

Contents

Contents	vii
List of Figures	ix
List of Tables	x
1 Introduction	1
1.1 Background	1
1.2 Motivation	3
1.3 Literature Review	3
2 Vector Space Model	5
2.1 Building Term Vectors in Document Space	5
2.2 Normalization of Term Vectors	7
2.3 Computation of Similarity between Document and Query	7
3 Related Work	9
3.1 Search Engine	9
3.1.1 Architecture of Search Engine	9
3.2 Ranking Algorithm	10
3.2.1 Description of PageRank	10
3.3 Major Operation	11
3.3.1 Crawling	11
3.3.2 Indexing	11
3.3.2.1 Indexing Documents into Barrels	11
3.3.2.2 Sorting	11
3.3.3 Searching	11
4 Implementing An Information Retrieval System	12
4.1 Implementation	12

4.1.1	Database Creation and Indexing	12
4.2	User Query and Snapshots of Output	13
4.3	Testing And Analysis of Result	15
4.4	Verification	16
5	Lexicon And Grammar	24
5.1	Types of Query	24
6	Conclusion And Future Work	27
6.1	Conclusion	27
6.2	Future Work	27

List of Figures

2.1	Document similarity under the vector space model	8
3.1	Architecture of Search Engine	10
4.1	Query result for: iitg departments	13
4.2	Query result for: message of iitg director	13
4.3	Query result for: iitg director	14
4.4	Query result for: head of departments in iitg	14
4.5	Query result for: faculty in iitk	14
4.6	Query result for: mba admission in iitk	15
5.1	Parse Tree	25

List of Tables

4.1	Result of Query : “iitg departments” from top to bottom ranked order	15
4.2	Result of Query : “message of iitg director” from top to bottom ranked order	16
4.3	Result of Query : “iitg director” from top to bottom ranked order . . .	17
4.4	Term frequency of word in the documents	18
4.5	Inverse document frequency of words	20
4.6	Term frequency * Inverse document frequency of words	21
4.7	Term frequency of the query	22
4.8	Termfrequency * Inverse document frequency of the query	23