

# Περιεχόμενα

<b>1</b>	<b>Εξόρυξη Δεδομένων</b>	<b>1</b>
1.1	Τί είναι Εξόρυξη Δεδομένων; . . . . .	1
1.1.1	Στατιστική Μοντελοποίηση . . . . .	2
1.1.2	Μηχανική Μάθηση . . . . .	2
1.1.3	Υπολογιστικές Προσεγγίσεις Μοντελοποίησης . . . . .	3
1.1.4	Σύνοψη . . . . .	4
1.1.5	Εξαγωγή Γνωρισμάτων . . . . .	4
1.2	Στατιστικά Όρια της Εξόρυξης Δεδομένων . . . . .	6
1.2.1	Συνολική Επίγνωση της Πληροφορίας . . . . .	6
1.2.2	Αρχή Bonferroni . . . . .	7
1.2.3	Παράδειγμα της Αρχής Bonferroni . . . . .	7
1.2.4	Ασκήσεις Ενότητας 1.2 . . . . .	9
1.3	Πράγματα χρήσιμα να γνωρίζουμε . . . . .	9
1.3.1	Σπουδαιότητα των Λέξεων σε ένα Έγγραφο . . . . .	10
1.3.2	Συναρτήσεις Κατακερματισμού . . . . .	11
1.3.3	Κατάλογοι . . . . .	13
1.3.4	Δευτερεύουσα Αποθήκευση . . . . .	15
1.3.5	Βάση των Φυσικών Λογαρίθμων . . . . .	16
1.3.6	Νόμος των Δυνάμεων . . . . .	17
1.3.7	Ασκήσεις Ενότητας 1.3 . . . . .	19
1.4	Περίγραμμα Βιβλίου . . . . .	20
1.5	Περίληψη Κεφαλαίου 1 . . . . .	22
1.6	Αναφορές Κεφαλαίου 1 . . . . .	23
<b>2</b>	<b>Συστήματα Αρχείων Μεγάλης Κλίμακας και Map-Reduce</b>	<b>25</b>
2.1	Κατανεμημένα Συστήματα Αρχείων . . . . .	26
2.1.1	Φυσική Οργάνωση Υπολογιστικών Κόμβων . . . . .	26
2.1.2	Οργάνωση Συστήματος Αρχείων Ευρείας Κλίμακας . . . . .	28
2.2	Map-Reduce . . . . .	29
2.2.1	Εργασίες Map . . . . .	30
2.2.2	Ομαδοποίηση και Συνάθροιση . . . . .	31
2.2.3	Εργασίες Reduce . . . . .	32
2.2.4	Συνδυαστές . . . . .	32
2.2.5	Λεπτομέρειες Εκτέλεσης του Map-Reduce . . . . .	33
2.2.6	Αντιμετώπιση Αποτυχιών Κόμβων . . . . .	35
2.3	Αλγόριθμοι που χρησιμοποιούν Map-Reduce . . . . .	35
2.3.1	Πολλαπλασιασμός Πίνακα με Διάνυσμα σε Map-Reduce . . . . .	36

2.3.2	Το Διάνυσμα δεν χωρά στην Κύρια Μνήμη . . . . .	37
2.3.3	Πράξεις Σχεσιακής Άλγεβρας . . . . .	38
2.3.4	Υπολογισμός Επιλογών σε Map-Reduce . . . . .	42
2.3.5	Υπολογισμός Προβολών σε Map-Reduce . . . . .	42
2.3.6	Ένωση, Τομή και Διαφορά σε Map-Reduce . . . . .	43
2.3.7	Υπολογισμός Φυσικής Σύνδεσης σε Map-Reduce . . . . .	44
2.3.8	Γενίκευση Αλγορίθμου Σύνδεσης . . . . .	45
2.3.9	Ομαδοποίηση και Συνάθροιση σε Map-Reduce . . . . .	45
2.3.10	Πολλαπλασιασμός Πινάκων . . . . .	46
2.3.11	Πολλαπλασιασμός Πινάκων σε ένα Βήμα . . . . .	47
2.3.12	Ασκήσεις Ενότητας 2.3 . . . . .	48
2.4	Επεκτάσεις του Map-Reduce . . . . .	49
2.4.1	Συστήματα Ροών Εργασίας . . . . .	50
2.4.2	Αναδρομικές Επεκτάσεις στο Map-Reduce . . . . .	51
