Fuzzy-Based Answer Ranking in Question Answering Communities

B.A. Ojokoh, Department of Computer Science, Federal University of Technology, Akure, Ondo State, Nigeria
P.I. Ayokunle, Department of Computer Science, Federal University of Technology, Akure, Ondo State, Nigeria

ABSTRACT
Owing to the vast amount of information readily available on the World Wide Web, there has been a significant increase in the number of online question answering (QA) systems. A branch of QA systems that has seen such remarkable growth is the community-based question answering (CQA) systems. In this paper, the authors propose a method that is proactive enough to provide answers to questions and additionally offers word definitions, with the aim of reducing the time lag that results from askers having to wait for answers to a question from various users. Additionally, it designs a method to evaluate and predict the quality of an answer in a CQA setting, based on experts’ rating. It uses fuzzy logic to aggregate the ratings and provide ranked answers in return. Experimental results with computing-related datasets from Yahoo! Answers demonstrate the effectiveness of the proposed techniques.

Keywords: Answerers, Askers, Community-Based Question Answering (CQA), Experts, Fuzzy Logic, Question Answering (QA), Voters

1. INTRODUCTION
Question answering is a computer science discipline within the fields of information retrieval and natural language processing. It is the task of automatically answering a question posed in natural language. QA systems deliver users short, succinct answers instead of overloading them with a large number of irrelevant documents (Lin and Katz 2003). This is the goal of every QA system. More commonly, QA systems pull answers from unstructured collection of natural language documents. Community Question Answering (CQA) services are dedicated platforms for users to respond to other users’ questions, resulting in the building of social communities where users share and interactively give ratings to questions and answers (Wang and Zhang 2011; Pal and Konstan 2010). From observation, more users are becoming inclined to this new area of QA systems as they can obtain a near perfect answer to their

DOI: 10.4018/jdls.2012070105
questions from other users rather than a list of likely documents containing result(s) to their questions that are provided by the system. While current web search engines enjoy huge commercial success and demonstrate good performance, especially for homepage-finding queries, their ability to find relevant information for hard queries such as those asking for opinions or summaries is far from satisfactory (Jeon et al., 2005).

Many of the existing works on QA systems have focused on retrieving high quality answers to an asker’s question and recommending answerers (Hieber and Riezler 2011). However, the time lag that results from askers having to wait for answers to a question from various users as identified by Jeon et al. (2005) have not received much consideration. In addition, while browsing the internet, we often spend large amount of time searching for a particular answer that best answers our question, and end up getting too large materials that are not specific and so difficult to extract specific and quality answers to the questions we have. If we later extract some line of text or a paragraph of text we are not even sure if the extracted text is the best, in this case an interactive question answering system that enables the asker get specific community rated expert answers would have been better. Answering communities such as Yahoo! Answers offer great intelligence to the users who have questions in either their daily lives or academic affairs. Discovering experts in these communities is a very important research problem that have been explored by a number of researchers (Liu et al., 2005; Jurczyk and Agichtein, 2007; Pal et al., 2012). Participants can express their judgments towards answers by voting for the answer they feel is the best among the expert provided answers. The need to obtain aggregate opinions from these experts relating to a particular question in an effective way forms one of the motivations for this research work.

Community Question Answering (QA) portals contain questions and answers contributed by hundreds of millions of users. These databases of questions and answers are of great value if they can be used directly to answer questions from any user. This paper therefore aims at using this venue to reduce the time lag between when a question is asked and an answer is found to a bare minimum by selecting the most relevant question and answer pair (from previously answered questions) to an asker’s query before other users are available to provide answers to such query. It uses text similarity measures (Dice metric and a modified form of Levenshtein distance) for relevant answer retrieval. It additionally offers definitions of words and their synonyms when the “define” keyword precedes the query using the WordNet database. Consequently, this additional feature brings dictionary functionality into proposed the QA system. It also designs a fuzzy logic based method to aggregate the rating of answers provided by expert users in a cQA and afterwards produce ranked answers to a particular question. Section two reviews related works. Section three describes the architecture of the proposed system. Section four contains the experiments and evaluation while section five concludes the work and proposes areas of further research.

2. RELATED WORK

Question Answering Systems have been in existence for several decades now (Green et al., 1961; Wood et al., 1973; Wilensky et al., 1989). Katz et al. (2002) developed the first question answering system (called START, SynTactic Analysis using Reversible Transformations) for the World Wide Web. START has been in continuous operation since December, 1993 and has answered millions of questions from hundreds of thousands of users all over the world. There has been extensive research on question answering (QA) since then (Burger et al., 2002; Dang and Tuyen 2009). The traditional QA solutions such as in Agichtein (2002) are often times content-based and focus on searching and extracting answers from a static collection of text segments or web data (Suryanto et al., 2009). Some QA works are
Related Content

Reports Generation with Koha Integrated Library System (ILS): Examples from Bowen University Library, Nigeria
[www.igi-global.com/article/reports-generation-with-koha-integrated-library-system-ils/174456?camid=4v1a](www.igi-global.com/article/reports-generation-with-koha-integrated-library-system-ils/174456?camid=4v1a)

Social Media and Copyright in Digital Libraries
[www.igi-global.com/chapter/social-media-and-copyright-in-digital-libraries/188541?camid=4v1a](www.igi-global.com/chapter/social-media-and-copyright-in-digital-libraries/188541?camid=4v1a)

First Person Singular: A Digital Library Collection that Helps Second Language Learners Express Themselves
[www.igi-global.com/article/first-person-singular/39035?camid=4v1a](www.igi-global.com/article/first-person-singular/39035?camid=4v1a)
Libraries as Publishers of Digital Video
www.igi-global.com/chapter/libraries-publishers-digital-video/5951?camid=4v1a