



ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΣΧΟΛΗ ΨΗΦΙΑΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ
Μ.Π.Σ. «ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗ»

**Αλγόριθμος για την εξαγωγή συνόλων συστάσεων από
δεδομένα προτιμήσεων**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Παναγιώτης Ε. Κουρής

Επιβλέπων Καθηγητής: Ηρακλής Βαρλάμης

Φεβρουάριος 2016



ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΣΧΟΛΗ ΨΗΦΙΑΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ
Μ.Π.Σ. «ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗ»

**Αλγόριθμος για την εξαγωγή συνόλων συστάσεων από
δεδομένα προτιμήσεων**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Παναγιώτης Ε. Κουρής

Επιβλέπων Καθηγητής: Ηρακλής Βαρλάμης

Μέλη Τριμελούς Επιτροπής: Μάρα Νικολαΐδου
Δημήτριος Μιχαήλ

Φεβρουάριος 2016

Αλγόριθμος για την εξαγωγή συνόλων συστάσεων από δεδομένα προτιμήσεων.
Algorithm for composing packages of recommendations using preference data.

Copyright © Παναγιώτης Ε. Κουρής, 2016
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Περίληψη

Τα τελευταία χρόνια όλο και περισσότερες εφαρμογές πληροφορικής έχουν ανάγκη την αυτοματοποιημένη πρόταση αντικειμένων λαμβάνοντας υπόψη τις προτιμήσεις των χρηστών. Η ανάγκη αυτή οδήγησε στην ανάπτυξη ερευνητικής δραστηριότητας στον τομέα αυτό ο οποίος αποτελεί τον ερευνητικό χώρο της παρούσας εργασίας.

Ο σκοπός της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη ενός αλγόριθμου συστάσεων που δεν προτείνει μεμονωμένα αντικείμενα μόνο αλλά συνθέτει σύνολα από αντικείμενα σύμφωνα με τα δεδομένα προτιμήσεων των χρηστών. Ο αλγόριθμος είναι ευέλικτος και δεν περιορίζεται σε κάποια συγκεκριμένη πρακτική εφαρμογή αλλά ενσωματώνει μια σειρά από παραμέτρους που η κατάλληλη επιλογή τους οδηγεί στην προσαρμογή του. Τα προτεινόμενα σύνολα αντικειμένων που εξάγονται περιλαμβάνουν αντικείμενα από έναν ορισμένο αριθμό κατηγοριών. Η παρούσα προσέγγιση χρησιμοποιεί την τεχνική του συνεργατικού φιλτραρίσματος και βασίζεται στον αλγόριθμο ροής ελάχιστου κόστους για τη βέλτιστη αντιστοίχιση των αντικειμένων με τις κατηγορίες και την εξαγωγή συνόλων αντικειμένων σύμφωνα με τις παραμέτρους.

Για την εφαρμογή και την αξιολόγηση της παρούσας προσέγγισης σχεδιάστηκε και υλοποιήθηκε κατάλληλο λογισμικό που υλοποιεί ένα ολοκληρωμένο σύστημα συστάσεων. Το λογισμικό αυτό περιλαμβάνει ένα εύχρηστο και φιλικό εποπτικό περιβάλλον χρήστη μέσα από το οποίο εκτελούνται όλες οι απαραίτητες λειτουργίες που αφορούν στην εισαγωγή του συνόλου δεδομένων, στην επιλογή των παραμέτρων, στο συνεργατικό φιλτράρισμα, στην εξαγωγή των k-καλύτερων πακέτων αντικειμένων και στην αξιολόγηση.

Η αξιολόγηση βοήθησε στην κατανόηση της συμπεριφοράς του συστήματος και έδειξε την αποδοτικότητα και την αποτελεσματικότητα τόσο για το χρόνο εκτέλεσης όσο και για την ακρίβεια που επιτυγχάνεται.

Λέξεις Κλειδιά: Αλγόριθμος συστάσεων, Σύστημα συστάσεων, Σύσταση συνόλων αντικειμένων, Σύσταση των k-καλύτερων πακέτων, Συνεργατικό φιλτράρισμα.

Abstract

In recent years more and more computer applications are in need of an automatic recommendation system which proposes items by taking user preferences into account. This need has led to the development of research activity in this area, which is the research area of this study.

This diploma thesis aims at the development of a recommendation algorithm that does not recommend individual items only, but it also composes packages of items according to the data based on the users' preference. The algorithm is flexible and is not limited to a particular practical application; it includes a number of parameters. Proper selection of these parameters leads to adaptation of the algorithm in a particular application. The proposed packages of items that are produced include items from a stated number of categories. This approach uses the technique of collaborative filtering and is based on the minimum cost flow algorithm for optimal matching of items to categories with aim of composing the recommendation packages according to the parameters.

An appropriate software that implements an integrated recommendation system was designed and developed for the testing and evaluation of this approach. This software includes a useful, friendly and graphical user interface through which all the necessary functions are executed. These functions include input of the dataset, setting of the parameters, collaborative filtering, recommendation of top-k packages and evaluation of this approach.

The evaluation enabled the system behavior to become comprehensible and showed the efficiency and effectiveness of both the execution time and the precision that it is achieved.

Keywords: Recommendation algorithm, Recommendation system, Package recommendations, Recommendation of the top-k packages, Collaborative filtering.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον καθηγητή μου κ. Ηρακλή Βαρλάμη που μου ανέθεσε αυτήν τη διπλωματική εργασία και μου έδωσε την ευκαιρία να έρθω σε επαφή με τον εξαιρετικά ενδιαφέροντα χώρο των συστημάτων συστάσεων. Ιδιαίτερα θα ήθελα να τον ευχαριστήσω για την καθοδήγηση, την πολύτιμη βοήθεια, τις συμβουλές, τις παρατηρήσεις και τον χρόνο που διέθεσε καθ' όλη τη διάρκεια της εκπόνησης αυτής της εργασίας.

Επίσης, θα ήθελα να ευχαριστήσω όλους όσους με βοήθησαν ή μου συμπαράσταθηκαν στο δύσκολο αλλά και δημιουργικό αυτό διάστημα των μεταπτυχιακών μου σπουδών στο Χαροκόπειο Πανεπιστήμιο και ιδιαίτερα τη Λένα Σιδηροπούλου για την υπομονή, τη βοήθεια και τη συμπαράστασή της.

Πίνακας Περιεχομένων

Περίληψη	7
Abstract.....	9
Ευχαριστίες	11
Πίνακας Περιεχομένων.....	13
Κατάλογος Πινάκων.....	15
Κατάλογος Εικόνων	19
Κεφάλαιο 1: Εισαγωγή.....	23
1.1 Γενικά.....	23
1.2 Σκοπός.....	24
1.3 Αντικείμενο	24
1.4 Φάσεις Υλοποίησης.....	26
1.5 Δομή της Εργασίας.....	26
1.6 Σύνοψη Αποτελεσμάτων.....	27
Κεφάλαιο 2: Επιστημονικό και Τεχνολογικό Υπόβαθρο	29
2.1 Γενικά.....	29
2.2 Συστήματα Συστάσεων	29
2.3 Συνεργατικό Φιλτράρισμα	31
2.4 Σχετικές Εργασίες - Συστήματα συστάσεων για σύνολα αντικειμένων.....	34
2.4.1 Συστήματα με Αυστηρούς Περιορισμούς.....	35
2.4.2 Ευέλικτα Συστήματα	39
2.5 Μετρικές Αξιολόγησης Συστημάτων Συστάσεων.....	43
2.6 Απαιτούμενες Τεχνολογίες και Συστήματα Λογισμικού	45
Κεφάλαιο 3: Σχεδίαση Αλγόριθμου.....	47
3.1 Γενικά.....	47
3.2 Καθορισμός Προβλήματος.....	47
3.3 Μοντελοποίηση Αλγόριθμου	52
3.3.1 Είσοδος Δεδομένων.....	53
3.3.2 Επιλογή του Τρέχοντα Χρήστη.....	53
3.3.3 Εκτίμηση της Βαθμολογίας Προτίμησης των Αντικειμένων	53
3.3.4 Επιλογή των Παραμέτρων.....	53
3.3.5 Βέλτιστη Αντιστοίχιση των Αντικειμένων με τις Κατηγορίες	61
3.3.6 Ενημέρωση του Γραφήματος Μοντελοποίησης του Προβλήματος	67
3.3.7 Εξαγωγή και Προβολή των N-καλύτερων Πακέτων Αντικειμένων	67
Κεφάλαιο 4: Σχεδίαση Λογισμικού.....	69
4.1 Γενικά.....	69
4.2 Αρχιτεκτονική του Συστήματος Λογισμικού	69
4.3 Συμπεριφορά του Συστήματος Λογισμικού	70
Κεφάλαιο 5: Ανάπτυξη Λογισμικού.....	75
5.1 Γενικά.....	75
5.2 Πηγαίος Κώδικας.....	75
5.3 Γραφικό Περιβάλλον	81
5.3.1 Αρχική Οθόνη	81
5.3.2 Οθόνη Επιλογής του Τρόπου Λειτουργίας.....	86

5.3.3 Οθόνη Επιλογής των Παραμέτρων Μοντελοποίησης	90
5.3.4 Οθόνη Αναλογικής Κατανομής των Αντικειμένων στις Κατηγορίες	93
5.3.5 Οθόνη Προβολής των N-καλύτερων Προτεινόμενων Πακέτων	95
5.3.6 Οθόνη Αξιολόγησης.....	97
Κεφάλαιο 6: Εφαρμογή και Αξιολόγηση	99
6.1 Γενικά.....	99
6.2 Σύνολο Δεδομένων	99
6.3 Σχέδιο Αξιολόγησης	100
6.4 Προετοιμασία Αξιολόγησης.....	103
6.5 Μετρήσεις Αξιολόγησης.....	108
6.5.1 Αριθμός Αντικειμένων Συνεργατικού Φιλτραρίσματος	108
6.5.2 Συντελεστής Δημοφιλίας των Αντικειμένων	112
6.5.3 Αριθμός Προτεινόμενων Πακέτων	114
6.5.4 Αριθμός Κατηγοριών	117
6.5.5 Αριθμός Αντικειμένων ανά Πακέτο	119
6.5.6 Μοντέλο Παραγωγής των Καλύτερων Προτεινόμενων Πακέτων.....	125
6.5.7 Μοντέλο Επιλογής των Δημοφιλέστερων Κατηγοριών	129
6.6 Αποτελέσματα Αξιολόγησης.....	131
Κεφάλαιο 7: Συμπεράσματα και Προοπτικές	137
7.1 Συμπεράσματα.....	137
7.2 Προοπτικές	139
Αναφορές	141
Παράρτημα	145
A. Τεκμηρίωση του Πηγαίου Κώδικα.....	145
A.1 Κλάσεις Βασικών Οντοτήτων	145
A.2 Κλάσεις Οντοτήτων για το Σχηματισμό του Γραφήματος	147
A.3 Κλάσεις Διαχείρισης των Αρχείων του Συνόλου Δεδομένων	150
A.4 Κλάση Συνεργατικού Φιλτραρίσματος.....	151
A.5 Κλάσεις Επίλυσης του Προβλήματος Γραμμικού Προγραμματισμού	152
A.6 Κλάσεις Γραφικού Περιβάλλοντος.....	152

Κατάλογος Πινάκων

Πίνακας 2.4	Οι σχετικές επιστημονικές εργασίες που μελετήθηκαν.	41
Πίνακας 3.3.4.3(α)	Παράδειγμα με το πλήθος των αντικειμένων ανά κατηγορία και το μέγεθος του προτεινόμενου πακέτου.	57
Πίνακας 3.3.4.3(β)	Παράδειγμα εκτέλεσης του αλγόριθμου αναλογικής κατανομής των αντικειμένων στις κατηγορίες.	57
Πίνακας 3.3.4.3(γ)	Η αναλογική κατανομή των αντικειμένων στις τρεις κατηγορίες μετά την εκτέλεση του σχετικού αλγόριθμου.	58
Πίνακας 3.3.5(α)	Παράδειγμα βαθμολογιών προτίμησης και υπολογισμού κόστους.	64
Πίνακας 3.3.5 (β)	Η λύση του παραδείγματος της εικόνας 3.3.5(δ).	66
Πίνακας 6.4(α)	Το μέσο απόλυτο και το μέσο τετραγωνικό σφάλμα των εκτιμώμενων βαθμολογιών προτίμησης σε σχέση με το ποσοστό απόκρυψης αντικειμένων για χρήστες που έχουν βαθμολογήσει περίπου 500 και 1000 αντικείμενα, αντίστοιχα. Υπολογίζεται και το διάστημα εμπιστοσύνης ($\pm$ από τη μέση τιμή του δείγματος) για επίπεδο εμπιστοσύνης 95%.	104
Πίνακας 6.4(β)	Η τυπική απόκλιση του μέσου απόλυτου και του μέσου τετραγωνικού σφάλματος των προβλεπόμενων βαθμολογιών προτίμησης σε σχέση με το ποσοστό απόκρυψης αντικειμένων για χρήστες που έχουν βαθμολογήσει περίπου 500 και 1000 αντικείμενα, αντίστοιχα.	105
Πίνακας 6.4(γ)	Το μέσο απόλυτο και το μέσο τετραγωνικό σφάλμα των εκτιμώμενων βαθμολογιών προτίμησης σε σχέση με την ελάχιστη βαθμολογία προτίμησης των κρυμμένων αντικειμένων για χρήστες που έχουν βαθμολογήσει περίπου 500 και 1000 αντικείμενα, αντίστοιχα. Αναγράφεται και το διάστημα εμπιστοσύνης για επίπεδο εμπιστοσύνης 95%.	107
Πίνακας 6.5.1(α)	Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης για τον αριθμό των αντικειμένων του συνεργατικού φιλτραρίσματος.	110
Πίνακας 6.5.1(β)	Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης που αφορούν στον αριθμό των αντικειμένων του συνεργατικού φιλτραρίσματος και στον συντελεστή δημοφιλίας των αντικειμένων.	110

Πίνακας 6.5.1(γ)	Οι μετρήσεις για τον χρόνο εκτέλεσης του αλγόριθμου, για τη συνολική βαθμολογία προτίμησης (<i>rating</i>) των αντικειμένων που περιέχονται στα προτεινόμενα πακέτα και για το δείκτη <i>precision</i> (με τα 95% διαστήματα εμπιστοσύνης) σε σχέση με τον αριθμό των εκτιμώμενων αντικειμένων που προέρχονται από το συνεργατικό φιλτράρισμα.	110
Πίνακας 6.5.2(α)	Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης του αλγόριθμου που αφορούν στον συντελεστή δημοφιλίας των αντικειμένων.	113
Πίνακας 6.5.2(β)	Ο δείκτης <i>Precision</i> για τις διάφορες τιμές του συντελεστή δημοφιλίας των αντικειμένων.	114
Πίνακας 6.5.3(α)	Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης του αλγόριθμου που αφορούν στον αριθμό των πακέτων, στον αριθμό των κατηγοριών, στον αριθμό των αντικειμένων ανά πακέτο, στο μοντέλο παραγωγής πακέτων και στο μοντέλο επιλογής των δημοφιλέστερων κατηγοριών.	115
Πίνακας 6.5.3(β)	Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης που αφορούν στον αριθμό των προτεινόμενων πακέτων.	115
Πίνακας 6.5.3(γ)	Ο χρόνος εκτέλεσης του αλγόριθμου και ο δείκτης <i>Precision</i> , του χειρότερου πακέτου, του καλύτερου πακέτου και η μέση τιμή με το 95% διάστημα εμπιστοσύνης, σε σχέση με τον αριθμό των παραγόμενων πακέτων.	115
Πίνακας 6.5.4(α)	Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης που αφορούν στη μεταβολή του αριθμού των κατηγοριών.	117
Πίνακας 6.5.4(β)	Ο χρόνος εκτέλεσης του αλγόριθμου και ο δείκτης <i>Precision</i> , του χειρότερου πακέτου, του καλύτερου πακέτου και της μέσης τιμής με τα 95% διαστήματα εμπιστοσύνης, σε σχέση με τον αριθμό των κατηγοριών των αντικειμένων που περιέχονται στα προτεινόμενα πακέτα.	118
Πίνακας 6.5.5.1(α)	Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης που αφορούν στη μεταβολή του αριθμού των αντικειμένων ανά κατηγορία προτεινόμενου πακέτου.	120
Πίνακας 6.5.5.1(β)	Ο χρόνος εκτέλεσης του αλγόριθμου και ο δείκτης <i>Precision</i> , του χειρότερου πακέτου, του καλύτερου πακέτου και της μέσης τιμής με το 95% διάστημα εμπιστοσύνης, σε σχέση με τον αριθμό των αντικειμένων ανά κατηγορία ή τον αριθμό των αντικειμένων ανά πακέτο.	120

Πίνακας 6.5.5.2(α)	Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης που αφορούν στη μεταβολή του αριθμού των αντικειμένων ανά κατηγορία προτεινόμενου πακέτου.	122
Πίνακας 6.5.5.2(β)	Ο <i>χρόνος εκτέλεσης</i> του αλγόριθμου και ο δείκτης <i>Precision</i> , του χειρότερου πακέτου, του καλύτερου πακέτου και της μέσης τιμής στο σύνολο των πακέτων με το 95% διάστημα εμπιστοσύνης, σε σχέση με τον αριθμό των αντικειμένων ανά πακέτο για αναλογική κατανομή των αντικειμένων στις κατηγορίες.	123
Πίνακας 6.5.6(α)	Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης που αφορούν στα διάφορα μοντέλα παραγωγής πακέτων.	125
Πίνακας 6.5.6(β)	Μετρήσεις για τον <i>χρόνο εκτέλεσης</i> του αλγόριθμου, το <i>Rating</i> και το <i>Precision</i> για το χειρότερο πακέτο, το καλύτερο πακέτο και για τη μέση τιμή όλων των προτεινόμενων πακέτων με τα 95% διαστήματα εμπιστοσύνης, σε σχέση με τα έξι μοντέλα παραγωγής των προτεινόμενων πακέτων.	126
Πίνακας 6.5.7(α)	Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης που αφορούν στα μοντέλα επιλογής των δημοφιλέστερων κατηγοριών.	129
Πίνακας 6.5.7(β)	Μετρήσεις για το <i>χρόνο εκτέλεσης</i> του αλγόριθμου, το <i>Rating</i> και το <i>Precision</i> , για το χειρότερο πακέτο, το καλύτερο πακέτο και για τη μέση τιμή όλων των προτεινόμενων πακέτων με τα 95% διαστήματα εμπιστοσύνης, σε σχέση με τα τρία μοντέλα προσδιορισμού των δημοφιλέστερων κατηγοριών.	130
Πίνακας 6.6	Οι βέλτιστες τιμές των παραμέτρων σύμφωνα με τις μετρήσεις αξιολόγησης.	135
Πίνακες Παραρτήματος:		
Πίνακας A.3 (α)	Οι μέθοδοι της κλάσης <i>ReadFiles</i> .	150
Πίνακας A.3 (β)	Οι μέθοδοι της κλάσης <i>ConvertFiles</i> .	151
Πίνακας A.4	Οι μέθοδοι της κλάσης <i>CollaborativeFiltering</i> .	151
Πίνακας A.5	Οι μέθοδοι της κλάσης <i>LinearProgramming</i> .	152
Πίνακας A.6 (α)	Οι μέθοδοι της κλάσης <i>StartForm</i> .	153
Πίνακας A.6 (β)	Οι μέθοδοι της κλάσης <i>RunningModeForm</i> .	154
Πίνακας A.6 (γ)	Οι μέθοδοι της κλάσης <i>ProblemModelingForm</i> .	155
Πίνακας A.6 (δ)	Οι μέθοδοι της κλάσης <i>ItemsAllocationToCategoriesForm</i> .	156
Πίνακας A.6 (ε)	Οι μέθοδοι της κλάσης <i>ProblemSolutionNPackagesForm</i> .	156
Πίνακας A.6 (στ)	Οι μέθοδοι της κλάσης <i>EvaluationMeasuresForm</i> .	157

Κατάλογος Εικόνων

Εικόνα 3.3	Διάγραμμα ροής για τη λειτουργία του αλγόριθμου.	52
Εικόνα 3.3.5(α)	Διμερές γράφημα $G(V, E)$ αντικειμένων και κατηγοριών.	62
Εικόνα 3.3.5(β)	Γράφημα μέγιστης ροής για τη βέλτιστη αντιστοίχιση μεταξύ αντικειμένων και κατηγοριών.	62
Εικόνα 3.3.5(γ)	Χαρακτηρισμός ακμής.	64
Εικόνα 3.3.5(δ)	Γράφημα ροής ελάχιστου κόστους.	65
Εικόνα 4.2	Το γενικό αρχιτεκτονικό διάγραμμα της εφαρμογής λογισμικού που υλοποιεί την λειτουργικότητα του αλγόριθμου.	69
Εικόνα 4.3(α)	Το συνεργατικό διάγραμμα του συστήματος συστάσεων πακέτων αντικειμένων για την κανονική λειτουργία.	71
Εικόνα 4.3(β)	Το συνεργατικό διάγραμμα του συστήματος συστάσεων πακέτων αντικειμένων για τη λειτουργία αξιολόγησης.	72
Εικόνα 5.2(α)	Το <i>Component</i> διάγραμμα του συστήματος λογισμικού.	76
Εικόνα 5.2(β)	Το διάγραμμα των πακέτων λογισμικού της java εφαρμογής.	77
Εικόνα 5.2(γ)	Το διάγραμμα κλάσεων του πακέτου <code>entityBasicClasses</code> .	77
Εικόνα 5.2(δ)	Το διάγραμμα των κλάσεων του πακέτου λογισμικού <code>entityGraphClasses</code> .	78
Εικόνα 5.2(ε)	Το διάγραμμα των κλάσεων του πακέτου λογισμικού <code>fileManagerClasses</code> .	78
Εικόνα 5.2(στ)	Το διάγραμμα της κλάσης <code>CollaborativeFiltering</code> .	79
Εικόνα 5.2(ζ)	Το διάγραμμα κλάσεων του πακέτου λογισμικού <code>linearProgrammingClasses</code> .	79
Εικόνα 5.2(η)	Το διάγραμμα των κλάσεων του πακέτου λογισμικού <code>formsAndFunctionality</code> .	80
Εικόνα 5.2(θ)	Το συνολικό διάγραμμα των κλάσεων του λογισμικού.	81
Εικόνα 5.3.1(α)	Η αρχική οθόνη της εφαρμογής.	82
Εικόνα 5.3.1(β)	Το μενού <code>file</code> .	83
Εικόνα 5.3.1(γ)	Μήνυμα τερματισμού.	83
Εικόνα 5.3.1(δ)	Το μήνυμα του μενού <code>About</code> .	83
Εικόνα 5.3.1(ε)	Το παράθυρο επιλογής των αρχείων των δεδομένων.	84
Εικόνα 5.3.1(στ)	Μήνυμα λανθασμένης επιλογής αρχείου δεδομένων.	84
Εικόνα 5.3.1(ζ)	Μήνυμα προτροπής για την επιλογή μιας εγγραφής.	85
Εικόνα 5.3.1(η)	Μήνυμα επιβεβαίωσης για την επιλογή του συγκεκριμένου χρήστη.	85
Εικόνα 5.3.2(α)	Η οθόνη για την επιλογή του τρόπου λειτουργίας της εφαρμογής και την εκτέλεση του συνεργατικού φιλτραρίσματος.	86
Εικόνα 5.3.2(β)	Το μενού <code>file</code> .	87
Εικόνα 5.3.2(γ)	Μήνυμα ενημέρωσης για την εκτέλεση του συνεργατικού φιλτραρίσματος.	89

Εικόνα 5.3.3	Η οθόνη για την επιλογή των παραμέτρων λειτουργίας του αλγόριθμου και την μοντελοποίηση του προβλήματος.	91
Εικόνα 5.3.4	Η οθόνη για την αναλογική κατανομή των αντικειμένων στις δημοφιλέστερες για το χρήστη κατηγορίες.	94
Εικόνα 5.3.5(α)	Η οθόνη προβολής των N-καλύτερων προτεινόμενων πακέτων αντικειμένων.	95
Εικόνα 5.3.5(β)	Αναφορά με τα N-καλύτερα προτεινόμενα πακέτα αντικειμένων.	97
Εικόνα 5.3.6	Η οθόνη προβολής των δεικτών αξιολόγησης του αλγόριθμου.	97
Εικόνα 6.3	Το διάγραμμα ροής της διαδικασίας αξιολόγησης.	100
Εικόνα 6.4(α)	Το μέσο απόλυτο (<i>MAE</i>) και το μέσο τετραγωνικό σφάλμα (<i>RMSE</i>) των εκτιμώμενων βαθμολογιών προτίμησης σε σχέση με το ποσοστό απόκρυψης αντικειμένων για χρήστες που έχουν βαθμολογήσει περίπου 500 και 1000 αντικείμενα, αντίστοιχα.	105
Εικόνα 6.4(β)	Η τυπική απόκλιση του μέσου απόλυτου και του μέσου τετραγωνικού σφάλματος των προβλεπόμενων βαθμολογιών προτίμησης σε σχέση με το ποσοστό απόκρυψης αντικειμένων για χρήστες που έχουν βαθμολογήσει περίπου 500 και 1000 αντικείμενα, αντίστοιχα.	106
Εικόνα 6.4(γ)	Το μέσο απόλυτο (<i>MAE</i>) και το μέσο τετραγωνικό σφάλμα (<i>RMSE</i>) των προβλεπόμενων βαθμολογιών προτίμησης σε σχέση με την ελάχιστη βαθμολογία προτίμησης των κρυμμένων αντικειμένων για χρήστες που έχουν βαθμολογήσει περίπου 500 και 1000 αντικείμενα, αντίστοιχα.	107
Εικόνα 6.5.1(α)	Η μέση τιμή του δείκτη <i>Precision</i> για τα προτεινόμενα πακέτα σε σχέση με τον αριθμό των αντικειμένων για τον σχηματισμό πακέτων.	111
Εικόνα 6.5.1(β)	Ο χρόνος εκτέλεσης του αλγόριθμου σε σχέση με τον αριθμό των αντικειμένων για τον σχηματισμό πακέτων.	111
Εικόνα 6.5.2	Η μεταβολή του δείκτη <i>precision</i> σε σχέση με τη μεταβολή του συντελεστή δημοφιλίας των αντικειμένων.	114
Εικόνα 6.5.3(α)	Η μεταβολή του δείκτη <i>Precision</i> , του χειρότερου πακέτου, του καλύτερου πακέτου και η μέση τιμή στο σύνολο των πακέτων, σε σχέση με τον αριθμό των παραγόμενων πακέτων.	116
Εικόνα 6.5.3(β)	Η μεταβολή του χρόνου εκτέλεσης του αλγόριθμου σε σχέση με τον αριθμό των παραγόμενων πακέτων.	117
Εικόνα 6.5.4(α)	Η μεταβολή του δείκτη <i>Precision</i> , του χειρότερου πακέτου, του καλύτερου πακέτου και της μέσης τιμής στο σύνολο των προτεινόμενων πακέτων, σε σχέση με τον αριθμό των κατηγοριών των αντικειμένων που περιέχονται στα προτεινόμενα πακέτα.	118
Εικόνα 6.5.4(β):	Η μεταβολή του χρόνου εκτέλεσης του αλγόριθμου σε σχέση με τον αριθμό των κατηγοριών των αντικειμένων που περιέχονται στα προτεινόμενα πακέτα.	119
Εικόνα 6.5.5.1(α)	Η μεταβολή του δείκτη <i>Precision</i> , του χειρότερου πακέτου, του καλύτερου πακέτου και της μέσης τιμής, σε σχέση με τον αριθμό των αντικειμένων ανά κατηγορία.	121
Εικόνα 6.5.5.1(β)	Η μεταβολή του χρόνου εκτέλεσης του αλγόριθμου σε σχέση με τον αριθμό των αντικειμένων ανά κατηγορία.	122

Εικόνα 6.5.5.2(α)	Η μεταβολή του δείκτη <i>Precision</i> , του χειρότερου πακέτου, του καλύτερου πακέτου και η μέση τιμή, σε σχέση με τον αριθμό των αντικειμένων ανά πακέτο για αναλογική κατανομή των αντικειμένων σε τέσσερις κατηγορίες.	123
Εικόνα 6.5.5.2(β)	Η μεταβολή του χρόνου εκτέλεσης του αλγόριθμου σε σχέση με τον αριθμό των αντικειμένων ανά πακέτο για αναλογική κατανομή των αντικειμένων σε τέσσερις κατηγορίες.	124
Εικόνα 6.5.6(α)	Οι τιμές του <i>Rating</i> για το χειρότερο πακέτο, το καλύτερο πακέτο και για τη μέση τιμή όλων των προτεινόμενων πακέτων, για τα έξι μοντέλα παραγωγής των προτεινόμενων πακέτων.	126
Εικόνα 6.5.6(β)	Ο δείκτης <i>Precision</i> για το χειρότερο πακέτο, το καλύτερο πακέτο και για τη μέση τιμή όλων των προτεινόμενων πακέτων σε σχέση με τα έξι μοντέλα παραγωγής των προτεινόμενων πακέτων.	127
Εικόνα 6.5.6(γ)	Ο χρόνος εκτέλεσης του αλγόριθμου για τα έξι μοντέλα παραγωγής των προτεινόμενων πακέτων.	128
Εικόνα 6.5.7(α)	Οι τιμές του <i>Rating</i> για το χειρότερο πακέτο, το καλύτερο πακέτο και για τη μέση τιμή όλων των προτεινόμενων πακέτων σε σχέση με τα τρία μοντέλα προσδιορισμού και επιλογής των δημοφιλέστερων κατηγοριών.	130
Εικόνα 6.5.7(β):	Οι τιμές του δείκτη <i>Precision</i> , για το καλύτερο πακέτο και για τη μέση τιμή όλων των προτεινόμενων πακέτων, σε σχέση με τα τρία μοντέλα προσδιορισμού και επιλογής των δημοφιλέστερων κατηγοριών.	131

Κεφάλαιο 1: Εισαγωγή

1.1 Γενικά

Τα τελευταία χρόνια έχει προκύψει η ανάγκη ανάπτυξης συστημάτων πληροφορικής που έχουν σκοπό να κάνουν συστάσεις ή προτάσεις σε χρήστες για αγαθά ή υπηρεσίες. Αυτά τα συστήματα βασίζονται, κυρίως, σε δεδομένα χρήσης ή προτιμήσεων των χρηστών και προτείνουν αντικείμενα ή πακέτα αντικειμένων που είναι επιθυμητά ή χρήσιμα στους χρήστες. Εξαιτίας αυτής της ανάγκης που έχει προκύψει, η οποία γίνεται όλο και πιο έντονη, υπάρχει μια σχετικά αυξημένη ερευνητική δραστηριότητα στον τομέα ανάπτυξης τέτοιων συστημάτων. Αυτή η ερευνητική δραστηριότητα, τις περισσότερες φορές, ωθείται από μία έντονη εμπορική και επιχειρηματική δράση που αναζητά την ανάπτυξη και την ενσωμάτωση τέτοιων συστημάτων αλλά και από τη γενικότερη ανάγκη που προκύπτει για τη διαχείριση και την αξιοποίηση μεγάλων όγκων από δεδομένα.

Τα συστήματα συστάσεων έχουν εφαρμογή σε πάρα πολλούς τομείς και είναι χρήσιμα όπου απαιτείται η αυτοματοποιημένη σύσταση αντικειμένων ή υπηρεσιών. Για παράδειγμα ένα ηλεκτρονικό κατάστημα θα μπορούσε να προτείνει βιβλία ή ταινίες σε χρήστες ανάλογα με το προφίλ τους, τις προηγούμενες προτιμήσεις τους, τις προτιμήσεις των φίλων τους ή γενικότερα τις προτιμήσεις των υπόλοιπων χρηστών.

Η εξέλιξη των συστημάτων συστάσεων αλλά και οι απαιτήσεις των εφαρμογών οδήγησαν στην ανάγκη να αναπτυχθούν συστήματα συστάσεων που προτείνουν σύνολα αντικειμένων. Η σχεδίαση και η ανάπτυξη ενός τέτοιου συστήματος αποτελεί το αντικείμενο αυτής της διπλωματικής εργασίας. Τα συστήματα συστάσεων που προτείνουν πακέτα αντικειμένων είναι χρήσιμα σε πολλές εφαρμογές. Κάποιες από αυτές τις εφαρμογές μπορεί να είναι η πρόταση πακέτων μαθημάτων σε φοιτητές, η πρόταση πακέτων με φαγητά για μια εβδομαδιαία διαίτα, η πρόταση πακέτων ταξιδιών, η πρόταση συνόλων από ταινίες ή η πρόταση συνόλων από αντικείμενα που είναι απαραίτητα για το σπίτι.

Σε αυτή την εργασία πραγματοποιείται μια εκτεταμένη έρευνα σχετικά με τα συστήματα συστάσεων και αναπτύσσεται ένας αλγόριθμος που υλοποιεί ένα τέτοιο σύστημα το οποίο προτείνει πακέτα αντικειμένων.

1.2 Σκοπός

Η διπλωματική αυτή εργασία έχει σκοπό τη σχεδίαση ενός ευέλικτου αλγόριθμου που χρησιμοποιεί δεδομένα προτιμήσεων για την εξαγωγή των καλύτερων συνόλων από αντικείμενα, καθώς και την ανάπτυξη λογισμικού που έχει στόχο την εφαρμογή, τη δοκιμή και την αξιολόγηση του αλγόριθμου αυτού μέσα από ένα εύχρηστο γραφικό περιβάλλον χρήστη.

Διευκρινίζεται ότι τα δεδομένα προτιμήσεων προέρχονται από χρήστες και τα καλύτερα σύνολα αντικειμένων προτείνονται στον τρέχοντα κάθε φορά χρήστη. Τα καλύτερα σύνολα αντικειμένων είναι εκείνα που αναμένεται να έχουν το μεγαλύτερο βαθμό προτίμησης από τον χρήστη στον οποίο προτείνονται. Το λογισμικό ενσωματώνει όλη την απαραίτητη λειτουργικότητα και την ευελιξία που υπαγορεύεται από τον αλγόριθμο και περιλαμβάνει ένα εύχρηστο και λειτουργικό γραφικό περιβάλλον χρήστη.

1.3 Αντικείμενο

Το αντικείμενο της διπλωματικής εργασίας αφορά τη σχεδίαση ενός αλγόριθμου και την υλοποίηση ενός ολοκληρωμένου συστήματος συστάσεων που προτείνει σύνολα αντικειμένων σε χρήστες. Ο αλγόριθμος που αναπτύσσεται παίρνει ως είσοδο τις προτιμήσεις ή τις επιλογές του τρέχοντα χρήστη και των υπόλοιπων χρηστών, καθώς και το σύνολο των αντικειμένων. Μετά από κατάλληλη διεργασία, ο αλγόριθμος επιστρέφει τα N-καλύτερα σύνολα αντικειμένων για τον τρέχοντα χρήστη. Κάθε σύνολο αντικειμένων που εξάγεται περιλαμβάνει αντικείμενα που ανήκουν σε μία ή περισσότερες κατηγορίες. Οι κατηγορίες είναι προκαθορισμένες και επιλέγονται με τέτοιο τρόπο ώστε να είναι οι δημοφιλέστερες για τον τρέχοντα χρήστη σύμφωνα με τις προτιμήσεις του. Κάθε αντικείμενο μπορεί να ανήκει σε μία ή περισσότερες κατηγορίες. Τα προτεινόμενα σύνολα αντικειμένων αναμένεται να είναι και τα καλύτερα σύνολα, εκείνα που πετυχαίνουν τον υψηλότερο βαθμό προτίμησης, σύμφωνα με τις εκτιμήσεις για τον τρέχοντα χρήστη.

Η λειτουργία του αλγόριθμου απαιτεί τη χρησιμοποίηση του κατάλληλου συνόλου δεδομένων και μετά από μια σειρά από διεργασίες εξάγονται τα N-καλύτερα για τον τρέχοντα χρήστη πακέτα αντικειμένων. Η πρώτη κατά σειρά διεργασία που εκτελείται είναι η πρόβλεψη της βαθμολογίας προτίμησης των νέων για τον τρέχοντα χρήστη αντικειμένων. Η πρόβλεψη της βαθμολογίας προτίμησης των αντικειμένων γίνεται με τη χρησιμοποίηση της τεχνικής του συνεργατικού φιλτραρίσματος (*collaborative filtering*). Για τον σχηματισμό των N-καλύτερων πακέτων

αντικειμένων απαιτείται να πραγματοποιηθεί η βέλτιστη αντιστοίχιση των αντικειμένων με τις κατηγορίες ώστε να προκύψουν τα πακέτα που πετυχαίνουν τους υψηλότερους βαθμούς προτίμησης.

Ένα μεγάλο πλεονέκτημα αυτού του αλγόριθμου είναι ότι μπορεί να χαρακτηριστεί ως ευέλικτος γιατί περιλαμβάνει έναν αριθμό από παραμέτρους που μπορούν να μεταβληθούν και να οδηγήσουν στην προσαρμοσμένη χρήση. Η αλλαγή των παραμέτρων αυτών οδηγεί σε αλλαγή της συμπεριφοράς του αλγόριθμου με αποτέλεσμα να είναι δυνατή η προσαρμογή του σε ένα εύρος από εφαρμογές ή προβλήματα. Μερικές από αυτές τις παραμέτρους αφορούν στο μέγεθος του πακέτου, στον αριθμό των κατηγοριών ανά πακέτο, στον αριθμό των αντικειμένων ανά κατηγορία, στο μοντέλο σχηματισμού των πακέτων ή σε άλλες περισσότερο εξειδικευμένες παραμέτρους που αναφέρονται στο κεφάλαιο 3. Ο αλγόριθμος σχεδιάζεται με τρόπο που να μπορεί να χαρακτηριστεί ως γενικός και ευέλικτος, με στόχο να μπορεί να χρησιμοποιηθεί από πολλές εφαρμογές με την κατάλληλη επιλογή των παραμέτρων που διαθέτει.

Η εφαρμογή, η δοκιμή και η αξιολόγηση του αλγόριθμου απαιτούν τη σχεδίαση ενός λογισμικού που ενσωματώνει όλη την απαραίτητη λειτουργικότητα και εκπληρώνει τους στόχους χρησιμοποίησης της τεχνολογίας αυτής. Στα πλαίσια αυτής της εργασίας σχεδιάζεται και αναπτύσσεται λογισμικό, το οποίο περιλαμβάνει ένα εύχρηστο και γραφικό περιβάλλον χρήστη που δίνει τη δυνατότητα της αποτελεσματικής διαχείρισης των απαιτούμενων λειτουργιών. Οι κύριες λειτουργίες του λογισμικού είναι η αρχική φόρτωση του συνόλου των δεδομένων, η επιλογή του τρέχοντα χρήστη, η επιλογή των παραμέτρων λειτουργίας, η εκτέλεση του συνεργατικού φιλτραρίσματος, ο σχηματισμός των N-καλύτερων πακέτων αντικειμένων και ο υπολογισμός των απαιτούμενων δεικτών αξιολόγησης.

Για την αξιολόγηση του συστήματος χρησιμοποιούνται δεδομένα εκπαίδευσης και δεδομένα ελέγχου με τη βοήθεια των οποίων γίνεται η δοκιμή και ο έλεγχος για την αποτελεσματικότητα και την ακρίβεια που επιτυγχάνεται. Επίσης, η αξιολόγησή του βοηθάει στην περεταίρω βελτίωση της τεχνολογίας αυτής, μέσα από μια διαδικασία αξιολόγησης και βελτίωσης που επαναλαμβάνεται μέχρι το σύστημα να απαντά με αξιοπιστία στο πρόβλημα της πρότασης συνόλων αντικειμένων, τα οποία ικανοποιούν και ανταποκρίνονται στις επιθυμίες των χρηστών.

Σε αυτή την ενότητα έγινε μια γενική περιγραφή του προβλήματος που επιχειρεί να αντιμετωπίσει η παρούσα εργασία. Περισσότερες λεπτομέρειες για τον αλγόριθμο, το λογισμικό, την εφαρμογή και την αξιολόγηση του συστήματος αναφέρονται στα επόμενα κεφάλαια.

1.4 Φάσεις Υλοποίησης

Η εκπόνηση της διπλωματικής εργασίας πραγματοποιήθηκε μεταξύ Σεπτεμβρίου 2015 και Φεβρουαρίου 2016 και η πορεία αυτής ακολούθησε τις φάσεις που αναφέρονται παρακάτω.

Φάση 1^η – Βιβλιογραφική αναζήτηση: Έγινε μία εκτεταμένη βιβλιογραφική αναζήτηση σε βιβλία και επιστημονικά άρθρα με σκοπό τη διερεύνηση του αντικειμένου των συστημάτων συστάσεων και ειδικότερα των συστημάτων που προτείνουν σύνολα αντικειμένων με τη χρησιμοποίηση δεδομένων προτιμήσεων από χρήστες.

Φάση 2^η – Οργάνωση και καταγραφή των σχετικών εργασιών: Μέσα από τη βιβλιογραφική αναζήτηση έγινε η οργάνωση και η καταγραφή του επιστημονικού υπόβαθρου και των σχετικών ερευνητικών εργασιών.

Φάση 3^η – Σχεδίαση αλγόριθμου: Σχεδιάστηκε ο αλγόριθμος για την εξαγωγή συνόλων συστάσεων από δεδομένα προτιμήσεων ο οποίος υπακούει στο πρόβλημα συστάσεων που καλείται να αντιμετωπίσει.

Φάση 4^η – Σχεδίαση και Ανάπτυξη λογισμικού: Σχεδιάστηκε και αναπτύχθηκε το απαιτούμενο λογισμικό που ενσωματώνει τη λειτουργικότητα του αλγόριθμου και έχει σκοπό την εφαρμογή, τη δοκιμή και την αξιολόγησή του.

Φάση 5^η – Εφαρμογή και αξιολόγηση: Πραγματοποιήθηκε η εφαρμογή και η αξιολόγηση του αλγόριθμου με τη βοήθεια του λογισμικού που αναπτύχθηκε για διάφορα σενάρια χρήσης.

Φάση 6^η – Εξαγωγή συμπερασμάτων και προοπτικών: Εξήχθησαν τα σημαντικότερα συμπεράσματα και έγινε αναφορά στις προοπτικές που ανοίγει η παρούσα διπλωματική εργασία.

1.5 Δομή της Εργασίας

Η παρούσα εργασία είναι οργανωμένη σε επτά κεφάλαια. Το κεφάλαιο 1 περιλαμβάνει την εισαγωγή, το κεφάλαιο 2 αναφέρει το θεωρητικό και τεχνολογικό υπόβαθρο καθώς και τις σχετικές εργασίες, το κεφάλαιο 3 περιλαμβάνει τη σχεδίαση του αλγόριθμου, το κεφάλαιο 4 τη σχεδίαση του λογισμικού, το κεφάλαιο 5 αναφέρεται στην υλοποίηση του λογισμικού, το κεφάλαιο 6 στην εφαρμογή και την αξιολόγηση του αλγορίθμου και στο κεφάλαιο 7 αναφέρονται τα συμπεράσματα που προκύπτουν και οι προοπτικές που ανοίγονται. Η εργασία περιλαμβάνει και παράρτημα που εκεί περιγράφεται ο πηγαίος κώδικας του λογισμικού. Αναλυτικότερα:

Κεφάλαιο 1 – Εισαγωγή: Περιλαμβάνονται γενικά στοιχεία για το αντικείμενο της παρούσας εργασίας και αναφέρεται ο σκοπός, το αντικείμενο, οι φάσεις υλοποίησης, η δομή του κειμένου και η σύνοψη των αποτελεσμάτων.

Κεφάλαιο 2 – Επιστημονικό και τεχνολογικό υπόβαθρο: Περιγράφεται το θεωρητικό υπόβαθρο για τα συστήματα συστάσεων, το συνεργατικό φιλτράρισμα και τους δείκτες αξιολόγησης τέτοιων συστημάτων. Αναφέρονται οι σχετικές ερευνητικές εργασίες και οι τεχνολογίες που χρησιμοποιούνται.

Κεφάλαιο 3 – Σχεδίαση Αλγόριθμου: Καθορίζεται το πρόβλημα που καλείται να αντιμετωπίσει ο προς σχεδίαση αλγόριθμος ο οποίος μοντελοποιείται και περιγράφεται.

Κεφάλαιο 4 – Σχεδίαση Λογισμικού: Περιγράφεται η σχεδίαση του λογισμικού το οποίο ενσωματώνει τη λειτουργικότητα του αλγόριθμου. Παρουσιάζεται η αρχιτεκτονική του συστήματος και τα συνεργατικά διαγράμματα που αφορούν στη σχεδίαση του λογισμικού.

Κεφάλαιο 5 – Ανάπτυξη Λογισμικού: Περιγράφεται ο πηγαίος κώδικας και το γραφικό περιβάλλον του λογισμικού που αναπτύχθηκε.

Κεφάλαιο 6 – Εφαρμογή και Αξιολόγηση: Δίνονται όλες οι απαραίτητες λεπτομέρειες που αφορούν στην εφαρμογή και στην αξιολόγηση του συστήματος. Περιγράφεται το σύνολο δεδομένων και η διαδικασία αξιολόγησης. Πραγματοποιούνται οι μετρήσεις αξιολόγησης και παρουσιάζονται τα αποτελέσματα.

Κεφάλαιο 7 – Συμπεράσματα και Προοπτικές: Παραθέτονται τα συμπεράσματα που προκύπτουν από την παρούσα διπλωματική εργασία και αναφέρονται οι προοπτικές που ανοίγονται.

Παράρτημα Α - Τεκμηρίωση του πηγαίου κώδικα: Δίνεται μια περιγραφή του πηγαίου κώδικα αναφέροντας τα χαρακτηριστικά των κλάσεων και των μεθόδων τους.

1.6 Σύνοψη Αποτελεσμάτων

Μετά την ολοκλήρωση αυτής της εργασίας, φαίνεται ότι το τελικό αποτέλεσμα αποζημιώνει τους κόπους και εκπληρώνει τις προσδοκίες σχετικά με την αποτελεσματικότητα και την αποδοτικότητα του αλγόριθμου που σχεδιάστηκε. Ο αλγόριθμος δοκιμάστηκε και αξιολογήθηκε με τη βοήθεια του σχετικού λογισμικού που αναπτύχθηκε. Το λογισμικό εκπληρώνει τον σκοπό του καθώς ενσωματώνει όλα τα χαρακτηριστικά του αλγόριθμου και επιτρέπει τη διαχείριση, την εφαρμογή και την αξιολόγησή του μέσα από ένα εύχρηστο γραφικό περιβάλλον. Σύμφωνα με τις μετρήσεις αξιολόγησης οι οποίες περιγράφονται στο κεφάλαιο 6, ο αλγόριθμος που

αναπτύχθηκε φαίνεται ότι λειτουργεί με αποδοτικό τρόπο και επιτυγχάνει υψηλά ποσοστά ακρίβειας.

Τα θετικά αυτά αποτελέσματα μπορούν να καταστήσουν την τεχνολογία αυτή χρήσιμη για πάρα πολλές εφαρμογές. Το πιο σημαντικό είναι ότι η τεχνολογία αυτή μπορεί να αποτελέσει αφετηρία για την ανάπτυξη άλλων, περισσότερο σύνθετων ή εξειδικευμένων συστημάτων συστάσεων που προτείνουν σύνολα αντικειμένων.

Κεφάλαιο 2: Επιστημονικό και Τεχνολογικό Υπόβαθρο

2.1 Γενικά

Σε αυτό το κεφάλαιο περιγράφεται το θεωρητικό και το τεχνολογικό υπόβαθρο στο οποίο βασίζεται η παρούσα διπλωματική εργασία. Στις επόμενες παραγράφους γίνεται μια γενική περιγραφή των συστημάτων συστάσεων (*Recommendation Systems*) και ειδικότερα των συστημάτων συστάσεων συνεργατικού φιλτραρίσματος (*Collaborative Filtering Recommendation Systems*). Καταγράφονται οι σχετικές ερευνητικές εργασίες που αφορούν σε συστήματα προτάσεων που εξάγουν σύνολα αντικειμένων τα οποία κυρίως βασίζονται σε δεδομένα προτιμήσεων. Περιγράφονται οι βασικοί δείκτες και οι μετρικές που χρησιμοποιούνται για την αξιολόγηση τέτοιων συστημάτων. Τέλος, αναφέρονται τα συστήματα λογισμικού και οι τεχνολογίες που χρησιμοποιούνται με σκοπό τη δημιουργία ενός ολοκληρωμένου συστήματος συστάσεων για την εφαρμογή και τον έλεγχο του αλγόριθμου που προτείνεται.

2.2 Συστήματα Συστάσεων

Τα συστήματα συστάσεων (*Recommendation Systems*) εξάγουν προβλέψεις σύμφωνα με τις προτιμήσεις των χρηστών (*users*) και έχουν σκοπό την πρόταση αντικειμένων (*items*) σε χρήστες. Η πρόβλεψη των προτιμήσεων γίνεται με βάση τη γνώση που διαθέτει το σύστημα από προηγούμενες προτιμήσεις χρηστών ή την προηγούμενη εμπειρία του συστήματος από τα δεδομένα που έχει συλλέξει σχετικά με τα αντικείμενα και τους χρήστες.

Σύμφωνα με την εργασία των Adomavicius και Tuzhilin (2005), τα συστήματα συστάσεων έχουν αρχίσει να απασχολούν την επιστήμη και να γίνονται ανεξάρτητη επιστημονική περιοχή έρευνας από τα μέσα της δεκαετίας του 1990. Επίσης, σύμφωνα με τους Adomavicius και Tuzhilin (2005) και τους Melville και Sindhvani (2010), το θεωρητικό υπόβαθρο των συστημάτων συστάσεων βασίζεται σε μία σειρά από επιστημονικές περιοχές όπως τη στατιστική, τη γνωστική επιστήμη (Rich, 1979), τη θεωρία προβλέψεων, τη μηχανική μάθηση και την περιοχή της εξόρυξης γνώσης. Όπως γίνεται εύκολα φανερό, η προϋπάρχουσα επιστημονική μελέτη σε αυτούς τους

βασικούς τομείς έχει συμβάλει στην ανάπτυξη των συστημάτων συστάσεων που χρησιμοποιούν δεδομένα χρήσης ή προτιμήσεων από χρήστες.

Το πρόβλημα που καλούνται να αντιμετωπίσουν τα συστήματα συστάσεων, τις περισσότερες φορές και σύμφωνα με τους Adomavicius και Tuzhilin (2005) και τους Melville και Sindhvani (2010), μοντελοποιείται ως ένα πρόβλημα που έχει σκοπό την εκτίμηση μιας βαθμολογίας προτίμησης (*rating*) για τα αντικείμενα που ο τρέχων χρήστης δεν έχει βαθμολογήσει και θα μπορούσε το σύστημα να του προτείνει. Η πρόβλεψη αυτή βασίζεται σε προηγούμενες αξιολογήσεις για την προτίμηση αντικειμένων που έχει κάνει ο τρέχων χρήστης και σε βαθμολογίες προτίμησης άλλων χρηστών. Όταν ολοκληρώνεται η διαδικασία της εκτίμησης των βαθμολογιών προτίμησης για τα αντικείμενα που ο τρέχων χρήστης δεν έχει βαθμολογήσει, τότε το σύστημα συστάσεων του προτείνει τα αντικείμενα με την υψηλότερη εκτιμώμενη βαθμολογία προτίμησης.

Για την εκτίμηση των βαθμολογιών προτίμησης των νέων αντικειμένων για κάποιο χρήστη έχουν προταθεί διάφορες μέθοδοι που βασίζονται σε διαφορετικές μεταξύ τους προσεγγίσεις για την αντιμετώπιση του προβλήματος της πρόβλεψης των προτιμήσεων και της πρότασης των αντικειμένων. Με βάση τα διαφορετικά μοντέλα που έχουν παρουσιαστεί, και σύμφωνα με τους Kantor et al. (2011), τα συστήματα συστάσεων ταξινομούνται ανάλογα με τον τρόπο που προσδιορίζουν την εκτιμώμενη βαθμολογία προτίμησης των νέων αντικειμένων, ως εξής:

Content-Based Recommendation Systems: Συστήματα προτάσεων που βασίζονται στο περιεχόμενο. Πρόκειται για συστήματα που προτείνουν αντικείμενα στους χρήστες παρόμοια με εκείνα που οι χρήστες έχουν επιλέξει στο παρελθόν.

Collaborative Filtering Recommendation Systems: Συστήματα προτάσεων συνεργατικού φιλτραρίσματος που βασίζονται στις προτιμήσεις άλλων χρηστών που έχουν παρόμοια συμπεριφορά προτιμήσεων με τον τρέχοντα χρήστη. Τα συστήματα αυτά προτείνουν αντικείμενα στους χρήστες με βάση τα αντικείμενα που έχουν επιλέξει ή βαθμολογήσει άλλοι χρήστες και ταυτόχρονα οι δεύτεροι έχουν παρόμοια χαρακτηριστικά επιλογής αντικειμένων με τους πρώτους.

Demographic Recommendation Systems: Συστήματα προτάσεων που βασίζονται στα δημογραφικά χαρακτηριστικά των χρηστών. Τα συστήματα

αυτά βασίζονται στην αρχή ότι οι χρήστες που ανήκουν στην ίδια δημογραφική περιοχή έχουν παρόμοια κριτήρια επιλογής αντικειμένων.

Knowledge-based Recommendation Systems: Συστήματα προτάσεων τα οποία βασίζονται σε συγκεκριμένη γνώση σχετικά με τον βαθμό που οι χρήστες έχουν ανάγκη κάποια συγκεκριμένα χαρακτηριστικά των αντικειμένων. Η γνώση αυτή οδηγεί στο συμπέρασμα εξακρίβωσης του βαθμού χρησιμότητας των αντικειμένων από τους χρήστες με σκοπό την πρόταση των αντικειμένων με το μεγαλύτερο βαθμό χρησιμότητας.

Community-based Recommendation Systems: Συστήματα προτάσεων που βασίζονται στις προτιμήσεις μιας κοινότητας χρηστών. Αυτά τα συστήματα προτείνουν αντικείμενα στους χρήστες σύμφωνα με τις προτιμήσεις των φίλων τους ή τις προτιμήσεις των υπόλοιπων μελών της ίδιας κοινότητας.

Hybrid Recommendation Systems: Υβριδικά συστήματα που συνδυάζουν τις παραπάνω μεθόδους και τεχνικές με σκοπό να πετύχουν το βέλτιστο αποτέλεσμα για το πρόβλημα προτάσεων που έχουν να αντιμετωπίσουν.

Στην επόμενη ενότητα γίνεται μια αναφορά στην τεχνική του συνεργατικού φιλτραρίσματος (*Collaborative Filtering*) και στα συστήματα συστάσεων συνεργατικού φιλτραρίσματος, καθώς είναι μια τεχνική που χρησιμοποιείται στην παρούσα εργασία.

2.3 Συνεργατικό Φιλτράρισμα

Τα συστήματα συστάσεων συνεργατικού φιλτραρίσματος προτείνουν αντικείμενα στους χρήστες με βάση τα αντικείμενα που προτιμούν άλλοι χρήστες με παρόμοιο κριτήριο προτιμήσεων. Το πρόβλημα που αντιμετωπίζουν τα συστήματα αυτά, σύμφωνα με τις εργασίες των Ekstrand et al. (2010) και Schafer et al. (2007), περιλαμβάνει δύο διεργασίες. Η πρώτη αφορά στη διαδικασία πρόβλεψης και απαντά στο ερώτημα:

Με δεδομένο έναν χρήστη και ένα αντικείμενο, ποιος είναι ο βαθμός προτίμησης του αντικειμένου αυτού από τον χρήστη;

Η δεύτερη διεργασία αφορά στη διαδικασία εξαγωγής συστάσεων και απαντά στο ερώτημα:

Με δεδομένο έναν χρήστη και ένα σύνολο από αντικείμενα, ποια είναι τα αντικείμενα εκείνα που έχουν τον υψηλότερο βαθμό προτίμησης από τον χρήστη;

Σύμφωνα με τους Melville και Sindhvani (2010), τα συστήματα συστάσεων συνεργατικού φιλτραρίσματος μπορούν να διακριθούν σε δύο γενικές κατηγορίες, στα Neighborhood-based και στα Model-based συστήματα. Αναλυτικότερα:

Neighborhood-based Collaborative Filtering Recommendation Systems: Συστήματα συστάσεων συνεργατικού φιλτραρίσματος που βασίζονται στους κοντινότερους γείτονες των χρηστών. Οι κοντινότεροι γείτονες είναι ένα υποσύνολο από χρήστες που επιλέγεται με βάση την ομοιότητα που έχουν με τον τρέχοντα χρήστη. Ο συνδυασμός των προτιμήσεων αυτού του υποσυνόλου χρησιμοποιείται για να παραχθούν προβλέψεις προτιμήσεων και προτάσεων για τον τρέχοντα χρήστη. Αυτά τα συστήματα μπορούν με τη σειρά τους να διακριθούν στα *User-based* και στα *Item-based* συστήματα συστάσεων συνεργατικού φιλτραρίσματος.

