

Τεχνητή Νοημοσύνη

Μια σύγχρονη προσέγγιση

ΔΕΥΤΕΡΗ ΑΜΕΡΙΚΑΝΙΚΗ ΕΚΔΟΣΗ



Περιεχόμενα

1	Εισαγωγή	31
1.1	Τι Είναι η Τεχνητή Νοημοσύνη.....	31
	Ανθρώπινη δράση: Η προσέγγιση με τη δοκιμασία Turing	32
	Ανθρώπινη σκέψη: Η προσέγγιση με γνωστικά μοντέλα.....	33
	Ορθολογική σκέψη: Η προσέγγιση με τους “νόμους της σκέψης”	34
	Ορθολογική δράση: Η προσέγγιση με ορθολογικούς πράκτορες	34
1.2	Βασικές Αρχές της Τεχνητής Νοημοσύνης	36
	Φιλοσοφία (428 π.Χ. – σήμερα).....	36
	Μαθηματικά (περ. 800 – σήμερα).....	38
	Οικονομικά (1776 – σήμερα)	40
	Νευροεπιστήμες (1861 – σήμερα).....	41
	Ψυχολογία (1879 – σήμερα).....	43
	Τεχνολογία υπολογιστών (1940 – σήμερα).....	44
	Θεωρία ελέγχου και κυβερνητική (1948 – σήμερα).....	46
	Γλωσσολογία (1957 – σήμερα).....	46
1.3	Ιστορία της Τεχνητής Νοημοσύνης.....	47
	Η κυοφορία της τεχνητής νοημοσύνης (1943 – 1955).....	47
	Η γέννηση της τεχνητής νοημοσύνης (1956).....	48
	Πρώιμος ενθουσιασμός, μεγάλες προσδοκίες (1952 – 1969)	49
	Μια δόση ρεαλισμού (1966 – 1973)	52
	Συστήματα βασισμένα στη γνώση: Το κλειδί της επιτυχίας; (1969 – 1979).....	53
	Η τεχνητή νοημοσύνη γίνεται βιομηχανία (1980 – σήμερα)	55
	Η επιστροφή των νευρωνικών δικτύων (1986 – σήμερα).....	56
	Η τεχνητή νοημοσύνη γίνεται επιστήμη (1987 – σήμερα).....	56
	Η εμφάνιση των ευφύων πρακτόρων (1995 – σήμερα)	58
1.4	Σύγχρονη Τεχνολογία	58
1.5	Σύνοψη	60
	Βιβλιογραφικές και ιστορικές σημειώσεις.....	61
	Ασκήσεις.....	61
2	Ευφυείς Πράκτορες	64
2.1	Πράκτορες και Περιβάλλοντα.....	64
2.2	Καλή Συμπεριφορά: Η Έννοια της Ορθολογικότητας	67
	Μέτρα της απόδοσης.....	67
	Ορθολογικότητα	68
	Παντογνωσία, μάθηση και αυτονομία.....	69
2.3	Φύση των Περιβαλλόντων	70
	Καθορισμός του περιβάλλοντος εργασιών.....	70
	Ιδιότητες των περιβαλλόντων εργασιών	73
2.4	Η Δομή των Πρακτόρων	77
	Προγράμματα πρακτόρων.....	77
	Απλοί αντανakλαστικοί πράκτορες.....	79
	Αντανakλαστικοί πράκτορες βασισμένοι σε μοντέλο	81
	Πράκτορες βασισμένοι στο στόχο.....	82
	Πράκτορες βασισμένοι στη χρησιμότητα	84
	Πράκτορες που μαθαίνουν	84
2.5	Σύνοψη	87

	Βιβλιογραφικές και ιστορικές σημειώσεις.....	88
	Ασκήσεις.....	89
3	Επίλυση Προβλημάτων με Αναζήτηση.....	92
3.1	Πράκτορες Επίλυσης Προβλημάτων.....	92
	Καλά ορισμένα προβλήματα και λύσεις.....	95
	Διατύπωση προβλημάτων	96
3.2	Παραδείγματα Προβλημάτων	97
	Προβλήματα-παιχνίδια.....	97
	Προβλήματα του πραγματικού κόσμου.....	100
3.3	Αναζήτηση Λύσεων	102
	Μέτρηση της απόδοσης στην επίλυση προβλημάτων	106
3.4	Στρατηγικές Απληροφόρητης Αναζήτησης	107
	Αναζήτηση πρώτα σε πλάτος	107
	Αναζήτηση πρώτα σε βάθος	110
	Αναζήτηση περιορισμένου βάθους.....	111
	Αναζήτηση πρώτα σε βάθος με επαναληπτική εκβάθυνση	112
	Αμφίδρομη αναζήτηση.....	114
	Σύγκριση στρατηγικών απληροφόρητης αναζήτησης.....	116
3.5	Αποφυγή Επαναλαμβανόμενων Καταστάσεων.....	116
3.6	Αναζήτηση με Μερική Πληροφόρηση	118
	Προβλήματα χωρίς αισθητήρες.....	119
	Προβλήματα ενδεχομένων	121
3.7	Σύνοψη	122
	Βιβλιογραφικές και ιστορικές σημειώσεις.....	123
	Ασκήσεις.....	125
4	Πληροφορημένη αναζήτηση και εξερεύνηση	131
4.1	Στρατηγικές Πληροφορημένης (Ευρετικής) Αναζήτησης	131
	Άπληστη αναζήτηση πρώτα στο καλύτερο	132
	Αναζήτηση A*: Ελαχιστοποίηση του ολικού εκτιμώμενου κόστους λύσης	134
	Ευρετική αναζήτηση περιορισμένης μνήμης	139
	Εκμάθηση καλύτερων τρόπων αναζήτησης	143
4.2	Ευρετικές Συναρτήσεις	143
	Επίδραση της ευρετικής ακρίβειας στην απόδοση.....	144
	Επινόηση παραδεκτών ευρετικών συναρτήσεων	145
	Εκμάθηση ευρετικών μηχανισμών από την εμπειρία.....	148
4.3	Αλγόριθμοι Τοπικής Αναζήτησης και Προβλήματα Βελτιστοποίησης	149
	Αναζήτηση με αναρρίχηση λόφων	150
	Αναζήτηση με προσομοιωμένη απόπτηση	154
	Τοπική ακτινική αναζήτηση	155
	Γενετικοί αλγόριθμοι.....	156
4.4	Τοπική Αναζήτηση σε Συνεχείς Χώρους	160
4.5	Πράκτορες Online Αναζήτησης και Άγνωστα Περιβάλλοντα	162
	Προβλήματα online αναζήτησης.....	163
	Πράκτορες online αναζήτησης.....	165
	Τοπική online αναζήτηση	166
	Μάθηση στην online αναζήτηση.....	169
4.6	Σύνοψη	169

	Βιβλιογραφικές και ιστορικές σημειώσεις.....	170
	Ασκήσεις.....	175
5	Προβλήματα Ικανοποίησης Περιορισμών	179
5.1	Προβλήματα Ικανοποίησης Περιορισμών	179
5.2	Αναζήτηση με Υπαναχώρηση στα Προβλήματα Ικανοποίησης Περιορισμών	184
	Προτεραιότητα μεταβλητών και τιμών	186
	Διάδοση πληροφοριών μέσω περιορισμών	187
	Ευφυής υπαναχώρηση: Κοιτάζοντας πίσω.....	192
5.3	Τοπική Αναζήτηση στα Προβλήματα Ικανοποίησης Περιορισμών.....	193
5.4	Η Δομή των Προβλημάτων	195
5.5	Σύνοψη	199
	Βιβλιογραφικές και ιστορικές σημειώσεις.....	200
	Ασκήσεις.....	202
6	Αναζήτηση με Αντιπαλότητα.....	205
6.1	Παιχνίδια	205
6.2	Βέλτιστες Αποφάσεις στα Παιχνίδια.....	206
	Βέλτιστες στρατηγικές	207
	Ο αλγόριθμος minimax	209
	Βέλτιστες αποφάσεις στα παιχνίδια πολλών παικτών.....	210
6.3	Κλάδεμα Άλφα-Βήτα.....	211
6.4	Ατελείς Αποφάσεις σε Πραγματικό Χρόνο.....	215
	Συναρτήσεις αξιολόγησης.....	216
	Αποκοπή αναζήτησης.....	218
6.5	Παιχνίδια που Εμπεριέχουν Στοιχείο Τύχης.....	220
	Αξιολόγηση θέσης σε παιχνίδια με κόμβους τύχης.....	222
	Πολυπλοκότητα της expectiminimax	223
	Παιχνίδια με τραπουλόχαρτα	224
6.6	Προηγμένα Προγράμματα Παιχνιδιών.....	226
6.7	Συζήτηση	229
6.8	Σύνοψη	231
	Βιβλιογραφικές και ιστορικές σημειώσεις.....	231
	Ασκήσεις.....	235
7	Λογικοί Πράκτορες	240
7.1	Πράκτορες Βασισμένοι στη Γνώση	241
7.2	Ο Κόσμος του Wumpus	243
7.3	Λογική.....	247
7.4	Προτασιακή Λογική: Μια Πολύ Απλή Λογική	251
	Σύνταξη	251
	Σημασιολογία	252
	Μια απλή βάση γνώσης.....	254
	Συμπερασμός.....	255
	Ισοδυναμία, εγκυρότητα και ικανοποιησιμότητα.....	257
7.5	Πρότυπα Συλλογιστικής στην Προτασιακή Λογική.....	258
	Ανάλυση	260
	Αλυσίδα εκτέλεσης προς τα εμπρός και προς τα πίσω.....	265
7.6	Αποδοτικός Προτασιακός Συμπερασμός	269
	Ένας πλήρης αλγόριθμος υπαναχώρησης	269

	Αλγόριθμοι τοπικής αναζήτησης.....	271
	Δύσκολα προβλήματα ικανοποιησιμότητας.....	272
7.7	Πράκτορες Βασισμένοι στην Προτασιακή Λογική.....	273
	Εύρεση γουβών και wumpus με τη χρήση λογικού συμπερασμού.....	274
	Παρακολούθηση της θέσης και του προσανατολισμού.....	275
	Πράκτορες βασισμένοι σε κύκλωμα.....	276
	Μια σύγκριση.....	280
7.8	Σύνοψη.....	281
	Βιβλιογραφικές και ιστορικές σημειώσεις.....	282
	Ασκήσεις.....	285
8	Λογική Πρώτης Τάξης.....	289
8.1	Μια Δεύτερη Ματιά στην Αναπαράσταση.....	289
8.2	Σύνταξη και Σημασιολογία της Λογικής Πρώτης Τάξης.....	294
	Μοντέλα για τη λογική πρώτης τάξης.....	294
	Σύμβολα και ερμηνείες.....	295
	Όροι.....	297
	Ατομικές προτάσεις.....	298
	Σύνθετες προτάσεις.....	298
	Ποσοδείκτες.....	298
	Ισότητα.....	302
8.3	Χρήση της Λογικής Πρώτης Τάξης.....	303
	Ισχυρισμοί και ερωτήματα στη λογική πρώτης τάξης.....	303
	Το πεδίο των συγγενειών.....	304
	Αριθμοί, σύνολα και λίστες.....	306
	Ο κόσμος του wumpus.....	308
8.4	Μηχανική Γνώσης στη Λογική Πρώτης Τάξης.....	311
	Η διαδικασία της μηχανικής της γνώσης.....	311
	Το πεδίο των ηλεκτρονικών κυκλωμάτων.....	313
8.5	Σύνοψη.....	317
	Βιβλιογραφικές και ιστορικές σημειώσεις.....	317
	Ασκήσεις.....	318
9	Συμπερασμός στη Λογική Πρώτης Τάξης.....	322
9.1	Προτασιακός Συμπερασμός και Συμπερασμός Πρώτης Τάξης.....	322
	Κανόνες συμπερασμού για ποσοδείκτες.....	323
	Αναγωγή στον προτασιακό συμπερασμό.....	324
9.2	Ενοποίηση και Ανύψωση.....	325
	Ένας κανόνας συμπερασμού πρώτης τάξης.....	325
	Ενοποίηση.....	327
	Αποθήκευση και ανάκτηση.....	329
9.3	Προς τα Εμπρός Αλυσίδα Εκτέλεσης.....	330
	Οριστικές προτάσεις πρώτης τάξης.....	331
	Ένας απλός αλγόριθμος προς τα εμπρός αλυσίδας εκτέλεσης.....	332
	Αποδοτική προς τα εμπρός αλυσίδα εκτέλεσης.....	334
	Αυξητική προς τα εμπρός αλυσίδα εκτέλεσης.....	336
9.4	Προς τα Πίσω Αλυσίδα Εκτέλεσης.....	339
	Ένας αλγόριθμος προς τα πίσω αλυσίδας εκτέλεσης.....	339
	Λογικός προγραμματισμός.....	340
	Αποδοτική υλοποίηση λογικών προγραμμάτων.....	342
	Πλεονάζων συμπερασμός και ατέρμονες βρόχοι.....	344

	Λογικός προγραμματισμός με περιορισμούς	346
9.5	Ανάλυση	347
	Συζευκτική κανονική μορφή για τη λογική πρώτης τάξης.....	348
	Ο κανόνας συμπερασμού για την ανάλυση.....	349
	Παραδείγματα απόδειξης	350
	Πληρότητα ανάλυσης.....	352
	Χειρισμός της ισότητας.....	355
	Στρατηγικές ανάλυσης	357
	Προτίμηση μοναδιαίων προτάσεων	357
	Σύνολο υποστήριξης	358
	Ανάλυση εισόδου	358
	Υπαγωγή.....	359
	Αποδείκτες θεωρημάτων.....	359
	Επέκταση της Prolog.....	360
9.6	Σύνοψη	363
	Βιβλιογραφικές και ιστορικές σημειώσεις.....	364
	Ασκήσεις.....	369
10	Αναπαράσταση της Γνώσης	374
10.1	Οντολογική Μηχανική	374
10.2	Κατηγορίες και Αντικείμενα.....	377
	Φυσική σύνθεση.....	378
	Μετρήσεις	380
	Ουσίες και αντικείμενα	382
10.3	Ενέργειες, Καταστάσεις, και Συμβάντα	383
	Η οντολογία του λογισμού καταστάσεων	383
	Περιγραφή ενεργειών στο λογισμό καταστάσεων	385
	Επίλυση του προβλήματος πλαισίου αναπαράστασης.....	387
	Επίλυση του προβλήματος πλαισίου συμπερασμού.....	388
	Χρόνος και λογισμός συμβάντων.....	389
	Γενικευμένα συμβάντα.....	390
	Διαδικασίες.....	392
	Διαστήματα	394
	Ρευστά και αντικείμενα.....	395
10.4	Διανοητικά Συμβάντα και Διανοητικά Αντικείμενα.....	396
	Μια τυπική θεωρία πεποιθήσεων.....	397
	Γνώση και πεποίθηση.....	399
	Γνώση, χρόνος, και δράση	399
10.5	Ο Κόσμος Των Διαδικτυακών Αγορών	400
	Σύγκριση προσφορών.....	404
10.6	Συστήματα Συλλογιστικής για τις Κατηγορίες	406
	Σημασιολογικά δίκτυα	406
	Περιγραφικές λογικές.....	409
10.7	Συλλογιστική με Πληροφορίες Προεπιλογών.....	411
	Ανοικτοί και κλειστοί κόσμοι.....	411
	Η άρνηση ως αποτυχία και η σημασιολογία ευσταθούς μοντέλου	413
	Οροθέτηση και λογική προεπιλογών.....	415
10.8	Συστήματα Διατήρησης Αλήθειας.....	417
10.9	Σύνοψη	419
	Βιβλιογραφικές και Ιστορικές Σημειώσεις	420
	Ασκήσεις.....	427

