

# Contents

<b>Preface</b>	<b>ix</b>
<b>Chapter 1. Introduction to Web Search Engines</b>	<b>1</b>
1.1 A Short History of Information Retrieval	1
1.2 An Overview of Traditional Information Retrieval	5
1.3 Web Information Retrieval	9
<b>Chapter 2. Crawling, Indexing, and Query Processing</b>	<b>15</b>
2.1 Crawling	15
2.2 The Content Index	19
2.3 Query Processing	21
<b>Chapter 3. Ranking Webpages by Popularity</b>	<b>25</b>
3.1 The Scene in 1998	25
3.2 Two Theses	26
3.3 Query-Independence	30
<b>Chapter 4. The Mathematics of Google's PageRank</b>	<b>31</b>
4.1 The Original Summation Formula for PageRank	32
4.2 Matrix Representation of the Summation Equations	33
4.3 Problems with the Iterative Process	34
4.4 A Little Markov Chain Theory	36
4.5 Early Adjustments to the Basic Model	36
4.6 Computation of the PageRank Vector	39
4.7 Theorem and Proof for Spectrum of the Google Matrix	45
<b>Chapter 5. Parameters in the PageRank Model</b>	<b>47</b>
5.1 The $\alpha$ Factor	47
5.2 The Hyperlink Matrix $H$	48
5.3 The Teleportation Matrix $E$	49
<b>Chapter 6. The Sensitivity of PageRank</b>	<b>57</b>
6.1 Sensitivity with respect to $\alpha$	57

6.2	Sensitivity with respect to $\mathbf{H}$	62
6.3	Sensitivity with respect to $\mathbf{v}^T$	63
6.4	Other Analyses of Sensitivity	63
6.5	Sensitivity Theorems and Proofs	66
<b>Chapter 7. The PageRank Problem as a Linear System</b>		<b>71</b>
7.1	Properties of $(\mathbf{I} - \alpha\mathbf{S})$	71
7.2	Properties of $(\mathbf{I} - \alpha\mathbf{H})$	72
7.3	Proof of the PageRank Sparse Linear System	73
<b>Chapter 8. Issues in Large-Scale Implementation of PageRank</b>		<b>75</b>
8.1	Storage Issues	75
8.2	Convergence Criterion	79
8.3	Accuracy	79
8.4	Dangling Nodes	80
8.5	Back Button Modeling	84
<b>Chapter 9. Accelerating the Computation of PageRank</b>		<b>89</b>
9.1	An Adaptive Power Method	89
9.2	Extrapolation	90
9.3	Aggregation	94
9.4	Other Numerical Methods	97
<b>Chapter 10. Updating the PageRank Vector</b>		<b>99</b>
10.1	The Two Updating Problems and their History	100
10.2	Restarting the Power Method	101
10.3	Approximate Updating Using Approximate Aggregation	102
10.4	Exact Aggregation	104
10.5	Exact vs. Approximate Aggregation	105
10.6	Updating with Iterative Aggregation	107
10.7	Determining the Partition	109
10.8	Conclusions	111
<b>Chapter 11. The HITS Method for Ranking Webpages</b>		<b>115</b>
11.1	The HITS Algorithm	115
11.2	HITS Implementation	117
11.3	HITS Convergence	119
11.4	HITS Example	120
11.5	Strengths and Weaknesses of HITS	122
11.6	HITS's Relationship to Bibliometrics	123
11.7	Query-Independent HITS	124
11.8	Accelerating HITS	126
11.9	HITS Sensitivity	126

<b>Chapter 12. Other Link Methods for Ranking Webpages</b>	<b>131</b>
12.1 SALSA	131
12.2 Hybrid Ranking Methods	135
12.3 Rankings based on Traffic Flow	136
<b>Chapter 13. The Future of Web Information Retrieval</b>	<b>139</b>
13.1 Spam	139
13.2 Personalization	142
13.3 Clustering	142
13.4 Intelligent Agents	143
13.5 Trends and Time-Sensitive Search	144
13.6 Privacy and Censorship	146
13.7 Library Classification Schemes	147
13.8 Data Fusion	148
<b>Chapter 14. Resources for Web Information Retrieval</b>	<b>149</b>
14.1 Resources for Getting Started	149
14.2 Resources for Serious Study	150
<b>Chapter 15. The Mathematics Guide</b>	<b>153</b>
15.1 Linear Algebra	153
15.2 Perron–Frobenius Theory	167
15.3 Markov Chains	175
15.4 Perron Complementation	186
15.5 Stochastic Complementation	192
15.6 Censoring	194
15.7 Aggregation	195
15.8 Disaggregation	198
<b>Chapter 16. Glossary</b>	<b>201</b>
<b>Bibliography</b>	<b>207</b>
<b>Index</b>	<b>219</b>