Model-based Collaborative Filtering Recommendation Systems: Συστήματα συστάσεων συνεργατικού φιλτραρίσματος που βασίζονται σε στατιστικά μοντέλα. Σύμφωνα με τους Melville και Sindhvani (2010), είναι συστήματα που χρησιμοποιούν στατιστικά μοντέλα για να εξαγάγουν προβλέψεις για τις προτιμήσεις των χρηστών. Πιο συγκεκριμένα, θα μπορούσε ένα τέτοιο σύστημα να προτείνει αντικείμενα σε χρήστες μοντελοποιώντας το πρόβλημα αυτό ως ένα πρόβλημα κατηγοριοποίησης (*Classification*). Σύμφωνα με το μοντέλο αυτό, το σύστημα αποφασίζει τον βαθμό προτίμησης των αντικειμένων για τον τρέχοντα χρήστη σύμφωνα με τα χαρακτηριστικά της συστάδας που αυτός ανήκει. Με ανάλογο τρόπο χρησιμοποιούνται και άλλες τεχνικές που βασίζονται σε στατιστικά μοντέλα για την εκτίμηση των προβλέψεων και την εξαγωγή προτάσεων.

Σύμφωνα με τους Ekstrand et al. (2010) μια ειδικότερη κατάταξη των συστημάτων συστάσεων συνεργατικού φιλτραρίσματος περιλαμβάνει δύο σημαντικές κατηγορίες, τα *User-Based* και τα *Item-Based* συστήματα. Είναι συστήματα που βασίζονται στους χρήστες ή στα αντικείμενα αντίστοιχα. Όπως έχει ήδη αναφερθεί οι δύο αυτές ομάδες συστημάτων περιλαμβάνονται στην κατηγορία των *Neighborhood-based* συστημάτων που αναφέρθηκε παραπάνω. Αναλυτικότερα:

Τα συστήματα συνεργατικού φιλτραρίσματος που βασίζονται στους χρήστες εκτιμούν τις βαθμολογίες προτίμησης του τρέχοντα χρήστη για τα νέα αντικείμενα χρησιμοποιώντας τις βαθμολογίες προτίμησης των άλλων χρηστών, οι οποίοι έχουν παρόμοια συμπεριφορά προτιμήσεων με τον τρέχοντα χρήστη. Η εξαγωγή των προβλέψεων γίνεται σε δύο βασικά βήματα. Στο πρώτο βήμα, το σύστημα αναζητά τους χρήστες εκείνους που έχουν παρόμοια συμπεριφορά στη βαθμολόγηση των αντικειμένων με τον τρέχοντα χρήστη. Αυτό προκύπτει από τον υπολογισμό της ομοιότητας των άλλων χρηστών με τον τρέχοντα χρήστη σύμφωνα με τις προτιμήσεις τους. Στο δεύτερο βήμα, χρησιμοποιούνται οι βαθμολογίες προτίμησης των χρηστών που έχουν υψηλή ομοιότητα με τον τρέχοντα χρήστη για τον υπολογισμό των προβλέψεων και την εξαγωγή των προτάσεων. Τα συστήματα αυτά παρουσιάζουν πρόβλημα επεκτασιμότητας (*scalability*) με την αύξηση του αριθμού των χρηστών. Για το λόγο αυτό είναι προτιμότερο να χρησιμοποιούνται σε περιπτώσεις που έχουμε σχετικά μικρό αριθμό χρηστών και γενικότερα σε εφαρμογές που ο αριθμός των χρηστών είναι μικρότερος από τον αριθμό των αντικειμένων.

Τα συστήματα συνεργατικού φιλτραρίσματος που βασίζονται στα αντικείμενα (Ekstrand et al. 2010; Sarwar et al. 2001) εκτιμούν και αυτά τις βαθμολογίες προτίμησης του τρέχοντα χρήστη για τα νέα αντικείμενα λαμβάνοντας υπόψη την ομοιότητα που έχουν τα αντικείμενα που έχει βαθμολογήσει ο χρήστης αυτός σε σχέση με εκείνα που δεν έχει βαθμολογήσει. Η πρόβλεψη των βαθμολογιών προτίμησης γίνεται σε δύο βήματα. Στο πρώτο βήμα δημιουργείται ένας πίνακας αντικειμένων που περιλαμβάνει την ομοιότητα (*similarity*) μεταξύ όλων των ζευγαριών των αντικειμένων. Αν δύο αντικείμενα τείνουν να έχουν τις ίδιες βαθμολογίες προτίμησης από χρήστες, τότε τα αντικείμενα αυτά έχουν υψηλή ομοιότητα. Στο δεύτερο βήμα με τη χρήση του πίνακα ομοιοτήτων και τις βαθμολογίες προτίμησης του τρέχοντα χρήστη υπολογίζεται ο βαθμός προτίμησης για κάθε νέο αντικείμενο και εξάγονται οι προτάσεις. Τα συστήματα συστάσεων που βασίζονται στα αντικείμενα, σε αντίθεση με εκείνα που βασίζονται στους χρήστες, είναι κατάλληλα για περιπτώσεις που ο αριθμός των αντικειμένων είναι μικρότερος από τον αριθμό των χρηστών.

2.4 Σχετικές Εργασίες - Συστήματα συστάσεων για σύνολα αντικειμένων

Σε αυτή την ενότητα γίνεται αναφορά σε άλλες εργασίες που επιχειρούν να αναπτύξουν συστήματα συστάσεων που προτείνουν σύνολα από αντικείμενα. Όπως έχει γίνει ήδη φανερό, η παρούσα εργασία επιδιώκει τη σχεδίαση και την υλοποίηση ενός συστήματος συστάσεων που προτείνει σύνολα αντικειμένων σε χρήστες χρησιμοποιώντας δεδομένα προτιμήσεων. Για τον λόγο αυτό, η ενότητα αυτή έχει ιδιαίτερη σημασία καθώς επικεντρώνεται σε προηγούμενη έρευνα στην ίδια επιστημονική περιοχή που εντάσσεται και η παρούσα εργασία.

Η ανάπτυξη συστημάτων που εξάγουν σύνολα συστάσεων οδηγείται από την ανάγκη που έχει προκύψει για την αυτοματοποιημένη και στοχευμένη πρόταση συνόλων αντικειμένων από έναν μεγάλο αριθμό πρακτικών εφαρμογών. Με βάση τα χαρακτηριστικά και τις απαιτήσεις των εφαρμογών αυτών, έχουν σχεδιαστεί και έχουν υλοποιηθεί συστήματα που προτείνουν σύνολα αντικειμένων με σκοπό την κάλυψη των εκάστοτε αναγκών. Τα συστήματα συστάσεων που προτείνουν σύνολα αντικειμένων μπορούν να είναι χρήσιμα σε εφαρμογές που προτείνουν ταξιδιωτικά πακέτα, σε εφαρμογές που προτείνουν πακέτα φαγητών για μια εβδομαδιαία διαίτα που το κάθε πακέτο της περιλαμβάνει πρωινό, μεσημεριανό και βραδινό ή σε εφαρμογές που προτείνουν πακέτα μαθημάτων σε σπουδαστές σύμφωνα με τις προτιμήσεις ή τις κλίσεις τους. Αυτά είναι μόνο τρία παραδείγματα αλλά σίγουρα υπάρχουν περισσότερες πρακτικές εφαρμογές που έχουν ανάγκη από τέτοια συστήματα. Γενικότερα, και όπως προκύπτει από την έρευνα που διεξάγεται, οι περισσότερες εφαρμογές που απαιτούν συστήματα συνόλων συστάσεων διακρίνονται σε εφαρμογές σύστασης ταξιδιωτικών πακέτων (Angel et al., 2009; Xie et al., April, 2011; Xie et al., 2013), σε εφαρμογές σύστασης πακέτων προσφερόμενων μαθημάτων (Parameswaran & Garcia-Molina, October, 2009; Courserank, n.d.; Parameswaran et al., 2011) και γενικά σε εφαρμογές που αφορούν στο εμπόριο.

Σύμφωνα με τον Xie, (2015) τα συστήματα συστάσεων για σύνολα αντικειμένων διακρίνονται σε συστήματα ή μοντέλα συστημάτων που έχουν αυστηρούς περιορισμούς και σε συστήματα ή μοντέλα συστημάτων που έχουν ευέλικτους περιορισμούς. Στα συστήματα με αυστηρούς περιορισμούς μπορεί να περιλαμβάνονται συστήματα με σταθερό μέγεθος πακέτου και συστήματα με μεταβλητό ή ευέλικτο μέγεθος πακέτου. Τα συστήματα με αυστηρούς περιορισμούς προτείνουν σύνολα αντικειμένων που υπακούν σε περιορισμούς ή απαιτήσεις που υπάρχουν μεταξύ των αντικειμένων. Για παράδειγμα, όταν προτείνεται ένα ταξιδιωτικό πακέτο πρέπει το ξενοδοχείο να βρίσκεται στην πόλη προορισμού και τα

αξιοθέατα σε κοντινή απόσταση. Αυτό το παράδειγμα δείχνει πως πρέπει να λαμβάνονται υπόψη κάποιοι σημαντικοί περιορισμοί για να έχουν νόημα και πρακτική αξία τα προτεινόμενα ταξιδιωτικά πακέτα. Επίσης, υπάρχουν συστήματα με αυστηρούς περιορισμούς που πληρούν απαιτήσεις προαπαιτούμενων αντικειμένων όπως για παράδειγμα η πρόταση ενός πακέτου μαθημάτων όταν πρέπει να έχουν προηγηθεί άλλα προαπαιτούμενα μαθήματα. Από την άλλη πλευρά, τα συστήματα με ευέλικτους περιορισμούς έχουν μεγαλύτερη ελαστικότητα και είναι περισσότερο προσαρμόσιμα στο πρόβλημα που καλούνται να αντιμετωπίσουν. Να σημειωθεί ότι η παρούσα εργασία επιχειρεί τη σχεδίαση ενός συστήματος που ανήκει στην κατηγορία των συστημάτων με ευέλικτους περιορισμούς.

Η εργασία των Deng et al. (2013) μελετά θέματα που έχουν θεμελιώδη σημασία σχετικά τα συστήματα συστάσεων για σύνολα αντικειμένων. Οι συγγραφείς της εργασίας αυτής μοντελοποιούν το πρόβλημα των συστημάτων συστάσεων ενός συνόλου αντικειμένων και μελετάνε τις περιπτώσεις συστημάτων που προτείνουν ένα σύνολο από σημεία ενδιαφέροντος (*Points Of Interest - POI*), όπως ταξιδιωτικά πακέτα. Επιπλέον, οι συγγραφείς αναγνωρίζουν και μελετάνε τα προβλήματα που ανακύπτουν σε τέτοια συστήματα και προσπαθούν να τα προσεγγίσουν ή και να τα αντιμετωπίσουν. Μερικά από τα προβλήματα αυτά είναι οι αποφάσεις που παίρνονται για το εάν ένα προτεινόμενο πακέτο αποτελεί την καλύτερη επιλογή και για το εάν η βαθμολογία προτίμησης που προκύπτει είναι η μέγιστη δυνατή για την επιλογή των k -καλύτερων πακέτων. Επίσης, κάποια άλλα προβλήματα που εξετάζουν είναι σχετικά με το εάν θα πρέπει να χαλαρώσουν οι περιορισμοί για να προκύψει ένα πακέτο που είναι επιθυμητό στους χρήστες και πώς θα μπορούσε να αλλάξει το όριο του αριθμού των αντικειμένων ενός πακέτου με τέτοιο τρόπο ώστε να ικανοποιούνται οι απαιτήσεις των χρηστών. Οι Deng et al. (2013) στην εργασία τους προσπαθούν να ορίσουν τη συνδυασμένη πολυπλοκότητα αλλά και την πολυπλοκότητα των δεδομένων για όλα τα προαναφερόμενα προβλήματα που θέτουν, αναλύουν και εξετάζουν.

Ακολουθούν περισσότερες λεπτομέρειες για τα συστήματα που έχουν αναπτυχθεί στα πλαίσια ερευνητικών εργασιών και ανήκουν στις κατηγορίες που προαναφέρθηκαν.

2.4.1 Συστήματα με Αυστηρούς Περιορισμούς

Όπως ήδη έχει αναφερθεί, μια κατηγορία συστημάτων συνόλου συστάσεων είναι τα συστήματα με αυστηρούς περιορισμούς. Αυτά τα συστήματα μπορούν να διακριθούν σε συστήματα με σταθερό μέγεθος πακέτου (Xie et al., 2011; Angel et al., 2009; Parameswaran et al., Oktober, 2010; Parameswaran et al., 2011) και σε

συστήματα με ευέλικτο ή μεταβλητό μέγεθος πακέτου (Xie et al., 2010; Xie et al. April, 2011; Lappas et al., 2009). Ακολουθούν οι λεπτομέρειες και τα χαρακτηριστικά τέτοιων συστημάτων που έχουν αναπτυχθεί στα πλαίσια ερευνητικών εργασιών.

Στην εργασία των Xie et al. (2011) περιγράφεται ένα σύστημα συστάσεων για σύνολα αντικειμένων με αυστηρούς περιορισμούς και σταθερό μέγεθος πακέτου που ενσωματώνει την τεχνική των *Rank Join Queries* (Finger & Polyzotis, 2009; Ilyas et al., 2004). Η τεχνική των *Rank Join Queries* περιλαμβάνει σχέσεις (σχεσιακούς πίνακες) που κάθε σχέση έχει μία ή περισσότερες ιδιότητες αξίας (value attributes), έναν αριθμό k και μία συνάρτηση συνάθροισης (*aggregation function*) που συνδυάζει τις ατομικές τιμές αξίας. Το ζητούμενο είναι να βρεθούν τα k -καλύτερα (*top-k*) αποτελέσματα μετά την εκτέλεση μιας συνένωσης (*join*) με τη μεγαλύτερη αξία. Σύμφωνα με τους Xie et al. (2011) μια γενίκευση της τεχνικής των *Rank Join Queries* είναι η αναζήτηση των k -καλύτερων ερωτημάτων (*top-k queries*) που αναζητούν τα k -καλύτερα αντικείμενα ικανοποιώντας τα κριτήρια και τους περιορισμούς. Στην εργασία των Xie et al. (2011) αναπτύσσονται αποδοτικοί πιθανολογικοί και ντετερμινιστικοί αλγόριθμοι που επιλέγουν τα k -καλύτερα αντικείμενα λαμβάνοντας υπόψη τους περιορισμούς, οι οποίοι βασίζονται στο πλαίσιο της τεχνικής των *Rank Join Queries*.

Στην εργασία των Angel et al. (2009) οι συγγραφείς εισάγουν την τεχνική “*Entity Package Finder*”, όπως την ονομάζουν, που αναζητά ερωτήματα, οι απαντήσεις των οποίων οδηγούν στην εξαγωγή των k -καλύτερων συνόλων αντικειμένων τα οποία πληρούν τους περιορισμούς. Για το σκοπό αυτό μοντελοποιούν το πρόβλημα του “*Entity Package Finder*” και αναπτύσσουν έναν αποδοτικό αλγόριθμο που μπορεί να εφαρμοστεί πάνω σε μεγάλο όγκο δεδομένων. Τα κύρια συστατικά του συστήματος αυτού είναι τα έγγραφα (*documents*) με τα δεδομένα, οι οντότητες (*entities*) που δημιουργούνται από τα έγγραφα, οι σχέσεις μεταξύ των εγγράφων (*documents*) και των οντοτήτων και οι συσχετίσεις (*associations*) μεταξύ διαφορετικών τύπων οντοτήτων. Το σύστημα αυτό μπορεί να χρησιμοποιηθεί κυρίως για την πρόταση ταξιδιωτικών πακέτων, καθώς αυτές οι εφαρμογές απαιτούν την εξαγωγή πακέτων με καθορισμένο μέγεθος που τα αντικείμενά τους πρέπει να ικανοποιούν συγκεκριμένους αυστηρούς περιορισμούς.

Στα συστήματα συνόλων συστάσεων με αυστηρούς περιορισμούς και σταθερό μέγεθος πακέτου θα μπορούσαν να περιλαμβάνονται και οι ιδέες των Parameswaran και Garcia-Molina (2009, October) και των Parameswaran et al. (2010, October) που αφορούν στην πρόταση συνόλου αντικειμένων μεγέθους N με τον περιορισμό ότι υπάρχουν προϋποθέσεις ή προαπαιτούμενα αντικείμενα που θα πρέπει να τηρηθούν για να επιλεγεί και να προταθεί ένα σύνολο αντικειμένων. Για παράδειγμα, στην

περίπτωση που το σύστημα προτείνει πακέτα μαθημάτων, δεν μπορεί να προτείνει τη Φυσική ΙΙ αν ο εκπαιδευόμενος δεν έχει διδαχθεί τη Φυσική Ι και ΙΙ, καθώς αυτά τα δύο μαθήματα αποτελούν προαπαιτούμενα μαθήματα για τη Φυσική ΙΙΙ. Να σημειωθεί ότι το ενδιαφέρον για αυτές τις ερευνητικές εργασίες προέκυψε από τη λειτουργία του CourseRank (n.d.), ενός κοινωνικού εργαλείου που αναπτύχθηκε από φοιτητές του Πανεπιστημίου *Stanford* και έχει σκοπό την αξιολόγηση μαθημάτων και την πρόταση κύκλων μαθημάτων σε σπουδαστές. Στα πλαίσια αυτά, οι Parameswaran et al., (2010, October) επεκτείνουν τη δουλειά που έχουν κάνει στην προηγούμενη τους εργασία (Parameswaran, & Garcia-Molina, October, 2009) και προτείνουν έναν αλγόριθμο δυναμικού προγραμματισμού, κάποιους προσεγγιστικούς αλγόριθμους, καθώς και έναν αλγόριθμο που συνδυάζει τα αποτελέσματα των συνόλων αντικειμένων των άλλων αλγόριθμων και εξάγει ένα καινούριο καλύτερο σύνολο συστάσεων. Οι αλγόριθμοι αυτοί αντιμετωπίζουν με αρκετά αποτελεσματικό τρόπο το πρόβλημα της πρότασης ενός συνόλου αντικειμένων, πληρώντας τις προϋποθέσεις και τους περιορισμούς που τίθενται σχετικά με τα προαπαιτούμενα αντικείμενα.

Στην εργασία τους οι Parameswaran et al. (2011) μελετάνε το πρόβλημα της πρότασης συνόλων αντικειμένων όταν τα αντικείμενα που προτείνονται πρέπει να ικανοποιούν περιορισμούς ή απαιτήσεις στα πλαίσια της σύστασης ενός συνόλου μαθημάτων. Να σημειωθεί ότι και αυτή η εργασία υλοποιείται στα πλαίσια του CourseRank (n.d.). Η σύσταση πακέτων με μαθήματα ικανοποιεί περιορισμούς όπως είναι η επιλογή συγκεκριμένου αριθμού μαθημάτων από ένα ευρύτερο σύνολο μαθημάτων ή η επιλογή ενός συγκεκριμένου αριθμού μαθημάτων από διάφορες γνωστικές περιοχές με σκοπό την ολοκλήρωση ενός προγράμματος σπουδών. Το σύστημα που αναπτύσσουν οι Parameswaran et al. (2011), εκτός από το να προτείνει μαθήματα που ικανοποιούν μόνο τους περιορισμούς, στοχεύει και στο να προτείνει ελκυστικά μαθήματα που ικανοποιούν τις προτιμήσεις και τις επιθυμίες των σπουδαστών. Για το σκοπό αυτό οι συγγραφείς αναπτύσσουν ένα μοντέλο για την πρόταση συνόλων μαθημάτων που ικανοποιούν τους περιορισμούς και ταυτόχρονα είναι επιθυμητά από τους σπουδαστές στους οποίους απευθύνονται. Να σημειωθεί ότι στο μοντέλο αυτό οι συγγραφείς εισάγουν τον αλγόριθμο μέγιστης ροής (Max Flow) που η εκτέλεσή του πάνω σε ένα δίκτυο ροής, το οποίο συνδέει τα αντικείμενα (μαθήματα) με τους περιορισμούς, εξάγει το σύνολο των μαθημάτων που ικανοποιούν τους περιορισμούς και έχουν ταυτόχρονα το μέγιστο βαθμό προτίμησης από τους φοιτητές. Με αφορμή αυτή την εργασία των Parameswaran et al. (2011), η παρούσα εργασία χρησιμοποιεί τον αλγόριθμο ροής ελάχιστου κόστους (*minimum cost flow algorithm*) ο οποίος αναζητά το καλύτερο πακέτο αντικειμένων κάθε φορά

που εκτελείται. Περισσότερες λεπτομέρειες για τον αλγόριθμο της παρούσας εργασίας δίνονται στο κεφάλαιο 3.

Στην εργασία των Xie et al. (2010) αναπτύσσεται ένα σύστημα συστάσεων που προτείνει σύνολα αντικειμένων και ανήκει στην κατηγορία των συστημάτων με αυστηρούς περιορισμούς και ευέλικτο μέγεθος πακέτου. Κάθε πρόταση του συστήματος αυτού περιλαμβάνει ένα σύνολο από αντικείμενα όπου κάθε αντικείμενο χαρακτηρίζεται από μία βαθμολογία προτίμησης (*rating*) και ένα κόστος (*cost*). Οι χρήστες σε αυτό το σύστημα προσδιορίζουν την μέγιστη τιμή κόστους (*budget*) για το σύνολο των προτεινόμενων αντικειμένων και το σύστημα εξάγει σύνολα αντικειμένων, που η συνολική τους τιμή κόστους δεν υπερβαίνει τη μέγιστη τιμή κόστους και ταυτόχρονα τα αντικείμενα του πακέτου έχουν αθροιστικά τη μέγιστη δυνατή βαθμολογία προτίμησης. Το σύστημα αυτό χρησιμοποιεί ένα σύστημα συστάσεων που εκτιμά τις βαθμολογίες προτίμησης των αντικειμένων σύμφωνα με τις προτιμήσεις του τρέχοντα χρήστη. Στη συνέχεια χρησιμοποιείται ο αλγόριθμος του σακιδίου (*knapsack algorithm*) ο οποίος εξάγει το καλύτερο δυνατό σύνολο αντικειμένων ικανοποιώντας τον περιορισμό του μέγιστου επιτρεπτού κόστους. Το πρόβλημα που αντιμετωπίζουν οι ερευνητές (Xie et al., 2010) στην εργασία τους είναι ένα NP-Complete πρόβλημα και για αυτό το λόγο επινόησαν προσεγγιστικούς αλγόριθμους που προτείνουν τα k-καλύτερα πακέτα αντικειμένων στους χρήστες. Οι ίδιοι συγγραφείς σε επόμενη τους εργασία (Xie et al., April, 2011) έχουν αναπτύξει ένα σύστημα που προτείνει τα k-καλύτερα πακέτα ταξιδιών χρησιμοποιώντας το μοντέλο που έχουν ήδη αναπτύξει στην προηγούμενη τους εργασία (Xie et al., 2010). Δηλαδή, εφαρμόζουν το θεωρητικό μοντέλο που ανέπτυξαν σε μία πρακτική εφαρμογή που προτείνει πακέτα ταξιδιών, τα οποία ικανοποιούν ένα μέγιστο επιτρεπτό κόστος που έχει τη δυνατότητα να διαθέσει ο τρέχων χρήστης.

Στην εργασία τους οι Lappas et al. (2009, June) προσπαθούν να λύσουν το πρόβλημα της σύστασης μίας ομάδας από ειδικούς (*team of experts*) που όλοι μαζί πρέπει να διαθέτουν ένα σύνολο από προκαθορισμένες δεξιότητες (*skills*). Οι ειδικοί αυτοί συνδέονται και επικοινωνούν μεταξύ τους στα πλαίσια ενός κοινωνικού δικτύου (*social network*). Επίσης, οι ειδικοί ανά δύο μεταξύ τους έχουν ένα κόστος επικοινωνίας που συνδέεται με τις ικανότητές τους. Το χαμηλότερο κόστος επικοινωνίας σημαίνει καλύτερη συνεργασία. Το σύστημα αυτό έχει σκοπό να προτείνει μία ομάδα από ειδικούς που καλύπτουν ένα σύνολο από ικανότητες και ταυτόχρονα έχουν μικρό κόστος επικοινωνίας μεταξύ τους. Να σημειωθεί ότι κάθε ειδικός έχει μια ή περισσότερες ικανότητες και πρέπει η ομάδα των ειδικών που θα προταθεί να καλύπτει το σύνολο των ικανοτήτων που απαιτούνται. Οι συγγραφείς μοντελοποιούν το πρόβλημα αυτό χρησιμοποιώντας ένα γράφημα που οι κόμβοι του

αντιπροσωπεύουν τους ειδικούς και η επικοινωνία μεταξύ των ειδικών αναπαρίσταται με τις ακμές του γραφήματος οι οποίες περιλαμβάνουν το κόστος επικοινωνίας. Για την επίλυση του προβλήματος κατάλληλοι είναι δύο αλγόριθμοι. Ο πρώτος είναι ο αλγόριθμος διαμέτρου (*Diameter*) που αναζητά το μεγαλύτερο σε μήκος μονοπάτι με το μικρότερο κόστος μεταξύ οποιωνδήποτε δύο κόμβων του γραφήματος. Ο δεύτερος είναι ο αλγόριθμος εύρεσης του συνδετικού δέντρου με το ελάχιστο κόστος (*Minimum spanning tree*). Οι συγγραφείς αποδεικνύουν ότι αυτοί οι δύο αλγόριθμοι είναι NP-hard και για να αντιμετωπίσουν το πρόβλημα σχεδιάζουν κατάλληλους προσεγγιστικούς αλγόριθμους που αντιστοιχούν στον αλγόριθμο της διαμέτρου και στον αλγόριθμο του ελάχιστου συνδετικού δέντρου. Σύμφωνα με τους Lappas et al. (2009, June), η συνεισφορά αυτής της εργασίας είναι σημαντική καθώς είναι η πρώτη που μοντελοποιεί το πρόβλημα της σύστασης μιας ομάδας από ειδικούς στα πλαίσια ενός κοινωνικού δικτύου.

2.4.2 Ευέλικτα Συστήματα

Μια ακόμη κατηγορία συστημάτων συστάσεων που προτείνουν πακέτα αντικειμένων είναι τα ευέλικτα συστήματα που δεν υπακούν σε αυστηρούς περιορισμούς ή οι περιορισμοί τους προσαρμόζονται σύμφωνα με το πρόβλημα που έχουν να αντιμετωπίσουν. Τέτοια συστήματα περιγράφονται στις εργασίες των Xie et al. (2013), των Xie et al. (2014) και των Interdonato et al. (2013).

Οι Xie et al. (2013) προτείνουν ένα διαδραστικό με τους χρήστες σύστημα συστάσεων που προτείνει ταξιδιωτικά πακέτα. Το σύστημα αυτό προτείνει διάφορα πακέτα ταξιδιού που εμφανίζονται πάνω σε έναν χάρτη και εμπλέκει τους χρήστες στη διαδικασία του προσδιορισμού του κατάλληλου πακέτου μέσα από μία απλή αλληλεπίδραση με το σύστημα. Σε αυτή την αλληλεπίδραση ο χρήστης συγκρίνει πακέτα και σβήνει ή προσθέτει προορισμούς σε ένα πακέτο. Το σύστημα αυτό περιλαμβάνει δύο στάδια λειτουργίας. Στο πρώτο στάδιο ο χρήστης επιλέγει τις αρχικές προτιμήσεις με σκοπό το σύστημα να του προτείνει τα αρχικά πακέτα ταξιδιών και στο δεύτερο στάδιο ο χρήστης αλληλεπιδρά με το σύστημα. Κατά το στάδιο της επιλογής των αρχικών προτιμήσεων ο χρήστης επιλέγει τον προορισμό (*destination*), το είδος των σημείων ενδιαφέροντος και τον βαθμό ενδιαφέροντος για το κόστος (*budget*) ή τη βαθμολογία προτίμησης (*rating*). Το σύστημα εμφανίζει πάνω σε έναν χάρτη τα πακέτα ταξιδιού και αλληλεπιδρά με τον χρήστη ο οποίος μπορεί να προσθέσει ή να διαγράψει σημεία ενδιαφέροντος στα πακέτα των προτάσεων. Μέσα από αυτή την αλληλεπίδραση το σύστημα τείνει να προτείνει τα καλύτερα για τον χρήστη ταξιδιωτικά πακέτα.

Οι Xie et al. (2013) προχωρούν ένα βήμα παραπέρα και στην επόμενη τους εργασία (Xie et al., 2014) προτείνουν ένα μοντέλο συστήματος που προτείνει τα k-καλύτερα πακέτα αντικειμένων μέσα από διαδικασίες εκμαίευσης των προτιμήσεων των χρηστών. Οι συγγραφείς της εργασίας αυτής (Xie et al., 2014) εμπνεύστηκαν για το σύστημα της εκμαίευσης των προτιμήσεων των χρηστών από τις εργασίες των Chajewska et al. (2000, July) και του Boutilier (2002, July), οι οποίοι μελετάνε το πρόβλημα της εκμαίευσης των προτιμήσεων. Οι Xie et al. (2014) χρησιμοποιούν την τεχνική της συνάρτησης χρησιμότητας (*utility function*) που συνδυάζει την ποιότητα (*rating*) και το κόστος (*cost*) των αντικειμένων ενός πακέτου, με σκοπό το σύστημα να προτείνει πακέτα με την υψηλότερη ποιότητα αντικειμένων και το χαμηλότερο κόστος. Η συνάρτηση χρησιμότητας είναι μία γραμμική συνάρτηση που το σύστημα προσπαθεί να την προσαρμόσει ανάλογα με τα χαρακτηριστικά των χρηστών χωρίς να ζητάει από τους χρήστες να δώσουν τις παραμέτρους. Το σύστημα εκπαιδεύεται και προσαρμόζει τη συνάρτηση χρησιμότητας ανάλογα με την ανατροφοδότηση από τους χρήστες. Για το σκοπό αυτό οι συγγραφείς αναπτύσσουν έναν αποδοτικό αλγόριθμο που παράγει τα k-καλύτερα πακέτα αντικειμένων. Πιο συγκεκριμένα αναπτύσσεται ένα απλό σύστημα σύγκρισης πακέτων με έναν μηχανισμό ερωτημάτων εκμαίευσης των προτιμήσεων των χρηστών. Αρχικά, στον τρέχοντα χρήστη προτείνεται ένα σύνολο από πακέτα που περιλαμβάνουν τα καλύτερα πακέτα σύμφωνα με την τρέχουσα συνάρτηση χρησιμότητας για τις προτιμήσεις του χρήστη και ένα σύνολο από τυχαία πακέτα για την εξακρίβωση των περαιτέρω προτιμήσεών του. Ο χρήστης επιλέγει τα πακέτα που του ταιριάζουν και το σύστημα εκπαιδεύεται προσαρμόζοντας τη συνάρτηση χρησιμότητας, με σκοπό να βελτιώσει τις προτάσεις και να προτείνει τα k-καλύτερα πακέτα αντικειμένων.

Στην εργασία τους οι Interdonato et al. (2013, November) αναπτύσσουν ένα ευέλικτο σύστημα συστάσεων που προτείνει πακέτα αντικειμένων. Σε αντίθεση με άλλες εργασίες, οι Interdonato et al. (2013) αναπτύσσουν ένα σύστημα που είναι ευέλικτο και δεν αναφέρεται σε κάποια συγκεκριμένη εφαρμογή αλλά προσαρμόζεται στις εκάστοτε απαιτήσεις των χρηστών. Το σύστημά τους βασίζεται στο περιεχόμενο και προτείνει πακέτα αντικειμένων σύμφωνα με τους περιορισμούς που θέτουν οι χρήστες. Ο αλγόριθμος εκπαιδεύεται από την προηγούμενη γνώση των προτάσεων που έχουν γίνει και από δεδομένα που παρέχονται από τους χρήστες. Ο τρέχων χρήστης εκτός από περιορισμούς μπορεί να θέσει και ένα μέγιστο κόστος (*budget*) πακέτου. Τα πακέτα που εξάγονται κατατάσσονται ανάλογα με το στόχο του τρέχοντα χρήστη, δηλαδή κατατάσσονται βάσει του κατά πόσο πληρούνται οι περιορισμοί και το μέγιστο κόστος που έχει θέσει ο χρήστης αλλά και από τη χρησιμοποίηση της προηγούμενης γνώσης. Το σύστημα αυτό βασίζεται στο

περιεχόμενο, αυτό το κάνει να διαφοροποιείται από την προσέγγιση της παρούσας εργασίας που προτείνει αντικείμενα χρησιμοποιώντας συνεργατικό φιλτράρισμα για την εκτίμηση των προτιμήσεων των χρηστών. Ωστόσο, παραμένει μια ενδιαφέρουσα δουλειά που εντάσσεται στην κατηγορία των ευέλικτων συστημάτων συστάσεων.

Στον πίνακα 2.4 συνοψίζονται όλες οι σχετικές επιστημονικές εργασίες που μελετήθηκαν και περιγράφηκαν στην ενότητα 2.4.

A/A	Επιστημονική εργασία	Επιστημονικό υπόβαθρο ή Βασική ιδέα	Περιγραφή
1	Lappas et al. (2009).	Αλγόριθμοι της διαμέτρου και του ελάχιστου συνδετικού δέντρου	Μοντελοποιείται το πρόβλημα της σύστασης μιας ομάδας από ειδικούς που πρέπει να καλύπτουν ένα σύνολο από δεξιότητες και προτείνονται προσεγγιστικοί αλγόριθμοι που λύνουν το πρόβλημα.
2	Angel et al. (2009)	<i>Entity Package Finder</i>	Σύσταση των k-καλύτερων πακέτων που ικανοποιούν τους περιορισμούς με την αναγνώριση των k-καλύτερων πακέτων που ταιριάζουν με ένα δοσμένο σύνολο απαιτήσεων. Κατάλληλο για προτάσεις ταξιδιωτικών πακέτων.
3	Parameswaran & Garcia-Molina (2009)	Ανάπτυξη αποδοτικών αλγόριθμων στο πλαίσιο του <i>CourseRank</i>	Σύσταση συνόλων αντικειμένων με τήρηση προϋποθέσεων ή περιορισμών που αφορούν σε προαπαιτούμενα αντικείμενα. Κατάλληλο για σύσταση μαθημάτων.
4	Parameswaran et al., (2010)	Επέκταση της ιδέας των Parameswaran & Garcia-Molina (2009)	Ανάπτυξη αλγόριθμου δυναμικού προγραμματισμού, κάποιοι προσεγγιστικοί αλγόριθμοι και ένας αλγόριθμος που συνδυάζει τα αποτελέσματα των υπολοίπων για την πρόταση ενός συνόλου αντικειμένων που πληρούν προϋποθέσεις και περιορισμούς προαπαιτούμενων αντικειμένων.
5	Parameswaran et al. (2011)	Ο αλγόριθμος της μέγιστης ροής.	Ανάπτυξη μοντέλου για τη σύσταση συνόλων μαθημάτων που ικανοποιούν τους περιορισμούς ενός προγράμματος σπουδών και ταυτόχρονα είναι επιθυμητά στους σπουδαστές.

6	Xie et al. (2010)	Ο αλγόριθμος του σακιδίου.	Προτείνονται προσεγγιστικοί αλγόριθμοι για την πρόταση των N-καλύτερων πακέτων αντικειμένων τα οποία πληρούν περιορισμούς και έχουν ένα ελάχιστο κόστος (<i>budget</i>).
7	Xie et al. (2011, April)	Βασίζεται στην ιδέα των Xie et al., 2010.	Ανάπτυξη συστήματος για την πρόταση των k-καλύτερων ταξιδιωτικών πακέτων που ικανοποιούν ένα μέγιστο διαθέσιμο κόστος.
8	Xie et al., (2011)	<i>Rank Join Queries</i> .	Ανάπτυξη αλγόριθμων για την αναζήτηση των k-καλύτερων ερωτημάτων που αναζητούν τα k-καλύτερα πακέτα αντικειμένων τα οποία ικανοποιούν τους περιορισμούς.
9	Xie et al. (2013)	Κατάταξη πακέτων βασισμένη σε πολλαπλά κριτήρια.	Προτείνεται ένα διαδραστικό σύστημα συστάσεων για ταξιδιωτικά πακέτα το οποίο εμφανίζει τα σημεία ενδιαφέροντος πάνω σε έναν χάρτη και μετά από αλληλεπίδραση με τους χρήστες εξάγονται τα καλύτερα.
10	Deng et al. (2013)	Μοντελοποίηση του προβλήματος των συστάσεων συνόλων αντικειμένων.	Μελέτη θεμάτων θεμελιώδους σημασίας σχετικά τα συστήματα συστάσεων για σύνολα αντικειμένων και ιδιαίτερα για σύνολα με σημεία ενδιαφέροντος όπως ταξιδιωτικά πακέτα.
11	Interdonato et al. (2013)	Μεθοδολογία για το <i>package ranking</i> .	Προτείνεται ένα ευέλικτο μοντέλο που εξάγει τα καλύτερα πακέτα αντικειμένων. Τα μοντέλο αυτό βασίζεται σε έναν αλγόριθμο που εκπαιδεύεται από τις προηγούμενες προτάσεις. Η προσέγγιση τους είναι γενική και δεν απευθύνεται σε συγκεκριμένη εφαρμογή.
12	Xie et al. (2014)	Βασίζεται στην ιδέα των Xie et al., 2013 και στις τεχνικές <i>εκμαίευσης των προτιμήσεων</i> .	Ανάπτυξη μοντέλου σύγκρισης πακέτων με έναν μηχανισμό εκμαίευσης των προτιμήσεων. Το σύστημα εκπαιδεύεται χρησιμοποιώντας μια συνάρτηση χρησιμότητας που συνδυάζει την ποιότητα και το κόστος των αντικειμένων.

Πίνακας 2.4: Οι σχετικές επιστημονικές εργασίες που μελετήθηκαν.

2.5 Μετρικές Αξιολόγησης Συστημάτων Συστάσεων

Σε αυτή την ενότητα αναφέρονται οι σημαντικότεροι δείκτες και οι μετρικές που χρησιμοποιούνται για την αξιολόγηση των συστημάτων συστάσεων. Οι μετρικές που περιγράφονται παρακάτω και χρησιμοποιούνται στην αξιολόγηση της παρούσας εργασίας είναι το απόλυτο σφάλμα, το μέσο τετραγωνικό σφάλμα, ο δείκτης *Precision* και ο δείκτης *Recall*. Ακολουθούν οι μαθηματικές σχέσεις που δίνουν την τιμή των μετρικών που προαναφέρθηκαν.

Το μέσο απόλυτο σφάλμα δίνεται από τον τύπο (Witten & Frank, 2005):

$$MAE = \frac{1}{n} \sum_{i=1}^n |X_i - Y_i|$$

Όπου X_i : η εκτιμώμενη τιμή και Y_i : η πραγματική τιμή.

Το μέσο τετραγωνικό σφάλμα δίνεται από τον τύπο (Witten & Frank, 2005):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_i - Y_i)^2}{n}}$$

Όπου X_i : η εκτιμώμενη τιμή και Y_i : η πραγματική τιμή.

Ο δείκτης *Precision* δίνεται από τον τύπο (Olson & Delen, 2008):

$$Precision = \frac{TP}{TP + FP}$$

Όπου *TP* (*True Positive*): ο αριθμός των σωστά ανακτημένων αντικειμένων,

FP (*False Positive*): ο αριθμός των λανθασμένα ανακτημένων αντικειμένων.

Ο δείκτης *Recall* δίνεται από τον τύπο (Olson & Delen, 2008):

$$Recall = \frac{TP}{TP + FN}$$

Όπου *FN* (*False Negative*): ο αριθμός των λανθασμένα μη ανακτημένων αντικειμένων.

Για την περιγραφή του δείγματος των στατιστικών μετρήσεων και την εξαγωγή χρήσιμων συμπερασμάτων για ολόκληρο τον πληθυσμό χρησιμοποιείται η τυπική απόκλιση και τα διαστήματα εμπιστοσύνης. Ακολουθεί η περιγραφή της τυπικής απόκλισης του τυχαίου δείγματος και του διαστήματος εμπιστοσύνης της μέσης τιμής του πληθυσμού με άγνωστη τη διασπορά του πληθυσμού.

Η τυπική απόκλιση του δείγματος δίνεται από τη σχέση (Κοκολάκης & Σπηλιώτης, 2010, σελ. 227):

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n - 1}}$$

Όπου x_i : η τιμή του i -οστού στοιχείου του δείγματος

$\bar{X}$: η μέση τιμή του δείγματος

n : το μέγεθος του δείγματος

Σύμφωνα με τους Κοκολάκη και Σπηλιώτη (2010, σελ. 292-293) το διάστημα εμπιστοσύνης (*confidence interval*) για τη μέση τιμή του πληθυσμού (μ) με βαθμό εμπιστοσύνης (*confidence level*) $\gamma=1-\alpha$ και άγνωστη διασπορά είναι:

$$\bar{X} - t_{n-1, \frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{n-1, \frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$$

Όπου $t_{n-1, \frac{\alpha}{2}}$: η τιμή της μεταβλητή της κατανομής t του Student,

$\bar{X}$: η μέση τιμή του δείγματος,

S : η τυπική απόκλιση του δείγματος και

n : το μέγεθος του δείγματος

Για $n \gg 1$, επειδή η κατανομή t του Student προσεγγίζει την κανονική κατανομή, μπορεί να χρησιμοποιηθεί η προσέγγιση (Κοκολάκης & Σπηλιώτης, 2010, σελ. 293):

$$\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$$

Όπου $z_{\frac{\alpha}{2}}$: η τιμή της μεταβλητής Z της κανονικής κατανομής για την

οποία η πιθανότητα της μεταβλητής Z να είναι

μεγαλύτερη από την τιμή $z_{\frac{\alpha}{2}}$ είναι: $P(Z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$.

2.6 Απαιτούμενες Τεχνολογίες και Συστήματα Λογισμικού

Σε αυτή την ενότητα παρουσιάζονται οι τεχνολογίες και τα συστήματα λογισμικού που χρησιμοποιούνται στην παρούσα εργασία για την ανάπτυξη ενός ολοκληρωμένου συστήματος συστάσεων. Η ανάπτυξη του λογισμικού ξεκίνησε μετά από μία μελέτη που έγινε για τον καθορισμό και την καταγραφή των απαιτήσεων που πρέπει να καλύπτει το προς υλοποίηση λογισμικό. Μετά από αυτήν τη μελέτη και την καταγραφή ξεκίνησε μία αναζήτηση για τον προσδιορισμό των κατάλληλων εργαλείων που θα μπορούσαν να καλύψουν τις ανάγκες και τις απαιτήσεις της εφαρμογής. Ένα από τα πιο βασικά ζητήματα για το οποίο θα έπρεπε να βρεθεί λύση ήταν η επιλογή της κατάλληλης γλώσσας προγραμματισμού και του ολοκληρωμένου περιβάλλοντος ανάπτυξης λογισμικού. Με μεθοδική έρευνα των δυνατοτήτων και των χαρακτηριστικών των διάφορων εργαλείων ανάπτυξης λογισμικού προέκυψε η τελική επιλογή τους.

Για την ανάπτυξη του λογισμικού επιλέχτηκε η γλώσσα προγραμματισμού *java* και το ολοκληρωμένο περιβάλλον ανάπτυξης λογισμικού *NetBeans*. Η *java* επιλέχτηκε καθώς είναι μια συνεχώς αναπτυσσόμενη γλώσσα, καλά υποστηριζόμενη και ευρέως χρησιμοποιούμενη. Υπάρχει ένας αρκετά μεγάλος αριθμός βιβλιοθηκών λογισμικού που την υποστηρίζουν επαρκώς και αυτό βοηθάει στη σχετικά εύκολη και αποτελεσματική αντιμετώπιση τυχόν προγραμματιστικών προβλημάτων που ανακύπτουν. Το *NetBeans* επιλέχτηκε καθώς είναι και αυτό ένα ευρέως χρησιμοποιούμενο και δημοφιλές εργαλείο που παρέχει ευκολία στη συγγραφή, στη διόρθωση και στη δοκιμή του κώδικα. Επίσης, παρέχει τη δυνατότητα δημιουργίας γραφικού περιβάλλοντος χρήστη με σχετικά απλό τρόπο, κάτι που είναι πολύ χρήσιμο στην παρούσα εργασία.

Πιο συγκεκριμένα η γλώσσα προγραμματισμού, τα εργαλεία ανάπτυξης, οι προγραμματιστικές τεχνολογίες, τα υποσυστήματα, οι χρησιμοποιούμενες βιβλιοθήκες και οι εκδόσεις τους αναφέρονται παρακάτω.

Γλώσσα προγραμματισμού: Java

Πακέτο ανάπτυξης (Java Development Kit): JDK 1.8.0_66

Ολοκληρωμένο περιβάλλον ανάπτυξης: NetBeans IDE 8.1

Το σύστημα επίλυσης προβλημάτων γραμμικού προγραμματισμού: IBM ILOG CPLEX Optimization Studio v.12.2

Η τεχνολογία της java για την ανάπτυξη του γραφικού περιβάλλοντος: Swing

Οι βιβλιοθήκες της java:

Apache Mahout 0.11.1: Για την υλοποίηση του συνεργατικού φιλτραρίσματος.

cplex-12.2: Για την επίλυση των προβλημάτων γραμμικού προγραμματισμού.

Στην ενότητα αυτή αναφέρθηκαν οι τεχνολογίες που χρησιμοποιούνται στην παρούσα εργασία για την υλοποίηση του ολοκληρωμένου συστήματος συστάσεων. Στα κεφάλαια 4 και 5 δίνονται όλες οι λεπτομέρειες για τη σχεδίαση και την τελική υλοποίηση του λογισμικού που αναπτύχθηκε.

Κεφάλαιο 3: Σχεδίαση Αλγόριθμου

3.1 Γενικά

Σε αυτό το κεφάλαιο καθορίζεται το πρόβλημα συστάσεων που καλείται να αντιμετωπίσει η παρούσα εργασία και περιγράφεται ο αλγόριθμος που αναπτύσσεται με σκοπό την αντιμετώπιση αυτού του προβλήματος. Ο καθορισμός του προβλήματος αφορά στην περιγραφή και στην οριοθέτηση των απαιτήσεων και των προδιαγραφών βάσει των οποίων πρέπει να βρεθεί κάποια αποτελεσματική λύση. Η λύση αυτή δίνεται από τον αλγόριθμο που αναπτύσσεται και αφορά στην ανάπτυξη ενός μοντέλου αποδοτικής αλλά και αποτελεσματικής αντιμετώπισης του προβλήματος συστάσεων που τίθεται. Ο σκοπός της σχεδίασης του αλγόριθμου είναι η μοντελοποίηση και η ανάπτυξη ενός ολοκληρωμένου συστήματος συστάσεων που συνθέτει σύνολα αντικειμένων και ταυτόχρονα λειτουργεί με αποδοτικό και αποτελεσματικό τρόπο πετυχαίνοντας υψηλά ποσοστά ακρίβειας στα αποτελέσματά του.

3.2 Καθορισμός Προβλήματος

Όπως είναι ήδη γνωστό, αυτή η εργασία έχει σκοπό την ανάπτυξη ενός αλγόριθμου για τη λειτουργία ενός συστήματος συστάσεων το οποίο προτείνει σύνολα αντικειμένων, ωστόσο δεν είναι ακόμη ξεκάθαρο το πρόβλημα που καλείται να αντιμετωπίσει αυτός ο αλγόριθμος. Σε αυτή την ενότητα θα καθοριστεί το πρόβλημα σύμφωνα με το οποίο θα σχεδιαστεί ο αλγόριθμος ο οποίος δίνει τη λύση.

Για τον καθορισμό του προβλήματος θεωρείται ότι δίνεται ένα σύνολο από βαθμολογίες προτίμησης χρηστών (*ratings*) που αφορούν σε συγκεκριμένα αντικείμενα (*items*). Τα αντικείμενα αυτά είναι γνωστά και το καθένα ανήκει σε μία ή περισσότερες κατηγορίες (*categories*).

Το ζητούμενο του προβλήματος είναι η σύσταση των N-καλύτερων πακέτων αντικειμένων για έναν δεδομένο χρήστη λαμβάνοντας υπόψη τα διαθέσιμα δεδομένα προτιμήσεων.

Η επίλυση αυτού του προβλήματος ακολουθεί κάποια βασικά βήματα. Για την πρόταση ενός πακέτου (*package*) από αντικείμενα σε κάποιον χρήστη πρέπει αρχικά να γίνει εκτίμηση της βαθμολογίας προτίμησης που θα έβαζε ο τρέχων χρήστης σε κάθε νέο για αυτόν αντικείμενο. Στη συνέχεια, με τις βαθμολογίες προτίμησης του

κάθε αντικείμενου να είναι δεδομένες, θα πρέπει να γίνει μια διμερής αντιστοίχιση μεταξύ των αντικειμένων και των κατηγοριών, με τέτοιο τρόπο, ώστε να προκύψουν τα ζητούμενα αντικείμενα ανά κατηγορία και ταυτόχρονα το προτεινόμενο πακέτο αντικειμένων να έχει τη βέλτιστη βαθμολογία προτίμησης. Αυτή η διαδικασία που είναι η ζητούμενη για τον αλγόριθμο της παρούσας εργασίας μπορεί να γενικευτεί στην εύρεση και στην πρόταση των N-καλύτερων πακέτων αντικειμένων για τον εκάστοτε χρήστη.

Σύμφωνα με τα προηγούμενα, τα δεδομένα του προβλήματος είναι τα εξής:

Ένα σύνολο από χρήστες:

$$U = \{u_1, u_2, u_3, \dots, u_m\}, \text{ όπου } u_x: \text{ ο χρήστης } x.$$

Ένα σύνολο από αντικείμενα:

$$I = \{i_1, i_2, i_3, \dots, i_n\}, \text{ όπου } i_x: \text{ το αντικείμενο } x.$$

Ένα σύνολο από κατηγορίες στις οποίες ανήκουν τα αντικείμενα:

$$C = \{c_1, c_2, c_3, \dots, c_t\}, \text{ όπου } c_x: \text{ η κατηγορία } x.$$

Το σύνολο των κατηγοριών στις οποίες ανήκει κάθε αντικείμενο:

$$C_{i_x} = \{c_p, c_q, c_r\}, \text{ όπου το αντικείμενο } x \text{ ανήκει στις κατηγορίες } p, q \text{ και } r.$$

Ένα σύνολο με τις βαθμολογίες προτίμησης των χρηστών:

$$R = \{r_{u_x, i_y}, r_{u_z, i_k}, \dots\}, \text{ όπου } r_{u_x, i_y}: \text{ η βαθμολογία του χρήστη } u_x \text{ στο αντικείμενο } i_y.$$

Από το σύνολο των βαθμολογιών προτίμησης R εξάγεται το σύνολο με τις βαθμολογίες προτίμησης κάθε χρήση:

$$R_{u_x} = \{r_{u_x, i_y}, r_{u_x, i_z}, r_{u_x, i_t}\}, \text{ όπου ο χρήστης } x \text{ έχει δώσει βαθμολογίες προτίμησης για τα αντικείμενα } y, z \text{ και } t.$$

Ο αριθμός των προτεινόμενων πακέτων:

$$N$$

Ο αριθμός των αντικειμένων που περιλαμβάνει κάθε προτεινόμενο πακέτο:

$$n$$

Ο αριθμός των κατηγοριών που περιλαμβάνει κάθε προτεινόμενο πακέτο:

$$n_c$$

Ο αριθμός των αντικειμένων ανά κατηγορία:

$$n_{i/C}$$

Ο μέγιστος αριθμός εμφανίσεων κάθε αντικειμένου στα προτεινόμενα πακέτα:

$$n_i$$

Το ποσοστό της δημοφιλίας του κάθε αντικειμένου:

$$Popularity(x) = \frac{\text{number of ratings of item } x}{\text{number of users}}$$

Όπου ο βαθμός δημοφιλίας του αντικειμένου x δίνεται από το λόγο του αριθμού των βαθμολογήσεων από ξεχωριστούς χρήστες για το αντικείμενο x προς τον συνολικό αριθμό των χρηστών.

Ο συντελεστής δημοφιλίας του κάθε αντικειμένου:

$PopularityFactor \in \{0, 1, 2, \dots, 10\}$, όπου ο συντελεστής δημοφιλίας πολλαπλασιάζεται με το ποσοστό της δημοφιλίας και ο αριθμός που προκύπτει προστίθεται στον εκτιμώμενο βαθμό προτίμησης του κάθε αντικειμένου για να προκύψει ένας νέος βαθμός προτίμησης για τα αντικείμενα. Ο τύπος που δίνει το νέο βαθμό προτίμησης για το αντικείμενο x είναι:

$$Rating_{New}(x) = Rating(x) + PopularityFactor \cdot Popularity(x)$$

Όπου $Rating(x)$: η εκτιμώμενη βαθμολογία προτίμησης του αντικειμένου x .

Ο νέος αυτός βαθμός προτίμησης δίνει προτεραιότητα στα δημοφιλέστερα αντικείμενα σύμφωνα με την τιμή του συντελεστή δημοφιλίας που μπορεί να πάρει τιμές από μηδέν έως δέκα. Η τιμή μηδέν σημαίνει ότι δεν λαμβάνεται υπόψη η δημοφιλία των αντικειμένων, ενώ η τιμή 10 σημαίνει ότι δίνεται μεγάλη προτεραιότητα στα δημοφιλέστερα αντικείμενα. Να σημειωθεί ότι η νέα βαθμολογία προτίμησης κανονικοποιείται στην κλίμακα της εκάστοτε βαθμολογίας προτίμησης.

Το ζητούμενο του προβλήματος είναι να προταθούν τα καλύτερα σύνολα αντικειμένων που ικανοποιούν τις ακόλουθες προτάσεις και περιορισμούς:

Ο αριθμός των κατηγοριών (n_c) από τις οποίες πρόκειται να προταθούν αντικείμενα είναι προκαθορισμένος και σταθερός.

$$n_c \leq |C|$$

Ο αριθμός των αντικειμένων ανά κατηγορία ($n_{i/c}$) είναι προκαθορισμένος και σταθερός. Αυτός ο αριθμός μπορεί να είναι ο ίδιος για όλες τις κατηγορίες ή κάθε κατηγορία να έχει τον δικό της αριθμό αντικειμένων.

$$\begin{aligned} n_{i/c_1} &= n_{i/c_2} = n_{i/c_3} = \dots = n_{i/c} \\ &\quad \text{ή} \\ n_{i/c_1}, &\quad n_{i/c_2}, &\quad n_{i/c_3}, &\quad \dots \end{aligned}$$

Ο αριθμός των αντικειμένων ανά πακέτο (n) είναι προκαθορισμένος και σταθερός.

$$\begin{aligned} n &= n_{i/c} \cdot n_c \\ &\quad \text{ή} \\ n &= \sum_{j=1}^m n_{i/c_j} \end{aligned}$$

Εναλλακτικά μπορεί να δοθεί και ένας μέγιστος αριθμός αντικειμένων ανά πακέτο (n') που είναι μικρότερος από τον προαναφερόμενο αριθμό n ($n' < n$). Αυτό είναι χρήσιμο στην περίπτωση της σύστασης πακέτων που περιλαμβάνουν ένα μικρότερο σύνολο αντικειμένων από το προκαθορισμένο. Αυτό το μικρότερο σύνολο αντικειμένων περιλαμβάνει τα καλύτερα αντικείμενα από αυτά που θα προτεινόταν αν δεν είχε ληφθεί υπόψη ο μέγιστος αριθμός αντικειμένων ανά πακέτο (n').

Ο αριθμός των πακέτων (N) που πρόκειται να προταθούν είναι επίσης προκαθορισμένος και σταθερός.

$$N \leq \frac{|I|}{n}$$

Ο μέγιστος αριθμός εμφανίσεων του κάθε αντικειμένου (n_i) στα προτεινόμενα πακέτα είναι προκαθορισμένος και σταθερός. Αυτός ο αριθμός μπορεί να είναι ο ίδιος για όλα τα αντικείμενα ή κάθε αντικείμενο να έχει τον δικό του μέγιστο αριθμό εμφανίσεων.