11	Σχεδιασμός.....	433
11.1	Το Πρόβλημα του Σχεδιασμού.....	433
	Η γλώσσα των προβλημάτων σχεδιασμού.....	435
	Εκφραστικότητα και επεκτάσεις.....	437
	Παράδειγμα: Αερομεταφορά φορτίων.....	439
	Παράδειγμα: Το πρόβλημα της ρεζέρβας.....	439
	Παράδειγμα: Ο κόσμος των κύβων.....	440
11.2	Σχεδιασμός με Αναζήτηση στο Χώρο Καταστάσεων.....	441
	Προς τα εμπρός αναζήτηση στο χώρο καταστάσεων.....	441
	Προς τα πίσω αναζήτηση στο χώρο καταστάσεων.....	443
	Ευρετικές συναρτήσεις για προς τα πίσω αναζήτηση στο χώρο καταστάσεων.....	445
11.3	Σχεδιασμός Μερικής Διάταξης.....	446
	Ένα παράδειγμα προβλήματος σχεδιασμού μερικής διάταξης.....	451
	Σχεδιασμός μερικής διάταξης με μη δεσμευμένες μεταβλητές.....	453
	Ευρετικές συναρτήσεις για το σχεδιασμό μερικής διάταξης.....	454
11.4	Γραφήματα Σχεδιασμού.....	455
	Γραφήματα σχεδιασμού για εκτίμηση ευρετικού μηχανισμού.....	457
	Ο αλγόριθμος GRAPHPLAN.....	459
	Τερματισμός του GRAPHPLAN.....	461
11.5	Σχεδιασμός με Προτασιακή Λογική.....	462
	Περιγραφή προβλημάτων σχεδιασμού με προτασιακή λογική.....	463
	Πολυπλοκότητα των προτασιακών κωδικοποιήσεων.....	466
11.6	Ανάλυση των Προσεγγίσεων Σχεδιασμού.....	468
11.7	Σύνοψη.....	469
	Βιβλιογραφικές και ιστορικές σημειώσεις.....	470
	Ασκήσεις.....	474
12	Σχεδιασμός και Δράση στον Πραγματικό κόσμο.....	479
12.1	Χρόνος, Χρονοδιαγράμματα, και Πόροι.....	479
	Χρονοπρογραμματισμός με περιορισμούς πόρων.....	482
12.2	Σχεδιασμός με Ιεραρχικά Δίκτυα Εργασιών.....	484
	Αναπαράσταση των αποσυνθέσεων ενεργειών.....	485
	Τροποποίηση του πράκτορα σχεδιασμού για τις αποσυνθέσεις.....	488
	Συζήτηση.....	490
12.3	Σχεδιασμός και Δράση σε μη Αιτιοκρατικά Πεδία.....	493
12.4	Υπό Συνθήκη Σχεδιασμός.....	496
	Υπό συνθήκη σχεδιασμός σε πλήρως παρατηρήσιμα περιβάλλοντα.....	496
	Υπό συνθήκη σχεδιασμός σε μερικώς παρατηρήσιμα περιβάλλοντα.....	500
12.5	Παρακολούθηση Εκτέλεσης και Επανασχεδιασμός.....	505
12.6	Συνεχής Σχεδιασμός.....	510
12.7	Πολυπρακτορικός Σχεδιασμός.....	514
	Συνεργασία: συνδυασμένοι στόχοι και πλάνα.....	515
	Πολυσωματικός σχεδιασμός.....	516
	Μηχανισμοί συντονισμού.....	518
	Ανταγωνισμός.....	519
12.8	Σύνοψη.....	520
	Βιβλιογραφικές και ιστορικές σημειώσεις.....	521
	Ασκήσεις.....	525

13	Αβεβαιότητα.....	529
13.1	Δράση Υπό Αβεβαιότητα	529
	Χειρισμός της αβέβαιης γνώσης	530
	Αβεβαιότητα και ορθολογικές αποφάσεις.....	532
	Σχεδίαση πράκτορα που χρησιμοποιεί τη θεωρία αποφάσεων	533
13.2	Βασική Σημειογραφία Πιθανοτήτων.....	534
	Προτάσεις.....	534
	Ατομικά συμβάντα	535
	Εκ των προτέρων πιθανότητα	536
	Υπό συνθήκη πιθανότητα.....	537
13.3	Τα Αξιώματα των Πιθανοτήτων.....	539
	Χρήση των αξιωμάτων των πιθανοτήτων	541
	Γιατί είναι εύλογα τα αξιώματα των πιθανοτήτων.....	541
13.4	Συμπερασμοί Με Πλήρεις Συνδυασμένες Κατανομές Πιθανότητας	543
13.5	Ανεξαρτησία	546
13.6	Ο Κανόνας του Bayes και η Χρήση του.....	548
	Εφαρμογή του κανόνα του Bayes: Η απλή περίπτωση	548
	Χρήση του κανόνα του Bayes: Συνδυασμός μαρτυριών.....	549
13.7	Εκ Νέου Επίσκεψη στον Κόσμο του Wumpus	552
13.8	Σύνοψη	555
	Βιβλιογραφικές και ιστορικές σημειώσεις.....	556
	Ασκήσεις.....	558
14	Πιθανοτική Συλλογιστική	561
14.1	Αναπαράσταση Γνώσης σε ένα Αβέβαιο Πεδίο	561
14.2	Η Σημασιολογία των Δικτύων Bayes.....	564
	Αναπαράσταση της πλήρους συνδυασμένης κατανομής.....	564
	Μια μέθοδος για την κατασκευή δικτύων Bayes	565
	Συμπαγής αναπαράσταση και διάταξη κόμβων	566
	Σχέσεις υπό συνθήκη ανεξαρτησίας σε δίκτυα Bayes.....	568
14.3	Αποδοτική Αναπαράσταση Για Υπό Συνθήκη Κατανομές.....	569
	Δίκτυα Bayes με συνεχείς μεταβλητές.....	571
14.4	Ακριβής Συμπερασμός σε Δίκτυα Bayes	574
	Συμπερασμός με απαρίθμηση	574
	Ο αλγόριθμος απαλοιφής μεταβλητών.....	577
	Η πολυπλοκότητα του ακριβούς συμπερασμού	579
	Αλγόριθμοι ομαδοποίησης.....	580
14.5	Προσεγγιστικός Συμπερασμός σε Δίκτυα Bayes	581
	Μέθοδοι άμεσης δειγματοληψίας.....	581
	Απορριπτική δειγματοληψία σε δίκτυα Bayes	583
	Στάθμιση πιθανοφάνειας.....	584
	Συμπερασμός με προσομοίωση αλυσίδας Markov	587
14.6	Επέκταση των Πιθανοτήτων σε Αναπαραστάσεις Πρώτης Τάξης	590
14.7	Άλλες Προσεγγίσεις στην Αβέβαιη Συλλογιστική.....	594
	Μέθοδοι βασισμένες σε κανόνες για την αβέβαιη συλλογιστική	595
	Αναπαράσταση της άγνοιας: Η θεωρία Dempster-Shafer	597
	Αναπαράσταση της αοριστίας: Ασαφή σύνολα και ασαφής λογική.....	598
14.8	Σύνοψη	600
	Βιβλιογραφικές και ιστορικές σημειώσεις.....	601
	Ασκήσεις.....	605

15	Πιθανοτική Συλλογιστική στο Χρόνο	610
15.1	Χρόνος και Αβεβαιότητα	610
	Καταστάσεις και παρατηρήσεις	611
	Στάσιμες διαδικασίες και η υπόθεση Markov	612
15.2	Συμπερασμός σε Χρονικά Μοντέλα	615
	Φιλτράρισμα και πρόβλεψη	616
	Εξομάλυνση.....	618
	Εύρεση της πλέον πιθανής ακολουθίας.....	621
15.3	Κρυφά Μοντέλα Markov	623
	Απλοποιημένοι αλγόριθμοι πίνακα	623
15.4	Φίλτρα Kalman.....	626
	Ενημέρωση κατανομών Gauss	627
	Ένα απλό μονοδιάστατο παράδειγμα	628
	Η γενική περίπτωση	631
	Εφαρμοσιμότητα του φιλτραρίσματος Kalman	632
15.5	Δυναμικά Δίκτυα Bayes	634
	Κατασκευή δικτύων DBN	635
	Ακριβής συμπερασμός στα δίκτυα DBN	639
	Προσεγγιστικός συμπερασμός στα δίκτυα DBN	640
15.6	Αναγνώριση Ομιλίας	643
	Έχοι ομιλίας.....	645
	Λέξεις	648
	Προτάσεις.....	650
	Δημιουργία ενός συστήματος αναγνώρισης ομιλίας.....	653
15.7	Σύνοψη	654
	Βιβλιογραφικές και Ιστορικές Σημειώσεις	654
	Ασκήσεις.....	657
16	Λήψη Απλών Αποφάσεων.....	660
16.1	Συνδυασμός Πεποιθήσεων και Επιθυμιών υπό Αβεβαιότητα.....	660
16.2	Η Βάση της Θεωρίας Χρησιμότητας.....	662
	Περιορισμοί ως προς τις ορθολογικές προτιμήσεις	662
	Και εγένετο Χρησιμότητα	664
16.3	Συναρτήσεις Χρησιμότητας	665
	Η χρησιμότητα των χρημάτων	665
	Κλίμακες χρησιμότητας και αποτίμηση χρησιμότητας.....	668
16.4	Πολυκριτηριακές Συναρτήσεις Χρησιμότητας	670
	Κυριαρχία.....	670
	Δομή προτίμησης και πολυκριτηριακή χρησιμότητα.....	673
	Προτιμήσεις χωρίς αβεβαιότητα	673
	Προτιμήσεις με αβεβαιότητα	674
16.5	Δίκτυα Αποφάσεων	675
	Αναπαράσταση ενός προβλήματος απόφασης με ένα δίκτυο αποφάσεων.....	675
	Αποτίμηση δικτύων αποφάσεων	677
16.6	Η Αξία των Πληροφοριών.....	677
	Ένα απλό παράδειγμα	678
	Ένας γενικός τύπος.....	678
	Ιδιότητες της αξίας των πληροφοριών	680
	Υλοποίηση ενός πράκτορα συλλογής πληροφοριών.....	681
16.7	Έμπειρα Συστήματα της Θεωρίας Αποφάσεων	682
16.8	Σύνοψη	685

Βιβλιογραφικές και Ιστορικές Σημειώσεις	685
Ασκήσεις.....	687
17 Λήψη Σύνθετων Αποφάσεων	691
17.1 Ακολουθιακά Προβλήματα Αποφάσεων	691
Ένα παράδειγμα	691
Βέλτιστη συμπεριφορά στα ακολουθιακά προβλήματα αποφάσεων	694
17.2 Επανάληψη Αξιών.....	697
Χρησιμότητες καταστάσεων	697
Ο αλγόριθμος επανάληψης αξιών	699
Σύγκλιση της επανάληψης αξιών	699
17.3 Επανάληψη Πολιτικών.....	703
17.4 Μερικώς Παρατηρήσιμες Διαδικασίες Απόφασης Markov	705
17.5 Πράκτορες της Θεωρίας Αποφάσεων	709
17.6 Αποφάσεις με Πολλούς Πράκτορες: Θεωρία Παιγνίων.....	711
17.7 Σχεδίαση Μηχανισμών.....	721
17.8 Σύνοψη	724
Βιβλιογραφικές και Ιστορικές Σημειώσεις	725
Ασκήσεις.....	728
18 Μάθηση από Παρατηρήσεις	731
18.1 Μορφές Μάθησης	731
18.2 Επαγωγική Μάθηση	733
18.3 Μάθηση Δένδρων Αποφάσεων	735
Δένδρα αποφάσεων ως στοιχεία εκτέλεσης	736
Εκφραστικότητα των δένδρων αποφάσεων	737
Επαγωγή δένδρων αποφάσεων από παραδείγματα	738
Επιλογή χαρακτηριστικών ελέγχου	742
Αποτίμηση της απόδοσης του αλγόριθμου μάθησης	743
Θόρυβος και υπερπροσαρμογή	745
Επέκταση της εφαρμοσιμότητας των δένδρων αποφάσεων.....	747
18.4 Συλλογική Μάθηση.....	748
18.5 Γιατί Αποδίδει η Μάθηση: Θεωρία Υπολογιστικής Μάθησης.....	752
Πόσα παραδείγματα χρειάζονται;	752
Μάθηση λίστας αποφάσεων.....	754
Συζήτηση.....	756
18.6 Σύνοψη	757
Βιβλιογραφικές και Ιστορικές Σημειώσεις	758
Ασκήσεις.....	760
19 Η Γνώση στη Μάθηση.....	763
19.1 Μια Λογική Διατύπωση της Μάθησης.....	763
Παραδείγματα και υποθέσεις	764
Αναζήτηση της τρέχουσας βέλτιστης υπόθεσης.....	765
Αναζήτηση ελάχιστης δέσμευσης.....	768
19.2 Η Γνώση στη Μάθηση	772
Μερικά απλά παραδείγματα	773
Μερικά γενικά σχήματα	773
19.3 Μάθηση Βασισμένη στις Εξηγήσεις.....	775
Εξαγωγή γενικών κανόνων από παραδείγματα	776
Βελτίωση της αποδοτικότητας	778

19.4	Μάθηση με Χρήση Πληροφοριών Συνάφειας	780
	Προσδιορισμός του χώρου υποθέσεων	780
	Μάθηση και χρήση των πληροφοριών συνάφειας	781
19.5	Επαγωγικός Λογικός Προγραμματισμός	783
	Παράδειγμα	785
	Μέθοδοι επαγωγικής μάθησης από επάνω προς τα κάτω	786
	Επαγωγική μάθηση με αντίστροφη παραγωγή.....	789
	Πραγματοποίηση ανακαλύψεων με τον επαγωγικό λογικό προγραμματισμό	791
19.6	Σύνοψη	793
	Βιβλιογραφικές και Ιστορικές Σημειώσεις	794
	Ασκήσεις.....	796
20	Στατιστικές Μέθοδοι Μάθησης.....	798
20.1	Στατιστική Μάθηση	798
20.2	Μάθηση με Πλήρη Δεδομένα.....	802
	Μάθηση παραμέτρων μέγιστης πιθανοφάνειας: Διακριτά μοντέλα.....	802
	Απλοϊκά μοντέλα Bayes.....	805
	Μάθηση παραμέτρων μέγιστης πιθανοφάνειας: Συνεχή μοντέλα	806
	Μάθηση παραμέτρων κατά Bayes	807
	Μάθηση των δομών για δίκτυα Bayes	809
20.3	Μάθηση με Κρυφές Μεταβλητές: Ο Αλγόριθμος EM.....	811
	Μη επιβλεπόμενη ομαδοποίηση: Μάθηση μειγμάτων Gauss.....	812
	Μάθηση δικτύων Bayes με κρυφές μεταβλητές	814
	Μάθηση κρυφών μοντέλων Markov	817
	Η γενική μορφή του αλγόριθμου EM.....	818
	Μάθηση δομών δικτύων Bayes με κρυφές μεταβλητές	819
20.4	Μάθηση Βασισμένη σε Στιγμιότυπα.....	820
	Μοντέλα πλησιέστερου γείτονα.....	820
	Μοντέλα πυρήνων	823
20.5	Νευρωνικά Δίκτυα.....	824
	Μονάδες στα νευρωνικά δίκτυα.....	824
	Δομές δικτύου	826
	Νευρωνικά δίκτυα ενός επιπέδου με προς τα εμπρός τροφοδότηση σήματος (perceptrons)	827
	Νευρωνικά δίκτυα πολλών επιπέδων με προς τα εμπρός τροφοδότηση του σήματος.....	832
	Μάθηση δομών νευρωνικών δικτύων	836
20.6	Μηχανές Πυρήνων	837
20.7	Μελέτη Περίπτωσης: Αναγνώριση Χειρόγραφων Ψηφίων.....	840
20.8	Σύνοψη	843
	Βιβλιογραφικές και Ιστορικές Σημειώσεις	844
	Ασκήσεις.....	848
21	Ενισχυτική Μάθηση	852
21.1	Εισαγωγή.....	852
21.2	Παθητική Ενισχυτική Μάθηση	854
	Άμεση εκτίμηση χρησιμότητας.....	855
	Προσαρμόσιμος δυναμικός προγραμματισμός	856
	Μάθηση χρονικών διαφορών	857
21.3	Ενεργητική Ενισχυτική Μάθηση	860

