Chapter 5

Vertical Result Page Generation for Academic Web Searching:
A Summary-Based Approach

ABSTRACT

Vertical search engines are meant for answering a user's web query within a specific domain such as news, media, and academic web searching. One main difference between vertical and horizontal web searching is that in vertical web searching, unlike horizontal web searching, a subset of entire web is engaged. The chapter investigates the state-of-the-art in academic web searching and points out shortcomings in this particular domain. Lastly, the authors aimed to propose a summary-based recommender to respond to a user's query by retrieving and ranking them according to their similarity merits on the basis of papers' summaries. Results of the evaluations revealed the fact that the proposed framework has outperformed the state-of-the-art in different metrics such as unanimous ranks and F1 measures.

INTRODUCTION

A comprehensive study is always a hard target to achieve considering the huge amount of online documents available. As a result, numerous approaches have been introduced to overcome this difficulty by recommender systems

based on user profile (Hong, K. et al. 2013a, Hong, K. et al. 2013b), using citation data to recommend scholarly documents (Ma. N et al. 2008, Bogers, T. and van den Bosch, A. 2008), a combination of page rank algorithm and citation network (Nykl, M. et al. 2014). In the meantime, we can argue the correctness of these systems based on previous findings (Vellino, A. 2010, He, Q et al. 2010) that using page rank values will not necessarily improve relevancy judgments as well as user profiling in the context of scholarly recommendation would lead to prepare a huge set of data to work with.

In the meantime, a citation-based recommender, i.e., a recommender which considers quality of a paper by counting its citation score, suffers from several issues based on the following facts:

- Pohl, S. et al (2007) showed that a paper needs at least 2 years for gaining enough citation scores for being considered by a citation-based scholarly recommender. This means that a new and relevant paper will be ignored for at least 2 years.
- In addition, *Matthew Effects* (Stanovich, K., E. 1986, Merton, R., K. 1986) would be another critical issue of such systems were a paper with more citation counts would always get more attention and more cites while a relevant paper with less citation score will be otherwise.
- Accuracy will be another spot for a citation-based recommender. This is shown that coverage of a recommender is a critical factor for its accuracy (Good, N. et al. 1999). This is also proven that a citation-based recommender ignores a huge set of documents (He, Q. et al. 2010). As a result, the coverage would be degraded as well.
- These systems also suffer from manual or complicated NLP techniques (Pohl, S. et al 2007).
- Finally, *self-citation* may affect the recommendation process in a negative way where authors and/or coauthor(s) of scientific papers cite each others' publications (Hyland, K. 2003, Tagliacozzo, R. 1977)

Consequently, we are motivated to introduce a recommender which does not suffer from stated issues. We proposed to use summary of a paper as a metric for relevancy judgment. This has been proven that a full-text metric for the aim of scholarly recommendation is slow and time-intensive (He, Q et al. 2010). As a result, using summary of a paper will be a good substitute.

Since our recommender uses textual content of papers, consequently issues of state-of-the-art are avoided.

Related Content

Outsourcing Issues in Web Development

Clif Kussmauland Roger Jack (2008). *Software Engineering for Modern Web Applications: Methodologies and Technologies (pp. 217-238).*

www.irma-international.org/chapter/outsourcing-issues-web-development/29586

Autonomic Healing for Service Specific Overlay Networks

Ibrahim Al-Oqily, Bassam Subaih, Saad Bani-Mohammad, Jawdat Jamil Alshaerand Mohammed Refai (2012). *International Journal of Information Technology and Web Engineering (pp. 46-59).*

www.irma-international.org/article/autonomic-healing-service-specific-overlay/70385

Incremental Learning for Interactive E-Mail Filtering

Ding-Yi Chen, Xue Li, Zhao Yang Dongand Xia Chen (2009). *Agent Technologies and Web Engineering: Applications and Systems (pp. 134-152).*

www.irma-international.org/chapter/incremental-learning-interactive-mail-filtering/5031

Cloud Security Threats and Techniques to Strengthen Cloud Computing Adoption Framework

Nabeel Khanand Adil Al-Yasiri (2016). *International Journal of Information Technology and Web Engineering (pp. 50-64).*

www.irma-international.org/article/cloud-security-threats-and-techniques-to-strengthen-cloud-computing-adoption-framework/164471

Context-Aware Service Provisioning in Next-Generation Networks: An Agent Approach

Vedran Podobnik, Krunoslav Trzecand Gordan Jezic (2009). *Agent Technologies and Web Engineering: Applications and Systems (pp. 19-38).*

www.irma-international.org/chapter/context-aware-service-provisioning-next/5025