2.4.3	Σύστημα Pregel . . . . .	55
2.4.4	Ασκήσεις Ενότητας 2.4 . . . . .	56
2.5	Απόδοση Αλγορίθμων Υπολογισμού Συστοιχίας . . . . .	57
2.5.1	Επικοινωνιακό Κόστος σε Υπολογισμό Συστοιχίας . . . . .	57
2.5.2	Χρόνος Επικοινωνιακού Κόστους . . . . .	60
2.5.3	Συνδέσεις Πολλαπλών Δρόμων . . . . .	60
2.5.4	Ασκήσεις Ενότητας 2.5 . . . . .	65
2.6	Περίληψη Κεφαλαίου 2 . . . . .	66
2.7	Αναφορές Κεφαλαίου 2 . . . . .	69
<b>3</b>	<b>Αναζήτηση Όμοιων Αντικειμένων</b>	<b>73</b>
3.1	Εφαρμογές Αναζήτησης Κοντινότερου Γείτονα . . . . .	74
3.1.1	Μετρική Jaccard για την Ομοιότητα Συνόλων . . . . .	74
3.1.2	Ομοιότητα Εγγράφων . . . . .	74
3.1.3	Η Συνεργατική Διήθηση ως Πρόβλημα Ομοιότητας . . . . .	76
3.1.4	Ασκήσεις Ενότητας 3.1 . . . . .	78
3.2	Θρυματοποίηση Εγγράφων . . . . .	78
3.2.1	k-θύμμα . . . . .	79
3.2.2	Επιλογή Μεγέθους Θρύματος . . . . .	79
3.2.3	Κατακερματισμός Θρυμάτων . . . . .	80
3.2.4	Κατασκευή Θρυμάτων από Λέξεις . . . . .	81
3.2.5	Ασκήσεις Ενότητας 3.2 . . . . .	82
3.3	Συνόψεις που διατηρούν την Ομοιότητα . . . . .	82
3.3.1	Αναπαράσταση Πινάκων για Σύνολα . . . . .	83
3.3.2	Minhashing . . . . .	84
3.3.3	Minhashing και Ομοιότητα Jaccard . . . . .	85
3.3.4	Υπογραφές Minhash . . . . .	86
3.3.5	Υπολογισμός Υπογραφών Minhash . . . . .	86
3.3.6	Ασκήσεις Ενότητας 3.3 . . . . .	90
3.4	Κατακερματισμός Διατήρησης Τοπικότητας για Έγγραφα . . . . .	91
3.4.1	LSH για Υπογραφές Minhash . . . . .	92
3.4.2	Ανάλυση της Τεχνικής Λωριδοποίησης . . . . .	94
3.4.3	Συνδυασμός των Τεχνικών . . . . .	96
3.4.4	Ασκήσεις Ενότητας 3.4 . . . . .	97
3.5	Μετρικές Απόστασης . . . . .	98
3.5.1	Ορισμός Μετρικής Απόστασης . . . . .	98

3.5.2	Ευκλείδεια Απόσταση . . . . .	99
3.5.3	Απόσταση Jaccard . . . . .	100
3.5.4	Απόσταση Συνημιτόνου . . . . .	101
3.5.5	Απόσταση Διόρθωσης . . . . .	102
3.5.6	Απόσταση Hamming . . . . .	104
3.5.7	Ασκήσεις Ενότητας 3.5 . . . . .	104
3.6	Θεωρία Συναρτήσεων Διατήρησης Τοπικότητας . . . . .	106
3.6.1	Συναρτήσεις Διατήρησης της Τοπικότητας . . . . .	107
3.6.2	Οικογένειες Συναρτήσεων Διατήρησης Τοπικότητας για Απόσταση Jaccard . . . . .	109
3.6.3	Ενίσχυση Πιθανότητας μίας Οικογένειας Συναρτήσεων Διατήρησης Τοπικότητας . . . . .	109
3.6.4	Ασκήσεις Ενότητας 3.6 . . . . .	112
3.7	Οικογένειες Συναρτήσεων LSH για Άλλες Μετρικές Απόστασης . . . . .	113
3.7.1	Συναρτήσεις LSH για Απόσταση Hamming . . . . .	113