	Εξερεύνηση	861
	Μάθηση μιας συνάρτησης αξίας ενεργειών	865
21.4	Γενίκευση στην Ενισχυτική Μάθηση.....	866
	Εφαρμογές στα παιχνίδια	870
	Εφαρμογή στον έλεγχο ρομπότ	870
21.5	Αναζήτηση Πολιτικής.....	871
21.6	Σύνοψη	874
	Βιβλιογραφικές και Ιστορικές Σημειώσεις	875
	Ασκήσεις.....	878
22	Επικοινωνία	881
22.1	Η Επικοινωνία ως Ενέργεια	882
	Θεμελιώδεις αρχές της γλώσσας.....	883
	Τα συνιστώντα βήματα της επικοινωνίας.....	884
22.2	Τυπική Γραμματική για ένα Τμήμα της Αγγλικής Γλώσσας.....	887
	Το Λεξικό της ϵ_0	887
	Η γραμματική της ϵ_0	888
22.3	Συντακτική Ανάλυση.....	889
	Αποδοτική συντακτική ανάλυση.....	891
22.4	Επαυξημένη Γραμματική	897
	Υποκατηγοριοποίηση ρημάτων.....	900
	Παραγωγική ικανότητα των επαυξημένων γραμματικών	901
22.5	Σημασιολογική Ερμηνεία.....	902
	Η σημασιολογία ενός αποσπάσματος σε Αγγλική γλώσσα	904
	Χρόνος και ρηματικοί χρόνοι.....	904
	Ποσοτικοποίηση.....	905
	Πραγματολογική Ερμηνεία	907
	Παραγωγή γλώσσας με γραμματικές DCG.....	910
22.6	Αμφισημία και Αποσαφήνιση	910
	Αποσαφήνιση	913
22.7	Κατανόηση Πραγματείας	914
	Ανάλυση αναφορών.....	914
	Η δομή μιας συνεκτικής πραγματείας.....	915
22.8	Επαγωγή Γραμματικής.....	917
22.9	Σύνοψη	919
	Βιβλιογραφικές και Ιστορικές Σημειώσεις	920
	Ασκήσεις.....	924
23	Πιθανοτική Επεξεργασία Γλώσσας.....	927
23.1	Πιθανοτικά Μοντέλα Γλωσσών	927
	Πιθανοτικές γραμματικές ανεξάρτητες από τα συμφραζόμενα.....	930
	Μάθηση των πιθανοτήτων για τις γραμματικές PCFG	932
	Μάθηση της δομής κανόνων για τις γραμματικές PCFG.....	933
23.2	Ανάκτηση Πληροφοριών.....	933
	Αξιολόγηση συστημάτων IR.....	936
	Βελτιώσεις IR.....	938
	Παρουσίαση των συνόλων αποτελεσμάτων.....	939
	Υλοποίηση συστημάτων IR	940
23.3	Εξαγωγή Πληροφοριών.....	942
23.4	Μηχανική Μετάφραση.....	945
	Συστήματα μηχανικής μετάφρασης	947

	Στατιστική μηχανική μετάφραση	948
	Μάθηση πιθανοτήτων για τη μηχανική μετάφραση.....	951
23.5	Σύνοψη	953
	Βιβλιογραφικές και Ιστορικές Σημειώσεις	953
	Ασκήσεις.....	956
24	Αντίληψη.....	959
24.1	Εισαγωγή.....	959
24.2	Σχηματισμός Εικόνων	961
	Εικόνες χωρίς φακούς: φωτογραφική μηχανή σημειακής οπής	961
	Συστήματα με φακό.....	962
	Φως: η φωτομετρία του σχηματισμού εικόνας.....	963
	Χρώμα: η φασματοφωτομετρία του σχηματισμού εικόνας.....	964
24.3	Πρώιμες Λειτουργίες Επεξεργασίας Εικόνας.....	965
	Ανίχνευση ακμών	966
	Κατάτμηση εικόνας	969
24.4	Εξαγωγή Τρισδιάστατων Πληροφοριών	970
	Κίνηση.....	972
	Διόφθαλμη στερεοσκοπική όραση	974
	Κλίσεις υφής	976
	Σκίαση	977
	Περίγραμμα	978
24.5	Αναγνώριση Αντικειμένων	982
	Αναγνώριση με βάση τη φωτεινότητα.....	985
	Αναγνώριση με βάση τα χαρακτηριστικά	986
	Εκτίμηση στάσης.....	989
24.6	Χρήση της Όρασης για το Χειρισμό και την Πλοήγηση	991
24.7	Σύνοψη	993
	Βιβλιογραφικές και Ιστορικές Σημειώσεις	993
	Ασκήσεις.....	997
25	Ρομποτική.....	1000
25.1	Εισαγωγή.....	1000
25.2	Υλικό των Ρομπότ	1002
	Αισθητήρες.....	1002
	Συσκευές δράσης.....	1004
25.3	Ρομποτική Αντίληψη	1007
	Εντοπισμός.....	1008
	Χαρτογράφηση.....	1013
	Άλλοι τύποι αντίληψης.....	1016
25.4	Σχεδιασμός Κίνησης	1017
	Χώρος διαμόρφωσης.....	1017
	Μέθοδοι αποσύνθεσης σε κελιά.....	1021
	Μέθοδοι σκελετοποίησης.....	1023
25.5	Σχεδιασμός Αβέβαιων Κινήσεων	1024
	Εύρωστες Μέθοδοι.....	1026
25.6	Κίνηση.....	1028
	Δυναμική και έλεγχος.....	1029
	Έλεγχος πεδίου δυναμικού	1031
	Αντιδραστικός έλεγχος	1032
25.7	Αρχιτεκτονικές Λογισμικού Ρομποτικής.....	1034

Αρχιτεκτονική υπαγωγή.....	1034
Αρχιτεκτονική τριών επιπέδων	1036
Γλώσσες προγραμματισμού ρομποτικής.....	1036
25.8 Πεδία Εφαρμογής.....	1038
25.9 Σύνοψη	1041
Βιβλιογραφικές και Ιστορικές Σημειώσεις	1042
Ασκήσεις.....	1045
26 Φιλοσοφική Θεμελίωση	1050
26.1 Ασθενής TN: Μπορούν να Ενεργήσουν Ευφυώς οι Μηχανές;	1050
Το επιχείρημα από την αδυναμία	1052
Η μαθηματική ένσταση	1053
Το επιχείρημα περί μη τυποποίησης.....	1054
26.2 Ισχυρή TN: Μπορούν Πράγματι να Σκεφτούν οι Μηχανές;	1056
Το πρόβλημα του νου και του σώματος	1058
Το πείραμα του "εγκεφάλου μέσα σε κάδο"	1059
Το πείραμα προσθετικής εγκεφάλου	1060
Το Κινέζικο δωμάτιο	1062
26.3 Η Ηθική και οι Κίνδυνοι της Ανάπτυξης Τεχνητής Νοημοσύνης	1064
26.4 Σύνοψη	1068
Βιβλιογραφικές και Ιστορικές Σημειώσεις	1068
Ασκήσεις.....	1071
27 TN: Παρόν και Μέλλον.....	1072
27.1 Συνιστώσες Πράκτορα	1072
27.2 Αρχιτεκτονικές Πρακτόρων	1075
27.3 Οδεύουμε Προς Τη Σωστή Κατεύθυνση;	1076
27.4 Τι Θα Συμβεί Εάν Πετύχει Η TN;	1079
A Μαθηματικό Υπόβαθρο	1080
A.1 Ανάλυση Πολυπλοκότητας και Σημειογραφία $O()$	1080
Ασυμπτωτική ανάλυση.....	1080
NP και εγγενώς δύσκολα προβλήματα	1081
A.2 Διανύσματα, Πίνακες, και Γραμμική Άλγεβρα	1082
A.3 Κατανομές Πιθανοτήτων.....	1084
Βιβλιογραφικές και Ιστορικές Σημειώσεις	1086
B Σημειώσεις Επί Των Γλωσσών Και Των Αλγορίθμων	1088
B.1 Ορισμός Γλωσσών Με Τη Μορφή Backus–Naur (BNF)	1088
B.2 Περιγραφή Αλγορίθμων με Ψευδοκώδικα	1089
B.3 Βοήθεια στον Ιστό.....	1090
Βιβλιογραφία	1091
Γλωσσάρι.....	1145
Ευρετήριο	1165

Πρόλογος

Η **Τεχνητή Νοημοσύνη, TN** (Artificial Intelligence, AI), είναι ένα μεγάλο πεδίο, και το βιβλίο αυτό είναι επίσης μεγάλο. Προσπαθήσαμε να εξερευνήσουμε ολόκληρο το εύρος του πεδίου, το οποίο συνδυάζει τη λογική, τις πιθανότητες και τα συνεχή μαθηματικά, καθώς και την αντίληψη, τη συλλογιστική, τη μάθηση και τη δράση, και ακόμα όλο το φάσμα από τις μικροηλεκτρονικές συσκευές μέχρι τους ρομποτικούς πλανητικούς εξερευνητές. Το βιβλίο είναι επίσης μεγάλο επειδή προχωρά σε αρκετό βάθος στην παρουσίαση των πρακτικών αποτελεσμάτων, αν και κάναμε προσπάθεια να καλύψουμε μόνο τις πιο κεντρικές ιδέες στο κύριο μέρος του κάθε κεφαλαίου. Παραπομπές για περισσότερα πρακτικά αποτελέσματα δίνονται στις βιβλιογραφικές σημειώσεις στο τέλος του κάθε κεφαλαίου.

Ο δευτερεύων τίτλος του βιβλίου είναι “Μια σύγχρονη προσέγγιση”. Αυτό που θέλαμε να δηλώσουμε με αυτή τη μάλλον κενή φράση είναι ότι προσπαθήσαμε να συνθέσουμε όλα όσα γνωρίζουμε σήμερα σε ένα κοινό πλαίσιο, αντί να επιχειρήσουμε να παρουσιάσουμε το κάθε επιμέρους πεδίο της TN μέσα στο δικό του ιστορικό πλαίσιο. Ζητούμε συγγνώμη από εκείνους που εργάζονται σε πεδία που έγιναν γι’ αυτόν το λόγο λιγότερο αναγνωρίσιμα από ό,τι ίσως θα ήταν διαφορετικά.

Το κύριο ενοποιητικό θέμα είναι η ιδέα του **ευφυή πράκτορα** (intelligent agent). Ορίζουμε την TN ως μελέτη των πρακτόρων (agents) οι οποίοι προσλαμβάνουν αντιλήψεις (percepts) από το περιβάλλον και πραγματοποιούν ενέργειες. Κάθε τέτοιος πράκτορας υλοποιεί μια συνάρτηση που αντιστοιχίζει ακολουθίες αντιλήψεων σε ενέργειες, και εδώ καλύπτουμε διάφορους τρόπους αναπαράστασης αυτών των συναρτήσεων, όπως τα συστήματα παραγωγής (production systems), τους αντιδραστικούς πράκτορες (reactive agents), τα προγράμματα υπο-συνθήκη σχεδιασμού σε πραγματικό χρόνο (real-time conditional planners), τα νευρωνικά δίκτυα (neural networks) και τα συστήματα της θεωρίας των αποφάσεων (decision-theoretic systems). Εξηγούμε το ρόλο της μάθησης ως επέκταση της πρόσβασης του σχεδιαστή σε άγνωστα περιβάλλοντα και δείχνουμε πώς αυτός ο ρόλος θέτει περιορισμούς στη σχεδίαση πρακτόρων, ευνοώντας τη ρητή αναπαράσταση γνώσης και συλλογιστική. Αντιμετωπίζουμε τη ρομποτική και την όραση όχι ως ανεξάρτητα οριζόμενα προβλήματα, αλλά ως μέσα στην υπηρεσία της επίτευξης στόχων. Τονίζουμε τη σημασία του περιβάλλοντος της εργασίας για τον προσδιορισμό της κατάλληλης σχεδίασης πρακτόρων.

Ο πρωταρχικός μας σκοπός είναι να μεταδώσουμε τις ιδέες που έχουν αναδυθεί τα τελευταία πενήντα χρόνια έρευνας στην TN, καθώς και στις προηγούμενες δύο χιλιετίες σχετικής δουλειάς. Προσπαθήσαμε να αποφύγουμε την υπερβολική τυπικότητα στην παρουσίαση αυτών των ιδεών, διατηρώντας όμως την ακρίβεια. Όπου ήταν σκόπιμο, συμπεριλάβαμε αλγόριθμους ψευδοκώδικα για να συγκεκριμενοποιήσουμε τις ιδέες· ο ψευδοκώδικας περιγράφεται συνοπτικά στο Παράρτημα Β. Μπορείτε να βρείτε υλοποιήσεις σε πολλές γλώσσες προγραμματισμού στην ιστοσελίδα του βιβλίου, στη διεύθυνση aima.cs.berkeley.edu.

Το βιβλίο αυτό προορίζεται κυρίως για χρήση σε προπτυχιακά μαθήματα. Μπορεί να χρησιμοποιηθεί και σε μεταπτυχιακά μαθήματα (ίσως με την προσθήκη μερικών από τις κύριες πηγές που προτείνονται στις βιβλιογραφικές σημειώσεις). Λόγω της εκτεταμένης κάλυψης του θέματος και των πολλών λεπτομερών αλγορίθμων, είναι χρήσιμο ως πρωτεύον βιβλίο αναφοράς για τους μεταπτυχιακούς φοιτητές της TN και τους επαγγελματίες που θέλουν να προχωρήσουν πέρα από το επιμέρους πεδίο τους. Το μόνο προαπαιτούμενο είναι η οικειότητα με τις βασικές έννοιες της επιστήμης των υπολογιστών (αλγόριθμοι, δομές δεδομένων, πολυπλοκότητα) σε επίπεδο δευτεροετούς φοιτητή. Ο μαθηματικός λογισμός πρωτοετούς επιπέδου είναι χρήσιμος για

την λεπτομερή κατανόηση των νευρωνικών δικτύων και της στατιστικής μάθησης. Μερικές από τις απαιτούμενες μαθηματικές γνώσεις υποβάθρου παρέχονται στο Παράρτημα Α.

Επισκόπηση του βιβλίου

Το βιβλίο διαιρείται σε οκτώ μέρη. Το Μέρος I, με τίτλο **Τεχνητή νοημοσύνη**, παρουσιάζει μια άποψη του εγχειρήματος της ΤΝ που βασίζεται στην ιδέα των ευφυών πρακτόρων (intelligent agents) — συστήματα που μπορούν να αποφασίζουν τι να κάνουν και μετά να το κάνουν. Το Μέρος II, **Επίλυση προβλημάτων**, εστιάζεται σε μεθόδους με τις οποίες μπορεί κανείς να αποφασίζει τι να κάνει όταν χρειάζεται να προβλέψει πολλά βήματα — για παράδειγμα, όταν ταξιδεύει σε μια χώρα ή όταν παίζει σκάκι. Το Μέρος III, **Γνώση και συλλογιστική**, εξετάζει τρόπους αναπαράστασης γνώσης που αφορά τον κόσμο — πώς λειτουργεί, ποια είναι η τρέχουσα μορφή του, τι αποτέλεσμα μπορεί να έχουν οι πράξεις κάποιου — και τρόπους λογικής συλλογιστικής με αυτή τη γνώση. Το Μέρος IV, **Σχεδιασμός**, εξετάζει στη συνέχεια το πώς μπορεί να χρησιμοποιεί κανείς αυτούς τους τρόπους συλλογιστικής για να αποφασίζει τι να κάνει, ιδιαίτερα με τη δημιουργία *πλάνων*. Το Μέρος V, **Αβέβαιη γνώση και συλλογιστική**, είναι ανάλογο με τα Μέρη III και IV, αλλά εστιάζεται στη συλλογιστική και τη λήψη αποφάσεων όταν υπάρχει *αβεβαιότητα* για τον κόσμο, όπως θα μπορούσε να συμβαίνει, για παράδειγμα, με ένα σύστημα ιατρικής διάγνωσης και θεραπείας.

