search: the query factor. In: Fuhr, N., Kamps, J., Lalmas, M., Trotman, A. (eds.) INEX 2007. LNCS, vol. 4862, pp. 359–372. Springer, Heidelberg (2008)
25. Warner, D., Procaccino, J.D.: Toward wellness: Women seeking health information. Journal of American Society for Information Science and Technology 55(8), 709–730 (2004)
26. White, M.D.: Questioning behavior on a consumer health electronic list. Library Quarterly 70(3), 302–334 (2000)
27. White, R.W., Dumais, S., Teevan, J.: How medical expertise influences web search interaction. In: Proceedings of SIGIR 2008, pp. 791–792 (2008)
28. Wildemuth, B.M., Bliek, R., Friedman, C.P., Miya, T.S.: Information-seeking behaviors of medical students: a classification of questions asked of librarians and physicians. Bulletin of Medical Library Association 82(3), 295–304 (1994)
29. Zeng, Q.T., Crowell, J., Plovnick, R.M., Kim, E., Ngo, L., Dibble, E.: Assisting consumer health information retrieval with query recommendations. Journal of American Medical Informatics Association 13, 80–90 (2006)
30. Zickuhr, K.: Generations 2010. Pew Internet & American Life project (2010), `http://www.pewinternet.org/~/media//Files/Reports/2010/PIP_Generations_and_Tech10.pdf` (retrieved December 21)