

---

# Contents

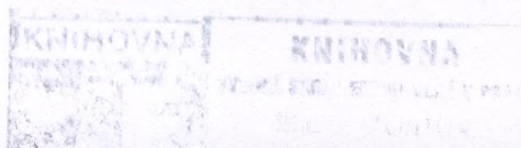
List of Figures	xxi
List of Tables	xxv
Symbol Description	xxvii
<b>I Basic Concepts</b>	<b>1</b>
<b>1 Ontologies and Applications of Ontologies in Biomedicine</b>	<b>3</b>
1.1 What Is an Ontology? . . . . .	3
1.2 Ontologies and Bio-Ontologies . . . . .	5
1.3 Ontologies for Data Organization, Integration, and Searching . . . . .	6
1.4 Computer Reasoning with Ontologies . . . . .	9
1.5 Typical Applications of Bio-Ontologies . . . . .	10
<b>2 Mathematical Logic and Inference</b>	<b>11</b>
2.1 Representation and Logic . . . . .	11
2.2 Propositional Logic . . . . .	13
2.3 First-Order Logic . . . . .	18
2.4 Sets . . . . .	25
2.5 Description Logic . . . . .	29
2.5.1 Description Language $\mathcal{AL}$ . . . . .	29
2.5.2 Description Language $\mathcal{ALC}$ . . . . .	31
2.5.3 Further Description Logic Constructors . . . . .	32
2.6 Exercises and Further Reading . . . . .	36
<b>3 Probability Theory and Statistics for Bio-Ontologies</b>	<b>41</b>
3.1 Probability Theory . . . . .	41
3.1.1 Hypothesis Testing . . . . .	43
3.1.2 $p$ -Values and Probability Distributions . . . . .	44
3.1.3 Multiple-Testing Correction . . . . .	51
3.2 Bayes' Theorem . . . . .	56
3.3 Introduction to Graphs . . . . .	58
3.4 Bayesian Networks . . . . .	62
3.5 Exercises and Further Reading . . . . .	64



<b>4</b>	<b>Ontology Languages</b>	<b>67</b>
4.1	OBO . . . . .	67
4.1.1	OBO Stanzas . . . . .	68
4.1.2	Intersections: Computable Definitions . . . . .	70
4.2	OWL and the Semantic Web . . . . .	71
4.2.1	Resource Description Framework . . . . .	72
4.2.2	RDF Schema . . . . .	78
4.2.3	The Web Ontology Language OWL . . . . .	85
4.2.4	OBO, RDF, RDFS, and OWL . . . . .	95
4.3	Exercises and Further Reading . . . . .	99
<b>II</b>	<b>Bio-Ontologies</b>	<b>113</b>
<b>5</b>	<b>The Gene Ontology</b>	<b>115</b>
5.1	A Tool for the Unification of Biology . . . . .	115
5.2	Three Subontologies . . . . .	117
5.2.1	Molecular Function . . . . .	117
5.2.2	Biological Process . . . . .	117
5.2.3	Cellular Component . . . . .	118
5.3	Relations in GO . . . . .	120
5.4	GO Annotations . . . . .	121
5.4.1	Evidence for Gene Functions . . . . .	124
5.4.2	Inferred from Electronic Annotation . . . . .	127
5.4.3	The True Path Rule and Propagation of Annotations . . . . .	128
5.5	GO Slims . . . . .	133
5.6	Exercises and Further Reading . . . . .	134
<b>6</b>	<b>Upper-Level Ontologies</b>	<b>139</b>
6.1	Basic Formal Ontology . . . . .	139
6.2	The Big Divide: Continuants and Occurrents . . . . .	140
6.2.1	Continuants . . . . .	141
6.2.2	Occurrents . . . . .	142
6.3	Universals and Particulars . . . . .	142
6.4	Relation Ontology . . . . .	143
6.5	Revisiting Gene Ontology . . . . .	147
6.6	Revisiting GO Annotations . . . . .	148
6.7	Exercises and Further Reading . . . . .	150
<b>7</b>	<b>A Selective Survey of Bio-Ontologies</b>	<b>153</b>
7.1	OBO Foundry . . . . .	153
7.2	The National Center for Biomedical Ontology . . . . .	155
7.3	Bio-Ontologies . . . . .	155
7.3.1	Ontologies for Anatomy: The FMA and Model Organisms . . . . .	156
7.3.2	Cell Ontology . . . . .	159
7.3.3	Chemical Entities of Biological Interest . . . . .	160



