Health Conversational System based on Contextual Matching of Community-Driven Question-Answer Pairs

Wilson Wong, John Thangarajah, Lin Padgham
School of Computer Science and Information Technology
RMIT University
{wilson.wong,john.thangarajah,lin.padgham}@rmit.edu.au

ABSTRACT

More and more people are turning to the World Wide Web for learning and sharing information about their health using search engines, forums and question answering systems. In this demonstration, we look at a new way of delivering health information to the end-users via coherent conversations. The proposed conversational system allows the end-users to vaguely express and gradually refine their information needs using only natural language questions or statements as input. We provide example scenarios in this demonstration to illustrate the inadequacies of current delivery mechanisms and highlight the innovative aspects of the proposed conversational system.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*search process*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*natural language*

General Terms

Algorithms, Human Factors

Keywords

QA Pairs, Conversational Agent, Health Information

1. INTRODUCTION

The World Wide Web has revolutionised the ways we access information, and health-related information is no exception [5]. The sense of empowerment derived from the any-time, anywhere access of online information coupled with the rising cost of healthcare have contributed to our increasing reliance on the Web for purposes ranging from self-diagnosis to sharing patient experiences. Although the use of online information for health-related purposes has been greatly criticised by health practitioners [8, 7], benefits have

also been suggested. Information on the Web has the ability to reach hard to reach groups such as teenagers and raise their awareness on health related issues. The ability to provide information for those with minor health problems that may be managed without the need for medical consultation [5] also helps to free up the time of health practitioners to deal with people requiring attention.

Currently, the end-users have a range of online services to rely on for their health information needs. These services range from live chat with medical practitioners (e.g., `AskTheDoctor.com`) and self-help communities for sharing patient experiences (e.g., `PatientsLikeMe.com`) to Web search engines (e.g., Yahoo!, Bing) and specialised question answering systems (e.g., HONqa). It is clear from the proliferation of these websites that people find it beneficial and valuable to be able to inquire about or search for health information at their own convenience. Some of the shortcomings of these services, however, prohibit the effective delivery of health information to the end-users. The lack of concise responses and the difficulty of representing complex health information needs as Boolean queries make search engines far from ideal [10]. Question answering systems provide the end-users with concise answers for factual questions. This specialisation on answering factual questions (e.g., *"what"*, *"when"*, *"where"*), however, greatly restricts their applications to other input types such as *why*-questions or even statements. The need to express something as multi-faceted as health concerns all in a single natural language question may not be ideal. It is our view that a conversational ability that supports context-dependent, open-ended health information seeking is beneficial to allow the end-users to gradually explore a complex information need.

In this demonstration, we present a Web-based system that engages end-users in coherent discussions about their health concerns. This system is our first attempt at addressing some of the shortcomings of existing delivery mechanisms discussed above. Although there exist some health conversational agents in the literature, they focus on very specific tasks such as promoting behavioural change for long-term health [1] or acquiring the objective (e.g., weight) and subjective (e.g., pain) aspects of a patient's conditions for in-home monitoring [3]. These systems, which are based on deep discourse analysis, require the intent of the end-users to be recognised and the various states of the conversations to be anticipated and represented. These systems are not suitable for managing more open-ended natural language input about personal health and other related information.

Our conversational system allows the end-users to vaguely

express and gradually build up over time their health information needs. The innovative aspects of this system, which set it apart from existing conversational agents, are:

- To the best of our knowledge, it is the first system to use disjointed question-answer (QA) pairs from community-driven websites to build coherent conversations.[1]

- We use a novel scoring process based on exponential decay and contextual information to rank QA pairs and select the best answers to be used as system responses.

- The system supports open-ended conversations about health and related issues, in real time, as opposed to task-specific interactions.

The demonstrated system provides far more concise responses than document retrieval systems, and while it may not deliver the precise answers available from a specialised question answering system, it is able to carry on a meaningful and natural interaction in response to both questions and statements. The coherence of the content is far greater than standard chatterbots such as ALICE[2].

