

Table of Contents

1. Web Intelligence (WI):

A New Paradigm for Developing the Wisdom Web and Social Network Intelligence

Ning Zhong, Jiming Liu, and Yiyu Yao	1
1.1 Introduction	1
1.2 The Wisdom Web	2
1.2.1 A Minimalist Wisdom Web Scenario	2
1.2.2 Fundamental Capabilities of the Wisdom Web	3
1.3 Levels of WI vs. Social Intelligence	4
1.3.1 Levels of WI vs. Social Intelligence	4
1.3.2 Social Network Intelligence for Enterprise Portals	6
1.4 Extensional Description of WI	7
1.5 An Overview of This Book	9
1.6 Conclusions	15
References	15

Part I. Web Agents

2. Agent-Based Characterization of Web Regularities

Jiming Liu, Shiwu Zhang, and Yiming Ye	19
2.1 Introduction	19
2.1.1 Empirical Regularities on the World-Wide Web	19
2.1.2 Regularity Characterization	20
2.2 Problem Statements	20
2.3 An Overview of Foraging Agent-Based Web Characterization	21
2.4 Foraging Agents	21
2.4.1 Artificial Web Space	21
2.4.2 Interests of Foraging Agents	23
2.4.3 Motivational Support Aggregation	23
2.4.4 Characterization of Foraging Decisions	24
2.4.5 Motivational Support	25
2.5 An Outline of the Algorithm	26
2.6 Experiments	27

VIII Table of Contents

2.6.1	A Comparison with Real-World Log Data.....	29
2.6.2	Further Remarks	30
2.7	Conclusions	33
	References	35
 3. Agent-Based Composite Services in DAML-S: the Behavior-Oriented Design of an Intelligent Semantic Web		
	Joanna J. Bryson, David Martin, Sheila A. McIlraith, and Lynn Andrea Stein	37
3.1	Introduction: Intelligence and the Semantic Web	37
3.2	Definitions: Agents and Services	39
3.3	Bringing Services onto the Semantic Web	40
3.3.1	DAML-S Processes	42
3.4	Semantic Web Development and Software Agent Architecture	44
3.4.1	Modularity.....	45
3.4.2	Action Selection	46
3.5	Web Services as Agent Behavior.....	47
3.5.1	Services as Behavior Modules	47
3.5.2	Composite Services as Action Selection	47
3.5.3	Program, Agent, or Multi-Agent System?	48
3.6	Implications for DAML-S	49
3.6.1	Data	49
3.6.2	Primitives	50
3.6.3	Sequences	50
3.6.4	Basic Reactive Plans	51
3.6.5	Agent-Level Control	52
3.7	Conclusions	53
3.8	Appendix A – Basic Reactive Plans	54
	References	56
 4. Designing Scenarios for Social Agents		
	Toru Ishida and Hideyuki Nakanishi	59
4.1	Introduction	59
4.2	Describing Scenarios	60
4.2.1	Overview	60
4.2.2	Cue and Action	60
4.2.3	Guarded Command	61
4.2.4	Scenario	62
4.2.5	Agent	63
4.3	Q for Legacy Agents	63
4.3.1	Microsoft Agents	63
4.3.2	FreeWalk Agents	67
4.4	Designing Scenarios	68
4.4.1	Q Architecture	68
4.4.2	Design Process	69



X Table of Contents

6.6.2	Efficiency	122
6.7	Related Work	123
6.8	Conclusions	125
	References	125
7. Discovery of Indirect Associations from Web Usage Data		
	Pang-Ning Tan and Vipin Kumar	128
7.1	Introduction	128
7.1.1	Related Work	132
7.2	Preliminaries	133
7.2.1	Definition	133
7.2.2	NonSequential Indirect Association	135
7.2.3	Sequential Indirect Association	137
7.3	Implementation	139
7.3.1	The INDIRECT Algorithm	139
7.3.2	Combining Indirect Associations	140
7.4	Experimental Evaluation	142
7.4.1	Non-sequential Indirect Association	142
7.4.2	Sequential Indirect Association	144
7.4.3	Performance	146
7.4.4	Threshold Selection	148
7.5	Conclusions	148
	References	150
8. Knowledge-Based Wrapper Induction for Intelligent Web Information Extraction		
	Jaeyoung Yang and Joongmin Choi	153
8.1	Introduction	153
8.1.1	Classification of Wrapper Generation	153
8.1.2	Our Approach	155
8.2	XTROS System Overview	156
8.3	Domain Knowledge Specification by XML	157
8.4	Knowledge-Based Wrapper Generation	160
8.4.1	Converting HTML Sources into Logical Lines	160
8.4.2	Determining the Meaning of Logical Lines	160
8.4.3	Finding the Most Frequent Pattern	161
8.4.4	Constructing an XML-Based Wrapper	163
8.4.5	Interpreting the Wrapper	164
8.5	Implementation and Evaluation	164
8.6	Conclusions	171
	References	171