7.3.4	OBI	162
7.3.5	The Protein Ontology	162
7.3.6	The Sequence Ontology	165
7.3.7	Mammalian Phenotype Ontology (MPO)	166
7.3.8	Human Phenotype Ontology (HPO)	167
7.3.9	MPATH	171
7.3.10	PATO	172
7.4	What Makes a Good Ontology?	175
7.5	Exercises and Further Reading	177
<b>III Graph Algorithms for Bio-Ontologies</b>		<b>179</b>
<b>8</b>	<b>Overrepresentation Analysis</b>	<b>181</b>
8.1	Definitions	182
8.2	Term-for-Term	183
8.3	Multiple Testing Problem	185
8.4	Term-for-Term Analysis: An Extended Example	188
8.5	Inferred Annotations Lead to Statistical Dependencies in Ontology DAGs	192
8.6	Parent-Child Algorithms	195
8.7	Parent-Child Analysis: An Extended Example	198
8.8	Topology-Based Algorithms	200
8.8.1	Elim	200
8.8.2	Weight	202
8.9	Topology-elim: An Extended Example	205
8.10	Other Approaches	207
8.11	Summary	209
8.12	Exercises and Further Reading	209
<b>9</b>	<b>Model-Based Approaches to GO Analysis</b>	<b>219</b>
9.1	A Probabilistic Generative Model for GO Enrichment Analysis	219
9.2	A Bayesian Network Model	222
9.2.1	Maximum a posteriori	226
9.2.2	Monte Carlo Markov Chain Algorithm	227
9.2.3	MGSA Algorithm with Unknown Parameters	230
9.3	MGSA: An Extended Example	233
9.4	Summary	234
9.5	Exercises and Further Reading	235
<b>10</b>	<b>Semantic Similarity</b>	<b>237</b>
10.1	Information Content in Ontologies	237
10.2	Semantic Similarity of Genes and Other Items Annotated by Ontology Terms	247
10.2.1	Graph-Based and Set-Based Measures of Semantic Similarity	248
10.2.2	Applications of Semantic Similarity in Bioinformatics	249





10.2.3 Applications of Semantic Similarity for Clinical Diagnostics . . . . .	251
10.3 Statistical Significance of Semantic Similarity Scores . . . . .	252
10.4 Exercises and Further Reading . . . . .	255
<b>11 Frequency-Aware Bayesian Network Searches in Attribute Ontologies</b>	<b>261</b>
11.1 Modeling Queries . . . . .	262
11.1.1 High-Level Description of the Model . . . . .	262
11.1.2 Annotation Propagation Rule for Bayesian Networks . . . . .	264
11.1.3 LPDs of Hidden Term States . . . . .	266
11.1.4 LPDs of Observed Term States . . . . .	266
11.2 Probabilistic Inference for the Items . . . . .	270
11.3 Parameter-Augmented Network . . . . .	272
11.4 The Frequency-Aware Network . . . . .	273
11.5 Benchmark . . . . .	274
<b>IV Inference in Ontologies</b>	<b>279</b>
<b>12 Inference in the Gene Ontology</b>	<b>281</b>
12.1 Inference over GO Edges . . . . .	281
12.2 Cross-Products and Logical Definitions . . . . .	285
12.2.1 Intra-GO Cross-Product Definitions . . . . .	286
12.2.2 External Cross-Product Definitions . . . . .	288
12.2.3 Reasoning with Cross-Product Definitions . . . . .	288
12.3 Exercises and Further Reading . . . . .	289
<b>13 RDFS Semantics and Inference</b>	<b>293</b>
13.1 Definitions . . . . .	293
13.2 Interpretations . . . . .	294
13.3 RDF Entailment . . . . .	303
13.4 RDFS Entailment . . . . .	305
13.5 Entailment Rules . . . . .	307
13.6 Summary . . . . .	314
13.7 Exercises and Further Reading . . . . .	314
<b>14 Inference in OWL Ontologies</b>	<b>317</b>
14.1 The Semantics of Equality . . . . .	317
14.2 The Semantics of Properties . . . . .	320
14.3 The Semantics of Classes . . . . .	325
14.4 The Semantics of the Schema Vocabulary . . . . .	331
14.5 Conclusions . . . . .	333
14.6 Exercises and Further Reading . . . . .	333



<b>15 Algorithmic Foundations of Computational Inference</b>	<b>335</b>
15.1 The Tableau Algorithm	336
15.1.1 Negative Normal Form	338
15.1.2 Algorithm for ABox	339
15.1.3 Adding Support for the TBox	344
15.2 Developer Libraries	346
15.3 Exercises and Further Reading	348
<b>16 SPARQL</b>	<b>357</b>
16.1 SPARQL Queries	357
16.2 Combining RDF Graphs	365
16.3 Conclusions	368
16.4 Exercises and Further Reading	369
<b>Appendices</b>	<b>375</b>
<b>A An Overview of R</b>	<b>377</b>
<b>B Information Content and Entropy</b>	<b>395</b>
<b>C W3C Standards: XML, URIs, and RDF</b>	<b>399</b>
<b>D W3C Standards: OWL</b>	<b>427</b>
<b>Glossary</b>	<b>457</b>
<b>Bibliography</b>	<b>461</b>
<b>Index</b>	<b>485</b>