We allow the QA pairs retrieved to be manually edited to improve their quality, though certain aspects such as spelling correction can be automated to some extent. In our demonstration system, the QA pairs are unedited in our DB, though in the examples shown only the initial sentences are used for readability. Although this demonstration focuses on health-related conversations, the system is generic and can be extended to converse about other topics by downloading the corresponding QA pairs.

2. CONVERSATIONAL SCENARIOS

To illustrate the use and features of the demonstration system and its differences from existing delivery mechanisms, let us consider the following scenario. A user accidentally cuts her finger whilst cooking and has no knowledge of basic first aid. If she is one of the two billion Internet users worldwide today[3], she may first go to the Web for information. Her first stop would be to use one of the many popular Web search engines today such as Google, Yahoo! or Bing. Using Google in this scenario to search for *"I've accidentally cut myself. What do I do?"*, close to 400 million results were returned with Figure 2 showing the top 3. The results are clearly not the concise response desired in such a situation. A user may then try alternative keywords to find more suitable results. Even then, the results obtained are generally not concise as the end-user has to click on each of the individual links and filter through the webpages looking for the appropriate answer.

The above is typical for search engines as they are simply document retrieval systems. Let us instead consider using the two question answering systems HONqa[4], that specialises in health information, and the more generic START[5]. Entering the same natural language question *"I accidentally cut myself. What do I do?"*, we find, as shown in Figures

---

[1]There is work using QA pairs for automated question answering [2, 9], but our focus is on coherent conversations.

[2]http://alice.pandorabots.com/

[3]http://www.itu.int/net/pressoffice/press_releases/2010/39.aspx

[4]http://www.hon.ch/QA

[5]http://start.csail.mit.edu

Figure 1: The results of searching the Web for information about treatments for cuts using Google.

2(a) and 2(b), that the two systems are unable to 'understand' the input. This is because these systems work better with more straightforward inputs in the form of single direct questions. After reformulating the user's concern to *"How to treat cut fingers?"*, START still returned no answers and HONqa returned a summary *"If the bleeding is heavy, bright red or spurting then follow the steps to control bleeding. If the finger is amputated, put pressure on it to control bleeding and follow the steps for treating an amputation. Fundamental First Aid Cuts and Bleeding First Aid Kits and Supplies Related Articles Also from About.com"*. Even with this response from HONqa, the user still has to click on a link to About.com to read an article about stopping bleeding.

Another approach would be to ask a real medical practitioner on websites such as AskTheDoctor.com. However the delay in getting a response, and the need to disclose personal information such as name and email address may make this a less desirable option. Finally the user could attempt to engage in a conversation regarding her concern with a conversational agent such as ALICE. Figure 2(c) illustrates a conversation that was initiated by the same question *"I cut myself. What do I do?"*. We can observe from the figure that the resulting conversation is more likely to frustrate or perhaps amuse the user than to be of any assistance.

The scenario described above illustrates the lack of a conversational system that allows a user to express, clarify and refine complex information needs in natural language. With our system, a user is offered some of these capabilities to improve their health information seeking experience. Figure 3 shows a conversation with the proposed system, initiated by the question *"I've accidentally cut myself. What do I do?"*. The system concisely responds along the lines of *"From a first aid perspective...(do ... else)...just go to your doctor."*. Given the suggestion of using antiseptic, the user, who has knowledge of alcohol and peroxide, inquires *"Can I rub some alcohol on?"* and later *"How about peroxide?"*. Notice how the system is able to interpret these two inputs in the context of the previous user utterance concerning accidental cuts. Instead of requiring the end-user to fully specify their questions, e.g., *"Can I rub some alcohol on my wound?"* or *"Can I use peroxide on the wound that I've just accidentally created?"*, the system systematically takes into account keywords from previous utterances to allow the end-users to concisely represent their information needs, and to ensure the best system responses actually correspond to the previ-

HONQA | I accidentally cut myself. What do I do?

START's reply

===> I accidentally cut myself.

100 %

You asked me to add the following assertion to my knowledge base: I ACCIDENTALLY CUT MYSELF. At present, however, remote users are only allowed to ask questions. Please send mail to Boris@csail.mit.edu for more information.

No answers found, please try with **HONcodeHunt**

===> What do I do?