Στα Μέρη II – V περιγράφεται το μέρος του ευφυή πράκτορα που είναι υπεύθυνο για τη λήψη αποφάσεων. Το Μέρος VI, **Μάθηση**, περιγράφει μεθόδους για την παραγωγή της γνώσης που απαιτούν αυτά τα στοιχεία λήψης αποφάσεων. Το Μέρος VII, **Επικοινωνία, Αντίληψη, και Δράση**, περιγράφει τρόπους με τους οποίους ένας ευφυής πράκτορας μπορεί να αντιλαμβάνεται το περιβάλλον του ώστε να γνωρίζει τι συμβαίνει, είτε με την όραση, είτε με την αφή, είτε με την ακοή, είτε με την κατανόηση της γλώσσας, καθώς και τρόπους με τους οποίους μπορεί να μετατρέπει τα σχέδιά του σε πραγματικές ενέργειες, είτε με ρομποτική κίνηση είτε με φράσεις φυσικής γλώσσας. Τέλος, το Μέρος VIII, **Συμπεράσματα**, αναλύει το παρελθόν και το μέλλον της ΤΝ και τις φιλοσοφικές και ηθικές επιπτώσεις της.

Αλλαγές από την πρώτη έκδοση

Πολλά έχουν αλλάξει στην ΤΝ από τον καιρό της πρώτης έκδοσης του βιβλίου το 1995, και πολλά έχουν επίσης αλλάξει στο βιβλίο. Όλα τα κεφάλαια έχουν αναθεωρηθεί σε μεγάλο βαθμό, για να ληφθεί υπόψη η πιο πρόσφατη δουλειά στο πεδίο, για να επανερμηνευτεί η παλαιότερη δουλειά με έναν τρόπο πιο συνεπή με τα νέα ευρήματα, και για να βελτιωθεί από παιδαγωγική άποψη η ροή των ιδεών. Οι ασχολούμενοι με την ΤΝ θα πρέπει να ενθαρρυνθούν να ασχοληθούν με τις σημερινές τεχνικές, οι οποίες είναι πολύ πιο πρακτικές από εκείνες του 1995· για παράδειγμα οι αλγόριθμοι σχεδιασμού της πρώτης έκδοσης του βιβλίου μπορούσαν να παράγουν μόνο πλάνα με μερικές δεκάδες βήματα, ενώ οι αλγόριθμοι αυτής της έκδοσης φτάνουν σε πλάνα με κλίμακα μεγέθους δεκάδων χιλιάδων βημάτων. Παρόμοια, συναντάμε πολύ μεγάλες βελτιώσεις στον πιθανοτικό συμπερασμό (probabilistic inference), στην επεξεργασία γλώσσας και σε άλλα επιμέρους πεδία. Οι σημαντικότερες αλλαγές στο βιβλίο είναι οι παρακάτω:

- Στο Μέρος I, αναγνωρίζουμε την ιστορική συνεισφορά της θεωρίας ελέγχου, της θεωρίας παιγνίων, των οικονομικών επιστημών και των νευροεπιστημών (neuroscience). Αυτό μας βοηθά να προετοιμάσουμε το έδαφος για μια πιο ολοκληρωμένη κάλυψη αυτών των ιδεών σε επόμενα κεφάλαια.

- Στο Μέρος II, καλύπτονται οι αλγόριθμοι online αναζήτησης και έχει προστεθεί ένα νέο κεφάλαιο για την ικανοποίηση περιορισμών. Αυτό το τελευταίο μας παρέχει μια φυσική διασύνδεση με την ύλη που ασχολείται με τη λογική.
- Στο Μέρος III, η προτασιακή λογική, που στην πρώτη έκδοση παρουσιάστηκε ως προοίμιο της λογικής πρώτης τάξης, παρουσιάζεται τώρα ως μια αφεαυτής χρήσιμη γλώσσα αναπαράστασης, με γρήγορους αλγόριθμους συμπερασμού και με σχεδιάσεις πρακτόρων βασισμένες σε κυκλώματα. Τα κεφάλαια για τη λογική πρώτης τάξης έχουν αναδιοργανωθεί έτσι ώστε να παρουσιάζεται η ύλη με μεγαλύτερη σαφήνεια και έχουμε προσθέσει ως παράδειγμα την εμπορική περιοχή του Διαδικτύου.
- Στο Μέρος IV, συμπεριλάβαμε νεότερες μεθόδους σχεδιασμού, όπως ο αλγόριθμος GRAPHPLAN και ο σχεδιασμός με βάση την ικανοποιησιμότητα (satisfiability-based planning), ενώ καλύψαμε πιο εκτεταμένα το χρονοπρογραμματισμό (scheduling), τον υπό συνθήκη σχεδιασμό (conditional planning), τον ιεραρχικό σχεδιασμό (hierarchical planning) και τον πολυπρακτορικό σχεδιασμό (multiagent planning).
- Στο Μέρος V, επαυξήσαμε την ύλη για τα δίκτυα Bayes με νέους αλγόριθμους, όπως οι αλγόριθμοι απαλοιφής μεταβλητών (variable elimination) και Monte Carlo αλυσίδας Markov (MCMC), ενώ προσθέσαμε ένα νέο κεφάλαιο για την αβέβαιη χρονική συλλογιστική το οποίο καλύπτει τα κρυφά μοντέλα Markov, τα φίλτρα Kalman, και τα δυναμικά δίκτυα Bayes. Οι διαδικασίες αποφάσεων Markov καλύπτονται σε μεγαλύτερο βάθος, και έχουν προστεθεί ενότητες για τη θεωρία των παιγνίων και τη σχεδίαση μηχανισμών.
- Στο Μέρος VI, συνδυάσαμε τις εργασίες πάνω στη στατιστική, τη συμβολική και την νευρωνική μάθηση, και προσθέσαμε ενότητες για τους αλγόριθμους ενδυνάμωσης (boosting), τον αλγόριθμο μεγιστοποίησης αναμονής (EM), τη μάθηση με βάση τα στιγμιότυπα (instance-based), και τις μεθόδους πυρήνων (μηχανές διανυσμάτων υποστήριξης).
- Στο Μέρος VII, στην κάλυψη της επεξεργασίας γλώσσας, έχουν προστεθεί ενότητες για την επεξεργασία λόγου και την επαγωγή γραμματικών, καθώς και ένα κεφάλαιο για τα πιθανοτικά μοντέλα γλώσσας, με εφαρμογές στην ανάκτηση πληροφοριών και τη μηχανική μετάφραση. Η κάλυψη της ρομποτικής δίνει έμφαση στην αφομοίωση αβέβαιων δεδομένων αισθητήρων και το κεφάλαιο για την όραση έχει ενημερωθεί με ύλη για την αναγνώριση αντικειμένων.
- Στο Μέρος VIII, έχουμε προσθέσει μια ενότητα για τις ηθικές επιπτώσεις της ΤΝ.

Πώς να χρησιμοποιήσετε το βιβλίο

Το βιβλίο έχει 27 κεφάλαια, κάθε ένα από τα οποία απαιτεί περίπου μία εβδομάδα διαλέξεων, και επομένως, για να καλυφθεί ολόκληρο το βιβλίο χρειάζονται δύο εξάμηνα. Εναλλακτικά, μπορεί να καταρτιστεί ένα πρόγραμμα προσαρμοσμένο στα ενδιαφέροντα του καθηγητή και του φοιτητή. Με το ευρύ φάσμα που καλύπτει, το βιβλίο μπορεί να χρησιμοποιηθεί για την υποστήριξη τέτοιων μαθημάτων, είτε πρόκειται για σύντομα εισαγωγικά προπτυχιακά μαθήματα είτε για εξειδικευμένα μεταπτυχιακά μαθήματα σε προχωρημένα θέματα. Μπορείτε να βρείτε στον Ιστό παραδείγματα προγραμμάτων σπουδών από περισσότερα από 600 πανεπιστήμια και κολέγια που υιοθέτησαν την πρώτη έκδοση, στη διεύθυνση aima.cs.berkeley.edu, με υποδείξεις που θα σας βοηθήσουν να καταλήξετε σε μια σειρά μαθημάτων κατάλληλη για τις ανάγκες σας.



Το βιβλίο περιέχει 385 ασκήσεις. Οι ασκήσεις που απαιτούν σοβαρό προγραμματισμό είναι σημειωμένες με ένα εικονίδιο πληκτρολογίου. Οι ασκήσεις αυτές μπορούν να απαντηθούν καλύτερα αν εκμεταλλευτεί κανείς τον κώδικα που είναι διαθέσιμος στην ιστοσελίδα aima.cs.berkeley.edu. Μερικές από αυτές είναι αρκετά μεγάλες για να θεωρηθούν εξαμηνιαίες εργασίες. Αρκετές από τις ασκήσεις απαιτούν κάποια έρευνα στη βιβλιογραφία· αυτές είναι σημειωμένες με ένα εικονίδιο βιβλίου.

Σε ολόκληρο το βιβλίο, τα σημαντικά σημεία είναι σημειωμένα με ένα εικονίδιο χεριού που δείχνει. Έχουμε συμπεριλάβει ένα εκτεταμένο ευρετήριο με κάπου 10.000 εγγραφές για να διευκολύνουμε την αναζήτηση όρων στο βιβλίο. Όπου ορίζεται για πρώτη φορά ένας **νέος όρος**, σημειώνεται επίσης στο περιθώριο.

ΝΕΟΣ ΟΡΟΣ

Πώς να χρησιμοποιήσετε την ιστοσελίδα

Στην ιστοσελίδα aima.cs.berkeley.edu θα βρείτε:

- υλοποιήσεις των αλγορίθμων του βιβλίου σε πολλές γλώσσες προγραμματισμού,
- μια λίστα με πάνω από 600 σχολές που έχουν χρησιμοποιήσει το βιβλίο, σε πολλές περιπτώσεις με παραπομπές σε ύλη μαθημάτων διαθέσιμη από τον Ιστό,
- μια λίστα με σχόλια, η οποία περιλαμβάνει πάνω από 800 παραπομπές σε ιστοσελίδες με χρήσιμο περιεχόμενο για την ΤΝ,
- μια λίστα συμπληρωματικής ύλης και παραπομπών, κεφάλαιο προς κεφάλαιο,
- οδηγίες για το πώς μπορείτε να ενταχθείτε σε μια ομάδα συζήτησης για το βιβλίο,
- οδηγίες για το πώς μπορείτε να επικοινωνήσετε με τους συγγραφείς για ερωτήσεις ή σχόλια,
- οδηγίες για το πώς μπορείτε να αναφέρετε λάθη του βιβλίου, τα οποία βέβαια είναι πιθανό να υπάρχουν, και
- τις εικόνες του βιβλίου, καθώς και διαφάνειες και άλλα υλικά για το διδάσκοντα.

Επιπρόσθετο υλικό, που αφορά την ελληνική έκδοση του βιβλίου, μπορείτε να βρείτε στην ιστοσελίδα aima.uom.gr.

Ευχαριστίες

Ο Jitendra Malik έγραψε το μεγαλύτερο μέρος του Κεφαλαίου 24 (για την όραση). Το μεγαλύτερο μέρος του Κεφαλαίου 25 (για τη ρομποτική) το έγραψε ο Sebastian Thrun σε αυτή την έκδοση και ο John Canny στην πρώτη έκδοση. Ο Doug Edwards έκανε την έρευνα για τις ιστορικές σημειώσεις στην πρώτη έκδοση. Οι Tim Huang, Mark Paskin, και Cynthia Bruyns βοήθησαν στη μορφοποίηση των διαγραμμάτων και των αλγορίθμων. Οι Alan Apt, Sondra Chavez, Toni Holm, Jake Warde, Irwin Zucker και Camille Trentacoste από την Prentice Hall έκαναν ό,τι μπορούσαν για να μας κρατήσουν μέσα στις χρονικές προθεσμίες και υπέβαλαν πολλές χρήσιμες υποδείξεις για τη σχεδίαση και το περιεχόμενο του βιβλίου.

Ο Stuart θα ήθελε να ευχαριστήσει τους γονείς του για τη συνεχή υποστήριξη και ενθάρρυνσή τους, και τη σύζυγό του, Loy Sheflott, για την ανεξάντλητη υπομονή της και τη μεγάλη σοφία της. Ελπίζει ότι ο Gordon και η Lucy θα είναι σύντομα σε θέση να διαβάσουν το βιβλίο. Η φοιτητική ομάδα RUGS (Russell's Unusual Group of Students) ήταν ασυνήθιστα πρόθυμη να δώσει τη βοήθειά της.

Ο Peter θα ήθελε να ευχαριστήσει τους γονείς του (Torsten και Gerda) στους οποίους οφείλει το ξεκίνημά του, και τη σύζυγό του (Kris), τα παιδιά και τους φίλους του, για την ενθάρ-

ρυνση και την ανοχή τους στις ατέλειωτες ώρες που έγραφε και τις ακόμα περισσότερες ώρες που ξαναέγραφε.

Είμαστε υπόχρεοι στους βιβλιοθηκάρχους του Berkeley, του Stanford, του MIT και της NASA, καθώς και στους σχεδιαστές του CiteSeer και του Google, οι οποίοι έφεραν επανάσταση στον τρόπο που κάνουμε έρευνα.

