# A WEIGHTED-BASED APPROACH TO PRIORITIZE USER PREFERENCES IN INFORMATION RETRIEVAL

A thesis submitted to the College of Art and Science
In partial fulfillment of the requirement for the master degree
Master of Science (Information Technology)
Universiti Utara Malaysia

By

Mashal Kasem Al-Qudah

**KOLEJ SASTERA DAN SAINS**
**(College of Arts and Sciences)**
**Universiti Utara Malaysia**

**PERAKUAN KERJA KERTAS PROJEK**
***(Certificate of Project Paper)***

Saya, yang bertandatangan, memperakukan bahawa
*(I, the undersigned, certify that)*

**MASHAL KASEM SULIEMAN ALQUDAH**
**(800107)**

calon untuk Ijazah
*(candidate for the degree of)* **MSc. (Information Technology)**

telah mengemukakan kertas projek yang bertajuk
*(has presented his/ her project paper of the following title)*

**A WEIGHTED-BASED APPROACH TO PRIORITIZE USER PREFERENCES IN INFORMATION RETRIEVAL**

seperti yang tercatat di muka surat tajuk dan kulit kertas projek
*(as it appears on the title page and front cover of project paper)*

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan.
*(that the project paper acceptable in form and content, and that a satisfactory knowledge of the field is covered by the project paper).*

Nama Penyelia Utama
*(Name of Main Supervisor):* **DR. YUHANIS YUSOF**

Tandatangan
*(Signature)* : _____ Tarikh (Date) : 18/11/08

Nama Penyelia Kedua
*(Name of 2nd Supervisor):* **ASSOC. PROF. DR. WAN ROZAINI SHEIK OSMAN**

Tandatangan
*(Signature)* : Rozaini Tarikh (Date) : 18/11/08

## PERMISSION TO USE

In presenting this thesis of the requirements for a Master of Science in Information Technology (MSc. IT) from University Utara Malaysia, I agree that the University library may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by my supervisor or in their absence, by the Dean of Graduate School. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to University Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Request for permission to copy or make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Research & Post Graduate Studies
University Utara Malaysia
College of Arts and Sciences
06010 Sintok
Kedah Darul Aman

Marshal

# ABSTRACT

Tourism statistics point towards increased tourist or citizen mobility in holiday destinations, with a subsequent rise in associated externalities, above all those related with the use of hire cars. Currently, there are a large number of car rental companies operating through the internet. They offer various services relating to vehicle hiring, nevertheless, these online business are not organize into a single retrieval system. Information retrieval (IR) is the art of searching for text, for information within text and for metadata about text, as well as that of searching the World Wide Web and relational databases. Information systems including the internet uses hypertext format. The critical idea carried out by a weighted based algorithm is that, when one attributes links to another, it is basically casting a vote for the other attribute. The higher the number of votes that are cast for an attribute, the higher the significance of the attribute. In order for customer to get a good bargain, he has to personally visit the car rental companies or their relevant websites and later manually perform the required comparison. Furthermore, the existing online car rental services do not include users' preferences in a rental booking.

The objective of this work is two-fold; to combine information on car rental services from three companies into a single system and to use a weighting scheme to prioritize a customer's preferences on the car to be hired. The proposed approach is intended to help customers to hire the required car from the company that offers the best services. Based on the result presented in the earlier subsections, we concluded that the respondents felt version B is better than version A. This shows that by applying a weighted-based approach, relevancy of the retrieved results is improved and at the same time fulfilled user preferences.

# ACKNOWLEDGEMENT

***By the Name of Allah, the Most Gracious and the Most Merciful***

First of all, I would like to express my gratitude to Allah, the Most Merciful and the Most Compassionate, Who has granted me the strength and will to start and complete this study. I do pray to His Greatness to inspire and enable me to continue the work for the benefits of humanity.

My most profound thank goes to my supervisor Dr. Yuhanis Yosef for her scientifically proven creativity and encouraging guidance. Honestly, she has been all the time a center of inspiration and guidance. I deeply thank her for her support and cooperation as being equipped to provide her best help. My deep appreciations also go to Assoc. Prof. Dr Wan Rozaini Bint Sheik Osman because she is very clever, polite and compassionate, when she supervised me from the beginning till I finish my study, with respect to Dr. Alaa AlMabhouh and all the lecturers whom helped me to collect my data during their classes. May Allah bless all of them.

