PRENTICE
HALL
INTERNATIONAL
EDITIONS

ARTIFICIAL INTELLIGENCE

A MODERN APPROACH

STUART
RUSSELL

PETER
NORVIG

Contents

I Artificial Intelligence 1

1 Introduction 3

1.1 What is AI? . . . 4
Acting humanly: The Turing Test approach . . . 5
Thinking humanly: The cognitive modelling approach . . . 6
Thinking rationally: The laws of thought approach . . . 6
Acting rationally: The rational agent approach . . . 7

1.2 The Foundations of Artificial Intelligence . . . 8
Philosophy (428 B.C.–present) . . . 8
Mathematics (c. 800–present) . . . 11
Psychology (1879–present) . . . 12
Computer engineering (1940–present) . . . 14
Linguistics (1957–present) . . . 15

1.3 The History of Artificial Intelligence . . . 16
The gestation of artificial intelligence (1943–1956) . . . 16
Early enthusiasm, great expectations (1952–1969) . . . 17
A dose of reality (1966–1974) . . . 20
Knowledge-based systems: The key to power? (1969–1979) . . . 22
AI becomes an industry (1980–1988) . . . 24
The return of neural networks (1986–present) . . . 24
Recent events (1987–present) . . . 25

1.4 The State of the Art . . . 26

1.5 Summary . . . 27

Bibliographical and Historical Notes . . . 28

Exercises . . . 28

2 Intelligent Agents 31

2.1 Introduction . . . 31

2.2 How Agents Should Act . . . 31
The ideal mapping from percept sequences to actions . . . 34
Autonomy . . . 35

2.3 Structure of Intelligent Agents . . . 35
Agent programs . . . 37
Why not just look up the answers? . . . 38
An example . . . 39
Simple reflex agents . . . 40
Agents that keep track of the world . . . 41
Goal-based agents . . . 42
Utility-based agents . . . 44

2.4 Environments . . . 45

Properties of environments 46
Environment programs 47
2.5 Summary 49
Bibliographical and Historical Notes 50
Exercises 50

II Problem-solving **53**

3 Solving Problems by Searching **55**
3.1 Problem-Solving Agents 55
3.2 Formulating Problems 57
Knowledge and problem types 58
Well-defined problems and solutions 60
Measuring problem-solving performance 61
Choosing states and actions 61
3.3 Example Problems 63
Toy problems 63
Real-world problems 68
3.4 Searching for Solutions 70
Generating action sequences 70
Data structures for search trees 72
3.5 Search Strategies 73
Breadth-first search 74
Uniform cost search 75
Depth-first search 77
Depth-limited search 78
Iterative deepening search 78
Bidirectional search 80
Comparing search strategies 81
3.6 Avoiding Repeated States 82
3.7 Constraint Satisfaction Search 83
3.8 Summary 85
Bibliographical and Historical Notes 86
Exercises 87

4 Informed Search Methods **92**
4.1 Best-First Search 92
Minimize estimated cost to reach a goal: Greedy search 93
Minimizing the total path cost: A* search 96
4.2 Heuristic Functions 101
The effect of heuristic accuracy on performance 102
Inventing heuristic functions 103
Heuristics for constraint satisfaction problems 104
4.3 Memory Bounded Search 106

Iterative deepening A* search (IDA*) . . . 106
SMA* search . . . 107
4.4 Iterative Improvement Algorithms . . . 111
Hill-climbing search . . . 111
Simulated annealing . . . 113
Applications in constraint satisfaction problems . . . 114
4.5 Summary . . . 115
Bibliographical and Historical Notes . . . 115
Exercises . . . 118

5 Game Playing **122**
5.1 Introduction: Games as Search Problems . . . 122
5.2 Perfect Decisions in Two-Person Games . . . 123
5.3 Imperfect Decisions . . . 126
Evaluation functions . . . 127
Cutting off search . . . 129
5.4 Alpha-Beta Pruning . . . 129
Effectiveness of alpha-beta pruning . . . 131
5.5 Games That Include an Element of Chance . . . 133
Position evaluation in games with chance nodes . . . 135
Complexity of expectiminimax . . . 135
5.6 State-of-the-Art Game Programs . . . 136
Chess . . . 137
Checkers or Draughts . . . 138
Othello . . . 138
Backgammon . . . 139
Go . . . 139
5.7 Discussion . . . 139
5.8 Summary . . . 141
Bibliographical and Historical Notes . . . 141
Exercises . . . 145

