



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΣΤΗ ΔΙΟΙΚΗΣΗ ΚΑΙ
ΟΙΚΟΝΟΜΙΚΗ ΤΩΝ ΤΗΛΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΔΙΚΤΥΩΝ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Χρήση Κοινωνικών Δικτύων ως Βάση ενός Συστήματος
Συστάσεων**

**Βανέσα – Αγγελική Χ. Λιανέρη
Ουρανία Σ. Στάμου**

**Επιβλέποντες: Ευστάθιος Χατζηευθυμιάδης, Αναπληρωτής Καθηγητής ΕΚΠΑ
Κωνσταντίνος Κολομβάτσος, Υποψήφιος Διδάκτωρ ΕΚΠΑ**

ΑΘΗΝΑ

Ιούνιος 2013

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Χρήση Κοινωνικών Δικτύων ως Βάση ενός Συστήματος Συστάσεων

Βανέσα – Αγγελική Χ. Λιανέρη

A.M.: ΜΟΠ304

Ουρανία Σ. Στάμου

A.M.: ΜΟΠ303

ΕΠΙΒΛΕΠΟΝΤΕΣ: Ευστάθιος Χατζηευθυμιάδης, Αναπληρωτής Καθηγητής ΕΚΠΑ
Κωνσταντίνος Κολομβάτσος, Υποψήφιος Διδάκτωρ ΕΚΠΑ

Ιούνιος 2013

ΠΕΡΙΛΗΨΗ

Σκοπός της παρούσας διπλωματικής εργασίας είναι η δημιουργία ενός ολοκληρωμένου συστήματος συστάσεων που βασίζεται σε πληροφορίες από κοινωνικά δίκτυα. Αρχικά, αναλύονται σε θεωρητικό επίπεδο τα είδη των συστημάτων συστάσεων μαζί με τα πλεονεκτήματα και τα μειονεκτήματά τους. Παράλληλα αναφέρονται τα βασικά χαρακτηριστικά των κοινωνικών δικτύων και πώς αυτά θα μπορούσαν να χρησιμοποιηθούν προκειμένου να βελτιωθούν οι συστάσεις προς τους χρήστες. Εν συνεχεία, περιγράφεται η αρχιτεκτονική ενός προτεινόμενου συστήματος συστάσεων και ο αλγόριθμος που εκτελεί για να παράγει αποτελέσματα. Το σύστημα απαρτίζεται από μια διεπαφή χρήστη που επιτελεί απλές λειτουργίες εισαγωγής δεδομένων και εμφάνισης αποτελεσμάτων σύστασης, μια βάση δεδομένων με πληροφορίες για τους χρήστες και τα προϊόντα και μια μηχανή συστάσεων. Ο προτεινόμενος αλγόριθμος σύστασης χρησιμοποιεί τεχνικές συνεργατικού φιλτραρίσματος μαζί με τεχνικές ομοιότητας ενώ ταυτόχρονα υιοθετεί έννοιες από τη θεωρία γράφων. Τα δεδομένα για τους χρήστες αντλούνται από το Facebook μέσω του Facebook Graph API το οποίο αναλύεται διεξοδικά. Η απόδοση του αλγορίθμου ελέγχεται με τη χρήση ευρέως γνωστών μετρικών ακρίβειας πρόβλεψης ύστερα από πραγματοποίηση πειραμάτων πάνω σε πραγματικά δεδομένα χρηστών. Η εργασία ολοκληρώνεται με την εξαγωγή γενικών συμπερασμάτων και προτάσεων για μελλοντικές προεκτάσεις.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Συστήματα Συστάσεων

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: Συνεργατικό Φιλτράρισμα, Συντελεστές συσχέτισης, Δέντρο επικάλυψης ελαχίστου κόστους, Facebook Graph API, Μετρικές απόδοσης

ABSTRACT

This thesis aims to the development of a complete recommendation system based on information retrieved by social networks. At the beginning, a survey on the available recommender systems is presented together with their advantages and disadvantages. The main features of the social networks are also described as well as how information retrieved by these networks could be used to improve recommendations. Consequently, both the architecture of the proposed recommendation system and the algorithm that is executed to produce ratings prediction are described. The system consists of: a) a user interface which provides simple functionality for data input and displays recommendation results, b) a database with information about users and products and c) a recommender engine. The proposed recommendation algorithm uses collaborative filtering techniques in combination with techniques that calculate similarity between users while in the same time it includes concepts from graph theory. In addition, the proposed system extracts data for registered users from Facebook through the Facebook Graph API which is described in detail. The performance of the algorithm is tested by using widely known prediction metrics after implementing experiments on a real dataset. Finally, we present our conclusions and suggestions for future work.