Last but not least, I wish to thank all my dearest family members, especially my mother, father, lovely brothers and sisters. I dedicate my admiration and thanks to all of them for supporting me to the completion of the thesis. My demonstrative appreciations are also to all my friends, colleagues, all FTM staff, and everyone who has put the hand either directly or indirectly to complete this thesis.

## إهـــداء

إلى من نحتت بدموعها تفاصيل فراقي ، وأضاءت بصلواتها ضلمات غربتي ،
ورسمت بأشواقها خطوات مستقبلي،
الى العظيمه دوما.
أمي الحبيبة

إلى من زرع في قلبي أسمى معاني العطاء، فكان لي نبراسا يشع بالجود والسخاء،
وخفف بدمعه آلام العناء،
إلى الغالي دوما.
والدي الحبيب

إلى ينابيع الانسانيه والعطاء اصحاب المشاعر النبيله، والنفوس الطيبه إلى من
لمحت في عيونهم ألم يأسي، وفرحة نجاحي.

عائلتي الكبيرة

أهدي هذا الجهد المتواضع تقديرا، ومحبة، وعرفانا.

مشعل

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1

# INTRODUCTION

## 1.1INTRODUCTION

This chapter provides a description to the undertaken study. It contains the background of the study, problem statement, and research questions, objectives of the study and scope and limitation of the study.

## 1.2 BACKGROUND TO THE STUDY

Information retrieval (IR) is the art of searching for text, for information within text and for metadata about text, as well as that of searching the World Wide Web and relational databases. It also partly covers the usage of the terms data retrieval, document retrieval, information retrieval, and text retrieval, but each also has its own corpse of literature, theory, praxis and technologies. The detailed information about this topic will be discussed further in literature review.

The importance of hypertext is growing during the last decade. Information systems and the internet uses hypertext format where the data is organized associatively rather than sequentially or relationally. In contrast to regular text, hypertext has a non-linear structure and the techniques of pattern matching for text cannot be directly applied to hypertext (Amir et al., 1997).

The contents of the thesis is for internal user only

Travel & Tours) is compared into a single list that provides users with a wider selection. Objective 1, to develop application using weighted-based approach, and objective 2, to evaluate the proposed approach, have been achieved and explained thoroughly. There were certain limitations during the study in terms of resources are identified to help the future works. Future work could be new lighter version of the system to work on mobile versions of computer.

# BIBLIOGRAPHY


Amir, A., Lewenstein, M., & Lewenstein, N., (1997). Pettern Matching in Hypertext, retrieved on August 20, 2008 available at: http://www.cs.biu.ac.il/~amir/Postscripts/hyper.ps.

Boyer R.S., & Moore J.S., (2003). Fast string searching algorithm, communications of the ACM, Vol.20, No.10, pp.762-772.

Black, P. E., (2006) Dijkstra's algorithm, in Dictionary of Algorithms and Data Structures, U.S. National Institute of Standards and Technology, retrieval at September 25, 2008 Available from: http://www.nist.gov/dads/HTML/dijkstraalgo.html.

Brin, S., & Page, L., (2000). The Anatomy of a Large-Scale Hypertextual Web Search Engine, Computer Science Department, Stanford University, Stanford, CA 94305.

Blanco, R., & Lioma, C., (2007). Random Walk Term Weighting for Information Retrieval, Amsterdam, Netherlands, SIGIR, ACM, pp. 23-27.

Bogaschewsky, R., & Hoppe, U., (1993). What Hypertext can do for Information Retrieval: Intelligent Information Retrieval: The Case of Astronomy and Related Space Sciences, 10.1007/978-0-585-33110-2, Springer Netherlands, pp. 81 - 102.

Colussi L., & Fastest, (1994). pattern matching in strings, Journal of Algorithms, Vol. 16, No. 2, pp 163-189.

Crochemore, M., & Rytter, W., (1994). The Forward Dawg Matching algorithm, Text Algorithms, Oxford University Press, available at: http://www-igm.univ-mlv.fr/~lecroq/string/fdm.html#SECTION00220

Coakes, S.J, & stead, L. (2007). SPSS version 14.0 for windows. Analysing without English. WILEY.