You are talking to the START natural language system developed by Boris Katz at the MIT Artificial Intelligence Laboratory.

Source: START KB

(a) START　　　　　　(b) HONqa

Human: I accidentally cut myself.
ALICE: Thanks for the information: unknown person said he cut him or herself.
Human: What do I do?
ALICE: You said you do cut yourself.
Human: I cut myself by accident.
ALICE: I haven't heard anything like that before: unknown person said he cut him or herself by accident.
Human: Please.
ALICE: Please what?
Human: I'm serious here.
ALICE: Does that make you feel superior?

You say: |

Benefits of Membership　　　Vote for ALICE!

(c) ALICE

Figure 2: The answers provided by (a) START, (b) HONqa and (c) ALICE in response to the input "I accidentally cut myself. What do I do?".

ask [clear] [switch to debug]

What is betadine?　　　　　　　　　　　　　　　ask

USER: *I cut myself. What do I do?*
SYSM: *From a first aid perspective, keep it clean with antiseptic and try not to use that part of your hand too roughly so it doesn't open back up. Keep the wound closed and covered with a dressing. If you have any doubt, just go to your doctor.*
USER: *Can I rub some alcohol on?*
SYSM: *Alcohol pads should not be used on open cuts or wounds, since they cause tissue damage and delay healing. They can be used to prep unbroken skin for injections etc. or to disinfect equipment such as thermometers.*
USER: *How about peroxide?*
SYSM: *The peroxide has been found to not be affective at disinfecting. The peroxide bubbles b/c of mixing with the waste that bacteria produce and make bubbles, but don't actually kill the bacteria. 70% rubbing alcohol kills the surface bacteria to decrease the chances of getting infected. So, if you get a shot, the bacteria don't get pushed under the skin with the needle. Peroxide just bubbles with the bacteria's poop.*
USER: *What antiseptic works best then?*
SYSM: *Good old soap and water!*
USER: *Seriously, my wound is quite large.*
SYSM: *Just wash the area with Betadine. This is a solution you can get over the counter. It is the solution used to prep a surgery site to keep the wound from getting infected.*
USER: *What is betadine?*
SYSM: *Betadine is the name of Purdue Pharma's brand of consumer-available povidone-iodine (PVPI) topical antiseptics. Betadine, like most PVPI products, is available in two formulations. First, a solution, sold over-the-counter (OTC) for cleaning minor wounds and used in hospitals to prepare a patient's skin prior to surgery. Second, a surgical scrub, which is a mixture of povidone-iodine and detergent, sold OTC as a skin cleaner and disinfectant hand wash and used for cleansing hands prior to surgery and other aseptic procedures.*

Figure 3: An example conversation with the proposed system which was initiated by the end-user using the input "I cut myself. What do I do?".

ous user inputs. After learning from the system that alcohol or peroxide should not be used on her wounds, she moved on to ask *"What antiseptic works best then?"*. The proposed system answered with *"Good old soap and water!"*, which is a valid response despite being somewhat cheeky. The user persists on obtaining more information by asking *"Seriously, my wound is quite large"*. Not knowing what *"Betadine"* is from the system's response, she further inquires *"What is betadine?"*, to which a reasonable answer was provided.

