UNIVERSITI PUTRA MALAYSIA

DOCUMENT RANKING USING INFORMATION QUALITY CRITERIA IN WEBLOG SEARCH ENGINE

FATEMEH AZIMZADEH

DOCUMENT RANKING USING INFORMATION QUALITY CRITERIA IN WEBLOG SEARCH ENGINE

By

FATEMEH AZIMZADEH

Thesis Submitted to the School of Graduate Studies, University Putra Malaysia, in Fulfilment of the Requirement for the Degree of Doctor of Philosophy

January 2013

To my parents, husband and children.

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirements for the degree of Doctor of Philosophy

DOCUMENT RANKING USING INFORMATION QUALITY CRITERIA IN WEBLOG SEARCH ENGINE

By

FATEMEH AZIMZADEH

January 2013

Chair: Associate Professor Abd Rahman Ramli, PhD

Faculty: Engineering

Social media has revolutionized the Web industry. Weblog medium, fundamentally, is an innovation in personal publishing. It has also come to engender a new form of social interaction on the web. Because much firsthand information is recorded in blog posts, more and more people tend to search their wanted information on blog sites. A major problem is that a weblog includes nontraditional features of the Web pages such as Weblog post, links, tags, and comments. Thus, the use of traditional rank algorithms like PageRank and HITS in general search engines are not appropriate to evaluate the Weblog posts because such algorithms do not consider the blog specific features.

On the other hand, information quality criteria are important factors for the users. From Weblogs, which have unfiltered information without expert peer review, users expect that search engines deliver quality information for their queries. There has

been little framework which consider information quality criteria in the Weblog search engine. This thesis establishes an integrated framework which incorporates information quality criteria into the ranking function of search engine on Persian weblogs. The presented framework rank Weblogs and posts based on the selected information quality criteria. Then, the ranking scores are merged with relevancy in the search engine. A ranking method is developed for the Weblog search engine where the post is considered as the document retrieved. This thesis proposes two ranking functions in the search engine which are combined with the information quality criteria, and then compared with a PageRank based ranking function. The results reveal that combination of quality criteria with relevancy, without suitable weight for each one, does not lead to user's satisfaction. Instead, applying proper weights to both information quality factors and relevancy intelligibly improve the results of the search engine and consequently lead to user satisfaction.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

KEDUDUKAN DOKUMENT DENGAN MENGGUNAKAN KRITERIA KUALITI MAKLUMAT DALAM ENJIN WEBLOG CARIAN

Oleh

FATEMEH AZIMZADEH

Januari 2013

Pengerusi: Profesor Madya Abd Rahman Ramli, PhD

Fakulti: Kejuruteraan

Media sosial telah merevolusikan industri Web. Medium Weblog pada asasnya adalah satu inovasi dalam penerbitan peribadi juga telah menghasilkan satu bentuk interaksi sosial yang baru dalam Web. Disebabkan banyak maklumat yang berterusan direkodkan dalam pos blog, semakin banyak orang ber kecenderungan untuk mencari maklumat yang diperlukan oleh mereka dalam laman blog. Satu masalah utama ialah sesebual Weblog memasukkan sifat-sifat yang bukan tradisi. Contohnya adalad pos Weblog, capaian, para tanda dan komen. Sejurus itu, algoritma tradisi dalam enjin carian yang biasa tidak sesuai untuk menilai pos-pos Weblog kerana algoritma tersebut tidak mempertimbangkan sifat blog yang spesifik. Selain daripada itu kriteria kualiti maklumat adalah satu faktor yang penting untuk pengguna. Tidak seperti Weblog yang maklumatnya tidak disaring tanpa ulasan teliti daripada pakar, pengguna-pengguna menjangkakan enjin carian itu mengeluarkan maklumat yang ber kualiti mengikut kehendak mereka. Setakat ini, walau bagaimanapun, telahpun ada beberapa rangka kerja yang mempertimbangkan

kriteria kualiti maklumat dalam enjin carian Weblog. Tesis ini memperkenalkan satu integrasi rangka kerja yang menggabungkan kriteria kualiti maklumat ke dalam fungsi kedudukan enjin carian dalam Weblog Parsi. Sistem yang dibangunkan ini adalah enjin carian pertama yang didedikasikan kepada Weblog Persian. Kedudukan rangka kerja Weblog dan pos yang dibentangkan terdiri daripada kriteria kualiti maklumat yang terpilih. Kemudian, skor kedudukan itu digabungkan dengan perkaitan dalam enjin carian itu. Tesis ini mencadangkan dua fungsi kedudukan dalam enjin pencarian yang digabungkan dengan kriteria kualiti maklumat, dan kemudian membandingkan mereka dengan fungsi PageRank berdasarkan kedudukan. Keputusan menunjukkan balawa penggabungan kriteria kualiti dengan perkaitan tanpa pemberat yang sesuai selalunya tidak memenuhi kehendak pengguna. Sebaliknya, penggunaan pemberat yang sesuai ke atas kedua-dua faktor kualiti maklumat dan perkaitan secara logiknya meningkatkan hasil carian enjin dan akibatnya menjurus kepada kepuasan pengguna.

ACKNOWLEDGEMENTS

Acknowledgement is not a play of words, but an attitude of mind. If words are considered as the symbol of approval and tokens of appreciation, then let the words play the heralding role to expressing my gratitude.

Thanks to Allah, who with His willing give me the opportunity to complete this thesis.

I am heartily thankful to my supervisor, Associate Professor Abdul Rahman Ramli, whose encouragement, guidance and support me; I would like to express my deepest thanks co- supervisors Professor Borhanuddin Mohd Ali and Professor Hamidah Ibrahim for their advice and constant support.

Deepest thanks and appreciation to my husband, parents, family, my children, and others for their cooperation, encouragement, constructive suggestion and full of support for the report completion, from the beginning till the end. Also thanks to all of my friends and everyone, those have been contributed by supporting my work and help myself during this four years thesis progress till it is fully completed.

I certify that a Thesis Examination Committee has met on 31 January 2013 to conduct the final examination of Fatemeh Azim Zadeh on her thesis entitled "Document Ranking Using Information Quality Criteria in Weblog Search Engine" in accordance with the Universities and University College Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Examination Committee were as follows:

M. Iqbal bin Saripan, PhD
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Chairman)

Wan Azizun binti Wan Adnan, PhD
Senior Lecturer
Faculty of Engineering
Universiti Putra Malaysia
(Internal Examiner)

Syamsiah binti Mashohor, PhD
Senior Lecturer
Faculty of Engineering
Universiti Putra Malaysia
(Internal Examiner)

Sachio Hirokawa, PhD
Professor
Kyushu University
Japan
(External Examiner)

SEOW HENG FONG, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 21 March 2013

This thesis submitted to the Senate of University Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy.

The members of the Supervisory Committee were as follows:

Abd Rahman Ramli, PhD
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Chairman)

Borhanuddin Mohd Ali, PhD
Professor
Faculty of Engineering
Universiti Putra Malaysia
(Member)

Hamidah Ibrahim, PhD
Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

_____

BUJANG BIN KIM HUAT, PhD
Professor and Dean
School of Graduate Studies
Universiti Pura Malaysia

Date:

DECLARATION

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently, submitted for any other degree at Universiti Putra Malaysia or at any other institution.

_____

FATEMEH AZIMZADEH

Date: 31 January 2013

TABLE OF CONTENTS