UNIVERSITI PUTRA MALAYSIA

COALESCENCE OF XML-BASED REALLY SIMPLE SYNDICATION AGGREGATOR FOR BLOGOSPHERE

TEH PHOEY LEE

FSKTM 2011 28

COALESCENCE OF XML-BASED REALLY SIMPLE SYNDICATION AGGREGATOR FOR BLOGOSPHERE

TEH PHOEY LEE

DOCTOR OF PHILOSOPHY
UNIVERSITI PUTRA MALAYSIA

2011

COALESCENCE OF XML-BASED REALLY SIMPLE SYNDICATION AGGREGATOR FOR BLOGOSPHERE

By

TEH PHOEY LEE

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

April 2011

DEDICATION

To my parents: Lean Hee and Bee Choon,

To my husband: Han Hauh

To my daughters: Ann Merlynn and Anyn Bliss

Phoey Lee

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the Degree of Doctor Philosophy

COALESCENCE OF XML-BASED REALLY SIMPLE SYNDICATION AGGREGATOR FOR BLOGOSPHERE

By

TEH PHOEY LEE

April 2011

Chairman: **Professor Abdul Azim b. Abdul Ghani, PhD**

Faculty: **Computer Science and Information Technology**

Really Simple Syndication (RSS) aggregator has been widely applied onto several applications starting from year 2000, such as news headline, poscasting, education, medical, geospatial and weblogs. The purpose of RSS is to enable users aggregating new content updates on the favorite site which has subscribed in the RSS feeder instead of visiting the sites individually. Blogging over the internet has become a hobby amongst the internet veteran, whether they are politicians, retired teachers, students, lawyers, journalist etc. The usage of RSS aggregator as a tool onto the blogging environment has become the latest form of internet phenomenon. Weblogs written in chronological order will discuss several different topics, while RSS serves as a tool to aggregate new content updates on the site subscribed. However, each of the independent readers have their different interests in several aspects, such as cooking,

computing, football, scholarly literature, political issues and etc. Relevant topics that had been raised by multiple writers from different sites aggregated onto current automated RSS aggregator do not completely satisfy the readers to find relevant topics from multiple websites. In this thesis, two major studies were carried out. The first study involved studying the different formats used in aggregator with aggregated result in terms of the coalescence of their metadata. Second study will cover the issues of ambiguity of the weblogs on the relevant topics aggregated by the user based on user interest. *PheRSS* help in resolving these problems hybriding the technique on the topics. Finally, experiment is done to prove the relevancy of outcome. Result is analyzed and compared and assessment of the RSS aggregated result based on user interest is done.

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

COALESCENCE OF XML-BASED REALLY SIMPLE SYNDICATION AGGREGATOR FOR BLOGOSPHERE

Oleh

TEH PHOEY LEE

April 2011

Pengerusi: **Profesor Abdul Azim b. Abdul Ghani, PhD**

Fakulti: **Sains Komputer dan Teknologi Maklumat**

Really Simple Syndication (RSS) agregator telah diterapkan secara luas ke beberapa aplikasi mulai dari tahun 2000, seperti berita utama, poscasting, pendidikan, kesihatan, Geospasial dan weblog. Objektif RSS adalah untuk membolehkan pengguna menggabungkan kemaskini kandungan baru di laman kegemaran yang telah melanggan dengan RSS feeder selain daripada melawat halaman satu per satu. Menulis blog melalui internet telah menjadi hobi di kalangan internet veteran. Tidak kira mereka adalah ahli politik, pesara guru, mahasiswa, peguam, wartawan dan sebagainya. Penggunaan RSS agregator sebagai alat kepada persekitaran blogging telah menjadi bentuk baru dari fenomena internet. Menulis blog melalui internet telah menjadi hobi di antara penulis. Weblogs ditulis dalam urutan kronologi membahas beberapa topik yang berbeza, sedangkan RSS berfungsi sebagai alat untuk agregat mengemaskini kandungan baru di laman melanggan. Namun, pembaca mempunyai perbincangan dalam beberapa aspek kepentingan yang berbeza masing-masing, sep-

erti memasak, komputer, bola sepak, literatur ilmiah, isu-isu politik dan sebagainya. Topik yang relevan dibincangkan oleh beberapa pengarang daripada laman yang berbeza dihimpunkan ke agregator RSS automatik masa sekarang tidak sepenuhnya memuaskan para pembaca untuk menemukan topik yang relevan daripada beberapa laman Web. Dalam tesis ini, dua kajian telah dilaksanakan. Kajian pertama melibatkan mempelajari pelbagai format yang digunakan dalam agregator dengan keputusan agregasi dalam jangka waktu koalesensi metadata mereka. Studi kedua akan merangkumi masalah kekaburan daripada weblog tentang topik yang relevan dikumpulkan oleh pembaca berdasarkan minat pengguna. PheRSS membantu dalam menyelesaikan masalah-masalah ini menghubungkan teknik tentang topik. Akhirnya, eksperimen dilakukan untuk membuktikan perkaitan keputusan. Keputusan dianalisis dan dibandingkan penilaian daripada keputusan dikumpulkan oleh RSS berdasarkan minat pengguna dilakukan.

ACKNOWLEDGEMENTS

First of all, I want to express my gratitude to my supervisor Prof. Dr. Abdul Azim Bin Abdul Ghani who allowed me to undertake my Ph.D. research in Universiti Putra Malaysia. I owe a debt of thanks to for his guidance, expertise, and meticulous editing for successful completion of this study. I am also indebted to Assoc. Prof Dr. Hamidah Ibrahim for her valuable advice, and her critical reading of the manuscript. I am also fortunate to have the support and guidance from co-adviser, Dr. Rodziah binti Atan.

My gratitude is also extended to my parents, brother and sisters, colleagues and friends, thank you for sharing all the pains and gains throughout the years. My special thanks and gratitude to my husband Chung Han Hauh and my children Ann Merlynn and Anyn Bliss, for their understanding, caring, support and patience.

I must extend my thanks to Madam Isabel for her advice and the support in work of publication. My thanks are also to UCSI University, for the flexi hours provided for researcher and sponsors on any related conferences and publication fees.

I certify that an Examination Committee has met on **8hb April 2011** to conduct the final examination of **Teh Phoey Lee** on her **Doctor of Philosophy** thesis entitled "**A Framework for Aggregating User-Assisted Keyword Search in Really Simple Syndication (RSS) for Relevant Weblogs**." in accordance with Universiti Putra Malaysia (Higher Degree) Act 1980 and Universiti Putra Malaysia (Higher Degree) Regulations 1981. The Committee recommends that the candidate be awarded the relevant degree.

Members of the Examination Committee are as follows:

Ali Mamat, PhD
Associate Professor
Computer Science Department
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Rusli Haji Abdullah, PhD
Associate Professor
Information System Department
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Muhamad Taufik Abdullah, PhD
Lecturer
Multimedia Department
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Prof. Dr. Fabio Antonio Crestani
Faculty of Informatics, University of Lugano (USI)
Via Buffi 13, CH-6904 Lugano, Switzerland
(External Examiner)

_____

NORITAH OMAR, PhD
Associate Professor/ Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia
Date: 5 May 2011

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee are as follows:

Abdul Azim Abd. Ghani, PhD
Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

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

Rodziah binti Atan., PhD
Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

_____

HASANAH MOHD GHAZALI, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra 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, and is not concurrently, submitted for any other degree at University Putra Malaysia or other institutions.

TEH PHOEY LEE

Date: 8 April 2011

TABLE OF CONTENTS