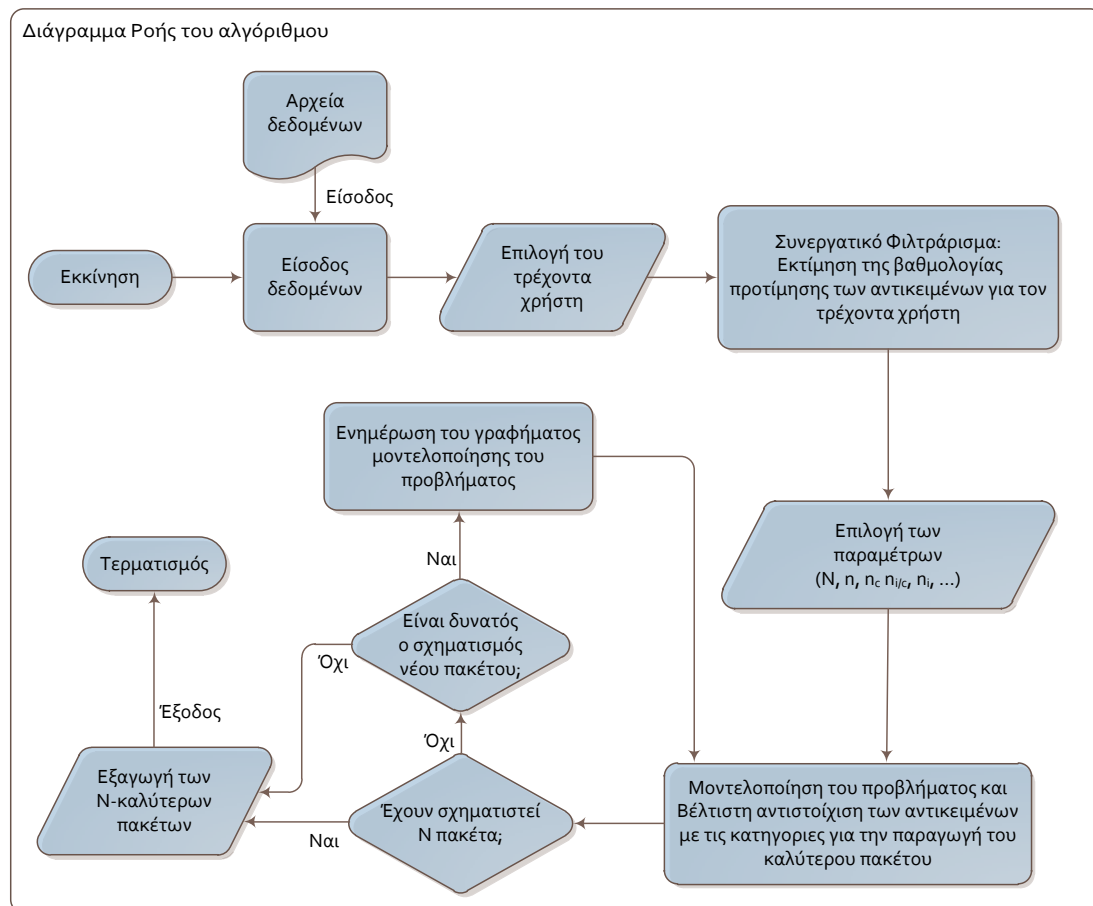
$$\begin{array}{c} n_{i_1} = n_{i_2} = n_{i_3} = \dots = n_i \\ \text{ή} \\ n_{i_1}, \quad n_{i_2}, \quad n_{i_3}, \quad \dots \end{array}$$

Τα προτεινόμενα σύνολα πακέτων σχηματίζονται διατεταγμένα σε φθίνουσα ταξινόμηση σύμφωνα με τον εκτιμώμενο βαθμό προτίμησης των αντικειμένων που περιλαμβάνουν.

Σύμφωνα με τα προηγούμενα, το πρόβλημα που τίθεται απαιτεί τη σχεδίαση ενός αλγόριθμου που δέχεται στην είσοδο του όλα τα απαραίτητα δεδομένα και μετά από κατάλληλη διεργασία προτείνονται τα N -καλύτερα πακέτα αντικειμένων σε έναν δεδομένο χρήστη.

3.3 Μοντελοποίηση Αλγόριθμου

Η μοντελοποίηση του αλγόριθμου περιλαμβάνει όλα τα στάδια που είναι απαραίτητα για τη λειτουργία ενός ολοκληρωμένου συστήματος συστάσεων που εξάγει σύνολα αντικειμένων. Ο αλγόριθμος λαμβάνει ως είσοδο τα δεδομένα που περιγράφονται στην προηγούμενη ενότητα και για κάποιο συγκεκριμένο χρήστη εξάγει στην έξοδο του τα N-καλύτερα πακέτα αντικειμένων. Τα προτεινόμενα πακέτα αντικειμένων παρατάσσονται σε φθίνουσα ταξινόμηση σύμφωνα με τον εκτιμώμενο βαθμό προτίμησης για κάθε πακέτο. Το διάγραμμα ροής με τα στάδια του αλγόριθμου παρουσιάζεται στην εικόνα 3.3.



Εικόνα 3.3: Διάγραμμα ροής με τα στάδια λειτουργίας του αλγόριθμου.

Όπως είναι φανερό από το διάγραμμα της εικόνας 3.3, το σύστημα περιλαμβάνει έναν αριθμό από στάδια λειτουργίας που κάθε στάδιο μπορεί να ενσωματώνει έναν επιμέρους αλγόριθμο. Στο υπόλοιπο μέρος αυτής της εργασίας θεωρούνται δεδομένοι οι όροι και ο συμβολισμός που εισήχθησαν στην ενότητα 3.2. Ακολουθούν τα χαρακτηριστικά και ο τρόπος λειτουργίας του κάθε επιμέρους σταδίου λειτουργίας.

3.3.1 Είσοδος Δεδομένων

Το αρχικό στάδιο αφορά στην είσοδο και στην επεξεργασία των δεδομένων. Τα δεδομένα είναι αποθηκευμένα σε αρχεία, το σύστημα διαβάζει τα αρχεία αυτά και διατηρεί τα δεδομένα σε κατάλληλες δομές για την μετέπειτα αξιοποίησή τους. Τα δεδομένα εισόδου αφορούν τα αντικείμενα, τους χρήστες και τις βαθμολογίες προτίμησης που έχει δώσει κάθε χρήστης ή τις βαθμολογίες προτίμησης που έχει δεχτεί κάθε αντικείμενο από τους χρήστες.

3.3.2 Επιλογή του Τρέχοντα Χρήστη

Το δεύτερο στάδιο αφορά στην επιλογή του χρήστη στον οποίο θα προταθούν τα N-καλύτερα πακέτα αντικειμένων. Τα δεδομένα εισόδου περιλαμβάνουν ένα σύνολο από χρήστες που έχουν βαθμολογήσει ένα σύνολο από αντικείμενα. Η επιλογή του τρέχοντα χρήστη γίνεται μεταξύ των χρηστών εκείνων που περιλαμβάνονται στα δεδομένα εισόδου. Δεν έχουμε λοιπόν την περίπτωση προτάσεων σε άγνωστους χρήστες, σε χρήστες δηλαδή που δεν περιλαμβάνονται στα δεδομένα εισόδου.

3.3.3 Εκτίμηση της Βαθμολογίας Προτίμησης των Αντικειμένων

Το τρίτο στάδιο αφορά στην εκτίμηση της βαθμολογίας προτίμησης (*rating*) των νέων αντικειμένων για τον τρέχοντα χρήστη. Για την πρόβλεψη της βαθμολογίας προτίμησης των αντικειμένων χρησιμοποιείται η τεχνική του συνεργατικού φιλτραρίσματος. Στα πλαίσια αυτής της εργασίας δεν αναπτύσσεται κάποιο σύστημα που να προβλέπει τη βαθμολογία προτίμησης των αντικειμένων αλλά χρησιμοποιείται ένα από τα ήδη ανεπτυγμένα τέτοια συστήματα, καθώς η τεχνολογία σε αυτό τον τομέα είναι αρκετά ώριμη και έχουν αναπτυχθεί αξιόπιστα εργαλεία.

3.3.4 Επιλογή των Παραμέτρων

Το τέταρτο στάδιο αφορά την επιλογή των παραμέτρων που χαρακτηρίζουν και καθορίζουν τη λειτουργία του αλγόριθμου, το είδος και τα χαρακτηριστικά των προτεινόμενων πακέτων. Οι παράμετροι αυτοί είναι οι εξής:

Ο αριθμός των καλύτερων πακέτων αντικειμένων που πρόκειται να προταθούν:

N

Ο αριθμός των κατηγοριών των αντικειμένων που περιλαμβάνονται σε κάθε προτεινόμενο πακέτο:

$$n_c$$

Το μοντέλο επιλογής των δημοφιλέστερων κατηγοριών.

Ο αριθμός των αντικειμένων ανά κατηγορία:

$$n_{i/c}$$

Ο αριθμός των αντικειμένων ανά προτεινόμενο πακέτο:

$$n$$

Ο μέγιστος αριθμός εμφάνισης κάθε αντικειμένου στα προτεινόμενα πακέτα:

$$n_i$$

Ο συντελεστής δημοφιλίας των αντικειμένων:

$$PopularityFactor$$

Το μοντέλο παραγωγής των N-καλύτερων προτεινόμενων πακέτων.

Στις επόμενες υποενότητες ακολουθεί μια αναλυτική περιγραφή των προαναφερόμενων παραμέτρων.

3.3.4.1 Αριθμός πακέτων (N) και κατηγοριών (n_c)

Η επιλογή του αριθμού των πακέτων (N) και του αριθμού των κατηγοριών (n_c) ανά πακέτο γίνεται με την προϋπόθεση ότι υπάρχει επαρκής αριθμός αντικειμένων ώστε να είναι δυνατή η συμπλήρωση όλων των πακέτων με τα απαιτούμενα αντικείμενα. Επίσης, για την επιλογή του αριθμού των πακέτων (N) και του αριθμού των κατηγοριών (n_c) ισχύουν οι εξής γενικοί περιορισμοί:

$$N \leq \frac{|I|}{n}$$

$$n_c \leq |C|$$

Όπου I: το σύνολο των αντικειμένων,

n: ο αριθμός των αντικειμένων ανά πακέτο και

C: το σύνολο των κατηγοριών

Συνεπώς, σύμφωνα με την επιλογή αυτή το σύστημα θα προτείνει τα N-καλύτερα πακέτα και ταυτόχρονα κάθε πακέτο θα περιλαμβάνει αντικείμενα από τις n_c δημοφιλέστερες για το χρήστη κατηγορίες.

3.3.4.2 Μοντέλο επιλογής των δημοφιλέστερων κατηγοριών

Για τον προσδιορισμό των δημοφιλέστερων για τον χρήστη κατηγοριών έχουν αναπτυχθεί τρία μοντέλα:

- Σύμφωνα με το πρώτο μοντέλο οι δημοφιλέστερες για τον χρήστη κατηγορίες είναι εκείνες από τις οποίες ο χρήστης έχει επιλέξει ή βαθμολογήσει τα περισσότερα αντικείμενα.
- Σύμφωνα με το δεύτερο μοντέλο αθροίζονται οι βαθμολογίες προτίμησης που έχει δώσει ο χρήστης στα αντικείμενα κάθε κατηγορίας και οι κατηγορίες που συγκεντρώνουν τον μεγαλύτερο βαθμό προτίμησης θεωρούνται οι δημοφιλέστερες.
- Η τρίτη προσέγγιση βασίζεται στα εκτιμώμενα αντικείμενα και οι δημοφιλέστερες κατηγορίες για τον χρήστη είναι εκείνες που συγκεντρώνουν τα περισσότερα αντικείμενα από αυτά που πρόκειται να προταθούν στον χρήστη.

Σύμφωνα με τα προαναφερόμενα μοντέλα προσδιορισμού των δημοφιλέστερων για τον χρήστη κατηγοριών υπάρχει η δυνατότητα της επιλογής του μοντέλου που ταιριάζει καλύτερα στην εκάστοτε εφαρμογή.

3.3.4.3 Αριθμός αντικειμένων ανά κατηγορία ($n_{i/c}$)

Ο αριθμός των αντικειμένων ανά κατηγορία ($n_{i/c}$) μπορεί να είναι σταθερός και ίσος για όλες τις κατηγορίες ή ανάλογος με τον βαθμό προτίμησης της κάθε κατηγορίας από το χρήστη. Στην πρώτη περίπτωση αρκεί η επιλογή του αριθμού των κατηγοριών (n_c) και του αριθμού των αντικειμένων ανά κατηγορία ($n_{i/c}$) για να προκύψει ο αριθμός των αντικειμένων ανά πακέτο:

$$n = n_c \cdot n_{i/c}$$

Ενώ στη δεύτερη περίπτωση, επιλέγεται ο αριθμός των κατηγοριών (n_c) και ο αριθμός των αντικειμένων ανά πακέτο (n) και ο αλγόριθμος υπολογίζει αυτόματα την κατανομή των n -αντικειμένων ανά πακέτο στις διάφορες κατηγορίες, ανάλογα με τον βαθμό προτίμησης της κάθε κατηγορίας από τον χρήστη. Για την αναλογική κατανομή των αντικειμένων στις κατηγορίες, ως βαθμός προτίμησης της κάθε κατηγορίας από τον χρήστη θεωρείται ο αριθμός των αντικειμένων που ο χρήστης έχει βαθμολογήσει ή επιλέξει από την εκάστοτε κατηγορία.

Για την αναλογική κατανομή των αντικειμένων στις διάφορες κατηγορίες έχει αναπτυχθεί ένας επαναληπτικός αλγόριθμος που περιγράφεται παρακάτω:

Αριθμός αντικειμένων ανά πακέτο: n
 Αριθμός διαθέσιμων αντικειμένων: $n_{rem} = n$
 Σύνολο με τις κατηγορίες και το πλήθος των αντικειμένων ανά κατηγορία:
 $A = \{ \langle c_1, n_{items, c_1} \rangle, \langle c_2, n_{items, c_2} \rangle, \dots \}$
 Επανάλαβε μέχρι να εξεταστούν όλα τα στοιχεία του συνόλου A :
 Άθροισμα αντικειμένων για όλες τις κατηγορίες του συνόλου A :

$$S = \sum_{n_{items, c_X} \in A} (n_{items, c_X})$$

 Για την κατηγορία (c_X) με το μικρότερο πλήθος αντικειμένων (n_{items, c_X}):
 Συντελεστής βαρύτητας κατηγορίας: $p_x = \frac{n_{items, c_X}}{S}$
 Κατανομή αριθμού αντικειμένων στην κατηγορία c_X :

$$n_{i/c_X} = \text{round}(p \cdot n_{rem})$$

 Εάν ($n_{i/c_X} == 0$) τότε $n_{i/c_X} = 1$
 Εναπομείναντα αντικείμενα: $n_{rem} = n_{rem} - n_{i/c_X}$
 Αφαίρεση του αντικειμένου της κατηγορίας c_X από το σύνολο A
 Επιστροφή της κατανομής του αριθμού των αντικειμένων ανά κατηγορία (n_{i/c_X}) για κάθε κατηγορία x του αρχικού συνόλου A .

Σύμφωνα με τον παραπάνω αλγόριθμο εξασφαλίζεται ότι η κατανομή των αντικειμένων θα γίνει με τέτοιο τρόπο ώστε κάθε κατηγορία αντικειμένων να περιλαμβάνει τουλάχιστον ένα αντικείμενο.

Για να γίνει περισσότερο κατανοητός ο τρόπος λειτουργίας του αλγόριθμου κατανομής των αντικειμένων στις κατηγορίες, ακολουθεί ένα παράδειγμα εκτέλεσης του αλγόριθμου αυτού με τα δεδομένα του πίνακα 3.3.4.3(α).

A/A	Κατηγορίες Αντικειμένων C_X	Πλήθος αντικειμένων ανά κατηγορία n_{items, c_X}	Πλήθος αντικειμένων ανά προτεινόμενο πακέτο n
1	C_1	60	12
2	C_2	30	
3	C_3	10	

Πίνακας 3.3.4.3(α): Παράδειγμα με το πλήθος των αντικειμένων ανά κατηγορία και το μέγεθος του προτεινόμενου πακέτου.

Βήμα Αλγόριθμου	1 ^ο	2 ^ο	3 ^ο
Αριθμός αντικειμένων ανά πακέτο: n	12	12	12
Υπόλοιπα αντικείμενα: n_{rem}	12	11	7
Σύνολο με τις κατηγορίες και το πλήθος των αντικειμένων ανά κατηγορία: A	{<c1, 60>, <c2, 30>, <c3, 10>}	{<c1, 60>, <c2, 30>}	{<c1, 60>}
Άθροισμα αντικειμένων για όλες τις κατηγορίες του συνόλου A : S	60+30+10 = 100	60+30 = 90	60
Η κατηγορία με τον μικρότερο αριθμό αντικειμένων	C_3	C_2	C_1
Συντελεστής Βαρύτητας για την κατηγορία x : p_x	$\frac{n_{items, c_3}}{S} = \frac{10}{100} = 0.1$	$\frac{30}{90} = 0.33$	$\frac{60}{60} = 1$
Κατανομή αριθμού αντικειμένων στην κατηγορία x : n_{i/c_X}	$round(p_3 \cdot n_{rem}) = round(0.1 \cdot 12) = 1$	$round(0.33 \cdot 11) = 4$	$round(1 \cdot 7) = 7$
Αφαίρεση από το σύνολο A της κατηγορίας:	C_3	C_2	C_1
Εναπομείναντα αντικείμενα: n_{rem}	$n_{rem} = n_{rem} - n_{i/c_3} = 12 - 1 = 11$	$11 - 4 = 7$	0

Πίνακας 3.3.4.3(β): Παράδειγμα εκτέλεσης του αλγόριθμου αναλογικής κατανομής των αντικειμένων στις κατηγορίες.

Στον πίνακα 3.3.4.3(β) παρουσιάζεται ένα παράδειγμα εκτέλεσης του αλγόριθμου αναλογικής κατανομής των αντικειμένων στις κατηγορίες σύμφωνα με τα δεδομένα του πίνακα 3.3.4.3(α). Στο παράδειγμα αυτό φαίνονται τα βήματα

εκτέλεσης του αλγόριθμου και οι τιμές που παίρνουν οι μεταβλητές κατά τη διάρκεια της εκτέλεσης. Η κατανομή των αντικειμένων σε κάθε κατηγορία, όπως υπολογίζεται από τον αλγόριθμο κατανομής, παρουσιάζεται στον πίνακα 3.3.4.3(γ).

A/A	Κατηγορίες Αντικειμένων C_X	Κατανομή αντικειμένων στις κατηγορίες n_i/c_X	Πλήθος αντικειμένων ανά πακέτο n
1	C_1	7	12
2	C_2	4	
3	C_3	1	

Πίνακας 3.3.4.3 (γ): Η αναλογική κατανομή των αντικειμένων στις τρεις κατηγορίες μετά την εκτέλεση του σχετικού αλγόριθμου.

3.3.4.4 Μέγιστος αριθμός εμφάνισης κάθε αντικειμένου (n_i)

Ο μέγιστος αριθμός εμφάνισης κάθε αντικειμένου στα προτεινόμενα πακέτα (n_i) μπορεί να είναι σταθερός και ίσος για όλα τα αντικείμενα ή κάθε αντικείμενο να έχει το δικό του μέγιστο αριθμό εμφάνισης. Στην πρώτη περίπτωση αρκεί η επιλογή μιας τιμής που είναι ίση για όλα τα αντικείμενα, ενώ στη δεύτερη περίπτωση ο μέγιστος αριθμός εμφάνισης κάθε αντικειμένου στα προτεινόμενα πακέτα λαμβάνεται από τα δεδομένα εισόδου και ειδικότερα από το αρχείο εισόδου που αφορά τα αντικείμενα.

Η τρίτη επιλογή που αφορά τον μέγιστο αριθμό εμφάνισης των αντικειμένων στα προτεινόμενα πακέτα είναι η επιλογή να μη λαμβάνεται υπόψη ο αριθμός αυτός. Σε αυτή την περίπτωση, ο αλγόριθμος ανάλογα με το μοντέλο λειτουργίας που έχει επιλεγεί, δε λαμβάνει υπόψη κανένα όριο στον αριθμό εμφανίσεων των αντικειμένων στα πακέτα εξόδου. Αυτό έχει σαν αποτέλεσμα να προτείνονται τα N-καλύτερα πακέτα αντικειμένων ανεξάρτητα από τον αριθμό εμφάνισης του κάθε αντικειμένου στα πακέτα αυτά.

Να σημειωθεί ότι οι περισσότερες από μια εμφανίσεις του κάθε αντικειμένου στα δεδομένα εξόδου αφορούν τις εμφανίσεις τους σε διαφορετικά πακέτα αντικειμένων. Αυτό σημαίνει ότι ανεξάρτητα από την επιλογή που θα γίνει για τον μέγιστο αριθμό εμφάνισης των αντικειμένων στο σύνολο των πακέτων εξόδου, κάθε αντικείμενο εμφανίζεται το πολύ μια φορά σε κάθε πακέτο, δηλαδή κάθε πακέτο περιλαμβάνει διακριτά αντικείμενα.

3.3.4.5 Αριθμός αντικειμένων ανά πακέτο

Η επιλογή του αριθμού των αντικειμένων σε κάθε προτεινόμενο πακέτο ακολουθεί τρεις προσεγγίσεις.

Η πρώτη προσέγγιση είναι αυτή που έχουμε σταθερή κατανομή των αντικειμένων στις κατηγορίες με αποτέλεσμα το μέγεθος του προτεινόμενου πακέτου να είναι ίσο με το γινόμενο του αριθμού των κατηγοριών (n_c) επί του αριθμού των αντικειμένων ανά κατηγορία ($n_{i/c}$).

$$n = n_c \cdot n_{i/c}$$

Η δεύτερη προσέγγιση ακολουθεί τη λογική της αναλογικής κατανομής των αντικειμένων στις δημοφιλέστερες κατηγορίες με αποτέλεσμα το μέγεθος του πακέτου να προκύπτει από το άθροισμα των αντικειμένων ανά κατηγορία (n_{i/c_j}).

$$n = \sum_j^m n_{i/c_j}$$

Όπου n_{i/c_j} : Ο αριθμός των αντικειμένων στην κατηγορία j

m : ο αριθμός των κατηγοριών

Η τρίτη προσέγγιση λαμβάνει υπόψη της τις προηγούμενες δύο αλλά τίθεται ένας αριθμός αντικειμένων ανά πακέτο n' μικρότερος από τον αριθμό n που προκύπτει από τις παραπάνω δύο προσεγγίσεις.

$$n' < n$$

Αυτή η επιλογή δίνεται για να περιορίσει το μέγεθος των προτεινόμενων πακέτων πέρα από αυτό που προκύπτει από τις άλλες παραμέτρους. Αυτό μπορεί να οδηγήσει στην εξαγωγή πακέτων που τα αντικείμενά τους είναι τα καλύτερα επιλεγμένα από μια ευρύτερη περιοχή κατηγοριών και αντικειμένων. Για παράδειγμα μπορεί να έχουμε επιλέξει τέσσερις κατηγορίες με ένα αντικείμενο ανά κατηγορία αλλά το μέγεθος του πακέτου να οριστεί σε 2. Ο αλγόριθμος τότε θα δημιουργήσει πακέτα με τα δύο καλύτερα αντικείμενα από τις τέσσερις δημοφιλέστερες κατηγορίες. Αυτό οδηγεί σε μια μεγαλύτερη ευελιξία και μπορεί να είναι χρήσιμο σε κάποιες εφαρμογές.

3.3.4.6 Συντελεστής δημοφιλίας των αντικειμένων

Ο συντελεστής δημοφιλίας του κάθε αντικειμένου είναι ο αριθμός που πολλαπλασιάζεται με το μοναδιαίο ποσοστό της δημοφιλίας (βλ. παρ. 3.2) κάθε

αντικειμένου για να προκύψει ένας παράγοντας που προστίθεται στον εκτιμώμενο βαθμό προτίμησης του αντικειμένου. Ο νέος αυτός εκτιμώμενος βαθμός προτίμησης δίνει προτεραιότητα στα δημοφιλέστερα αντικείμενα σύμφωνα με την τιμή του συντελεστή δημοφιλίας. Ο συντελεστής δημοφιλίας παίρνει τιμές από μηδέν έως δέκα.

$$PopularityFactor \in \{0, 1, 2, \dots, 10\}$$

Η τιμή μηδέν σημαίνει ότι δε λαμβάνεται υπόψη η δημοφιλία των αντικειμένων, ενώ η τιμή 10 σημαίνει ότι δίνεται μεγάλη προτεραιότητα στα δημοφιλέστερα αντικείμενα. Η δυνατότητα της κατάλληλης επιλογής αυτής της τιμής οδηγεί σε μεγαλύτερη ευελιξία του αλγόριθμου που μπορεί να προσαρμοστεί σύμφωνα με τις ανάγκες της εκάστοτε εφαρμογής.

3.3.4.7 Μοντέλο παραγωγής των προτεινόμενων πακέτων

Όπως θα γίνει περισσότερο κατανοητό στις επόμενες παραγράφους, το πρόβλημα της εύρεσης των N-καλύτερων πακέτων αντικειμένων μοντελοποιείται με χρήση ενός γραφήματος. Σύμφωνα με το διάγραμμα της εικόνας 3.3, κάθε φορά που ο αλγόριθμος παράγει το καλύτερο πακέτο αντικειμένων τροποποιείται το γράφημα μοντελοποίησης του προβλήματος για να παραχθεί το επόμενο καλύτερο πακέτο. Το γράφημα αυτό τροποποιείται διαγράφοντας ακμές και κόμβους ώστε να προκύψει νέος σχηματισμός πακέτου αντικειμένων. Οι ακμές και οι κόμβοι που διαγράφονται αφορούν το πακέτο που έχει σχηματιστεί ήδη και ο σκοπός της διαγραφής τους είναι να παραχθεί ένα επόμενο πακέτο που διαφοροποιείται από το προηγούμενο σύμφωνα με το μοντέλο που έχει επιλεγεί.

Για τον σχηματισμό των N-καλύτερων πακέτων έχουν θεωρηθεί διάφορα μοντέλα λειτουργίας του αλγόριθμου. Τα μοντέλα αυτά είναι τα εξής:

1. Διαγραφή της ακμής με τον μέγιστο βαθμό προτίμησης.
2. Διαγραφή της ακμής με τον ελάχιστο βαθμό προτίμησης.
3. Διαγραφή του 50% των ακμών με τον μέγιστο βαθμό προτίμησης.
4. Διαγραφή του 50% των ακμών με τον ελάχιστο βαθμό προτίμησης.
5. Διαγραφή του 100% των ακμών.
6. Παραγωγή Διακριτών πακέτων.

Στα διάφορα μοντέλα λειτουργίας του αλγόριθμου, η ακμή με το μεγαλύτερο βαθμό προτίμησης θεωρείται εκείνη η ακμή του γραφήματος που αντιστοιχεί στο αντικείμενο με τον μεγαλύτερο βαθμό προτίμησης. Σε περίπτωση που οι ακμές δύο ή

περισσότερων αντικειμένων παρουσιάζουν τον ίδιο βαθμό προτίμησης τότε διαγράφεται τυχαία μια ακμή ενός αντικειμένου με την πιθανότητα της διαγραφής να είναι ίση για όλες τις υποψήφιες προς διαγραφή ακμές. Με ανάλογο τρόπο διαγράφεται η ακμή που αντιστοιχεί στο αντικείμενο με τον μικρότερο βαθμό προτίμησης. Η τακτική αυτή γενικεύεται στην περίπτωση διαγραφής του 50% του αριθμού των ακμών με τον μεγαλύτερο ή τον μικρότερο βαθμό προτίμησης. Στην περίπτωση της διαγραφής του 100% των ακμών, διαγράφονται όλες οι ακμές των αντικειμένων που περιλαμβάνονται στον σχηματισμό του i -πακέτου με σκοπό να παραχθεί το $(i+1)$ -πακέτο.

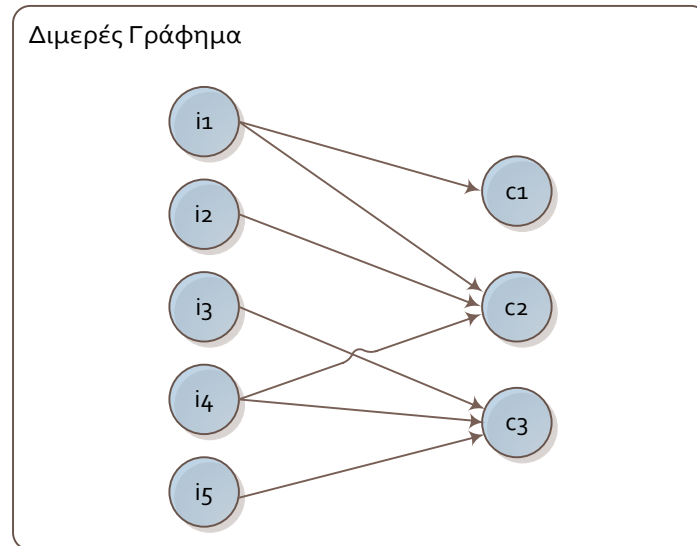
Να σημειωθεί ότι οι ακμές που αναφέρονται παραπάνω συνδέουν τα αντικείμενα με τις κατηγορίες όπως φαίνεται στην εικόνα 3.3.5(α). Η διαγραφή μιας ακμής σημαίνει ότι το αντικείμενο δεν θα εμφανιστεί ξανά σε αυτή την κατηγορία αλλά μπορεί να εμφανιστεί σε άλλη κατηγορία, αν πρόκειται βέβαια για αντικείμενο που περιλαμβάνεται σε δύο ή περισσότερες κατηγορίες.

Στο μοντέλο παραγωγής διακριτών πακέτων διαγράφονται όλοι οι κόμβοι των αντικειμένων με συνέπεια αυτά τα αντικείμενα να μην μπορούν πλέον να εμφανιστούν στον σχηματισμό επόμενων πακέτων. Αυτό έχει σαν συνέπεια το σχηματισμό διακριτών πακέτων αντικειμένων με το κάθε αντικείμενο να μπορεί να εμφανιστεί το πολύ σε ένα πακέτο.

3.3.5 Βέλτιστη Αντιστοίχιση των Αντικειμένων με τις Κατηγορίες

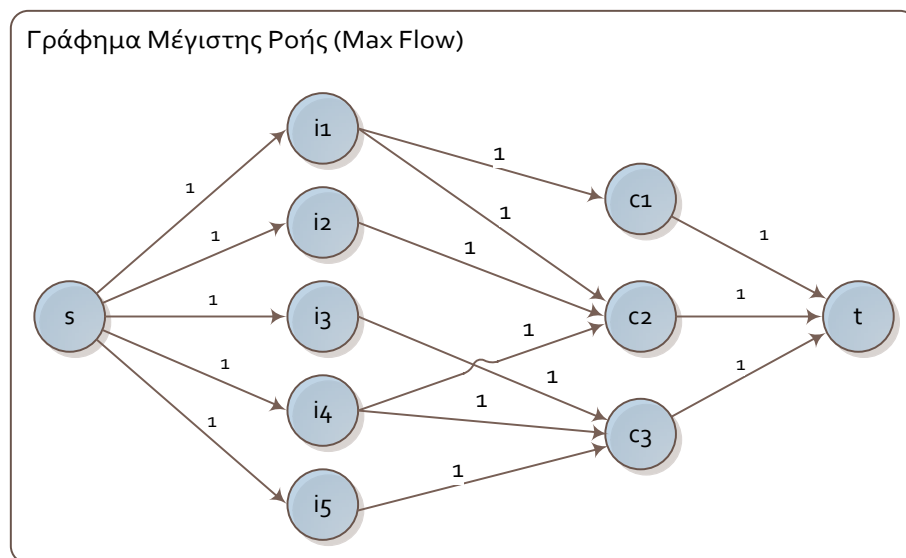
Το πέμπτο και πιο σημαντικό στάδιο του αλγόριθμου αφορά στη βέλτιστη αντιστοίχιση των αντικειμένων με τις κατηγορίες με τέτοιο τρόπο ώστε να προκύψει το καλύτερο πακέτο αντικειμένων. Επισημαίνεται ότι κάθε αντικείμενο ανήκε σε μία ή περισσότερες κατηγορίες αντικειμένων.

Από το σύνολο των αντικειμένων (I) και των κατηγοριών (C) προκύπτει ένας διμερές γράφημα $G(V, E)$ με διαμέριση κόμβων ($V = I \cup C$) οι οποίοι αντιστοιχούν στα αντικείμενα και τις κατηγορίες. Οι ακμές (E) του γραφήματος αναπαριστούν τη σχέση των αντικειμένων με τις κατηγορίες που ανήκουν. Στο σχήμα 3.3.5(α) φαίνεται ένα παράδειγμα ενός τέτοιου γραφήματος που περιλαμβάνει πέντε αντικείμενα και τρεις κατηγορίες.



Εικόνα 3.3.5 (α): Διμερές γράφημα $G(V, E)$ αντικειμένων και κατηγοριών.

Το ζητούμενο είναι κάθε κατηγορία να αντιστοιχιστεί με έναν αριθμό από αντικείμενα ($n_{i/C}$). Αυτό το πρόβλημα δημιουργίας αντιστοιχίσεων μπορεί να αναχθεί στο πρόβλημα της μέγιστης ροής (Cormen et al., 2001; Dasgupta et al., 2008). Για τη μετατροπή του διμερούς γραφήματος σε γράφημα μέγιστης ροής προσθέτουμε έναν αφετηριακό κόμβο s , με εξερχόμενες ακμές προς όλα τα αντικείμενα, και έναν τερματικό κόμβο t με εισερχόμενες ακμές από όλες τις κατηγορίες. Επίσης, στο γράφημα μέγιστης ροής προσθέτουμε τις χωρητικότητες των ακμών c_{ij} που δείχνουν τη μέγιστη επιτρεπόμενη ροή κάθε ακμής. Στην εικόνα 3.3.5(β) φαίνεται η μετατροπή του παραδείγματος της εικόνας 3.3.5(α) σε γράφημα μέγιστης ροής.



Εικόνα 3.3.5 (β): Γράφημα μέγιστης ροής για τη βέλτιστη αντιστοίχιση μεταξύ αντικειμένων και κατηγοριών.

Στο παράδειγμα της εικόνας 3.3.5(β) η εύρεση της μέγιστης ροής από τον αφετηριακό κόμβο s στον τερματικό κόμβο t μπορεί να οδηγήσει στο βέλτιστο ταίριασμα μεταξύ των αντικειμένων και των κατηγοριών. Όπως παρατηρούμε στο παράδειγμα της εικόνας 3.3.5(β), όλες οι ακμές έχουν χωρητικότητα ίση με ένα ($C_{ij} = 1 \ \forall i, j \in V$ και $i \neq j$), αυτό σημαίνει ότι θα προκύψει ένα βέλτιστο ταίριασμα με ένα αντικείμενο ανά κατηγορία. Το παράδειγμα αυτό μπορεί να γενικευθεί και να αποτελέσει ένα μοντέλο εύρεσης των βέλτιστων ταιριασμάτων μεταξύ αντικειμένων και κατηγοριών.

Ένα χρήσιμο θεώρημα που αφορά στα γραφήματα μέγιστης ροής, σύμφωνα με τους Cormen et al. (2001) και τους Dasgupta et al. (2008), είναι το θεώρημα της ακεραιότητας που διατυπώνεται ως εξής:

Εάν όλες οι χωρητικότητες των ακμών ενός γραφήματος μέγιστης ροής παίρνουν ακέραιες τιμές, τότε η μέγιστη ροή που προκύπτει έχει επίσης ακέραια τιμή.

Όπως γίνεται φανερό, το θεώρημα της ακεραιότητας είναι εκείνο που μας επιτρέπει, σε ένα γράφημα ροής όπως αυτό που εξετάζουμε, να αντιστοιχίσουμε βέλτιστα έναν ακέραιο αριθμό αντικειμένων με τις κατηγορίες με βάση την υπολογιζόμενη μέγιστη ροή.

Η βέλτιστη αντιστοίχιση μεταξύ των αντικειμένων και των κατηγοριών μπορεί να είναι εφικτή με τη δημιουργία ενός γραφήματος ροής και την εκτέλεση ενός αλγόριθμου εύρεσης της μέγιστης ροής. Αυτό όμως δε μας λύνει το πρόβλημα που τίθεται σε αυτή την εργασία, καθώς εμείς θέλουμε βέλτιστη αντιστοίχιση μεταξύ των αντικειμένων και των κατηγοριών διατηρώντας μια δεδομένη ροή και ταυτόχρονα τα αντικείμενα που αντιστοιχίζονται να έχουν μέγιστο βαθμό προτίμησης από τον χρήστη. Μια παραλλαγή του ζητούμενου προβλήματος μπορεί να επιτευχθεί με την αντιστροφή της βαθμολογίας προτίμησης κάθε αντικειμένου με σκοπό τη μετατροπή της σε κόστος σύμφωνα με τον τύπο:

$$\text{cost} = \frac{1}{\text{rating}}$$

Συνεπώς, το πρόβλημα μπορεί να αναχθεί σε πρόβλημα Ροής Ελάχιστου Κόστους (*minimum cost flow problem*) που ζητούμενο έχει την ελαχιστοποίηση του κόστους σε δεδομένη ροή.

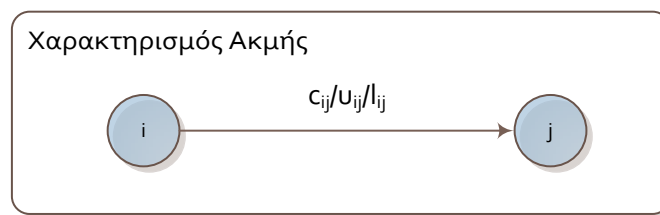
Σύμφωνα με τον Bertsekas (1998), για την μετατροπή του γραφήματος μέγιστης ροής σε γράφημα ροής ελάχιστου κόστους προσθέτουμε μια ακμή μεταξύ τερματικού (t) και αφετηριακού κόμβου (s) και τοποθετούμε σε κάθε ακμή τις τιμές:

c_{ij} : Κόστος ανά μονάδα διερχόμενης ροής για την ακμή μεταξύ των κόμβων i και j

u_{ij} : Μέγιστη ποσότητα διερχόμενης ροής για την ακμή μεταξύ των κόμβων i και j.

l_{ij} : Ελάχιστη ποσότητα διερχόμενης ροής για την ακμή μεταξύ των κόμβων i και j

Στην εικόνα 3.3.5(γ) φαίνονται δύο κόμβοι συνδεδεμένοι με μια ακμή που περιλαμβάνει το κόστος ανά μονάδα ροής, τη μέγιστη και την ελάχιστη διερχόμενη ροή.



Εικόνα 3.3.5 (γ): Χαρακτηρισμός ακμής.

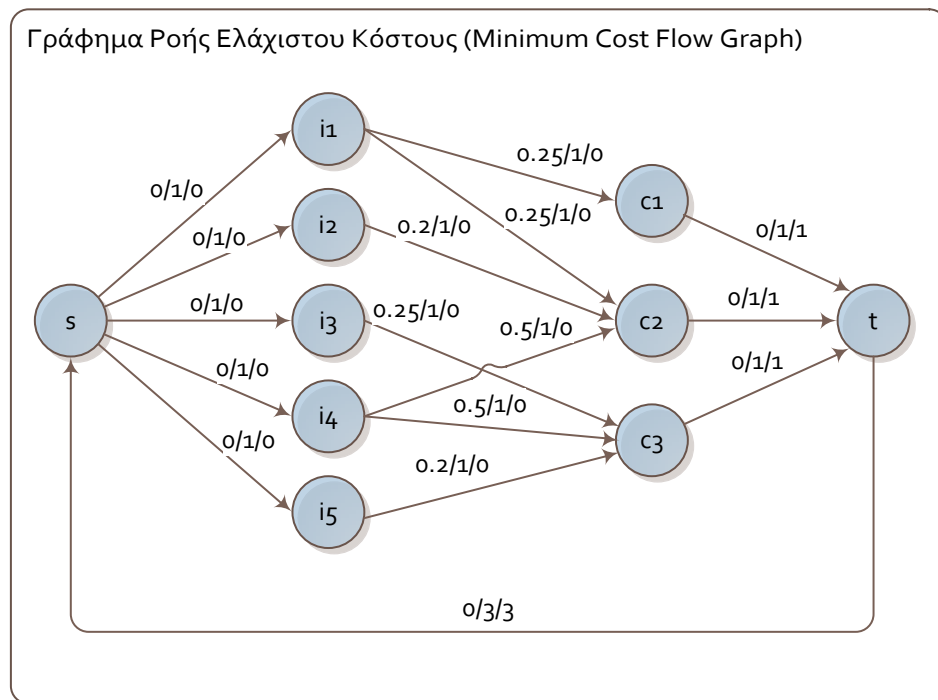
Σε ένα δίκτυο ροής ελάχιστου κόστους οι ακμές που εξέρχονται από τον αφετηριακό κόμβο, οι ακμές που εισέρχονται στον τερματικό κόμβο και η ακμή μεταξύ αφετηριακού και τερματικού κόμβου έχουν κόστος μηδέν. Όλες οι υπόλοιπες ακμές έχουν ένα δεδομένο κόστος μεγαλύτερο (ή ίσο) του μηδενός. Στην εικόνα 3.3.5(δ) φαίνεται ένα παράδειγμα γραφήματος ροής ελάχιστου κόστους που ανταποκρίνεται στο πρόβλημα που αντιμετωπίζουμε σε αυτή την εργασία.

Για το παράδειγμα της εικόνας 3.3.5(δ) θεωρούμε τις βαθμολογίες προτίμησης (*ratings*) των αντικειμένων και κατά συνέπεια τον υπολογισμό του κόστους σύμφωνα με τις τιμές που δίνονται στον πίνακα 3.3.5(α).

A/A	Αντικείμενο	Βαθμολογία προτίμησης (rating)	Κόστος (cost = 1/rating)
1	i1	4	0.25
2	i2	5	0.2
3	i3	4	0.25
4	i4	2	0.5
5	i5	5	0.2

Πίνακας 3.3.5(α): Παράδειγμα βαθμολογιών προτίμησης και υπολογισμού κόστους.

Η βαθμολογία προτίμησης των αντικειμένων μετατρέπεται σε κόστος και αναγράφεται πάνω στις ακμές που εξέρχονται από τα αντίστοιχα αντικείμενα. Επίσης, στο παράδειγμα της εικόνας 3.3.5(δ) είναι φανερό πως η ροή είναι δεδομένη, ίση με 3, και το ζητούμενο είναι η ελαχιστοποίηση του κόστους που θα δώσει τα βέλτιστα ταιριάσματα μεταξύ των αντικειμένων και των κατηγοριών. Σύμφωνα με αυτό το παράδειγμα, μετά την εύρεση του δικτύου ροής ελάχιστου κόστους θα προκύψει κάθε κατηγορία να ταιριάζει βέλτιστα με ένα αντικείμενο και αυτό θα οδηγήσει στον σχηματισμό ενός βέλτιστου πακέτου αντικειμένων μεγέθους $N=3$.



Εικόνα 3.3.5 (δ): Γράφημα ροής ελάχιστου κόστους.

Το πρόβλημα της ροής ελάχιστου κόστους, σύμφωνα με τον Bertsekas (1998) μπορεί να αναχθεί σε πρόβλημα γραμμικού προγραμματισμού, η λύση του οποίου θα δώσει το βέλτιστο ταίριασμα μεταξύ των κατηγοριών και των αντικειμένων.

Για ένα δεδομένο γράφημα $G(V, E)$, όπου V το σύνολο των κόμβων και E το σύνολο των ακμών, το πρόβλημα της ροής ελάχιστου κόστους μοντελοποιείται ως πρόβλημα γραμμικού προγραμματισμού ως εξής:

$$\text{Minimize } \sum_{\substack{i,j \in V \\ i \neq j}} c_{ij} \cdot x_{ij}$$

$$l_{ij} \leq x_{ij} \leq u_{ij}, \quad \forall i, j \in V \text{ \& } i \neq j$$

$$\sum_{\substack{\forall j \in V \\ i \neq j}} x_{ij} = 0, \forall i \in V$$

Όπου c_{ij} : Το κόστος ανά μονάδα διερχόμενης ροής από τον κόμβο i στον κόμβο j ,
 x_{ij} : Η ποσότητα διερχόμενης ροής από τον κόμβο i στον κόμβο j ,
 u_{ij} : Η μέγιστη ποσότητα διερχόμενης ροής από τον κόμβο i στον κόμβο j ,
 l_{ij} : Η ελάχιστη ποσότητα διερχόμενης ροής από τον κόμβο i στον κόμβο j ,
 V : Το σύνολο των κόμβων του γραφήματος ροής.

Μετά την επίλυση του προβλήματος οι μεταβλητές απόφασης (x_{ij}) θα λάβουν τιμές, βάσει των οποίων προκύπτει η βέλτιστη αντιστοίχιση των κατηγοριών με τα αντικείμενα. Η βέλτιστη αυτή λύση για το γράφημα ροής του παραδείγματος της εικόνας 3.3.5(δ) παρουσιάζεται στον πίνακα 3.3.5(β).

x_{ij}	Τιμές	Κόστος	$\sum_{\substack{i,j \in V \\ i \neq j}} c_{ij} \cdot x_{ij}$	Βέλτιστη αντιστοίχιση Αντικειμένων με κατηγορίες
$x_{s,i1}$	1	0	0.65	$i_1 - c_1$ $i_2 - c_2$ $i_5 - c_3$
$x_{s,i2}$	1	0		
$x_{s,i3}$	0	0		
$x_{s,i4}$	0	0		
$x_{s,i5}$	1	0		
$x_{i1,c1}$	1	0.25		
$x_{i1,c2}$	0	0		
$x_{i2,c2}$	1	0.2		
$x_{i3,c3}$	0	0		
$x_{i4,c2}$	0	0		
$x_{i4,c3}$	0	0		
$x_{i5,c3}$	1	0.2		
$x_{c1,t}$	1	0		
$x_{c2,t}$	1	0		
$x_{c3,t}$	1	0		
$x_{t,s}$	3	0		

Πίνακας 3.3.5 (β): Η λύση του παραδείγματος της εικόνας 3.3.5(δ).

Ο πίνακα 3.3.5(β) περιλαμβάνει τη λύση με τις τιμές των μεταβλητών, το κόστος για κάθε ακμή διερχόμενης ροής, την ελάχιστη τιμή της αντικειμενικής συνάρτησης

και τη βέλτιστη αντιστοίχιση μεταξύ των αντικειμένων και των κατηγοριών. Όπως είναι φανερό κάθε άλλη λύση αυξάνει την τιμή της αντικειμενικής συνάρτησης, συνεπώς η λύση που απεικονίζεται στον πίνακα 3.3.5(β) είναι η βέλτιστη λύση του προβλήματος.

3.3.6 Ενημέρωση του Γραφήματος Μοντελοποίησης του Προβλήματος

Σε αυτό το στάδιο ο αλγόριθμος ενημερώνει το γράφημα μοντελοποίησης του προβλήματος εύρεσης του καλύτερου πακέτου αντικειμένων, σύμφωνα με το μοντέλο εκτέλεσης του αλγόριθμου για την παραγωγή των πακέτων, με σκοπό να προκύψει το επόμενο καλύτερο πακέτο αντικειμένων. Ο αλγόριθμος εκτελεί το στάδιο αυτό υπό την προϋπόθεση ότι δεν έχουν σχηματιστεί ακόμη τα N-καλύτερα πακέτα αντικειμένων και ταυτόχρονα ο σχηματισμός νέου πακέτου αντικειμένου είναι εφικτός. Αυτές οι δύο προϋποθέσεις εκτέλεσης του σταδίου αυτού φαίνονται και στο διάγραμμα ροής με τα στάδια του αλγόριθμου στην εικόνα 3.3.

Το γράφημα μοντελοποίησης του προβλήματος τροποποιείται σύμφωνα με το μοντέλο λειτουργίας του αλγόριθμου που έχει επιλεγεί. Τα μοντέλα παραγωγής των πακέτων αναλύονται στην παράγραφο 3.3.4.7. Ωστόσο, σε αυτό το σημείο, που έχει παρουσιαστεί το στάδιο της βέλτιστης αντιστοίχισης των αντικειμένων με τις κατηγορίες, της ενότητας 3.3.5, μπορεί να γίνει εύκολα κατανοητός ο τρόπος λειτουργίας του κάθε μοντέλου της παραγράφου 3.3.4.7.

Επίσης, η ενημέρωση του γραφήματος λαμβάνει υπόψη της και τις υπόλοιπες παραμέτρους όπως την παράμετρο του μέγιστου αριθμού εμφάνισης ενός αντικειμένου. Σύμφωνα με αυτή την παράμετρο, αν ένα αντικείμενο εξαντλήσει τον μέγιστο αριθμό εμφανίσεων στα πακέτα εξόδου, τότε διαγράφεται ο αντίστοιχος κόμβος του γραφήματος, με αποτέλεσμα να μην είναι δυνατή η νέα εμφάνισή του στα προτεινόμενα πακέτα.

3.3.7 Εξαγωγή και Προβολή των N-καλύτερων Πακέτων Αντικειμένων

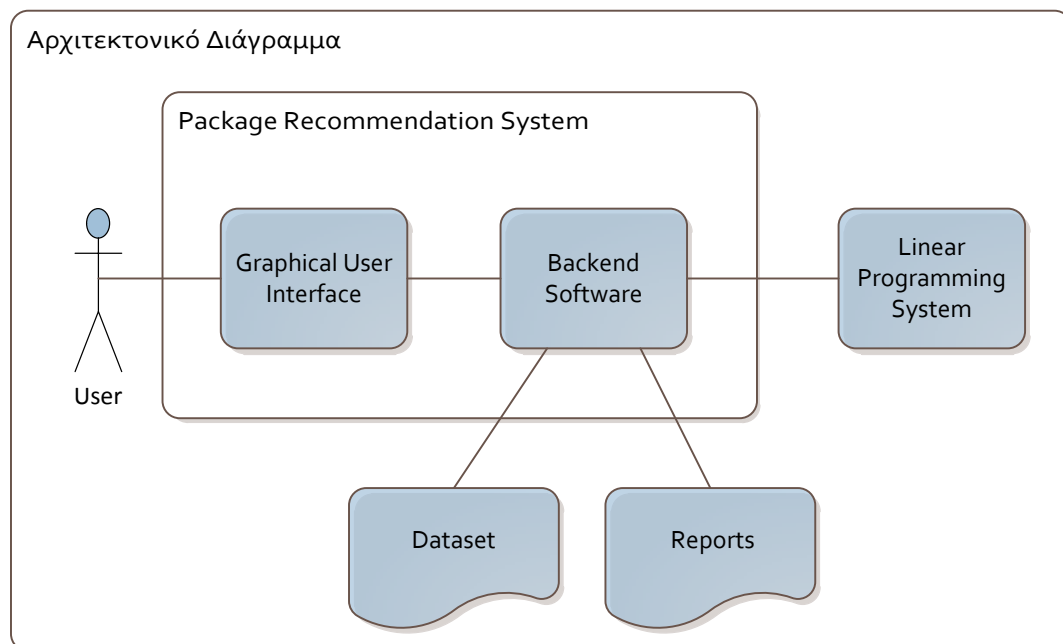
Το τελευταίο στάδιο του αλγόριθμου αφορά την εξαγωγή και την προβολή του αποτελέσματος των N-καλύτερων πακέτων αντικειμένων. Στο στάδιο αυτό προβάλλονται τα προτεινόμενα πακέτα αντικειμένων ταξινομημένα σε φθίνουσα σειρά προτίμησης σύμφωνα με τον συνολικό βαθμό προτίμησης των αντικειμένων που περιλαμβάνουν. Να σημειωθεί ότι η βαθμολογία προτίμησης κάθε πακέτου προκύπτει από το άθροισμα των βαθμών προτίμησης των αντικειμένων που περιλαμβάνονται στο αντίστοιχο πακέτο. Με το στάδιο αυτό ολοκληρώνεται η λειτουργία του αλγόριθμου συστάσεων που προτείνει σύνολα αντικειμένων.

Κεφάλαιο 4: Σχεδίαση Λογισμικού

4.1 Γενικά

Σε αυτό το κεφάλαιο περιγράφεται η σχεδίαση του λογισμικού που έχει σκοπό να ενσωματώσει όλη τη λειτουργικότητα του αλγόριθμου που αναπτύχθηκε. Στις επόμενες παραγράφους παρουσιάζεται η αρχιτεκτονική του συστήματος λογισμικού, με την περιγραφή των απαιτούμενων υποσυστημάτων και των λειτουργιών τους, και προσδιορίζεται η συμπεριφορά του συστήματος μέσα από τα σχετικά συνεργατικά διαγράμματα.

4.2 Αρχιτεκτονική του Συστήματος Λογισμικού



Εικόνα 4.2: Το γενικό αρχιτεκτονικό διάγραμμα της εφαρμογής λογισμικού που υλοποιεί την λειτουργικότητα του αλγόριθμου.

Στην εικόνα 4.2 φαίνεται το γενικό αρχιτεκτονικό διάγραμμα του συστήματος λογισμικού που έχει σκοπό να ενσωματώσει όλη την απαραίτητη λειτουργικότητα του αλγόριθμου που σχεδιάστηκε. Τα μέρη που απαρτίζουν το σύστημα λογισμικού και φαίνονται στην εικόνα 4.2 είναι τα εξής:

User: Ο χρήστης που έχει πρόσβαση και χειρίζεται το σύστημα μέσω του γραφικού περιβάλλοντος χρήστη (Graphical user interface).

Package Recommendation System: Η εφαρμογή Λογισμικού που υλοποιεί το γραφικό περιβάλλον χρήστη (Graphical user interface) και όλη την απαραίτητη λειτουργικότητα που ενσωματώνεται στο *backend* λογισμικό. Η εφαρμογή επικοινωνεί με το ανεξάρτητο σύστημα λογισμικού *Linear programming System* που λύνει το πρόβλημα γραμμικού προγραμματισμού για την παραγωγή των N-καλύτερων πακέτων αντικειμένων. Το *package Recommendation System* έχει πρόσβαση στα αρχεία του συνόλου των δεδομένων (*dataset*) και εξάγει τα αποτελέσματα των προτάσεων σε αρχεία αναφορών (*reports*).

Dataset: Το σύνολο δεδομένων που είναι απαραίτητο για τη λειτουργία της εφαρμογής το οποίο περιλαμβάνει τα αρχεία για την περιγραφή των αντικειμένων (*items*) και για τις βαθμολογίες προτίμησης (*ratings*) των χρηστών.

Reports: Τα αρχεία που περιλαμβάνουν τα προτεινόμενα πακέτα αντικειμένων για τον εκάστοτε χρήστη.

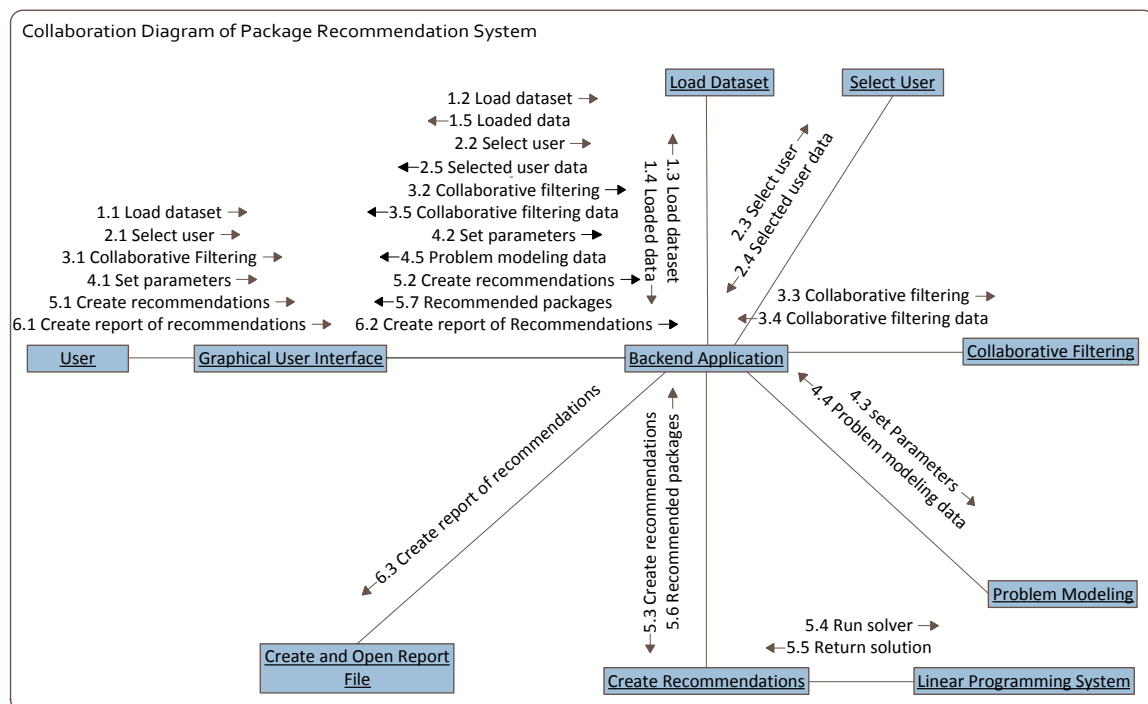
Linear Programming System: Το εξωτερικό και ανεξάρτητο σύστημα λογισμικού που επικοινωνεί με το *Package Recommendation System* με σκοπό την επίλυση του προβλήματος γραμμικού προγραμματισμού για την εύρεση των N-καλύτερων πακέτων αντικειμένων. Το εξωτερικό σύστημα γραμμικού προγραμματισμού που χρησιμοποιείται περιγράφεται στην ενότητα 2.6 του κεφαλαίου 2.

