Topic Participation Algorithm for Social Search Engine Based on Facebook Dataset

Hao-Ren Yao and I-Hsien Ting

Department of Information Management
National University of Kaohsiung, Taiwan
ghjk3695@gmail.com, iting@nuk.edu.tw

Abstract. With the rapid growth of users in social networking websites, large amount of data are aggregated. Users are tending to find information through their friends on social network such as Facebook, and this behavior leads to a new search paradigm called social search. However, the traditional search engine like Google cannot handle this kind of search. The data cannot be indexed because of the membership privacy setting and social network relationships. Under this situation, it is harder and harder for users to search information related to their social network.

In this paper, we therefore proposed a system architecture which can deal with this issue and using the data from Facebook as example. An algorithm is also proposed which is the core technique of the system which is called Topic Participation Algorithm (TPA). Furthermore, we will propose a novel implemented social search engine which is developed based on the concept of social network analysis, data mining techniques and searching techniques.

Keywords: Search Engine, Social Search, Social Network Analysis, Data Mining, Web Mining.

1 Introduction

With the rapid growth of the internet and the concept of Web2.0, the web has becoming a popular communication platform. Due to the advances of Web2.0 technology, social network sites like Facebook and Twitter make it possible for users to share their information instantaneously. The interactive nature gives users the chance to communicate and describe their daily life experience in a unique way. Take Facebook as an example, users can publish variety of data on their timeline including posts, photos, videos, and check-ins…, etc. The tagging scheme deeply combines the relationship between object and their friends. The like function can help users to express their attention on certain content such as pages, link sharing, or topics. In this situation, it is easy to figure out the users' preference by their activities on Facebook. [4]

In order to analyze the data on Facebook, social network analysis (SNA) [9] is an important research field due to it focuses on the analysis of social data and social relation. Social network is used to describe the relationship between people and also display what the role of an actor plays in a group. In particularly, the nature of

L.S.-L. Wang et al. (Eds.): MISNC 2014, CCIS 473, pp. 158–170, 2014.

network can easily detect a group of people who share similar interests by their interaction such as comments, likes, or tags. [8]

The importance of social search has been considered in Academia and Industrial due to the fast development of social network sites like Facebook and Twitter. Social relationship has been considered as the criteria for ranking the search results by some researchers.

The reason is in past years, traditional search engine rank the search result by keyword similarity or google page rank. These scheme has nothing to do with social relationship since no social interaction has been incorporated. [6][11][5]

In this case, our purpose would like to implement the social ranking approach that proposed by us in this paper called topic participation algorithm by using SNA to analyze the relationship on social network graph built on top of the user query. This can further lead to the discovery of not only explicit but also implicit users who have strong connection to the query. And the magnitude of these connections can be used as a basis of so-called "*Topic Participation Algorithm*". We then build up a system which implement this approach to evaluate how the social search runs in the real environment and to understand the cost, the requirement, and the presentation.

The rest of the paper is organized as following: In section 2, literatures that related to cloud computing, clustering and social search are reviewed as well as to provide some background of the paper. Then, the system architecture is introduced in section 3 and the explanation of each component in the architecture and the algorithm. In section 4, screenshots of the main functions of the system are presented to show the ability of the social search system. The paper is finally concluded in section 5.

2 Literature Review

2.1 Cloud Computing

Due to the definition from NIST(National Institute of Standards and Technology) of cloud computing, it is a model implementing the ubiquitous, convenient, and on-demand network access to a computing resources which is managed by the service provider. The computing resources typically are a collection of hardware and software which support the implementation of various service model. The infrastructure of the cloud includes broad network access, distributed storage, and a pool of computing server[10].

The service model includes SaaS(Software as a Service), PaaS(Platform as a Service), and IaaS(Infrastructure as a Service). The SaaS is to provide consumer application software usually based on the web by the provider's cloud infrastructure without knowing any detail about the software. The PaaS makes consumer possible to deploy their own application on the cloud infrastructure by the tools, or programming language provided by the service provider without knowing any detail about the platform. Finally, the IaaS gives consumer the usage on the cloud infrastructure including network, computing servers, or storage without managing or controlling the underlying infrastructure on their own[1].