3.7.2	Τυχαία Υπερεπίπεδα και Απόσταση Συνημιτόνου . . . . .	114
3.7.3	Σκίτσα . . . . .	116
3.7.4	Συναρτήσεις LSH για Ευκλείδεια Απόσταση . . . . .	117
3.7.5	Άλλες Οικογένειες Συναρτήσεων LSH για Ευκλείδεια Απόσταση . . . . .	118
3.7.6	Ασκήσεις Ενότητας 3.7 . . . . .	119
3.8	Εφαρμογές Κατακερματισμού με Διατήρηση Τοπικότητας . . . . .	121
3.8.1	Καθορισμός Οντότητας . . . . .	121
3.8.2	Παράδειγμα Καθορισμού Οντότητας . . . . .	122
3.8.3	Επαλήθευση Ταυριάσματος Εγγραφών . . . . .	123
3.8.4	Ταίριασμα Δακτυλικών Αποτυπωμάτων . . . . .	125
3.8.5	Μία Οικογένεια LSH για Ταίριασμα Δακτυλικών Αποτυπωμάτων . . . . .	126
3.8.6	Όμοια Άρθρα Εφημερίδων . . . . .	128
3.8.7	Ασκήσεις Ενότητας 3.8 . . . . .	130
3.9	Μέθοδοι για Υψηλό Βαθμό Ομοιότητας . . . . .	131
3.9.1	Εύρεση Ίδιων Αντικειμένων . . . . .	131
3.9.2	Αναπαράσταση Συνόλων με Αλφαριθμητικά . . . . .	132
3.9.3	Φιλτράρισμα με Βάση το Μήκος . . . . .	133
3.9.4	Δεικτοδότηση Προθέματος . . . . .	134
3.9.5	Χρήση της Πληροφορίας Θέσης . . . . .	136
3.9.6	Χρήση Θέσης και Μήκους για Δεικτοδότηση . . . . .	137
3.9.7	Ασκήσεις Ενότητας 3.9 . . . . .	140
3.10	Περίληψη Κεφαλαίου 3 . . . . .	142
3.11	Αναφορές Κεφαλαίου 3 . . . . .	145
<b>4</b>	<b>Εξόρυξη από Ροές Δεδομένων</b> . . . . .	<b>149</b>
4.1	Μοντέλο Δεδομένων Ροής . . . . .	150
4.1.1	Συστήματα Διαχείρισης Ροών Δεδομένων . . . . .	150
4.1.2	Πηγές Ροών Δεδομένων . . . . .	151
4.1.3	Ερωτήματα Ροών . . . . .	152
4.1.4	Βασικά Θέματα Επεξεργασίας Ροών . . . . .	154
4.2	Δειγματοληψία σε Ροή Δεδομένων . . . . .	155
4.2.1	Ενδεικτικό Παράδειγμα . . . . .	155
4.2.2	Θεώρηση ενός Αντιπροσωπευτικού Δείγματος . . . . .	157
4.2.3	Γενικό Πρόβλημα της Δειγματοληψίας . . . . .	157
4.2.4	Μεταβολή του Μεγέθους Δείγματος . . . . .	158

4.2.5	Ασκήσεις Ενότητας 4.2 . . . . .	159
4.3	Φιλτράρισμα Ροής . . . . .	159
4.3.1	Ενδεικτικό Παράδειγμα . . . . .	159
4.3.2	Φίλτρο Bloom . . . . .	161
4.3.3	Ανάλυση Φίλτρου Bloom . . . . .	161
4.3.4	Ασκήσεις Ενότητας 4.3 . . . . .	163
4.4	Μέτρηση Διακριτών Στοιχείων Ροής . . . . .	164
4.4.1	Πρόβλημα της Μέτρησης Διακριτών Στοιχείων . . . . .	164
4.4.2	Αλγόριθμος Flajolet-Martin . . . . .	165
4.4.3	Συνδυασμός Εκτιμήσεων . . . . .	166
4.4.4	Απαιτήσεις Χώρου . . . . .	167
4.4.5	Ασκήσεις Ενότητας 4.4 . . . . .	168
4.5	Εκτίμηση Ροπών . . . . .	168
4.5.1	Ορισμός Ροπών . . . . .	168
4.5.2	Αλγόριθμος Alon-Matias-Szegedy για Ροπές Δεύτερης Τάξης . . . . .	169