III Knowledge and reasoning **149**

6 Agents that Reason Logically **151**
6.1 A Knowledge-Based Agent . . . 151
6.2 The Wumpus World Environment . . . 153
Specifying the environment . . . 154
Acting and reasoning in the wumpus world . . . 155
6.3 Representation, Reasoning, and Logic . . . 157
Representation . . . 160
Inference . . . 163
Logics . . . 165
6.4 Propositional Logic: A Very Simple Logic . . . 166

Syntax . . . 166
Semantics . . . 168
Validity and inference . . . 169
Models . . . 170
Rules of inference for propositional logic . . . 171
Complexity of propositional inference . . . 173
6.5 An Agent for the Wumpus World . . . 174
The knowledge base . . . 174
Finding the wumpus . . . 175
Translating knowledge into action . . . 176
Problems with the propositional agent . . . 176
6.6 Summary . . . 178
Bibliographical and Historical Notes . . . 178
Exercises . . . 180

7 First-Order Logic **185**
7.1 Syntax and Semantics . . . 186
Terms . . . 188
Atomic sentences . . . 189
Complex sentences . . . 189
Quantifiers . . . 189
Equality . . . 193
7.2 Extensions and Notational Variations . . . 194
Higher-order logic . . . 195
Functional and predicate expressions using the λ operator . . . 195
The uniqueness quantifier $\exists!$ . . . 196
The uniqueness operator ι . . . 196
Notational variations . . . 196
7.3 Using First-Order Logic . . . 197
The kinship domain . . . 197
Axioms, definitions, and theorems . . . 198
The domain of sets . . . 199
Special notations for sets, lists and arithmetic . . . 200
Asking questions and getting answers . . . 200
7.4 Logical Agents for the Wumpus World . . . 201
7.5 A Simple Reflex Agent . . . 202
Limitations of simple reflex agents . . . 203
7.6 Representing Change in the World . . . 203
Situation calculus . . . 204
Keeping track of location . . . 206
7.7 Deducing Hidden Properties of the World . . . 208
7.8 Preferences Among Actions . . . 210
7.9 Toward a Goal-Based Agent . . . 211
7.10 Summary . . . 211

Bibliographical and Historical Notes . . . 212
Exercises . . . 213

8 Building a Knowledge Base **217**
8.1 Properties of Good and Bad Knowledge Bases . . . 218
8.2 Knowledge Engineering . . . 221
8.3 The Electronic Circuits Domain . . . 223
Decide what to talk about . . . 223
Decide on a vocabulary . . . 224
Encode general rules . . . 225
Encode the specific instance . . . 225
Pose queries to the inference procedure . . . 226
8.4 General Ontology . . . 226
Representing Categories . . . 229
Measures . . . 231
Composite objects . . . 233
Representing change with events . . . 234
Times, intervals, and actions . . . 238
Objects revisited . . . 240
Substances and objects . . . 241
Mental events and mental objects . . . 243
Knowledge and action . . . 247
8.5 The Grocery Shopping World . . . 247
Complete description of the shopping simulation . . . 248
Organizing knowledge . . . 249
Menu-planning . . . 249
Navigating . . . 252
Gathering . . . 253
Communicating . . . 254
Paying . . . 255
8.6 Summary . . . 256
Bibliographical and Historical Notes . . . 256
Exercises . . . 261

9 Inference in First-Order Logic **265**
9.1 Inference Rules Involving Quantifiers . . . 265
9.2 An Example Proof . . . 266
9.3 Generalized Modus Ponens . . . 269
Canonical form . . . 270
Unification . . . 270
Sample proof revisited . . . 271
9.4 Forward and Backward Chaining . . . 272
Forward-chaining algorithm . . . 273
Backward-chaining algorithm . . . 275

9.5 Completeness . . . 276
9.6 Resolution: A Complete Inference Procedure . . . 277
The resolution inference rule . . . 278
Canonical forms for resolution . . . 278
Resolution proofs . . . 279
Conversion to Normal Form . . . 281
Example proof . . . 282
Dealing with equality . . . 284
Resolution strategies . . . 284
9.7 Completeness of resolution . . . 286
9.8 Summary . . . 290
Bibliographical and Historical Notes . . . 291
Exercises . . . 294