4.3 Συμπεριφορά του Συστήματος Λογισμικού

Σε αυτή την ενότητα παρουσιάζεται η συμπεριφορά του συστήματος λογισμικού μέσα από τα σχετικά συνεργατικά διαγράμματα για την κανονική λειτουργία και τη λειτουργία αξιολόγησης. Το σύστημα λογισμικού αφορά στην κανονική λειτουργία η οποία στοχεύει στην πρόταση πακέτων αντικειμένων σε χρήστες, σύμφωνα με τα δεδομένα προτιμήσεων τους, και στη λειτουργία αξιολόγησης η οποία στοχεύει στην πρόταση πακέτων αντικειμένων με τον ταυτόχρονο υπολογισμό δεικτών αξιολόγησης. Όλες οι λεπτομέρειες για την αξιολόγηση του αλγόριθμου αναφέρονται

στον κεφάλαιο 6, εδώ δίνονται τα απαραίτητα στοιχεία που αφορούν στην κατανόηση της συμπεριφοράς του συστήματος.

Στην εικόνα 4.3(α) παρουσιάζεται το συνεργατικό διάγραμμα για την κανονική λειτουργία του συστήματος συστάσεων πακέτων αντικειμένων. Στο διάγραμμα αυτό φαίνεται όλη η απαραίτητη λειτουργικότητα του συστήματος λογισμικού και πώς οι διάφορες μονάδες επικοινωνούν μεταξύ τους. Το διάγραμμα περιλαμβάνει την αλληλεπίδραση του χρήστη με το γραφικό περιβάλλον της εφαρμογής και πώς αυτό επικοινωνεί με το *backend* λογισμικό που εκτελεί όλες τις λειτουργίες.



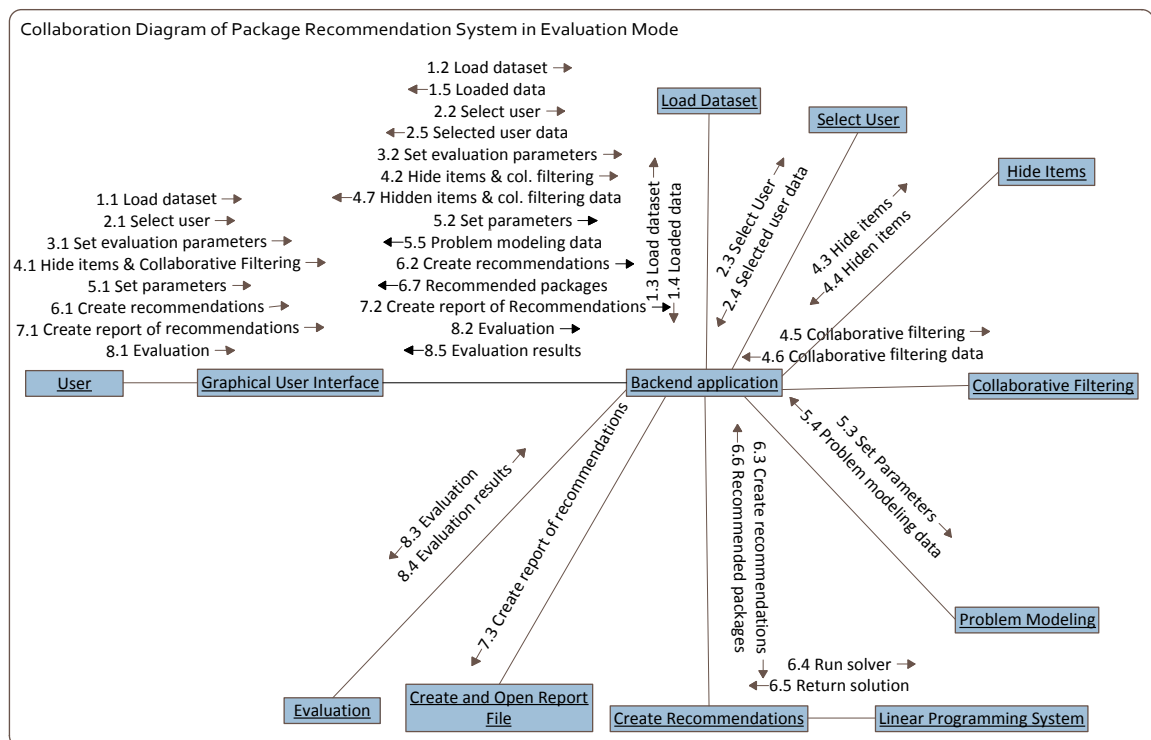
Εικόνα 4.3 (α): Το συνεργατικό διάγραμμα του συστήματος συστάσεων συνόλων αντικειμένων για την κανονική λειτουργία.

Οι διαδοχικές λειτουργίες που εκτελούνται με τη βοήθεια του χρήστη και περιγράφονται από το συνεργατικό διάγραμμα της εικόνας 4.3(α), για την κανονική λειτουργία του συστήματος, είναι οι εξής:

1. *Load Dataset*: Ο χρήστης επιλέγει το σύνολο δεδομένων που φορτώνεται στο σύστημα.
2. *Select User*: Επιλογή του χρήστη στον οποίο θα προταθούν τα N-καλύτερα πακέτα αντικειμένων.
3. *Collaborative Filtering*: Εκτέλεση του συνεργατικού φιλτραρίσματος για την πρόβλεψη της βαθμολογίας προτίμησης των αντικειμένων που θα χρησιμοποιηθούν για το σχηματισμό των καλύτερων προτεινόμενων πακέτων.

4. *Set parameters*: Επιλογή των παραμέτρων βάση των οποίων θα λειτουργήσει ο αλγόριθμος του συστήματος. Οι παράμετροι αυτές έχουν παρουσιαστεί στο κεφάλαιο 3.
5. *Create recommendations*: Εκτέλεση του αλγόριθμου και δημιουργία των N-καλύτερων πακέτων αντικειμένων σύμφωνα με τις παραμέτρους που τέθηκαν στο προηγούμενο βήμα.
6. *Create report of recommendations*: Δημιουργία, άνοιγμα και παρουσίαση του αρχείου που περιλαμβάνει την αναφορά με τα προτεινόμενα πακέτα αντικειμένων.

Στην εικόνα 4.3(β) παρουσιάζεται το συνεργατικό διάγραμμα για τη λειτουργία του συστήματος συστάσεων πακέτων αντικειμένων σε κατάσταση αξιολόγησης. Στο διάγραμμα αυτό φαίνεται όλη η απαραίτητη λειτουργικότητα του συστήματος λογισμικού. Το διάγραμμα περιλαμβάνει την αλληλεπίδραση του χρήστη με το γραφικό περιβάλλον της εφαρμογής και πώς αυτό επικοινωνεί με το *backend* λογισμικό που εκτελεί όλες τις λειτουργίες.



Εικόνα 4.3 (β): Το συνεργατικό διάγραμμα του συστήματος συστάσεων πακέτων αντικειμένων για τη λειτουργία αξιολόγησης.

Οι διαδοχικές λειτουργίες, που εκτελούνται με τη βοήθεια του χρήστη και περιγράφονται από το διάγραμμα της εικόνας 4.3(β) για την λειτουργία αξιολόγησης του συστήματος είναι οι εξής:

1. *Load Dataset*: Ο χρήστης επιλέγει το σύνολο δεδομένων που φορτώνεται στο σύστημα και διατηρείται στη μνήμη.
2. *Select User*: Επιλογή του χρήστη στον οποίο θα προταθούν τα N-καλύτερα πακέτα αντικειμένων και με βάση τα δεδομένα του θα πραγματοποιηθεί η αξιολόγηση του αλγόριθμου.
3. *Set evaluation parameters*: Επιλογή των παραμέτρων αξιολόγησης που είναι το ποσοστό απόκρυψης των αντικειμένων και η ελάχιστη βαθμολογία προτίμησης των βαθμολογημένων αντικειμένων. Περισσότερες λεπτομέρειες για αυτές τις παραμέτρους και τη διαδικασία αξιολόγησης αναφέρονται στο κεφάλαιο 6 που αφορά στην αξιολόγηση του συστήματος.
4. *Hide items and collaborative filtering*: Απόκρυψη αντικειμένων και εκτέλεση του συνεργατικού φιλτραρίσματος για την πρόβλεψη της βαθμολογίας προτίμησης των αντικειμένων που θα χρησιμοποιηθούν για το σχηματισμό των καλύτερων προτεινόμενων πακέτων. Τα κρυμμένα αντικείμενα θα χρησιμοποιηθούν για τον υπολογισμό των δεικτών αξιολόγησης που αφορούν στη σύγκριση μεταξύ των κρυμμένων αντικειμένων και των προτεινόμενων πακέτων αντικειμένων.
5. *Set parameters*: Επιλογή των παραμέτρων βάση των οποίων λειτουργεί ο αλγόριθμος του συστήματος συστάσεων. Οι παράμετροι αυτές έχουν παρουσιαστεί στο κεφάλαιο 3.
6. *Create recommendations*: Εκτέλεση του αλγόριθμου και δημιουργία των N-καλύτερων πακέτων αντικειμένων σύμφωνα με τις τιμές των παραμέτρων που έχουν τεθεί.
7. *Create report of recommendations*: Δημιουργία, άνοιγμα και παρουσίαση του αρχείου που περιλαμβάνει την αναφορά με τα προτεινόμενα πακέτα αντικειμένων.
8. *Evaluation*: Αξιολόγηση της προσέγγισης με τον υπολογισμό δεικτών για την ακρίβεια που επιτυγχάνουν τα προτεινόμενα πακέτα αντικειμένων σε σχέση με τα κρυμμένα αντικείμενα του τρέχοντα χρήστη.

Κεφάλαιο 5: Ανάπτυξη Λογισμικού

5.1 Γενικά

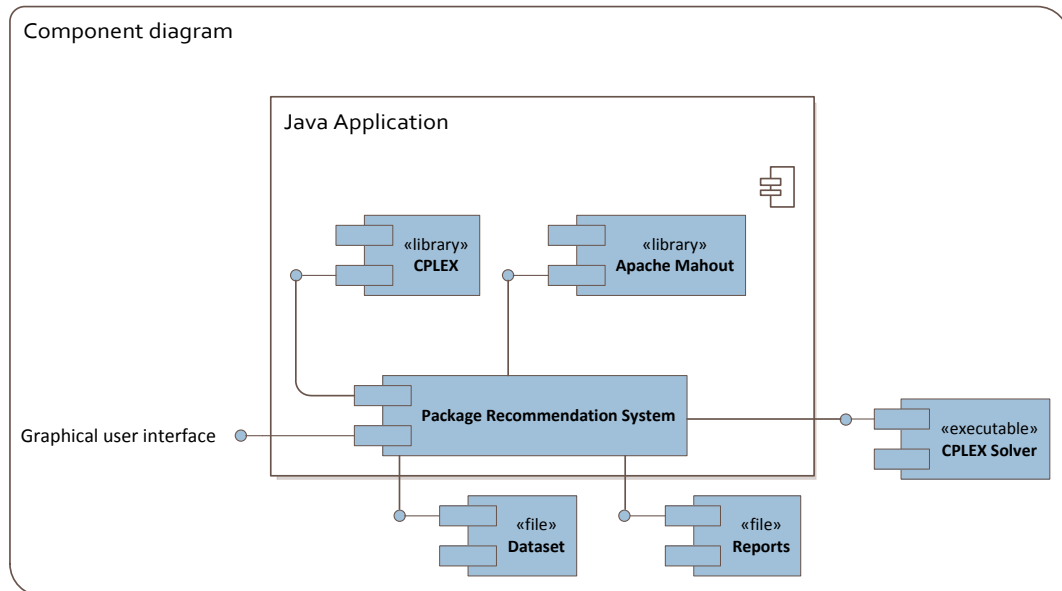
Σε αυτό το κεφάλαιο περιγράφεται το λογισμικό που αναπτύχθηκε με σκοπό την υλοποίηση ενός ολοκληρωμένου συστήματος συστάσεων το οποίο ενσωματώνει ολόκληρη τη λειτουργικότητα του αλγόριθμου της παρούσας εργασίας και επιπλέον περιλαμβάνει τις απαραίτητες λειτουργίες που αφορούν στην αξιολόγηση της προσέγγισης. Περιγράφονται οι λεπτομέρειες για την υλοποίηση του λογισμικού που αφορούν τον πηγαίο κώδικα και παρουσιάζονται οι οθόνες του γραφικού περιβάλλοντος της εφαρμογής. Οι λεπτομέρειες για τον πηγαίο κώδικα παρουσιάζονται σε μορφή διαγραμμάτων ενώ μια εκτενέστερη αναφορά περιλαμβάνεται στο παράρτημα Α της παρούσας εργασίας. Σχετικά με το γραφικό περιβάλλον της εφαρμογής περιγράφονται οι λειτουργίες και ο τρόπος χρήσης.

5.2 Πηγαίος Κώδικας

Σε αυτή την ενότητα περιγράφεται ο πηγαίος κώδικας της εφαρμογής λογισμικού η οποία ενσωματώνει όλη τη λειτουργικότητα ενός ολοκληρωμένου συστήματος συστάσεων σύμφωνα με την προσέγγιση της παρούσας εργασίας. Με τη βοήθεια της γλώσσας προγραμματισμού *Java* αναπτύχθηκε μία *standalone* εφαρμογή. Όλες οι λεπτομέρειες σχετικά με το σύστημα ανάπτυξης, τις βιβλιοθήκες, τις εκδόσεις και τα υποσυστήματα που χρησιμοποιήθηκαν περιγράφονται στην ενότητα 2.6 του κεφαλαίου 2. Εδώ θα παρουσιαστούν τα μέρη του συστήματος λογισμικού με τη βοήθεια ενός *component* διαγράμματος, τα πακέτα λογισμικού που ομαδοποιούν τις κλάσεις της *java* και θα δοθούν τα διαγράμματα κλάσεων για την περιγραφή του λογισμικού. Σε αυτή την ενότητα δίνεται μια συνοπτική περιγραφή του λογισμικού με τη βοήθεια των σχετικών διαγραμμάτων, ενώ μια εκτενέστερη περιγραφή των κλάσεων και της λειτουργικότητάς τους περιλαμβάνεται στο παράρτημα Α.

Στην εικόνα 5.2(α) παρουσιάζεται το *Component* διάγραμμα του συστήματος λογισμικού. Το διάγραμμα αυτό περιλαμβάνει την κυρίως εφαρμογή (*Java Application*) η οποία ενσωματώνει το λογισμικό που αναπτύχθηκε (*Package Recommendation System*) καθώς και τις βιβλιοθήκες *CPLEX* και *Apache Mahout* για την επίλυση του προβλήματος γραμμικού προγραμματισμού και το συνεργατικό φιλτράρισμα αντίστοιχα. Η εφαρμογή επικοινωνεί με το εξωτερικό σύστημα επίλυσης

των προβλημάτων γραμμικού προγραμματισμού (*CPLEX Solver*), με το σύνολο δεδομένων (*Dataset*), εξάγει τα αρχεία των αναφορών (*Reports*) με τις συστάσεις και παρέχει ένα γραφικό περιβάλλον για τη διαχείρισή της από τους χρήστες. Όλα τα προαναφερόμενα μέρη που απαρτίζουν το σύστημα λογισμικού απεικονίζονται στο διάγραμμα της εικόνας 5.2(α).



Εικόνα 5.2 (α): Το *Component* διάγραμμα του συστήματος λογισμικού.

Στο λογισμικό που αναπτύχθηκε οι κλάσεις της *java* ομαδοποιούνται σε πακέτα λογισμικού σύμφωνα με τη λειτουργικότητά τους. Τα πακέτα λογισμικού τα οποία απεικονίζονται στην εικόνα 5.2(β) είναι τα εξής:

entityBasicClasses: Το πακέτο περιλαμβάνει τις κλάσεις των βασικών οντοτήτων.

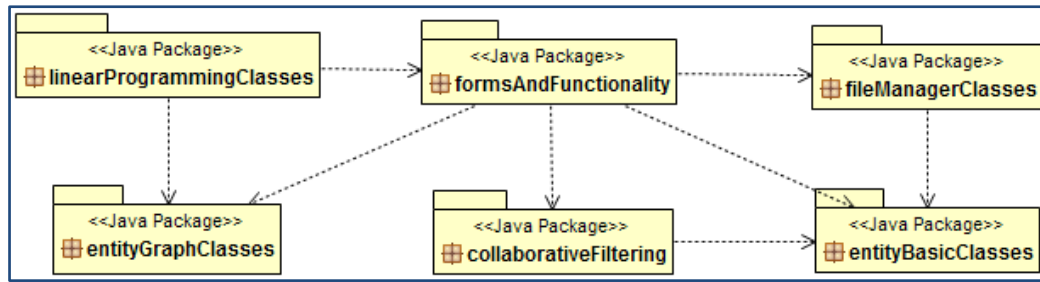
entityGraphClasses: Το πακέτο περιλαμβάνει τις κλάσεις οντοτήτων που υλοποιούν το γράφημα μοντελοποίησης του προβλήματος.

fileManagerClasses: Το πακέτο περιλαμβάνει τις κλάσεις διαχείρισης των αρχείων του συνόλου δεδομένων.

collaborativeFiltering: Το πακέτο περιλαμβάνει τις κλάσεις για το συνεργατικό φιλτράρισμα.

linearProgrammingClasses: Το πακέτο περιλαμβάνει τις κλάσεις για την επίλυση του προβλήματος γραμμικού προγραμματισμού.

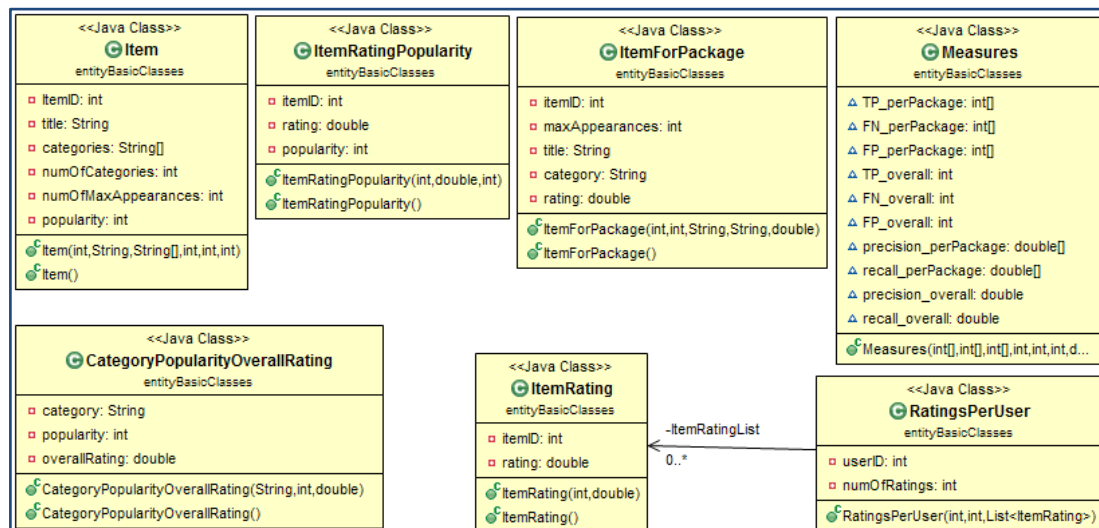
formsAndFunctionality: Το πακέτο περιλαμβάνει τις κλάσεις που υλοποιούν το γραφικό περιβάλλον και την απαραίτητη λειτουργικότητα που σχετίζεται με αυτό.



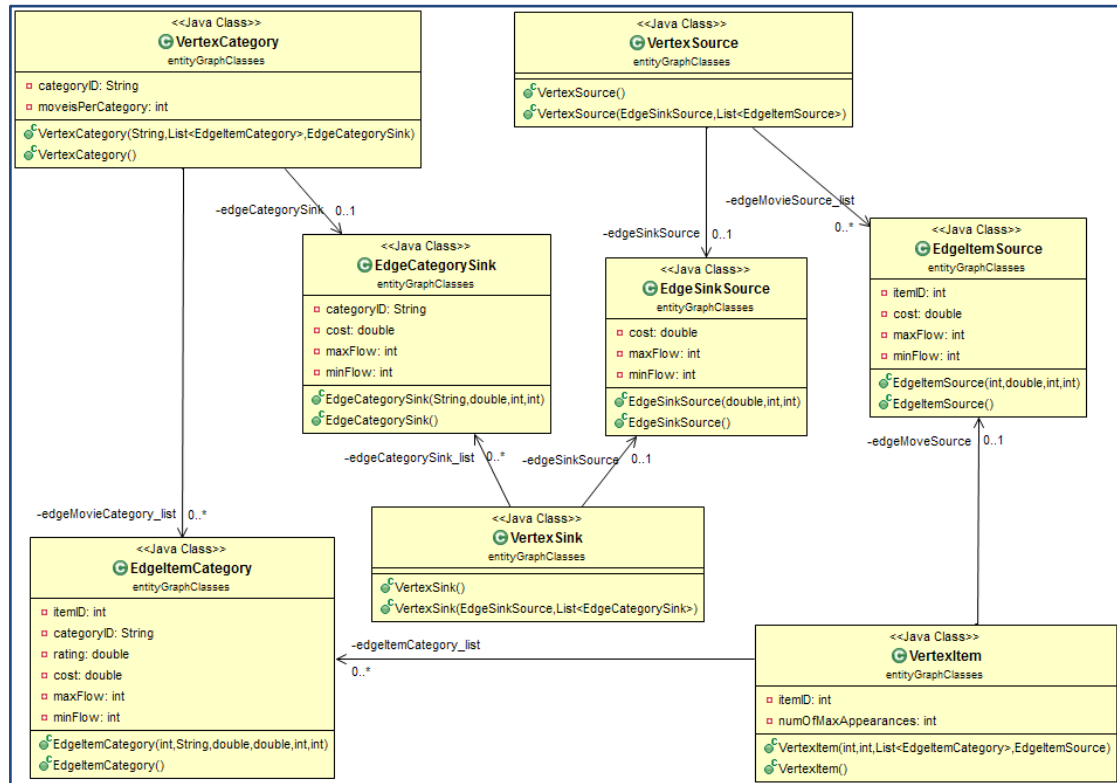
Εικόνα 5.2 (β): Το διάγραμμα των πακέτων λογισμικού της java εφαρμογής.

Επειδή είναι αδύνατη η παρουσίαση του πλήρους διαγράμματος κλάσεων για ολόκληρη την εφαρμογή λογισμικού λόγω του μεγάλου μεγέθους του, ακολουθεί η παρουσίαση των επιμέρους διαγραμμάτων κλάσεων για κάθε πακέτο λογισμικού ξεχωριστά. Να σημειωθεί ότι στα διαγράμματα κλάσεων που ακολουθούν, για οικονομία χώρου, δεν περιλαμβάνονται οι μέθοδοι *get* και *set* που αφορούν στην ανάκτηση και στην ανάθεση τιμών στα πεδία των κλάσεων αλλά αφήνεται να εννοηθεί ότι αυτές οι μέθοδοι έχουν υλοποιηθεί όπου απαιτείται. Μετά την παρουσίαση των επιμέρους διαγραμμάτων των κλάσεων παρουσιάζεται και το πλήρες διάγραμμα με όλες τις κλάσεις του λογισμικού μετά τις απαραίτητες αφαιρέσεις που έχουν γίνει για να μπορέσει να παρουσιαστεί το διάγραμμά αυτό.

Στην εικόνα 5.2(γ) παρουσιάζεται το διάγραμμα κλάσεων για τις κλάσεις που περιλαμβάνονται στο πακέτο λογισμικού *entityBasicClasses*. Οι κλάσεις αυτές υλοποιούν τις βασικές οντότητες που εμπλέκονται στη μοντελοποίηση του προβλήματος.

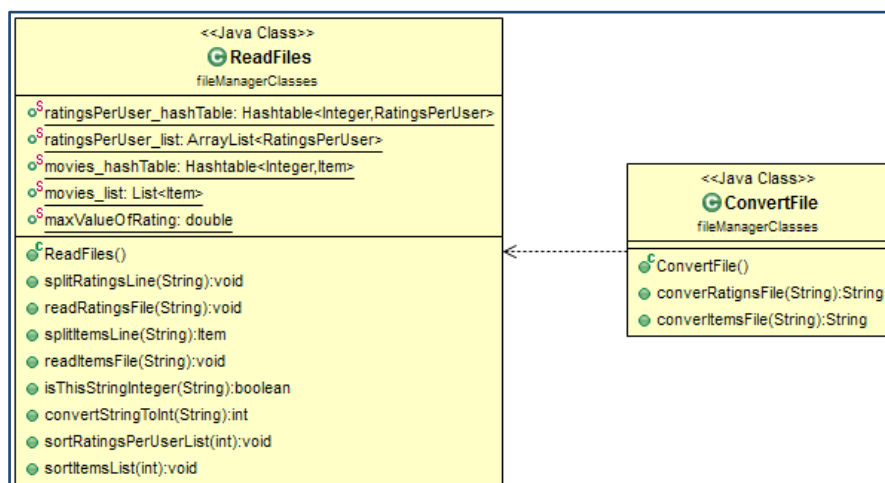
Εικόνα 5.2(γ): Το διάγραμμα κλάσεων του πακέτου *entityBasicClasses*.

Στην εικόνα 5.2(δ) παρουσιάζεται το διάγραμμα κλάσεων για τις κλάσεις που περιλαμβάνονται στο πακέτο λογισμικού *entityGraphClasses*. Οι κλάσεις αυτές σχετίζονται με τις οντότητες που αφορούν στο γράφημα μοντελοποίησης του προβλήματος των συστάσεων.



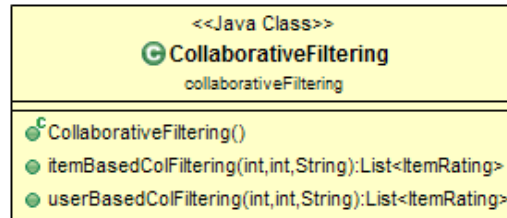
Εικόνα 5.2 (δ): Το διάγραμμα των κλάσεων του πακέτου λογισμικού *entityGraphClasses*.

Στην εικόνα 5.2(ε) παρουσιάζεται το διάγραμμα κλάσεων για τις κλάσεις που περιλαμβάνονται στο πακέτο λογισμικού *fileManagerClasses*. Οι κλάσεις αυτές διαχειρίζονται τα αρχεία του συνόλου δεδομένων.



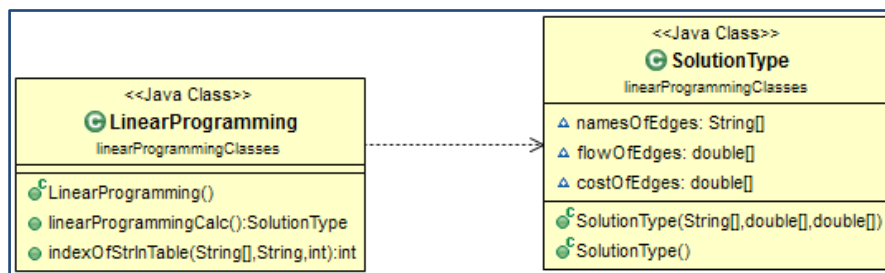
Εικόνα 5.2 (ε): Το διάγραμμα των κλάσεων του πακέτου λογισμικού *fileManagerClasses*.

Στην εικόνα 5.2(στ) παρουσιάζεται το διάγραμμα της κλάσης του πακέτο λογισμικού *collaborativeFiltering*. Η κλάση αυτή εκτελεί το συνεργατικό φιλτράρισμα.



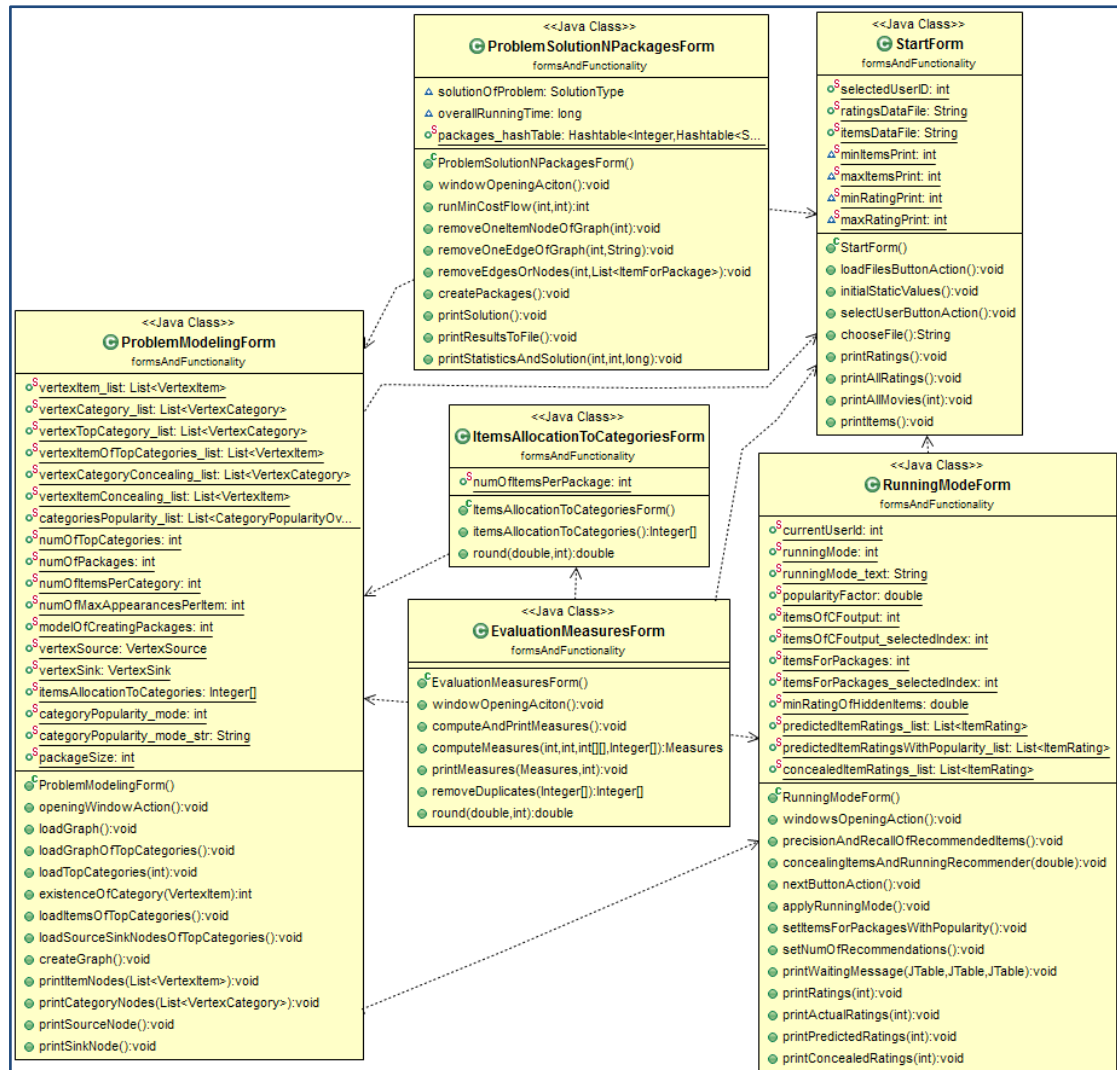
Εικόνα 5.2 (στ): Το διάγραμμα της κλάσης *CollaborativeFiltering*.

Στην εικόνα 5.2(ζ) παρουσιάζεται το διάγραμμα κλάσεων για τις κλάσεις που περιλαμβάνονται στο πακέτο λογισμικού *linearProgrammingClasses*. Οι κλάσεις αυτές εκτελούν τις λειτουργίες για την επίλυση του προβλήματος γραμμικού προγραμματισμού.



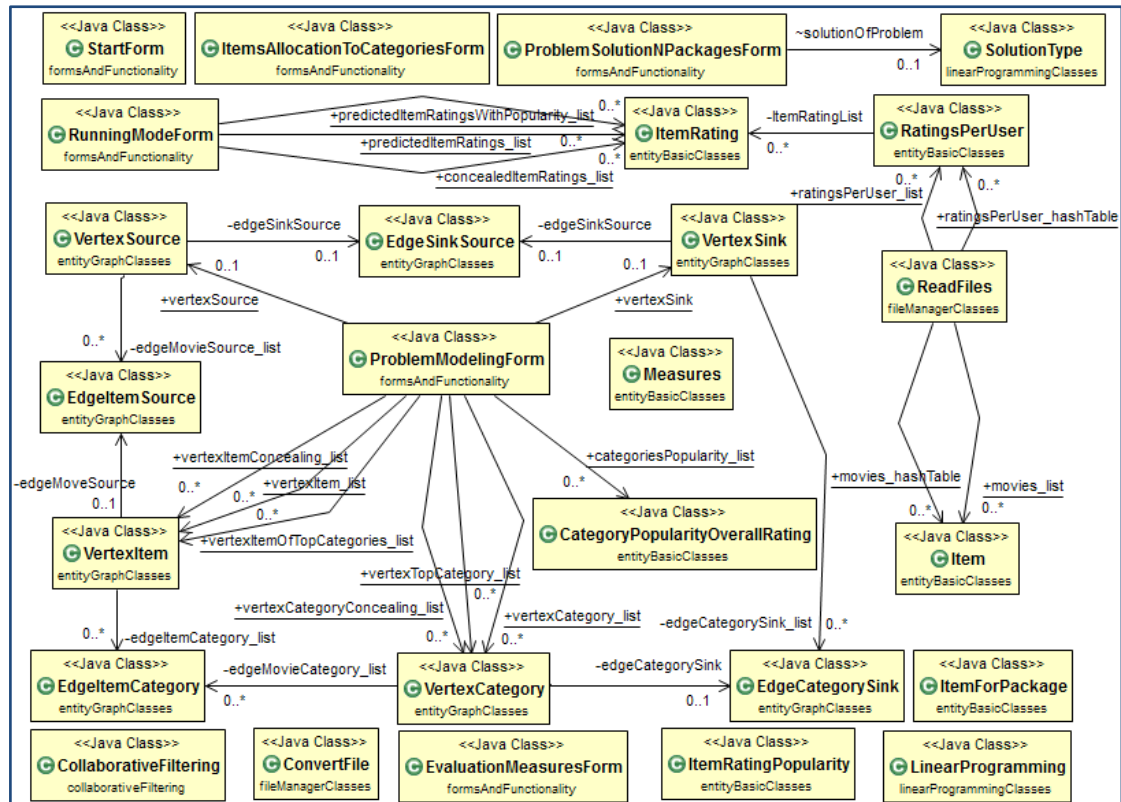
Εικόνα 5.2 (ζ): Το διάγραμμα κλάσεων του πακέτου λογισμικού *linearProgrammingClasses*.

Στο σχήμα 5.2(η) παρουσιάζεται το διάγραμμα κλάσεων για τις κλάσεις που περιλαμβάνονται στο πακέτο λογισμικού *formsAndFunctionality*. Οι κλάσεις αυτές υλοποιούν το γραφικό περιβάλλον της εφαρμογής και όλη την απαραίτητη λειτουργικότητα που ενσωματώνεται και εκτελείται από αυτό. Στο παρακάτω διάγραμμα κλάσεων δεν περιλαμβάνονται τα πεδία και οι μέθοδοι που αφορούν στην υλοποίηση του γραφικού περιβάλλοντος παρά μόνο τα πεδία και οι μέθοδοι που αφορούν στη βασική λειτουργικότητα των κλάσεων.



Εικόνα 5.2 (η): Το διάγραμμα των κλάσεων του πακέτου λογισμικού *formsAndFunctionality*.

Στο σχήμα 5.2(θ) παρουσιάζεται το πλήρες διάγραμμα όλων των κλάσεων του λογισμικού. Αυτό το διάγραμμα παρουσιάζεται με έναν συνοπτικό τρόπο καθώς για οικονομία χώρου δεν περιλαμβάνονται οι μέθοδοι, τα πεδία και οι εξαρτήσεις των κλάσεων. Στο διάγραμμα τις εικόνες 5.2(θ) μπορούμε να διακρίνουμε όλες τις κλάσεις της *java* που υλοποιούν την εφαρμογή, τα πακέτα στα οποία ανήκουν και τις μεταξύ τους σχέσεις.



Εικόνα 5.2 (θ): Το συνολικό διάγραμμα των κλάσεων του λογισμικού.

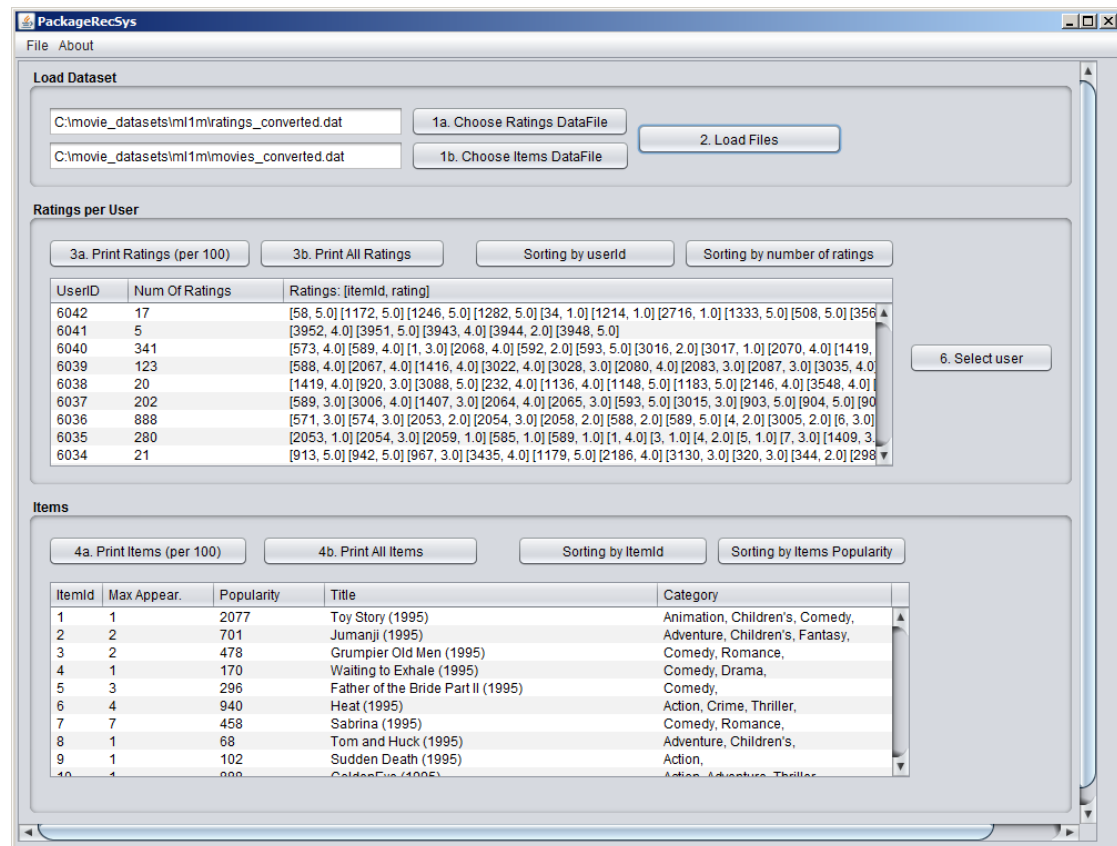
5.3 Γραφικό Περιβάλλον

Σε αυτή την ενότητα παρουσιάζεται το γραφικό περιβάλλον της εφαρμογής λογισμικού και δίνονται όλες οι απαραίτητες οδηγίες για τον τρόπο χρήσης. Η εφαρμογή λογισμικού που αναπτύχθηκε περιλαμβάνει ένα εύχρηστο γραφικό περιβάλλον χρήστη μέσα από το οποίο εκτελούνται όλες οι λειτουργίες του αλγόριθμου με σκοπό την εξαγωγή των N-καλύτερων πακέτων αντικειμένων. Το γραφικό περιβάλλον χρήστη περιλαμβάνει την αρχική οθόνη, την οθόνη καθορισμού του τρόπου λειτουργίας του προγράμματος, την οθόνη επιλογής των παραμέτρων του προβλήματος, την οθόνη αναλογικής κατανομής των αντικειμένων στις κατηγορίες, την οθόνη των προτάσεων των N-καλύτερων πακέτων αντικειμένων και την οθόνη αξιολόγησης. Ακολουθούν αναλυτικές πληροφορίες για τη λειτουργικότητα και τον τρόπο χρήσης του γραφικού περιβάλλοντος.

5.3.1 Αρχική Οθόνη

Η αρχική οθόνη της εφαρμογής παρουσιάζεται στην εικόνα 5.3.1(α). Μέσα από αυτήν υπάρχει η δυνατότητα φόρτωσης των αρχείων του συνόλου δεδομένων και παρουσίασης των δεδομένων αυτών στους σχετικούς πίνακες του παραθύρου. Στην

εικόνα 5.3.1(α) παρουσιάζεται ένα στιγμιότυπο του παραθύρου της εφαρμογής, στο οποίο έχει φορτωθεί το σύνολο δεδομένων και οι σχετικές εγγραφές φαίνονται στους δύο πίνακες της οθόνης. Για τη φόρτωση των κατάλληλων αρχείων δεδομένων υπάρχει η δυνατότητα της μετατροπής τους σε κατάλληλο τύπο. Επίσης, μέσα από το παράθυρο αυτό επιλέγεται ο χρήστης για τον οποίο θα προταθούν τα Ν-καλύτερα πακέτα αντικειμένων. Ακολουθεί ο τρόπος χρήσης και η περιγραφή των στοιχείων και των λειτουργιών του αρχικού παραθύρου της εφαρμογής.



Εικόνα 5.3.1 (α): Η αρχική οθόνη της εφαρμογής.

Λειτουργίες και τρόπος χρήσης της αρχικής οθόνης

Παρακάτω περιγράφονται οι λειτουργίες και ο τρόπος χρήσης των στοιχείων της αρχικής οθόνης που φαίνεται στην εικόνα 5.3.1(α).

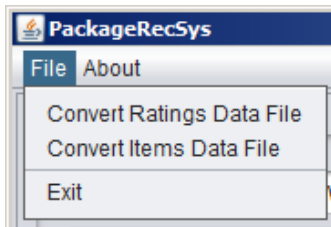
Η μπάρα του μενού περιλαμβάνει τα εξής:

File: Μετά το πάτημα του *File* εμφανίζονται οι επιλογές της εικόνας 5.3.1(β) που περιλαμβάνουν:

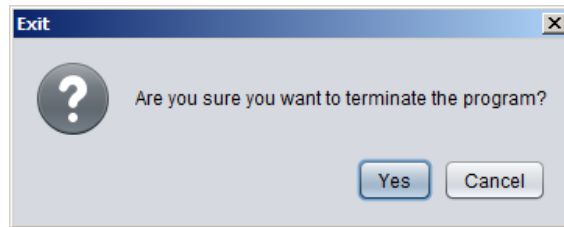
Convert Ratings Data File: Μετατροπή του περιεχομένου του αρχείου των ratings με αντικατάσταση των δύο άνω κάτω τελειών (::) με κόμμα (,).

Convert Items Data File: Μετατροπή του περιεχομένου του αρχείου των Items με αντικατάσταση των δύο άνω κάτω τελειών (::) με κόμμα (,).

Exit: Τερματισμός της εφαρμογής. Στην εικόνα 5.3.1(γ) παρουσιάζεται το μήνυμα επιβεβαίωσης για τον τερματισμό της εφαρμογής. Το μήνυμα αυτό εμφανίζεται κάθε φορά που επιλέγεται το *exit* από οποιοδήποτε παράθυρο της εφαρμογής.

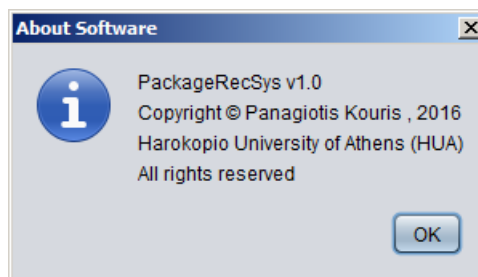


Εικόνα 5.3.1 (β): Το μενού file



Εικόνα 5.3.1 (γ): Μήνυμα τερματισμού.

About: Μετά το πάτημα του *About* το οποίο βρίσκεται στην μπάρα του μενού του παραθύρου εμφανίζεται το μήνυμα της εικόνας 5.3.1(δ) που περιλαμβάνει πληροφορίες σχετικές με το λογισμικό. Το ίδιο μήνυμα εμφανίζεται με το πάτημα του *About* από οποιοδήποτε παράθυρο της εφαρμογής.



Εικόνα 5.3.1 (δ): Το μήνυμα του μενού About.

Οι περιοχές του παραθύρου:

Load dataset: Περιλαμβάνει τις λειτουργίες για την φόρτωση των αρχείων του συνόλου δεδομένων.

Ratings per user: Στον πίνακα του πλαισίου αυτού εκτυπώνονται τα ratings για κάθε χρήστη.

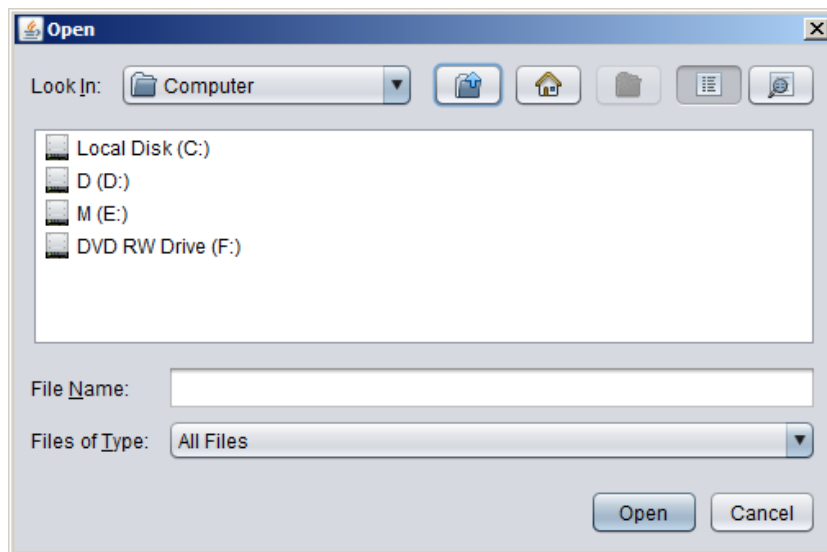
Items: Στον πίνακα του πλαισίου αυτού εκτυπώνονται οι πληροφορίες για τα αντικείμενα.

Τα κουμπιά του παραθύρου:

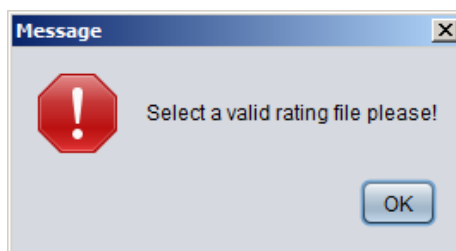
Choose Ratings Datafile: Επιλογή του αρχείου με τα δεδομένα των ratings. Με το πάτημα του κουμπιού εμφανίζεται το παράθυρο της εικόνας 5.3.1(ε) για την επιλογή του αρχείου των δεδομένων.

Choose Items Datafile: Επιλογή του αρχείου με τα δεδομένα των αντικειμένων (*items*). Με το πάτημα του κουμπιού εμφανίζεται το παράθυρο της εικόνας 5.3.1(ε) για την επιλογή του αρχείου των δεδομένων.

Load Datafiles: Φόρτωση των δύο αρχείων του συνόλου δεδομένων σε κατάλληλες δομές δεδομένων του προγράμματος. Αν δεν έχει επιλεγεί κατάλληλο αρχείο δεδομένων εμφανίζεται το μήνυμα της εικόνας 5.3.1(στ) που μπορεί να αφορά είτε στο αρχείο των *ratings* είτε στο αρχείο των *items*. Μετά τη φόρτωση των δεδομένων, αυτά εμφανίζονται στους δύο πίνακες του παραθύρου όπως φαίνονται στην εικόνα 1.3.5(α).



Εικόνα 5.3.1 (ε): Το παράθυρο επιλογής των αρχείων των δεδομένων.



Εικόνα 5.3.1 (στ): Μήνυμα λανθασμένης επιλογής αρχείου δεδομένων.

Print Ratings (per 100): Εκτύπωση των δεδομένων των *ratings* στον πίνακα του παραθύρου της εικόνας 5.3.1(α) ανά 100 εγγραφές. Αυτή είναι η προεπιλεγμένη επιλογή για την εκτύπωση μετά τη φόρτωση των δεδομένων, καθώς η εκτύπωση όλων των εγγραφών στην περίπτωση που τα *ratings* είναι της τάξης του ενός εκατομμυρίου ή περισσότερα απαιτεί αρκετό χρόνο.

Print All Rating: Εκτύπωση όλων των εγγραφών των δεδομένων των *ratings* στον αντίστοιχο πίνακα του παραθύρου της εικόνας 5.3.1(α).

Sorting by userId: Αύξουσα ταξινόμηση των εγγραφών των *ratings* σύμφωνα με το *userId*.

Sorting by number of ratings: Φθίνουσα ταξινόμηση των εγγραφών των *ratings* σύμφωνα με τον αριθμό των *ratings* που έχει κάθε χρήστης.

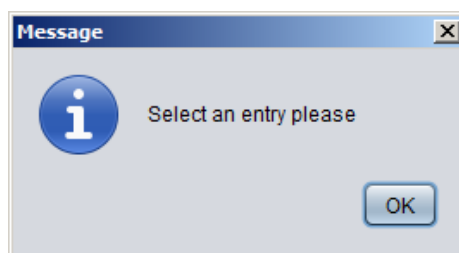
Print Items (per 100): Εκτύπωση των δεδομένων των *items* στον αντίστοιχο πίνακα του παραθύρου της εικόνας 5.3.1(α) ανά 100 εγγραφές.

Print All Items: Εκτύπωση όλων των εγγραφών των δεδομένων για τα αντικείμενα (*items*) στον αντίστοιχο πίνακα του παραθύρου της εικόνας 5.3.1(α). Αυτή είναι η προεπιλεγμένη επιλογή για την εκτύπωση μετά τη φόρτωση των δεδομένων, καθώς η εκτύπωση όλων των εγγραφών για τα αντικείμενα γίνεται σε λογικό χρόνο αφού το πλήθος τους είναι ίσο με μερικές χιλιάδες.

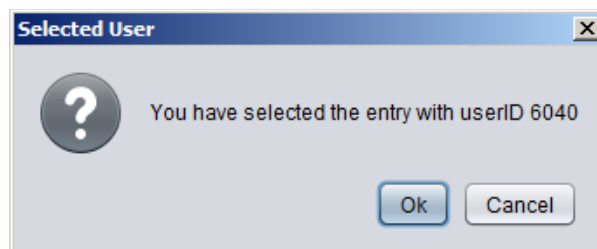
Sorting by itemId: Αύξουσα ταξινόμηση των εγγραφών των *items* σύμφωνα με το *itemId*.

Sorting by Items Popularity: Φθίνουσα ταξινόμηση των εγγραφών των *items* σύμφωνα με τον αριθμό των *ratings* που έχει δεχτεί κάθε αντικείμενο από χρήστες.

Select User: Επιλογή του χρήστη για τον οποίο θα προταθούν τα N-καλύτερα πακέτα αντικειμένων. Σε περίπτωση που πατηθεί το κουμπί χωρίς να επιλεγεί κάποιος χρήστης εμφανίζεται το μήνυμα της εικόνας 5.3.1(ζ). Μετά την επιτυχημένη επιλογή χρήστη εμφανίζεται το μήνυμα επιβεβαίωσης της εικόνας 5.3.1(η) και με το πάτημα του κουμπιού OK το πρόγραμμα οδηγείται στο επόμενο παράθυρο που αφορά στην επιλογή του τρόπου λειτουργίας της εφαρμογής.



Εικόνα 5.3.1 (ζ): Μήνυμα προτροπής για την επιλογή μιας εγγραφής.



Εικόνα 5.3.1 (η): Μήνυμα επιβεβαίωσης για την επιλογή του συγκεκριμένου χρήστη.

5.3.2 Οθόνη Επιλογής του Τρόπου Λειτουργίας

Η οθόνη για την επιλογή του τρόπου λειτουργίας του προγράμματος παρουσιάζεται στην εικόνα 5.3.2(α). Η οθόνη αυτή περιλαμβάνει τις επιλογές για τον καθορισμό του τρόπου λειτουργίας της εφαρμογής και την εκτέλεση του συνεργατικού φιλτραρίσματος με την παρουσίαση των αποτελεσμάτων στους σχετικούς πίνακες του παραθύρου. Η εικόνα 5.3.2(α) παρουσιάζει ένα στιγμιότυπο του παραθύρου σε λειτουργία αξιολόγησης μετά την εκτέλεση του συνεργατικού φιλτραρίσματος. Ακολουθεί ο τρόπος χρήσης και η επεξήγηση των στοιχείων και των λειτουργιών του παραθύρου.

The screenshot shows the PackageRecSys application window. The 'Running Mode' section on the left contains several dropdown menus and input fields for configuring the evaluation process. The 'Statistics' section on the right displays performance metrics. The 'Actual Ratings of User' section shows a table of user ratings.

Running Mode Configuration:

- Mode: Evaluation by Hiding 40% of items
- Items of CF output: 100%
- Items for packages: 500
- Item Popularity factor: 5
- Minimum rating of hidden items: 2

Statistics - measures of predicted items in evaluation mode:

- Selected Userid: 678
- CF Running Time: 41.67 sec
- MAE = 0.756 RMSE = 0.899
- TP = 143 FN = 256 FP = 357
- Precision = 0.286 Recall = 0.358 F = 0.318

Actual Ratings of User:

Num Of Ratings	Actual Ratings: [movieID, rating]
10	[1, 5.0] [2, 4.0] [4, 3.0] [10, 4.0] [12, 2.0] [15, 3.0] [16, 3.0] [24, 4.0] [29, 4.0] [31, 4.0]
20	[32, 4.0] [34, 5.0] [36, 3.0] [44, 4.0] [47, 5.0] [48, 4.0] [50, 5.0] [57, 3.0] [65, 2.0] [92, 3.0]
30	[110, 4.0] [125, 3.0] [137, 4.0] [141, 5.0] [144, 2.0] [145, 4.0] [150, 4.0] [151, 3.0] [153, 3.0] [158, 4.0]
40	[160, 3.0] [161, 4.0] [165, 3.0] [171, 5.0] [173, 2.0] [177, 4.0] [178, 4.0] [180, 4.0] [181, 4.0] [185, 3.0]
50	[186, 3.0] [187, 2.0] [188, 3.0] [193, 4.0] [195, 3.0] [196, 3.0] [198, 3.0] [202, 3.0] [203, 4.0] [208, 2.0]
60	[218, 4.0] [222, 4.0] [223, 4.0] [224, 3.0] [230, 4.0] [234, 3.0] [235, 3.0] [236, 3.0] [238, 3.0] [253, 4.0]
70	[260, 5.0] [261, 3.0] [266, 3.0] [273, 5.0] [275, 3.0] [277, 3.0] [282, 3.0] [288, 4.0] [296, 2.0] [299, 4.0]
80	[303, 4.0] [316, 4.0] [317, 4.0] [318, 5.0] [322, 2.0] [324, 4.0] [327, 2.0] [329, 4.0] [338, 3.0] [339, 4.0]
90	[342, 4.0] [345, 5.0] [348, 3.0] [351, 4.0] [353, 4.0] [355, 3.0] [356, 4.0] [357, 3.0] [364, 4.0] [366, 4.0]

Predicted Ratings of User:

Num Of Ratings	Predicted Ratings: [itemid, rating]
10	[6, 3.39] [11, 3.47] [17, 3.37] [21, 3.67] [25, 3.45] [28, 2.92] [39, 3.68] [53, 2.84] [58, 3.15] [62, 3.14]
20	[110, 4.38] [111, 3.63] [134, 2.89] [162, 3.12] [199, 2.89] [213, 2.89] [223, 3.71] [232, 3.06] [235, 3.4] [246, 3.29]
30	[247, 3.11] [265, 3.34] [272, 3.06] [293, 3.39] [296, 4.22] [300, 3.26] [306, 3.04] [307, 2.96] [308, 2.93] [314, 3.0]
40	[337, 3.26] [349, 3.47] [363, 2.85] [398, 2.91] [402, 3.33] [406, 2.89] [439, 2.89] [440, 3.41] [446, 2.98] [454, 3.32]
50	[457, 4.1] [474, 3.42] [480, 4.52] [497, 3.24] [509, 3.2] [515, 3.12] [529, 3.28] [531, 3.01] [545, 2.9] [549, 2.92]
60	[557, 2.91] [560, 2.85] [562, 3.1] [587, 3.51] [588, 3.68] [592, 3.74] [594, 3.32] [597, 3.47] [602, 2.85] [608, 4.44]
70	[615, 2.87] [624, 3.17] [648, 3.77] [658, 2.86] [684, 2.88] [720, 3.09] [728, 2.94] [729, 2.86] [733, 3.65] [736, 3.49]
80	[745, 3.25] [750, 3.72] [757, 2.82] [759, 2.86] [776, 2.86] [778, 3.29] [787, 2.81] [791, 2.97] [800, 3.23] [823, 2.94]
90	[826, 2.96] [838, 3.13] [858, 4.26] [872, 2.84] [889, 2.84] [898, 3.23] [899, 3.33] [900, 3.05] [902, 3.27] [903, 3.42]
100	[904, 3.52] [905, 3.06] [908, 3.69] [909, 3.09] [910, 3.37] [911, 3.01] [913, 3.51] [914, 3.25] [915, 3.05] [916, 3.1]

Hidden Ratings of User (Evaluation Mode):

Num Of Ratings	Concealed Ratings: [itemid, rating]
10	[2, 4.0] [110, 4.0] [203, 4.0] [218, 4.0] [222, 4.0] [223, 4.0] [234, 3.0] [235, 3.0] [236, 3.0] [253, 4.0]

Εικόνα 5.3.2 (α): Η οθόνη για την επιλογή του τρόπου λειτουργίας της εφαρμογής και την εκτέλεση του συνεργατικού φιλτραρίσματος.