Δεν μπορούμε να ευχαριστήσουμε όλους όσοι χρησιμοποίησαν το βιβλίο και έκαναν υποδείξεις, όμως θα θέλαμε να αναγνωρίσουμε τα ιδιαίτερα χρήσιμα σχόλια των Eyal Amir, Krzysztof Apt, Ellery Aziel, Jeff Van Baalen, Brian Baker, Don Barker, Tony Barrett, James Newton Bass, Don Beal, Howard Beck, Wolfgang Bibel, John Binder, Larry Bookman, David R. Boxall, Gerhard Brewka, Selmer Bringsjord, Carla Brodley, Chris Brown, Wilhelm Burger, Lauren Burka, Joao Cachopo, Murray Campbell, Norman Carver, Emmanuel Castro, Anil Chakravarthy, Dan Chisarick, Roberto Cipolla, David Cohen, James Coleman, Julie Ann Comparini, Gary Cottrell, Ernest Davis, Rina Dechter, Tom Dietterich, Chuck Dyer, Barbara Engelhardt, Doug Edwards, Kutluhan Erol, Oren Etzioni, Hana Filip, Douglas Fisher, Jeffrey Forbes, Ken Ford, John Fosler, Alex Franz, Bob Futrelle, Marek Galecki, Stefan Gerberding, Stuart Gill, Sabine Glesner, Seth Golub, Gosta Grahne, Russ Greiner, Eric Grimson, Barbara Grosz, Larry Hall, Steve Hanks, Othar Hansson, Ernst Heinz, Jim Hendler, Christoph Herrmann, Vasant Honavar, Tim Huang, Seth Hutchinson, Joost Jacob, Magnus Johansson, Dan Jurafsky, Leslie Kaelbling, Keiji Kanazawa, Surekha Kasibhatla, Simon Kasif, Henry Kautz, Gernot Kerschbaumer, Richard Kirby, Kevin Knight, Sven Koenig, Daphne Koller, Rich Korf, James Kurien, John Lafferty, Gus Larsson, John Lazzaro, Jon LeBlanc, Jason Leatherman, Frank Lee, Edward Lim, Pierre Louveaux, Don Loveland, Sridhar Mahadevan, Jim Martin, Andy Mayer, David McGrane, Jay Mendelsohn, Brian Milch, Steve Minton, Vibhu Mittal, Leora Morgenstern, Stephen Muggleton, Kevin Murphy, Ron Musick, Sung Myaeng, Lee Naish, Pandu Nayak, Bernhard Nebel, Stuart Nelson, XuanLong Nguyen, Illah Nourbakhsh, Steve Omohundro, David Page, David Palmer, David Parkes, Ron Parr, Mark Paskin, Tony Passera, Michael Pazzani, Wim Pijls, Ira Pohl, Martha Pollack, David Poole, Bruce Porter, Malcolm Pradhan, Bill Pringle, Lorraine Prior, Greg Provan, William Rapaport, Philip Resnik, Francesca Rossi, Jonathan Schaeffer, Richard Scherl, Lars Schuster, Soheil Shams, Stuart Shapiro, Jude Shavlik, Satinder Singh, Daniel Sleator, David Smith, Bryan So, Robert Sproull, Lynn Stein, Larry Stephens, Andreas Stolcke, Paul Stradling, Devika Subramanian, Rich Sutton, Jonathan Tash, Austin Tate, Michael Thielscher, William Thompson, Sebastian Thrun, Eric Tiedemann, Mark Torrance, Randall Upham, Paul Utgoff, Peter van Beek, Hal Varian, Sunil Vemuri, Jim Waldo, Bonnie Webber, Dan Weld, Michael Wellman, Michael Dean White, Kamin Whitehouse, Brian Williams, David Wolfe, Bill Woods, Alden Wright, Richard Yen, Weixiong Zhang, Shlomo Zilberstein, και των ανώνυμων σχολιαστών της Prentice Hall.

Σχετικά με το εξώφυλλο

Η εικόνα για το εξώφυλλο σχεδιάστηκε από τους συγγραφείς και υλοποιήθηκε από τη Lisa Marie Sardegna και τη Maryann Simmons, με χρήση του SGI InventorTM και του Adobe PhotoshopTM. Το εξώφυλλο απεικονίζει τα εξής στοιχεία από την ιστορία της TN:

1. Αλγόριθμος σχεδιασμού του Αριστοτέλη, από το *De Motu Animalium* (περ. 400 π.Χ.).
2. Γεννήτρια εννοιών του Ramon Lull από το *Ars Magna* (περ. 1300).
3. Μηχανή Διαφορών του Charles Babbage, ένα πρωτότυπο για τον πρώτο οικουμενικό υπολογιστή (1848).

8.4 ΜΗΧΑΝΙΚΗ ΓΝΩΣΗΣ ΣΤΗ ΛΟΓΙΚΗ ΠΡΩΤΗΣ ΤΑΞΗΣ

ΜΗΧΑΝΙΚΗ ΤΗΣ
ΓΝΩΣΗΣ

Στην προηγούμενη ενότητα επιδείξαμε τη χρήση της λογικής πρώτης τάξης για την αναπαράσταση της γνώσης σε τρία απλά πεδία. Σε αυτή την ενότητα περιγράφεται η γενική διαδικασία της κατασκευής βάσεων γνώσης — μια διαδικασία που ονομάζεται **μηχανική της γνώσης** (knowledge engineering). Ένας μηχανικός της γνώσης είναι κάποιος που ερευνά ένα συγκεκριμένο πεδίο, μαθαίνει ποιες έννοιες είναι σημαντικές στο πεδίο αυτό και δημιουργεί μια τυπική αναπαράσταση των αντικειμένων και των σχέσεων του πεδίου. Θα επιδείξουμε τη διαδικασία της μηχανικής της γνώσης σε ένα πεδίο των ηλεκτρονικών κυκλωμάτων που μάλλον σας είναι ήδη αρκετά οικείο, ώστε να μπορέσουμε να εστιάσουμε την προσοχή μας στα σχετικά ζητήματα αναπαράστασης. Η προσέγγιση που θα ακολουθήσουμε είναι κατάλληλη για την ανάπτυξη βάσεων γνώσης *ειδικής χρήσης* των οποίων το πεδίο είναι προσεκτικά οροθετημένο και το φάσμα των ερωτημάτων είναι γνωστό από πριν. Οι βάσεις γνώσης *γενικής χρήσης*, που προορίζονται για την υποστήριξη ερωτημάτων που καλύπτουν όλο το φάσμα της ανθρώπινης γνώσης, εξετάζονται στο Κεφάλαιο 10.

Η διαδικασία της μηχανικής της γνώσης

Τα έργα μηχανικής της γνώσης παρουσιάζουν ευρεία ποικιλία ως προς το περιεχόμενο, την εμβέλεια και τη δυσκολία, όμως όλα αυτά τα έργα περιλαμβάνουν τα παρακάτω βήματα:

1. *Αναγνώριση της εργασίας.* Ο μηχανικός γνώσης πρέπει να σκιαγραφήσει το φάσμα των ερωτήσεων που θα υποστηρίζει η βάση γνώσης και τα είδη των γεγονότων που θα είναι διαθέσιμα για κάθε συγκεκριμένο στιγμιότυπο προβλήματος. Για παράδειγμα, χρειάζεται η βάση γνώσης του wumpus να έχει τη δυνατότητα να επιλέγει ενέργειες ή απαιτείται μόνο να απαντά σε ερωτήσεις για τα περιεχόμενα του περιβάλλοντος; Θα περιλαμβάνουν τα γεγονότα των αισθητήρων την τρέχουσα θέση; Η εργασία θα προσδιορίσει τι γνώση θα πρέπει να αναπαρίσταται για να συνδέονται τα στιγμιότυπα προβλημάτων με απαντήσεις. Το βήμα αυτό είναι ανάλογο με τη διαδικασία PEAS για τη σχεδίαση πρακτόρων, που είδαμε στο Κεφάλαιο 2.
2. *Συναρμολόγηση της συναφούς γνώσης.* Ο μηχανικός γνώσης μπορεί να είναι ήδη ειδικός στο πεδίο ή μπορεί να χρειάζεται να συνεργαστεί με πραγματικούς ειδικούς για να εξάγει ό,τι γνωρίζουν — μια διαδικασία που λέγεται **απόκτηση γνώσης** (knowledge acquisition). Στο στάδιο αυτό η γνώση δεν αναπαρίσταται τυπικά. Η ιδέα είναι να κατανοηθεί η εμβέλεια της βάσης γνώσης, όπως προσδιορίζεται από την εργασία, και να κατανοηθεί το πώς λειτουργεί πραγματικά το πεδίο.

Για τον κόσμο του wumpus, ο οποίος ορίζεται με ένα τεχνητό σύνολο κανόνων, η συναφής γνώση είναι εύκολο να προσδιοριστεί. (Προσέξτε όμως ότι ο ορισμός της γειτνίασης δεν δόθηκε ρητά στους κανόνες του κόσμου του wumpus.) Στα πραγματικά πεδία, το ζήτημα της συνάφειας μπορεί να είναι πολύ δύσκολο — για παράδειγμα, ένα σύστημα για την προσομοίωση σχεδιάσεων VLSI θα μπορούσε να χρειάζεται ή να μη χρειάζεται να παίρνει υπόψη τις χωρητικότητες διασποράς και τα φαινόμενα επιφανειακού ρεύματος.

3. *Απόφαση για ένα λεξιλόγιο κατηγορημάτων, συναρτήσεων και σταθερών.* Δηλαδή, μετάφραση των σημαντικών εννοιών του επιπέδου πεδίου σε ονόματα του επιπέδου λογικής. Αυτό περιλαμβάνει πολλές ερωτήσεις για το *στιλ* της μηχανικής της γνώσης. Όπως και το στιλ προγραμματισμού, αυτό μπορεί να έχει σοβαρές επιπτώσεις για την τελική επιτυχία του έργου. Για παράδειγμα, θα πρέπει οι γούβες να αναπαρίστανται με αντικείμενα ή με ένα μοναδιαίο κατηγορημα για τα τετράγωνα; Θα πρέπει ο προσανατολισμός του πράκτορα να είναι συνάρτηση ή κατηγορημα; Θα πρέπει η θέση του wumpus να εξαρτάται από

ΑΠΟΚΤΗΣΗ ΓΝΩΣΗΣ

ΟΝΤΟΛΟΓΙΑ

το χρόνο; Αφού γίνουν οι επιλογές, το αποτέλεσμα είναι ένα λεξιλόγιο που είναι γνωστό ως **οντολογία** (ontology) του πεδίου. Η λέξη *οντολογία* σημαίνει μια συγκεκριμένη θεωρία για τη φύση τού είναι ή της ύπαρξης. Η οντολογία προσδιορίζει τι είδη πραγμάτων υπάρχουν, αλλά δεν προσδιορίζει τις συγκεκριμένες ιδιότητές τους και τις αμοιβαίες σχέσεις τους.

4. *Κωδικοποίηση γενικής γνώσης για το πεδίο.* Ο μηχανικός γνώσης καταγράφει τα αξιώματα για όλους τους όρους του λεξιλογίου. Αυτό συγκεκριμενοποιεί (στο βαθμό που είναι δυνατό) το νόημα των όρων, επιτρέποντας στον ειδικό να εξετάζει το περιεχόμενο. Συχνά αυτό το βήμα αποκαλύπτει παρανοήσεις ή χάσματα στο λεξιλόγιο, που πρέπει να διορθωθούν με επιστροφή στο βήμα 3 και με επανάληψη της διαδικασίας.
5. *Κωδικοποίηση μιας περιγραφής του συγκεκριμένου στιγμιότυπου προβλήματος.* Αν η οντολογία είναι καλά μελετημένη, αυτό το βήμα θα είναι εύκολο. Περιλαμβάνει το γράψιμο απλών ατομικών προτάσεων για στιγμιότυπα εννοιών που είναι ήδη μέρος της οντολογίας. Σε ένα λογικό πράκτορα, τα στιγμιότυπα προβλημάτων παρέχονται από τους αισθητήρες ενώ μια “ασώματη” βάση γνώσης εφοδιάζεται με πρόσθετες προτάσεις με τον ίδιο τρόπο που τα παραδοσιακά προγράμματα τροφοδοτούνται με δεδομένα εισόδου.
6. *Υποβολή ερωτημάτων στη διαδικασία συμπερασμού και λήψη απαντήσεων.* Εδώ βρίσκεται η ανταμοιβή: μπορούμε να αφήσουμε τη διαδικασία συμπερασμού να λειτουργήσει με βάση τα αξιώματα και τα ειδικά για το πρόβλημα γεγονότα για να προκύψουν τα γεγονότα που μας ενδιαφέρει να γνωρίζουμε.
7. *Αποσφαλμάτωση της βάσης γνώσης.* Αλίμονο, οι απαντήσεις στα ερωτήματα σπάνια θα είναι σωστές στην πρώτη προσπάθεια. Για την ακρίβεια, οι απαντήσεις θα είναι σωστές για τη βάση γνώσης όπως είναι γραμμένη, αν θεωρήσουμε ότι η διαδικασία συμπερασμού είναι ορθή, αλλά δεν θα είναι εκείνες που περιμένει ο χρήστης. Για παράδειγμα, αν ένα αξίωμα λείπει, μερικά ερωτήματα δεν θα μπορούν να απαντηθούν από τη βάση γνώσης. Το επάκουλο μπορεί να είναι μια μεγάλη διαδικασία αποσφαλμάτωσης. Τα αξιώματα που λείπουν ή τα πολύ ασθενή αξιώματα μπορούν να προσδιορίζονται εύκολα με την επισήμανση σημείων όπου η αλυσίδα της συλλογιστικής σταματά απροσδόκητα. Για παράδειγμα, αν η βάση γνώσης περιλαμβάνει το ένα από τα διαγνωστικά αξιώματα για τις γούβες,

$$\forall s \text{ ΈχειΆδρα}(s) \Rightarrow \exists r \text{ Γειτονικό}(r, s) \wedge \text{Γούβα}(r),$$

αλλά όχι και το άλλο, τότε ο πράκτορας δε θα μπορεί ποτέ να αποδείξει την *απουσία* γουβών. Τα λανθασμένα αξιώματα μπορούν να προσδιορίζονται επειδή είναι ψευδείς δηλώσεις για τον κόσμο. Για παράδειγμα, η πρόταση

$$\forall x \text{ ΑριθμόςΠοδιών}(x, 4) \Rightarrow \text{Θηλαστικό}(x)$$

είναι ψευδής για τα ερπετά, τα αμφίβια και, το σημαντικότερο, για τα τραπεζία. Το ψεύδος αυτής της πρότασης μπορεί να προσδιοριστεί ανεξάρτητα από την υπόλοιπη βάση γνώσης. Αντίθετα, ένα σύνηθες σφάλμα σε ένα πρόγραμμα μοιάζει κάπως έτσι:

$$\text{απόσταση} = \text{θέση} + 1.$$

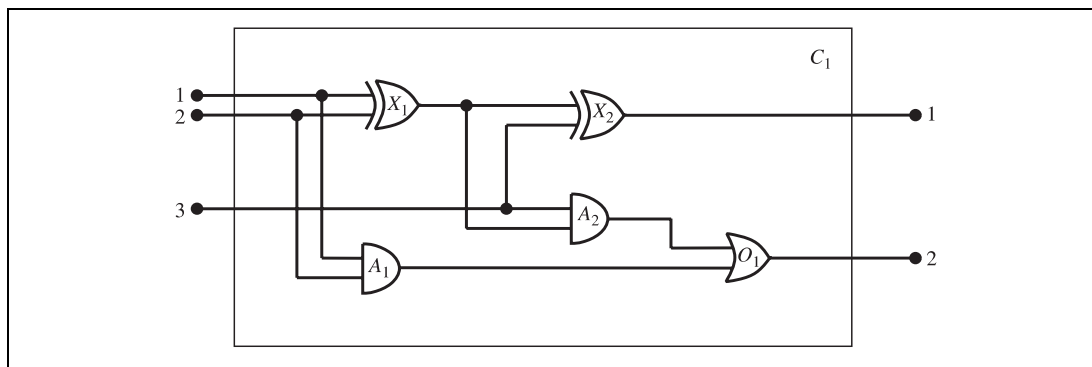
Είναι αδύνατο να πούμε αν αυτή η εντολή είναι σωστή χωρίς να κοιτάξουμε το υπόλοιπο πρόγραμμα για να δούμε, για παράδειγμα, αν η μεταβλητή *απόσταση* αναφέρεται στην τρέχουσα θέση ή σε ένα βήμα μετά την τρέχουσα θέση, ή αν η τιμή της μεταβλητής *θέση* μεταβάλλεται από μian άλλη εντολή και έτσι η *απόσταση* θα πρέπει επίσης να αλλάξει ξανά.

Για να γίνει καλύτερα κατανοητή αυτή η διαδικασία των επτά βημάτων, θα την εφαρμόσουμε τώρα σε ένα εκτεταμένο παράδειγμα — το πεδίο των ηλεκτρονικών κυκλωμάτων.



Το πεδίο των ηλεκτρονικών κυκλωμάτων

Θα αναπτύξουμε μια οντολογία και μια βάση γνώσης που μας επιτρέπουν να συλλογιστούμε για τα ψηφιακά κυκλώματα του είδους που παρουσιάζεται στην Εικόνα 8.4. Ακολουθούμε τη διαδικασία επτά βημάτων της μηχανικής γνώσης.