4.5.3	Γιατί Δουλεύει ο Αλγόριθμος Alon-Matias-Szegedy . . . . .	171
4.5.4	Ροπές Υψηλότερης Τάξης . . . . .	172
4.5.5	Αντιμετώπιση Άπειρων Ροπών . . . . .	172
4.5.6	Ασκήσεις Ενότητας 4.5 . . . . .	174
4.6	Μέτρηση Άσων σε Παράθυρο . . . . .	175
4.6.1	Κόστος Επακριβούς Μέτρησης . . . . .	175
4.6.2	Αλγόριθμος Datar-Gionis-Indyk-Motwani (DGIM) . . . . .	176
4.6.3	Απαιτήσεις Χώρου Αλγορίθμου DGIM . . . . .	178
4.6.4	Απάντηση Ερωτημάτων με τον Αλγόριθμο DGIM . . . . .	178
4.6.5	Διατήρηση των Κανόνων DGIM . . . . .	180
4.6.6	Περιορισμός Σφάλματος . . . . .	181
4.6.7	Επεκτάσεις στη Μέτρηση Άσων . . . . .	182
4.6.8	Ασκήσεις Ενότητας 4.6 . . . . .	183
4.7	Φθίνοντα Παράθυρα . . . . .	183
4.7.1	Πρόβλημα των πιο Κοινών Στοιχείων . . . . .	184
4.7.2	Ορισμός Φθίνοντος Παραθύρου . . . . .	184
4.7.3	Εύρεση Δημοφιλέστερων Στοιχείων . . . . .	186
4.8	Περίληψη Κεφαλαίου 4 . . . . .	187
4.9	Αναφορές Κεφαλαίου 4 . . . . .	189
<b>5</b>	<b>Ανάλυση Συνδέσμων</b> . . . . .	<b>191</b>
5.1	Ο Αλγόριθμος PageRank . . . . .	192
5.1.1	Πρώιμες Μηχανές και Ανεπιθύμητοι Όροι . . . . .	192
5.1.2	Ορισμός του PageRank . . . . .	194
5.1.3	Δομή του Παγκόσμιου Ιστού . . . . .	198
5.1.4	Αποφυγή Αδιεξόδων . . . . .	200
5.1.5	Αραχνοπαγίδες και Φορολόγηση . . . . .	204
5.1.6	Χρήση του PageRank σε Μηχανές Αναζήτησης . . . . .	207
5.1.7	Ασκήσεις Ενότητας 5.1 . . . . .	207
5.2	Αποδοτικός Υπολογισμός του PageRank . . . . .	209
5.2.1	Αναπαράσταση Πινάκων Μετάβασης . . . . .	210
5.2.2	Επανάληψη του Pagerank με Χρήση Map-Reduce . . . . .	211
5.2.3	Χρήση Συνδυαστών για τον Υπολογισμό Διανύσματος . . . . .	212
5.2.4	Αναπαράσταση Τμημάτων του Πίνακα Μεταβάσεων . . . . .	213
5.2.5	Άλλες Τεχνικές για τις Επαναλήψεις του PageRank . . . . .	214

5.2.6	Ασκήσεις Ενότητας 5.2 . . . . .	216
5.3	PageRank με Βάση το Θέμα . . . . .	217
5.3.1	Κίνητρο . . . . .	217
5.3.2	Μεροληπτικοί Τυχαίοι Περίπατοι . . . . .	218
5.3.3	Χρήση της Μεθόδου . . . . .	220
5.3.4	Εξαγωγή Θεμάτων από Όρους . . . . .	221
5.3.5	Ασκήσεις Ενότητας 5.3 . . . . .	222
5.4	Ανεπιθύμητοι Σύνδεσμοι . . . . .	222
5.4.1	Φάρμα Ανεπιθύμητων Ιστοσελίδων . . . . .	223
5.4.2	Ανάλυση της Φάρμας Ανεπιθύμητων Σελίδων . . . . .	224
5.4.3	Καταπολέμηση της Επίδρασης Ανεπιθύμητων Συνδέσμων . . . . .	225
5.4.4	TrustRank . . . . .	226
5.4.5	Ανεπιθύμητη Μάζα . . . . .	227
5.4.6	Ασκήσεις Ενότητας 5.4 . . . . .	228