Λειτουργίες και τρόπος χρήσης του παραθύρου

Παρακάτω περιγράφονται οι λειτουργίες και ο τρόπος χρήσης των στοιχείων και των επιλογών της οθόνης για την επιλογή του τρόπου λειτουργίας και της εκτέλεσης του συνεργατικού φιλτραρίσματος η οποία φαίνεται στην εικόνα 5.3.2(α).

Η μπάρα του μενού περιλαμβάνει:

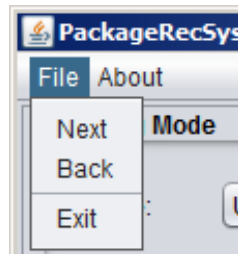
File: Μετά το πάτημα του *File* εμφανίζονται οι επιλογές της εικόνας 5.3.2(β) που περιλαμβάνουν:

Next: Μετάβαση στο επόμενο παράθυρο που περιλαμβάνει την επιλογή των παραμέτρων για τη μοντελοποίηση του προβλήματος. Η επιλογή του *Next* δίνεται και από το κουμπί που υπάρχει σε αυτό το παράθυρο.

Back: Μετάβαση στο αρχικό παράθυρο που περιλαμβάνει τη φόρτωση των δεδομένων και την επιλογή του τρέχοντα χρήστη. Η επιλογή του *Back* δίνεται και από το κουμπί που υπάρχει σε αυτό το παράθυρο.

Exit: Τερματισμός της εφαρμογής. Στην εικόνα 5.3.1(γ) παρουσιάζεται το μήνυμα επιβεβαίωσης για τον τερματισμό της εφαρμογής.

About: Εμφάνιση μηνύματος με πληροφορίες σχετικές με την εφαρμογή.



Εικόνα 5.3.2 (β): Το μενού file.

Οι περιοχές του παραθύρου:

Running mode: Περιλαμβάνει τις παραμέτρους βάσει των οποίων θα λειτουργήσει η εφαρμογή.

Statistics: Στο πλαίσιο με τα στατιστικά αναγράφονται οι τιμές κάποιων σημαντικών δεικτών αξιολόγησης του αλγόριθμου που χρησιμοποιούνται κυρίως στη λειτουργία αξιολόγησης. Οι δείκτες αυτοί είναι οι εξής:

CF Running Time: Χρόνος εκτέλεσης του συνεργατικού φιλτραρίσματος

MAE (Mean Absolute Error): Μέσο απόλυτο σφάλμα των αντικειμένων που προορίζονται για τον σχηματισμό των προτεινόμενων πακέτων σε σχέση με τα κρυμμένα αντικείμενα.

RMSE (Root Mean Squared Error): Μέσο τετραγωνικό σφάλμα των αντικειμένων που προορίζονται για τον σχηματισμό των προτεινόμενων πακέτων σε σχέση με τα κρυμμένα αντικείμενα.

TP (True Positive): Τα σωστά επιλεγμένα αντικείμενα του συνόλου των αντικειμένων που προορίζονται για τον σχηματισμό των προτεινόμενων πακέτων σε σχέση με τα κρυμμένα αντικείμενα.

FN (False Negative): Τα λανθασμένα μη επιλεγμένα αντικείμενα του συνόλου των αντικειμένων που προορίζονται για τον σχηματισμό των προτεινόμενων πακέτων σε σχέση με τα κρυμμένα αντικείμενα.

FP (False Positive): Τα λανθασμένα επιλεγμένα αντικείμενα του συνόλου των αντικειμένων που προορίζονται για τον σχηματισμό των προτεινόμενων πακέτων σε σχέση με τα κρυμμένα αντικείμενα.

Precision: Ο υπολογισμός του δείκτη *Precision* για το σύνολο των αντικειμένων που προορίζονται για τον σχηματισμό των προτεινόμενων πακέτων σε σχέση με τα κρυμμένα αντικείμενα.

Recall: Ο υπολογισμός του δείκτη *Recall* για το σύνολο των αντικειμένων που προορίζονται για τον σχηματισμό των προτεινόμενων πακέτων σε σχέση με τα κρυμμένα αντικείμενα.

F-measure: Ο αρμονικός μέσος (*harmonic mean*) του *Recall* και του *Precision* που δίνεται από τον τύπο:

$$F - measure = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

Actual ratings of user: Ο πίνακας του πλαισίου αυτού περιλαμβάνει τα αντικείμενα που ο χρήστης έχει βαθμολογήσει.

Predicted ratings: Ο πίνακας του πλαισίου αυτού περιλαμβάνει τα εκτιμώμενα αντικείμενα που έχουν παραχθεί από το συνεργατικό φιλτράρισμα.

Hidden ratings: Ο πίνακας του πλαισίου αυτού περιλαμβάνει τα κρυμμένα αντικείμενα του χρήστη για τη λειτουργία αξιολόγησης.

Οι παράμετροι για τον καθορισμό της λειτουργίας της εφαρμογής είναι οι εξής:

Mode: Επιλογή του τρόπου λειτουργίας της εφαρμογής μεταξύ των ακόλουθων επιλογών:

Using actual ratings: Λειτουργία με χρήση των αντικειμένων που περιλαμβάνονται στο σύνολο των προτιμήσεων του χρήστη χωρίς την εκτέλεση του συνεργατικού φιλτραρίσματος. Αυτή η λειτουργία δεν έχει κάποια σημαντική χρησιμότητα αλλά μπορεί να χρησιμοποιηθεί για περιπτώσεις ελέγχου της λειτουργίας της εφαρμογής παραλείποντας το βήμα του συνεργατικού φιλτραρίσματος.

Using predicted ratings: Κανονική λειτουργία της εφαρμογής με εκτέλεση του συνεργατικού φιλτραρίσματος. Η λειτουργία αυτή είναι η πιο σημαντική και στοχεύει στην εξαγωγή των N-καλύτερων προτεινόμενων πακέτων αντικειμένων που περιλαμβάνουν νέα για τον χρήστη αντικείμενα.

Evaluation by Hiding $p\%$ of items: Λειτουργία αξιολόγησης με την απόκρυψη του $p\%$ των αντικειμένων του τρέχοντα χρήστη. Εκτελείται το συνεργατικό φιλτράρισμα λαμβάνοντας υπόψη το υπόλοιπο $(100-p)\%$ των αντικειμένων του τρέχοντα χρήστη. Όπου ο αριθμός p παίρνει τιμές από 10 έως 90 με βήμα 10, $p \in \{10, 20, 30, 40, 50, 60, 70, 80, 90\}$. Η λειτουργία αξιολόγησης στοχεύει στη σύγκριση των αντικειμένων των προτεινόμενων πακέτων με το σύνολο των κρυμμένων αντικειμένων για την εξαγωγή συμπερασμάτων σχετικά με την ακρίβεια που το σύστημα πετυχαίνει. Περισσότερες λεπτομέρειες για την αξιολόγηση της εφαρμογής αναφέρονται στο κεφάλαιο 6.

Items of CF output: Επιλογή του αριθμού των αντικειμένων με τις υψηλότερες εκτιμώμενες βαθμολογίες προτίμησης που το συνεργατικό φιλτράρισμα εξάγει στην έξοδό του.

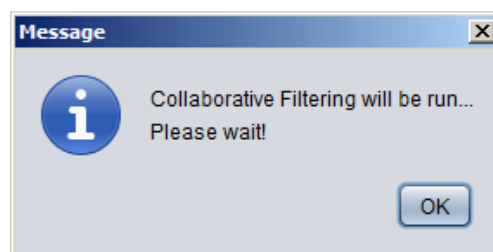
Items for packages: Επιλογή του αριθμού των αντικειμένων με τις υψηλότερες εκτιμώμενες βαθμολογίες προτίμησης που ο αλγόριθμος χρησιμοποιεί για τον σχηματισμό των N-καλύτερων πακέτων αντικειμένων.

Item popularity factor: Επιλογή της τιμής του συντελεστή δημοφιλίας των αντικειμένων. Ο σκοπός και η χρήση του συντελεστή της δημοφιλίας των αντικειμένων περιγράφεται στο κεφάλαιο 3.

Minimum rating of hidden items: Επιλογή της ελάχιστης βαθμολογίας προτίμησης των κρυμμένων αντικειμένων.

Τα κουμπιά του παραθύρου:

Apply: Εφαρμογή των επιλεγμένων παραμέτρων καθορισμού του τρόπου λειτουργίας της εφαρμογής. Σε περίπτωση που απαιτείται η λειτουργία του συνεργατικού φιλτραρίσματος εμφανίζεται το μήνυμα της εικόνας 5.3.2(γ) που ενημερώνει για την εκτέλεση του.



Εικόνα 5.3.2 (γ): Μήνυμα ενημέρωσης για την εκτέλεση του συνεργατικού φιλτραρίσματος.

Sorting by itemId: Αύξουσα ταξινόμηση των αντικειμένων σύμφωνα με το itemId και η εκτύπωση τους στους αντίστοιχους πίνακες του παραθύρου.

Sorting by rating: Φθίνουσα ταξινόμηση των αντικειμένων σύμφωνα με την τιμή rating και η εκτύπωση τους στους αντίστοιχους πίνακες του παραθύρου.

Μετά το πάτημα του κουμπιού *Apply* και την εκτέλεση του συνεργατικού φίλτραρίσματος συμπληρώνονται οι τρεις πίνακες του παραθύρου με τα σχετικά δεδομένα όπως φαίνεται στην εικόνα 5.3.2(α). Στον πρώτο πίνακα (*Actual Ratings of User*) περιλαμβάνονται τα ratings για τα αντικείμενα των προτιμήσεων του τρέχοντα χρήστη, στον δεύτερο πίνακα (*Predicted ratings*) περιλαμβάνονται τα εκτιμώμενα ratings που προκύπτουν από το συνεργατικό φιλτράρισμα και στον τρίτο πίνακα (*Hidden ratings*) περιλαμβάνονται τα ratings για τα κρυμμένα αντικείμενα που χρησιμοποιούνται για τη διαδικασία αξιολόγησης. Ο τρίτος πίνακας συμπληρώνεται μόνο στη λειτουργία αξιολόγησης που υπάρχουν κρυμμένα αντικείμενα και ο δεύτερος πίνακας συμπληρώνεται κάθε φορά που εκτελείται το συνεργατικό φιλτράρισμα.

5.3.3 Οθόνη Επιλογής των Παραμέτρων Μοντελοποίησης

Η οθόνη για τον καθορισμό των παραμέτρων λειτουργίας του αλγόριθμου και μοντελοποίησης του προβλήματος παρουσιάζεται στην εικόνα 5.3.3. Η εικόνα αυτή παρουσιάζει ένα στιγμιότυπο του παραθύρου με φορτωμένα τα δεδομένα του γραφήματος μοντελοποίησης του προβλήματος τα οποία φαίνονται στους σχετικούς πίνακες του παραθύρου. Ακολουθεί ο τρόπος χρήσης και η επεξήγηση των στοιχείων και των λειτουργιών του παραθύρου.

Λειτουργίες και τρόπος χρήσης του παραθύρου

Παρακάτω περιγράφονται οι λειτουργίες και ο τρόπος χρήσης των στοιχείων και των επιλογών του παραθύρου το οποίο φαίνεται στην εικόνα 5.3.3.

Η μπάρα του μενού περιλαμβάνει τα εξής:

File: Μετά το πάτημα του *File* εμφανίζονται οι επιλογές:

Back: Μετάβαση στο προηγούμενο παράθυρο. Η επιλογή αυτή δίνεται και από το σχετικό κουμπί που υπάρχει στην οθόνη αυτή.

Exit: Τερματισμός της εφαρμογής.

About: Εμφάνιση μηνύματος με πληροφορίες σχετικές με την εφαρμογή.

PackageRecSys

File About

Parameters

Number of Packages (N): 2
 Number of Categories (nc): 3
 Package Size: default
 Items per Category: (ni/c): 1
 Items Allocation to Categories: [button]
 Maximum Appearances/Item: Not Taken Account
 Model of Creating Packages: 1. Removing the Edge with Maximum Rating
 Category Popularity Mode: 1. Based on the number of items per category

Selected UserID: 678 [Back]
 Running Mode: Evaluation by Hiding 40% of items

1. Load Graph
 2. Load Graph with Top-N Categories
 3. Run Min Cost Flow Algorithm

Item Nodes

no	ItemId	Item-Source [cost, maxFlow, minFlow]	Item-Category Edges: [category, rating, cost, maxFlow, minFlow]
1	1	[0.0, 1, 0]	[Animation, 4.2, 0.24, 1, 0] [Children's, 4.2, 0.24, 1, 0] [Comedy, 4.2, 0.24, 1, 0]
2	6	[0.0, 1, 0]	[Action, 3.4, 0.29, 1, 0] [Crime, 3.4, 0.29, 1, 0] [Thriller, 3.4, 0.29, 1, 0]
3	11	[0.0, 1, 0]	[Comedy, 3.5, 0.29, 1, 0] [Drama, 3.5, 0.29, 1, 0] [Romance, 3.5, 0.29, 1, 0]
4	17	[0.0, 1, 0]	[Drama, 3.4, 0.3, 1, 0] [Romance, 3.4, 0.3, 1, 0]

Category Nodes

no	Category	Items/Category	Category-Sink [cost, maxFlow, minFlow]	Category-Item Edges: [itemId, rating, cost, maxFlow, minFlow]
1	Comedy	148	[0.0, 1, 1]	[1, 4.2, 0.24, 1, 0] [11, 3.5, 0.29, 1, 0] [21, 3.7, 0.27, 1, 0] [39, 3.7, 0.27, 1, 0]
2	Drama	252	[0.0, 1, 1]	[11, 3.5, 0.29, 1, 0] [17, 3.4, 0.3, 1, 0] [21, 3.7, 0.27, 1, 0] [25, 3.4, 0.29, 1, 0]
3	Action	66	[0.0, 1, 1]	[6, 3.4, 0.29, 1, 0] [21, 3.7, 0.27, 1, 0] [110, 4.4, 0.23, 1, 0] [349, 3.5, 0.29, 1, 0]

Source Node

Source-Sink Edge: [cost, maxFlow, minFlow]	Source-Items Edges: [itemId, cost, maxFlow, minFlow]
[0.0, 3, 0]	[1, 0.0, 1, 0] [6, 0.0, 1, 0] [11, 0.0, 1, 0] [17, 0.0, 1, 0] [21, 0.0, 1, 0] [25, 0.0, 1, 0] [39, 0.0, 1, 0]

Sink Node

Source-Sink Edge: [cost, maxFlow, minFlow]	Source-Items Edges: [itemId, cost, maxFlow, minFlow]
[0.0, 3, 0]	[Comedy, 0.0, 1, 1] [Drama, 0.0, 1, 1] [Action, 0.0, 1, 1]

Εικόνα 5.3.3: Η οθόνη για την επιλογή των παραμέτρων λειτουργίας του αλγόριθμου και την μοντελοποίηση του προβλήματος.

Οι παράμετροι που οι τιμές τους μπορούν να καθοριστούν από το παράθυρο της εικόνας 5.3.3 είναι οι εξής:

Number of packages (N): Ο αριθμός των N-καλύτερων πακέτων αντικειμένων.

Number of categories (nc): Ο αριθμός των δημοφιλέστερων κατηγοριών.

Items per category (ni/c): Ο αριθμός των αντικειμένων ανά κατηγορία. Υπάρχει η δυνατότητα αναλογικής κατανομής των αντικειμένων με το πάτημα του κουμπιού *Items Allocation to Categories* και τη μετάβαση στο παράθυρο που περιγράφεται στην ενότητα 5.3.4.

Package Size: Το μέγεθος των προτεινόμενων πακέτων. Αυτή η επιλογή χρησιμοποιείται για τον καθορισμό του μεγέθους των προτεινόμενων πακέτων αντικειμένων όταν απαιτείται μέγεθος πακέτου μικρότερο από τον αριθμό των αντικειμένων που προκύπτουν από τις υπόλοιπες περιμέτρους. Για παράδειγμα μπορεί να έχουν επιλεγεί τέσσερις κατηγορίες με ένα αντικείμενο ανά κατηγορία και το μέγεθος πακέτου να καθοριστεί σε 2, τότε κάθε

προτεινόμενο πακέτο θα περιλαμβάνει δύο αντικείμενα από τις τέσσερις αυτές κατηγορίες. Επίσης, η τιμή αυτής της παραμέτρου δεν μπορεί να γίνει μεγαλύτερη από το μέγιστο μέγεθος πακέτου που προκύπτει από τις υπόλοιπες παραμέτρους. Για το παράδειγμα που προαναφέρθηκε η μέγιστη τιμή του μεγέθους του πακέτου είναι ίση με 4. Η προκαθορισμένη τιμή για την παράμετρο αυτή είναι η *default* σύμφωνα με την οποία το μέγεθος του πακέτου προκύπτει από τις υπόλοιπες παραμέτρους.

Maximum Appearances per item: Ο μέγιστος αριθμός εμφάνισης κάθε αντικειμένου στα προτεινόμενα πακέτα αντικειμένων. Οι επιλογές είναι οι εξής:

Not taken Account: Δεν λαμβάνεται υπόψη αυτός ο αριθμός.

Reading from items file: Ανάγνωση αυτού του αριθμού από το αρχείο δεδομένων των αντικειμένων που μπορεί να είναι ξεχωριστός για κάθε αντικείμενο.

Αριθμητική τιμή: Μια αριθμητική τιμή που ορίζει τον αριθμό των μέγιστων εμφανίσεων κάθε αντικειμένου στα προτεινόμενα πακέτα η οποία είναι η ίδια για όλα τα αντικείμενα.

Model of creating packages: Μοντέλο παραγωγής των προτεινόμενων πακέτων αντικειμένων. Τα μοντέλα αυτά περιγράφονται αναλυτικά στο κεφάλαιο 3. Οι διαθέσιμες επιλογές είναι οι εξής:

1. *Removing the Edge with Maximum Rating*
2. *Removing the Edge with Minimum Rating*
3. *Removing the top 50% Rated Edges*
4. *Removing the bottom 50% Rated Edges*
5. *Removing all (100%) the Edges*
6. *Distinct Packages - Removing the Item Nodes*

Category popularity mode: Μοντέλο προσδιορισμού των δημοφιλέστερων κατηγοριών. Τα μοντέλα αυτά περιγράφονται αναλυτικά στο κεφάλαιο 3. Οι διαθέσιμες επιλογές είναι οι εξής:

1. *Based on the number of items per category*
2. *Based on the value of overall ratings per category*
3. *Based on the number of predicted items per category*

Το παράθυρο περιλαμβάνει τα εξής κουμπιά:

Load Graph: Σχηματισμός του γράφου μοντελοποίησης του προβλήματος για όλες τις κατηγορίες αντικειμένων. Ο γράφος αυτός αναπαρίσταται στους τέσσερις πίνακες του παραθύρου που κατά σειρά, από πάνω προς τα κάτω, προβάλλουν τους κόμβους των αντικειμένων, τους κόμβους των κατηγοριών, τον αρχικό κόμβο και τον τερματικό κόμβο. Για τον κάθε κόμβο του γραφήματος αναγράφονται οι ακμές που περιλαμβάνονται όπως φαίνεται στην εικόνα 5.3.3. Η λειτουργία αυτή είναι προαιρετική και χρησιμοποιείται για εποπτικούς λόγους.

Load Graph with Top-N categories: Σχηματισμός του γραφήματος μοντελοποίησης του προβλήματος που περιλαμβάνει τις N-δημοφιλέστερες για τον χρήστη κατηγορίες αντικειμένων. Ο γράφος αυτός αναπαρίσταται στους τέσσερις πίνακες του παραθύρου που κατά σειρά προβάλλουν τους κόμβους των αντικειμένων, τους κόμβους των δημοφιλέστερων κατηγοριών, τον αρχικό κόμβο και τον τερματικό κόμβο. Για τον κάθε κόμβο του γραφήματος αναγράφονται οι ακμές που περιλαμβάνονται όπως φαίνεται στην εικόνα 5.3.3. Η λειτουργία αυτή είναι προαιρετική και χρησιμοποιείται για εποπτικούς λόγους.

Run Min Cost Flow Algorithm: Μετάβαση στο επόμενο παράθυρο που προβάλλει τα N-καλύτερα πακέτα αντικειμένων μετά την επαναληπτική εκτέλεση του αλγόριθμου ροής ελάχιστου κόστους.

5.3.4 Οθόνη Αναλογικής Κατανομής των Αντικειμένων στις Κατηγορίες

Η οθόνη για την αναλογική κατανομή των αντικειμένων στις δημοφιλέστερες κατηγορίες παρουσιάζεται στην εικόνα 5.3.4. Η εικόνα αυτή παρουσιάζει ένα στιγμιότυπο του παραθύρου για την αναλογική κατανομή των 10 αντικειμένων στις τρεις δημοφιλέστερες για τον χρήστη κατηγορίες. Ακολουθεί ο τρόπος χρήσης και η επεξήγηση των στοιχείων και των λειτουργιών του παραθύρου.

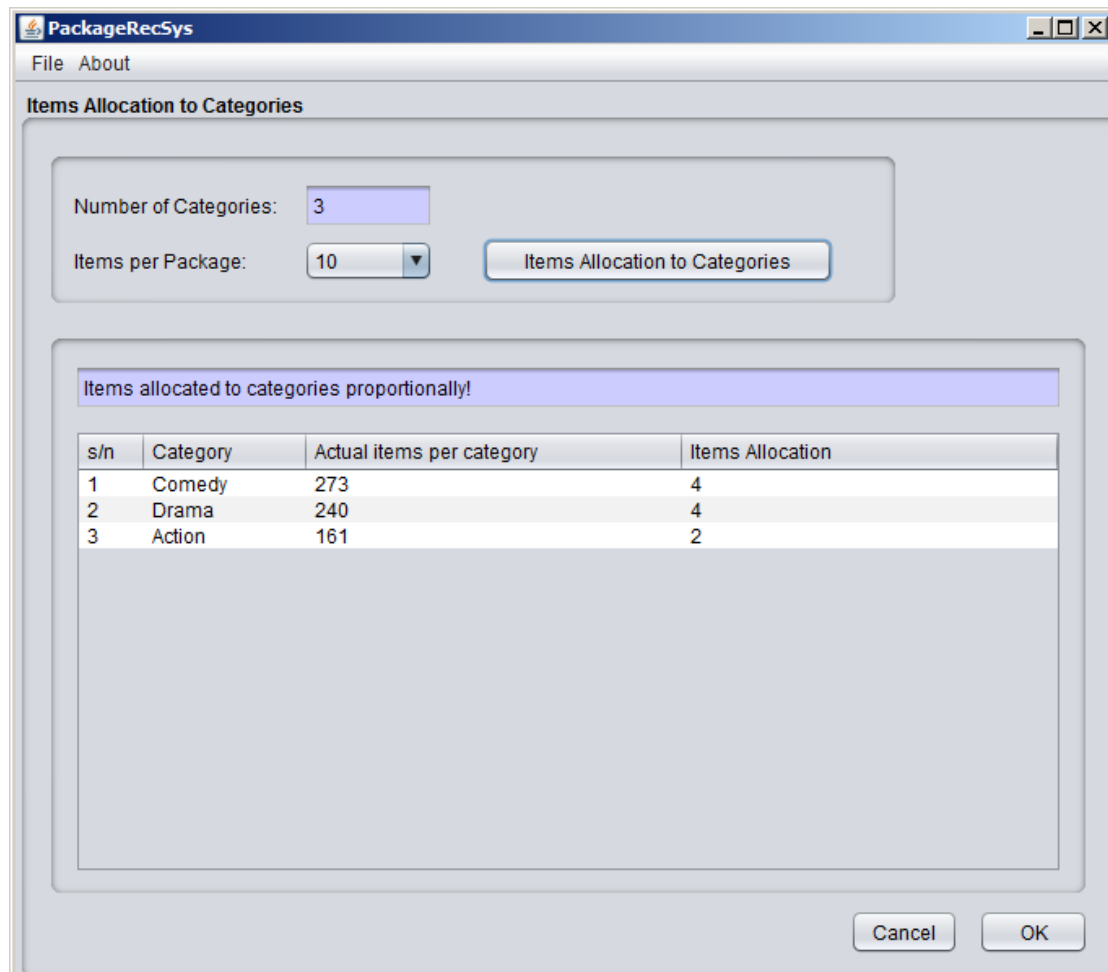
Λειτουργίες και τρόπος χρήσης του παραθύρου

Παρακάτω περιγράφονται οι λειτουργίες και ο τρόπος χρήσης των στοιχείων και των επιλογών του παραθύρου το οποίο φαίνεται στην εικόνα 5.3.4.

Η μπάρα του μενού περιλαμβάνει τα εξής:

File: Μετά το πάτημα του *File* που βρίσκεται στην μπάρα του μενού του παραθύρου εμφανίζεται η επιλογή *Exit για τον τερματισμό* της εφαρμογής.

About: Εμφάνιση μηνύματος με πληροφορίες σχετικές με την εφαρμογή.



Εικόνα 5.3.4: Η οθόνη για την αναλογική κατανομή των αντικειμένων στις δημοφιλέστερες για το χρήστη κατηγορίες.

Οι διαθέσιμες επιλογές του παραθύρου για το μέγεθος πακέτου:

Items per package: Επιλογή του αριθμού των αντικειμένων που περιλαμβάνονται στο πακέτο αντικειμένων και θα καταταξινομηθούν αναλογικά στις δημοφιλέστερες κατηγορίες.

Τα κουμπιά του παραθύρου:

Items Allocation to Categories: Αναλογική κατανομή των αντικειμένων στις δημοφιλέστερες για τον χρήστη κατηγορίες. Τα αποτελέσματα εμφανίζονται στον σχετικό πίνακα του παραθύρου όπως φαίνεται στην εικόνα 5.3.4.

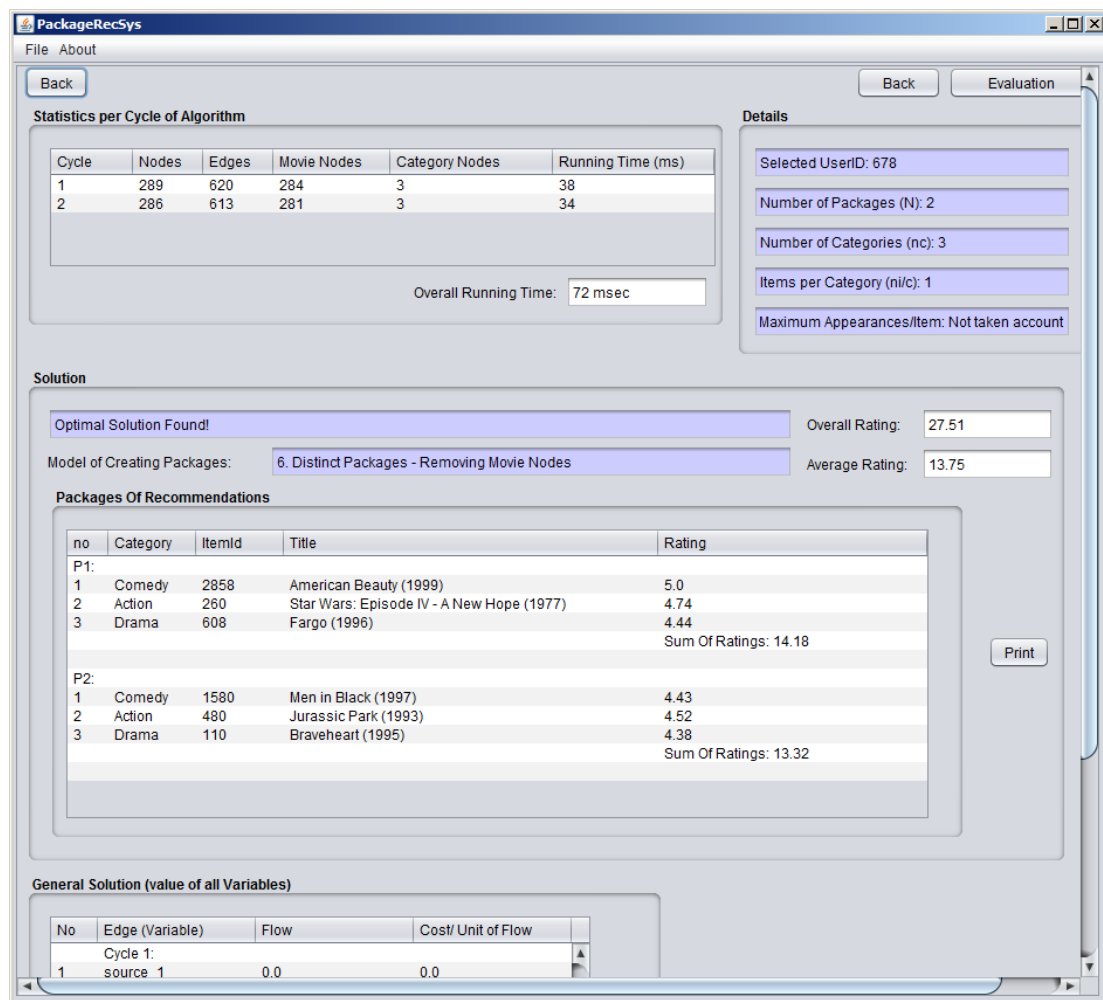
Cancel: Η κατανομή των αντικειμένων δεν πραγματοποιείται και το πρόγραμμα επιστρέφει στο προηγούμενο παράθυρο.

OK: Πραγματοποιείται η επιλεγμένη κατανομή των αντικειμένων και το πρόγραμμα μεταβαίνει στο προηγούμενο παράθυρο για την επιλογή των παραμέτρων και τη μοντελοποίηση του προβλήματος.

Εδώ υπάρχει ο περιορισμός ότι σε κάθε κατηγορία πρέπει να αντιστοιχεί τουλάχιστον ένα αντικείμενο. Συνεπώς, στην περίπτωση που επιλεγεί αριθμός αντικειμένων ανά πακέτο μικρότερος από τον αριθμό των κατηγοριών τότε το πρόγραμμα μας ενημερώνει πως είναι αδύνατη η κατανομή των αντικειμένων στις κατηγορίες.

5.3.5 Οθόνη Προβολής των N-καλύτερων Προτεινόμενων Πακέτων

Η οθόνη προβολής των N-καλύτερων προτεινόμενων πακέτων αντικειμένων παρουσιάζεται στην εικόνα 5.3.5(α). Η εικόνα αυτή παρουσιάζει ένα στιγμιότυπο του παραθύρου που εμφανίζει την πρόταση των δύο-καλύτερων πακέτων αντικειμένων που περιλαμβάνουν τρία αντικείμενα το καθένα. Ακολουθεί ο τρόπος χρήσης και η επεξήγηση των στοιχείων και των λειτουργιών του παραθύρου.



Εικόνα 5.3.5 (α): Η οθόνη προβολής των N-καλύτερων προτεινόμενων πακέτων αντικειμένων.

Λειτουργίες και τρόπος χρήσης του παραθύρου

Παρακάτω περιγράφονται οι λειτουργίες και ο τρόπος χρήσης των στοιχείων και των επιλογών του παραθύρου το οποίο φαίνεται στην εικόνα 5.3.5(α).

Η μπάρα του μενού περιλαμβάνει τα εξής:

File: Μετά το πάτημα του *File* που βρίσκεται στην μπάρα του μενού του παραθύρου εμφανίζονται οι επιλογές *Back* και *Exit* για την επιστροφή στο προηγούμενο παράθυρο ή τον τερματισμό της εφαρμογής αντίστοιχα. Η επιλογή του *Back* δίνεται και από τα σχετικά κουμπιά του παραθύρου.

About: Εμφάνιση μηνύματος με πληροφορίες σχετικές με την εφαρμογή.

Οι περιοχές του παραθύρου:

Statistics per Cycle of Algorithm: Στο πλαίσιο αυτό εμφανίζονται πληροφορίες σχετικές με το γράφημα που μοντελοποιεί το πρόβλημα και τον χρόνο εκτέλεσης κάθε κύκλου του αλγόριθμου. Επίσης, αναγράφεται και ο συνολικός χρόνος εκτέλεσης του αλγόριθμου.

Details: Στο πλαίσιο αυτό αναγράφονται κάποιες λεπτομέρειες για τις τιμές των παραμέτρων βάσει των οποίων λειτουργεί το σύστημα.

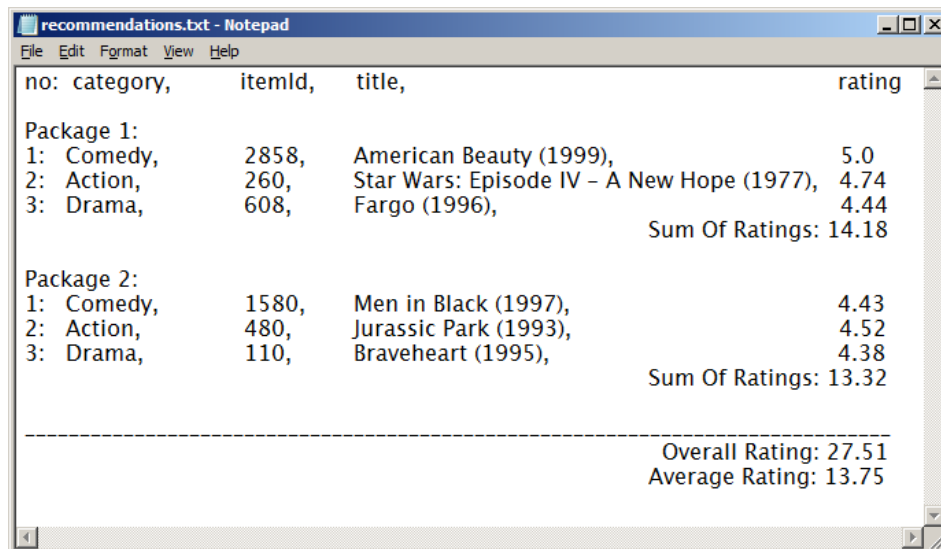
Solution: Στο πλαίσιο αυτό περιλαμβάνονται τα N-καλύτερα προτεινόμενα πακέτα αντικειμένων όπως μπορούμε να τα διακρίνουμε στην εικόνα 5.3.5(α). Αναγράφεται και το μοντέλο δημιουργίας των καλύτερων πακέτων που έχει επιλεγεί. Επίσης, δίνονται πληροφορίες για τα ratings ανά πακέτο, για το συνολικό άθροισμα των ratings και για τη μέση τιμή τους. Σε περίπτωση που το πρόβλημα γραμμικού προγραμματισμού δεν έχει εφικτή λύση τότε το σύστημα μας ενημερώνει σχετικά με αυτό.

General solution: Ο πίνακας που περιλαμβάνεται στην περιοχή αυτή προβάλλει τις τιμές όλων των μεταβλητών απόφασης του προβλήματος γραμμικού προγραμματισμού για την εύρεση των βέλτιστων πακέτων αντικειμένων μετά την επίλυσή του. Ο πίνακας αυτός χρησιμοποιείται για εποπτικούς λόγους.

Τα κουμπιά του παραθύρου:

Print: Εκτύπωση και άνοιγμα του αρχείου που περιλαμβάνει την αναφορά με τα N-Καλύτερα πακέτα αντικειμένων. Στην εικόνα 5.3.5(β) φαίνεται ένα στιγμιότυπο αυτής της αναφοράς.

Evaluation: Μετάβαση στο παράθυρο παρουσίασης των δεικτών αξιολόγησης του αλγόριθμου που περιγράφεται στην ενότητα 5.3.6. Η λειτουργία αυτή χρησιμοποιείται στη λειτουργία αξιολόγησης.

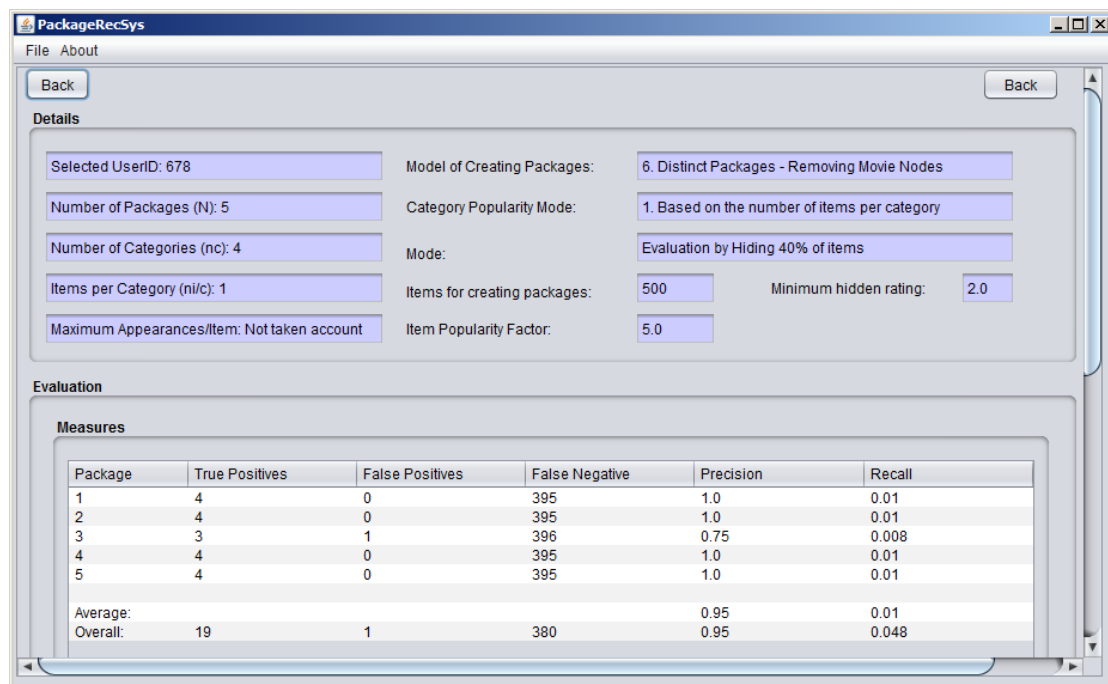


no:	category,	itemId,	title,	rating
Package 1:				
1:	Comedy,	2858,	American Beauty (1999),	5.0
2:	Action,	260,	Star Wars: Episode IV - A New Hope (1977),	4.74
3:	Drama,	608,	Fargo (1996),	4.44
				Sum Of Ratings: 14.18
Package 2:				
1:	Comedy,	1580,	Men in Black (1997),	4.43
2:	Action,	480,	Jurassic Park (1993),	4.52
3:	Drama,	110,	Braveheart (1995),	4.38
				Sum Of Ratings: 13.32
				Overall Rating: 27.51
				Average Rating: 13.75

Εικόνα 5.3.5 (β): Αναφορά με τα N-καλύτερα προτεινόμενα πακέτα αντικειμένων.

5.3.6 Οθόνη Αξιολόγησης

Η οθόνη προβολής των δεικτών αξιολόγησης παρουσιάζεται στην εικόνα 5.3.6. Η εικόνα αυτή παρουσιάζει ένα στιγμιότυπο του παραθύρου που εμφανίζει τους δείκτες αξιολόγησης για την πρόταση 5 πακέτων αντικειμένων που περιλαμβάνουν 4 αντικείμενα το καθένα. Ακολουθεί ο τρόπος χρήσης και η επεξήγηση των στοιχείων και των λειτουργιών του παραθύρου.



Package	True Positives	False Positives	False Negative	Precision	Recall
1	4	0	395	1.0	0.01
2	4	0	395	1.0	0.01
3	3	1	396	0.75	0.008
4	4	0	395	1.0	0.01
5	4	0	395	1.0	0.01
Average:				0.95	0.01
Overall:	19	1	380	0.95	0.048

Εικόνα 5.3.6: Η οθόνη προβολής των δεικτών αξιολόγησης του αλγόριθμου.

Λειτουργίες και τρόπος χρήσης του παραθύρου αξιολόγησης

Παρακάτω περιγράφονται οι λειτουργίες και ο τρόπος χρήσης των στοιχείων και των επιλογών του παραθύρου το οποίο φαίνεται στην εικόνα 5.3.6.

Η μπάρα του μενού περιλαμβάνει τα εξής:

File: Μετά το πάτημα του *File* εμφανίζονται οι επιλογές *Back* και *Exit* για την επιστροφή στο προηγούμενο παράθυρο ή τον τερματισμό της εφαρμογής αντίστοιχα. Η επιλογή του *Back* δίνεται και από τα σχετικά κουμπιά του παραθύρου.

About: Εμφάνιση μηνύματος με πληροφορίες σχετικές με την εφαρμογή.

Το παράθυρο της αξιολόγησης περιλαμβάνει τις περιοχές:

Details: Στο πλαίσιο αυτό αναγράφονται όλες οι λεπτομέρειες για τις τιμές των παραμέτρων βάσει των οποίων λειτουργεί το σύστημα.

Measures: Στον πίνακα παρουσιάζονται οι τιμές των δεικτών αξιολόγησης για κάθε προτεινόμενο πακέτο ξεχωριστά και για το σύνολο (*overall*) των προτεινόμενων πακέτων. Για τους δείκτες *precision* και *recall* υπολογίζεται και η μέση τιμή (*average*) για όλα τα προτεινόμενα πακέτα αντικειμένων.

Κεφάλαιο 6: Εφαρμογή και Αξιολόγηση

6.1 Γενικά

Σε αυτό το κεφάλαιο εφαρμόζεται και αξιολογείται η προσέγγιση της παρούσας εργασίας με χρήση πραγματικών δεδομένων. Για την εφαρμογή, τη δοκιμή και την αξιολόγηση αναπτύχθηκε το λογισμικό που περιγράφεται στο κεφάλαιο πέντε το οποίο ενσωματώνει όλη την απαιτούμενη λειτουργικότητα ενός ολοκληρωμένου συστήματος συστάσεων που βασίζεται στον αλγόριθμο που αναπτύχθηκε. Ο αλγόριθμος, όπως είναι ήδη γνωστό από το κεφάλαιο τρία, περιλαμβάνει ένα σύνολο από παραμέτρους που μπορούν να μεταβληθούν. Η μεταβολή αυτών των παραμέτρων μπορεί να οδηγήσει σε αλλαγή της συμπεριφοράς του συστήματος. Σε αυτό το κεφάλαιο παρουσιάζεται η δοκιμή και η αξιολόγηση του αλγόριθμου για διάφορα σενάρια λειτουργίας ή για διάφορες τιμές των παραμέτρων με σκοπό την εξαγωγή συμπερασμάτων για την απόδοση, την αποτελεσματικότητα και την ακρίβεια.

Στις επόμενες παραγράφους γίνεται αναφορά στο σύνολο των δεδομένων που χρησιμοποιείται, παρουσιάζεται η διαδικασία που ακολουθεί η αξιολόγηση, περιγράφονται οι μετρήσεις που έγιναν κατά την προετοιμασία της αξιολόγησης και τέλος καταγράφονται οι μετρήσεις αξιολόγησης και τα συμπεράσματα που προκύπτουν από αυτές.

6.2 Σύνολο Δεδομένων

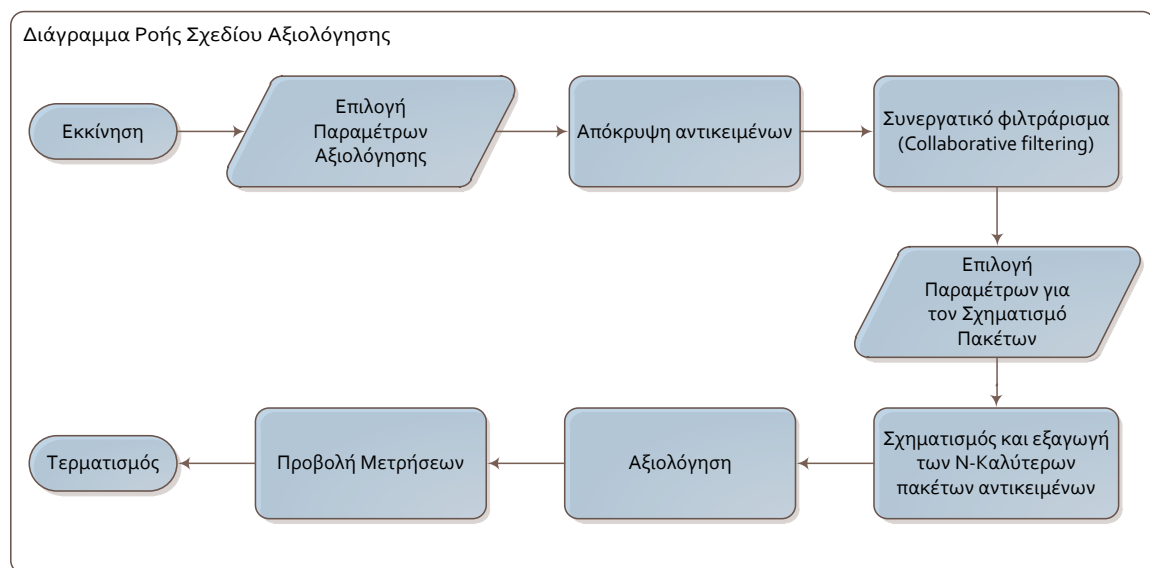
Για τη δοκιμή του συστήματος απαιτείται ένα κατάλληλο σύνολο δεδομένων που περιλαμβάνει πραγματικά δεδομένα με σκοπό την εξαγωγή έγκυρων συμπερασμάτων. Ένα τέτοιο σύνολο δεδομένων που υπακούει στις απαιτήσεις και τις προδιαγραφές της παρούσας προσέγγισης αρκεί να περιλαμβάνει ένα σύνολο από χρήστες, ένα σύνολο από αντικείμενα και ένα σύνολο από αξιολογήσεις χρηστών για τα αντικείμενα. Ο αλγόριθμος που αναπτύχθηκε είναι γενικός και δεν απευθύνεται σε κάποια συγκεκριμένη εφαρμογή, συνεπώς θα μπορούσαν να χρησιμοποιηθούν δεδομένα από οποιαδήποτε περιοχή όπως δεδομένα για πρόταση πακέτων ταινιών, δεδομένα για πρόταση μιας εβδομαδιαίας δίαιτας, δεδομένα για πρόταση βιβλίων ή δεδομένα για πρόταση πακέτων αντικειμένων για οποιαδήποτε

εμπορική εφαρμογή. Ωστόσο, η χρήση των δεδομένων και ο σκοπός της χρησιμοποίησης του συστήματος συστάσεων θα πρέπει να συμβαδίζουν με το πρόβλημα για το οποίο το σύστημα αυτό σχεδιάστηκε, σύμφωνα με όσα αναφέρονται στο κεφάλαιο 3.

Το σύνολο δεδομένων που χρησιμοποιείται είναι ένα γνωστό σύνολο από δεδομένα ταινιών που παρέχεται από την GroupLens (GroupLens Research datasets., n.d.) και ονομάζεται “*MovieLens 1M Dataset*”. Αυτό το σύνολο δεδομένων περιλαμβάνει 1.000.209 ανώνυμες αξιολογήσεις χρηστών για περίπου 3.900 ταινίες από 6.040 χρήστες του δικτυακού τόπου MovieLens (n.d.). Είναι ένα ευρέως διαδεδομένο σύνολο δεδομένων που χρησιμοποιείται για δοκιμή και αξιολόγηση συστημάτων συστάσεων και μπορεί να θεωρηθεί ως σημείο αναφοράς για τη σύγκριση τέτοιων συστημάτων. Όλες οι δοκιμές και οι μετρήσεις που ακολουθούν στις επόμενες παραγράφους γίνονται χρησιμοποιώντας αυτό το σύνολο δεδομένων.

6.3 Σχέδιο Αξιολόγησης

Η αξιολόγηση του αλγόριθμου βασίζεται κυρίως στη μέτρηση του *χρόνου εκτέλεσης* και του δείκτη *precision* για τα προτεινόμενα πακέτα αντικειμένων ακολουθώντας τη διαδικασία που απεικονίζεται στο διάγραμμα ροής της εικόνας 6.3. Η διαδικασία αυτή ξεκινάει από το σημείο που έχει επιλεγεί ο χρήστης για τον οποίο θα εξαχθούν τα N-καλύτερα πακέτα αντικειμένων με σκοπό την αξιολόγησή τους.



Εικόνα 6.3: Το διάγραμμα ροής της διαδικασίας αξιολόγησης.

Σύμφωνα με το διάγραμμα ροής της εικόνας 6.3, το πρώτο στάδιο της διαδικασίας αξιολόγησης είναι η επιλογή των παραμέτρων αξιολόγησης. Οι παράμετροι αυτές είναι οι εξής:

- *Ποσοστό απόκρυψης αντικειμένων*: Επιλογή του ποσοστού απόκρυψης των επιλεγμένων ή βαθμολογημένων από τον χρήστη αντικειμένων.
- *Ελάχιστος βαθμός προτίμησης κρυμμένων αντικειμένων*: Επιλογή της ελάχιστης τιμής για τον βαθμό προτίμησης των κρυμμένων αντικειμένων με συνέπεια να αποκρύπτονται αντικείμενα με ίσο ή μεγαλύτερο βαθμό προτίμησης από αυτή την τιμή.

Το δεύτερο στάδιο που ακολουθείται, στη διαδικασία αξιολόγησης, είναι η απόκρυψη ενός μέρους των αντικειμένων του τρέχοντα χρήστη, με σκοπό τον τελικό υπολογισμό της ακρίβειας των αντικειμένων που περιλαμβάνουν τα προτεινόμενα πακέτα σε σχέση με τα κρυμμένα αντικείμενα. Η απόκρυψη των αντικειμένων ακολουθεί μια διαδικασία τυχαίας απόκρυψης ενός προκαθορισμένου ποσοστού αντικειμένων για τον τρέχοντα χρήστη λαμβάνοντας υπόψη την ελάχιστη βαθμολογία προτίμησης των κρυμμένων αντικειμένων. Στις επόμενες παραγράφους γίνεται περισσότερο σαφής ο σκοπός της απόκρυψης των αντικειμένων.

Το τρίτο στάδιο της αξιολόγησης περιλαμβάνει το συνεργατικό φιλτράρισμα (*collaborative filtering*) το οποίο χρησιμοποιεί τα μη κρυμμένα αντικείμενα που έχει επιλέξει και βαθμολογήσει ο χρήστης με σκοπό την πρόβλεψη του βαθμού προτίμησης των μη βαθμολογημένων από τον χρήστη αντικειμένων. Τα αντικείμενα εξάγονται από το συνεργατικό φιλτράρισμα και ταξινομούνται σύμφωνα με τον εκτιμώμενο για το καθένα βαθμό προτίμησης. Σε αυτό το στάδιο, ανάλογα με την επιλογή της σχετικής παραμέτρου, υπάρχει η δυνατότητα να εξαχθεί ένας συγκεκριμένος αριθμός από τα καλύτερα για το χρήστη αντικείμενα ή να εξαχθούν όλα τα διαθέσιμα αντικείμενα. Να σημειωθεί ότι γίνεται συνεργατικό φιλτράρισμα βασιζόμενο στα αντικείμενα (*item-based collaborative filtering*), καθώς στις περισσότερες εφαρμογές τα αντικείμενα είναι λιγότερα από τους χρήστες και με αυτόν τον τρόπο επιτυγχάνονται υψηλότερες επιδόσεις.

Το τέταρτο στάδιο περιλαμβάνει την επιλογή των παραμέτρων για τον σχηματισμό των προτεινόμενων πακέτων. Η αναλυτική περιγραφή των παραμέτρων αυτών δίνεται στην ενότητα 3.3.4 του κεφαλαίου 3 που περιγράφεται ο αλγόριθμος, ωστόσο ακολουθεί μια σύντομη αναφορά στις παραμέτρους που είναι οι εξής:

- *Ο Μέγιστος αριθμός των προτεινόμενων αντικειμένων του συνεργατικού φιλτραρίσματος:* Επιλογή του μέγιστου αριθμού των αντικειμένων με τους μεγαλύτερους εκτιμώμενους βαθμούς προτίμησης που εξάγονται από το συνεργατικό φιλτράρισμα.
- *Ο συντελεστής δημοφιλίας των αντικειμένων:* Ο συντελεστής δημοφιλίας λειτουργεί ως πολλαπλασιαστικός παράγοντας στο ποσοστό δημοφιλίας κάθε αντικειμένου. Το ποσοστό δημοφιλίας κάθε αντικειμένου δίνεται από το πηλίκο των χρηστών που έχουν επιλέξει το συγκεκριμένο αντικείμενο προς τον συνολικό αριθμό των χρηστών. Το γινόμενο του συντελεστή δημοφιλίας επί του ποσοστού δημοφιλίας προστίθεται στον εκτιμώμενο βαθμό προτίμησης του κάθε αντικειμένου.
- *Ο αριθμός των N-καλύτερων πακέτων αντικειμένων.*
- *Ο αριθμός των δημοφιλέστερων κατηγοριών.*
- *Ο αριθμός των αντικειμένων ανά κατηγορία ή η επιλογή της αναλογικής κατανομής των αντικειμένων στις δημοφιλέστερες κατηγορίες.* Στη δεύτερη περίπτωση καθορίζεται ο αριθμός των αντικειμένων που περιλαμβάνει κάθε πακέτο.
- *Ο μέγιστος αριθμός εμφάνισης κάθε αντικειμένου στα προτεινόμενα πακέτα:* Η παράμετρος αυτή παίρνει την τιμή «δε λαμβάνεται υπόψη» σε όλες τις μετρήσεις που γίνονται κατά την αξιολόγηση του αλγόριθμου καθώς δεν έχει κάποια πρακτική εφαρμογή στην περίπτωση του συνόλου δεδομένων που χρησιμοποιείται για την αξιολόγηση.
- *Το μοντέλο παραγωγής των προτεινόμενων πακέτων.*
- *Το μοντέλο επιλογής των δημοφιλέστερων κατηγοριών.*

Τα δύο τελευταία στάδια της διαδικασίας αξιολόγησης περιλαμβάνουν τον υπολογισμό των δεικτών αξιολόγησης και την προβολή των αποτελεσμάτων. Για την αξιολόγηση, ανάλογα με το τι διερευνάται, υπολογίζεται ο δείκτης *precision*, ο εκτιμώμενος βαθμός προτίμησης των προτεινόμενων πακέτων και ο χρόνος εκτέλεσης του αλγόριθμου. Ο δείκτης *Precision* υπολογίζεται για κάθε προτεινόμενο πακέτο ξεχωριστά και εξάγεται η μέση τιμή του για όλα τα προτεινόμενα πακέτα. Επίσης, υπολογίζεται η τιμή του δείκτη *Precision* λαμβάνοντας υπόψη τα διακριτά αντικείμενα που περιλαμβάνονται στο σύνολο των προτεινόμενων πακέτων. Η βαθμολογία προτίμησης των αντικειμένων υπολογίζεται για κάθε πακέτο ξεχωριστά, αθροίζεται για το σύνολο των προτεινόμενων πακέτων και παράλληλα υπολογίζεται η μέση τιμή της στο σύνολο. Αντίστοιχα, ο χρόνος

εκτέλεσης του αλγόριθμου υπολογίζεται τόσο για κάθε προτεινόμενο πακέτο ξεχωριστά όσο και για το σύνολο των προτεινόμενων πακέτων.

Για τη μέτρηση του δείκτη *Precision* αποκρύπτεται ένα σύνολο από αντικείμενα που έχει βαθμολογήσει ο χρήστης και χρησιμοποιώντας τα υπόλοιπα αντικείμενα για τη λειτουργία του συνεργατικού φιλτραρίσματος, ο αλγόριθμος παράγει τα N-καλύτερα προτεινόμενα πακέτα αντικειμένων. Ο υπολογισμός του δείκτη *Precision* αντιστοιχεί στο λόγο του πλήθους των προτεινόμενων αντικειμένων που περιλαμβάνονται στα κρυμμένα αντικείμενα προς το πλήθος των προτεινόμενων αντικειμένων.

$$Precision = \frac{|\{\text{Προτεινόμενα αντικείμενα}\} \cap \{\text{Κρυμμένα αντικείμενα}\}|}{|\{\text{Προτεινόμενα αντικείμενα}\}|}$$

Αυτός ο υπολογισμός γίνεται για κάθε προτεινόμενο πακέτο και για το σύνολο των αντικειμένων που περιλαμβάνονται σε όλα τα προτεινόμενα πακέτα.