Εικόνα 8.4 Ψηφιακό κύκλωμα C_1 , που υποτίθεται ότι είναι ένας πλήρης αθροιστής του ενός bit. Οι πρώτες δύο εισόδους είναι τα δύο bit που προστίθενται και η τρίτη είσοδος είναι ένα bit κρατούμενου. Η πρώτη έξοδος είναι το άθροισμα και η δεύτερη έξοδος είναι ένα bit κρατούμενου για τον επόμενο αθροιστή. Το κύκλωμα περιέχει δύο πύλες XOR, δύο πύλες AND και μία πύλη OR.

Αναγνώριση της εργασίας

Υπάρχουν πολλές εργασίες συλλογιστικής σχετικές με τα ψηφιακά κυκλώματα. Στο ανώτατο επίπεδο, αναλύουμε τη λειτουργικότητα του κυκλώματος. Για παράδειγμα, προσθέτει πραγματικά σωστά το κύκλωμα της Εικόνας 8.4; Αν όλες οι εισόδους είναι σε κατάσταση *υψηλό* (high), τι έξοδο δίνει η πύλη A_2 ; Οι ερωτήσεις για τη δομή του κυκλώματος επίσης παρουσιάζουν ενδιαφέρον. Για παράδειγμα, τι είδους είναι όλες οι πύλες που είναι συνδεδεμένες στον πρώτο ακροδέκτη εισόδου; Περιέχει το κύκλωμα βρόχους ανάδρασης; Αυτές είναι οι εργασίες μας σε αυτή την ενότητα. Υπάρχουν πιο λεπτομερή επίπεδα ανάλυσης, μεταξύ των οποίων εκείνα που σχετίζονται με τις καθυστερήσεις χρονισμού, την επιφάνεια του κυκλώματος, την κατανάλωση ενέργειας, το κόστος παραγωγής κ.λπ. Κάθε ένα από αυτά τα επίπεδα θα απαιτούσε επιπλέον γνώση.

Συναρμολόγηση της συναφούς γνώσης

Τι γνωρίζουμε για τα ψηφιακά κυκλώματα; Για τους σκοπούς μας, αποτελούνται από αγωγούς και πύλες. Σήματα ρέουν μέσω των αγωγών στους ακροδέκτες εισόδου των πυλών και κάθε πύλη παράγει ένα σήμα στον ακροδέκτη εξόδου το οποίο ρέει μέσω ενός άλλου αγωγού. Για να προσδιορίσουμε τι θα είναι αυτά τα σήματα, χρειάζεται να γνωρίζουμε πώς μετασχηματίζουν οι πύλες τα σήματα εισόδου τους. Υπάρχουν τέσσερις τύποι πυλών: οι πύλες AND, OR και XOR έχουν δύο ακροδέκτες εισόδου, και οι πύλες NOT έχουν ένα. Όλες οι πύλες έχουν έναν ακροδέκτη εξόδου. Τα κυκλώματα, όπως και οι πύλες, έχουν ακροδέκτες εισόδου και εξόδου.

Για να συλλογιστούμε για τη λειτουργικότητα και τη συνδετικότητα, δε χρειάζεται να μιλάμε για τους ίδιους τους αγωγούς, τις διαδρομές που ακολουθούν ή τους κόμβους όπου ενώνονται δύο αγωγοί. Το μόνο που έχει σημασία είναι οι συνδέσεις μεταξύ ακροδεκτών — μπορούμε να λέμε ότι ένας ακροδέκτης εξόδου συνδέεται με έναν άλλο ακροδέκτη εισόδου χωρίς να χρειάζεται να αναφέρουμε τον αγωγό που τους συνδέει πραγματικά. Υπάρχουν πολλοί άλλοι παράγο-

ντες του πεδίου αυτού που είναι αδιάφοροι για την ανάλυσή μας, όπως το μέγεθος, το σχήμα, το χρώμα ή το κόστος των διαφόρων στοιχείων.

Αν ο σκοπός μας ήταν κάποιος άλλος, πέρα από την επαλήθευση σχεδιάσεων σε επίπεδο πυλών, η οντολογία θα ήταν διαφορετική. Για παράδειγμα, αν μας ενδιέφερε η αποσφαλμάτωση των ελαττωματικών κυκλωμάτων, τότε θα ήταν ίσως καλή ιδέα να συμπεριλάβουμε τους αγωγούς στην οντολογία, επειδή ένας ελαττωματικός αγωγός μπορεί να αλλοιώσει το σήμα που ρέει σε αυτόν. Για την αντιμετώπιση σφαλμάτων χρονισμού, θα χρειαζόταν να συμπεριλάβουμε τις καθυστερήσεις πυλών. Αν μας ενδιέφερε να σχεδιάσουμε ένα προϊόν που θα ήταν επικερδές, τότε το κόστος του κυκλώματος και η ταχύτητά του σε σύγκριση με άλλα προϊόντα της αγοράς θα ήταν σημαντικοί παράγοντες.

Απόφαση για ένα λεξιλόγιο

Γνωρίζουμε τώρα ότι θέλουμε να συζητήσουμε για κυκλώματα, ακροδέκτες, σήματα και πύλες. Το επόμενο βήμα είναι να επιλέξουμε συναρτήσεις, κατηγορήματα και σταθερές για να τα αναπαριστούμε. Θα ξεκινήσουμε από τις μεμονωμένες πύλες και θα ανεβούμε στα κυκλώματα.

Πρώτα χρειαζόμαστε τη δυνατότητα να διακρίνουμε μία πύλη από τις άλλες πύλες. Αυτό αντιμετωπίζεται με την ονομασία των πυλών με σταθερές: X_1, X_2 , κ.ο.κ. Αν και κάθε πύλη συνδέεται στο κύκλωμα με το δικό της τρόπο, η *συμπεριφορά* της — ο τρόπος που μετασχηματίζει τα σήματα εισόδου σε σήματα εξόδου — εξαρτάται μόνο από τον *τύπο* της. Μπορούμε να χρησιμοποιούμε μια συνάρτηση για να αναφερόμαστε στον τύπο της πύλης.¹¹ Για παράδειγμα, μπορούμε να γράφουμε $\text{Τύπος}(X_1) = \text{XOR}$. Η συνάρτηση αυτή εισάγει τη σταθερά XOR (αποκλειστικό ή) για ένα συγκεκριμένο τύπο πύλης: θα ονομάσουμε τις άλλες σταθερές OR (είτε), AND (και) και NOT (όχι). Η συνάρτηση Τύπος δεν είναι ο μόνος τρόπος για να κωδικοποιηθεί αυτή η οντολογική διάκριση. Θα μπορούσαμε να χρησιμοποιήσουμε ένα δυαδικό κατηγορήμα, $\text{Τύπος}(X_1, \text{XOR})$, ή πολλά μεμονωμένα κατηγορήματα τύπων, όπως $\text{XOR}(X_1)$. Οποιαδήποτε από αυτές τις επιλογές θα ήταν ικανοποιητική, αλλά επιλέγοντας τη συνάρτηση Τύπος αποφεύγουμε την ανάγκη για ένα αξίωμα που να λέει ότι κάθε μεμονωμένη πύλη μπορεί να έχει μόνο έναν τύπο. Η σημασιολογία των συναρτήσεων μας το εγγυάται ήδη.

Έπειτα εξετάζουμε του ακροδέκτες. Μια πύλη ή κύκλωμα μπορεί να έχει έναν ή περισσότερους ακροδέκτες εισόδου και έναν ή περισσότερους ακροδέκτες εξόδου. Θα μπορούσαμε απλώς να τους ονομάσουμε έναν-έναν με μια σταθερά, όπως ακριβώς ονομάσαμε τις πύλες. Έτσι, η πύλη X_1 θα μπορούσε να έχει τους ακροδέκτες $X_1\text{Eισ}_1, X_1\text{Eισ}_2$ και $X_1\text{Eξ}_1$. Η τάση αυτή να παράγονται μεγάλα σύνθετα ονόματα θα πρέπει όμως να αποφεύγεται. Ονομάζοντας κάτι $X_1\text{Eισ}_1$ δεν το κάνουμε αυτόματα πρώτη είσοδο του X_1 : χρειάζεται και πάλι να δηλώσουμε αυτό το γεγονός με ένα ρητό ισχυρισμό. Είναι μάλλον καλύτερο να ονομάσουμε την πύλη χρησιμοποιώντας μια συνάρτηση, όπως ακριβώς ονομάσαμε το αριστερό πόδι του Βασιλιά Ιωάννη *ΑριστερόΠόδι(Ιωάννης)*. Έστω λοιπόν ότι $\text{Eισ}(1, X_1)$ υποδηλώνει τον πρώτο ακροδέκτη εισόδου της πύλης X_1 . Μια παρόμοια συνάρτηση Eξ χρησιμοποιείται για τους ακροδέκτες εξόδου.

Η συνδετικότητα μεταξύ πυλών μπορεί να αναπαρίσταται με το κατηγορήμα *Συνδεδεμένο*, το οποίο δέχεται ως ορίσματα δύο ακροδέκτες, για παράδειγμα, $\text{Συνδεδεμένο}(\text{Eξ}(1, X_1), \text{Eισ}(1, X_2))$.

Τέλος χρειάζεται να γνωρίζουμε αν ένα σήμα είναι “ενεργό” (on) ή “ανενεργό” (off). Μία δυνατότητα είναι να χρησιμοποιηθεί ένα μοναδιαίο κατηγορήμα, το *Ενεργό*, το οποίο είναι αληθές όταν το σήμα σε έναν ακροδέκτη είναι “ενεργό”. Αυτό όμως δυσκολεύει κάπως την υποβολή ερωτήσεων όπως “Ποιες είναι όλες οι δυνατές τιμές των σημάτων στους ακροδέκτες εξόδου του κυκλώματος C1;” Γι’ αυτό θα εισαγάγουμε ως αντικείμενα δύο “τιμές σήματος”, τις 1 και 0, και

¹¹ Σημειώστε ότι χρησιμοποιούμε ονόματα που αρχίζουν με κατάλληλα γράμματα — A_1, X_1 , κ.ο.κ. — μόνο και μόνο για να κάνουμε το παράδειγμα πιο ευανάγνωστο. Η βάση γνώσης πρέπει και πάλι να περιέχει πληροφορίες τύπου για τις πύλες.

μια συνάρτηση $\Sigma\eta\mu\alpha$ που δέχεται ως όρισμα έναν ακροδέκτη και υποδηλώνει την τιμή του σήματος σε αυτόν τον ακροδέκτη.

Κωδικοποίηση γενικής γνώσης του πεδίου

Μία ένδειξη ότι έχουμε καλή οντολογία είναι το να υπάρχουν πολύ λίγοι γενικοί κανόνες που χρειάζεται να καθοριστούν. Μια ένδειξη ότι έχουμε καλό λεξιλόγιο είναι το ότι κάθε κανόνας μπορεί να διατυπωθεί με σαφήνεια και συνοπτικότητα. Στο παράδειγμά μας χρειαζόμαστε μόνο επτά απλούς κανόνες για να περιγράψουμε όλα όσα χρειάζεται να γνωρίζουμε για τα κυκλώματα:

1. Αν δύο ακροδέκτες είναι συνδεδεμένοι, τότε έχουν το ίδιο σήμα:

$$\forall t_1, t_2 \text{ Συνδεδεμένο}(t_1, t_2) \Rightarrow \Sigma\eta\mu\alpha(t_1) = \Sigma\eta\mu\alpha(t_2)$$

2. Το σήμα σε κάθε ακροδέκτη είναι είτε 1 είτε 0 (αλλά όχι και τα δύο):

$$\forall t \Sigma\eta\mu\alpha(t) = 1 \vee \Sigma\eta\mu\alpha(t) = 0 \\ 1 \neq 0$$

3. Το κατηγορημα Συνδεδεμένο είναι αντιμεταθετικό:

$$\forall t_1, t_2 \text{ Συνδεδεμένο}(t_1, t_2) \Leftrightarrow \text{Συνδεδεμένο}(t_2, t_1)$$

4. Η έξοδος μιας πύλης OR είναι 1 εάν και μόνο εάν οποιαδήποτε από τις εισόδους της είναι 1:

$$\forall g \text{ Τύπος}(g) = OR \Rightarrow \\ \Sigma\eta\mu\alpha(E\zeta(1, g)) = 1 \Leftrightarrow \exists n \Sigma\eta\mu\alpha(Eισ(n, g)) = 1$$

5. Η έξοδος μιας πύλης AND είναι 0 εάν και μόνο εάν οποιαδήποτε από τις εξόδους της είναι 0:

$$\forall g \text{ Τύπος}(g) = AND \Rightarrow \\ \Sigma\eta\mu\alpha(E\zeta(1, g)) = 0 \Leftrightarrow \exists n \Sigma\eta\mu\alpha(Eισ(n, g)) = 0$$

6. Η έξοδος μιας πύλης XOR είναι 1 εάν και μόνο εάν οι εισοδοί της είναι διαφορετικές:

$$\forall g \text{ Τύπος}(g) = XOR \Rightarrow \\ \Sigma\eta\mu\alpha(E\zeta(1, g)) = 1 \Leftrightarrow \Sigma\eta\mu\alpha(Eισ(1, g)) \neq \Sigma\eta\mu\alpha(Eισ(2, g))$$

7. Η έξοδος μια πύλης NOT είναι διαφορετική από την είσοδό της:

$$\forall g (\text{Τύπος}(g) = NOT) \Rightarrow \Sigma\eta\mu\alpha(E\zeta(1, g)) \neq \Sigma\eta\mu\alpha(Eισ(1, g))$$

Κωδικοποίηση του συγκεκριμένου στιγμιότυπου προβλήματος

Το κύκλωμα που εμφανίζεται στην Εικόνα 8.4 είναι κωδικοποιημένο ως κύκλωμα C_1 με την παρακάτω περιγραφή. Πρώτα κατατάσσουμε τις πύλες σε κατηγορίες:

$$\begin{array}{ll} \text{Τύπος}(X_1) = XOR & \text{Τύπος}(X_2) = XOR \\ \text{Τύπος}(A_1) = AND & \text{Τύπος}(A_2) = AND \\ \text{Τύπος}(O_1) = OR & \end{array}$$

Έπειτα δείχνουμε τις συνδέσεις μεταξύ τους:

$$\begin{array}{ll} \text{Συνδεδεμένο}(E\zeta(1, X_1), Eισ(1, X_2)) & \text{Συνδεδεμένο}(Eισ(1, C_1), Eισ(1, X_1)) \\ \text{Συνδεδεμένο}(E\zeta(1, X_1), Eισ(2, A_2)) & \text{Συνδεδεμένο}(Eισ(1, C_1), Eισ(1, A_1)) \\ \text{Συνδεδεμένο}(E\zeta(1, A_2), Eισ(1, O_1)) & \text{Συνδεδεμένο}(Eισ(2, C_1), Eισ(2, X_1)) \\ \text{Συνδεδεμένο}(E\zeta(1, A_1), Eισ(2, O_1)) & \text{Συνδεδεμένο}(Eισ(2, C_1), Eισ(2, A_1)) \end{array}$$

$$\begin{array}{ll} \text{Συνδεδεμένο}(E\zeta(1, X_2), E\zeta(1, C_1)) & \text{Συνδεδεμένο}(E\sigma(3, C_1), E\sigma(2, X_2)) \\ \text{Συνδεδεμένο}(E\zeta(1, O_1), E\zeta(2, C_1)) & \text{Συνδεδεμένο}(E\sigma(3, C_1), E\sigma(1, A_2)) \end{array}$$