5.5	Κέντρα και Αυθεντίες . . . . .	229
5.5.1	Διαίσθηση πίσω από τον HITS . . . . .	229
5.5.2	Υπολογισμός Βαθμολογιών Κέντρου και Αυθεντίας . . . . .	230
5.5.3	Ασκήσεις Ενότητας 5.5 . . . . .	234
5.6	Περίληψη Κεφαλαίου 5 . . . . .	234
5.7	Αναφορές Κεφαλαίου 5 . . . . .	238
<b>6</b>	<b>Συχνά Στοιχειοσύνολα</b> . . . . .	<b>241</b>
6.1	Μοντέλο Καλαθιού Αγορών . . . . .	242
6.1.1	Ορισμός Συχνών Στοιχειοσυνόλων . . . . .	242
6.1.2	Εφαρμογές Συχνών Στοιχειοσυνόλων . . . . .	244
6.1.3	Κανόνες Συσχέτισης . . . . .	247
6.1.4	Εύρεση Κανόνων Συσχέτισης με Υψηλή Εμπιστοσύνη . . . . .	248
6.1.5	Ασκήσεις Ενότητας 6.1 . . . . .	249
6.2	Καλάθια Αγορών και Αλγόριθμος A-Priori . . . . .	251
6.2.1	Αναπαράσταση Δεδομένων Καλαθιών Αγορών . . . . .	251
6.2.2	Χρήση Κύριας Μνήμης για την Καταμέτρηση Στοιχειοσυνόλων . . . . .	252
6.2.3	Μονοτονικότητα Στοιχειοσυνόλων . . . . .	255
6.2.4	Τυραννία της Καταμέτρησης Ζευγών . . . . .	256
6.2.5	Αλγόριθμος A-Priori . . . . .	256
6.2.6	A-Priori για Όλα τα Συχνά Στοιχειοσύνολα . . . . .	259
6.2.7	Ασκήσεις Ενότητας 6.2 . . . . .	261
6.3	Διαχείριση Μεγάλων Συνόλων Δεδομένων στην Κύρια Μνήμη . . . . .	262
6.3.1	Αλγόριθμος των Park, Chen και Yu . . . . .	263
6.3.2	Αλγόριθμος Multistage . . . . .	266
6.3.3	Αλγόριθμος Multihash . . . . .	268
6.3.4	Ασκήσεις Ενότητας 6.3 . . . . .	270
6.4	Αλγόριθμοι Περιορισμένων Περασμάτων . . . . .	273
6.4.1	Απλός Τυχαιοποιημένος Αλγόριθμος . . . . .	273
6.4.2	Αποφυγή Σφαλμάτων στους Αλγόριθμους Δειγματοληψίας . . . . .	275
6.4.3	Αλγόριθμος των Savasere, Omiecinski και Navathe . . . . .	276
6.4.4	Αλγόριθμος SON και Map-Reduce . . . . .	277
6.4.5	Αλγόριθμος Toivonen . . . . .	278
6.4.6	Γιατί Δουλεύει ο Αλγόριθμος Toivonen . . . . .	280
6.4.7	Ασκήσεις Ενότητας 6.4 . . . . .	280
6.5	Μέτρηση Συχνών Αντικειμένων σε μία Ροή Δεδομένων . . . . .	281

6.5.1	Μέθοδοι Δειγματοληψίας για Ροές . . . . .	282
6.5.2	Συχνά Στοιχειοσύνολα σε Φθίνοντα Παράθυρα . . . . .	283
6.5.3	Υβριδικές Μέθοδοι . . . . .	284
6.5.4	Ασκήσεις Ενότητας 6.5 . . . . .	286
6.6	Περίληψη Κεφαλαίου 6 . . . . .	286
6.7	Αναφορές Κεφαλαίου 6 . . . . .	289
<b>7</b>	<b>Ομαδοποίηση</b>	<b>291</b>
7.1	Εισαγωγή στις Τεχνικές Ομαδοποίησης . . . . .	291
7.1.1	Σημεία, Χώροι και Αποστάσεις . . . . .	292
7.1.2	Στρατηγικές Ομαδοποίησης . . . . .	293
7.1.3	Κατάρα της Διαστασιμότητας . . . . .	294
7.1.4	Ασκήσεις Ενότητας 7.1 . . . . .	296
