# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>INNER FIRST PAGE</td>
<td>i</td>
</tr>
<tr>
<td>DECLARATION BY THE SCHOLAR</td>
<td>ix</td>
</tr>
<tr>
<td>SUPERVISOR'S CERTIFICATE</td>
<td>x</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>xi</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>xii</td>
</tr>
<tr>
<td>LIST OF ACRONYMS AND ABBREVIATIONS</td>
<td>xiv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xix</td>
</tr>
<tr>
<td>CHAPTER 1</td>
<td></td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td></td>
</tr>
<tr>
<td>1.1 WWW AND INFORMATION RETRIEVAL TOOLS</td>
<td></td>
</tr>
<tr>
<td>1.1.1 WEB DIRECTORIES</td>
<td></td>
</tr>
<tr>
<td>1.1.2 META SEARCH ENGINE</td>
<td></td>
</tr>
<tr>
<td>1.1.3 SEARCH ENGINE</td>
<td></td>
</tr>
<tr>
<td>1.2 MOTIVATION</td>
<td></td>
</tr>
<tr>
<td>1.2.1 FOCUSED SEARCH ENGINE</td>
<td></td>
</tr>
<tr>
<td>1.2.2 ROLE OF CONTEXT IN FOCUSED SEARCH ENGINE</td>
<td></td>
</tr>
<tr>
<td>1.3 MAIN CONTRIBUTION OF THE THESIS</td>
<td></td>
</tr>
<tr>
<td>1.4 ORGANIZATION OF THESIS</td>
<td></td>
</tr>
<tr>
<td>CHAPTER 2</td>
<td></td>
</tr>
<tr>
<td>LITERATURE REVIEW</td>
<td>7</td>
</tr>
</tbody>
</table>
2.1 WORLD WIDE WEB

2.1.1 GROWTH OF WWW 7
2.1.2 THE SIZE OF WWW (SOME STATISTICS) 9

2.2 INFORMATION RETRIEVAL TOOLS 13

2.2.1 WEB DIRECTORY 15
2.2.2 META SEARCH ENGINE 15
2.2.3 SEARCH ENGINE 17
  2.2.3.1 USER INTERFACE 18
  2.2.3.2 QUERY PROCESSOR 19
  2.2.3.3 INDEX BUILDER 19
  2.2.3.4 RANKING MODULE 20
  2.2.3.5 WEB CRAWLER 21

2.3 USER SEARCH TRENDS ON INTERNET 21

2.4 FOCUSED SEARCH: TECHNIQUES AND EXISTING SYSTEMS 24

2.4.1 LINK STRUCTURE BASED APPROACHES 25
  2.4.1.1 FISH-SEARCH APPROACH 26
  2.4.1.2 SHARK SEARCH APPROACH 28
  2.4.1.3 CATEGORY TAXONOMY BASED APPROACH 29
  2.4.1.4 SIMILARITY COMPUTING BASED FOCUSED APPROACH 30
  2.4.1.5 PAGE RANK BASED SIMILARITY COMPUTING FOCUSED APPROACH 30
  2.4.1.6 HITS ALGORITHM BASED FOCUSED APPROACH 31
  2.4.1.7 AUTOMATIC RESOURCE COMPILER BASED FOCUSED APPROACH 32
  2.4.1.8 CONTEXT GRAPH BASED 33

2.4.2 CONTEXT BASED FOCUSED APPROACH 34
  2.4.2.1 PERSONALIZED SEARCH 34
  2.4.2.2 USING QUERY CONTEXT IN IR 35
  2.4.2.3 CONTEXT SENSITIVE INFORMATION 35
CHAPTER 3

PROPOSED ARCHITECTURE FOR CONTEXT BASED FOCUSED SEARCH ENGINE

3.1 INTRODUCTION

3.2 ARCHITECTURE OF A GENERAL SEARCH ENGINE

3.3 MOTIVATION FOR IMPROVEMENT

3.3.1 CONTEXTUAL SENSE: MEANING SCOPE AND EXAMPLES

3.3.1.1 SCOPE

3.3.1.2 MOTIVATING EXAMPLES

3.3.2 IMPROVEMENT IN EXISTING ARCHITECTURE

3.4 PROPOSED MODIFIED ARCHITECTURE OF SEARCH ENGINE

3.4.1 THE TOP LAYER

3.4.1.1 END USER MODE

3.4.1.1.1 USER INTERFACE

3.4.1.2 ADMINISTRATOR MODE

3.4.2 THE MIDDLE LAYER

3.4.2.1 INDEXER

3.4.2.1.1 CONTEXT BASED INDEX STRUCTURE

3.4.2.2 BACK-LINK EXTRACTOR

3.4.3 THE BOTTOM LAYER

3.4.3.1 CRAWL MANAGER

3.5 DATA FLOW IN VARIOUS COMPONENTS OF THE ARCHITECTURE
3.6 CONCLUSION

CHAPTER 4

CONTEXT BASED RELEVANCE EVALUATION OF WEB DOCUMENTS

4.1 INTRODUCTION

4.2 THE PROPOSED MECHANISM FOR CONTEXT BASED RELEVANCE CALCULATOR
   4.2.1 KEYWORD EXTRACTOR AND WEIGHT ASSIGNER MODULE
   4.2.2 EXTRACTION OF CONTEXTUAL SENSES USING WORDNET INTERFACE
      4.2.2.1 EXTRACTION OF CONTEXTUAL SENSES: ROLE OF WORDNET
   4.2.3 CONTEXT BASED RELEVANCE CALCULATOR