10 Logical Reasoning Systems **297**
10.1 Introduction . . . 297
10.2 Indexing, Retrieval, and Unification . . . 299
Implementing sentences and terms . . . 299
Store and fetch . . . 299
Table-based indexing . . . 300
Tree-based indexing . . . 301
The unification algorithm . . . 302
10.3 Logic Programming Systems . . . 304
The Prolog language . . . 304
Implementation . . . 305
Compilation of logic programs . . . 306
Other logic programming languages . . . 308
Advanced control facilities . . . 308
10.4 Theorem Provers . . . 310
Design of a theorem prover . . . 310
Extending Prolog . . . 311
Theorem provers as assistants . . . 312
Practical uses of theorem provers . . . 313
10.5 Forward-Chaining Production Systems . . . 313
Match phase . . . 314
Conflict resolution phase . . . 315
Practical uses of production systems . . . 316
10.6 Frame Systems and Semantic Networks . . . 316
Syntax and semantics of semantic networks . . . 317
Inheritance with exceptions . . . 319
Multiple inheritance . . . 320
Inheritance and change . . . 320
Implementation of semantic networks . . . 321
Expressiveness of semantic networks . . . 323

10.7 Description Logics . . . 323
Practical uses of description logics . . . 325
10.8 Managing Retractions, Assumptions, and Explanations . . . 325
10.9 Summary . . . 327
Bibliographical and Historical Notes . . . 328
Exercises . . . 332

IV Acting logically **335**

11 Planning **337**
11.1 A Simple Planning Agent . . . 337
11.2 From Problem Solving to Planning . . . 338
11.3 Planning in Situation Calculus . . . 341
11.4 Basic Representations for Planning . . . 343
Representations for states and goals . . . 343
Representations for actions . . . 344
Situation space and plan space . . . 345
Representations for plans . . . 346
Solutions . . . 349
11.5 A Partial-Order Planning Example . . . 349
11.6 A Partial-Order Planning Algorithm . . . 355
11.7 Planning with Partially Instantiated Operators . . . 357
11.8 Knowledge Engineering for Planning . . . 359
The blocks world . . . 359
Shakey's world . . . 360
11.9 Summary . . . 362
Bibliographical and Historical Notes . . . 363
Exercises . . . 364

12 Practical Planning **367**
12.1 Practical Planners . . . 367
Spacecraft assembly, integration, and verification . . . 367
Job shop scheduling . . . 369
Scheduling for space missions . . . 369
Buildings, aircraft carriers, and beer factories . . . 371
12.2 Hierarchical Decomposition . . . 371
Extending the language . . . 372
Modifying the planner . . . 374
12.3 Analysis of Hierarchical Decomposition . . . 375
Decomposition and sharing . . . 379
Decomposition versus approximation . . . 380
12.4 More Expressive Operator Descriptions . . . 381
Conditional effects . . . 381
Negated and disjunctive goals . . . 382

Universal quantification . . . 383
A planner for expressive operator descriptions . . . 384
12.5 Resource Constraints . . . 386
Using measures in planning . . . 386
Temporal constraints . . . 388
12.6 Summary . . . 388
Bibliographical and Historical Notes . . . 389
Exercises . . . 390

13 Planning and Acting **392**
13.1 Conditional Planning . . . 393
The nature of conditional plans . . . 393
An algorithm for generating conditional plans . . . 395
Extending the plan language . . . 398
13.2 A Simple Replanning Agent . . . 401
Simple replanning with execution monitoring . . . 402
13.3 Fully Integrated Planning and Execution . . . 403
13.4 Discussion and Extensions . . . 407
Comparing conditional planning and replanning . . . 407
Coercion and abstraction . . . 409
13.5 Summary . . . 410
Bibliographical and Historical Notes . . . 411
Exercises . . . 412