Craswell, N. E., (2000), Methods for Distributed Information Retrieval, ©Nicholas Eric Craswell, Typeset in Palatino by TEX and LATEX2e, August 29, 2008; Available: http://research.microsoft.com/users/nickcr/pubs/craswell_thesis00.pdf.

Deogun, J. S., Sever, H., & Raghavan,V. V., (1998). Structural Abstractions of Hypertext Documents for Web-based Retrieval, Ninth International Workshop, 10.1109/DEXA.1998.707429, IEEE, pp. 385-390.

Dietinger, T., Gütl, C., Maurer, H., & Pivec, M., (). Targeted Information Retrieval, September 16, 2008, Available on: http://www.iicm.edu/iicm_papers/targetedinformationretrieval/ICCE99_target_information.pdf

French, J. C., Knight, J. C., & Powell, A. L., (1997).Applying Hypertext Structures To Software Documentation, Information Processing & Management, Elsevier Science Ltd. Printed in Great Britain 0306-4573/97 $17 +0.00, Pp. 219-231.

Hassan, A. A., (2005). Mixed Heuristic Algorithm for Intelligent String Matching for Information Retrieval, Proceedings of the Sixth International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'05).

Horspool, R.N., (1980) Practical fast searching in strings, Software-Practice and Experience, Vol. 10, No. 6, pp. 501- 506.

Huang, J., Wang, G., & Wang, Z., (2008). Cross-Subject Page Ranking Based on Text Categorization, Zhangjiajie, China, IEEE International Conference on Information and Automation, pp. 363-368.

Hassan, S., Mihalcea, R., & Banea, C., (2006). Random Walk Term Weighting for Improved Text Classification, In Proceedings of TextGraps: 2nd Workshop on Graph Based Methods for Natural Language Processing, ACL, pp. 53-60.

Hawk Rent A Car, (2008). Kuala Lumpur Car Rental, Kuala Lumpur, Malaysia, retrieval July 25, 2008 available at: http://www.kuala-lumpur.ws/hawk.

Islam, Md. R., Sarker, B. D., & Islam, Md. R., (2008). An Effective Term Weighting Method Using Random Walk Model for Information Retrieval, Proceedings of

the International Conference on Computer and Communication Engineering, Kuala Lumpur, Malaysia, IEEE, pp.1357 - 1362.

Krovetz R, & Croft W.B., (1992). In proceeding ACM Transactions on lexical ambiguity and Information Systems, Vol 10,2, pp,115-141.

Kleinberg, J. K., (1998). Authoritative Sources in a Hyperlinked Environment, Proc. ACM-SIAM Symposium on Discrete Algorithms.

Kothari, C. R. (1985). Research Methodology, Methods and Technique. Delhi: Wiley Eastern Limited.

Kendall, K. E. & Kendall, J. (2005). Systems Analysis and Design (Sixth Edition), Prentice Hall.

Lynch, P., Luan, X., Prettyman, M., Mericle, L., Borkmann, E., & Schlaifer, J., (2004). An Evaluation of New and Old Similarity Ranking Algorithms, Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04) 0-7695-2108-8/04 $ 20.00, IEEE, pp. 148 – 149.

Lopresti, D. P., (1996). Robust Retrieval of Noisy Text, 0-8186-7402-4/96 $5.00 0 1996 IEEE Proceedings of ADL '96, August 29, 2008; Available: http://ieeexplore.ieee.org.eserv.uum.edu.my/iel3/3741/10939/00502518.pdf?tp=&arnumber=502518&isnumber=10939.

Mansor, T. A. T., (2007). Tourism. Emerging Malyasia 2007, Oxford Business Group, pp 143-150.

Mayflower, (n.d). Company Profile, Reviewed on 14 / 7 / 2008 avaliabile at: http://www.Mayflowercarrental.com.my/cr_about.asp.

Mosteo, A. R., (2006). Dijkstra's algorithm, retrieval at September 25, 2008 available from: http://coding.derkeiler.com/Archive/Ada/comp.lang.ada/2006-02/msg00138.html.

