Log Mining to Support Web Query Expansions

by

Ngok Man Chan

E-Commerce Technology Master Degree Program

2008



Faculty of Science and Technology
University of Macau

TABLE OF CONTENTS

List of Figures 8
List of Tables 12
Chapter 1: Introduction 14
1.1 The Age of Information Overload 14
1.2 Traditional Query Search Techniques 14
1.3 Improving Query Search Techniques 15
1.4 Organization of this Thesis 16
Chapter 2: Related Works 18
2.1 Traditional Methods for Query Expansions Using Query Log 18
2.2 New Developments 19
Chapter 3: Query Expansion: A Proposed Solution 20
3.1 Query Expansion 20
3.1.1 Local Context Analysis 20
3.1.2 Global Context Analysis 21
3.2 Query feature 22
3.2.1 Query Format 22
3.2.2 Query Log 22
3.2.3 Query Preprocessing 22
3.2.3.1 Stop Words 22
3.2.3.2 Stem Words 23
3.2.3.3 Same Case Letter 23
3.2.4 Query Feature Vector 23
3.3 Query Expansion by Association Rules 25
3.3.1 Query Association Mining: A Conceptual Discussion 25
3.3.1.1 Query Sequence 25
3.3.1.2 Assumption 25
3.3.1.3 Concept 25

3.3.2 Finding all Frequent Itemsets 27
3.3.2.1 Apriori 27
3.3.2.2 FP-Growth 27
3.3.3 Generating Strong Association Rules from Frequent Itemsets 28
3.3.4 Query Association Mining: An Illustration 29
3.3.4.1 Definition of User Sessions 29
3.3.4.2 Using FP-Growth Algorithm to Find Frequent Itemsets 30
3.3.4.3 Generating Strong Association Rules from Frequent Itemsets 32
3.3.4.4 Most Related Items 33
3.3.5 Term Association 33
3.4 Query Expansion by Query Information in the Query Log 35
3.4.1 Query Expansion by Similar Queries 35
3.4.2 Query Expansion by Same URL Queries 35
3.4.3 Query Expansion by Stemming, Stopword Removal Backward Algorithm 36
3.4.4 Query Expansion by Final Query in a Query Session 36
3.4.5 A Mixture of Query Information 37
3.5 Query Expansion by Wordnet 39
3.5.1 WordNet 39
3.5.1.1 Synonym 39
3.5.1.2 Hypernym 39
3.5.1.3 Hyponym 39
3.5.1.4 Siblings 39
3.5.2 Query Expansion 40
3.5.2.1 Selection Rule: the First Noun in a Set 40
3.5.2.2 Multi-term Query Refinement 40
Chapter 4: Experiments 42
4.1 Sample Query Log Adopted in this Study 42
4.2 Performance Evaluation 42
4.2.1 Rate of relevant documents for the Original Queries in the Sample Query Log 43

4.2.2 Rate of relevant documents for Queries Expanded by Association Rules from the Sample Query Log 43
4.2.3 Rate of relevant documents for Queries Expanded by Query Information from the Query Log 44
4.2.4 Rate of relevant documents for Queries Expanded by WordNet....45
4.3 Summary and Discussion of Results 46
4.3.1 WordNet....47
Chapter 5: Conclusions....48
5.1 What Have Been Done....48
5.2 Problems not yet Solved 48
5.2.1 Languages 48
5.2.2 Dictionary 49
5.3 Future Enhancements....49
5.4 Limitation....49
BIBLIOGRAPHY....50
APPENDIX A: Performance of Expanded Query Versus Original Query....53
APPENDIX B: Query Log Statistics 79
APPENDIX C: Sample Queries for Testing....81
APPENDIX D: Stop Words List....84