V Uncertain knowledge and reasoning **413**

14 Uncertainty **415**
14.1 Acting under Uncertainty . . . 415
Handling uncertain knowledge . . . 416
Uncertainty and rational decisions . . . 418
Design for a decision-theoretic agent . . . 419
14.2 Basic Probability Notation . . . 420
Prior probability . . . 420
Conditional probability . . . 421
14.3 The Axioms of Probability . . . 422
Why the axioms of probability are reasonable . . . 423
The joint probability distribution . . . 425
14.4 Bayes' Rule and Its Use . . . 426
Applying Bayes' rule: The simple case . . . 426
Normalization . . . 427
Using Bayes' rule: Combining evidence . . . 428
14.5 Where Do Probabilities Come From? . . . 430
14.6 Summary . . . 431
Bibliographical and Historical Notes . . . 431

Exercises . . . 433

15 Probabilistic Reasoning Systems **436**
15.1 Representing Knowledge in an Uncertain Domain . . . 436
15.2 The Semantics of Belief Networks . . . 438
Representing the joint probability distribution . . . 439
Conditional independence relations in belief networks . . . 444
15.3 Inference in Belief Networks . . . 445
The nature of probabilistic inferences . . . 446
An algorithm for answering queries . . . 447
15.4 Inference in Multiply Connected Belief Networks . . . 453
Clustering methods . . . 453
Cutset conditioning methods . . . 454
Stochastic simulation methods . . . 455
15.5 Knowledge Engineering for Uncertain Reasoning . . . 456
Case study: The Pathfinder system . . . 457
15.6 Other Approaches to Uncertain Reasoning . . . 458
Default reasoning . . . 459
Rule-based methods for uncertain reasoning . . . 460
Representing ignorance: Dempster–Shafer theory . . . 462
Representing vagueness: Fuzzy sets and fuzzy logic . . . 463
15.7 Summary . . . 464
Bibliographical and Historical Notes . . . 464
Exercises . . . 467

16 Making Simple Decisions **471**
16.1 Combining Beliefs and Desires Under Uncertainty . . . 471
16.2 The Basis of Utility Theory . . . 473
Constraints on rational preferences . . . 473
. . . and then there was Utility . . . 474
16.3 Utility Functions . . . 475
The utility of money . . . 476
Utility scales and utility assessment . . . 478
16.4 Multiattribute utility functions . . . 480
Dominance . . . 481
Preference structure and multiattribute utility . . . 483
16.5 Decision Networks . . . 484
Representing a decision problem using decision networks . . . 484
Evaluating decision networks . . . 486
16.6 The Value of Information . . . 487
A simple example . . . 487
A general formula . . . 488
Properties of the value of information . . . 489
Implementing an information-gathering agent . . . 490

16.7 Decision-Theoretic Expert Systems . . . 491
16.8 Summary . . . 493
Bibliographical and Historical Notes . . . 493
Exercises . . . 495

17 Making Complex Decisions **498**
17.1 Sequential Decision Problems . . . 498
17.2 Value Iteration . . . 502
17.3 Policy Iteration . . . 505
17.4 Decision-Theoretic Agent Design . . . 508
The decision cycle of a rational agent . . . 508
Sensing in uncertain worlds . . . 510
17.5 Dynamic Belief Networks . . . 514
17.6 Dynamic Decision Networks . . . 516
Discussion . . . 518
17.7 Summary . . . 519
Bibliographical and Historical Notes . . . 520
Exercises . . . 521

VI Learning **523**

18 Learning from Observations **525**
18.1 A General Model of Learning Agents . . . 525
Components of the performance element . . . 527
Representation of the components . . . 528
Available feedback . . . 528
Prior knowledge . . . 528
Bringing it all together . . . 529
18.2 Inductive Learning . . . 529
18.3 Learning Decision Trees . . . 531
Decision trees as performance elements . . . 531
Expressiveness of decision trees . . . 532
Inducing decision trees from examples . . . 534
Assessing the performance of the learning algorithm . . . 538
Practical uses of decision tree learning . . . 538
18.4 Using Information Theory . . . 540
Noise and overfitting . . . 542
Broadening the applicability of decision trees . . . 543
18.5 Learning General Logical Descriptions . . . 544
Hypotheses . . . 544
Examples . . . 545
Current-best-hypothesis search . . . 546
Least-commitment search . . . 549
Discussion . . . 552

18.6 Why Learning Works: Computational Learning Theory . . . 552
How many examples are needed? . . . 553
Learning decision lists . . . 555
Discussion . . . 557
18.7 Summary . . . 558
Bibliographical and Historical Notes . . . 559
Exercises . . . 560

