# Contents


1 Introduction ..... 1
1.1 Information Retrieval ..... 2
1.1.1 Brief History of Information Retrieval ..... 2
1.1.2 “Definition” of Information Retrieval ..... 7
1.2 Retrieval Methods ..... 11
1.3 Modern Algebra ..... 13
1.3.1 Equations ..... 13
1.3.2 Solving by Radicals ..... 14
1.3.3 Birth of Modern Algebra ..... 16
1.3.4 Modern Algebra ..... 18
1.4 Lattice ..... 19
1.5 Importance of Lattices ..... 21
1.6 Lattices in Information Retrieval ..... 22
1.6.1 Retrieval Systems ..... 22
1.6.2 Boolean Retrieval ..... 23
1.6.3 Vector Space Retrieval ..... 23
1.6.4 Fuzzy-Algebra-Based Retrieval Methods ..... 24
1.6.5 Probabilistic Retrieval ..... 25
1.6.6 Web Retrieval and Ranking ..... 25
1.7 Exercises and Problems ..... 26

2 Mathematics Basics ..... 27
2.1 Elements of Mathematical Logic ..... 28
2.1.1 Proposition ..... 28
2.1.2 Negation ..... 29
2.1.3 Conjunction ..... 29
2.1.4 Disjunction ..... 30
2.1.5 Implication ..... 30
2.1.6 Equivalence ..... 31
2.2 Elements of Set Theory ..... 32
2.2.1 Set ..... 32
2.2.2 Subset ..... 33
2.2.3 Equality of Sets ..... 34

2.2.4 Set Union ..... 34
2.2.5 Set Intersection ..... 35
2.2.6 Set Difference ..... 35
2.2.7 Cartesian Product ..... 36
2.2.8 Set Complement ..... 37
2.2.9 Powerset ..... 37
2.2.10 Cardinality of Set ..... 37
2.2.11 Properties of Set Operations ..... 38
2.3 Elements of Relations Theory ..... 38
2.3.1 Binary Relations ..... 39
2.3.2 Function ..... 40
2.3.3 Predicate ..... 41
2.3.4 Equivalence Relation ..... 41
2.3.5 Ordering Relation ..... 42
2.3.6 Partially Ordered Set ..... 42
2.3.7 Partition ..... 42
2.4 Exercises and Problems ..... 43
2.5 Bibliography ..... 44

3 Elements of Lattice Theory ..... 45
3.1 Lattice ..... 46
3.2 Lattice and Poset ..... 47
3.3 Duality ..... 48
3.4 Hasse Diagram ..... 48
3.5 Complete, Atomic Lattice ..... 50
3.6 Modular Lattice ..... 51
3.7 Sublattice ..... 53
3.8 Distributive Lattice ..... 53
3.9 Complemented, Orthomodular Lattice ..... 56
3.10 Boolean Algebra ..... 59
3.11 Important Lattices ..... 59
3.11.1 Powerset Lattice ..... 60
3.11.2 Lattice of Logical Propositions ..... 60
3.11.3 Lattice of Logical Predicates ..... 60
3.11.4 Lattice of Logical Implications ..... 61
3.11.5 Lattice Types ..... 61
3.12 Exercises and Problems ..... 62
3.13 Bibliography ..... 64

4 Basics of Information Retrieval Technology ..... 65
4.1 Documents ..... 66
4.2 Power Law ..... 66
4.3 Stoplist ..... 71
4.4 Stemming ..... 73
4.5 Inverted File Structure ..... 74
4.6 Term-Document Matrix ..... 76
4.7 General Architecture of a Retrieval System ..... 79
4.8 Elements of Web Retrieval Technology ..... 80
4.8.1 World Wide Web ..... 80
4.8.2 Major Characteristics of the Web ..... 80
4.8.3 General Architecture of a Web Search Engine ..... 84
4.8.4 General Architecture of a Web Metasearch Engine ..... 86
4.9 Measurement of Relevance Effectiveness ..... 87
4.9.1 Relevance ..... 87
4.9.2 Measures ..... 87
4.9.3 Precision-Recall Graph Method ..... 91
4.9.4 Uncertainty of Measurement ..... 93
4.10 Measurement of Search Engine Effectiveness ..... 98
4.10.1 M-L-S Method ..... 99
4.10.2 RP Method ..... 102
4.11 Exercises and Problems ..... 103

5 Lattice-Based Retrieval Systems ..... 105
5.1 Mooers' Model ..... 106
5.1.1 Lattice of Documents ..... 106
5.1.2 Lattice of Unstructured Queries ..... 106
5.1.3 Lattice of Term Hierarchies ..... 107
5.1.4 Lattice of Boolean Queries and Documents ..... 108
5.2 The FaIR System ..... 110
5.3 Galois (Concept) Lattice-Based Models ..... 112
5.3.1 Galois (Concept) Lattice ..... 112
5.3.2 Term-Document Matrix and Concept Lattice ..... 113
5.3.3 BR-Explorer System ..... 115
5.3.4 Rajapakse-Denham System ..... 115
5.3.5 The FooCA System ..... 116
5.3.6 Query Refinement, Thesaurus Representation ..... 116
5.4 Properties of the Lattices Applied ..... 117
5.5 Exercises and Problems ..... 123