Για την καλύτερη κατανόηση του τρόπου υπολογισμού του δείκτη *Precision* ακολουθεί ένα σχετικό παράδειγμα που περιλαμβάνει την πρόταση ενός πακέτου τεσσάρων αντικειμένων. Σε αυτό το παράδειγμα θεωρείται ότι το σύνολο των βαθμολογημένων αντικειμένων ενός χρήστη είναι $R=\{a, b, c, d, e, f, g, h\}$. Το σύστημα αξιολόγησης, από το σύνολο R , αποκρύπτει τα αντικείμενα $H=\{e, f, g, h\}$ και ο αλγόριθμος χρησιμοποιεί τα υπόλοιπα αντικείμενα $R'=R-H=\{a, b, c, d\}$ για την εκτίμηση και την παραγωγή των προτάσεων. Έστω ότι ο αλγόριθμος παράγει το προτεινόμενο πακέτο $P=\{f, g, h, i\}$, τότε ο δείκτης *Precision* για το προτεινόμενο αυτό πακέτο υπολογίζεται ως εξής:

$$Precision = \frac{|P \cap H|}{|P|} = \frac{|\{f, g, h, i\} \cap \{e, f, g, h\}|}{|\{f, g, h, i\}|} = \frac{|\{f, g, h\}|}{|\{f, g, h, i\}|} = \frac{3}{4} = 0.75$$

6.4 Προετοιμασία Αξιολόγησης

Για να πραγματοποιηθεί με εγκυρότητα η αξιολόγηση θα πρέπει να καθοριστούν οι παράμετροι αξιολόγησης με τρόπο που να μην επηρεάζουν την απόδοση και την αποτελεσματικότητα του αλγόριθμου. Για το σκοπό αυτό θα γίνει μια προσπάθεια προσδιορισμού της τιμής των παραμέτρων αξιολόγησης, που περιγράφονται στην προηγούμενη ενότητα, με τέτοιο τρόπο ώστε να μην επηρεάζεται η ακρίβεια και η αποτελεσματικότητα του συστήματος από τη διαδικασία της αξιολόγησης.

Η πρώτη παράμετρος, για την οποία θα διερευνηθεί η επιρροή της στην ακρίβεια πρόβλεψης των βαθμών προτίμησης, είναι η παράμετρος που καθορίζει το ποσοστό απόκρυψης των αντικειμένων. Για το σκοπό αυτό έγιναν μετρήσεις μεταβάλλοντας το ποσοστό απόκρυψης των αντικειμένων και υπολογίζοντας το μέσο απόλυτο σφάλμα (*Mean Absolute Error - MAE*) και το μέσο τετραγωνικό σφάλμα (*Root Mean Squared Error - RMSE*) της προβλεπόμενης βαθμολογίας σε σχέση με την πραγματική βαθμολογία των κρυμμένων αντικειμένων. Οι ορισμοί για το μέσο απόλυτο και το μέσο τετραγωνικό σφάλμα δίνονται στην ενότητα 2.5 του κεφαλαίου 2. Οι μετρήσεις έγιναν για δέκα χρήστες που έχουν βαθμολογήσει περίπου 500 αντικείμενα ο καθένας και για δέκα χρήστες που έχουν βαθμολογήσει περίπου 1000 αντικείμενα ο καθένας.

Τα αποτελέσματα από τη μέση τιμή των μετρήσεων με τα διαστήματα εμπιστοσύνης για τη μέση τιμή θεωρώντας 95% επίπεδο εμπιστοσύνης συγκεντρώνονται στον πίνακα 6.4(α) και απεικονίζονται στο διάγραμμα της εικόνας 6.4(α).

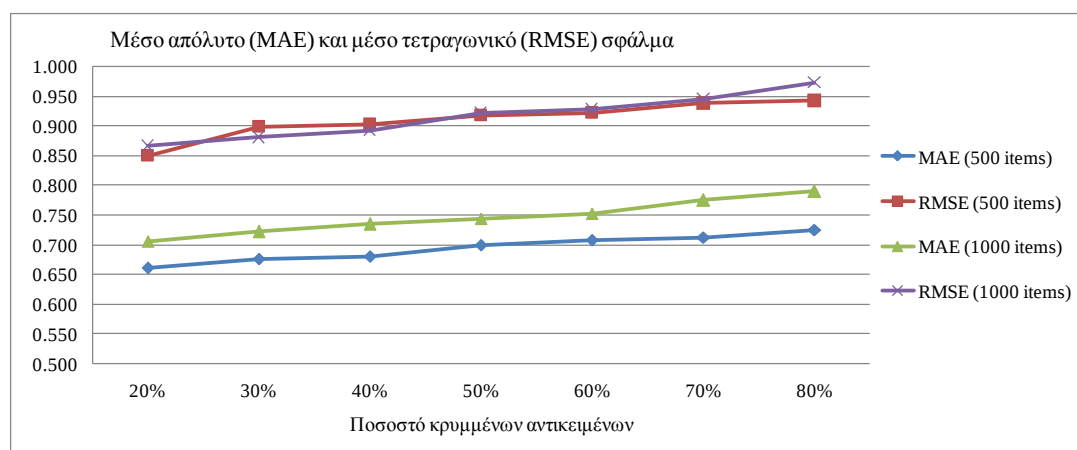
A/A	Ποσοστό κρυμμένων αντικειμένων	Μέσο απόλυτο σφάλμα (MAE)		Μέσο τετραγωνικό σφάλμα (RMSE)	
		500 Αντικείμενα	1000 Αντικείμενα	500 Αντικείμενα	1000 Αντικείμενα
1	20%	0.661±0.026	0.705±0.045	0.850±0.031	0.866±0.053
2	30%	0.676±0.022	0.722±0.051	0.899±0.029	0.881±0.055
3	40%	0.681±0.023	0.734±0.055	0.902±0.029	0.893±0.058
4	50%	0.699±0.024	0.744±0.049	0.918±0.035	0.921±0.056
5	60%	0.707±0.026	0.752±0.054	0.921±0.034	0.927±0.059
6	70%	0.712±0.028	0.775±0.072	0.938±0.033	0.946±0.069
7	80%	0.725±0.026	0.789±0.071	0.942±0.036	0.974±0.076

Πίνακας 6.4 (α): Το μέσο απόλυτο και το μέσο τετραγωνικό σφάλμα των εκτιμώμενων βαθμολογιών προτίμησης σε σχέση με το ποσοστό απόκρυψης αντικειμένων για χρήστες που έχουν βαθμολογήσει περίπου 500 και 1000 αντικείμενα, αντίστοιχα. Υπολογίζεται και το διάστημα εμπιστοσύνης ($\pm$ από τη μέση τιμή του δείγματος) για επίπεδο εμπιστοσύνης 95%.

Υπολογίστηκε και η τυπική απόκλιση για το δείγμα των αντίστοιχων μετρήσεων του μέσου απόλυτου και του μέσου τετραγωνικού σφάλματος. Οι τιμές που βρέθηκαν για την τυπική απόκλιση αναγράφονται στον πίνακα 6.4(β) και απεικονίζονται γραφικά στην εικόνα 6.4(β).

Α/Α	Ποσοστό κρυμμένων αντικειμένων	Τυπική απόκλιση (S)			
		Μέσου απόλυτου σφάλματος (Sm)		Μέσου τετραγωνικού σφάλματος (Sr)	
		500 Αντικείμενα	1000 Αντικείμενα	500 Αντικείμενα	1000 Αντικείμενα
1	20%	0.043	0.073	0.050	0.086
2	30%	0.035	0.083	0.047	0.088
3	40%	0.037	0.089	0.047	0.094
4	50%	0.039	0.080	0.056	0.090
5	60%	0.042	0.087	0.055	0.096
6	70%	0.045	0.117	0.053	0.111
7	80%	0.041	0.115	0.058	0.123

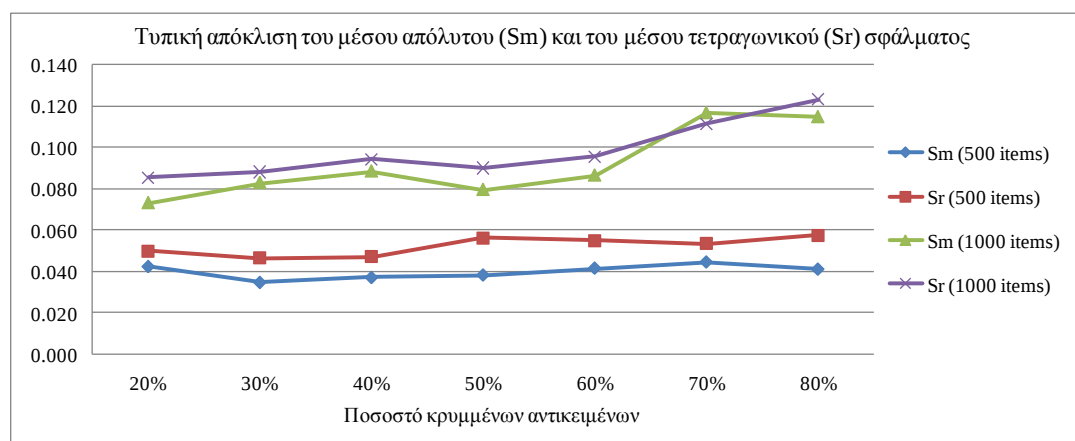
Πίνακας 6.4(β): Η τυπική απόκλιση του μέσου απόλυτου και του μέσου τετραγωνικού σφάλματος των προβλεπόμενων βαθμολογιών προτίμησης σε σχέση με το ποσοστό απόκρυψης αντικειμένων για χρήστες που έχουν βαθμολογήσει περίπου 500 και 1000 αντικείμενα, αντίστοιχα.



Εικόνα 6.4 (α): Το μέσο απόλυτο (MAE) και το μέσο τετραγωνικό σφάλμα (RMSE) των εκτιμώμενων βαθμολογιών προτίμησης σε σχέση με το ποσοστό απόκρυψης αντικειμένων για χρήστες που έχουν βαθμολογήσει περίπου 500 και 1000 αντικείμενα, αντίστοιχα.

Όπως παρατηρούμε στο διάγραμμα της εικόνας 6.4(α), η μεταβολή του ποσοστού των κρυμμένων αντικειμένων δεν επηρεάζει σημαντικά το μέσο απόλυτο και το μέσο τετραγωνικό σφάλμα του συνεργατικού φιλτραρίσματος καθώς παρατηρούμε μια μικρή αύξηση των σφαλμάτων σε σχέση με την αύξηση του ποσοστού απόκρυψης των αντικειμένων. Ωστόσο παρατηρούμε ότι έχουμε μια αύξηση των διαστημάτων εμπιστοσύνης ειδικά στην περίπτωση των 1000 αντικειμένων. Αυτό σημαίνει ότι ενώ φαίνεται πως υπάρχει μια σχετική ελευθερία για την επιλογή του ποσοστού των κρυμμένων αντικειμένων λόγω της μικρής αύξησης των μέσων σφαλμάτων, αυτό το ποσοστό δεν πρέπει να ξεπεράσει την τιμή του 60% λόγω της σχετικά μεγάλης αύξησης των διαστημάτων εμπιστοσύνης.

Στο διάγραμμα της εικόνας 6.4(β) παρουσιάζεται η μεταβολή της τυπικής απόκλισης του δείγματος για τις μετρήσεις του μέσου απόλυτου και του μέσου τετραγωνικού σφάλματος σε σχέση με το ποσοστό απόκρυψης των αντικειμένων. Για την τυπική απόκλιση παρατηρούμε ότι η τιμή της παραμένει σε χαμηλά επίπεδα με μια αυξητική τάση ειδικά για το δείγμα των 1000 αντικειμένων σε ποσοστά απόκρυψης μεγαλύτερα από 60%. Οι σχετικά μικρές τιμές της τυπικής απόκλισης δείχνουν τη σταθερότητα και την αξιοπιστία του συνεργατικού φιλτραρίσματος καθώς δεν παρατηρούνται μεγάλες διακυμάνσεις.

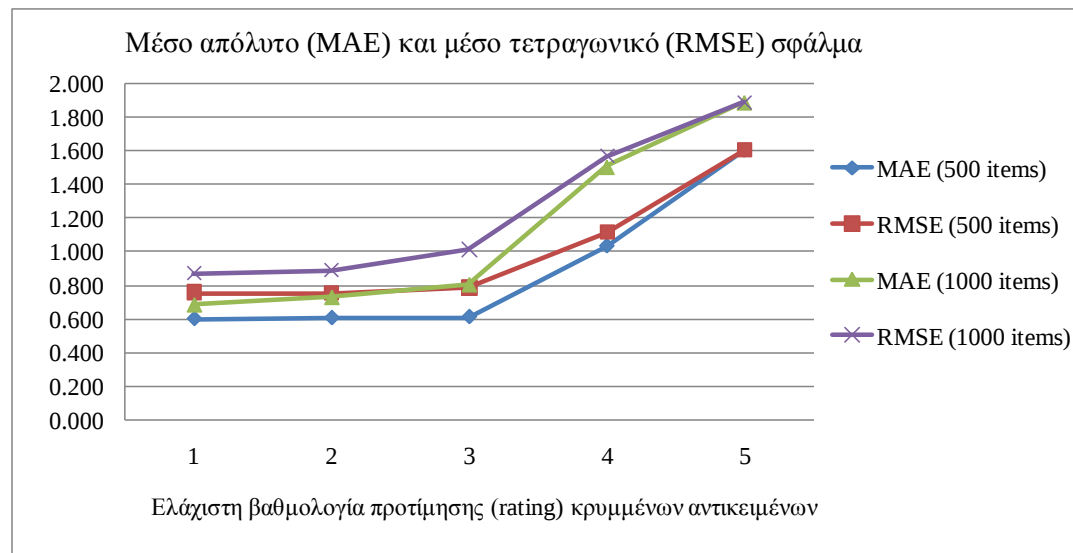


Εικόνα 6.4(β): Η τυπική απόκλιση του μέσου απόλυτου και του μέσου τετραγωνικού σφάλματος των προβλεπόμενων βαθμολογιών προτίμησης σε σχέση με το ποσοστό απόκρυψης αντικειμένων για χρήστες που έχουν βαθμολογήσει περίπου 500 και 1000 αντικείμενα, αντίστοιχα.

Ένας δεύτερος παράγοντας που πρέπει να εξεταστεί για το κατά πόσο επηρεάζει την ακρίβεια του συνεργατικού φιλτραρίσματος είναι η τιμή του ελάχιστου βαθμού προτίμησης των κρυμμένων αντικειμένων. Για τον σκοπό αυτό μετρήθηκε το μέσο απόλυτο και το μέσο τετραγωνικό σφάλμα του συνεργατικού φιλτραρίσματος διατηρώντας σταθερό το ποσοστό των κρυμμένων αντικειμένων στην τιμή του 40% και μεταβάλλοντας την ελάχιστη βαθμολογία προτίμησης αυτών. Για τον υπολογισμό του μέσου απόλυτου και του μέσου τετραγωνικού σφάλματος συγκρίθηκαν οι τιμές των προβλεπόμενων βαθμών προτίμησης με τις αντίστοιχες τιμές των κρυμμένων αντικειμένων. Οι μετρήσεις έγιναν για δέκα χρήστες που έχουν βαθμολογήσει περίπου 500 αντικείμενα και για δέκα χρήστες που έχουν βαθμολογήσει περίπου 1000 αντικείμενα ο καθένας. Τα αποτελέσματα από τη μέση τιμή των μετρήσεων με τα διαστήματα εμπιστοσύνης για επίπεδο εμπιστοσύνης ίσο με 95% συγκεντρώνονται στον πίνακα 6.4(γ) και απεικονίζονται αντίστοιχα στο διάγραμμα της εικόνας 6.4(γ).

A/A	Ελάχιστη βαθμολογία προτίμησης κρυμμένων αντικειμένων	Μέσο απόλυτο σφάλμα (MAE)		Μέσο τετραγωνικό σφάλμα (RMSE)	
		500 Αντικείμενα	1000 Αντικείμενα	500 Αντικείμενα	1000 Αντικείμενα
1	1	0.599±0.022	0.682±0.052	0.755±0.034	0.869±0.070
2	2	0.604±0.022	0.726±0.061	0.753±0.039	0.887±0.085
3	3	0.610±0.020	0.802±0.075	0.784±0.036	1.009±0.099
4	4	1.032±0.160	1.503±0.229	1.115±0.147	1.568±0.225
5	5	1.602±0.099	1.887±0.150	1.602±0.099	1.889±0.150

Πίνακας 6.4 (γ): Το μέσο απόλυτο και το μέσο τετραγωνικό σφάλμα των εκτιμώμενων βαθμολογιών προτίμησης σε σχέση με την ελάχιστη βαθμολογία προτίμησης των κρυμμένων αντικειμένων για χρήστες που έχουν βαθμολογήσει περίπου 500 και 1000 αντικείμενα, αντίστοιχα. Αναγράφεται και το διάστημα εμπιστοσύνης για επίπεδο εμπιστοσύνης 95%.



Εικόνα 6.4 (γ): Το μέσο απόλυτο (MAE) και το μέσο τετραγωνικό σφάλμα (RMSE) των προβλεπόμενων βαθμολογιών προτίμησης σε σχέση με την ελάχιστη βαθμολογία προτίμησης των κρυμμένων αντικειμένων για χρήστες που έχουν βαθμολογήσει περίπου 500 και 1000 αντικείμενα, αντίστοιχα.

Όπως παρατηρούμε στο διάγραμμα της εικόνας 6.4(γ), για απόκρυψη αντικειμένων με ελάχιστη βαθμολογία προτίμησης ίση με δύο δεν υπάρχει καμία επίπτωση στην ακρίβεια του συνεργατικού φιλτραρίσματος. Συνεπώς, για τις ανάγκες της αξιολόγησης τα κρυμμένα αντικείμενα μπορούν να έχουν ελάχιστη βαθμολογία προτίμησης ίση με δύο. Διαισθητικά φαίνεται ότι η απόκρυψη αντικειμένων με μεγαλύτερη βαθμολογία προτίμησης είναι ευνοϊκότερη καθώς το σύστημα αναμένεται να προτείνει τα καλύτερα πακέτα αντικειμένων με όσο το δυνατόν μεγαλύτερο βαθμό προτίμησης, και παράλληλα, η αξιολόγηση σκοπεύει να

εξακριβώσει σε ποιο βαθμό αυτά τα πακέτα περιέχουν αντικείμενα από το σύνολο των κρυμμένων αντικειμένων. Συνεπώς, το σύνολο των κρυμμένων αντικειμένων θα πρέπει να περιέχει αντικείμενα με όσο το δυνατόν μεγαλύτερη βαθμολογία προτίμησης, χωρίς βέβαια να επηρεάζεται η ακρίβεια του συνεργατικού φίλτραρίσματος.

Σύμφωνα με τις μετρήσεις που έγιναν κατά το στάδιο της προετοιμασίας της αξιολόγησης και με βάση τα συμπεράσματα που προέκυψαν από αυτές, οι τιμές των δύο παραμέτρων για τη διαδικασία της αξιολόγησης επιλέγονται ως εξής: για το ποσοστό απόκρυψης των αντικειμένων στο 40% και για την ελάχιστη τιμή της βαθμολογίας προτίμησης των κρυμμένων αντικειμένων στην τιμή 2. Οι τιμές αυτές διατηρούνται σταθερές σε όλη τη διάρκεια της εφαρμογής και της αξιολόγησης.

6.5 Μετρήσεις Αξιολόγησης

Στα πλαίσια της αξιολόγησης του συστήματος πραγματοποιήθηκαν μετρήσεις που έχουν σκοπό την εξακρίβωση της αποτελεσματικότητας του αλγόριθμου αλλά και τη διερεύνηση των παραμέτρων εκείνων που η επιλογή τους μπορεί να οδηγήσει στη βελτίωση της αποτελεσματικότητας και της ακρίβειας. Η αξιολόγηση του αλγόριθμου, όπως έχει ήδη αναφερθεί, γίνεται με τη βοήθεια του σχετικού λογισμικού που αναπτύχθηκε, μέσω του οποίου εκτελούνται όλες οι λειτουργίες και λαμβάνονται όλες οι απαραίτητες μετρήσεις. Ακολουθούν οι μετρήσεις αξιολόγησης, που αφορούν τόσο στον αλγόριθμο όσο και στο σύστημα συστάσεων που τον ενσωματώνει, για τις διάφορες τιμές των παραμέτρων που καθορίζουν τη συμπεριφορά του συστήματος.

6.5.1 Αριθμός Αντικειμένων Συνεργατικού Φίλτραρίσματος

Η λειτουργία του αλγόριθμου που αναπτύχθηκε, όπως είναι ήδη γνωστό από το κεφάλαιο 3, βασίζεται κυρίως στη βέλτιστη αντιστοίχιση των αντικειμένων με τις κατηγορίες με στόχο τον σχηματισμό των καλύτερων πακέτων. Είναι προφανές ότι όσο περισσότερα είναι τα αντικείμενα τόσο μεγαλύτερο το υπολογιστικό φορτίο αυτού του αλγόριθμου. Με δεδομένο ότι οι εφαρμογές ενός τέτοιου συστήματος απευθύνονται κυρίως σε χρήστες που τους προτείνονται σχετικά λίγα πακέτα με ένα μικρό αριθμό αντικειμένων το καθένα, τα σχηματιζόμενα πακέτα δεν μπορούν να είναι πάρα πολλά και ο αριθμός των αντικειμένων που περιέχουν διατηρείται και αυτός σε χαμηλά επίπεδα. Αυτό έχει ως συνέπεια στις περισσότερες εφαρμογές να

μην απαιτείται μεγάλος όγκος αντικειμένων για τον σχηματισμό των καλύτερων πακέτων.

Ο στόχος των μετρήσεων που γίνονται σε αυτή την ενότητα είναι η εύρεση μίας τιμής για το πλήθος των προτεινόμενων αντικειμένων, που προκύπτουν από το συνεργατικό φιλτράρισμα, τέτοιας ώστε να υπάρχει ικανός αριθμός αντικειμένων για τον σχηματισμό των απαιτούμενων πακέτων και ταυτόχρονα το σύστημα να λειτουργεί αποδοτικά. Δηλαδή, το πλήθος των προτεινόμενων αντικειμένων που προορίζονται για τον σχηματισμό πακέτων να διατηρείται σε όσο το δυνατόν χαμηλότερα επίπεδα χωρίς αυτό να έχει επιπτώσεις στην ακρίβεια των προτάσεων.

Για τον σκοπό αυτό, έγιναν μετρήσεις για τα δεδομένα 10 χρηστών που έχουν βαθμολογήσει περίπου 1000 αντικείμενα ο καθένας. Σύμφωνα με τις παραμέτρους οι οποίες διατηρούνται σταθερές, το σύστημα προτείνει 5 πακέτα αντικειμένων που περιέχουν 8 αντικείμενα το καθένα με αναλογική κατανομή στις τέσσερις δημοφιλέστερες για τους χρήστες κατηγορίες. Επιλέχτηκε η δημιουργία διακριτών πακέτων ως μοντέλο παραγωγής πακέτων και οι δημοφιλέστερες κατηγορίες αντικειμένων θεωρούνται εκείνες από τις οποίες οι χρήστες έχουν επιλέξει τα περισσότερα αντικείμενα. Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές κατά το πείραμα συνοψίζονται στους πίνακες 6.5.1(α) και 6.5.1(β).

Οι μετρήσεις για τον *χρόνο εκτέλεσης* του αλγόριθμου, για τη *συνολική βαθμολογία προτίμησης* των αντικειμένων που περιέχονται στα προτεινόμενα πακέτα και για το δείκτη *precision* (με τα 95% διαστήματα εμπιστοσύνης) σε σχέση με το πλήθος των αντικειμένων που προέρχονται από το συνεργατικό φιλτράρισμα συγκεντρώνονται στον πίνακα 6.5.1(γ).

Επισημαίνεται ότι η συνολική βαθμολογία προτίμησης των αντικειμένων που αναγράφεται στον πίνακα 6.5.1(γ) είναι ίση με το άθροισμα των βαθμολογιών προτίμησης των αντικειμένων που περιέχονται σε όλα τα προτεινόμενα πακέτα σύμφωνα με τον τύπο:

$$\text{Συνολική βαθμολογία προτίμησης} = \sum_{i=1}^k \sum_{j=1}^n r_{ij}$$

Όπου r_{ij} : η βαθμολογία προτίμησης του αντικειμένου j του πακέτου i ,
 k : ο αριθμός των προτεινόμενων πακέτων και
 n : ο αριθμός των αντικειμένων σε κάθε προτεινόμενο πακέτο

Ποσοστό απόκρυψης αντικειμένων	Συντελεστής δημοφιλίας αντικειμένων	Ελάχιστη βαθμολογία προτίμησης κρυμμένων αντικειμένων
40%	5	2

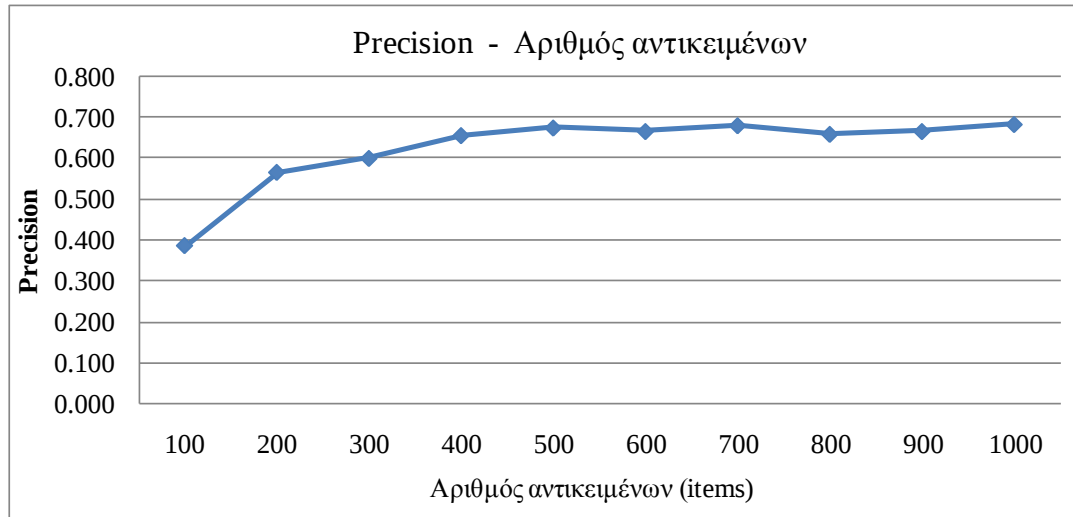
Πίνακας 6.5.1(α): Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης για τον αριθμό των αντικειμένων του συνεργατικού φιλτραρίσματος.

Αριθμός προτεινόμενων πακέτων	Αριθμός κατηγοριών	Αριθμός αντικειμένων ανά πακέτο	Μοντέλο παραγωγής πακέτων	Μοντέλο επιλογής των δημοφιλέστερων κατηγοριών
5	4	8 αντικείμενα αναλογικά καταναμεμένα στις κατηγορίες	6 Παραγωγή διακριτών πακέτων	1 Σύμφωνα με τον αριθμό των αντικειμένων που έχει επιλέξει ο χρήστης από κάθε κατηγορία

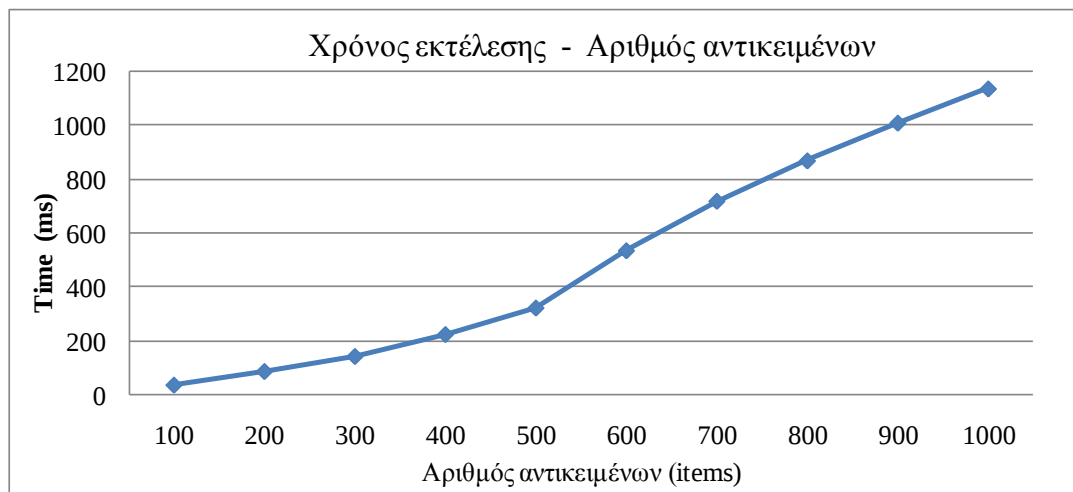
Πίνακας 6.5.1(β): Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης που αφορούν στον αριθμό αντικειμένων του συνεργατικού φιλτραρίσματος και στον συντελεστή δημοφιλίας των αντικειμένων.

A/A	Αριθμός προτεινόμενων αντικειμένων για τον σχηματισμό πακέτων	Χρόνος Εκτέλεσης (ms)	Συνολικός βαθμός προτίμησης αντικειμένων	Precision
1	100	35	131.95	0.384±0.062
2	200	86	136.00	0.564±0.036
3	300	142	142.80	0.599±0.046
4	400	224	142.84	0.655±0.048
5	500	323	142.67	0.674±0.063
6	600	536	143.21	0.665±0.045
7	700	720	142.97	0.680±0.053
8	800	871	143.61	0.658±0.061
9	900	1011	143.54	0.665±0.073
10	1000	1139	143.05	0.683±0.060

Πίνακας 6.5.1 (γ): Οι μετρήσεις για τον χρόνο εκτέλεσης του αλγόριθμου, για τη συνολική βαθμολογία προτίμησης (rating) των αντικειμένων που περιέχονται στα προτεινόμενα πακέτα και για το δείκτη precision (με τα 95% διαστήματα εμπιστοσύνης) σε σχέση με τον αριθμό των εκτιμώμενων αντικειμένων που προέρχονται από το συνεργατικό φιλτράρισμα.



Εικόνα 6.5.1 (α): Η μέση τιμή του δείκτη *Precision* για τα προτεινόμενα πακέτα σε σχέση με τον αριθμό των αντικειμένων για τον σχηματισμό πακέτων.



Εικόνα 6.5.1 (β): Ο χρόνος εκτέλεσης του αλγόριθμου σε σχέση με τον αριθμό των αντικειμένων για τον σχηματισμό πακέτων.

Στην εικόνα 6.5.1(α) φαίνεται η μεταβολή του δείκτη *precision* σε σχέση με την μεταβολή των προτεινόμενων αντικειμένων που είναι υποψήφια για τον σχηματισμό των καλύτερων πακέτων. Όπως παρατηρούμε από τη γραφική παράσταση η τιμή του δείκτη *precision* παραμένει σχεδόν σταθερή από τα 400 αντικείμενα και πάνω. Αυτό σημαίνει ότι σε εφαρμογές που απαιτούν την πρόταση ενός σχετικά μικρού αριθμού πακέτων αρκεί η χρησιμοποίηση των 500 ή 600 αντικειμένων με τη μεγαλύτερη αναμενόμενη βαθμολογία προτίμησης για τον σχηματισμό των καλύτερων πακέτων.

Στο διάγραμμα της εικόνας 6.5.1(β) φαίνεται η μεταβολή του χρόνου εκτέλεσης του αλγόριθμου για τη δημιουργία των καλύτερων πακέτων αντικειμένων σε σχέση

με τη μεταβολή του αριθμού των διαθέσιμων αντικειμένων. Σε αυτό το γράφημα φαίνεται ότι με την αύξηση των αντικειμένων αυξάνεται με μεγαλύτερο ρυθμό ο χρόνος εκτέλεσης του αλγόριθμου.

Το ζητούμενο από αυτές τις μετρήσεις είναι να βρεθεί ένας αριθμός αντικειμένων για τον οποίο θα επιτυγχάνονται ταυτόχρονα υψηλές επιδόσεις και υψηλή ακρίβεια. Για εφαρμογές που το μέγεθος των προτεινόμενων πακέτων είναι της τάξης εκείνων που χρησιμοποιούνται για τις παραπάνω μετρήσεις μπορούν να χρησιμοποιηθούν 500 αντικείμενα καθώς σε αυτή την τιμή επιτυγχάνεται χαμηλός χρόνος εκτέλεσης και η βέλτιστη ακρίβεια. Στις υπόλοιπες μετρήσεις αυτής της εργασίας τα προτεινόμενα αντικείμενα που προκύπτουν από το συνεργατικό φιλτράρισμα και χρησιμοποιούνται για τον σχηματισμό των πακέτων θα έχουν πλήθος ίσο με 500.

6.5.2 Συντελεστής Δημοφιλίας των Αντικειμένων

Σε αυτή την ενότητα θα διερευνηθεί η επιρροή του *συντελεστή της δημοφιλίας* των αντικειμένων στην ακρίβεια σχηματισμού πακέτων. Το ποσοστό της δημοφιλίας κάθε αντικειμένου δίνεται από τον λόγο των χρηστών που έχουν βαθμολογήσει το συγκεκριμένο αντικείμενο προς το συνολικό αριθμό των χρηστών. Το γινόμενο του συντελεστή δημοφιλίας επί το ποσοστό της δημοφιλίας, όπως αυτή έχει οριστεί, προστίθεται στον εκτιμώμενο βαθμό προτίμησης του κάθε αντικειμένου και προκύπτει ο νέος βαθμός προτίμησης για το αντικείμενο. Τα προτεινόμενα πακέτα σχηματίζονται χρησιμοποιώντας αυτόν τον νέο βαθμό προτίμησης που ενσωματώνει την εκτιμώμενη βαθμολογία προτίμησης και τη δημοφιλία των αντικειμένων. Η δημοφιλία των αντικειμένων, ο συντελεστής δημοφιλίας και ο *νέα βαθμολογία προτίμησης* που προκύπτει έχουν οριστεί στην ενότητα 3.2, ωστόσο για ευκολία και καλύτερη κατανόηση οι σχετικοί τύποι επαναλαμβάνονται και εδώ.

Το ποσοστό της δημοφιλίας του κάθε αντικειμένου:

$$Popularity(x) = \frac{\text{number of ratings of item } x}{\text{number of users}}$$

Όπου ο βαθμός δημοφιλίας του αντικειμένου x δίνεται από το λόγο του αριθμού των βαθμολογήσεων από ξεχωριστούς χρήστες για το αντικείμενο x προς τον συνολικό αριθμό των χρηστών.

Ο συντελεστής δημοφιλίας του κάθε αντικειμένου:

$$PopularityFactor \in \{0, 1, 2, \dots, 10\}$$

Ο τύπος που δίνει τη νέα βαθμολογία προτίμησης για το αντικείμενο x:

$$Rating_{New}(x) = Rating(x) + PopularityFactor \cdot Popularity(x)$$

Όπου $Rating(x)$: η εκτιμώμενη βαθμολογία προτίμησης του αντικειμένου x.

Η νέα αυτή βαθμολογία προτίμησης δίνει προτεραιότητα στα δημοφιλέστερα αντικείμενα σύμφωνα με την τιμή του συντελεστή δημοφιλίας που μπορεί να πάρει τιμές από μηδέν έως δέκα. Η τιμή μηδέν σημαίνει ότι δεν λαμβάνεται υπόψη η δημοφιλία των αντικειμένων, ενώ η τιμή 10 σημαίνει ότι δίνεται μεγάλη προτεραιότητα στα δημοφιλέστερα αντικείμενα. Να σημειωθεί ότι η νέα βαθμολογία προτίμησης κανονικοποιείται στην κλίμακα της εκάστοτε βαθμολογίας προτίμησης.

Για τις ανάγκες του πειράματος γίνονται μετρήσεις για τον σχηματισμό 5 διακριτών πακέτων που περιέχουν 8 αντικείμενα το καθένα από τις 4 δημοφιλέστερες για τους χρήστες κατηγορίες. Η κατανομή των 8 αντικειμένων στις 4 κατηγορίες είναι αναλογική σύμφωνα με τον αριθμό των αντικειμένων που έχουν βαθμολογήσει οι χρήστες από κάθε κατηγορία. Ως δημοφιλέστερες κατηγορίες αντικειμένων θεωρούνται εκείνες από τις οποίες οι χρήστες έχουν επιλέξει ή βαθμολογήσει τα περισσότερα αντικείμενα. Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές κατά το πείραμα συνοψίζονται στους πίνακες 6.5.1(β) και 6.5.2(α).

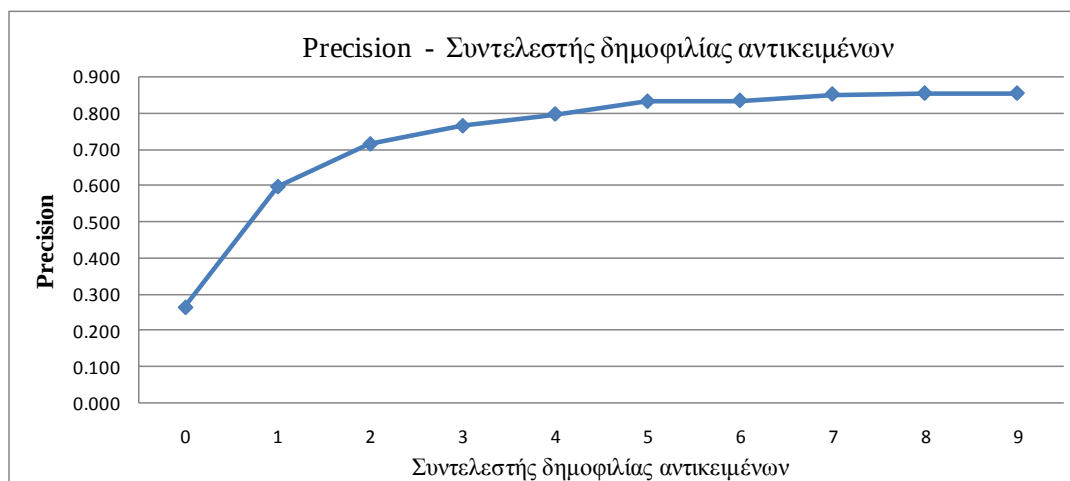
Ποσοστό απόκρυψης αντικειμένων	Αριθμός αντικειμένων που προέρχονται από το συνεργατικό φιλτράρισμα	Ελάχιστη βαθμολογία προτίμησης κρυμμένων αντικειμένων
40%	500	2

Πίνακας 6.5.2(α): Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης του αλγόριθμου που αφορούν στον συντελεστή δημοφιλίας των αντικειμένων.

Τα αποτελέσματα από τη μέση τιμή των μετρήσεων συγκεντρώνονται στον πίνακα 6.5.2(β) και απεικονίζονται στο διάγραμμα της εικόνας 6.5.2.

Συντελεστής δημοφιλίας των αντικειμένων	0	1	2	3	4	5	6	7	8	9
Precision	0.265	0.598	0.715	0.765	0.798	0.833	0.835	0.853	0.855	0.855

Πίνακας 6.5.2 (β): Ο δείκτης *Precision* για τις διάφορες τιμές του συντελεστή δημοφιλίας των αντικειμένων.



Εικόνα 6.5.2: Η μεταβολή του δείκτη *Precision* σε σχέση με τη μεταβολή του συντελεστή δημοφιλίας των αντικειμένων.

Όπως μπορούμε να διακρίνουμε στο διάγραμμα της εικόνας 6.5.2, η συμβολή του συντελεστή δημοφιλίας των αντικειμένων στη βελτίωση της ακρίβειας σχηματισμού των πακέτων είναι σημαντική. Το ζητούμενο εδώ είναι να προτείνονται πακέτα αντικειμένων που βασίζονται περισσότερο στα ιδιαίτερα χαρακτηριστικά προτιμήσεων του κάθε χρήστη και λιγότερο στα χαρακτηριστικά προτιμήσεων του συνόλου των χρηστών. Συνεπώς, η τιμή του συντελεστή αυτού θα πρέπει να είναι όσο το δυνατόν μικρότερη και ταυτόχρονα να επιτυγχάνεται η βέλτιστη συμβολή του στη βελτίωση της ακρίβειας του συστήματος. Όπως παρατηρούμε στο διάγραμμα 6.5.2 η μικρότερη τιμή του συντελεστή δημοφιλίας των αντικειμένων για την οποία έχουμε τη μεγαλύτερη βελτίωση του δείκτη *Precision* είναι η τιμή 5. Για μεγαλύτερες τιμές του συντελεστή αυτού φαίνεται να διατηρείται σχεδόν σταθερός ο δείκτης *Precision*.

Σύμφωνα με όλα τα προηγούμενα, στις υπόλοιπες μετρήσεις αυτής της εργασίας θα διατηρηθεί η τιμή 5 για τον συντελεστή της δημοφιλίας των αντικειμένων.

6.5.3 Αριθμός Προτεινόμενων Πακέτων

Σε αυτή την ενότητα γίνεται διερεύνηση για την ακρίβεια που ο αλγόριθμος πετυχαίνει και για το χρόνο εκτέλεσης που το σύστημα χρειάζεται σε σχέση με τον

αριθμό των προτεινόμενων πακέτων. Για τον σκοπό αυτό γίνονται μετρήσεις μεταβάλλοντας τον αριθμό των προτεινόμενων πακέτων και ταυτόχρονα διατηρώντας σταθερές τις υπόλοιπες παραμέτρους σύμφωνα με τους πίνακες 6.5.3(α) και 6.5.3(β).

Ποσοστό απόκρυψης αντικειμένων	Αριθμός αντικειμένων για τον σχηματισμό πακέτων που προέρχονται από το συνεργατικό φιλτράρισμα	Συντελεστής δημοφιλίας αντικειμένων	Ελάχιστη βαθμολογία προτίμησης κρυμμένων αντικειμένων
40%	500	5	2

Πίνακας 6.5.3(α): Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης του αλγόριθμου που αφορούν στον αριθμό των πακέτων, στον αριθμό των κατηγοριών, στον αριθμό των αντικειμένων ανά πακέτο, στο μοντέλο παραγωγής πακέτων και στο μοντέλο επιλογής των δημοφιλέστερων κατηγοριών.

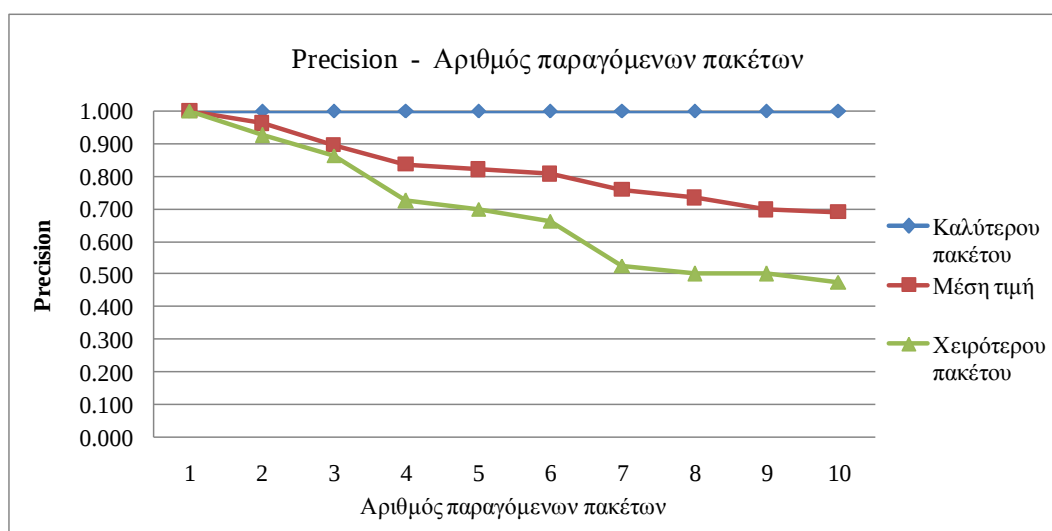
Αριθμός κατηγοριών	Αριθμός αντικειμένων ανά πακέτο	Μοντέλο παραγωγής πακέτων	Μοντέλο επιλογής των δημοφιλέστερων κατηγοριών
4	8 αντικείμενα αναλογικά κατανεμημένα στις κατηγορίες	6 Παραγωγή διακριτών πακέτων	1 Σύμφωνα με τον αριθμό των αντικειμένων που έχει επιλέξει ο χρήστης από κάθε κατηγορία

Πίνακας 6.5.3(β): Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης που αφορούν στον αριθμό των προτεινόμενων πακέτων.

Α/Α	Αριθμός παραγόμενων Πακέτων	Χρόνος Εκτέλεσης (ms)	Precision		
			Χειρότερου πακέτου	Καλύτερου πακέτου	Μέση τιμή
1	1	78	1.000	1.000	1.000±0
2	2	142	0.925	1.000	0.963±0.033
3	3	198	0.863	1.000	0.897±0.045
4	4	257	0.725	1.000	0.835±0.061
5	5	307	0.700	1.000	0.820±0.053
6	6	363	0.663	1.000	0.806±0.051
7	7	416	0.525	1.000	0.759±0.041
8	8	491	0.500	1.000	0.733±0.043
9	9	530	0.500	1.000	0.697±0.035
10	10	577	0.475	1.000	0.691±0.030

Πίνακας 6.5.3 (γ): Ο χρόνος εκτέλεσης του αλγόριθμου και ο δείκτης Precision, του χειρότερου πακέτου, του καλύτερου πακέτου και η μέση τιμή με το 95% διάστημα εμπιστοσύνης, σε σχέση με τον αριθμό των παραγόμενων πακέτων.

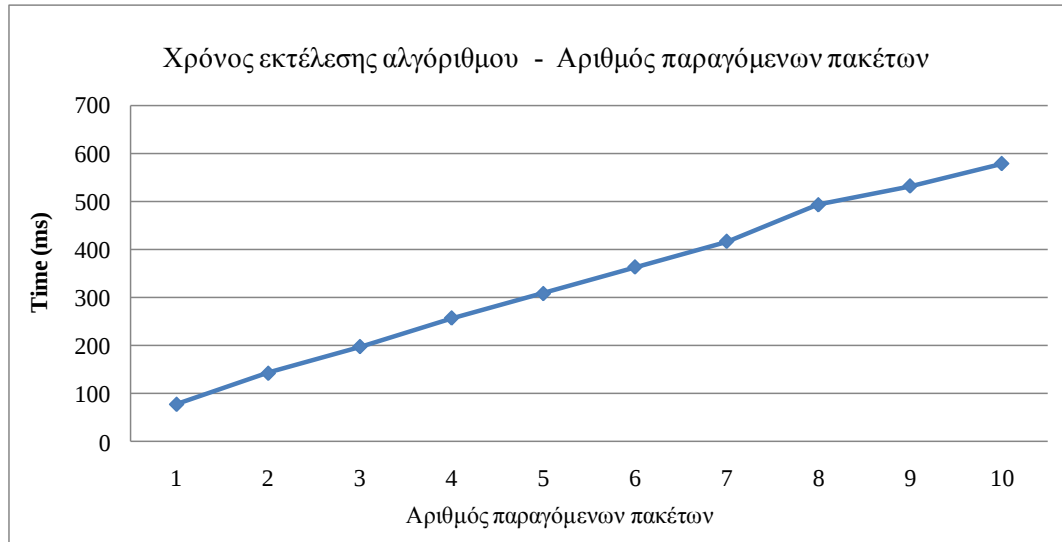
Διατηρώντας σταθερές τις παραμέτρους των πινάκων 6.5.3(α) και 6.5.3(β) πραγματοποιήθηκαν μετρήσεις για 10 χρήστες που έχουν βαθμολογήσει περίπου 1000 αντικείμενα ο καθένας και οι μέσοι όροι των τιμών συγκεντρώνονται στον πίνακα 6.5.3(γ). Οι τιμές των μετρήσεων αυτών απεικονίζονται γραφικά στα διαγράμματα των εικόνων 6.5.3(α) και 6.5.3(β).



Εικόνα 6.5.3 (α): Η μεταβολή του δείκτη Precision, του χειρότερου πακέτου, του καλύτερου πακέτου και η μέση τιμή στο σύνολο των πακέτων, σε σχέση με τον αριθμό των παραγόμενων πακέτων.

Σύμφωνα με το διάγραμμα της εικόνας 6.5.3(α), που απεικονίζει τη μεταβολή του δείκτη *precision* σε σχέση με τον αριθμό των προτεινόμενων πακέτων, παρατηρούμε ότι για το καλύτερο πακέτο έχουμε 100% ακρίβεια ανεξάρτητα από τον αριθμό των παραγόμενων πακέτων, ενώ για τη μέση τιμή του δείκτη *precision* όλων των πακέτων έχουμε μια μείωση της τιμής του με την αύξηση των προτεινόμενων πακέτων. Ωστόσο, όπως μπορούμε να διακρίνουμε από το διάγραμμα η τιμή του δείκτη *precision* διατηρείται σε υψηλά επίπεδα.

Στο διάγραμμα της εικόνας 6.5.3(β) παρατηρούμε μια γραμμική αύξηση του χρόνου εκτέλεσης σε σχέση με τον αριθμό των προτεινόμενων πακέτων που το σύστημα εξάγει. Αξίζει να αναφερθεί ότι ο χρόνος εκτέλεσης διατηρείται σε χαμηλά επίπεδα καθώς για την παραγωγή των 10 καλύτερων πακέτων δεν ξεπερνά τα 0,6 δευτερόλεπτα.



Εικόνα 6.5.3 (β): Η μεταβολή του χρόνου εκτέλεσης του αλγόριθμου σε σχέση με τον αριθμό των παραγόμενων πακέτων.

6.5.4 Αριθμός Κατηγοριών

Σε αυτή την ενότητα γίνεται διερεύνηση για την ακρίβεια που ο αλγόριθμος πετυχαίνει και για τον χρόνο εκτέλεσης που το σύστημα χρειάζεται σε σχέση με τον αριθμό των δημοφιλέστερων για τον χρήστη κατηγοριών που τα προτεινόμενα πακέτα περιλαμβάνουν. Για τον σκοπό αυτό γίνονται μετρήσεις μεταβάλλοντας τον αριθμό των κατηγοριών με την ταυτόχρονη διατήρηση των υπόλοιπων παραμέτρων σε σταθερή κατάσταση.

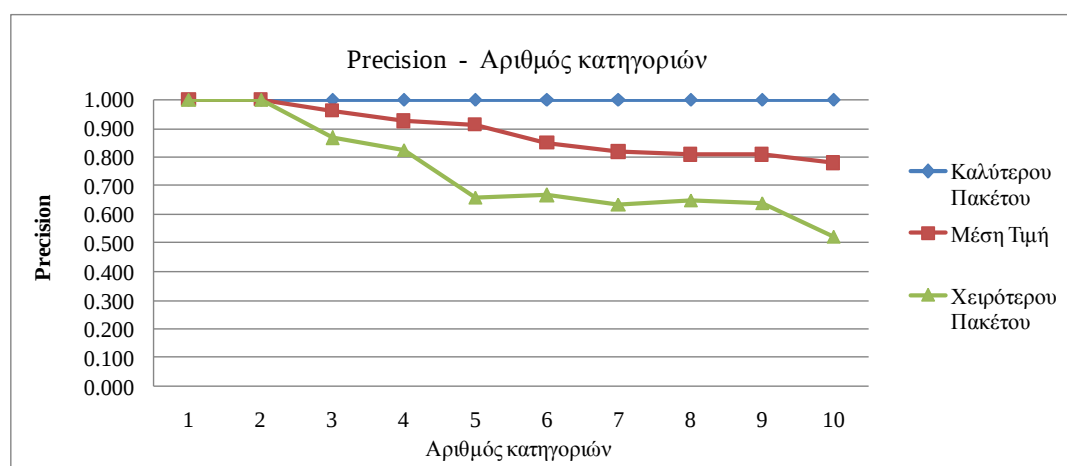
Για το πείραμα διατηρούνται σταθερές οι τιμές των παραμέτρων που αναγράφονται στους πίνακες 6.5.3(α) και 6.5.4(α). Πραγματοποιήθηκαν μετρήσεις για 10 χρήστες που έχουν βαθμολογήσει περίπου 1000 αντικείμενα ο καθένας και οι μέσοι όροι των τιμών συγκεντρώνονται στον πίνακα 6.5.4(β). Οι τιμές των μετρήσεων αυτών απεικονίζονται γραφικά στα διαγράμματα των εικόνων 6.5.4(α) και 6.5.4(β).

Αριθμός προτεινόμενων πακέτων	Αριθμός αντικειμένων ανά κατηγορία	Μοντέλο παραγωγής πακέτων	Μοντέλο επιλογής των δημοφιλέστερων κατηγοριών
5	1	6 Παραγωγή διακριτών πακέτων	1 Σύμφωνα με τον αριθμό των αντικειμένων που έχει επιλέξει ο χρήστης από κάθε κατηγορία

Πίνακας 6.5.4(α): Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης που αφορούν στη μεταβολή του αριθμού των κατηγοριών.

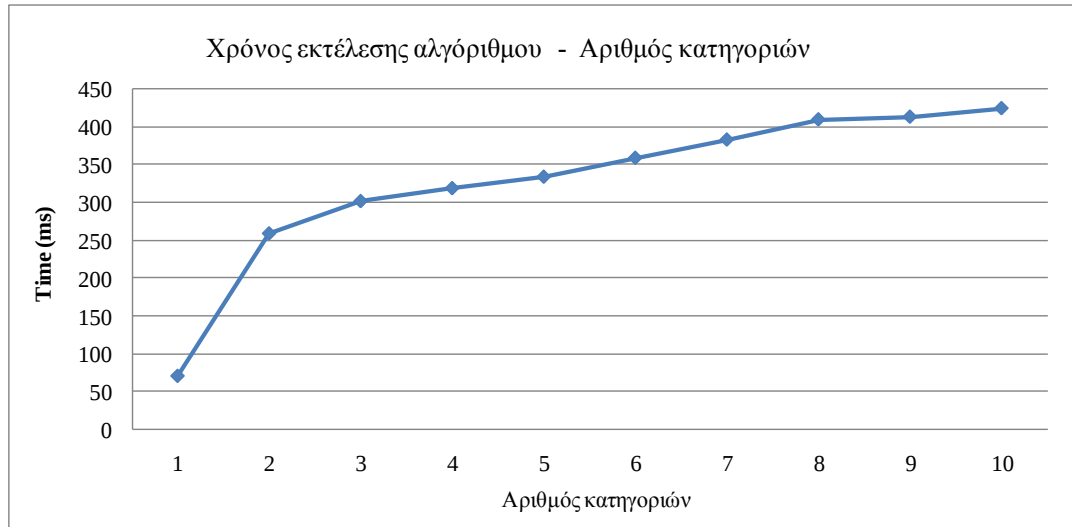
Α/Α	Αριθμός Κατηγοριών	Χρόνος Εκτέλεσης (ms)	Precision		
			Χειρότερου πακέτου	Καλύτερου πακέτου	Μέση τιμή
1	1	71	1.000	1.000	1.000±0
2	2	260	1.000	1.000	1.000±0
3	3	302	0.867	1.000	0.960±0.035
4	4	319	0.825	1.000	0.925±0.042
5	5	334	0.660	1.000	0.912±0.038
6	6	359	0.667	1.000	0.850±0.058
7	7	384	0.636	1.000	0.817±0.066
8	8	410	0.650	1.000	0.808±0.073
9	9	414	0.638	1.000	0.808±0.076
10	10	425	0.523	1.000	0.778±0.055

Πίνακας 6.5.4 (β): Ο χρόνος εκτέλεσης του αλγόριθμου και ο δείκτης *Precision*, του χειρότερου πακέτου, του καλύτερου πακέτου και της μέσης τιμής με τα 95% διαστήματα εμπιστοσύνης, σε σχέση με τον αριθμό των κατηγοριών των αντικειμένων που περιέχονται στα προτεινόμενα πακέτα.



Εικόνα 6.5.4 (α): Η μεταβολή του δείκτη *Precision*, του χειρότερου πακέτου, του καλύτερου πακέτου και της μέσης τιμής στο σύνολο των προτεινόμενων πακέτων, σε σχέση με τον αριθμό των κατηγοριών των αντικειμένων που περιέχονται στα προτεινόμενα πακέτα.

Σύμφωνα με το διάγραμμα της εικόνας 6.5.4(α), που απεικονίζει τη μεταβολή του δείκτη *precision* σε σχέση με τον αριθμό των κατηγοριών, παρατηρούμε ότι για το καλύτερο πακέτο έχουμε 100% ακρίβεια ανεξάρτητα από τον αριθμό των κατηγοριών, ενώ για τη μέση τιμή του δείκτη *precision* όλων των πακέτων έχουμε μια μείωση της τιμής του με την αύξηση των κατηγοριών. Ωστόσο, όπως φαίνεται στο διάγραμμα, η μέση τιμή του δείκτη *precision* διατηρείται υψηλή, σε επίπεδα μεγαλύτερα του 75%, ακόμη και στην περίπτωση που κάθε πακέτο περιλαμβάνει αντικείμενα από 10 κατηγορίες.