19 Learning in Neural and Belief Networks **563**
19.1 How the Brain Works . . . 564
Comparing brains with digital computers . . . 565
19.2 Neural Networks . . . 567
Notation . . . 567
Simple computing elements . . . 567
Network structures . . . 570
Optimal network structure . . . 572
19.3 Perceptrons . . . 573
What perceptrons can represent . . . 573
Learning linearly separable functions . . . 575
19.4 Multilayer Feed-Forward Networks . . . 578
Back-propagation learning . . . 578
Back-propagation as gradient descent search . . . 580
Discussion . . . 583
19.5 Applications of Neural Networks . . . 584
Pronunciation . . . 585
Handwritten character recognition . . . 586
Driving . . . 586
19.6 Bayesian Methods for Learning Belief Networks . . . 588
Bayesian learning . . . 588
Belief network learning problems . . . 589
Learning networks with fixed structure . . . 589
A comparison of belief networks and neural networks . . . 592
19.7 Summary . . . 593
Bibliographical and Historical Notes . . . 594
Exercises . . . 596

20 Reinforcement Learning **598**
20.1 Introduction . . . 598
20.2 Passive Learning in a Known Environment . . . 600
Naïve updating . . . 601
Adaptive dynamic programming . . . 603
Temporal difference learning . . . 604
20.3 Passive Learning in an Unknown Environment . . . 605
20.4 Active Learning in an Unknown Environment . . . 607

20.5 Exploration . . . 609
20.6 Learning an Action-Value Function . . . 612
20.7 Generalization in Reinforcement Learning . . . 615
Applications to game-playing . . . 617
Application to robot control . . . 617
20.8 Genetic Algorithms and Evolutionary Programming . . . 619
20.9 Summary . . . 621
Bibliographical and Historical Notes . . . 622
Exercises . . . 623

21 Knowledge in Learning **625**
21.1 Knowledge in Learning . . . 625
Some simple examples . . . 626
Some general schemes . . . 627
21.2 Explanation-Based Learning . . . 629
Extracting general rules from examples . . . 630
Improving efficiency . . . 631
21.3 Learning Using Relevance Information . . . 633
Determining the hypothesis space . . . 633
Learning and using relevance information . . . 634
21.4 Inductive Logic Programming . . . 636
An example . . . 637
Inverse resolution . . . 639
Top-down learning methods . . . 641
21.5 Summary . . . 644
Bibliographical and Historical Notes . . . 645
Exercises . . . 647

VII Communicating, perceiving, and acting **649**

22 Agents that Communicate **651**
22.1 Communication as Action . . . 652
Fundamentals of language . . . 654
The component steps of communication . . . 655
Two models of communication . . . 659
22.2 Types of Communicating Agents . . . 659
Communicating using Tell and Ask . . . 660
Communicating using formal language . . . 661
An agent that communicates . . . 662
22.3 A Formal Grammar for a Subset of English . . . 662
The Lexicon of $\mathcal{E}_0$ . . . 664
The Grammar of $\mathcal{E}_0$ . . . 664
22.4 Syntactic Analysis (Parsing) . . . 664
22.5 Definite Clause Grammar (DCG) . . . 667

22.6 Augmenting a Grammar . . . 668
Verb Subcategorization . . . 669
Generative Capacity of Augmented Grammars . . . 671
22.7 Semantic Interpretation . . . 672
Semantics as DCG Augmentations . . . 673
The semantics of "John loves Mary" . . . 673
The semantics of $\mathcal{E}_1$ . . . 675
Converting quasi-logical form to logical form . . . 677
Pragmatic Interpretation . . . 678
22.8 Ambiguity and Disambiguation . . . 680
Disambiguation . . . 682
22.9 A Communicating Agent . . . 683
22.10 Summary . . . 684
Bibliographical and Historical Notes . . . 685
Exercises . . . 688