9. Web Log Mining

Zhiyong Lu, Yiyu Yao, and Ning Zhong	173
9.1 Introduction	173
9.2 Overview of Web Mining	174
9.2.1 Classification of Web Mining	174
9.2.2 Web Content Mining	175
9.2.3 Web Structure Mining	176
9.2.4 Web Usage/Log Mining	176
9.2.5 Combinations of Web Content, Structure, and Usage Mining	177
9.3 Data Preparation	177
9.3.1 Data Collection	178
9.3.2 Data Preprocessing	180
9.3.3 Data Abstraction	180
9.4 Data Mining and Pattern Analysis	183
9.4.1 Statistical Information	183
9.4.2 Association Rules	183
9.4.3 Classification and Clustering	184
9.4.4 Sequential Patterns	185
9.4.5 Dependency Modeling	185
9.4.6 Data Warehousing and OLAP	185
9.4.7 Pattern and Rule Evaluation	186
9.5 Applications	186
9.5.1 Web Pre-fetching and Caching	187
9.5.2 Improved Website Design and Organization	187
9.5.3 Web Personalization and Recommendation	187
9.5.4 Adaptive Websites and Pages	188
9.5.5 Intelligent Web Agents	189
9.6 Conclusions	189
References	189

Part III. Web Information Retrieval

10. Personalized and Focused Web Spiders

Michael Chau and Hsinchun Chen	197
10.1 Introduction	197
10.1.1 Web Spider Research	197
10.1.2 Applications of Web Spiders	198
10.1.3 Analysis of Web Content and Structure	199
10.1.4 Graph Traversal Algorithms	202
10.2 Web Spiders for Personal Search	203
10.2.1 Personal Web Spiders	203
10.2.2 Case Study	205
10.3 Using Web Spiders to Create Specialized Search Engines	206

XII Table of Contents

10.3.1 Specialized Search Engines	207
10.3.2 Focused Spidering Algorithms for Specialized Search Engines	207
10.3.3 Case Study	208
10.4 Conclusions	211
10.5 Appendix A: URLs of Spiders and Search Engines	212
References	213
 11. Exploiting the Web as Parallel Corpora for Cross-Language Information Retrieval	
Jian-Yun Nie and Jiang Chen	218
11.1 Introduction	218
11.1.1 Query Translation	219
11.1.2 The Need for Parallel Corpora	220
11.2 Mining for Parallel Texts - PTMiner	221
11.2.1 General Principle of Automatic Mining	221
11.2.2 Identification of Candidate Websites	223
11.2.3 File Name Fetching	224
11.2.4 Host Crawling	224
11.2.5 Pair Scan by Names	225
11.2.6 Filtering by Content	226
11.2.7 PTMiner Implementation	227
11.2.8 Generated Corpora	228
11.3 Training Statistical Translation Models on Parallel Corpora	229
11.3.1 Sentence Alignment	229
11.3.2 Processing of Words	230
11.3.3 Model Training	231
11.4 Evaluation of the Translation Models	232
11.5 CLIR Experiments	234
11.5.1 English-French CLIR	234
11.5.2 English-Chinese CLIR	235
11.5.3 Discussions	236
11.6 Conclusions	237
References	238

Part IV. Web Knowledge Management

12. Knowledge Representation, Sharing, and Retrieval on the Web	
Philippe Martin	243
12.1 Introduction	243
12.2 Elements and Landmarks of Knowledge Representation and Sharing on the Web	246
12.2.1 Exchange Formats and Programming Interfaces	246