Εικόνα 6.5.4 (β): Η μεταβολή του χρόνου εκτέλεσης του αλγόριθμου σε σχέση με τον αριθμό των κατηγοριών των αντικειμένων που περιέχονται στα προτεινόμενα πακέτα.

Στο διάγραμμα της εικόνας 6.5.4(β) παρατηρούμε ότι αυξάνεται ο χρόνος εκτέλεσης με την αύξηση του αριθμού των κατηγοριών. Ο ρυθμός της αύξησης του χρόνου εκτέλεσης μειώνεται με την αύξηση των κατηγοριών. Αυτό μπορεί να συμβαίνει γιατί οι λιγότερο δημοφιλείς για το χρήστη κατηγορίες περιλαμβάνουν λιγότερα αντικείμενα με συνέπεια να μειώνεται το υπολογιστικό φορτίο που απαιτείται για αυτές τις κατηγορίες. Αξίζει και εδώ να αναφερθεί ότι για την παραγωγή των 5 καλύτερων πακέτων που το καθένα περιέχει 10 αντικείμενα από τις 10 δημοφιλέστερες για τον τρέχοντα χρήστη κατηγορίες, ο χρόνος εκτέλεσης διατηρείται στα χαμηλά επίπεδα της τάξης των 0.4 δευτερολέπτων.

6.5.5 Αριθμός Αντικειμένων ανά Πακέτο

Σε αυτή την ενότητα γίνεται διερεύνηση για την ακρίβεια που ο αλγόριθμος πετυχαίνει και για το χρόνο εκτέλεσης που το σύστημα χρειάζεται σε σχέση με τον αριθμό των αντικειμένων ανά προτεινόμενο πακέτο. Η διερεύνηση αυτή διενεργείται σε δύο μέρη, στο πρώτο μεταβάλλεται ο αριθμός των αντικειμένων ανά πακέτο με τη μεταβολή του αριθμού των αντικειμένων ανά κατηγορία και στο δεύτερο έχουμε αναλογική κατανομή των αντικειμένων στις δημοφιλέστερες για το χρήστη κατηγορίες. Ακολουθούν οι μετρήσεις και τα αποτελέσματα για τους δύο τρόπους μεταβολής του αριθμού των αντικειμένων που περιλαμβάνουν τα προτεινόμενα πακέτα.

6.5.5.1 Μεταβολή του αριθμού των αντικειμένων ανά κατηγορία

Διερευνήθηκε η μεταβολή για τον χρόνο εκτέλεσης του αλγόριθμου και τον δείκτη *Precision* του χειρότερου πακέτου, του καλύτερου πακέτου και της μέσης τιμής στο σύνολο των προτεινόμενων πακέτων, σε σχέση με τον αριθμό των αντικειμένων ανά κατηγορία. Εδώ να σημειωθεί ότι με τη μεταβολή του αριθμού των αντικειμένων ανά κατηγορία μεταβάλλεται και ο αριθμός των αντικειμένων ανά πακέτο καθώς το κάθε πακέτο περιλαμβάνει έναν προκαθορισμένο αριθμό κατηγοριών.

Αριθμός προτεινόμενων πακέτων	Αριθμός Κατηγοριών	Μοντέλο παραγωγής πακέτων	Μοντέλο επιλογής των δημοφιλέστερων κατηγοριών
5	4	6 Παραγωγή διακριτών πακέτων	1 Σύμφωνα με τον αριθμό των αντικειμένων που έχει επιλέξει ο χρήστης από κάθε κατηγορία

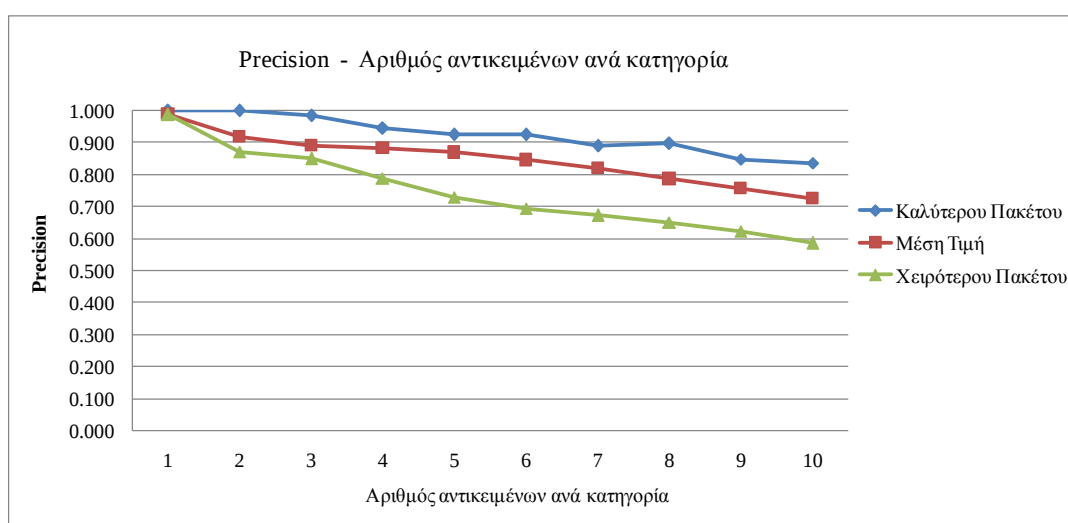
Πίνακας 6.5.5.1 (α): Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης που αφορούν στη μεταβολή του αριθμού των αντικειμένων ανά κατηγορία προτεινόμενου πακέτου.

A/A	Αριθμός αντικειμένων ανά κατηγορία	Αριθμός αντικειμένων ανά πακέτο	Χρόνος εκτέλεσης (ms)	Precision		
				Χειρότερου πακέτου	Καλύτερου πακέτου	Μέση τιμή
1	1	4	351	0.988	1.000	0.990±0.013
2	2	8	347	0.873	1.000	0.920±0.036
3	3	12	335	0.850	0.983	0.890±0.048
4	4	16	329	0.788	0.944	0.884±0.045
5	5	20	328	0.730	0.925	0.870±0.048
6	6	24	325	0.695	0.925	0.846±0.039
7	7	28	325	0.673	0.889	0.820±0.039
8	8	32	323	0.650	0.897	0.787±0.041
9	9	36	323	0.623	0.847	0.757±0.030
10	10	40	321	0.586	0.835	0.724±0.036

Πίνακας 6.5.5.1 (β): Ο χρόνος εκτέλεσης του αλγόριθμου και ο δείκτης *Precision*, του χειρότερου πακέτου, του καλύτερου πακέτου και της μέσης τιμής με το 95% διάστημα εμπιστοσύνης, σε σχέση με τον αριθμό των αντικειμένων ανά κατηγορία ή τον αριθμό των αντικειμένων ανά πακέτο.

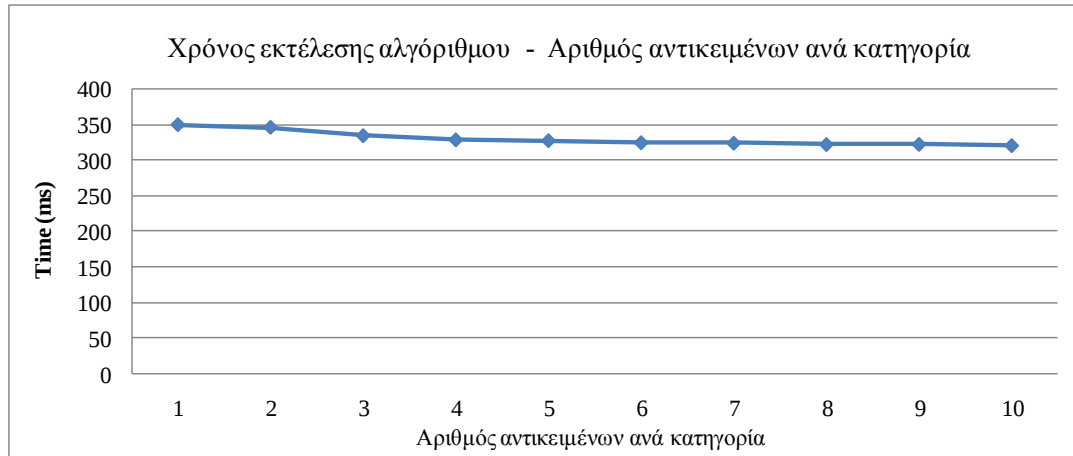
Για τις μετρήσεις διατηρήθηκαν σταθερές οι τιμές των παραμέτρων που αναγράφονται στους πίνακες 6.5.3(α) και 6.5.5.1(α) και πραγματοποιήθηκαν μετρήσεις για 10 χρήστες που έχουν βαθμολογήσει περίπου 1000 αντικείμενα ο καθένας και οι μέσοι όροι των μετρήσεων συγκεντρώνονται στον πίνακα 6.5.5.1(β). Οι τιμές των μετρήσεων αυτών απεικονίζονται γραφικά στα διαγράμματα των εικόνων 6.5.5.1(α) και 6.5.5.1(β).

Στο διάγραμμα της εικόνας 6.5.5.1(α), που απεικονίζει τη μεταβολή του δείκτη *precision* σε σχέση με τον αριθμό των αντικειμένων ανά κατηγορία, παρατηρούμε ότι με την αύξηση του αριθμού των αντικειμένων έχουμε μείωση της τιμής του δείκτη *Precision* τόσο για το χειρότερο και το καλύτερο προτεινόμενο πακέτο όσο και για τη μέση τιμή στο σύνολο των προτεινόμενων πακέτων. Ωστόσο, όπως φαίνεται στο διάγραμμα η μέση τιμή του δείκτη *Precision* διατηρείται υψηλή, σε επίπεδα μεγαλύτερα του 70%, ακόμη και στην περίπτωση που κάθε πακέτο περιλαμβάνει 10 αντικείμενα ανά κατηγορία με το μέγεθος του πακέτου να φτάνει τα 40 αντικείμενα.



Εικόνα 6.5.5.1 (α): Η μεταβολή του δείκτη *Precision*, του χειρότερου πακέτου, του καλύτερου πακέτου και της μέσης τιμής, σε σχέση με τον αριθμό των αντικειμένων ανά κατηγορία.

Στο διάγραμμα της εικόνας 6.5.5.1(β) παρατηρούμε ότι με την αύξηση του αριθμού των αντικειμένων ανά κατηγορία διατηρείται πρακτικά σταθερός ο χρόνος εκτέλεσης του αλγόριθμου. Μία μικρή μείωση που παρατηρείται στο χρόνο εκτέλεσης με την αύξηση του αριθμού των αντικειμένων ανά κατηγορία μπορεί να οφείλεται στη μείωση του φορτίου του αλγόριθμου, καθώς με την αύξηση των αντικειμένων ανά κατηγορία μειώνονται τα περιθώρια εξέτασης εναλλακτικών λύσεων. Ωστόσο, αυτή η μείωση, της τάξης των 25ms, μπορεί να θεωρηθεί πρακτικά αμελητέα. Αυτό σημαίνει ότι η αύξηση του αριθμού των αντικειμένων ανά κατηγορία δεν επηρεάζει σημαντικά τις επιδόσεις του αλγόριθμου.



Εικόνα 6.5.5.1 (β): Η μεταβολή του χρόνου εκτέλεσης του αλγόριθμου σε σχέση με τον αριθμό των αντικειμένων ανά κατηγορία.

6.5.5.2 Αναλογικά κατανομημένα αντικείμενα στις δημοφιλέστερες κατηγορίες

Διερευνήθηκε η μεταβολή του χρόνου εκτέλεσης του αλγόριθμου και του δείκτη *Precision*, του καλύτερου πακέτου, του χειρότερου πακέτου και του μέσου όρου του συνόλου των προτεινόμενων πακέτων, σε σχέση με τον αριθμό των αντικειμένων ανά πακέτο όταν τα αντικείμενα κατανέμονται αναλογικά στις κατηγορίες.

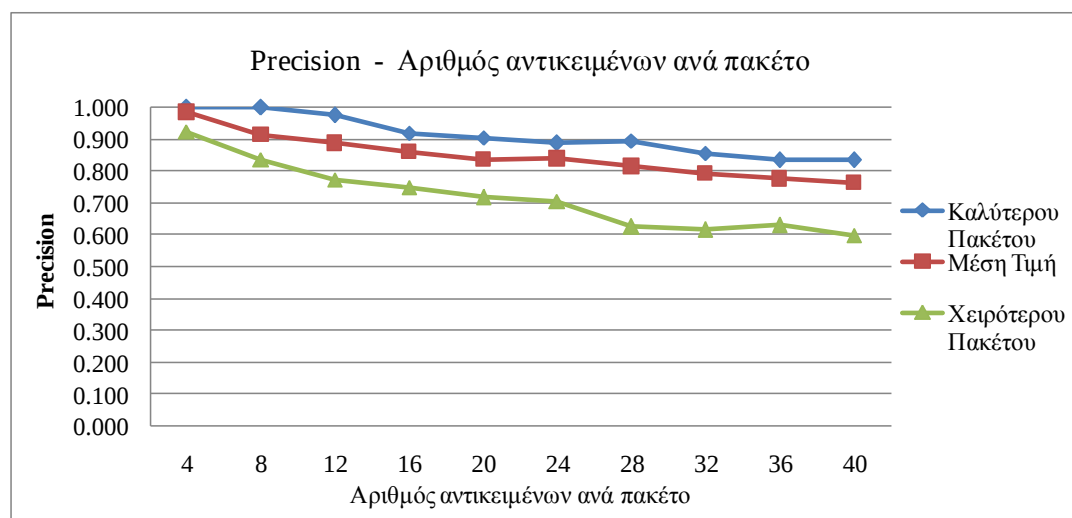
Για τις ανάγκες του πειράματος διατηρήθηκαν σταθερές οι τιμές των παραμέτρων που αναγράφονται στους πίνακες 6.5.3(α) και 6.5.5.2(α). Πραγματοποιήθηκαν μετρήσεις για 10 χρήστες που έχουν βαθμολογήσει περίπου 1000 αντικείμενα ο καθένας και οι μέσοι όροι των μετρήσεων συγκεντρώνονται στον πίνακα 6.5.5.2(β). Οι τιμές των μετρήσεων αυτών απεικονίζονται γραφικά στα διαγράμματα των εικόνων 6.5.5.2(α) και 6.5.5.2(β).

Αριθμός προτεινόμενων πακέτων	Αριθμός Κατηγοριών	Μοντέλο παραγωγής πακέτων	Μοντέλο επιλογής των δημοφιλέστερων κατηγοριών
5	4	6 Παραγωγή διακριτών πακέτων	1 Σύμφωνα με τον αριθμό των αντικειμένων που έχει επιλέξει ο χρήστης από κάθε κατηγορία

Πίνακας 6.5.5.2 (α): Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης που αφορούν στη μεταβολή του αριθμού των αντικειμένων ανά προτεινόμενο πακέτο με αναλογική κατανομή των αντικειμένων στις δημοφιλέστερες κατηγορίες.

Α/Α	Αριθμός αντικειμένων ανά πακέτο	Χρόνος εκτέλεσης (ms)	Precision		
			Χειρότερου πακέτου	Καλύτερου πακέτου	Μέση τιμή
1	4	333	0.925	1.000	0.985±0.015
2	8	339	0.838	1.000	0.915±0.038
3	12	338	0.775	0.975	0.889±0.042
4	16	329	0.750	0.919	0.863±0.052
5	20	331	0.718	0.905	0.836±0.053
6	24	327	0.705	0.888	0.840±0.044
7	28	331	0.625	0.896	0.816±0.044
8	32	331	0.615	0.856	0.792±0.047
9	36	335	0.631	0.835	0.776±0.047
10	40	321	0.598	0.835	0.763±0.052

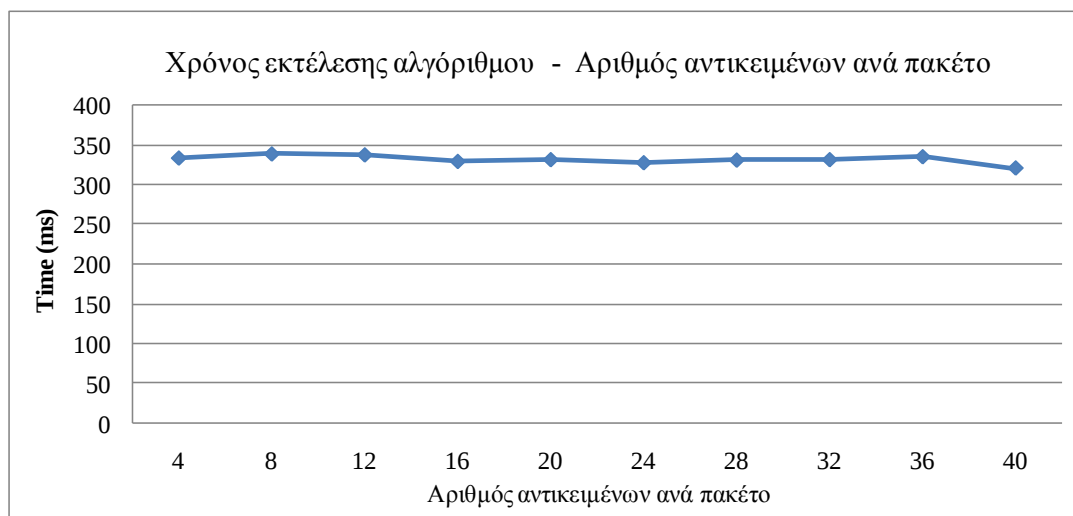
Πίνακας 6.5.5.2 (β): Ο χρόνος εκτέλεσης του αλγόριθμου και ο δείκτης *Precision*, του χειρότερου πακέτου, του καλύτερου πακέτου και της μέσης τιμής στο σύνολο των πακέτων με το 95% διάστημα εμπιστοσύνης, σε σχέση με τον αριθμό των αντικειμένων ανά πακέτο για αναλογική κατανομή των αντικειμένων στις κατηγορίες.



Εικόνα 6.5.5.2 (α): Η μεταβολή του δείκτη *Precision*, του χειρότερου πακέτου, του καλύτερου πακέτου και η μέση τιμή, σε σχέση με τον αριθμό των αντικειμένων ανά πακέτο για αναλογική κατανομή των αντικειμένων σε τέσσερις κατηγορίες.

Στο διάγραμμα της εικόνας 6.5.5.2(α), που απεικονίζει τη μεταβολή του δείκτη *precision* σε σχέση με τον αριθμό των αντικειμένων ανά πακέτο, παρατηρούμε ότι με την αύξηση του αριθμού των αντικειμένων έχουμε μείωση της τιμής του δείκτη *Precision* τόσο για το χειρότερο και το καλύτερο προτεινόμενο πακέτο όσο και για τη μέση τιμή στο σύνολο των προτεινόμενων πακέτων. Ωστόσο και εδώ, όπως φαίνεται στο διάγραμμα, η μέση τιμή του δείκτη *Precision* διατηρείται υψηλή, σε

επίπεδα μεγαλύτερα του 75%, ακόμη και στην περίπτωση που κάθε πακέτο περιλαμβάνει 40 αντικείμενα.



Εικόνα 6.5.5.2 (β): Η μεταβολή του χρόνου εκτέλεσης του αλγόριθμου σε σχέση με τον αριθμό των αντικειμένων ανά πακέτο για αναλογική κατανομή των αντικειμένων σε τέσσερις κατηγορίες.

Στο διάγραμμα της εικόνας 6.5.5.2(β) παρατηρούμε ότι με την αύξηση του αριθμού των αντικειμένων ανά πακέτο διατηρείται σχεδόν σταθερός ο χρόνος εκτέλεσης του αλγόριθμου. Αυτό σημαίνει ότι η αύξηση του αριθμού των αντικειμένων ανά κατηγορία δεν επηρεάζει σημαντικά τις επιδόσεις του αλγόριθμου.

Σε μία σύγκριση των διαγραμμάτων των εικόνων 6.5.5.1(α) 6.5.5.2(α) που απεικονίζουν τον δείκτη *Precision* σε σχέση με τον αριθμό των αντικειμένων ανά πακέτο παρατηρούμε ότι στην δεύτερη περίπτωση, που έχουμε αναλογική κατανομή των αντικειμένων στις δημοφιλέστερες για τον χρήστη κατηγορίες, επιτυγχάνονται υψηλότερες τιμές για τον δείκτη *precision*. Αυτό είναι αναμενόμενο καθώς στη δεύτερη περίπτωση τα προτεινόμενα πακέτα προσεγγίζουν περισσότερο τις προτιμήσεις των χρηστών, περιλαμβάνοντας περισσότερα αντικείμενα από τις δημοφιλέστερες για τους χρήστες κατηγορίες, οδηγώντας τον αλγόριθμο σε υψηλότερα ποσοστά ακρίβειας.

Επίσης, σε μία σύγκριση των διαγραμμάτων των εικόνων 6.5.5.1(β) και 6.5.5.2(β) που απεικονίζουν τον χρόνο εκτέλεσης του αλγόριθμου σε σχέση με το μέγεθος των προτεινόμενων πακέτων παρατηρούμε ότι και στις δύο περιπτώσεις έχουμε πρακτικά τον ίδιο χρόνο εκτέλεσης. Συνεπώς, ο χρόνος εκτέλεσης δεν φαίνεται να επηρεάζεται από τη μεταβολή του μεγέθους των προτεινόμενων

πακέτων αλλά ούτε και από τον τρόπο που τα αντικείμενα των πακέτων κατανέμονται στις κατηγορίες.

6.5.6 Μοντέλο Παραγωγής των Καλύτερων Προτεινόμενων Πακέτων

Μια σημαντική παράμετρος που μεταβάλλει τη συμπεριφορά του αλγόριθμου είναι το χρησιμοποιούμενο μοντέλο για την παραγωγή των προτεινόμενων πακέτων. Όπως έχουν ήδη περιγραφεί στο κεφάλαιο 3, τα μοντέλα παραγωγής πακέτων είναι τα εξής:

1. Διαγραφή μέγιστης ακμής (*DelMax*)
2. Διαγραφή ελάχιστης ακμής (*DelMin*)
3. Διαγραφή του 50% των μέγιστων ακμών (*DelHalfMax*)
4. Διαγραφή του 50% των ελάχιστων ακμών (*DelHalfMin*)
5. Διαγραφή του 100% των ακμών (*DelAll*)
6. Παραγωγή διακριτών πακέτων (*Distinct*)

Για τη διενέργεια του πειράματος διατηρήθηκαν σταθερές οι τιμές των παραμέτρων που αναγράφονται στους πίνακες 6.5.3(α) και 6.5.6(α) και πραγματοποιήθηκαν μετρήσεις για 10 χρήστες που έχουν βαθμολογήσει περίπου 1000 αντικείμενα ο καθένας και οι μέσοι όροι των μετρήσεων συγκεντρώνονται στον πίνακα 6.5.6(β). Επίσης, οι τιμές των μετρήσεων για το *Rating*, το *Precision* και τον χρόνο εκτέλεσης απεικονίζονται στα διαγράμματα των εικόνων 6.5.6(α), 6.5.6(β) και 6.5.6(γ).

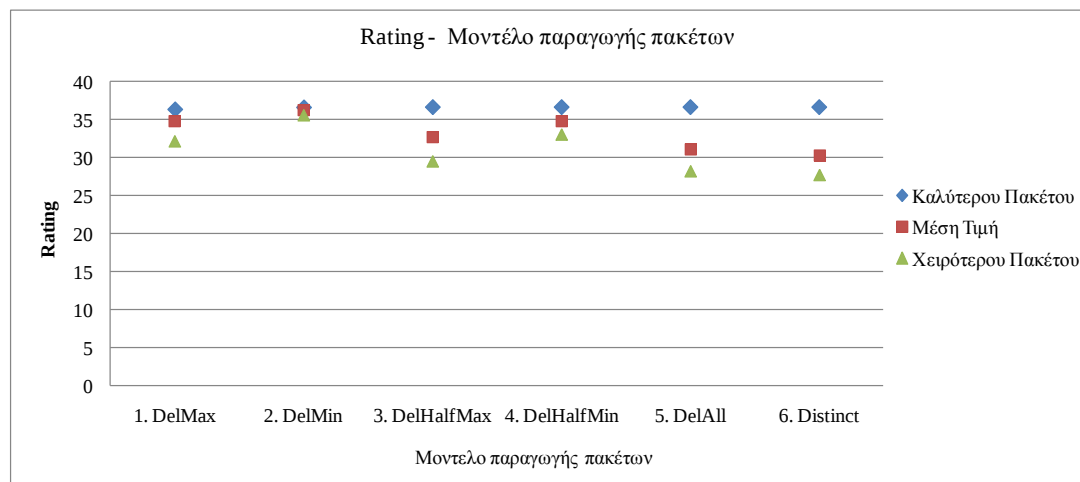
Αριθμός προτεινόμενων πακέτων	Αριθμός Κατηγοριών	Αριθμός αντικειμένων ανά πακέτο	Μοντέλο επιλογής των δημοφιλέστερων κατηγοριών
10	4	8 Αναλογικά κατανεμημένα στις κατηγορίες	1 Σύμφωνα με τον αριθμό των αντικειμένων που έχει επιλέξει ο χρήστης από κάθε κατηγορία

Πίνακας 6.5.6 (α): Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης που αφορούν στα διάφορα μοντέλα παραγωγής πακέτων.

Στον πίνακα 6.5.6(β) συγκεντρώνονται οι μετρήσεις για τον χρόνο εκτέλεσης του αλγόριθμου, για το *Rating* και το *Precision* του χειρότερου πακέτου, του καλύτερου πακέτου και της μέσης τιμής στο σύνολο των πακέτων, σε σχέση με το χρησιμοποιούμενο μοντέλο παραγωγής πακέτων.

Μοντέλο παραγωγής πακέτων	Χρόνος εκτέλεσης (ms)	Rating			Precision		
		Χειρότερου πακέτου	Καλύτερου πακέτου	Μέση Τιμή	Χειρότερου πακέτου	Καλύτερου πακέτου	Μέση τιμή
1. DelMax	654	32.10	36.35	34.70±0.46	0.938	1.000	0.970±0.02
2. DelMin	662	35.55	36.61	36.17±0.43	0.900	1.000	0.983±0.01
3. DelHalfMax	641	29.45	36.65	32.62±0.48	0.750	1.000	0.902±0.04
4. DelHalfMin	631	33.00	36.65	34.76±0.46	0.763	1.000	0.928±0.03
5. DelAll	612	28.14	36.65	31.04±0.44	0.588	1.000	0.784±0.04
6. Distinct	577	27.65	36.65	30.26±0.66	0.513	1.000	0.743±0.04

Πίνακας 6.5.6 (β): Μετρήσεις για τον χρόνο εκτέλεσης του αλγόριθμου, το *Rating* και το *Precision* για το χειρότερο πακέτο, το καλύτερο πακέτο και για τη μέση τιμή όλων των προτεινόμενων πακέτων με τα 95% διαστήματα εμπιστοσύνης, σε σχέση με τα έξι μοντέλα παραγωγής των προτεινόμενων πακέτων.

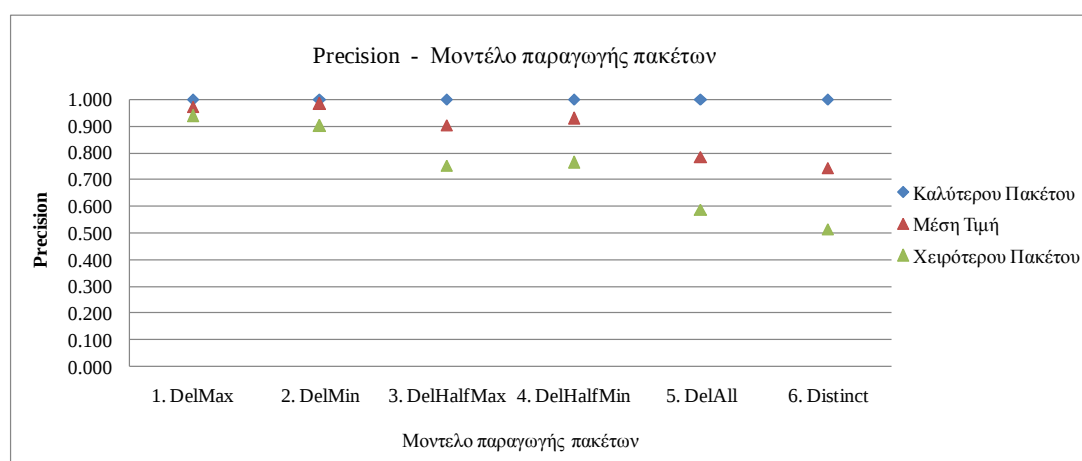


Εικόνα 6.5.6 (α): Οι τιμές του *Rating* για το χειρότερο πακέτο, το καλύτερο πακέτο και για τη μέση τιμή όλων των προτεινόμενων πακέτων, για τα έξι μοντέλα παραγωγής των προτεινόμενων πακέτων.

Στο διάγραμμα της εικόνας 6.5.6(α), που απεικονίζει τις τιμές του *Rating* σε σχέση με το χρησιμοποιούμενο μοντέλο παραγωγής πακέτων, παρατηρούμε ότι τα μοντέλα 1 (*DelMax*) και 2 (*DelMin*) επιτυγχάνουν τις υψηλότερες βαθμολογίες προτίμησης με τα μοντέλα 5 (*DelAll*) και 6 (*Distinct*) να έχουν τις χαμηλότερες. Αυτό είναι αναμενόμενο καθώς στα μοντέλα 1 και 2 έχουμε διαγραφή μίας μόνο ακμής με αποτέλεσμα να αλλάζουν ελάχιστα τα πακέτα και να μειώνεται ελάχιστα ο βαθμός προτίμησης μεταξύ των διαδοχικών προτεινόμενων πακέτων. Από την άλλη, στα μοντέλα 5 και 6 έχουμε μεγάλη αλλαγή στα αντικείμενα των πακέτων με αποτέλεσμα να έχουμε μεγαλύτερη μείωση της βαθμολογίας προτίμησης των διαδοχικών προτεινόμενων πακέτων.

Εδώ αξίζει να γίνει μία σύγκριση μεταξύ των μοντέλων 1 και 2 καθώς και μεταξύ των μοντέλων 3 (*DelHalfMax*) και 4 (*DelHalfMin*). Από το διάγραμμα της

εικόνας 6.5.6(α) παρατηρούμε ότι το μοντέλο 2 επιτυγχάνει υψηλότερες βαθμολογίες προτίμησης από το μοντέλο 1 και αντίστοιχα το μοντέλο 4 επιτυγχάνει υψηλότερες βαθμολογίες προτίμησης από το μοντέλο 3. Αυτό δείχνει ότι διαγράφοντας τη χειρότερη ακμή, με τον μικρότερο βαθμό προτίμησης, ή το 50% των χειρότερων ακμών, με τους μικρότερους βαθμούς προτίμησης, επιτυγχάνονται μεγαλύτεροι βαθμοί προτίμησης για τα σχηματιζόμενα πακέτα σε σχέση με τους βαθμούς προτίμησης που επιτυγχάνονται με τη διαγραφή της καλύτερης ακμής ή του 50% των καλύτερων ακμών, αντίστοιχα. Συνεπώς είναι προτιμότερη η διαγραφή ακμών με τους μικρότερους βαθμούς προτίμησης με σκοπό την παραγωγή πακέτων αντικειμένων με υψηλότερους βαθμούς προτίμησης. Ωστόσο αυτή η διαφορά φαίνεται να είναι πάρα πολύ μικρή, της τάξης του 3%.

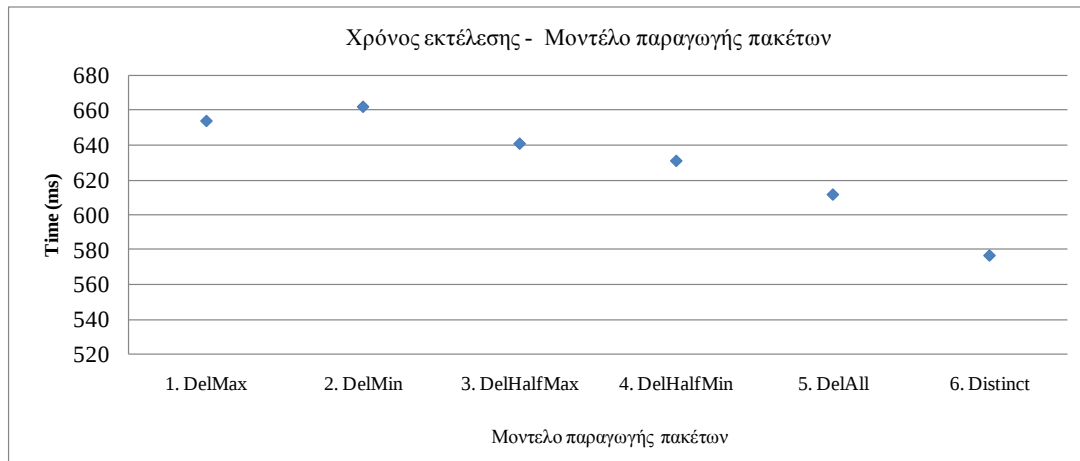


Εικόνα 6.5.6 (β): Ο δείκτης *Precision* για το χειρότερο πακέτο, το καλύτερο πακέτο και για τη μέση τιμή όλων των προτεινόμενων πακέτων σε σχέση με τα έξι μοντέλα παραγωγής των προτεινόμενων πακέτων.

Στο διάγραμμα της εικόνας 6.5.6(β), που απεικονίζει την τιμή του δείκτη *Precision* σε σχέση με το χρησιμοποιούμενο μοντέλο παραγωγής πακέτων, παρατηρούμε ότι τα μοντέλα 1 και 2 επιτυγχάνουν τις υψηλότερες τιμές ακρίβειας με τα μοντέλα 5 και 6 να επιτυγχάνουν τις χαμηλότερες. Αυτό είναι αναμενόμενο καθώς στα μοντέλα 1 και 2 έχουμε διαγραφή μίας μόνο ακμής με αποτέλεσμα να αλλάζουν ελάχιστα τα πακέτα και να μειώνεται ελάχιστα η ακρίβεια μεταξύ των προτεινόμενων πακέτων. Από την άλλη, στα μοντέλα 5 και 6 έχουμε μεγάλη αλλαγή στα αντικείμενα των πακέτων με αποτέλεσμα να έχουμε μεγαλύτερη μείωση του δείκτη *Precision*.

Σε αντιστοιχία με αυτό που συμβαίνει με τις βαθμολογίες προτίμησης, στο διάγραμμα της εικόνας 6.5.6(β) παρατηρούμε ότι το μοντέλο 2 επιτυγχάνει υψηλότερες τιμές του δείκτη *Precision* από το μοντέλο 1 και αντίστοιχα το μοντέλο

4 επιτυγχάνει υψηλότερη ακρίβεια από το μοντέλο 3. Συνεπώς, είναι προτιμότερη η διαγραφή ακμών με τις μικρότερες βαθμολογίες προτίμησης καθώς αυτό οδηγεί στην παραγωγή πακέτων με υψηλότερες τιμές του δείκτη *Precision*. Ωστόσο, αυτή η διαφορά φαίνεται να είναι πάρα πολύ μικρή, της τάξης του 0,02 ή 2% στην κλίμακα του δείκτη *Precision*.



Εικόνα 6.5.6 (γ): Ο χρόνος εκτέλεσης του αλγόριθμου για τα έξι μοντέλα παραγωγής των προτεινόμενων πακέτων.

Στο διάγραμμα 6.5.6(γ) απεικονίζεται η γραφική παράσταση του χρόνου εκτέλεσης του αλγόριθμου σε σχέση με το χρησιμοποιούμενο μοντέλο παραγωγής πακέτων. Ο χρόνος εκτέλεσης βλέπουμε να είναι υψηλότερος για τα δύο πρώτα μοντέλα, να παίρνει ενδιάμεσες τιμές για τα μοντέλα 3 και 4 και να έχει τη μικρότερη τιμή στα μοντέλα 5 και 6. Οι μικρότεροι χρόνοι εκτέλεσης οφείλονται στη μείωση του χώρου αναζήτησης του αλγόριθμου. Φαίνεται να υπάρχει μια σχέση σύνδεσης μεταξύ του χώρου αναζήτησης και του χρόνου εκτέλεσης του αλγόριθμου. Στα μοντέλα 3, 4, 5 και ιδιαίτερα στο 6, που το γράφημα μοντελοποίησης του αλγόριθμου μειώνεται με γρηγορότερους ρυθμούς με την παραγωγή των διαδοχικών πακέτων, παρατηρούμε μειωμένους χρόνους εκτέλεσης. Ωστόσο, ιδιαίτερα για τα 4 πρώτα μοντέλα, δεν φαίνεται να υπάρχει κάποια σημαντική διαφορά στον χρόνο εκτέλεσης. Καθώς η διακύμανση μεταξύ μέγιστου και ελάχιστου χρόνου εκτέλεσης είναι της τάξης των 90ms για την παραγωγή 10 πακέτων με 8 αντικείμενα το καθένα.

6.5.7 Μοντέλο Επιλογής των Δημοφιλέστερων Κατηγοριών

Σε αυτή την ενότητα θα διερευνηθεί η συμπεριφορά του αλγόριθμου για τα τρία μοντέλα προσδιορισμού και επιλογής των δημοφιλέστερων για τους χρήστες κατηγοριών. Τα μοντέλα αυτά έχουν αναλυθεί στο κεφάλαιο 2 και είναι τα εξής:

1. Επιλογή των δημοφιλέστερων κατηγοριών σύμφωνα με τον αριθμό των αντικειμένων που έχει βαθμολογήσει ο χρήστης από κάθε κατηγορία (*ItemsPerCat.*).
2. Επιλογή των δημοφιλέστερων κατηγοριών σύμφωνα με τη συνολική βαθμολογία προτίμησης του χρήστη για τα αντικείμενα που ανήκουν σε κάθε κατηγορία (*RatingPerCat.*).
3. Επιλογή των δημοφιλέστερων κατηγοριών σύμφωνα με τον αριθμό των αντικειμένων κάθε κατηγορίας από το σύνολο των προβλεπόμενων αντικειμένων που προορίζονται για τον σχηματισμό πακέτων (*PredictedItems*).

Στον πίνακα 6.5.7(β) συγκεντρώνονται οι μετρήσεις για τον χρόνο εκτέλεσης του αλγόριθμου, για το *Rating* και τον δείκτη *Precision* του χειρότερου πακέτου, του καλύτερου πακέτου και του μέσου όρου στο σύνολο των πακέτων, σε σχέση με το χρησιμοποιούμενο μοντέλο επιλογής των δημοφιλέστερων κατηγοριών.

Για τη διενέργεια του πειράματος διατηρήθηκαν σταθερές οι τιμές των παραμέτρων που αναγράφονται στους πίνακες 6.5.3(α) και 6.5.7(α) και πραγματοποιήθηκαν μετρήσεις για 10 χρήστες που έχουν βαθμολογήσει περίπου 1000 αντικείμενα ο καθένας. Οι μέσοι όροι των μετρήσεων αυτών με τα 95% διαστήματα εμπιστοσύνης όπου κρίθηκε χρήσιμο συγκεντρώνονται στον πίνακα 6.5.7(β). Επίσης, οι τιμές των μετρήσεων για το *Rating* και το *Precision* απεικονίζονται στα διαγράμματα των εικόνων 6.5.7(α) και 6.5.7(β).

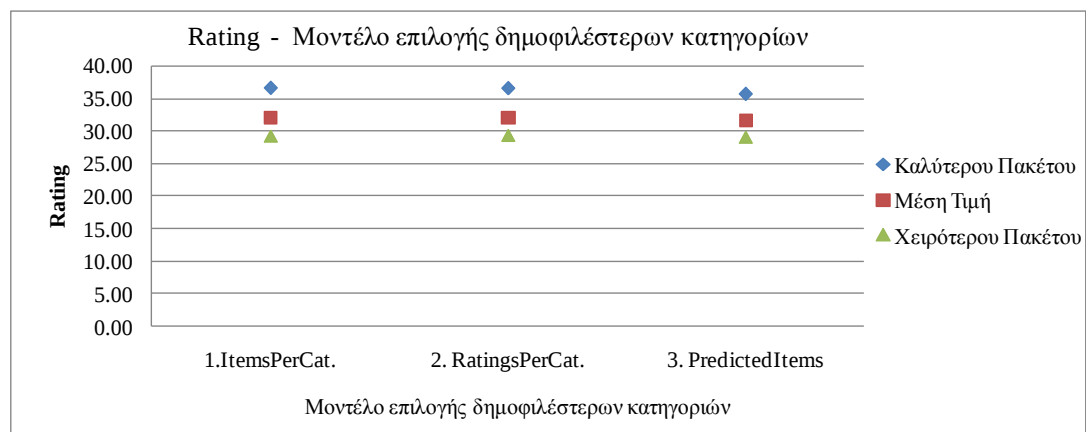
Αριθμός προτεινόμενων πακέτων	Αριθμός Κατηγοριών	Αριθμός αντικειμένων ανά πακέτο	Μοντέλο παραγωγής πακέτων
5	5	8 Αναλογικά κατανεμημένα στις κατηγορίες	6 Παραγωγή διακριτών πακέτων

Πίνακας 6.5.7 (α): Οι παράμετροι που οι τιμές τους διατηρούνται σταθερές στις μετρήσεις αξιολόγησης που αφορούν στα μοντέλα επιλογής των δημοφιλέστερων κατηγοριών.

Μοντέλο επιλογής των δημοφιλέστερων κατηγοριών	Χρόνος εκτέλεσης (ms)	Rating			Precision		
		Χειρότερου πακέτου	Καλύτερου Πακέτου	Μέση Τιμή	Χειρότερου πακέτου	Καλύτερου πακέτου	Μέση τιμή
1. ItemsPerCat.	324	29.21	36.59	32.01±0.48	0.763	1.000	0.835±0.06
2. RatingPerCat.	336	29.33	36.54	32.10±0.55	0.725	1.000	0.838±0.05
3. PredictedItems	387	29.03	35.69	31.61±0.34	0.638	1.000	0.778±0.05

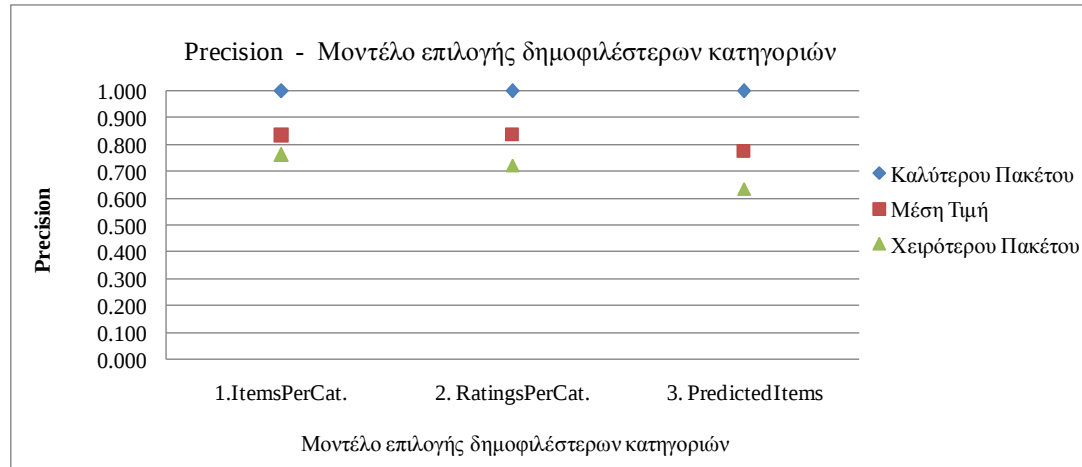
Πίνακας 6.5.7(β): Μετρήσεις για το χρόνο εκτέλεσης του αλγόριθμου, το *Rating* και το *Precision*, για το χειρότερο πακέτο, το καλύτερο πακέτο και για τη μέση τιμή όλων των προτεινόμενων πακέτων με τα 95% διαστήματα εμπιστοσύνης, σε σχέση με τα τρία μοντέλα προσδιορισμού των δημοφιλέστερων κατηγοριών.

Στο διάγραμμα 6.5.7(α) παρατηρούμε ότι η βαθμολογία προτίμησης των προτεινόμενων πακέτων είναι σχεδόν ίδια για τα δύο πρώτα μοντέλα και ελάχιστα μικρότερη για το τρίτο μοντέλο. Αυτό είναι αναμενόμενο καθώς τα δύο πρώτα μοντέλα αναφέρονται στις δημοφιλέστερες για το χρήστη κατηγορίες ενώ το τρίτο μοντέλο αφορά τα προβλεπόμενα αντικείμενα που προτείνονται από το συνεργατικό φιλτράρισμα για τον σχηματισμό των καλύτερων πακέτων.



Εικόνα 6.5.7 (α): Οι τιμές του *Rating* για το χειρότερο πακέτο, το καλύτερο πακέτο και για τη μέση τιμή όλων των προτεινόμενων πακέτων σε σχέση με τα τρία μοντέλα προσδιορισμού και επιλογής των δημοφιλέστερων κατηγοριών.

Αντίστοιχα, στο διάγραμμα 6.5.7(β) παρατηρούμε ότι ο δείκτης *Precision* των προτεινόμενων πακέτων έχει σχεδόν την ίδια τιμή για τα δύο πρώτα μοντέλα και ελάχιστα μικρότερη τιμή για το τρίτο μοντέλο. Αυτό είναι αναμενόμενο καθώς τα δύο πρώτα μοντέλα υπακούν περισσότερο στις πραγματικές προτιμήσεις του χρήστη για την επιλογή των δημοφιλέστερων κατηγοριών.



Εικόνα 6.5.7 (β): Οι τιμές του δείκτη *Precision*, για το καλύτερο πακέτο και για τη μέση τιμή όλων των προτεινόμενων πακέτων, σε σχέση με τα τρία μοντέλα προσδιορισμού και επιλογής των δημοφιλέστερων κατηγοριών.

Από τις μετρήσεις που έγιναν για το *Rating* και τον δείκτη *Precision*, των προτεινόμενων πακέτων σε σχέση με το μοντέλο επιλογής των δημοφιλέστερων κατηγοριών, φαίνεται να μην υπάρχει ουσιαστική διαφορά μεταξύ των δύο πρώτων μοντέλων. Αντίθετα, το τρίτο μοντέλο υστερεί έναντι των δύο πρώτων. Η διαφορά του δείκτη *Precision* μεταξύ του καλύτερου και του χειρότερου μοντέλου είναι της τάξης του 6%. Συνεπώς, είναι προτιμότερο να χρησιμοποιείται ένα από τα δύο πρώτα μοντέλα επιλογής της δημοφιλέστερης κατηγορίας.

6.6 Αποτελέσματα Αξιολόγησης

Στα πλαίσια της αξιολόγησης του αλγόριθμου μέσω του λογισμικού που τον ενσωματώνει, έγιναν οι απαραίτητες μετρήσεις με σκοπό τη διερεύνηση της αποτελεσματικότητας, της ακρίβειας και της αποδοτικότητας που επιτυγχάνεται. Εξετάστηκε η συμπεριφορά του αλγόριθμου σε διάφορες συνθήκες λειτουργίας με τη διερεύνηση των τιμών των παραμέτρων εκείνων που η επιλογή τους οδηγεί στη βελτίωση της αποδοτικότητας και της ακρίβειας του συστήματος.

Για τη διενέργεια μιας επιτυχημένης αξιολόγησης απαραίτητη προϋπόθεση είναι η ίδια η διαδικασία της αξιολόγησης να μην έχει επιπτώσεις στην ακρίβεια των αποτελεσμάτων. Για να επιτευχθεί αυτό, προσδιορίστηκε το ποσοστό απόκρυψης των αντικειμένων και ο ελάχιστος βαθμός προτίμησης των κρυμμένων αντικειμένων με τρόπο που να μην προκαλούν επιπτώσεις στην ακρίβεια του συνεργατικού φιλτραρίσματος. Μέσα από τις μετρήσεις που έγιναν, κατά την προετοιμασία της αξιολόγησης, βρέθηκε ότι το μέσο τετραγωνικό σφάλμα και το

μέσο απόλυτο σφάλμα του συνεργατικού φιλτραρίσματος δεν επηρεάζονται για ποσοστό απόκρυψης των αντικειμένων ίσο με 40% και για ελάχιστη τιμή της βαθμολογίας προτίμησης των κρυμμένων αντικειμένων ίση με 2. Οι τιμές αυτές διατηρήθηκαν σταθερές σε όλες τις μετρήσεις της αξιολόγησης.

Για να λειτουργεί αποδοτικά ο αλγόριθμος απαιτείται ένας όσο το δυνατόν μικρός αριθμός αντικειμένων που προέρχονται από το συνεργατικό φιλτράρισμα και προορίζονται για το σχηματισμό των καλύτερων πακέτων. Σύμφωνα με τις μετρήσεις που έγιναν, για πακέτα που το μέγεθος και το πλήθος τους είναι ανάλογο με αυτό των μετρήσεων, ο αριθμός των απαιτούμενων αντικειμένων προσδιορίστηκε στην τιμή των 500. Με τη χρησιμοποίηση 500 αντικειμένων για τον σχηματισμό πακέτων επιτυγχάνονται βέλτιστοι χρόνοι λειτουργίας χωρίς απώλεια στην ακρίβεια. Στις μετρήσεις χρησιμοποιήθηκε ο αριθμός των 500 αντικειμένων ακόμη και στην περίπτωση που τα προτεινόμενα πακέτα περιελάμβαναν συνολικά 200 αντικείμενα, με τον δείκτη *Precision* να παίρνει υψηλές τιμές, μεγαλύτερες από 75%.

Η θεώρηση του συντελεστή δημοφιλίας των αντικειμένων και η ενσωμάτωση της τιμής που προκύπτει από αυτόν στον βαθμό προτίμησης των αντικειμένων φαίνεται να λειτουργεί θετικά και να οδηγεί σε μεγάλη αύξηση του δείκτη *Precision*. Σύμφωνα με τις μετρήσεις η βέλτιστη τιμή του συντελεστή δημοφιλίας των αντικειμένων προσδιορίστηκε ίση με 5. Με αυτή την τιμή επιτυγχάνεται η βέλτιστη αύξηση του δείκτη *Precision* με την ταυτόχρονη διατήρηση της σύστασης αντικειμένων σύμφωνα με τα χαρακτηριστικά προτιμήσεων του τρέχοντα χρήστη. Σύμφωνα με τις μετρήσεις που έγιναν, είναι χαρακτηριστικό ότι με την αύξηση του συντελεστή δημοφιλίας των αντικειμένων από την τιμή 0 στην τιμή 5 έχουμε μια σημαντική αύξηση του δείκτη *Precision* από το ποσοστό του 30% στο 85%.

Η αύξηση του αριθμού των προτεινόμενων πακέτων, όπως είναι αναμενόμενο, επιφέρει μια αύξηση στον χρόνο εκτέλεσης του αλγόριθμου και μια μείωση στην μέση τιμή του δείκτη *Precision* για τα προτεινόμενα πακέτα αντικειμένων. Η αύξηση στον χρόνο εκτέλεσης του αλγόριθμου είναι γραμμική και ανάλογη με τον αριθμό των προτεινόμενων πακέτων, καθώς αυξάνεται το υπολογιστικό φορτίο του αλγόριθμου. Ο χρόνος εκτέλεσης για την παραγωγή ενός προτεινόμενου πακέτου είναι μικρότερος από 100ms ενώ για τα 10 προτεινόμενα πακέτα δεν ξεπερνά τα 600ms. Ο δείκτης *Precision* για το καλύτερο πακέτο παραμένει στο 100% ανεξάρτητα από τον αριθμό των προτεινόμενων πακέτων, ενώ για το σύνολο των 10 προτεινόμενων πακέτων με 8 αντικείμενα το καθένα, η μέση τιμή του δείκτη *Precision* είναι της τάξης του 70%. Συνεπώς, η τιμή του δείκτη *Precision* διατηρείται σε υψηλά επίπεδα. Αν έπρεπε να δώσουμε ένα νούμερο για τον αριθμό

των προτεινόμενων πακέτων σύμφωνα με τον οποίο θα πετυχαίναμε υψηλή ακρίβεια με μικρούς χρόνους εκτέλεσης, αυτό το νούμερο θα είχε την τιμή 5. Με τη σύσταση 5 πακέτων έχουμε ακρίβεια 70% για το χειρότερο πακέτο και χρόνους εκτέλεσης της τάξης των 300ms.

Η αύξηση των κατηγοριών που περιλαμβάνουν τα προτεινόμενα πακέτα επιφέρει μια αύξηση στο *χρόνο εκτέλεσης* και μια μείωση στον δείκτη *Precision*. Ο ρυθμός αύξησης του χρόνου εκτέλεσης μειώνεται με την αύξηση των κατηγοριών και, σύμφωνα με τις μετρήσεις, στην ακραία περίπτωση των 10 κατηγοριών ο *χρόνος εκτέλεσης* διατηρείται σε χαμηλά επίπεδα, της τάξης των 0.4sec. Αυτό συμβαίνει γιατί οι λιγότερο δημοφιλείς κατηγορίες περιλαμβάνουν μικρότερο αριθμό αντικειμένων και είναι μειωμένο το φορτίο του αλγόριθμου που αφορά σε αυτές τις κατηγορίες. Ο δείκτης *Precision* για το καλύτερο πακέτο παραμένει στο 100% ανεξάρτητα από τον αριθμό των κατηγοριών. Επίσης, με την αύξηση των κατηγοριών στις 10 η μέση τιμή του δείκτη *Precision* δεν είναι μικρότερη από 75%. Μία βέλτιστη τιμή για τον αριθμό των κατηγοριών θα μπορούσε να είναι η τιμή 4 καθώς η περαιτέρω αύξηση των κατηγοριών φαίνεται να προκαλεί μια μεγαλύτερη πτώση της ακρίβειας.

Με την αύξηση του αριθμού των αντικειμένων ανά πακέτο έχουμε μείωση του δείκτη *Precision*, με τη μικρότερη μείωση να παρατηρείται στην περίπτωση της αναλογικής κατανομής των αντικειμένων στις δημοφιλέστερες κατηγορίες. Αυτό είναι λογικό, καθώς το σύστημα κατανέμει τα αντικείμενα στις κατηγορίες σύμφωνα με τις προτιμήσεις του εκάστοτε χρήστη με αποτέλεσμα να βελτιώνεται η ακρίβεια. Από την άλλη πλευρά, στην περίπτωση της αύξησης του αριθμού των αντικειμένων ενός πακέτου με την αύξηση του αριθμού των αντικειμένων ανά κατηγορία έχουμε μεγαλύτερη μείωση του δείκτη *Precision*, καθώς για τις λιγότερο δημοφιλείς για τον χρήστη κατηγορίες επιβάλλεται ένας σχετικά μεγάλος αριθμός από αντικείμενα που οδηγεί σε μείωση της ακρίβειας. Ο *χρόνος εκτέλεσης* του αλγόριθμου σε σχέση με την αύξηση των αντικειμένων του πακέτου φαίνεται να διατηρείται πρακτικά σταθερός και για τις μετρήσεις του πειράματος δεν ξεπερνά τα 350ms με μία διακύμανση της τάξης των 25ms.

Τα μοντέλα παραγωγής των προτεινόμενων πακέτων στοχεύουν στην προσαρμογή της συμπεριφοράς του συστήματος σύμφωνα με τις διάφορες απαιτήσεις των εφαρμογών. Ωστόσο, από τις μετρήσεις που έγιναν μπορούν να εξαχθούν χρήσιμα συμπεράσματα. Φαίνεται ότι για το μοντέλο 2, σύμφωνα με το οποίο διαγράφεται η ελάχιστη ακμή, επιτυγχάνονται μεγαλύτερα ποσοστά στον δείκτη *Precision* με την ταυτόχρονη αύξηση του *rating* των προτεινόμενων πακέτων συγκρινόμενο με το μοντέλο 1 στο οποίο διαγράφεται η μέγιστη ακμή.

Επίσης, κάτι αντίστοιχο παρατηρείται κατά τη σύγκριση των μοντέλων 3 και 4. Είναι προτιμότερη η εφαρμογή του μοντέλου 4, που διαγράφει το 50% των ελάχιστων ακμών, καθώς επιτυγχάνει μεγαλύτερες τιμές τόσο για το δείκτη *Precision* όσο και για το *rating* των προτεινόμενων πακέτων συγκρινόμενο με το μοντέλο 3 το οποίο διαγράφεται το 50% των μέγιστων ακμών. Το μοντέλο 6 που παράγει διακριτά πακέτα, όπως είναι αναμενόμενο, παρουσιάζει τις μικρότερες τιμές τόσο για τον δείκτη *Precision* όσο και για το *rating*. Ο χρόνος εκτέλεσης είναι μεγαλύτερος για τα δύο πρώτα μοντέλα και μικρότερος για τα δύο τελευταία. Αυτό συμβαίνει καθώς το υπολογιστικό φορτίο στην περίπτωση των μοντέλων 1 και 2, που διαγράφεται μόνο μια ακμή, μειώνεται ελάχιστα μεταξύ διαδοχικών παραγόμενων πακέτων. Από την άλλη, ιδιαίτερα για το μοντέλο 6, κάθε φορά που παράγεται ένα νέο πακέτο αντικειμένων μειώνεται αρκετά ο χώρος αναζήτησης του αλγόριθμου και αυτό συμβαίνει γιατί μειώνεται το γράφημα μοντελοποίησης του προβλήματος με την διαγραφή των αντικειμένων που έχουν χρησιμοποιηθεί. Η διακύμανση του χρόνου εκτέλεσης μεταξύ των μοντέλων παραγωγής πακέτων, σύμφωνα με τις μετρήσεις που έγιναν δεν ξεπερνά τα 100ms.