Υποβολή ερωτημάτων στη διαδικασία συμπερασμού

Ποιοι συνδυασμοί εισόδων θα έκαναν την πρώτη έξοδο του C_1 (το bit αθροίσματος) να γίνει 0 και τη δεύτερη έξοδο του C_1 (το bit κρατουμένου) να γίνει 1;

$$\begin{aligned} \exists i_1, i_2, i_3 \quad \text{Σήμα}(E\sigma(1, C_1)) = i_1 \wedge \text{Σήμα}(E\sigma(2, C_1)) = i_2 \wedge \text{Σήμα}(E\sigma(3, C_1)) = i_3 \\ \wedge \text{Σήμα}(E\zeta(1, C_1)) = 0 \wedge \text{Σήμα}(E\zeta(2, C_1)) = 1. \end{aligned}$$

Οι απαντήσεις είναι αντικαταστάσεις των μεταβλητών i_1, i_2 και i_3 τέτοιες ώστε η πρόταση που προκύπτει να καλύπτεται από τη βάση γνώσης. Υπάρχουν τρεις τέτοιες αντικαταστάσεις:

$$\{i_1/1, i_2/1, i_3/0\} \quad \{i_1/1, i_2/0, i_3/1\} \quad \{i_1/0, i_2/1, i_3/1\}.$$

Ποια είναι τα δυνατά σύνολα τιμών όλων των ακροδεκτών για το κύκλωμα του αθροιστή;

$$\begin{aligned} \exists i_1, i_2, i_3, o_1, o_2 \quad \text{Σήμα}(E\sigma(1, C_1)) = i_1 \wedge \text{Σήμα}(E\sigma(2, C_1)) = i_2 \\ \wedge \text{Σήμα}(E\sigma(3, C_1)) = i_3 \wedge \text{Σήμα}(E\zeta(1, C_1)) = o_1 \wedge \text{Σήμα}(E\zeta(2, C_1)) = o_2. \end{aligned}$$

Το τελευταίο ερώτημα θα μας επιστρέψει έναν πλήρη πίνακα εισόδων-εξόδων για τη συσκευή, ο οποίος μπορεί μετά να χρησιμοποιηθεί για να ελεγχθεί αν πραγματικά προσθέτει τις εισόδους της σωστά. Αυτό είναι ένα απλό παράδειγμα **επαλήθευσης κυκλώματος** (circuit verification). Μπορούμε επίσης να χρησιμοποιήσουμε τον ορισμό του κυκλώματος για να κατασκευάσουμε μεγαλύτερα ψηφιακά συστήματα, για τα οποία μπορεί να πραγματοποιηθεί το ίδιο είδος διαδικασίας επαλήθευσης (δείτε την Άσκηση 8.17). Πολλά πεδία επιδέχονται το ίδιο είδος δομημένης ανάπτυξης βάσεων γνώσης, όπου οι πιο σύνθετες έννοιες ορίζονται πάνω σε απλούστερες έννοιες.

Αποσφαλμάτωση της βάσης γνώσης

Μπορούμε να διαταράξουμε τη βάση γνώσης με διάφορους τρόπους για να δούμε τι είδους λανθασμένες συμπεριφορές εμφανίζονται. Για παράδειγμα, ας υποθέσουμε ότι παραλείπουμε τον ισχυρισμό $1 \neq 0$.¹² Ξαφνικά το σύστημα θα είναι ανίκανο να αποδείξει οποιεσδήποτε εξόδους του κυκλώματος, εκτός από τις περιπτώσεις με εισόδους 000 και 110. Μπορούμε να εντοπίσουμε το πρόβλημα ζητώντας τις εξόδους της κάθε πύλης. Για παράδειγμα, μπορούμε να ζητήσουμε το

$$\exists i_1, i_2, o \quad \text{Σήμα}(E\sigma(1, C_1)) = i_1 \wedge \text{Σήμα}(E\sigma(2, C_1)) = i_2 \wedge \text{Σήμα}(E\zeta(1, X_1)) = o$$

το οποίο αποκαλύπτει ότι καμία έξοδος δεν είναι γνωστή στην πύλη X_1 για τις περιπτώσεις των εισόδων 10 και 01. Έπειτα κοιτάζουμε το αξίωμα για τις πύλες XOR, όπως εφαρμόζεται στην X_1 :

$$\text{Σήμα}(E\zeta(1, X_1)) = 1 \Leftrightarrow \text{Σήμα}(E\sigma(1, X_1)) \neq \text{Σήμα}(E\sigma(2, X_1)).$$

Αν οι εισοδοί είναι γνωστό ότι είναι, ας πούμε, 1 και 0, τότε η τελευταία ανάγεται σε:

$$\text{Σήμα}(E\zeta(1, X_1)) = 1 \Leftrightarrow 1 \neq 0.$$

Τώρα το πρόβλημα είναι φανερό: το σύστημα είναι ανίκανο να συμπεράνει ότι $\text{Σήμα}(E\zeta(1, X_1)) = 1$, γι' αυτό χρειάζεται να του πούμε ότι $1 \neq 0$.

¹² Οι παραλήψεις αυτού του είδους είναι πολύ συνηθισμένες επειδή οι άνθρωποι γενικά θεωρούν ότι τα διαφορετικά ονόματα αναφέρονται σε διαφορετικά πράγματα. Τα συστήματα λογικού προγραμματισμού που περιγράφονται στο Κεφάλαιο 9 κάνουν επίσης αυτή την παραδοχή.

8.5 ΣΥΝΟΨΗ

Σε αυτό το κεφάλαιο παρουσιάσαμε τη **λογική πρώτης τάξης** (first-order logic), μια γλώσσα αναπαράστασης πολύ ισχυρότερη από την προτασιακή λογική. Τα σημαντικά σημεία είναι τα παρακάτω:

- Οι γλώσσες αναπαράστασης γνώσης πρέπει να είναι δηλωτικές, συνθετικές, εκφραστικές, ανεξάρτητες από τα συμφοραζόμενα και χωρίς αμφισημίες.
- Οι λογικές διαφέρουν στις **οντολογικές δεσμεύσεις** και στις **επιστημολογικές δεσμεύσεις** τους. Ενώ η προτασιακή λογική δεσμεύεται μόνο για την ύπαρξη γεγονότων, η λογική πρώτης τάξης δεσμεύεται για την ύπαρξη αντικειμένων και σχέσεων και έτσι υπερέχει σε εκφραστική ισχύ.
- Ένας **δυνατός κόσμος** ή **μοντέλο** στη λογική πρώτης τάξης ορίζεται από ένα σύνολο αντικειμένων, από τις σχέσεις μεταξύ τους και από τις συναρτήσεις που μπορούν να εφαρμοζονται σε αυτά.
- Τα **σύμβολα σταθερών** ονομάζουν αντικείμενα, τα **σύμβολα κατηγορημάτων** ονομάζουν σχέσεις και τα **σύμβολα συναρτήσεων** ονομάζουν συναρτήσεις. Μια **ερμηνεία** καθορίζει μια απεικόνιση από τα σύμβολα στο μοντέλο. Οι **σύνθετοι όροι** εφαρμόζουν σύμβολα συναρτήσεων σε όρους για να ονομάσουν ένα αντικείμενο. Με δεδομένη μια ερμηνεία και ένα μοντέλο, προσδιορίζεται η αλήθεια μιας πρότασης.
- Μια **ατομική πρόταση** αποτελείται από ένα κατηγορημα που εφαρμόζεται σε έναν ή περισσότερους όρους· είναι αληθής μόνο όταν η σχέση που κατονομάζεται από το κατηγορημα ισχύει μεταξύ των αντικειμένων που κατονομάζονται από τους όρους. Οι **σύνθετες προτάσεις** χρησιμοποιούν συνδετικά όπως ακριβώς και στην προτασιακή λογική, και οι **ποσοτικοποιημένες προτάσεις** επιτρέπουν να εκφράζονται γενικοί κανόνες.
- Η ανάπτυξη μιας βάσης γνώσης στη λογική πρώτης τάξης απαιτεί μια προσεκτική διαδικασία ανάλυσης του πεδίου, επιλογής λεξιλογίου και κωδικοποίησης των αξιωμάτων που απαιτούνται για την υποστήριξη των επιθυμητών συμπερασμών.

ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΚΑΙ ΙΣΤΟΡΙΚΕΣ ΣΗΜΕΙΩΣΕΙΣ

Αν και ακόμα και η λογική του Αριστοτέλη καταπιάνεται με γενικεύσεις για τα αντικείμενα, η γνήσια λογική πρώτης τάξης χρονολογείται από την εισαγωγή των ποσοδεικτών στο *Begriffsschrift* (“Γραφή εννοιών” ή “Εννοιολογική σημειογραφία”) του Gottlob Frege (1879). Η ικανότητα του Frege να ενθέτει τους ποσοδείκτες ήταν ένα μεγάλο βήμα προς τα εμπρός, όμως χρησιμοποίησε μια δύσχρηστη σημειογραφία. (Ένα παράδειγμα φαίνεται στο μπροστινό εξώφυλλο αυτού του βιβλίου.) Η σημερινή σημειογραφία της λογικής πρώτης τάξης οφείλεται σε σημαντικό βαθμό στον Giuseppe Peano (1889), αλλά η σημασιολογία είναι ουσιαστικά ταυτόσημη με του Frege. Είναι μάλλον αξιοπερίεργο ότι τα αξιώματα του Peano οφείλονται σε μεγάλο βαθμό στους Grassmann (1861) και Dedekind (1888).

Ένα σημαντικό εμπόδιο στην ανάπτυξη της λογικής πρώτης τάξης ήταν η εστίαση της προσοχής στα μονοθεσικά (one-place) κατηγορήματα σε βάρος των πολυθεσικών (many-place) σχεσιακών κατηγορημάτων. Αυτή η εμμονή στα μονοθεσικά κατηγορήματα ήταν σχεδόν καθολική στα λογικά συστήματα από τον Αριστοτέλη μέχρι και τον Boole. Η πρώτη συστηματική πραγματεία των σχέσεων παρουσιάστηκε από τον Augustus De Morgan (1864), ο οποίος επικαλέστηκε το εξής παράδειγμα για να δείξει τι είδους συμπερασμούς δεν μπορούσε να χειριστεί η λογική του Αριστοτέλη: “Όλα τα άλογα είναι ζώα· επομένως το κεφάλι ενός αλόγου είναι κεφάλι ζώου.” Ο συμπερασμός αυτός είναι απρόσιτος για τον Αριστοτέλη, επειδή οποιοσδήποτε έγκυρος κανόνας που μπορεί να υποστηρίξει αυτόν το συμπερασμό θα πρέπει πρώτα να αναλύσει την

πρόταση χρησιμοποιώντας το διθεσικό κατηγορήμα “το x είναι η κεφαλή του y ”. Η λογική των σχέσεων μελετήθηκε σε βάθος από τον Charles Sanders Peirce (1870), ο οποίος επίσης ανέπτυξε τη λογική πρώτης τάξης ανεξάρτητα από τον Frege, αν και κάπως αργότερα (Peirce, 1883).

Ο Leopold Löwenheim (1915) παρουσίασε μια συστηματική μελέτη της θεωρίας των μοντέλων για τη λογική πρώτης τάξης το 1915. Η εργασία αυτή αντιμετώπισε επίσης το σύμβολο της ισότητας ως αναπόσπαστο μέρος της λογικής. Τα αποτελέσματα του Löwenheim επεκτάθηκαν περισσότερο από τον Thoralf Skolem (1920). Ο Alfred Tarski (1935, 1956) παρουσίασε ένα ρητό ορισμό της αλήθειας και της ικανοποίησης κατά τη θεωρία των μοντέλων στη λογική πρώτης τάξης, χρησιμοποιώντας τη θεωρία των συνόλων.

Ο McCarthy (1958) ήταν ο κυρίως υπεύθυνος για την εισαγωγή της λογικής πρώτης τάξης ως εργαλείο για την οικοδόμηση συστημάτων ΤΝ. Οι προοπτικές για τη βασισμένη στη λογική ΤΝ ενισχύθηκαν σημαντικά από την ανάπτυξη της ανάλυσης (resolution) από τον Robinson (1965), μιας πλήρους διαδικασίας για το συμπερασμό πρώτης τάξης που περιγράφεται στο Κεφάλαιο 9. Η λογικιστική προσέγγιση έπιασε ρίζες στο Stanford. Ο Cordell Green (1969a, 1969b) ανέπτυξε ένα σύστημα συλλογιστικής πρώτης τάξης, το QA3, το οποίο οδήγησε στις πρώτες απόπειρες να κατασκευαστεί ένα λογικό ρομπότ στο SRI (Fikes and Nilsson, 1971). Η λογική πρώτης τάξης εφαρμόστηκε από τους Zohar Manna και Richard Waldinger (1971) στη συλλογιστική για προγράμματα και αργότερα από τον Michael Genesereth (1984) στη συλλογιστική για κυκλώματα. Στην Ευρώπη, αναπτύχθηκε ο λογικός προγραμματισμός (μια περιορισμένη μορφή συλλογιστικής πρώτης τάξης) για τη γλωσσική ανάλυση (Colmerauer *et al.*, 1973) και για τα γενικά δηλωτικά συστήματα (Kowalski, 1974). Η υπολογιστική λογική επίσης παγιώθηκε καλά στο Εδιμβούργο, μέσω του έργου LCF (Logic for Computable Functions — Gordon *et al.*, 1979). Οι εξελίξεις αυτές χρονογραφούνται περισσότερο στα Κεφάλαια 9 και 10.

Υπάρχουν κάμποσα καλά σύγχρονα εισαγωγικά εγχειρίδια για τη λογική πρώτης τάξης. Το βιβλίο του Quine (1982) είναι από τα πιο ευκολοδιάβαστα. Το βιβλίο του Enderton (1972) δίνει μια πιο μαθηματική προοπτική. Μια πολύ τυπική διαπραγμάτευση της λογικής πρώτης τάξης, μαζί με πολλά πιο προχωρημένα θέματα της λογικής, παρέχεται από τους Bell και Machover (1977). Οι Manna και Waldinger (1985) κάνουν μια ευκολοδιάβαστη εισαγωγή στη λογική, από την άποψη της επιστήμης των υπολογιστών. Το βιβλίο του Gallier (1986) παρέχει μια εξαιρετικά αυστηρή μαθηματική παρουσίαση της λογικής πρώτης τάξης, συνοδευόμενη από πάρα πολλή ύλη για τη χρήση της στην αυτοματοποιημένη συλλογιστική. Το *Logical Foundations of Artificial Intelligence* (Genesereth and Nilsson, 1987) παρέχει τόσο μια σοβαρή εισαγωγή στη λογική όσο και την πρώτη συστηματική μελέτη των λογικών πρακτόρων με αντιλήψεις και ενέργειες.

ΑΣΚΗΣΕΙΣ

8.1 Μια λογική βάση γνώσης αναπαριστά τον κόσμο χρησιμοποιώντας ένα σύνολο προτάσεων χωρίς κάποια ρητή δομή. Μια **αναλογική** (analogical) αναπαράσταση, από την άλλη, έχει φυσική δομή που αντιστοιχεί άμεσα στη δομή του πράγματος που αναπαρίσταται. Θεωρήστε έναν οδικό χάρτη της χώρας σας ως αναλογική αναπαράσταση γεγονότων για τη χώρα. Η διδιάστατη δομή του χάρτη αντιστοιχεί στη διδιάστατη επιφάνεια της περιοχής.

- α.** Δώστε πέντε παραδείγματα *συμβόλων* στη γλώσσα του χάρτη.
- β.** *Ρητή* (explicit) πρόταση είναι μια πρόταση που ο δημιουργός της αναπαράστασης την καταγράφει πραγματικά. *Υπονοούμενη* (implicit) πρόταση είναι μια πρόταση που προκύπτει από ρητές προτάσεις λόγω ιδιοτήτων της αναλογικής αναπαράστασης. Δώστε τρία παραδείγματα *ρητών* και τρία παραδείγματα *υπονοούμενων* προτάσεων στη γλώσσα του χάρτη.