7.2	Ιεραρχική Ομαδοποίηση . . . . .	297
7.2.1	Ιεραρχική Ομαδοποίηση σε Ευκλείδειο Χώρο . . . . .	297
7.2.2	Αποδοτικότητα της Ιεραρχικής Ομαδοποίησης . . . . .	302
7.2.3	Εναλλακτικοί Κανόνες Ελέγχου Ιεραρχικής Ομαδοποίησης . . . . .	303
7.2.4	Ιεραρχική Ομαδοποίηση σε μη Ευκλείδειους Χώρους . . . . .	305
7.2.5	Ασκήσεις Ενότητας 7.2 . . . . .	307
7.3	Αλγόριθμοι $K$ -Μέσων . . . . .	308
7.3.1	Τα βασικά των $K$ -Μέσων . . . . .	308
7.3.2	Αρχικοποίηση Ομάδων στους $K$ -Μέσους . . . . .	309
7.3.3	Επιλογή Σωστής Τιμής του $k$ . . . . .	310
7.3.4	Αλγόριθμος των Bradley, Fayyad και Reina . . . . .	312
7.3.5	Επεξεργασία Δεδομένων στον Αλγόριθμο BFR . . . . .	314
7.3.6	Ασκήσεις Ενότητας 7.3 . . . . .	317
7.4	Αλγόριθμος CURE . . . . .	318
7.4.1	Αρχικοποίηση του CURE . . . . .	319
7.4.2	Ολοκλήρωση Αλγόριθμου CURE . . . . .	321
7.4.3	Ασκήσεις Ενότητας 7.4 . . . . .	322
7.5	Ομαδοποίηση σε Μη Ευκλείδειους Χώρους . . . . .	322
7.5.1	Αναπαράσταση Ομάδων στον Αλγόριθμο GRGPF . . . . .	323
7.5.2	Αρχικοποίηση του Δένδρου Ομάδων . . . . .	324
7.5.3	Πρόσθεση Σημείων στον Αλγόριθμο GRGPF . . . . .	324
7.5.4	Διάσπαση και Συγχώνευση Ομάδων . . . . .	326
7.5.5	Ασκήσεις Ενότητας 7.5 . . . . .	328
7.6	Ομαδοποίηση για Ροές και Παράλληλισμό . . . . .	328
7.6.1	Μοντέλο Υπολογισμού Ροών . . . . .	329
7.6.2	Αλγόριθμος Ομαδοποίησης Ροών . . . . .	329
7.6.3	Αρχικοποίηση Κάδων . . . . .	330
7.6.4	Συγχώνευση Κάδων . . . . .	331
7.6.5	Απάντηση Ερωτημάτων . . . . .	334
7.6.6	Ομαδοποίηση σε Παράλληλο Περιβάλλον . . . . .	335
7.6.7	Ασκήσεις Ενότητας 7.6 . . . . .	335
7.7	Περίληψη Κεφαλαίου 7 . . . . .	336
7.8	Αναφορές Κεφαλαίου 7 . . . . .	340
<b>8</b>	<b>Διαφήμιση στο Διαδίκτυο</b>	<b>343</b>
8.1	Θέματα Διαδικτυακής Διαφήμισης . . . . .	344
8.1.1	Ευκαιρίες Διαφήμισης . . . . .	344

8.1.2	Άμεση Τοποθέτηση Διαφημίσεων . . . . .	345
8.1.3	Θέματα Εμφάνισης Διαφημίσεων . . . . .	346
8.2	Αλγόριθμοι Άμεσης Επεξεργασίας . . . . .	347
8.2.1	Αλγόριθμοι Άμεσης και Ολικής Επεξεργασίας . . . . .	348
8.2.2	Άπληστοι Αλγόριθμοι . . . . .	350
8.2.3	Ανταγωνιστικός Λόγος . . . . .	350
8.2.4	Ασκήσεις Ενότητας 8.2 . . . . .	351
8.3	Το Πρόβλημα του Ταιριάσματος . . . . .	352
8.3.1	Ταιριάσματα και Τέλεια Ταιριάσματα . . . . .	352
8.3.2	Άπληστος Αλγόριθμος Μέγιστου Ταιριάσματος . . . . .	353
8.3.3	Ανταγωνιστικός Λόγος για Άπληστο Τάιριασμα . . . . .	354