The responses produced by the proposed system in the above scenario are all answers from Yahoo!Answers' QA pairs that were truncated for presentation. The response *"Good old soap and water!"*, for instance, is the answer to the question *"What works best to clean cuts alcohol,hydrogen peroxide, hand sanitizer ect?"*, which together form a QA pair from Yahoo!Answers[6]. As for the system response *"Just wash the area with Betadine..."* to the user input *"Seriously, my wound is quite large"*, it is the answer to the question *"What works better for large cuts... What else is good to use?"*[7]. These two and in fact, all other QA pairs from Yahoo!Answers are distinct answers to individual questions. This is why we refer to the QA pairs as disjointed. By themselves, they are unable to contribute much to forming coherent conversations. However, with the scoring process based on exponential decay and contextual information to be briefly described in the next Section 3, the proposed system is able to engage users in coherent and relevant conversations about their health concerns.
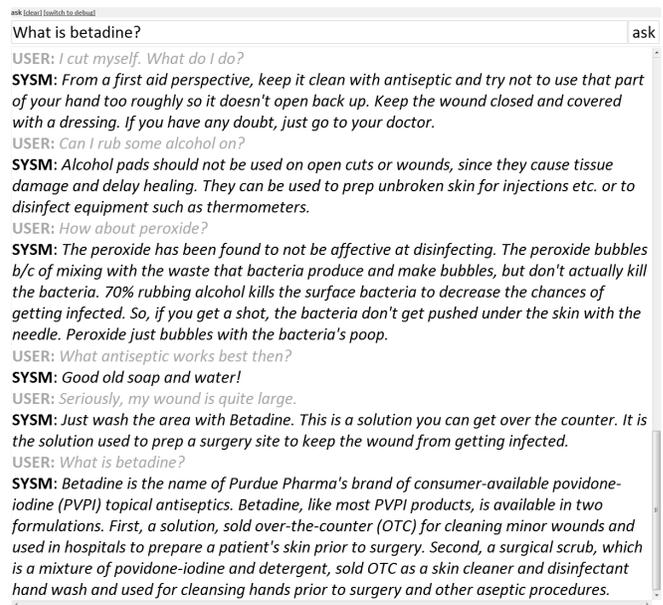
---

[6]http://answers.yahoo.com/question/index?qid=20080722161422AAxIRpe
[7]http://answers.yahoo.com/question/index?qid=20060726190740AAXdL1L

3. SYSTEM COMPONENTS

In this section, we discuss the two main components, as shown in Figure 4, of the proposed conversational system, namely, the collection of question and answer (QA) pairs, and the process through which we use the QA pairs for generating coherent responses.
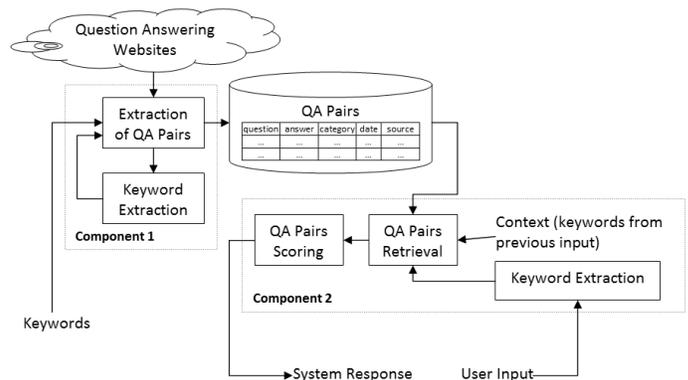

Figure 4: Components of the proposed conversational system.

The extraction of QA pairs from community-driven question-and-answer (Q&A) websites is performed *offline*. That is, they are downloaded prior to, and are independent of, the real-time process of engaging in conversation. A QA pair is simply the pairing of a question posted by a human together with the possible answers contributed by other volunteers. The QA pairs are currently extracted from Yahoo!Answers

using the API[8] provided by Yahoo!. The use of QA pairs is ideal considering the fact that data from Q&A websites are not optimal for factoid input but instead, are a preferred choice for complex information needs such as opinion or advice [4]. We are also motivated by the availability of APIs to ease the implementation process. QA pairs may also be obtained from other community-driven Q&A websites such as `Answers.com`. Many of these websites, however, do not permit the non-commercial, automated extraction of data from their sites, hence preventing their use here.

A GUI has been created to allow the system administrator to manage the various aspects of gathering QA pairs from Yahoo!Answers. Using this interface, the administrator is able to provide keywords for extraction, examine the current composition of the QA pairs, and search and edit the QA pairs currently in the database. Other metadata such as category and the date of posting, which are not used in this version of the system, can also be extracted and used during the scoring process to improve the quality of the responses.

Next, we look at the process of using inputs in a conversation to create gradually-decaying context for retrieving and scoring QA pairs in our database with the aim of locating the best responses. Every input is first analysed for keywords. In this version of the system, we consider a keyword as a word or phrase that is content-bearing, that is, not a function word. The words in an input are first chunked using parts of speech and later filtered using a stopword list[9]. At each turn in a conversation where the user provides an input, the keywords from previous user utterances (i.e., context) and the current keywords are unioned and weighted using a decay model. Intuitively, a keyword that no longer appears in the more recent inputs will have its weight decayed. If a keyword is constantly repeated up till the most recent input, its weight will be more.