12.2.2	Ontologies and Knowledge Bases	247
12.2.3	Ontology Servers	248
12.2.4	Knowledge Within Web Documents	250
12.3	Requirements for a Viable Semantic Web	252
12.3.1	Need for a Standard Library of Ontological Primitives	252
12.3.2	Need for Expressive Notations	253
12.3.3	Need for High-Level (and Expressive) Notations	254
12.3.4	Need for Lexical/Structural/Ontological Conventions	256
12.3.5	Need for Flexible Ways to Refer to a Category	258
12.3.6	Need for a Shared Natural Language Ontology	261
12.3.7	Need for More Centralization	262
12.4	Mechanisms for Cooperatively Editing a Shared KB	263
12.4.1	Control on Graph Additions	265
12.5	Search Interfaces and Mechanisms	265
12.5.1	Searching Categories and Links	267
12.5.2	Accessing or Adding Graphs Via Generated Interfaces	267
12.5.3	Mechanisms for Searching Graphs	270
12.6	Conclusions	273
	References	275

13. On-To-Knowledge: Semantic Web-Enabled Knowledge Management

	York Sure, Hans Akkermans, Jeen Broekstra, John Davies, Ying Ding, Alistair Duke, Robert Engels, Dieter Fensel, Ian Horrocks, Victor Iosif, Arjohn Kampman, Atanas Kiryakov, Michel Klein, Thorsten Lau, Damyan Ognyanov, Ulrich Reimer, Kiril Simov, Rudi Studer, Jos van der Meer, and Frank van Harmelen	277
13.1	Introduction	277
13.2	Tool Environment for Ontology-Based Knowledge Management ...	279
13.2.1	RDFferret: Full Text Searching plus RDF Querying	280
13.2.2	OntoShare: Community Support	281
13.2.3	Spectacle: Information Presentation	281
13.2.4	OntoEdit: Ontology Development	283
13.2.5	Ontology Middleware Module: Integration Platform	285
13.2.6	OntoView: Change Management for Ontologies	287
13.2.7	Sesame: Repository for Ontologies and Data	288
13.2.8	CORPORUM: Information Extraction	289
13.3	OIL: Inference Layer for the Semantic World-Wide Web	290
13.3.1	Combining Description Logics with Frame Languages	290
13.3.2	Web Interface	291
13.3.3	Layering	292
13.3.4	Current Status	293
13.3.5	Future Developments	293
13.4	Business Applications in Semantic Information Access	294

XIV Table of Contents

13.4.1 On-To-Knowledge Methodology	294
13.4.2 Information Search	296
13.4.3 Skills Management	296
13.4.4 Exchanging Knowledge in a Virtual Organization	296
13.5 Conclusions	297
References	299
 14. Ontology Learning Part One — on Discovering Taxonomic Relations from the Web	
Alexander Maedche, Viktor Pekar, and Steffen Staab	301
14.1 Introduction	301
14.2 Survey of Symbolic Approaches	302
14.2.1 Extraction of Taxonomic Relations	302
14.2.2 Refinement of Taxonomic Relations	304
14.3 Survey of Statistics-Based Approaches	304
14.3.1 Statistics-Based Extraction of Taxonomic Relations	305
14.3.2 Statistics-Based Refinement of Taxonomic Relations	309
14.4 Making Use of the Structure of the Ontology	310
14.4.1 Tree Descending Algorithm	310
14.4.2 Tree Ascending Algorithm	311
14.5 Data and Settings of the Experiments	312
14.6 Evaluation Method	313
14.7 Results	313
14.8 Conclusions	317
References	318
<hr/> Part V. Infrastructure for Web Intelligent Systems <hr/>	
15. Algorithmic Aspects of Web Intelligent Systems	
Dimitrios Kalles, Athanasios Papagelis, and Christos Zaroliagis	323
15.1 Introduction	323
15.2 An Overview of the System	325
15.2.1 User Interface	325
15.2.2 Performance	328
15.2.3 Users and Authentication Techniques	331
15.2.4 Agent's Inference Engine	331
15.3 Algorithms	332
15.3.1 Data Characteristics and Generic Handling Techniques	332
15.3.2 Choosing the Next Document	333
15.3.3 Finding Interesting Object Collections and Predicting Votes by Matching Users	335
15.3.4 Finding an Interesting Documents Collection and Predict- ing Votes Using Naïve Bayes Analysis	336

