Social Information Retrieval for Online Community Question Answering Services

Jun Choi Lee[1], Yu-N Cheah[2], and Wai Shiang Cheah[3]

[1]*Universiti Malaysia Sarawak. Malaysia, {jclee; wscheach}@unimas.my*

[2]*Universiti Sains Malaysia, Malaysia, yncheah@usm.my*

ABSTRACT

Online Community Question Answering Services is a platform that allows users to seek for answers anytime and anywhere. The services cover a large collection of unstructured knowledge and become a knowledge repository for problem solving. The answers in the services are generated by other public user. This environment creates a large collection of unstructured knowledge, which can be use as knowledge source for other intelligent processes. From the review, it is a challenge to access accurate information from the Online Community Question Answering services. This study propose a social information retrieval method to access the knowledge from the services. The proposed method query and retrieve information the online services through query expansion. Then, the retrieved information is filtered through a similarity analysis between a question with a query statement and semantic relatedness between answers and the query statement. The accuracy of the proposed approach is evaluated through human evaluation.

Keywords: Social Information Retrieval, Online Community Question Answering Services, Semantic Relatedness.

I INTRODUCTION

Online Community Question Answering Services (CQA) like Yahoo! Answers, Quora and Baidu Knows are online platforms to guide users in seeking for answers. A question is posted online. Then, people can post answers based on their expertise and personal experience. Lots of studies are conducted based on online CQA. These studies ranged from artificial intelligent to human behavior studies. However, most of these studies are conducted based on pre-collected data from online CQA data. We argue that the pre-collected CQA data is not suitable for real-life applications. This is because the pre-collected data is only a snapshot of the online CQA. Application of pre-collected data in real life application will require the pre-collected data to be updated from time-to-time. Therefore, knowledge in pre-collected data cannot represent the overall knowledge in online CQA.

On the other hand, it is a time consuming and complex process in working on pre-collected data. Besides limitation of knowledge, application of pre-collected CQA data in real-life applications also required the data to be stored in a structured storage which allows it to be query and retrieved. This is a resource consuming task as online CQA constantly receive huge amount of user generated data daily.

Social Information Retrieval (SIR) is the study to apply social media in information retrieval. The process can be conducted by either query and retrieve information from social media, or applying social profile data in performing information retrieval. In this study, a social information retrieval framework is proposed to query and retrieve relevant information from selected online community question answering service.

This study presents a Social Information Retrieval to query and retrieve relevant information from the selected online CQA. Section two explains the background and motivation of the proposed SIR, while section three provides a detailed description of the proposed SIR. The evaluation of the proposed SIR and its result are presented in the section four. The paper is concluded in Section five.

II BACKGROUND AND MOTIVATION

Information Retrieval (IR) is science to gain access and retrieve relevant information from different information sources in order to satisfy information thirst of a user. This definition is shared by Salton (1968), Baeza-Yates and Ribeiro-Neto (2010), and Bouadjenek, Hacid and Bouzeghoub (2015). Boudjenek (2013) defined SIR as "The process of leveraging social information (both social relationships and the social content), extracted from social platforms, to perform an IR task with an objective of better meeting users' needs." Goh and Foo (2007) introduced a platform to share knowledge through Social Information Retrieval. Bouadjenek, Hacid and Bouzeghoub (2015) reported on IR approaches and platforms for social networks. From the review, most of the studies on SIR were on expending the IR model with social network applications. These studies provide fundamentals for involving data from social network in various stages in IR architecture.