- γ. Δώστε τρία παραδείγματα γεγονότων για τη φυσική δομή της χώρας σας, τα οποία δεν μπορούν να αναπαρασταθούν στη γλώσσα του χάρτη.
- δ. Δώστε δύο παραδείγματα γεγονότων που είναι πολύ ευκολότερο να εκφραστούν στη γλώσσα του χάρτη από ό,τι στη λογική πρώτης τάξης.
- ε. Δώστε δύο άλλα παραδείγματα χρήσιμων αναλογικών αναπαραστάσεων. Ποια είναι τα πλεονεκτήματα και τα μειονεκτήματα κάθε μίας από αυτές τις γλώσσες;

8.2 Έστω μια βάση γνώσης που περιέχει μόνο δύο προτάσεις: τις $P(a)$ και $P(b)$. Καλύπτει λογικά αυτή η βάση γνώσης την $\forall x P(x)$; Εξηγήστε την απάντησή σας με όρους μοντέλων.

8.3 Είναι έγκυρη η πρόταση $\exists x, y x = y$; Εξηγήστε.

8.4 Γράψτε μια λογική πρόταση τέτοια ώστε κάθε κόσμος στον οποίο είναι αληθής να περιέχει ακριβώς ένα αντικείμενο.

8.5 Θεωρήστε ένα λεξιλόγιο συμβόλων που περιέχει c σύμβολα σταθερών, p_k σύμβολα κατηγορημάτων κάθε τάξης (arity) k , και f_k σύμβολα συναρτήσεων κάθε τάξης k , όπου $1 \leq k \leq A$. Έστω ότι το πεδίο έχει σταθερό μέγεθος D . Για οποιονδήποτε δεδομένο συνδυασμό ερμηνείας–μοντέλου, κάθε σύμβολο κατηγορήματος ή συνάρτησης αντιστοιχίζεται σε μια σχέση ή συνάρτηση, αντίστοιχα, της ίδιας τάξης. Μπορείτε να θεωρήσετε ότι οι συναρτήσεις του μοντέλου επιτρέπουν κάποιες πλειάδες (tuples) εισόδου να μην έχουν τιμή για τη συνάρτηση (δηλαδή η τιμή να είναι το αόρατο αντικείμενο). Βρείτε ένα μαθηματικό τύπο για τον αριθμό των δυνατών συνδυασμών ερμηνειών–μοντέλων για ένα πεδίο με D στοιχεία. Μη σας απασχολεί η απαλοιφή των πλεοναζόντων συνδυασμών.

8.6 Αναπαραστήστε τις παρακάτω προτάσεις στη λογική πρώτης τάξης, χρησιμοποιώντας ένα συνεπές λεξιλόγιο (το οποίο πρέπει να ορίσετε):

- α. Μερικοί φοιτητές εγγράφηκαν στο μάθημα των Γαλλικών την άνοιξη του 2001.
- β. Κάθε φοιτητής που εγγράφεται στο μάθημα των Γαλλικών το περνά.
- γ. Μόνο ένας φοιτητής εγγράφηκε στο μάθημα των Ελληνικών την άνοιξη του 2001.
- δ. Ο καλύτερος βαθμός στα Ελληνικά είναι πάντα μεγαλύτερος από τον καλύτερο βαθμό στα Γαλλικά.
- ε. Κάθε άνθρωπος που αγοράζει ασφαλιστικό συμβόλαιο είναι έξυπνος.
- στ. Κανένας άνθρωπος δεν αγοράζει ακριβό ασφαλιστικό συμβόλαιο.
- ζ. Υπάρχει ένας πράκτορας που πουλά ασφαλιστικά συμβόλαια μόνο σε άτομα που δεν είναι ασφαλισμένα.
- η. Υπάρχει ένας κουρέας που ξυρίζει όλους τους άνδρες της πόλης που δεν ξυρίζονται μόνοι τους.
- θ. Ένας άνθρωπος γεννημένος στο Ηνωμένο Βασίλειο που κάθε ένας από τους γονείς του είναι είτε πολίτης του Η.Β. είτε κάτοικος του Η.Β. είναι πολίτης του Η.Β. από γέννηση.
- ι. Ένας άνθρωπος γεννημένος εκτός του Η.Β. που ένας από τους γονείς του είναι πολίτης του Η.Β. από γέννηση είναι πολίτης του Η.Β. από καταγωγή.
- ια. Οι πολιτικοί μπορούν εξαπατούν μερικούς ανθρώπους πάντοτε, και μπορούν να εξαπατούν όλους τους ανθρώπους μερικές φορές, αλλά δεν μπορούν να εξαπατούν όλους τους ανθρώπους πάντοτε.

8.7 Αναπαραστήστε την πρόταση “Όλοι οι Γερμανοί μιλούν τις ίδιες γλώσσες” στον κατηγορηματικό λογισμό. Χρησιμοποιήστε το κατηγορήμα $Milá(x, l)$, που σημαίνει ότι ο άνθρωπος x μιλά τη γλώσσα l .

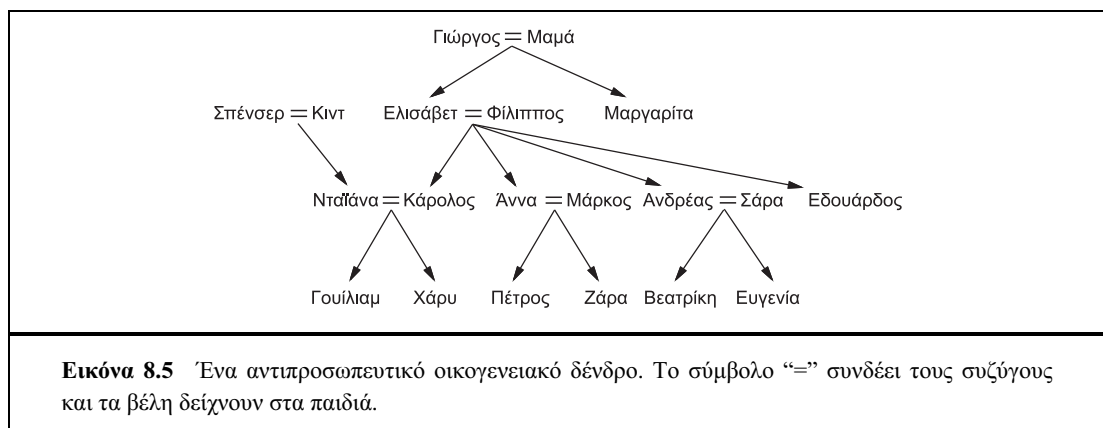
8.8 Ποιο αξίωμα χρειάζεται για να εξαχθεί ως συμπέρασμα το γεγονός *Θηλυκό(Λάουρα)* με δεδομένα τα γεγονότα *Αρσενικό(Δημήτρης)* και *Σύζυγος(Δημήτρης, Λάουρα)*;

8.9 Γράψτε ένα γενικό σύνολο γεγονότων και αξιωμάτων που να αναπαριστούν τον ισχυρισμό “Ο Ουέλινγκτον άκουσε για το θάνατο του Ναπολέοντα” και να απαντούν σωστά στην ερώτηση “Άκουσε ο Ναπολέων για το θάνατο του Ουέλινγκτον;”

8.10 Επαναδιατυπώστε τα προτασιακά γεγονότα του κόσμου του wumpus από την Ενότητα 7.5 στη λογική πρώτης τάξης. Πόσο πιο συμπαγής είναι αυτή η εκδοχή;

8.11 Γράψτε αξιώματα που να περιγράφουν τα κατηγορήματα *Δισέγγονο*, *Προ-Προγονιός*, *Αδελφός*, *Αδελφή*, *Κόρη*, *Γιος*, *Θεία*, *Θεός*, *Κουνιάδος*, *Κουνιάδα*, και *ΠρώτοΞαδέλφι*. Βρείτε τον κατάλληλο ορισμό για το *m*-οστό ξαδέλφι με διαφορά *n* γενιές,¹³ και γράψτε αυτόν τον ορισμό στη λογική πρώτης τάξης.

Τώρα καταγράψτε τα βασικά γεγονότα που απεικονίζονται στο οικογενειακό δένδρο της Εικόνας 8.5. Χρησιμοποιώντας ένα κατάλληλο σύστημα λογικής συλλογιστικής, ενημερώστε το (TELL) με όλες προτάσεις που έχετε καταγράψει και ρωτήστε το (ASK) ποια είναι τα εγγόνια της Ελισάβετ, οι κουνιάδοι της Νταϊάνα και οι προ-προγονιοί της Ζάρα.



8.12 Γράψτε μια πρόταση που να ισχυρίζεται ότι το + είναι αντιμεταθετική συνάρτηση. Προκύπτει η πρότασή σας από τα αξιώματα του Peano; Αν ναι, εξηγήστε γιατί· αν όχι, βρείτε ένα μοντέλο στο οποίο τα αξιώματα να είναι αληθή και η πρότασή σας να είναι ψευδής.

8.13 Εξηγήστε τι πρόβλημα έχει ο παρακάτω προτεινόμενος ορισμός του κατηγορήματος μέλους συνόλου $\in$:

$$\begin{aligned} \forall x, s \quad x \in \{x \mid s\} \\ \forall x, s \quad x \in s \Rightarrow \forall y \quad x \in \{y \mid s\} . \end{aligned}$$

8.14 Χρησιμοποιώντας ως παράδειγμα τα αξιώματα των συνόλων, γράψτε αξιώματα για το πεδίο των λιστών, συμπεριλαμβάνοντας όλες τις σταθερές, τις συναρτήσεις και τα κατηγορήματα που αναφέρθηκαν σε αυτό το κεφάλαιο.

8.15 Εξηγήστε τι πρόβλημα έχει ο παρακάτω προτεινόμενος ορισμός των γειτονικών τετραγώνων στον κόσμο του wumpus:

$$\forall x, y \quad \text{Γειτονικό}([x, y], [x+1, y]) \wedge \text{Γειτονικό}([x, y], [x, y+1]) .$$

¹³ Σ.τ.Μ.: Τα πρώτα ξαδέλφια έχουν κανονικά τους ίδιους παππούδες, τα δεύτερα ξαδέλφια έχουν τους ίδιους προ-παππούδες κ.ο.κ. Όσοι ασχολούνται με τα γενεαλογικά δένδρα χρησιμοποιούν όρους όπως “πρώτος ξάδελφος με διαφορά 1 γενιά” (1 time removed) για τον αδελφό ενός γονέα και “πρώτος ξάδελφος με διαφορά 2 γενιές” (2 times removed) για τον αδελφό ενός από τους παππούδες.



Τεχνητή Νοημοσύνη

Μια σύγχρονη προσέγγιση

Stuart Russell • Peter Norvig

ΔΕΥΤΕΡΗ ΑΜΕΡΙΚΑΝΙΚΗ ΕΚΔΟΣΗ

Η πρώτη αμερικανική έκδοση του βιβλίου *Τεχνητή Νοημοσύνη: Μια σύγχρονη προσέγγιση* έχει γίνει κλασική για τη βιβλιογραφία της ΤΝ. Το βιβλίο από το έχουν υιοθετήσει περισσότερα από 800 πανεπιστήμια σε 89 χώρες, και το έχουν υμνήσει ως την απόλυτη σύνθεση του πεδίου αυτού. Ας δούμε τι είχαν σχετικά διάφοροι άνθρωποι:

"Η έκδοση αυτού του εγχειριδίου ήταν ένα μεγάλο βήμα προς τα εμπρός, όχι μόνο για τη διδασκαλία της ΤΝ, αλλά και για την ενσωματωμένη θεωρία του πεδίου που εισάγει αυτό το βιβλίο. Ακόμα και για ειδικούς του πεδίου υπάρχουν σημαντικές διορθωτικές παρατηρήσεις σε όλα σχεδόν τα κεφάλαια." — Καθ. Thomas Dietterich (Oregon State)

"Απίστευτα καταπληκτικό. Το βιβλίο που πάντοτε περιμένα ... η βίβλος της ΤΝ για την επόμενη δεκαετία." — Καθ. Gerd Brewka (Βιέννη)

"Ένα θαυμάσιο επίτευγμα, ένα πραγματικά υπέροχο βιβλίο!" — Καθ. Selmer Brångsjord (RPI)

"Ένα σπουδαίο βιβλίο, με απίστευτο εύρος και βάθος και πολύ κατανοητό. Όλοι οι γνωστοί μου που το έχουν χρησιμοποιήσει στις τάξεις τους το έχουν λατρέψει." — Καθ. Haym Hirsh (Rutgers)

"Έχω εντυπωσιαστεί βαθιά από την άμεση προηγουμένη ποιότητα στην παρουσίαση μιας συνεκτικής, ισορροπημένης, ειρικής και βαθιάς και απλοποιητικής επίθεσης του πεδίου της ΤΝ. Θα γίνει το κλασικό σύγγραμμα για τα χρόνια που θα ακολουθήσουν." — Καθ. Wolfgang Bibel (Darmstadt)

"Καταπληκτικό! Κατανοημένο και καλά οργανωμένο, με περιεκτική κάλυψη της ελκς που θα πρέπει να γνωρίζουν όλοι οι σπουδαστές της ΤΝ." — Καθ. Martha Pollack (Michigan)

"Έξοχο ... Οι περιγραφές του είναι εξαιρετικά βαθύρες και εννοητικές. Η οργάνωσή του είναι θαυμάσια. Τα παραδείγματα του είναι παρατηρητικά. Και η κάλυψη των θεμάτων είναι συστηματική και πλήρης! ... Θα χρησιμοποιήσω επάξια στο πεδίο αυτό για οριστό καιρό." — Καθ. Nik Nilsson (Stanford)

"Το καλύτερο βιβλίο που υπάρχει διαθέσιμο αυτή τη στιγμή... Είναι σχεδόν εξίσου καλό με το βιβλίο που γράψαμε ο Chatteak και εγώ, αλλά είναι πιο ενημερωμένο. (Εντάξει, θα το παραδεχτώ, μπορεί ακόμα και να είναι καλύτερο από το βιβλίο μας.)" — Καθ. Drew McDermott (Yale)

"Μια επιβλητική ειρική θεωρία ολόκληρου του πεδίου της Τεχνητή Νοημοσύνης, η οποία θα διαφωτίσει τόσο τους καθηγητές όσο και τους σπουδαστές" — Δρ. Alan Kay

"Αυτό είναι το βιβλίο που με έκανε να αγαπήσω την ΤΝ." — Σπουδαστής (Ινδονησία)

Στη δεύτερη αμερικανική έκδοση έχουν εκτενώς ξαναγραφτεί όλα τα κεφάλαια. Έχει προστεθεί σημαντικό νέο υλικό που καλύπτει τομείς όπως η ικανοποίηση περιορισμών, ο γρήγορος προτασιακός συμπερασμός, τα γραφήματα σχεδιασμού, οι πράκτορες στο Internet, ο ακριβής πιθανοτικός συμπερασμός, οι τεχνικές Monte Carlo αλυσίδων Markov, τα φίλτρα Kalman, οι μέθοδοι συλλογικής μάθησης, η στατιστική μάθηση, τα πιθανοτικά μοντέλα φυσικής γλώσσας, η πιθανοτική ρομποτική, και οι ηθικές πτυχές της ΤΝ.

Το βιβλίο υποστηρίζεται από ένα σύνολο ηλεκτρονικά διαθέσιμων πόρων, στους οποίους περιλαμβάνεται πηγής κώδικας, εικόνες, διαφάνειες διαλέξεων, ένας κατάλογος με περισσότερους από 800 συνδέσμους για την "ΤΝ στον Ιστό", και μια ηλεκτρονική ομάδα συζητήσεων. Όλα αυτά είναι διαθέσιμα στη διεύθυνση aima.cs.berkeley.edu

Αντίστοιχες πληροφορίες και πρόσθετο υλικό για την ελληνική έκδοση του βιβλίου, θα βρείτε στη διεύθυνση aima.uom.gr