

Contents

List of Figures	xv
List of Tables	xxiii
Preface	xxv
I	KNOWLEDGE GRAPH FUNDAMENTALS
1	Introduction to Knowledge Graphs 3
1.1	Graphs 3
1.2	Representing Knowledge as Graphs 5
1.3	Examples of Knowledge Graphs 10
1.3.1	Example 1: Scientific Publications and Academics 10
1.3.2	Example 2: E-Commerce, Products, and Companies 11
1.3.3	Example 3: Social Networks 12
1.3.4	Example 4: Geopolitical Events 12
1.4	How to Read This Text 14
1.5	Concluding Notes 15
1.6	Software and Resources 15
1.7	Bibliographic Notes 16
1.8	Exercises 18
2	Modeling and Representing Knowledge Graphs 21
2.1	Introduction 21
2.1.1	Resource Description Framework 23
2.1.2	RDF Serializations 26
2.2	RDF Schema 28
2.2.1	RDFS Classes 28
2.2.2	RDFS Properties 30

7	Nontraditional Information Extraction	149
7.1	Introduction	149
7.2	Open Information Extraction	151
7.2.1	KnowItAll	157
7.2.2	TextRunner	158
7.2.3	Evaluating and Comparing Open Information Extraction Systems	159
7.3	Social Media Information Extraction	161
7.3.1	TWICAL	163
7.3.2	TwitIE	164
7.4	Other Kinds of Nontraditional Information Extraction	166
7.5	Concluding Notes	167
7.6	Software and Resources	169
7.7	Bibliographic Notes	170
7.8	Exercises	171
III	KNOWLEDGE GRAPH COMPLETION	
8	Instance Matching	175
8.1	Introduction	175
8.2	Formalism	178
8.3	Why Is Instance Matching Challenging?	179
8.4	Two-Step Pipeline	180
8.4.1	Blocking	180
8.4.2	Similarity	190
8.5	Evaluating the Two-Step Pipeline	194
8.5.1	Evaluating Blocking	195
8.5.2	Evaluating Similarity	197
8.6	Postsimilarity Steps	198
8.6.1	Clustering and Transitive Closure	198
8.6.2	Entity Name System	201
8.7	Formalizing Instance Matching: Swoosh	203
8.8	A Note on Research Frontiers	205
8.9	Data Cleaning beyond Instance Matching	208
8.10	Concluding Notes	212
8.11	Software and Resources	213
8.12	Bibliographic Notes	215

8.13 Exercises	217
9 Statistical Relational Learning	221
9.1 Introduction	221
9.2 Modeling Dependencies	223
9.3 Statistical Relational Learning Frameworks	224
9.3.1 Markov Logic Networks	225
9.3.2 Probabilistic Soft Logic	231
9.4 Knowledge Graph Identification	232
9.4.1 Representing Uncertain Extractions	233
9.4.2 Representing Instance Matching Outputs	233
9.4.3 Enforcement of Ontological Constraints	234
9.4.4 Putting It Together: Probabilistic Distributions over Uncertain Knowledge Graphs	234
9.4.5 A Note on Experimental Performance	235
9.5 Other Applications	235
9.5.1 Collective Classification	235
9.5.2 Link Prediction	236
9.5.3 Social Network Modeling	236
9.6 Advanced Research: Data Programming	236
9.7 Concluding Notes	237
9.8 Software and Resources	238
9.9 Bibliographic Notes	238
9.10 Exercises	239
10 Representation Learning for Knowledge Graphs	241
10.1 Introduction	241
10.2 Embedding Architectures: A Primer	243
10.2.1 Continuous Bag of Words Model	245
10.2.2 Skip-Gram Model	246
10.3 Embeddings beyond Words	246
10.4 Knowledge Graph Embeddings	248
10.4.1 Energy Functions	250
10.5 Influential KGE Systems	251
10.5.1 Structured Embeddings	252
10.5.2 Neural Tensor Networks	254
10.5.3 Translational Embedding Models	256
10.5.4 TransE	256

10.5.5 Other Trans* Algorithms	259
10.6 Extrafactual Contexts	262
10.6.1 Entity Types	263
10.6.2 Textual Data	264
10.6.3 Beyond Text and Concepts: Other Information Sets	265
10.7 Applications	267
10.7.1 Link Prediction	267
10.7.2 Triple Classification	268
10.7.3 Entity Classification	268
10.7.4 Revisiting Instance Matching	269
10.7.5 Other Applications	270
10.8 Concluding Notes	270
10.9 Software and Resources	271
10.10 Bibliographic Notes	272
10.11 Exercises	273
IV ACCESSING KNOWLEDGE GRAPHS	
11 Reasoning and Retrieval	279
11.1 Introduction	279
11.2 Reasoning	281
11.2.1 Description Logics: A Brief Primer	283
11.2.2 Web Ontology Language	284
11.2.3 Sample Reasoning Framework: Protégé	289
11.3 Retrieval	291
11.3.1 Term Frequency and Weighting	292
11.4 Retrieval versus Reasoning	293
11.4.1 Evaluation	295
11.4.2 Sample Information Retrieval Framework: Lucene	300
11.5 Concluding Notes	302
11.6 Software and Resources	303
11.7 Bibliographic Notes	303
11.8 Exercises	305
12 Structured Querying	307
12.1 Introduction	307
12.2 SPARQL	308
12.2.1 Subqueries	310