Για την επιλογή των δημοφιλέστερων για τον χρήστη κατηγοριών προέκυψε ότι τα καταλληλότερα μοντέλα είναι το 1 ή το 2. Αυτά τα δύο μοντέλα χρησιμοποιούν τα δεδομένα προτίμησης του τρέχοντα χρήστη ενώ το μοντέλο 3 χρησιμοποιεί τα εκτιμώμενα δεδομένα. Για το μοντέλο 1 ο προσδιορισμός των δημοφιλέστερων για τον χρήστη κατηγοριών γίνεται σύμφωνα με τα αντικείμενα που έχει βαθμολογήσει ο χρήστης από κάθε κατηγορία, ενώ για το μοντέλο 2 σύμφωνα με το συνολικό *rating* των αντικειμένων κάθε κατηγορίας. Τα δύο αυτά μοντέλα φαίνεται να είναι πρακτικά ισοδύναμα, αν εξαιρέσουμε μια πολύ μικρή υπεροχή του μοντέλου 2. Το τρίτο μοντέλο, που χρησιμοποιεί τα εκτιμώμενα δεδομένα, φαίνεται να υστερεί έναντι των δύο πρώτων, ωστόσο η διαφορά είναι μικρή, της τάξης του 6%. Στην περίπτωση των μοντέλων επιλογής των δημοφιλέστερων κατηγοριών, οι διακυμάνσεις στο χρόνο εκτέλεσης δεν φαίνεται να μας απασχολούν ιδιαίτερα γιατί προέκυψαν πολύ μικρές και δεν δικαιολογούν κάποια ουσιαστική διαφορά, ωστόσο το μοντέλο 1 φαίνεται να ανταποκρίνεται ελάχιστα πιο γρήγορα από τα υπόλοιπα μοντέλα.

Στον πίνακα 6.6 συνοψίζονται οι βέλτιστες τιμές ή οι βέλτιστες επιλογές των παραμέτρων για την αποδοτική και την αποτελεσματική λειτουργία του συστήματος συστάσεων σύμφωνα με όσα αναφέρονται παραπάνω.

A/A	Παράμετρος	Τιμή
1	Αριθμός αντικειμένων για τον σχηματισμό πακέτων	500
2	Συντελεστής δημοφιλίας αντικειμένων	5
3	Αριθμός προτεινόμενων πακέτων	5
4	Αριθμός κατηγοριών	4
5	Τρόπος κατανομής του αριθμού των αντικειμένων ανά κατηγορία	Αναλογική κατανομή στις δημοφιλέστερες κατηγορίες
6	Μοντέλο παραγωγής πακέτων	Τα DelMin και DelHalfMin υπερέχουν έναντι των DelMax και DelHalfMax.
7	Μοντέλο επιλογής των δημοφιλέστερων κατηγοριών.	ItemsPerCat. ή RatingsPerCat.

Πίνακας 6.6: Οι βέλτιστες τιμές των παραμέτρων σύμφωνα με τις μετρήσεις αξιολόγησης.

Οι μετρήσεις που έγιναν αφορούν σε μια συγκεκριμένη εφαρμογή με τη χρήση ενός συγκεκριμένου συνόλου δεδομένων, ωστόσο τα γενικά συμπεράσματα για τη συμπεριφορά του συστήματος σε σχέση με τη μεταβολή των παραμέτρων μπορούν να γενικευτούν. Αυτό που δεν μπορεί να γενικευτεί, και δεν είναι δυνατό να υποστηριχτεί ότι ισχύει σε κάθε πρόβλημα συστάσεων, είναι οι βέλτιστες τιμές των παραμέτρων που προσδιορίστηκαν στο συγκεκριμένο πρόβλημα. Οι συγκεκριμένες τιμές του πίνακα 6.6 μπορούν να διαφοροποιηθούν σε διάφορα άλλα προβλήματα συστάσεων. Για παράδειγμα ο αριθμός των αντικειμένων για τον σχηματισμό των πακέτων σε ένα άλλο πρόβλημα που τα πακέτα περιλαμβάνουν πολύ περισσότερα αντικείμενα σίγουρα θα πρέπει να πάρει μεγαλύτερη τιμή. Το ίδιο ισχύει και για τον βέλτιστο αριθμό των προτεινόμενων πακέτων, τον συντελεστή δημοφιλίας και τον αριθμό των κατηγοριών καθώς σε κάποιο άλλο σύνολο δεδομένων ή σε κάποιο άλλο πρόβλημα μπορεί οι τιμές των παραμέτρων αυτών να διαφοροποιηθούν. Αντίθετα για τον τρόπο κατανομής των αντικειμένων στις κατηγορίες, το μοντέλο παραγωγής των πακέτων και το μοντέλο επιλογής των δημοφιλέστερων κατηγοριών μπορεί να υποστηριχθεί ότι τα συμπεράσματα που εξήχθησαν παραπάνω και συνοψίζονται στον πίνακα 6.6 ισχύουν στα περισσότερα προβλήματα συστάσεων. Συνεπώς, οι μετρήσεις αξιολόγησης μας βοήθησαν περισσότερο στην κατανόηση της συμπεριφοράς του συστήματος παρά στον προσδιορισμό συγκεκριμένων τιμών για τις παραμέτρους. Αυτό σημαίνει ότι η παρούσα προσέγγιση θα πρέπει κάθε φορά να προσαρμόζεται σε κάθε διαφορετικό πρόβλημα συστάσεων που καλείται να αντιμετωπίσει.

Για τη διερεύνηση της αποτελεσματικότητας και της αποδοτικότητας του συστήματος θα ήταν ωφέλιμη μια σύγκριση με άλλες σχετικές προσεγγίσεις. Ωστόσο αυτό δεν κατέστη εφικτό καθώς οι σχετικές εργασίες που μελετήθηκαν και

αναφέρονται στην ενότητα 2.4 δεν θα μπορούσαν να συγκριθούν με την παρούσα εργασία. Αυτό συμβαίνει γιατί οι σχετικές εργασίες είτε αναφέρονται σε μοντέλα που περιλαμβάνουν περιορισμούς μεταξύ των αντικειμένων είτε αναφέρονται σε εξειδικευμένες εφαρμογές που δεν ταιριάζουν με τη φιλοσοφία της παρούσας προσέγγισης. Επίσης, οι περισσότερες αξιολογήσεις των σχετικών εργασιών βασίζονται μόνο στον χρόνο εκτέλεσης των αλγόριθμων και όχι στην ακρίβεια που επιτυγχάνεται. Η παρούσα ιδέα δεν έχει σκοπό να βελτιώσει ή να επεκτείνει κάποιο ήδη ανεπτυγμένο μοντέλο για να μπορέσει να συγκριθεί μαζί του αλλά μπορεί να θεωρηθεί ως μια καινοτομία που ενδεχομένως συμπληρώνει κάποια κενά στον συγκεκριμένο ερευνητικό χώρο. Ωστόσο, είναι σίγουρο ότι, χωρίς να είναι εφικτή κάποια σύγκριση με άλλα σχετικά μοντέλα, μπορούν να εξαχθούν μόνο θετικά συμπεράσματα από τις μετρήσεις αξιολόγησης που πραγματοποιήθηκαν.

Σύμφωνα με τις μετρήσεις που έγιναν προέκυψε ότι ο αλγόριθμος που σχεδιάστηκε λειτουργεί με αποδοτικό και αποτελεσματικό τρόπο και επιτυγχάνει υψηλά ποσοστά ακρίβειας. Επίσης, το εργαλείο λογισμικού, που ενσωματώνει όλη την απαιτούμενη λειτουργικότητα του αλγόριθμου, εκπληρώνει τον σκοπό για τον οποίο σχεδιάστηκε επιτρέποντας τον έλεγχο και την αποδοτική λειτουργία.

Κεφάλαιο 7: Συμπεράσματα και Προοπτικές

7.1 Συμπεράσματα

Στα πλαίσια της παρούσας εργασίας σχεδιάστηκε ένας αλγόριθμος συστάσεων για την πρόταση πακέτων αντικειμένων σύμφωνα με τις προτιμήσεις των χρηστών. Για την εφαρμογή και την αξιολόγηση του αλγόριθμου αναπτύχθηκε ένα ολοκληρωμένο σύστημα συστάσεων που περιλαμβάνει ένα εύχρηστο και γραφικό εποπτικό περιβάλλον χρήστη και ενσωματώνει όλη την απαιτούμενη λειτουργικότητα. Η εφαρμογή και οι μετρήσεις αξιολόγησης έδειξαν ότι ο αλγόριθμος εκπληρώνει τον σκοπό του καθώς είναι αποδοτικός και αποτελεσματικός.

Η παρούσα προσέγγιση είναι γενική και δεν αναφέρεται σε κάποια συγκεκριμένη εφαρμογή αλλά περιλαμβάνει μια σειρά από παραμέτρους σύμφωνα με τις οποίες μπορεί να προσαρμοστεί σε διάφορα προβλήματα συστάσεων. Όλες οι λεπτομέρειες για τη συμπεριφορά του συστήματος σε σχέση με αυτές τις παραμέτρους περιγράφονται στο κεφάλαιο 6 που αφορά στην αξιολόγηση. Ακολουθούν κάποια χρήσιμα συμπεράσματα που προέκυψαν από την εφαρμογή και την αξιολόγηση τα οποία συνοψίζουν και συμπληρώνουν τα αποτελέσματα των μετρήσεων αξιολόγησης του κεφαλαίου 6. Να σημειωθεί ότι τα συμπεράσματα αυτά γενικεύονται για την εφαρμογή της ιδέας σε κάθε σύνολο δεδομένων ή σε κάθε πρόβλημα συστάσεων που υπακούει στην παρούσα προσέγγιση.

Η μείωση του αριθμού των αντικειμένων που εξέρχονται από το συνεργατικό φιλτράρισμα και χρησιμοποιούνται για τον σχηματισμό των καλύτερων πακέτων οδηγεί στη μείωση του χρόνου εκτέλεσης που είναι το ζητούμενο για την αποδοτική λειτουργία του συστήματος. Συνεπώς, η κατάλληλη επιλογή αυτού του αριθμού, ο οποίος μπορεί να επιλεγεί σύμφωνα με το εκάστοτε πρόβλημα συστάσεων, οδηγεί σε αύξηση της αποδοτικότητας του συστήματος με τη μείωση του χρόνου εκτέλεσης.

Ένα κρίσιμο σημείο που φαίνεται η υπεροχή της προσέγγισης είναι η σωστή κατάταξη των προτεινόμενων πακέτων (*package ranking*) σε φθίνουσα ταξινόμηση σύμφωνα με τον βαθμό προτίμησής τους. Οι μετρήσεις αξιολόγησης έδειξαν ότι η κατάταξη αυτή γίνεται με αποτελεσματικό τρόπο καθώς τα πακέτα που βρίσκονται υψηλότερα στη λίστα προτάσεων επιτυγχάνουν υψηλότερη ακρίβεια.

Ένας παράγοντας που βοήθησε πολύ στην αύξηση της ακρίβειας των προτάσεων είναι η θεώρηση του *συντελεστή δημοφιλίας* των αντικειμένων. Σύμφωνα με αυτόν τον συντελεστή ένας παράγοντας που βασίζεται στη *δημοφιλία* του κάθε αντικειμένου προστίθεται στον εκτιμώμενο βαθμό προτίμησης. Ο παράγοντας αυτός, όπως αποδείχθηκε, βοηθάει να ξεχωρίσουν τα δημοφιλέστερα αντικείμενα και οδηγεί σε υψηλά ποσοστά ακρίβειας.

Ο αριθμός των κατηγοριών των προτεινόμενων αντικειμένων καθορίζεται σύμφωνα με το εκάστοτε πρόβλημα συστάσεων, ωστόσο μπορεί να υποστηριχθεί ότι είναι εφικτό να οριστεί κάποιος αριθμός κατηγοριών για τον οποίο επιτυγχάνεται η βέλτιστη ακρίβεια με τον ελάχιστο χρόνο εκτέλεσης. Αυτή η απόφαση καθορίζει την αποδοτικότητα του αλγόριθμου καθώς τα πακέτα που περιλαμβάνουν μικρό αριθμό κατηγοριών επιτυγχάνουν μεγαλύτερα ποσοστά ακρίβειας και μικρότερους χρόνους εκτέλεσης από εκείνα που περιλαμβάνουν περισσότερες κατηγορίες.

Ο αριθμός των αντικειμένων ανά πακέτο καθορίζεται και αυτός από το εκάστοτε πρόβλημα συστάσεων αλλά πρέπει να γνωρίζουμε ότι με την αύξηση του αριθμού των αντικειμένων έχουμε μείωση της ακρίβειας, ενώ ο χρόνος εκτέλεσης διατηρείται σχεδόν σταθερός. Επισημαίνεται ότι στην περίπτωση που ο αριθμός των αντικειμένων κάθε προτεινόμενου πακέτου είναι μεγαλύτερος από τον αριθμό των κατηγοριών που το πακέτο περιλαμβάνει είναι προτιμότερη η αναλογική κατανομή των αντικειμένων στις δημοφιλέστερες κατηγορίες, με την προϋπόθεση βέβαια ότι η εφαρμογή το επιτρέπει.

Να σημειωθεί ότι υπάρχει η δυνατότητα επιλογής μεταξύ έξι διαφορετικών μοντέλων παραγωγής προτεινόμενων πακέτων. Αυτή η δυνατότητα οδηγεί σε μεγαλύτερη ευελιξία καθώς ανάλογα με την εφαρμογή μπορεί να χρησιμοποιηθεί το κατάλληλο μοντέλο παραγωγής πακέτων.

Για την επιλογή των δημοφιλέστερων για τον χρήστη κατηγοριών έχει δημιουργηθεί ένας μηχανισμός ο οποίος βασίζεται σε τρεις διαφορετικές προσεγγίσεις. Οι δύο πρώτες λαμβάνουν υπόψη τους τα αντικείμενα που ο χρήστης έχει βαθμολογήσει ενώ η τρίτη βασίζεται στα νέα αντικείμενα που προέρχονται από το συνεργατικό φιλτράρισμα και έχει γίνει η εκτίμηση για τη βαθμολογία προτίμησης τους. Σύμφωνα με την αξιολόγηση, τα καταλληλότερα μοντέλα επιλογής των δημοφιλέστερων κατηγοριών είναι τα δύο πρώτα για τα οποία δεν προκύπτει κάποια ουσιαστική διαφορά μεταξύ τους.

Η προτεινόμενη προσέγγιση μπορεί να θεωρηθεί ότι αποτελεί μια καινοτομία καθώς συνδυάζει την υψηλή ακρίβεια του συνεργατικού φιλτραρίσματος με τη βέλτιστη σύνθεση των εξαγόμενων πακέτων μέσω του αλγόριθμου ροής ελάχιστου κόστους. Η ιδέα περιλαμβάνει και επιμέρους παράγοντες που οδηγούν στην

επιτυχημένη σύσταση αντικειμένων όπως είναι ο συντελεστής δημοφιλίας και όλες οι υπόλοιπες παράμετροι οι οποίες βοηθούν στην προσαρμογή του συστήματος στο εκάστοτε πρόβλημα συστάσεων. Η παρούσα προσέγγιση δε βελτιώνει ούτε εξελίσσει κάποιο ήδη ανεπτυγμένο μοντέλο αλλά μπορεί να θεωρηθεί ως μια καινοτομία που ενδεχομένως συμπληρώνει κάποιο κενό στον συγκεκριμένο ερευνητικό χώρο. Βέβαια, υπάρχει χώρος για εξέλιξη και βελτίωση καθώς η εργασία αυτή αποτελεί μια πρώτη προσπάθεια ανάπτυξης ενός τέτοιου συστήματος που σίγουρα έχει περιθώρια για περαιτέρω επεκτάσεις, μερικές από τις οποίες αναφέρονται στην επόμενη ενότητα.

Η ολοκλήρωση της παρούσας εργασίας δικαιώνει τους κόπους και τις προσδοκίες καθώς το μοντέλο που προτείνεται φαίνεται πως εκπληρώνει τον σκοπό του και μπορεί να υιοθετηθεί από πρακτικές εφαρμογών επιτυγχάνοντας το ζητούμενο αποτέλεσμα.

7.2 Προοπτικές

Η παρούσα προσέγγιση αποτελεί μια δομημένη και καλά καθορισμένη ιδέα που όπως αποδείχθηκε εκπληρώνει με αρκετά αποτελεσματικό τρόπο τον σκοπό της, ωστόσο θα ήταν χρήσιμο να αναφερθούν κάποιες επεκτάσεις ή προοπτικές που ανοίγονται και αφορούν στην εξέλιξη και στη βελτίωσή της.

Ο αλγόριθμος που αναπτύχθηκε είναι γενικός, ωστόσο θα μπορούσε να αποτελέσει μια βάση για να αναπτυχθούν πιο σύνθετα και εξειδικευμένα μοντέλα που απευθύνονται σε κάποια συγκεκριμένη εφαρμογή. Αυτό σημαίνει ότι η παρούσα προσέγγιση μπορεί να εξελιχθεί ή να αποτελέσει την αφετηρία για την ανάπτυξη συστημάτων συστάσεων με υψηλότερες απαιτήσεις και εξειδικευμένες προδιαγραφές. Ένα παράδειγμα στο οποίο θα μπορούσε να προσαρμοστεί και να χρησιμοποιηθεί ο αλγόριθμος αυτός είναι η ανάπτυξη ενός συστήματος συστάσεων που προτείνει σύνολα γευμάτων για μια ημερήσια ή εβδομαδιαία διαίτα.

Μία δεύτερη επέκταση που θα ήταν πολύ χρήσιμη είναι η κατάλληλη τροποποίηση του αλγόριθμου με σκοπό να δίνει προτεραιότητα στα νέα αντικείμενα. Αυτό σημαίνει ότι θα είχαν μεγαλύτερη βαρύτητα τα νέα αντικείμενα με αποτέλεσμα να προτείνονται με μεγαλύτερη συχνότητα. Μια τέτοια επέκταση απαιτεί τη γνωστοποίηση της ημερομηνίας ή της χρονολογίας που ένα αντικείμενο δημιουργείται ή διατίθεται. Αυτό θα ήταν χρήσιμο στις περιπτώσεις συστημάτων που απευθύνονται σε εφαρμογές σχετικές με το εμπόριο αλλά και σε γενικότερες εφαρμογές καθώς ένα νέο αντικείμενο όπως μια ταινία, ένα βιβλίο ή ένα φαγητό, είναι πάντα ελκυστικότερο.

Μια ακόμη επέκταση που θα διεύρυνε το χώρο χρήσης του αλγόριθμου είναι η εισαγωγή περιορισμών. Η παρούσα προσέγγιση δε λαμβάνει υπόψη της τους τυχόν περιορισμούς μεταξύ των αντικειμένων και αυτό την κάνει να περιορίζεται σε έναν μικρότερο αριθμό εφαρμογών. Μια τέτοια επέκταση απαιτεί από το σύνολο δεδομένων να περιλαμβάνει τις πληροφορίες για τους περιορισμούς μεταξύ των αντικειμένων που μπορεί να αφορούν σε προαπαιτούμενα αντικείμενα ή σε μεταξύ τους αποκλεισμούς. Αυτή η επέκταση θα ήταν χρήσιμη σε εφαρμογές όπως εκείνες που προτείνουν ταξιδιωτικά πακέτα στα οποία τα σημεία ενδιαφέροντος πρέπει να βρίσκονται σε κοντινή απόσταση μεταξύ τους ή σε εφαρμογές που προτείνουν πακέτα μαθημάτων που μπορεί να έχουν περιορισμούς προαπαιτούμενων μαθημάτων ή περιορισμούς σχετικούς με το ακολουθούμενο πρόγραμμα σπουδών. Είναι φανερό ότι η ενσωμάτωση περιορισμών θα πρέπει να γίνει με τρόπο που να είναι σύμφωνη με κάποια συγκεκριμένη εφαρμογή ή με μια ομάδα εφαρμογών.

Τέλος, να αναφερθεί ότι οποιαδήποτε προσέγγιση που έχει σχέση με την πληροφορική, λόγω των τεχνολογικών εξελίξεων και των αλλαγών στις ανάγκες χρήσης, για να συνεχίσει να υφίσταται, χρειάζεται μία συνεχή προσπάθεια εξέλιξης και επέκτασης. Συνεπώς και η παρούσα προσέγγιση θα πρέπει να συνεχίσει να αναπτύσσεται, να εξελίσσεται και να προσαρμόζεται στα νέα δεδομένα για να επιτυγχάνει το σκοπό της.

Αναφορές

- Adomavicius G., & Tuzhilin A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734-749. doi:10.1109/TKDE.2005.99.
- Angel A., Chaudhuri S., Das G., & Koudas N. (2009). Ranking objects based on relationships and fixed associations. Paper presented at the *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology, EDBT'09*, 910-921. doi:10.1145/1516360.1516464.
- Bertsekas D. P. (1998). *Network optimization: Continuous and discrete models*. Massachusetts, Belmont, MA: Athena Scientific.
- Boutilier C. (2002, July). A POMDP formulation of preference elicitation problems. In *AAAI/IAAI* (pp. 239-246).
- Chajewska U., Koller D., & Parr R. (2000, July). Making rational decisions using adaptive utility elicitation. In *AAAI/IAAI* (pp. 363-369).
- Cormen T., Leiserson C, Rivest R., & Stein C. (2001). *Introduction to algorithms. 2nd edition*, The Massachusetts Institute of Technology - MIT Press.
- CourseRank. (n.d.) Retrieved October 20, 2015, from <http://studentapps.stanford.edu/courserank.html>
- Dasgupta S., Papadimitriou C. H., & Vazirani U. V. (2008). *Algorithms*. Boston: McGraw-Hill Higher Education.
- Deng T., Fan W., & Geerts F. (2013). On the complexity of package recommendation problems. *SIAM Journal on Computing*, 42(5), 1940-1986.
- Ekstrand M. D., Riedl J. T. & Konstan J. A. (2010). Collaborative Filtering Recommender Systems. *Foundations and Trends in Human-Computer Interaction*, 4(2), 81-173. doi: 10.1561/1100000009.
- Finger J., & Polyzotis N. (2009). Robust and efficient algorithms for rank join evaluation. Paper presented at the *SIGMOD-PODS'09 - Proceedings of the International Conference on Management of Data and 28th Symposium on Principles of Database Systems*, 415-428. doi:10.1145/1559845.1559890.

- GroupLens Research datasets. (n.d.). Retrieved December 29, 2015, from <http://grouplens.org/datasets/movielens/>
- Ilyas I. F., Aref W. G., & Elmagarmid A. K. (2004). Supporting top-k join queries in relational databases. *The VLDB Journal—The International Journal on Very Large Data Bases*, 13(3), 207-221.
- Interdonato R., Romeo S., Tagarelli A., & Karypis G. (2013, November). A Versatile Graph-Based Approach to Package Recommendation. In *Tools with Artificial Intelligence (ICTAI), 2013 IEEE 25th International Conference on* (pp. 857-864). IEEE.
- Kantor P. B., Rokach L., Ricci F., & Shapira B. (2011). *Recommender systems handbook*. Springer.
- Lappas T., Liu K., & Terzi E. (2009, June). Finding a team of experts in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 467-476). ACM.
- Melville P. & Sindhvani V. (2010). Recommender Systems. In Sammut C. & Webb G. I. (Eds). *Encyclopedia of Machine Learning* (pp. 829-838). New York: Springer.
- MovieLens. (n.d.). Retrieved December 29, 2015, from <https://movielens.org/>
- Olson, D. L., & Delen, D. (2008). *Advanced data mining techniques*. Berlin: Springer.
- Parameswaran A. G., & Garcia-Molina H. (2009, October). Recommendations with prerequisites. In *Proceedings of the third ACM conference on Recommender systems* (pp. 353-356). ACM.
- Parameswaran, A. G., Garcia-Molina, H., & Ullman, J. D. (2010, October). Evaluating, combining and generalizing recommendations with prerequisites. In *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 919-928). ACM.
- Parameswaran, A., Venetis, P., & Garcia-Molina, H. (2011). Recommendation systems with complex constraints: A course recommendation perspective. *ACM Transactions on Information Systems (TOIS)*, 29(4), 20.
- Rich, E. (1979). User modeling via stereotypes*. *Cognitive science*, 3(4), 329-354.
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001, April). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web* (pp. 285-295). ACM.

- Schafer J.B., Frankowski D., Herlocker J., & Sen S. (2007). Collaborative filtering recommender systems. In Brusilovsky P., Kobsa A., & Nejdl W. (Eds.). *The Adaptive Web*. (pp. 291–324). Berlin: Springer.
- Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*. Amsterdam: Morgan Kaufman.
- Xie M., Lakshmanan L.V., & Wood P.T. (2011). Efficient rank join with aggregation constraints. *Proceedings of the VLDB Endowment*, 4(11), 1200-1212.
- Xie M., Lakshmanan L.V.S., & Wood P.T. (2010). Breaking out of the box of recommendations: From items to packages. Paper presented at the *RecSys'10 - Proceedings of the 4th ACM Conference on Recommender Systems*, 151-158. doi:10.1145/1864708.1864739.
- Xie, M. (2015). Composite recommendation: semantics and efficiency.
- Xie, M., Lakshmanan, L. V. S., & Wood, P. T. (2013). IPS: An interactive package configuration system for trip planning. *Proceedings of the VLDB Endowment*, 6(12), 1362-1365.
- Xie, M., Lakshmanan, L. V. S., & Wood, P. T. (2014). Generating top-k packages via preference elicitation. Paper presented at the *Proceedings of the VLDB Endowment*, 7(14) 1941-1952.
- Xie, M., Lakshmanan, L. V., & Wood, P. T. (2011, April). Comprec-trip: A composite recommendation system for travel planning. In *Data Engineering (ICDE), 2011 IEEE 27th International Conference on* (pp. 1352-1355). IEEE.
- Κοκολάκης, Γ. & Σπηλιώτης, Ι. (2010). *Θεωρία πιθανοτήτων και στατιστική με εφαρμογές*. Αθήνα: Συμεών.

Παράρτημα

A. Τεκμηρίωση του Πηγαίου Κώδικα

Στο παράρτημα αυτό περιλαμβάνονται οι λεπτομέρειες για τον πηγαίο κώδικα της εφαρμογής λογισμικού που υλοποιεί το σύστημα συστάσεων που αναπτύχθηκε. Για την ανάπτυξη του λογισμικού χρησιμοποιήθηκε η γλώσσα προγραμματισμού *java* και οι κλάσεις της *java* ομαδοποιούνται στα εξής πακέτα λογισμικού:

- *entityBasicClasses*: Το πακέτο περιλαμβάνει τις κλάσεις των βασικών οντοτήτων.
- *entityGraphClasses*: Το πακέτο περιλαμβάνει τις κλάσεις οντοτήτων που υλοποιούν το γράφημα μοντελοποίησης του προβλήματος.
- *fileManagerClasses*: Το πακέτο περιλαμβάνει τις κλάσεις διαχείρισης των αρχείων του συνόλου δεδομένων.
- *collaborativeFiltering*: Το πακέτο περιλαμβάνει τις κλάσεις για το συνεργατικό φιλτράρισμα.
- *linearProgrammingClasses*: Το πακέτο περιλαμβάνει τις κλάσεις για την επίλυση του προβλήματος γραμμικού προγραμματισμού.
- *formsAndFunctionality*: Το πακέτο περιλαμβάνει τις κλάσεις που υλοποιούν το γραφικό περιβάλλον και την απαραίτητη λειτουργικότητα που σχετίζεται με αυτό.

Στις υποενότητες που ακολουθούν παρουσιάζονται οι κλάσεις της *java* σύμφωνα με το πακέτο λογισμικού στο οποίο ανήκουν και περιγράφεται η λειτουργικότητά τους, τα πεδία και οι μέθοδοι που περιλαμβάνουν.

A.1 Κλάσεις Βασικών Οντοτήτων

Σε αυτή την ενότητα παρουσιάζονται οι κλάσεις της *java* που υλοποιούν τις βασικές οντότητες που εμπλέκονται στη μοντελοποίηση του αλγόριθμου.

Κλάση: *Item*

(public class *Item*)

Περιγραφή: Η κλάση υλοποιεί την οντότητα για τα αντικείμενα (*items*) του προβλήματος συστάσεων.

Πεδία: `int itemId`, `String title`, `String[] categories`, `int numOfCategories`, `int numOfMaxAppearances`, `int popularity`

Κατασκευαστής: Item(int itemId, String title, String[] categories, int numOfCategories, int numOfMaxAppearances, int popularity)

Μέθοδοι: Οι getters και οι setters για τα πεδία της κλάσης αυτής.

Κλάση: ItemForPackage

(public class ItemForPackage)

Περιγραφή: Η κλάση υλοποιεί την οντότητα για τα αντικείμενα (*items*) που περιλαμβάνονται στα πακέτα συστάσεων.

Πεδία: int itemId, int maxAppearances, String category, String title, double rating

Κατασκευαστής: ItemForPackage(int itemId, int maxAppearances, String category, String title, double rating)

Μέθοδοι: Οι getters και οι setters για τα πεδία της κλάσης.

Κλάση: ItemRating

(public class ItemRating)

Περιγραφή: Η κλάση υλοποιεί την οντότητα για τα αντικείμενα (*items*) και τις βαθμολογίες προτίμησής (*ratings*) τους.

Πεδία: int movieID, double rating

Κατασκευαστής: ItemRating(int movieID, double rating)

Μέθοδοι: Οι getters και οι setters για τα πεδία της κλάσης.

Κλάση: ItemRatingPopularity

(public class ItemRatingPopularity)

Περιγραφή: Η κλάση υλοποιεί την οντότητα για τα αντικείμενα (*items*), τις βαθμολογίες προτίμησης (*ratings*) και τον αριθμό της δημοφιλίας (*popularity*) τους.

Πεδία: int movieID, double rating, int popularity

Κατασκευαστής: ItemRatingPopularity(int movieID, double rating, int popularity)

Μέθοδοι: Οι getters και οι setters για τα πεδία της κλάσης.

Κλάση: RatingsPerUser

(public class RatingsPerUser)

Περιγραφή: Η κλάση υλοποιεί την οντότητα για τις βαθμολογίες προτίμησης (*ratings*) ανά χρήστη.

Πεδία: int userID, int numOfRatings, List<ItemRating> movieRatingList

Κατασκευαστής: RatingsPerUser(int userID, int numOfRatings, List<ItemRating> movieRatingList)

Μέθοδοι: Οι getters και οι setters για τα πεδία της κλάσης.

Κλάση: Measures

(public class Measures)

Περιγραφή: Η κλάση υλοποιεί την οντότητα που περιλαμβάνει τις μετρήσεις αξιολόγησης του αλγόριθμου.

Πεδία: int[] TP_perPackage, int[] FN_perPackage, int[] FP_perPackage, int TP_overall, int FN_overall, int FP_overall, double[] precision_perPackage, double[] recall_perPackage, double precision_overall, double recall_overall

Κατασκευαστής: Measures(int[] TP_perPackage, int[] FN_perPackage, int[] FP_perPackage, int TP_overall, int FN_overall, int FP_overall, double[] precision_perPackage, double[] recall_perPackage, double precision_overall, double recall_overall)

Μέθοδοι: Οι getters και οι setters για τα πεδία της κλάσης.

Κλάση: CategoryPopularityOverallRating

(public class CategoryPopularityOverallRating)

Περιγραφή: Η κλάση χρησιμοποιείται για τον χαρακτηρισμό της δημοφιλίας της κάθε κατηγορίας αντικειμένων. Τα αντικείμενα της κλάσης αυτής περιλαμβάνουν πληροφορίες για την κατηγορία αντικειμένων (*category*), τον αριθμό των αντικειμένων που ανήκουν σε αυτή την κατηγορία (*popularity*) και τη συνολική βαθμολογία προτίμησης των αντικειμένων της κατηγορίας (*overallRating*).

Πεδία: String category, int popularity, double overallRating

Κατασκευαστής: CategoryPopularityOverallRating(String category, int popularity, double overallRating)

Μέθοδοι: Οι getters και οι setters για τα πεδία της κλάσης αυτής.

A.2 Κλάσεις Οντοτήτων για το Σχηματισμό του Γραφήματος

Σε αυτή την ενότητα παρουσιάζονται οι κλάσεις της java που σχετίζονται με τις οντότητες που αφορούν στο γράφημα μοντελοποίησης του προβλήματος των συστάσεων.

Κλάση: EdgeCategorySink

(public class EdgeCategorySink)

Περιγραφή: Η κλάση χρησιμοποιείται για τη δημιουργία των οντοτήτων που αντιστοιχούν στις ακμές του γραφήματος μεταξύ των κατηγοριών (*categories*) και του τερματικού κόμβου (*sink*).

Πεδία: String categoryID, double cost, int maxFlow, int minFlow

Κατασκευαστής: EdgeCategorySink(String categoryID, double cost, int maxFlow, int minFlow)

Μέθοδοι: Οι getters και οι setters που σχετίζονται με τα πεδία της κλάσης.

Κλάση: EdgeItemCategory

(public class EdgeItemCategory)

Περιγραφή: Η κλάση χρησιμοποιείται για τη δημιουργία των οντοτήτων που αντιστοιχούν στις ακμές του γραφήματος μεταξύ των αντικειμένων (*items*) και των κατηγοριών (*categories*).

Πεδία: int itemId, String categoryID, double rating, double cost, int maxFlow, int minFlow

Κατασκευαστής: EdgeItemCategory(int itemId, String categoryID, double rating, double cost, int maxFlow, int minFlow)

Μέθοδοι: Οι getters και οι setters που σχετίζονται με τα πεδία της κλάσης.

Κλάση: EdgeItemSource

(public class EdgeItemSource)

Περιγραφή: Η κλάση χρησιμοποιείται για τη δημιουργία των οντοτήτων που αντιστοιχούν στις ακμές του γραφήματος μεταξύ των αντικειμένων (*items*) και του αρχικού κόμβου (*source*).

Πεδία: int itemId, double cost, int maxFlow, int minFlow

Κατασκευαστής: EdgeItemSource(int itemId, double cost, int maxFlow, int minFlow)

Μέθοδοι: Οι getters και οι setters που σχετίζονται με τα πεδία της κλάσης.

Κλάση: EdgeSinkSource

(public class EdgeSinkSource)

Περιγραφή: Η κλάση χρησιμοποιείται για τη δημιουργία των οντοτήτων που αντιστοιχούν στις ακμές του γραφήματος μεταξύ του τερματικού (*sink*) και του αρχικού (*source*) κόμβου.

Πεδία: double cost, int maxFlow, int minFlow

Κατασκευαστής: EdgeSinkSource(double cost, int maxFlow, int minFlow)

Μέθοδοι: οι getters και οι setters που σχετίζονται με τα πεδία της κλάσης.

Κλάση: VertexCategory

(public class VertexCategory)

Περιγραφή: Η κλάση χρησιμοποιείται για τη δημιουργία των οντοτήτων που αντιστοιχούν στους κόμβους των κατηγοριών του γραφήματος.

Πεδία: String categoryID, List<EdgeItemCategory> edgeMovieCategory_list, EdgeCategorySink edgeCategorySink

Κατασκευαστής: VertexCategory(String categoryID, List<EdgeItemCategory> edgeMovieCategory_list, EdgeCategorySink edgeCategorySink)

Μέθοδοι: Οι getters και οι setters που σχετίζονται με τα πεδία της κλάσης.

Κλάση: VertexItem

(public class VertexItem)

Περιγραφή: Η κλάση χρησιμοποιείται για τη δημιουργία των οντοτήτων που αντιστοιχούν στους κόμβους των αντικειμένων του γραφήματος.

Πεδία: int itemId, int numAppearances,
List<EdgeItemCategory> edgeMovieCategory_list,
EdgeItemSource edgeMovieSource

Κατασκευαστής: VertexItem(int itemId, int numAppearances,
List<EdgeItemCategory> edgeMovieCategory_list,
EdgeItemSource edgeMovieSource)

Μέθοδοι: Οι getters και οι setters που σχετίζονται με τα πεδία της κλάσης.

Κλάση: VertexSink

(public class VertexSink)

Περιγραφή: Η κλάση χρησιμοποιείται για τη δημιουργία της οντότητας που αντιστοιχεί στον τερματικό κόμβο (sink) του γραφήματος.

Πεδία: EdgeSinkSource edgeSinkSource,
List<EdgeCategorySink> edgeCategorySink_list

Κατασκευαστής: VertexSink(EdgeSinkSource edgeSinkSource,
List<EdgeCategorySink> edgeCategorySink_list)

Μέθοδοι: Οι getters και οι setters που σχετίζονται με τα πεδία της κλάσης.

Κλάση: VertexSource

(public class VertexSource)

Περιγραφή: Η κλάση χρησιμοποιείται για τη δημιουργία της οντότητας που αντιστοιχεί στον αρχικό κόμβο (source) του γραφήματος.

Πεδία: EdgeSinkSource edgeSinkSource,
List<EdgeItemSource> edgeItemSource_list

Κατασκευαστής: VertexSource(EdgeSinkSource edgeSinkSource,
List<EdgeItemSource> edgeItemSource_list)

Μέθοδοι: οι getters και οι setters που σχετίζονται με τα πεδία της κλάσης.

A.3 Κλάσεις Διαχείρισης των Αρχείων του Συνόλου Δεδομένων

Σε αυτή την ενότητα παρουσιάζονται οι κλάσεις της java που διαχειρίζονται τα αρχεία του συνόλου δεδομένων.

Κλάση: ReadFiles

(public class ReadFiles)

Περιγραφή: Η κλάση χρησιμοποιείται για το διάβασμα των αρχείων του συνόλου δεδομένων και τη διατήρηση της πληροφορίας σε κατάλληλες δομές δεδομένων.

Πεδία: Hashtable<Integer, RatingsPerUser> ratingsPerUser_hashTable,
ArrayList<RatingsPerUser> ratingsPerUser_list,
Hashtable<Integer,Item> movies_hashTable,
List<Item> movies_list, double maxValueOfRating

Κατασκευαστής: ReadFiles()

Μέθοδοι: οι getters και οι setters που σχετίζονται με τα πεδία της κλάσης και οι μέθοδοι που περιγράφονται στον πίνακα A.3(α).

Τύπος Μεθόδου	Μέθοδος	Περιγραφή
public void	readRatingsFile(String path)	Διάβασμα του αρχείου των ratings και διατήρηση των δεδομένων σε κατάλληλες δομές.
public void	readItemsFile(String path)	Διάβασμα του αρχείου των αντικειμένων και διατήρηση των δεδομένων σε κατάλληλες δομές.
public boolean	isThisStringInteger(String str)	Εξέταση για το εάν το String str μπορεί να γίνει ακέραιος αριθμός.
public int	convertStringToInt(String str)	Μετατροπή ενός String σε Integer.
public void	sortItemsList(int mode)	Ταξινόμηση της λίστας των αντικειμένων βάσει του itemId ή του popularity σύμφωνα με την τιμή του mode.
public void	sortRatingsPerUserList(int mode)	Ταξινόμηση της λίστας των ratings βάσει του userId ή του αριθμού των αντικειμένων ανά χρήστη σύμφωνα με την τιμή του mode.

Πίνακας A.3 (α): Οι μέθοδοι της κλάσης ReadFiles .

Κλάση: ConvertFiles

(public class ConvertFiles)

Περιγραφή: Η κλάση χρησιμοποιείται για τη μετατροπή των αρχείων του συνόλου δεδομένων σε κατάλληλη μορφή για την περαιτέρω χρησιμοποίησή τους με τη μετατροπή των δύο άνω κάτω τελειών (::) σε κόμμα (,).

Πεδία: -

Κατασκευαστής: ConvertFiles()

Μέθοδοι: Περιγράφονται στον πίνακα Α.3(β).

Τύπος Μεθόδου	Μέθοδος	Περιγραφή
public String	converRatignsFile(String inputFileName)	Μετατροπή του αρχείου των ratings σε κατάλληλο τύπο αρχείου και επιστροφή του path για το νέο παραγόμενο αρχείο.
public String	converItemsFile(String inputFileName)	Μετατροπή του αρχείου των αντικειμένων στον κατάλληλο τύπο αρχείου και επιστροφή του path για το νέο παραγόμενο αρχείο.

Πίνακας Α.3 (β): Οι μέθοδοι της κλάσης *ConvertFiles***Α.4 Κλάση Συνεργατικού Φιλτραρίσματος**

Σε αυτή την ενότητα παρουσιάζεται η κλάση της java που εκτελεί το συνεργατικό φιλτράρισμα.

Κλάση: CollaborativeFiltering

(public class CollaborativeFiltering)

Περιγραφή: Η κλάση χρησιμοποιείται για την εκτέλεση του συνεργατικού φιλτραρίσματος.

Πεδία: -

Κατασκευαστής: CollaborativeFiltering()

Μέθοδοι: Περιγράφονται στον πίνακα Α.4.

Τύπος Μεθόδου	Μέθοδος	Περιγραφή
public List<ItemRating>	itemBasedColFiltering(int userId, int numOfRecommendations, String dataset)	Εκτέλεση του συνεργατικού φιλτραρίσματος βασισμένο στα αντικείμενα.
public List<ItemRating>	userBasedColFiltering(int userId, int numOfRecommendations, String dataset)	Εκτέλεση του συνεργατικού φιλτραρίσματος βασισμένο στους χρήστες.

Πίνακας Α.4: Οι μέθοδοι της κλάσης *CollaborativeFiltering*.

A.5 Κλάσεις Επίλυσης του Προβλήματος Γραμμικού Προγραμματισμού

Σε αυτή την ενότητα παρουσιάζονται οι κλάσεις της java που εκτελούν τις λειτουργίες για την επίλυση του προβλήματος γραμμικού προγραμματισμού.

Κλάση: SolutionType

(public class SolutionType)

Περιγραφή: Η κλάση δημιουργεί την οντότητα η οποία περιλαμβάνει τον τύπο της λύσης που επιστρέφεται μετά την επίλυση του προβλήματος γραμμικού προγραμματισμού.

Πεδία: String[] namesOfEdges, double[] flowOfEdges, double[] costOfEdges

Κατασκευαστής: SolutionType(String[] namesOfEdges, double[] flowOfEdges, double[] costOfEdges)

Μέθοδοι: Οι getters και οι setters για τα πεδία της κλάσης.

Κλάση: LinearProgramming

(public class LinearProgramming)

Περιγραφή: Η κλάση χρησιμοποιείται για τη μοντελοποίηση και επίλυση του προβλήματος γραμμικού προγραμματισμού.

Πεδία: -

Κατασκευαστής: LinearProgramming()

Μέθοδοι: Περιγράφονται στον πίνακα A.5.

Τύπος Μεθόδου	Μέθοδος	Περιγραφή
public SolutionType	linearProgrammingCalc()	Μοντελοποίηση, επίλυση και επιστροφή της λύσης για το πρόβλημα γραμμικού προγραμματισμού.
public int	indexOfStrInTable(String[] str_table, String str, int size_of_table)	Επιστροφή της θέσης του str μέσα στον πίνακα str_table ή -1 σε περίπτωση που δεν υπάρχει το str στον πίνακα.

Πίνακας A.5: Οι μέθοδοι της κλάσης *LinearProgramming*.

A.6 Κλάσεις Γραφικού Περιβάλλοντος

Σε αυτή την ενότητα παρουσιάζονται οι κλάσεις της java που υλοποιούν το γραφικό περιβάλλον της εφαρμογής και όλη την απαραίτητη λειτουργικότητα που εκτελείται από αυτό. Για τις κλάσεις αυτές αναφέρονται μόνο τα πεδία και οι μέθοδοι που υλοποιούν τη βασική λειτουργικότητα των κλάσεων ενώ παραλείπονται τα παιδιά και οι μέθοδοι που αφορούν στην υλοποίηση και λειτουργία του γραφικού περιβάλλοντος.

Κλάση: StartForm

(public class StartForm extends JFrame)

Περιγραφή: Η κλάση δημιουργεί το αρχικό παράθυρο της εφαρμογής και ενσωματώνει όλη την απαραίτητη λειτουργικότητα που εκτελείται από αυτό το παράθυρο.

Πεδία: int selectedUserID, String ratingsDataFile, String itemsDataFile, int minItemsPrint, int maxItemsPrint, int minRatingPrint, int maxRatingPrint

Κατασκευαστής: public StartForm()

Μέθοδοι: Περιγράφονται στον πίνακα Α.6(α)

Τύπος Μεθόδου	Μέθοδος	Περιγραφή
public void	loadFilesButtonAction()	Φόρτωση των αρχείων των δεδομένων μετά το πάτημα του σχετικού κουμπιού.
public void	initialStaticValues()	Αρχικοποίηση των τιμών των μεταβλητών για την εκτύπωση των δεδομένων.
public void	selectUserButtonAction()	Εκτέλεση των εργασιών για την επιλογή του τρέχοντα χρήστη μετά το πάτημα του σχετικού κουμπιού.
public String	chooseFile()	Επιλογή αρχείου.
public void	printRatings()	Εκτύπωση στην οθόνη των εγγραφών για τα ratings ανά 100.
public void	printAllRatings()	Εκτύπωση στην οθόνη όλων των εγγραφών για τα ratings.
public void	printItems()	Εκτύπωση στην οθόνη των εγγραφών για τα αντικείμενα ανά 100.
public void	printAllItems()	Εκτύπωση στην οθόνη όλων των εγγραφών για τα αντικείμενα.

Πίνακας Α.6 (α): Οι μέθοδοι της κλάσης *StartForm*.

Κλάση: RunningModeForm

(public class RunningModeForm extends JFrame)

Περιγραφή: Η κλάση υλοποιεί το δεύτερο κατά σειρά παράθυρο για την επιλογή του τρόπου λειτουργίας της εφαρμογής και ενσωματώνει όλη την απαραίτητη λειτουργικότητα που εκτελείται από αυτό.

Πεδία: int currentUserId, int runningMode, int runningMode_text, double popularityFactor, int itemsOfCFoutput, int itemsOfCFoutput_selectedIndex, int itemsForPackages, int itemsForPackages_selectedIndex, double minRatingOfHiddenItems, List<ItemRating> predictedItemRatings_list, List<ItemRating> predictedItemRatingsWithPopularity_list, List<ItemRating> concealedItemRatings_list

Κατασκευαστής: public RunningModeForm()

Μέθοδοι: Περιγράφονται στον πίνακα Α.6(β).

Τύπος Μεθόδου	Μέθοδος	Περιγραφή
public void	windowsOpeningAction()	Εκτέλεση των απαραίτητων ενεργειών κατά το άνοιγμα του παραθύρου.
public void	precisionAndRecallOfRecommendedItems()	Υπολογισμός και εκτύπωση στην οθόνη των δεικτών precision και recall των εκτιμώμενων αντικειμένων.
public void	concealingItemsAndRunningRecommender(double probabilityOfConcealedItems)	Απόκρυψη αντικειμένων και εκτέλεση του συνεργατικού φιλτραρίσματος.
public void	nextButtonAction()	Υλοποίηση της λειτουργικότητας του κουμπιού Next.
public void	applyRunningMode()	Εφαρμογή του επιλεγμένου μοντέλου λειτουργίας.
public void	setItemsForPackagesWithPopularity()	Δημιουργία των αντικειμένων για τα πακέτα λαμβάνοντας υπόψη το popularity στον υπολογισμό των ratings.
public void	setNumOfRecommendations()	Ορισμός του αριθμού των αντικειμένων που θα χρησιμοποιηθούν για τον σχηματισμό των προτεινόμενων πακέτων.
public void	printWaitingMessage()	Εκτύπωση μηνύματος σχετικά με την αναμονή για το συνεργατικό φιλτράρισμα.
public void	printRatings(int sortingMode)	Εκτύπωση στην οθόνη των αντικειμένων με τις βαθμολογίες προτίμησης.
public void	printActualRatings(int sortingMode)	Εκτύπωση στην οθόνη των προτιμήσεων του χρήστη.
public void	printPredictedRatings(int sortingMode)	Εκτύπωση στην οθόνη των εκτιμήσεων του συνεργατικού φιλτραρίσματος.
public void	printHiddenRatings(int sortingMode)	Εκτύπωση στην οθόνη των κρυμμένων αντικειμένων.

Πίνακας Α.6 (β): Οι μέθοδοι της κλάσης *RunningModeForm*.

Κλάση: ProblemModelingForm

(public class ProblemModelingForm extends JFrame)

Περιγραφή: Η κλάση υλοποιεί το τρίτο κατά σειρά παράθυρο για την επιλογή των παραμέτρων λειτουργίας του αλγόριθμου και ενσωματώνει όλη την απαραίτητη λειτουργικότητα που εκτελείται από αυτό.

Πεδία: VertexSink vertexSink, VertexSource vertexSource,

List<VertexItem> vertexItem_list,

List<VertexItem> vertexItemOfTopCategories_list,

List<VertexItem> vertexItemConcealing_list,

List<VertexCategory> vertexCategory_list,
 List<VertexCategory> vertexTopCategory_list
 List<VertexCategory> vertexCategoryConcealing_list,
 List<CategoryPopularityOverallRating> categoriesPopularity_list,
 int numOfPackages, int numOfTopCategories, int packageSize,
 int numOfItemsPerCategory, int numOfMaxAppearancesPerItem,
 int categoryPopularity_mode, String categoryPopularity_mode_str,
 Integer[] itemsAllocationToCategories, int modelOfCreatingPackages,
Κατασκευαστής: public ProblemModelingForm()
Μέθοδοι: Περιγράφονται στον πίνακα Α.6(γ).

Τύπος Μεθόδου	Μέθοδος	Περιγραφή
public void	windowsOpeningAction()	Εκτέλεση των απαραίτητων ενεργειών κατά το άνοιγμα του παραθύρου.
public void	loadGraph()	Φόρτωση του γραφήματος μοντελοποίησης του προβλήματος.
public void	loadGraphOfTopCategories()	Φόρτωση του γραφήματος με τις δημοφιλέστερες κατηγορίες.
public void	createGraph()	Δημιουργία του γραφήματος.
public void	loadItemsOfTopCategories()	Φόρτωση των κόμβων των αντικειμένων για τις δημοφιλέστερες κατηγορίες.
public void	loadSourceSinkNodesOfTopCategories()	Φόρτωση του τερματικού κόμβου για τις δημοφιλέστερες κατηγορίες.
public void	loadTopCategories(int numOfCategories)	Φόρτωση των δημοφιλέστερων για τον χρήστη κατηγοριών σύμφωνα με την παράμετρο της μεθόδου.
public void	existenceOfCategory(VertexItem vertexItem)	Έλεγχος για την ύπαρξη της κατηγορίας στον κόμβο του αντικειμένου vertexItem.
public void	printCategoryNodes(List<VertexCategory> vertexCategory_list)	Εκτύπωση στην οθόνη των κόμβων του γραφήματος για τις κατηγορίες.
public void	printItemNodes(List<VertexItem> vertexItem_list)	Εκτύπωση στην οθόνη των κόμβων του γραφήματος για τα αντικείμενα.
public void	printSinkNode()	Εκτύπωση στην οθόνη του τερματικού κόμβου του γραφήματος.
public void	printSourceNode()	Εκτύπωση στην οθόνη του αρχικού κόμβου του γραφήματος.
public void	round(double number, int decimals)	Στρογγυλοποίηση ενός δεκαδικού αριθμού σε συγκεκριμένο αριθμό δεκαδικών ψηφίων.

Πίνακας Α.6 (γ): Οι μέθοδοι της κλάσης *ProblemModelingForm*.

Κλάση: ItemsAllocationToCategoriesForm

(public class ItemsAllocationToCategoriesForm extends JFrame)

Περιγραφή: Η κλάση υλοποιεί το παράθυρο για την αναλογική κατανομή των αντικειμένων στις κατηγορίες και ενσωματώνει όλη την απαραίτητη λειτουργικότητα που εκτελείται από αυτό.

Πεδία: int numOfItemsPerPackage

Κατασκευαστής: public ItemsAllocationToCategoriesForm()

Μέθοδοι: Περιγράφονται στον πίνακα A.6(δ).

Τύπος Μεθόδου	Μέθοδος	Περιγραφή
public Integer[]	itemsAllocationToCategories()	Κατανομή των αντικειμένων ενός πακέτου στις κατηγορίες σύμφωνα με τον βαθμό προτίμησης κάθε κατηγορίας.
public void	round(double number, int decimals)	Στρογγυλοποίηση ενός δεκαδικού αριθμού σε συγκεκριμένο αριθμό δεκαδικών ψηφίων.

Πίνακας A.6 (δ): Οι μέθοδοι της κλάσης *ItemsAllocationToCategoriesForm*.**Κλάση:** ProblemSolutionNPackagesForm

(public class ProblemSolutionNPackagesForm extends JFrame)

Περιγραφή: Η κλάση υλοποιεί το παράθυρο που παρουσιάζει τα προτεινόμενα πακέτα αντικειμένων και ενσωματώνει όλη την απαιτούμενη λειτουργικότητα.

Πεδία: SolutionType solutionOfProblem, long overallRunningTime,

Hashtable<Integer, Hashtable<String, List<ItemForPackage>>>

packages_hashTable

Κατασκευαστής: public ProblemSolutionNPackagesForm()

Μέθοδοι: Περιγράφονται στον πίνακα A.6(ε).

Τύπος Μεθόδου	Μέθοδος	Περιγραφή
public void	windowsOpeningAction()	Εκτέλεση των απαραίτητων ενεργειών κατά το άνοιγμα του παραθύρου.
public void	createPackages()	Δημιουργία των καλύτερων πακέτων αντικειμένων.
public void	runMinCostFlow(int ackageNumber, int numOfCategories)	Εκτέλεση του minimum cost flow αλγόριθμου για την εξαγωγή του καλύτερου πακέτου.
public void	removeOneItemNodeOfGraph(int movieID)	Διαγραφή ενός κόμβου του γραφήματος για συγκεκριμένο αντικείμενο.

public void	removeOneEdgeOfGraph(int movieID, String category)	Διαγραφή μίας ακμής του γραφήματος με τα χαρακτηριστικά των παραμέτρων της μεθόδου.
public void	removeEdgesOrNodes(int modelOfPackage, List<ItemForPackage> moviesForRemovingEdges_list)	Διαγραφή ακμών ή κόμβων από το γράφημα σύμφωνα με το μοντέλο παραγωγής των πακέτων.
public void	printSolution()	Εκτύπωση στην οθόνη της λύσης με τα N-καλύτερα πακέτα αντικειμένων
public void	printStatisticsAndSolution(int packageNumber, int numOfCategories, long runningTime)	Εκτύπωση στην οθόνη των στατιστικών και των τιμών των μεταβλητών μετά τη λύση του προβλήματος γραμμικού προγραμματισμού.
public void	printResultsToFile()	Εκτύπωση σε αρχείο της λύσης με τα N-καλύτερα πακέτα αντικειμένων και άνοιγμα του αρχείου.

Πίνακας Α.6 (ε): Οι μέθοδοι της κλάσης *ProblemSolutionNPackagesForm*.**Κλάση:** EvaluationMeasuresForm

(public class EvaluationMeasuresForm extends JFrame)

Περιγραφή: Η κλάση υλοποιεί το παράθυρο που παρουσιάζει τα αποτελέσματα για την αξιολόγηση του συστήματος και ενσωματώνει όλη την απαιτούμενη λειτουργικότητα.

Πεδία: -

Κατασκευαστής: public EvaluationMeasuresForm()

Μέθοδοι: Περιγράφονται στον πίνακα Α.6(στ).

Τύπος Μεθόδου	Μέθοδος	Περιγραφή
public void	windowsOpeningAction()	Εκτέλεση των απαραίτητων ενεργειών κατά το άνοιγμα του παραθύρου.
public Measures	computeMeasures(int itemsPerPackage, int numberOfPackages, int[][] ItemsPerPackages, Integer[] ItemsOfAllPackages)	Υπολογισμός των δεικτών αξιολόγησης του συστήματος
public void	computeAndPrintMeasures()	Υπολογισμός των δεικτών αξιολόγησης του συστήματος και εκτύπωσης των αποτελεσμάτων στην οθόνη
public void	printMeasures(Measures measures, int numOfPackages)	Εκτύπωση στην οθόνη των αποτελεσμάτων αξιολόγησης.
public Integer[]	removeDuplicates(Integer[] array)	Διαγραφή των διπλότυπων του πίνακα array και επιστροφή του νέου πίνακα

Πίνακας Α.6 (στ): Οι μέθοδοι της κλάσης *EvaluationMeasuresForm*.

