Chapter 11

Ranking Tagged Resources Using Social Semantic Relevance

Anjali Thukral
University of Delhi, India

Hema Banati
University of Delhi, India

Punam Bedi
University of Delhi, India

ABSTRACT

The WWW today is overwhelmed with information on almost every topic. Therefore, relevance ranking of web pages to a user’s expectations is a challenge, rather than retrieving a collection of thousands of web pages selected by keyword matching. This paper presents an approach to rank tagged web pages retrieved from a Social Bookmarking Site for a learner who needs web resources containing content on a given topic. Besides the popularity of the web page in the community, the relevance of a web page for ranking is computed based on the semantic distance between tags and a given topic using domain ontology. An experimental study has been conducted to evaluate the ranks generated by the proposed approach. The test collection was created using a questionnaire which was designed to judge the crawled web pages for their graded relevance on a topic.

1. INTRODUCTION

The World Wide Web has evolved as a unanimous powerful information portal that allows a learner to retrieve and access available resources or documents on any subject or topic. Various search engines, with their rational design and organizational efficiency retrieve web pages on a user’s query or search topic. However, the resulted web pages are usually very large in number which consequently leaves the user with the only option to search required content by visiting each web page. Such large volume of search results therefore needs a ranking according to the user’s information requirements. The ranking algorithms used to rank search results often relay on the formation of user’s query, and its Boolean and syntactic keyword matching in the web pages’ content. This means that the document / resource or web page relevance is determined by using the words
in the topic or the query which is matched (keyword based) with the content of the web pages. The web page retrieval and ranking algorithms (Yi, Liangjie, Ruihua, Jian-Yun, & Ji-Rong, 2009; Chawla & Bedi, 2008; Greg, Chowdhury, & Torgeson) based on classification, clustering of queries and documents, click logs, terms co-occurrence (Matsuo and Ishizuka 2003), etc., are the few used by many search engines. Apart from these, the algorithms based on the term frequency (TF, IDF) (Salton and Buckley 1987), internal, external links, and clickthrough of HIT count methods (Kleinberg, 1999; Kazunari, Kenji, Masatoshi, & Shunsuke, 2003) with their various modified versions, that uses syntactic keyword matching (Agrahri, Anand, & Riedl, 2008) are also used to search and rank relevant documents. These algorithms work best for a well-formed query with specific keywords because of their syntactic searching potential. However, the ranking criterion built on the algorithms like clickthrough usually gets biased interpretations in selecting links. This is apparent from the study (Chi and Mytkowicz 2008) that shows more than 90% of search sessions do not go beyond the first result page. As the result, most of the time, the users click links available only on first few pages, ignoring various links that exist at the end pages. Moreover, people have also been found biased in the selection of search engines and their search activities (Shenghua, et al. 2007). Besides this, many links that exist at end pages never get a chance to get their HIT. All above studies infer that the users’ clicks are biased in nature and are influenced by many factors like placing few links at the top on the first page of search results, selection of search engine and placing popular keywords in the web pages. However while computing relevance to rank resources, if semantically related terms to the search topic or query and some sort of the community’s feedback regarding web resources are incorporated, the ranking order of resources can be improved.

This paper proposes Social Semantic Ranking (SSR) to rank tagged web resource using Social Semantic Relevance (S2R) which is based on Vector Space Model. It uses tags which are bookmarked in Social Bookmarking Site (SBS), and pre-existing semantic knowledge from ontologies for computing Semantic Relevance of the crawled web resources on a search topic.

The rest of the paper is organized as follows: In Section 2 the related work and some basic features used in the approach have been discussed. The proposed approach for Social Semantic Relevance ranking of tagged resources with respect to a search topic is explained in Section 3. Section 4 presents the implementation of the approach and evaluation of its ranked results in the experimental study. Finally Section 5 concludes the paper.

2. RELATED WORK

Tagging is a very powerful source of information which emerges from a collaborative effort of numerous users of similar community. These tags are usually placed by users after analyzing the content of web resources and therefore can be considered as a useful feedback. This implicit feedback in the form of tags given by the users can be considered more appropriate than the computational based classification approaches where the attempt is made to compute the meaning of a resource. Besides, a collaborative bookmarking system facilitates to locate many relevant resources, which a normal crawler in a search engine is usually unable to find. Using tags as the base for deciding which URL fits best for a concept makes the decision more fruitful. Research on Social annotation (Bischoff, et al. 2008) has shown promising results in this direction. Experimental results (Firan, Nejdl and Paiu 2007) have also shown that the tag based search algorithms yield improved results and faster recommendations, which consequently can improve