6 Boolean Retrieval ..... 125
6.1 Boolean Retrieval Method ..... 126
6.2 Technology of Boolean Retrieval ..... 128
6.3 Lattice-Based Boolean Retrieval ..... 129
6.4 Exercises and Problems ..... 132

7 Lattices of Subspaces and Projectors ..... 135
7.1 Metric Space ..... 136
7.2 Complete Metric Space ..... 137
7.3 Linear Space ..... 139
7.4 Subspace of Linear Space ..... 141
7.5 Linear Operator ..... 142
7.6 Banach Space ..... 143
7.7 Hilbert Space ..... 145
7.8 Euclidean Space ..... 146
7.9 Projection Theorem ..... 147
7.10 Projector ..... 149
7.11 Basis of Subspace ..... 151
7.12 Lattice of Subspaces ..... 152
7.13 Exercises and Problems ..... 153
7.14 Bibliography ..... 154

8 Vector Space Retrieval ..... 157
8.1 Introduction ..... 158
8.2 Lattices in Vector Space Retrieval ..... 159
8.2.1 Vector Space Retrieval ..... 159
8.2.2 Technology of Vector Space Retrieval ..... 163
8.3 Calculation of Meaning Using the Hilbert Lattice ..... 165
8.3.1 Queries with Negation ..... 165
8.3.2 Queries with Disjunction ..... 166
8.4 Compatibility of Relevance Assessments ..... 167
8.5 Vector Space Retrieval: Lattice-Lattice Mapping ..... 168
8.6 Discussion ..... 173
8.6.1 Query Lattice and Free Will ..... 173
8.6.2 Vector Space Retrieval? ..... 173
8.6.3 Vector Space Retrieval and Quantum Mechanics ..... 174
8.7 Exercises ..... 177

9 Fuzzy Algebra-Based Retrieval ..... 179
9.1 Elements of Tensor Algebra ..... 180
9.2 Similarity Measure and Scalar Product ..... 182
9.3 Latent Semantic Indexing Retrieval ..... 186
9.3.1 Eigenvalue, Eigenvector ..... 186
9.3.2 Singular Value Decomposition ..... 188
9.3.3 Latent Semantic Indexing ..... 188
9.4 Generalized Vector Space Retrieval ..... 191
9.5 Principle of Invariance ..... 192
9.6 Elements of Fuzzy Set Theory ..... 193
9.6.1 Fuzzy Set ..... 193
9.6.2 Fuzzy Intersection ..... 195
9.6.3 Fuzzy Union ..... 195
9.6.4 Fuzzy Complement ..... 195
9.6.5 Fuzzy Subset ..... 195
9.7 Retrieval Using Linear Space ..... 196
9.8 Fuzzy Algebra-Based Retrieval Methods ..... 199
9.8.1 Fuzzy Jordan Measure ..... 200
9.8.2 Fuzzy Entropy Retrieval Method ..... 203
9.8.3 Fuzzy Probability Retrieval Method ..... 204
9.8.4 Experimental Results ..... 206
9.9 Discussion ..... 207
9.9.1 More on Measures ..... 207
9.9.2 More on Algebra, Entropy, and Probability ..... 208
9.9.3 Information Retrieval and Integration Theory ..... 209
9.9.4 Principle of Invariance and String Theory ..... 210
9.10 Exercises and Problems ..... 212

10 Probabilistic Retrieval ..... 215
10.1 Elements of Probability Theory ..... 216
10.2 Principles of Probabilistic Retrieval ..... 218
10.3 Probabilistic Retrieval Method ..... 220
10.4 Language Model Retrieval Method ..... 224
10.5 Lattice Theoretical Framework for Probabilistic Retrieval ..... 226
10.6 Bayesian Network Retrieval ..... 231
10.7 Exercises ..... 235

11 Web Retrieval and Ranking ..... 237
11.1 Web Graph ..... 238
11.2 Link Structure Analysis ..... 246

11.3 The PageRank Method ..... 249
11.4 The HITS Method ..... 255
11.4.1 Application of the HITS Method in Web Retrieval ..... 257
11.4.2 Latent Semantic Indexing and HITS ..... 259
11.5 The SALSA Method ..... 260
11.6 The Associative Interaction Method ..... 263
11.6.1 Artificial Neural Networks ..... 263
11.6.2 Associative Interaction Method ..... 266
11.6.3 Application of the Associative Interaction Method in Web Retrieval ..... 270
11.7 Combined Methods ..... 270
11.7.1 Similarity Merge ..... 271
11.7.2 Belief Network ..... 272
11.7.3 Inference Network ..... 274
11.7.4 Aggregated Method ..... 274
11.8 Lattice-Based View of Web Ranking ..... 282
11.8.1 Web Lattice ..... 282
11.8.2 Chain ..... 283
11.8.3 Ranking ..... 284
11.8.4 Global Ranking ..... 284
11.8.5 Structure-Based Ranking ..... 288
11.9 P2P Retrieval ..... 292
11.9.1 P2P Network ..... 292
11.9.2 Information Retrieval ..... 293
11.9.3 Lattice-Based Indexing ..... 296
11.10 Exercises and Problems ..... 298

Solutions to Exercises and Problems ..... 301

Reference ..... 307

Index ..... 321

http://www.springer.com/978-3-540-77658-1

The Modern Algebra of Information Retrieval

Dominich, S.

2008, XIV, 330 p., Hardcover
ISBN: 978-3-540-77658-1