American Journal of Infection Control 40 (2012) 211-7

Contents lists available at ScienceDirect

American Journal of Infection Control

journal homepage: www.ajicjournal.org

Major article

Assessment of H1N1 questions and answers posted on the Web

Sujin Kim PhD [a,b,c,*], Thomas Pinkerton BA [b], Nithya Ganesh MS [b]

[a] Division of Biomedical Informatics, College of Public Health, University of Kentucky, Lexington, KY
[b] School of Library and Information Science, College of Communication and Information Studies, University of Kentucky, Lexington, KY
[c] Department of Pathology and Laboratory Medicine, School of Medicine, University of Kentucky, Lexington, KY

Key Words:
Influenza A FAQ, H1N1 surveillance
Text mining
PubMed
Influenza pandemic
Medical Internet research
Consumer health information

Background: A novel strain of human influenza A (H1N1) posed a serious pandemic threat worldwide during 2009. The public's fear of pandemic flu often raises awareness and discussion of such events.
Objectives: The goal of this study was to characterize major topical matters of H1N1 questions and answers raised by the online question and answer community Yahoo! Answers during H1N1 outbreak.
Methods: The study used Text Mining for SPSS Clementine (v.12; SPSS Inc., Chicago, IL) to extract the major concepts of the collected Yahoo! questions and answers. The original collections were retrieved using "H1N1" in search, keyword and then filtered for only "resolved questions" in the "health" category submitted within the past 2 years.
Results: The most frequently formed categories were as follows: general health (health, disease, medicine, investigation, evidence, problem), flu-specific terms (H1N1, swine, shot, fever, cold, infective, throat), and nonmedical issues (feel, North American, people, child, nations, government, states, help, doubt, emotion). The study found that URL data are fairly predictable: those providing answers are divided between ones dedicated to giving trustworthy information—from news organizations and the government, for instance—and those looking to espouse a more biased point of view.
Conclusion: Critical evaluation of online sources should be taught to select the quality of information and improve health literacy. The challenges of pandemic prevention and control, therefore, demand both e-surveillance and better informed "Netizens."

A novel strain of human influenza A (H1N1) posed a serious pandemic threat worldwide during 2009. The public's fear of pandemic flu often creates voluminous online or offline discussions about diseases ranging from laboratory confirmation status, age, relative severity, exposure history, onset of symptoms, and contact history.[1] Pandemic response to the novel flu has highlighted the importance of online information dissemination for the support of disease control and surveillance. However, the information given at the majority of the public institutions is packaged to serve unidirectional announcements, in the hope of reaching out to the majority of the public. Online forums, on the other hand, are regarded as highly interactive communication and are populated by people who can both post and answer questions. These communities have formed to share information and to fill knowledge gaps in health matters. Considering the large number of people who use Web resources for seeking health information, such an application could be an important vehicle for disseminating information and interacting with the goal of serving health-related information questions—in the case of this study, H1N1.[2,3] Using this context, the primary goal of this study was to characterize the major topical matters of H1N1 questions and answers raised by the online question and answer community Yahoo! Answers during the 2009 H1N1 outbreak. The following section discusses 2 research streams: influenza information sources and services and online knowledge sharing through online questions and answers.

INFLUENZA INFORMATION, SOURCES, AND SERVICES

Frequently, available information on influenza covers conventional topics of infectious disease, with information sources and

* Address correspondence to Sujin Kim, PhD, Division of Biomedical Informatics, College of Public Health, University of Kentucky, 339 Lucille Little Fine Art Library, Lexington, KY 40506-0224.
E-mail address: sujinkim@uky.edu (S. Kim).
Supported in part by a grant (RE-04-08-0069-08) from the Institute of Museum and Library Service (IMLS) and by grant number P20RR-16481 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH).
Conflicts of interest: None to report.

0196-6553/$36.00 - Copyright © 2012 by the Association for Professionals in Infection Control and Epidemiology, Inc. Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.ajic.2011.03.028

services running the following gamut: general overviews of the influenza virus, vaccination, assessment, laboratory testing, treatment, infection prevention and control measures, pandemic influenza planning and preparation, human resource issues, work absences, and travel.[4] Depending on the organization, one can find tailored information for the general public as well as health care workers, caregivers, and policy makers that are targeted for specific flu-concerned events and activities such as planning for businesses, the community, and school events, as well as domestic and international traveling. The information contained in the popular influenza sources on the Web is prepared to respond to some frequently asked questions (FAQs) about vaccine and vaccine development, what to do if one gets flu-like symptoms, how to care for a sick person, what pregnant women should know, community strategy for pandemic influenza mitigation, and national strategy for pandemic influenza.[4-6] The FAQ's goal, of course, is to fulfill the user's information need, as well as provide an outlet for the public to become knowledgeable—at least at a general level—on the subject of pandemic influenzas. However, no access mechanism is provided on these popular flu Web sites for searching the FAQ questions, which means users must go through the hyperlinked question list linearly. Despite the fact that the FAQs—as localized compendiums of influenza resources and services—are intended to help the public as well as health care workers, the original information sourced from the major Web sites is difficult to navigate (to find answers to specific questions) because of its size and lack of search mechanisms.