8.3.4	Ασκήσεις Ενότητας 8.3 . . . . .	355
8.4	Το Πρόβλημα Adwords . . . . .	356
8.4.1	Ιστορία της Διαφήμισης με Αναζήτηση . . . . .	356
8.4.2	Ορισμός του Προβλήματος Adwords . . . . .	357
8.4.3	Άπληστη Μέθοδος για το Πρόβλημα Adwords . . . . .	358
8.4.4	Αλγόριθμος Ισορροπίας . . . . .	359
8.4.5	Κάτω Φράγμα Ανταγωνιστικού Λόγου Αλγόριθμου Ισορροπίας . . . . .	361
8.4.6	Αλγόριθμος Ισορροπίας για Πολλούς Ανταγωνιστές . . . . .	364
8.4.7	Γενικευμένος Αλγόριθμος Ισορροπίας . . . . .	366
8.4.8	Μερικές Τελικές Παρατηρήσεις για το Πρόβλημα Adwords . . . . .	367
8.4.9	Ασκήσεις Ενότητας 8.4 . . . . .	368
8.5	Υλοποίηση του Προβλήματος Adwords . . . . .	369
8.5.1	Τάιριασμα Προσφορών και Ερωτήσεων Αναζήτησης . . . . .	369
8.5.2	Πιο Πολύπλοκα Προβλήματα Ταιριάσματος . . . . .	370
8.5.3	Αλγόριθμος Ταιριάσματος για Έγγραφα και Προσφορές . . . . .	371
8.6	Περίληψη Κεφαλαίου 8 . . . . .	374
8.7	Αναφορές Κεφαλαίου 8 . . . . .	377
<b>9</b>	<b>Συστήματα Συστάσεων</b> . . . . .	<b>379</b>
9.1	Ένα Μοντέλο Συστημάτων Συστάσεων . . . . .	380
9.1.1	Πίνακας Ωφελείας . . . . .	380
9.1.2	Η Μακριά Ουρά . . . . .	382
9.1.3	Εφαρμογές Συστημάτων Συστάσεων . . . . .	382
9.1.4	Συμπλήρωση του Πίνακα Ωφελείας . . . . .	384
9.2	Συστάσεις Βασισμένες στο Περιεχόμενο . . . . .	385
9.2.1	Προφίλ Στοιχείων . . . . .	385
9.2.2	Ανακαλύπτοντας Γνωρίσματα Εγγράφων . . . . .	386
9.2.3	Απόκτηση Γνωρισμάτων Στοιχείων από Επισημειώσεις . . . . .	388
9.2.4	Αναπαριστώντας Προφίλ Στοιχείων . . . . .	389
9.2.5	Προφίλ Χρηστών . . . . .	391
9.2.6	Συστήνοντας Στοιχεία σε Χρήστες με Βάση το Περιεχόμενο . . . . .	392
9.2.7	Αλγόριθμοι Κατηγοριοποίησης . . . . .	393
9.2.8	Ασκήσεις Ενότητας 9.2 . . . . .	396
9.3	Συνεργατική Διήθηση . . . . .	398
9.3.1	Μέτρηση Ομοιότητας . . . . .	398
9.3.2	Δυαδικότητα της Ομοιότητας . . . . .	401
9.3.3	Ομαδοποίηση Χρηστών και Στοιχείων . . . . .	403
9.3.4	Ασκήσεις Ενότητας 9.3 . . . . .	405
9.4	Μείωση Διαστάσεων . . . . .	406

9.4.1	Αποσύνθεση-UV . . . . .	407
9.4.2	Λάθος Μέσης Τετραγωνικής Ρίζας . . . . .	408
9.4.3	Υπολογιστική Αύξηση της Αποσύνθεσης-UV . . . . .	409
9.4.4	Βελτιστοποιώντας ένα Αυθαίρετο Στοιχείο . . . . .	412
9.4.5	Δημιουργία Ολοκληρωμένου Αλγορίθμου Αποσύνθεσης-UV . . . . .	414
9.4.6	Ασκήσεις Ενότητας 9.4 . . . . .	417
9.5	Η Πρόκληση του NetFlix . . . . .	418
9.6	Περίληψη Κεφαλαίου 9 . . . . .	419
9.7	Αναφορές Κεφαλαίου 9 . . . . .	422