A simple structured query to the database is then performed to obtain all QA pairs containing at least one keyword from the unioned and weighted set of keywords at the current turn. Each pair is assigned a score using four criteria based on the frequency and location of the occurrences of keywords as well as other heuristics involving edit distance and so on. The details of these criteria and the scoring process in general are outlined in [11][10].

4. DEMONSTRATION PLAN

In this demonstration, we will show the system interfaces for conversing about health issues with the end-users, and for populating the database with QA pairs to cope with evolving conversations. In particular, we will demonstrate how an end-user can easily interact with the proposed conversational system over a range of health issues. During the demonstration, we will restrict conversations with the system to 10 health and medical-related concepts (e.g., *"lung cancer"*, *"sinusitis"*, *"dry mouth"*) from the list of 150 used for evaluating health-specific question answering systems [6]. This restriction will allow us to examine the capabilities and limitations of the system together with the audience within a controlled space, showing the effect of the depth and breadth of QA pairs on the system's conversations. We have prepared four videos[11] as a preview of our demonstration.

5. REFERENCES

[1] BICKMORE, T., SCHULMAN, D., AND SIDNER, C. A reusable framework for health counseling dialogue systems based on a behavioral medicine ontology. *Journal of Biomedical Informatics 44*, 2 (2011), 183–197.

[2] CONG, G., WANG, L., LIN, C., SONG, Y., AND SUN, Y. Finding question-answer pairs from online forums. In *Proceedings of the 31st Annual International ACM SIGIR Conference* (Singapore, 2008).

[3] FERGUSON, G., ALLEN, J., GALESCU, L., QUINN, J., AND SWIFT, M. Cardiac: An intelligent conversational assistant for chronic heart failure patient heath monitoring. In *Proceedings of the AAAI Fall Symposium on Virtual Healthcare Interaction* (Arlington, USA, 2009).

[4] LIU, Y., AND AGICHTEIN, E. On the evolution of the yahoo! answers qa community. In *Proceedings of the 31st Annual International ACM SIGIR Conference* (Singapore, 2008).

[5] MCDAID, D., AND PARK, A. Online health: Untangling the web. Report Bupa Health Pulse 2010, London School of Economics, 2011.

[6] OLVERA-LOBO, M., AND GUTIERREZ-ARTACHO, J. Open- vs. restricted-domain qa systems in the biomedical field. *Journal of Information Science 37*, 2 (2011), 152–162.

[7] ROBERTSON, N., AND HARRISON, P. Whats wrong with me? concerns about online medical self-diagnosis. In *Proceedings of the Australia and New Zealand Marketing Academy Conference (ANZMAC)* (Monash University, Australia, 2009).

[8] RYAN, A., AND WILSON, S. Internet healthcare: Do self-diagnosis sites do more harm than good? *Expert Opinion on Drug Safety 7*, 3 (2008), 227–229.

[9] WANG, B., WANG, X., SUN, C., LIU, B., AND SUN, L. Modeling semantic relevance for question-answer pairs in web social communities. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)* (Uppsala, Sweden, 2010).

[10] WHITE, R., AND HORVITZ, E. Web to world: Predicting transitions from self-diagnosis to the pursuit of local medical assistance in web search. In *Proceedings of the AMIA Annual Symposium* (Washington DC, USA, 2010).

[11] WONG, W., THANGARAJAH, J., AND PADGHAM, L. Open-ended health conversational system. Technical report, RMIT University, 2011.

---

[8] `http://developer.yahoo.com/answers`

[9] In later versions, we will examine the need for more advanced noun phrase chunking and term extraction.

[10] The report is in the process of being submitted for publication.

---

[11] `http://www.youtube.com/watch?v=n-liJ0hsklw`
`http://www.youtube.com/watch?v=v6JSK6nvkno`
`http://www.youtube.com/watch?v=c0goNkKDT18`
`http://www.youtube.com/watch?v=owjajxRX-d4`