12.3	Relational Processing of Queries over Knowledge Graphs	311
12.3.1	Triple (Vertical) Stores	312
12.3.2	Property Table Stores	313
12.3.3	Horizontal Stores	314
12.4	NoSQL	316
12.4.1	Key-Value Stores	316
12.4.2	Graph Databases	321
12.4.3	NoSQL Databases with Extreme Scalability	328
12.5	Concluding Notes	330
12.6	Software and Resources	330
12.7	Bibliographic Notes	332
12.8	Exercises	333
13	Question Answering	337
13.1	Introduction	337
13.2	Question Answering as a Stand-Alone Application	339
13.2.1	Learning from Conversational Dialogue: KnowBot	339
13.2.2	Bidirectional Encoder Representations from Transformers	341
13.2.3	Necessity of Knowledge Graphs	345
13.3	Question Answering as Knowledge Graph Querying	346
13.3.1	Challenges and Solutions	347
13.3.2	Template-Based Solutions	355
13.3.3	Evaluation of SQA	356
13.4	Concluding Notes	358
13.5	Software and Resources	358
13.5.1	BERT and Language Model-Based Question Answering	359
13.5.2	HOBBIT	359
13.6	Bibliographic Notes	360
13.7	Exercises	362
V	KNOWLEDGE GRAPH ECOSYSTEMS	
14	Linked Data	367
14.1	Introduction	367
14.1.1	Principle 1: Use Uniform Resource Identifiers for Naming Things	372
14.1.2	Principle 2: Use HTTP Uniform Resource Identifiers	373

14.1.3	Principle 3: Provide Useful Information on Lookup Using Standards	375
14.1.4	Principle 4: Link New Data to Existing Data	376
14.2	Impact and Adoption of Linked Data Principles	377
14.2.1	Overall Impact	379
14.3	Important Knowledge Graphs in Linked Open Data	379
14.3.1	DBpedia	380
14.3.2	GeoNames	380
14.3.3	YAGO	381
14.3.4	Wikidata	382
14.3.5	Upper Mapping and Binding Exchange Layer	383
14.3.6	Friend of a Friend	384
14.3.7	Other Examples	385
14.4	Concluding Notes	386
14.5	Software and Resources	386
14.6	Bibliographic Notes	387
14.7	Exercises	388
15	Enterprise and Government	391
15.1	Introduction	391
15.2	Enterprise	392
15.2.1	Knowledge Vault	393
15.2.2	Social Media and Open Graph Protocol	399
15.2.3	Schema.org	399
15.3	Governments and Nonprofits	402
15.3.1	Open Government Data	402
15.3.2	BBC	405
15.3.3	OpenStreetMap	405
15.4	Where Is the Future Headed?	407
15.5	Concluding Notes	408
15.6	Software and Resources	409
15.7	Bibliographic Notes	410
15.8	Exercises	412
16	Knowledge Graphs and Ontologies in Science	415
16.1	Introduction	415
16.2	Biology	417
16.2.1	Gene Ontology	417

16.3	Chemistry	423
16.3.1	Chemical Entities of Biological Interest	423
16.3.2	PubChem	426
16.4	Earth, Environment, and Geosciences	427
16.4.1	Semantic Web for Earth and Environmental Terminology	427
16.4.2	The GEON Portal and OpenTopography	427
16.4.3	Environment Ontology	431
16.5	Concluding Notes	433
16.6	Software and Resources	434
16.7	Bibliographic Notes	435
16.8	Exercises	436
17	Knowledge Graphs for Domain-Specific Social Impact	439
17.1	Introduction	439
17.2	Domain-Specific Insight Graphs	441
17.2.1	Domain Setup	442
17.2.2	Domain Exploration	446
17.3	Alternative System: DeepDive	450
17.4	Applications and Use-Cases	452
17.4.1	Investigative Domains	452
17.4.2	Crisis Informatics	456
17.4.3	COVID-19 and Medical Informatics	458
17.5	Concluding Notes	460
17.6	Software and Resources	461
17.7	Bibliographic Notes	463
17.8	Exercises	465
	Bibliography	467
	Index	511