The social media technologies have become widely studied in several academic disciplines as a new method of disease surveillance, by identifying online communities for targeted pandemic communications. Corley et al, in 2010, identified flu trends posted online and correlated then with the Centers for Disease Control and Prevention (CDC) Influenza-like Illness Surveillance Program (ILINet) data by using text and structural data mining techniques.[7] User search keywords sent to the popular search engines Google and Yahoo! were studied to track influenza-like illness (ILI).[8,9] Other influenza-concerned information services have also been investigated by analyzing Web access logs and telephone triage service data, used to detect ILI symptoms.[10,11] All of these studies show evidence of social media technologies being used in health communication. These media have become critical information dissemination vehicles during pandemic outbreaks among people seeking influenza information.

Sharing knowledge online has become a major topic of research in the computer sciences and in other information-intensive domains: the health sciences, library and information science, and others. Web applications such as Yahoo! Answers are built on the assumption that everyone knows something and that, for the most part, people who know something are willing to share their knowledge with those who seek information. However, the authority behind these answers is a primary concern of trained health information specialists because few are experts on these topics. Some applications address this by providing a function that shows the best answers, as rated by fellow users, to assist people in identifying reliable and high-quality answers. However, this mechanism alone cannot filter the most trustworthy information for online questioners. This is particularly important when it comes to serious matters such as health issues because bad information can result in serious consequences. Knowledge-sharing applications should implement methods for highlighting the best information in a more objective manner because their biggest challenge is to understand exactly who posts questions and who answers them. In addition, this current research is examining what questions are raised and how these questions are answered and by whom. For instance, Adamic et al, in 2008, reported that diverse types of questions are often asked by Yahoo! Answers users and that these types of questions can be predicted by categorizing the questions posted by a questioner.[12] In addition, to this point, there has been a lack of research that investigates users of knowledge-sharing systems and the information posted by them.

In the field of library and information science, question asking has been studied as a way of understanding library user's information needs for facilitating reference services, as well as being employed in designing information retrieval systems.[13] Library and information science researchers have heavily studied the characteristics of reference questions, which then facilitate the interaction between human searchers and bibliographic retrieval systems. Within the context of information retrieval studies, the questions are expressed in a few search keywords, as a manifestation of searcher's queries submitted to receive a set of relevant results of bibliographic records (eg, articles, books, and others). Professional searchers are trained to analyze the questions posed by end-users to construct efficient search strategies, by first identifying major topics, and then determining a set of subheadings. Some of the reasons people pose questions online are to update timely information, to write and read compiled or multiple messages in one place, to communicate with others who are also seeking the same information, to expand knowledge, to validate information received from other sources, and to prepare background information before visiting health care provider. Apart from medical knowledge, some people simply seek emotional support and encouragement to cope with the disease.[14]

METHODS

Research questions (RQ)

This study investigates the following 3 research questions to understand questions and answers posted by the public in search of H1N1-related information through the online question/answer discussion service Yahoo! Answers.

- RQ1: What are the topical characteristics of questions posted by questioners on Yahoo! Answers?
- RQ2: What are the major resources referenced in the answers gathered from respondents on Yahoo! Answers?
- RQ3: What are the mapping results of the extracted concepts and descriptors into the vocabularies in the Unified Medical Language System (UMLS) Metathesaurus?

Data collection

The original collection of the questions and answers were keyword searched using "H1N1" and then filtered for only "resolved questions" in the "health" category submitted within the past 2 years. The data collection was completed on April 30, 2010. The answer set only included the "best answer" selected by the questioner. The original collection of entries from Yahoo! Answers was 6,578; however, it contained many items that were either inappropriate within the scope of this study or duplicates. The entries were proofread to make the text mining process more efficient: excess punctuation was removed, and major headings such as "H1N1" were normalized (such outliers as "HINI" were edited). Next, the list was narrowed further by eliminating duplicate questions. When eliminating duplicates, the entire question and answer collection to be examined was narrowed from 6,578 to 5,400.