Marchiori, M., (1997). The Quest for Correct Information on the Web: Hyper Search Engines. The Sixth International WWW Conference (WWW). Santa Clara, USA, April 7-11.

Mamtani, G., & Green, G., (2006). Reliability Risk Evaluation during the Conceptual Design Phase, Czech Technical University in Prague, Acts Polytechnica Vol. 46 No., pp.8-14.

Navigli,R., & Velardi , P., (2003). An Analysis of Ontology-based Query Expansion Strategies. In Proceedings of the 14th European Conference on Machine Learning.

Roberto Navigli and Paola Velardi(2003). An Analysis of Ontology-based Query Expansion Strategies. In Proceedings of the 14th European Conference on Machine Learning.

Russell, S., (1995). Artificial Intelligence A Modern Approach, Prentice-Hall,Inc, (UK).

Rob, P., & Coronel, C., (2009). Database System: Design, Implementation, and Management, Eight Edition, THOMSON: Course Technology, US.

Riloff., E., & Hollaar, L., (1996). Text Databases and Information Retrieval, ACM Computing Surveys, pp.133-136.

Ranade, A., (1998). A Simple Optimal List Ranking Algorithm, 5th International Conference, IEEE, pp. 60 – 64.

Sommerville, I. (2001). Hands-On Prototyping in System Analysis and Design Software Engineering, Addison Wesley.

Salmeron, J. L., & Smarandach, F., (n.d.). Redesigning Decision Matrix Method with an indeterminacy-based inference process, September 27, 2008 from: http://arxiv.org/ftp/cs/papers/0703/0703060.pdf

Shajasa Travel & Tours, (2008). Car Rental Services and Corporate Vehicle Hire, retrieved August 5, 2008 available at: http://www.malaysiacarrental.com/

Spertus, E., (1997). ParaSite: Mining Structural Information on the Web. The Sixth International WWW Conference (WWW). Santa Clara, USA, April 7-11.

Search, F., & Practices. B., (2006). Demystifying relevance and ranking, UC Berkeley, School of Information Management & Systems,

Toms, E. G., (n.d.). Serendipitous Information Retrieval, September 23, 2008, available on: http://www.ercim.org/publication/ws-proceedings/DelNoe01/3_Toms.pdf.

Toms, E. G., (1998). „Information exploration of the third kind: the concept of chance encounters. "A position paper for the CHI 98 Workshop on Innovation and evaluation in Information Exploration Interfaces. (http://www.fxpal.com/chi98ie/submissions/long/toms/index.htm)

Tomassen, S. L., (n.d.). Research on Ontology-Driven Information Retrieval, August 27, 2008; Available: http://folk.ntnu.no/steint/papers/2006-OTM-ADC-SLT.pdf.

Vaishnavi & Kuechler. (2007). Design Research in information system. Retrieved January 28, 2008 from http://www.isworld.org/Researchdesign/drisISworld.htm.

Weiss, R., Bienvenido, V., Mark, A. S., Chanathip, M., Peter, S., Andrzej, D., David, K., & Gifford, H., (1996). A Hierarchical Network Search Engine that Exploits

Content-Link Hypertext Clustering. Proceedings of the 7th ACM Conference on Hypertext. New York.

Xapian, (2007). Stemming Algorithms, Reviewed on 18 / 7 / 2008 avaliabile at: http://xapian.org/docs/stemming.html.

Yan, T. W., Jacobsen, M., Garcia-Molina., H., & Dayal, U., (1996). From user access pattems to dynamic hypertext linking, Computer Networks and ISDN Systems 28 (1996) 1007-1014

Yuwono, B., & Lee, D. L., (1996). Search and Ranking Algorithms for Locating Resources on the World Wide Web, 1063-6382/96, IEEE, pp. 164-171.

Zeinalipour-Yazti, D., Kalogeraki, V., & Gunopulos, D., (2004). Information Retrieval Techniques for Peer-To-Peer Networks, Computing In Science & Engineering, Pp 20-26.

Zhao, Y. & Karypis, G., (2002). Improve Precategorized Collection Retrieval by Using Supervised Term Weighting Schemes, Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC.02), pp 1-6.