4.3 EMPIRICAL EVALUATION OF COMPUTING CONTEXT SCORE
   4.3.1 RESULTS (CONTEXT SCORE COMPUTATION OF A WEB PAGE BASED ON VARIOUS CONTEXTUAL SENSES)
   4.3.2 VALIDATION OF CBR CALCULATOR

4.4 CONCLUSION

CHAPTER 5

CONTEXT BASED RANKING OF WEB PAGES

5.1 INTRODUCTION
5.2 RANKING OF WEB DOCUMENTS
5.3 CONTEXT BASED RANKING
5.4 EMPIRICAL EVALUATION OF CONTEXT BASED RANKING
   5.4.1 RANKING WEB PAGES BASED ON CONTEXT SCORE
   5.4.2 COMPARISON OF PROPOSED RANKING MECHANISM
7.3 USER QUERY PROCESSING
   7.3.1 USER INTERFACE 140
   7.3.2 SEARCH MODULE 140
   7.3.3 RANKING MODULE 141
7.4 RESULT ANALYSIS 143
7.5 CONCLUSION 146

CHAPTER 8

CONCLUSION AND SCOPE FOR FUTURE WORK 148

8.1 INTRODUCTION 148
8.2 CONCLUSION 148
8.3 FUTURE WORK 150

REFERENCES 151

APPENDIX A 161
A.1 COMPLEXITY ANALYSIS OF CBR CALCULATOR ALGORITHM 161
   A.1.2 ALGORITHM: CBR CALCULATOR 161
   A.1.2 ILLUSTRATION 162
A.2 COMPLEXITY ANALYSIS OF BACK-LINK EXTRCITION ALGORITHM 163
   A.2.1 ALGORITHM: EXTRACTION OF BACK-LINKS 163
   A.2.2 ILLUSTRATION 163
   A.2.3 COMPLEXITY ANALYSIS 165
A.3 COMPLEXITY ANALYSIS OF BACK-LINK RELEVANCE EVALUATION 166
   A.3.1 ILLUSTRATION 167

APPENDIX B 169
B.1 COMPARISON OF PROPOSED CONTEXT BASED RANKING WITH PAGE RANK 169
   B.1.1 COMPARISON OF URLS ON TOPIC ‘MOUSE’ IN SENSE ‘COMPUTER MOUSE’ 170
   B1.2 COMPARISON OF URLS ON TOPIC ‘MOUSE’ 171
IN SENSE ‘MOUSE RODENT’

B 1.3 COMPARISON OF URLS ON TOPIC ‘CRANE’
IN SENSE ‘CRANE BIRD’ 172

B 1.4 COMPARISON OF URLS ON TOPIC ‘CRANE’
IN SENSE ‘CRANE MACHINE’ 173

B 1.5 COMPARISON OF URLS ON TOPIC ‘JAVA’
IN SENSE ‘JAVA PROGRAMMING LANGUAGE’ 174

B 1.6 COMPARISON OF URLS ON TOPIC ‘JAVA’
IN SENSE ‘JAVA ISLAND’ 175

B 1.7 COMPARISON OF URLS ON TOPIC ‘JAVA’
IN SENSE ‘JAVA COFFEE’ 176

B 1.8 COMPARISON OF URLS ON TOPIC ‘LION’
IN SENSE ‘LION ANIMAL’ 177

B.2 COMPARISON OF PROPOSED CONTEXT BASED RANKING
WITH Google’s RANKING 178

B.2.1 COMPARISON OF URLS POSITIONS ON TOPIC
‘JAVA PROG. LANG.’ 179

B 2.2 COMPARISON OF URLS ON TOPIC ‘JAVA ISLAND’ 180
B 2.3 COMPARISON OF URLS ON TOPIC ‘JAVA COFFEE’ 181
B 2.4 COMPARISON OF URLS ON TOPIC ‘CRANE BIRD’ 182
B 2.5 COMPARISON OF URLS ON TOPIC ‘CRANE MACHINE’ 183
B 2.6 COMPARISON OF URLS ON TOPIC ‘LION ANIMAL’ 184
B 2.7 COMPARISON OF URLS ON TOPIC ‘COLT HORSE’ 185

LIST OF AUTHOR’S PUBLICATIONS 186

SYNOPSIS