23 Practical Natural Language Processing **691**
23.1 Practical Applications . . . 691
Machine translation . . . 691
Database access . . . 693
Information retrieval . . . 694
Text categorization . . . 695
Extracting data from text . . . 696
23.2 Efficient Parsing . . . 696
Extracting parses from the chart: Packing . . . 701
23.3 Scaling Up the Lexicon . . . 703
23.4 Scaling Up the Grammar . . . 705
Nominal compounds and apposition . . . 706
Adjective phrases . . . 707
Determiners . . . 708
Noun phrases revisited . . . 709
Clausal complements . . . 710
Relative clauses . . . 710
Questions . . . 711
Handling agrammatical strings . . . 712
23.5 Ambiguity . . . 712
Syntactic evidence . . . 713
Lexical evidence . . . 713
Semantic evidence . . . 713
Metonymy . . . 714
Metaphor . . . 715
23.6 Discourse Understanding . . . 715
The structure of coherent discourse . . . 717
23.7 Summary . . . 719

Bibliographical and Historical Notes . . . 720
Exercises . . . 721

24 Perception **724**
24.1 Introduction . . . 724
24.2 Image Formation . . . 725
Pinhole camera . . . 725
Lens systems . . . 727
Photometry of image formation . . . 729
Spectrophotometry of image formation . . . 730
24.3 Image-Processing Operations for Early Vision . . . 730
Convolution with linear filters . . . 732
Edge detection . . . 733
24.4 Extracting 3-D Information Using Vision . . . 734
Motion . . . 735
Binocular stereopsis . . . 737
Texture gradients . . . 742
Shading . . . 743
Contour . . . 745
24.5 Using Vision for Manipulation and Navigation . . . 749
24.6 Object Representation and Recognition . . . 751
The alignment method . . . 752
Using projective invariants . . . 754
24.7 Speech Recognition . . . 757
Signal processing . . . 758
Defining the overall speech recognition model . . . 760
The language model: P(words) . . . 760
The acoustic model: P(signal|words) . . . 762
Putting the models together . . . 764
The search algorithm . . . 765
Training the model . . . 766
24.8 Summary . . . 767
Bibliographical and Historical Notes . . . 767
Exercises . . . 771

25 Robotics **773**
25.1 Introduction . . . 773
25.2 Tasks: What Are Robots Good For? . . . 774
Manufacturing and materials handling . . . 774
Gofer robots . . . 775
Hazardous environments . . . 775
Telepresence and virtual reality . . . 776
Augmentation of human abilities . . . 776
25.3 Parts: What Are Robots Made Of? . . . 777

Effectors: Tools for action . . . 777
Sensors: Tools for perception . . . 782
25.4 Architectures . . . 786
Classical architecture . . . 787
Situated automata . . . 788
25.5 Configuration Spaces: A Framework for Analysis . . . 790
Generalized configuration space . . . 792
Recognizable Sets . . . 795
25.6 Navigation and Motion Planning . . . 796
Cell decomposition . . . 796
Skeletonization methods . . . 798
Fine-motion planning . . . 802
Landmark-based navigation . . . 805
Online algorithms . . . 806
25.7 Summary . . . 809
Bibliographical and Historical Notes . . . 809
Exercises . . . 811

VIII Conclusions **815**

26 Philosophical Foundations **817**
26.1 The Big Questions . . . 817
26.2 Foundations of Reasoning and Perception . . . 819
26.3 On the Possibility of Achieving Intelligent Behavior . . . 822
The mathematical objection . . . 824
The argument from informality . . . 826
26.4 Intentionality and Consciousness . . . 830
The Chinese Room . . . 831
The Brain Prosthesis Experiment . . . 835
Discussion . . . 836
26.5 Summary . . . 837
Bibliographical and Historical Notes . . . 838
Exercises . . . 840

27 AI: Present and Future **842**
27.1 Have We Succeeded Yet? . . . 842
27.2 What Exactly Are We Trying to Do? . . . 845
27.3 What If We Do Succeed? . . . 848

A Complexity analysis and O() notation **851**
A.1 Asymptotic Analysis . . . 851
A.2 Inherently Hard Problems . . . 852
Bibliographical and Historical Notes . . . 853

B Notes on Languages and Algorithms **854**
B.1 Defining Languages with Backus–Naur Form (BNF) 854
B.2 Describing Algorithms with Pseudo-Code 855
Nondeterminism 855
Static variables 856
Functions as values 856
B.3 The Code Repository 857
B.4 Comments 857

Bibliography **859**

Index **905**