

Contents

1	Information Retrieval System Functions	1
1.1	Introduction	1
1.1.1	Primary Information Retrieval Problems	3
1.1.2	Objectives of Information Retrieval System	6
1.2	Functional Overview of Information Retrieval Systems	10
1.2.1	Selective Dissemination of Information	11
1.2.2	Alerts	12
1.2.3	Items and Item Index	13
1.2.4	Indexing and Mapping to a Taxonomy	13
1.3	Understanding Search Functions	14
1.3.1	Boolean Logic	15
1.3.2	Proximity	16
1.3.3	Contiguous Word Phrases	17
1.3.4	Fuzzy Searches	18
1.3.5	Term Masking	18
1.3.6	Numeric and Date Ranges	19
1.3.7	Vocabulary Browse	20
1.3.8	Multimedia Search	20
1.4	Relationship to Database Management Systems	20
1.5	Digital Libraries and Data Warehouses	22
1.6	Processing Subsystem Overview	24
1.7	Summary	25
1.8	Exercises	26
2	Data Structures and Mathematical Algorithms	27
2.1	Data Structures	27
2.1.1	Introduction to Data Structures	27
2.1.2	Inverted File Structure	29
2.1.3	N-Gram Data Structures	31
2.1.4	PAT Data Structure	34
2.1.5	Signature File Structure	38

2.1.6	Hypertext and XML Data Structures	40
2.1.7	XML	43
2.2	Mathematical Algorithms	44
2.2.1	Introduction	44
2.2.2	Bayesian Mathematics	45
2.2.3	Shannon's Theory of Information	47
2.2.4	Latent Semantic Indexing	48
2.2.5	Hidden Markov Models	53
2.2.6	Neural Networks	56
2.2.7	Support Vector Machines	58
2.3	Summary	59
2.4	Exercises	60
3	Ingest	63
3.1	Introduction to Ingest	63
3.2	Item Receipt	64
3.3	Duplicate Detection	67
3.4	Item Normalization	71
3.5	Zoning and Creation of Processing Tokens	72
3.6	Stemming	76
3.6.1	Introduction to the Stemming Process	77
3.6.2	Porter Stemming Algorithm	79
3.6.3	Dictionary Look-Up Stemmers	80
3.6.4	Successor Stemmers	81
3.6.5	Conclusions on Stemming	83
3.7	Entity Processing	84
3.7.1	Entity Identification	85
3.7.2	Entity Normalization	86
3.7.3	Entity Resolution	86
3.7.4	Information Extraction	87
3.8	Categorization	88
3.9	Citational Metadata	92
3.10	Summary	92
3.11	Exercises	93
4	Indexing	95
4.1	What is Indexing	95
4.1.1	History	96
4.1.2	Objectives	97
4.2	Manual Indexing Process	99
4.2.1	Scope of Indexing	99
4.2.2	Precoordination and Linkages	100
4.3	Automatic Indexing of Text	102
4.3.1	Statistical Indexing	105
4.3.2	Natural Language	120
4.3.3	Concept Indexing	125

4.4	Automatic Indexing of Multimedia	129
4.4.1	Introduction to Multimedia Indexing	130
4.4.2	Audio Indexing	131
4.4.3	Image Indexing	134
4.4.4	Video Indexing	136
4.5	Summary	137
4.6	Exercises	139
5	Search	141
5.1	Introduction	141
5.2	Similarity Measures and Ranking	142
5.2.1	Similarity Measures	144
5.3	Hidden Markov Models Techniques	152
5.4	Ranking Algorithms	153
5.5	Relevance Feedback	154
5.6	Selective Dissemination of Information Search	157
5.7	Weighted Searches of Boolean Systems	163
5.8	Multimedia Searching	167
5.9	Summary	168
5.10	Exercises	170
6	Document and Term Clustering	171
6.1	Introduction to Clustering	171
6.2	Thesaurus Generation	174
6.2.1	Manual Clustering	175
6.2.2	Automatic Term Clustering	176
6.3	Item Clustering	184
6.4	Hierarchy of Clusters	186
6.4.1	Automatic Hierarchical Cluster Algorithms	189
6.5	Measure of Tightness for Cluster	193
6.6	Issues with Use of Hierarchical Clusters for Search	194
6.7	Summary	195
6.8	Exercises	197
7	Information Presentation	199
7.1	Information Presentation Introduction	199
7.2	Presentation of the Hits	199
7.2.1	Sequential Listing of Hits	200
7.2.2	Cluster View	201
7.2.3	Network View	205
7.2.4	Timeline Presentation	208
7.3	Display of the Item	210
7.3.1	Indicating Search Terms in Display	210
7.3.2	Text Summarization	211
7.4	Collaborative Filtering	213
7.4.1	Page Ranking as Collaborative Filtering	215

7.5	Multimedia Presentation	216
7.5.1	Audio Presentation	216
7.5.2	Image Item Presentation	219
7.5.3	Video Presentation	223
7.6	Human Perception and Presentation	225
7.6.1	Introduction to Information Visualization	226
7.6.2	Cognition and Perception	229
7.7	Summary	233
7.8	Exercises	234
8	Search Architecture	235
8.1	Index Search Optimization	235
8.1.1	Pruning the Index	236
8.1.2	Champion Lists	236
8.2	Text Search Optimization	237
8.2.1	Software Text Search Algorithms	239
8.2.2	Hardware Text Search Systems	244
8.3	GOOGLE Scalable Multiprocessor Architecture	249
8.4	Summary	251
8.5	Exercises	252
9	Information System Evaluation	253
9.1	Introduction to Information System Evaluation	253
9.2	Measures Used in System Evaluations	259
9.3	Multimedia Information Retrieval Evaluation	269
9.4	Measurement Example: TREC Evolution	271
9.5	Summary	279
9.6	Exercises	280
	Bibliography	283
	Index	301



<http://www.springer.com/978-1-4419-7715-1>

Information Retrieval Architecture and Algorithms

Kowalski, G.

2011, XII, 305 p., Hardcover

ISBN: 978-1-4419-7715-1