Concept based term weighting (CBW) in online second hand car trading

Said Mohammad Al Tahat

University Utara Malaysia

2009

Concept based term weighting (CBW) in online second hand car trading

A thesis submitted to the College of Arts and Sciences in Partial in Fulfillment of the requirement for the degree Master of Science (Information Technology)

Universiti Utara Malaysia

October 2009

By

Said Mohammad Al Tahat

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)

SAID MOHAMMAD ALTAHAT
(801618)

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)

CONCEP BASED TERM WEIGHTING (CBW) IN
ONLINE SECOND HAND CAR TRADING

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): **MR. NURNASRAN PUTEH**

Tandatangan
(Signature)   :

Tarikh
(Date)   :   16 Nov 2009

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this project in any manner, in whole or in part, for scholarly purposes may be granted by my supervisor(s) or, in their absence, by the Dean of the 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 to make other use of materials in this thesis, in Whole or in part, should be addressed to:

<div align="center">

Dean of the Graduate School

Universiti Utara Malaysia

06010 UUM Sintok

Kedah Darul Aman

</div>

ABSTRACT

With the increasing of life demands and challenge people face problems in buy or sale second hand car. Therefore online second hand car by using internet technology is proposed to minimize such challenges and demands, by allowing people to buy or sale car more efficiently and easily. The objective of this work is two-fold; to combine information on online second hand car services from many second hand car into a single system and to use a weighting scheme to prioritize a customer's preferences on the car to be buy or sale The proposed approach is intended help customers in sale or buy the required car from second hand car trading that offers good price.

ACKNOWLEDGME NTS

"In the Name of Allah the Most Gracious and Most Merciful"

First and foremost, I am forever indebted and thankful to Allah for his blessings without which the whole would not have been possible.

I will always be especially thankful to my supervisor Mr. Nurnasran, for his guidance, patience and faith in me, he was always positive, supportive and encouraging, he fostered my academic growth by challenging and inspiring me to each deeper, to learn more, to expand my viewpoint, and to think critically. Yet, he allowed me to express my views openly and to disagree even when I was wrong. Also I would like to thanks all members of CAS college.

The final and the most important acknowledgment are owed to my family specially to my brothers Ahmad, Dr. Yousef and Ali to the source of my light and pleasure to the one who enlightens my life, to my dear mother, to all my sisters and all my brothers.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

1.1 INTRODUCTION

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

1.2 BACKGROUND TO THE STUDY

Information retrieval (IR) is the science 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 in chapter two.

The contents of the thesis is for internal user only

REFERNCES

Amir, A., Lewenstein, M., & Lewenstein, N., (1997). Pattern Matching in Hypertext, retrieved on September 20, 2009 available at: http://www.cs.biu.ac.il/~amir/Postscripts/hyper.ps.

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, 2009 Available from: http://www.nist.gov/dads/HTML/dijkstraalgo.html.

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

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

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.

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

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

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

Craswell, N. E., (2000), Methods for Distributed Information Retrieval, ©Nicholas Eric Craswell, Typeset in Palatino by TEX and LATEX2e, August 29, 2009; Available: http://research.microsoft.com/users/nickcr/pubs/craswell_thesis00.pdf.

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

Davis, F. D., Bagozzi, R. P., and Warshaw, P. R., (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," Management Science, 35, , 982-1003.

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., (1999). Targeted Information Retrieval, September 16, 2009, Available on: http://www.iicm.edu/iicm_papers/targetedinformationretrieval/ICCE99_target_inf ormation.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).

Craswell, N. E., (2000), Methods for Distributed Information Retrieval, ©Nicholas Eric Craswell, Typeset in Palatino by TEX and LATEX2e, August 29, 2009; Available: http://research.microsoft.com/users/nickcr/pubs/craswell_thesis00.pdf.

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

Davis, F. D., Bagozzi, R. P., and Warshaw, P. R., (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," Management Science, 35, , 982-1003.

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., (1999). Targeted Information Retrieval, September 16, 2009, Available on: http://www.iicm.edu/iicm_papers/targetedinformationretrieval/ICCE99_target_inf ormation.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).

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, 2009 available at: http://www.kuala-lumpur.ws/hawk.

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.

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.

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

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

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, 2009; Available: http://ieeexplore.ieee.org.eserv.uum.edu.my/iel3/3741/10939/00502518.pdf?tp=&a rnumber=502518&isnumber=10939.

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.

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.

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

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.

Mosteo, A. R., (2006). Dijkstra's algorithm, retrieval at September 25, 2009 available from:Http://coding.derkeiler.com/Archive/Ada/comp.lang.ada/2006-02/msg00138 .html.

Nielsen, J. (2000). Designing Web Usability: The Practice of Simplicity. Indianapolis: New Riders Publishing. Retrieved: September 2009, Available from: http://www.idemployee.id.tue.nl/g.w.m.rauterberg/lecture notes/ISO9241part11.pdfRanade, A., (1998). A Simple Optimal List Ranking Algorithm, 5th International Conference, IEEE, pp. 60 – 64.

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

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

80

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).

Salmeron, J. L., & Smarandach, F., (2007). Redesigning Decision Matrix Method with an indeterminacy-based inference process, September 27, 2009 from: http://arxiv.org/ftp/cs/papers/0703/0703060.pdf

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

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

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

Toms, E. G., (2000). Serendipitous Information Retrieval, September 23, 2009, 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)

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.

Williams, J, R. (2009). Usability Testing. Retrieved April 1, , from Synergetic Applications web site: http://www.synergeticapplications.com/usability.htm.

Xapian, (2007). Stemming Algorithms, Reviewed on July 8, 2009 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.

Zakos, J., & Verma, B. (2005). Concept-Based term weighting for web information retrieval [Electronic Version]. Retrieved August 9, 2009.

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