15.3.5 Matching Related Documents	338
15.4 Conclusions	343
References	343
16. Web Document Prefetching on the Internet	
Xin Chen and Xiaodong Zhang	345
16.1 Introduction: Prefetching at Different Stages	345
16.1.1 DNS Prefetching	345
16.1.2 TCP Connection Prefetching	345
16.1.3 Content Prefetching	346
16.2 Conditions of Content Prefetching	347
16.2.1 History Information for Prefetching	347
16.2.2 Expiration Time	347
16.2.3 Time for Prefetching	347
16.3 Classifying Prefetching Methods	348
16.3.1 Client-Based Prefetching	348
16.3.2 Proxy-Based Prefetching	349
16.3.3 Server-Based Prefetching	350
16.3.4 Cooperative Prefetching	351
16.4 Prefetching Structure and Optimization	353
16.4.1 PPM	353
16.4.2 Longest Repeating Subsequence	355
16.4.3 Popularity-Based PPM	355
16.5 Performance Evaluations on Prefetching	357
16.5.1 Latency Reduction Bounds	357
16.5.2 Prefetching Effects on Networks	358
16.5.3 Tradeoff Analysis	359
16.6 Other Variants of Prefetching	360
16.6.1 Real-Time Prediction	360
16.6.2 Prefetching for Multimedia on the Internet	360
16.6.3 Predict HTTP Requests for Dynamic Content	361
16.7 Related Applications	361
16.7.1 Search Engines	361
16.7.2 Recommender Systems	361
16.8 Conclusions	361
References	362

Part VI. Social Network Intelligence

17. Social Networks: From the Web to Knowledge Management	
Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew Tomkins	367
17.1 Introduction	367

XVI Table of Contents

17.2 Link Analysis of the Web	368
17.3 Communities on the Web	370
17.4 Connectivity and the Diameter of the Web	371
17.5 Fractal Nature of the Web	374
17.6 Social Networks for Knowledge Management	375
17.6.1 Enterprise Knowledge Management	377
17.7 Conclusions	378
References	378
 18. A Ranking Algorithm Based on Graph Topology to Generate Reputation or Relevance	
Josep M. Pujol, Ramon Sangüesa, and Jordi Delgado	380
18.1 Introduction	380
18.2 Social Networks	381
18.3 Ranking Algorithms	383
18.3.1 Overview of <i>Pagerank</i>	383
18.3.2 Overview of <i>HITS</i>	384
18.3.3 Our Proposal: the <i>NodeRanking</i> Algorithm	385
18.3.4 Comparisons	387
18.4 Experiments About Ranking, Reputation, and Relevance	388
18.4.1 Extracting Reputation from Social Networks	388
18.4.2 Extracting Relevance from the Web	391
18.5 Conclusions	392
References	393
 19. Communityware That Facilitates Knowledge Interactions	
Yasuyuki Sumi and Kenji Mase	395
19.1 Introduction	395
19.2 PalmGuide: Personal Tour Assistant	398
19.3 Semantic Map: Visual Explorer of Community Information	401
19.4 AgentSalon: Facilitating Face-to-Face Conversations	403
19.5 Experiments and Evaluation	410
19.6 Related Work	415
19.7 Conclusions	416
References	417
 20. Social Intelligence Design for Web Intelligence	
Toyoaki Nishida	419
20.1 Introduction: Social Intelligence Design for Web Intelligence	419
20.2 Overview of Social Intelligence Design	421
20.2.1 Groups and Communities	421
20.2.2 Issues of Social Intelligence Design	422
20.2.3 Applications of Social Intelligence Design	424

Table of Contents XVII

20.3 The Traveling Conversation Model	425
20.4 A Broadcast-Based Approach	425
20.5 A Conversational Agent-Based Approach	429
20.6 A Smart Environment-Based Approach	431
20.7 Psychological Evaluation	433
20.8 Technical Issues	435
20.9 Conclusions	435
References	436
Author Index	438
Subject Index	439



<http://www.springer.com/978-3-540-44384-1>

Web Intelligence

Zhong, N.; Liu, J.; Yao, Y. (Eds.)

2003, XXIV, 440 p., Hardcover

ISBN: 978-3-540-44384-1