SUBJECT AREA: Recommender Systems

KEYWORDS: Collaborative filtering, Correlation coefficients, Minimum cost spanning tree, Facebook Graph API, Evaluation metrics

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΡΟΛΟΓΟΣ	10
1. ΕΙΣΑΓΩΓΗ	11
2. ΣΥΣΤΗΜΑΤΑ ΣΥΣΤΑΣΕΩΝ	13
2.1 Διαδικασία σύστασης	13
2.2 Μορφές της διαδικασίας σύστασης	15
2.3 Προσεγγίσεις σύστασης	16
2.3.1 Μη εξατομικευμένη προσέγγιση (non-personalized).....	17
2.3.2 Φιλτράρισμα βασισμένο στο περιεχόμενο (content-based filtering - CB).....	17
2.3.3 Συνεργατικό ή κοινωνικό φιλτράρισμα (collaborative filtering – CF or social filtering).....	18
2.3.4 Συνεργατικό φιλτράρισμα με βάση τη μνήμη (memory-based collaborative filtering).....	20
2.3.5 Συνεργατικό φιλτράρισμα με βάση το μοντέλο (model-based collaborative filtering).....	21
2.3.6 Φιλτράρισμα βασισμένο στη γνώση (knowledge-based filtering - KB).....	22
2.3.7 Φιλτράρισμα βασισμένο στα δημογραφικά στοιχεία (Demography-based filtering - DF).....	23
2.3.8 Υβριδικό Φιλτράρισμα (hybrid filtering).....	23
2.4 Δημιουργία συστάσεων με χρήση του Κοινωνικού Ιστού	24
2.4.1 Βασικές οντότητες του Κοινωνικού Ιστού	26
3. ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΣΥΣΤΗΜΑΤΟΣ	28
3.1 Διεπαφή σύστασης (Recommender Interface Component - RIC)	29
3.2 Βάση Δεδομένων (DataBase - DB)	38
3.3 Μηχανή συστάσεων (Recommender Engine - RE)	42
4. ΜΗΧΑΝΙΣΜΟΣ ΣΥΣΤΑΣΕΩΝ	43
4.1 Η χρήση του Facebook	43
4.1.1 Facebook Graph API.....	44
4.1.2 Πρόσβαση δεδομένων.....	45
4.1.3 Κλήσεις στο Graph API (Graph API calls).....	47
4.1.4 Υλοποίηση του Graph API σε Java	56
4.1.5 Αξιολογήσεις χρηστών	59

4.2	Περιγραφή του αλγορίθμου	60
4.2.1	Ομοιότητα χρηστών	61
4.2.2	Τελικός υπολογισμός συστάσεων.....	63
5.	ΑΞΙΟΛΟΓΗΣΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ.....	66
5.1	Σύνολο δεδομένων	66
5.2	Μετρικές απόδοσης.....	67
5.3	Μοντέλο αξιολόγησης	68
5.4	Πειραματικά αποτελέσματα	70
6.	ΣΥΜΠΕΡΑΣΜΑΤΑ	81
	ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ	83
	ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ	85
	ΑΝΑΦΟΡΕΣ.....	86

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Σχήμα 1: Λειτουργία του συστήματος	28
Σχήμα 2: Διάγραμμα οντοτήτων-συσχετίσεων	39
Σχήμα 3: Διάγραμμα μετρικών απόδοσης για ίσες τιμές παραμέτρων.....	71
Σχήμα 4: Διάγραμμα μετρικών απόδοσης για $w_1=0.4$, $w_2=0.3$ και $w_3=0.3$	72
Σχήμα 5: Διάγραμμα μετρικών απόδοσης για $w_1=0.3$, $w_2=0.4$ και $w_3=0.3$	73
Σχήμα 6: Διάγραμμα μετρικών απόδοσης για $w_1=0.3$, $w_2=0.3$ και $w_3=0.4$	74
Σχήμα 7: Διάγραμμα μετρικών απόδοσης για $w_1=0.5$, $w_2=0.3$ και $w_3=0.2$	75
Σχήμα 8: Διάγραμμα μετρικών απόδοσης για $w_1=0.5$, $w_2=0.2$ και $w_3=0.3$	76
Σχήμα 9: Διάγραμμα μετρικών απόδοσης για $w_1=0.3$, $w_2=0.5$ και $w_3=0.2$	77
Σχήμα 10: Διάγραμμα μετρικών απόδοσης για $w_1=0.2$, $w_2=0.5$ και $w_3=0.3$	78
Σχήμα 11: Διάγραμμα μετρικών απόδοσης για $w_1=0.3$, $w_2=0.2$ και $w_3=0.5$	79
Σχήμα 12: Διάγραμμα μετρικών απόδοσης για $w_1=0.2$, $w_2=0.3$ και $w_3=0.5$	80

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1: Αρχική σελίδα εισόδου	29
Εικόνα 2: Φόρμα καταχώρησης στοιχείων.....	30
Εικόνα 3: Επιλογή κατηγορίας προϊόντος.....	30
Εικόνα 4: Καταχώρηση προτιμήσεων για βιβλία.....	31
Εικόνα 5: Καταχώρηση προτιμήσεων για μουσική	32
Εικόνα 6: Καταχώρηση προτιμήσεων για ταινίες	33
Εικόνα 7: Σελίδα συνδεδεμένου χρήστη	34
Εικόνα 8: Εμφάνιση αποτελεσμάτων	35
Εικόνα 9: Εμφάνιση προσωπικών στοιχείων	36
Εικόνα 10: Επεξεργασία προσωπικών στοιχείων	36
Εικόνα 11: Ιστορικό συστάσεων	37
Εικόνα 12: Εγχειρίδιο βοήθειας	38
Εικόνα 13: Δημιουργία εφαρμογής	57
Εικόνα 14: Κωδικοί εφαρμογής.....	57
Εικόνα 15: Η σελίδα αξιολογήσεων	60
Εικόνα 16: Δημιουργία δοκιμαστικών χρηστών	67

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1: Πεδία του αντικειμένου User	50
Πίνακας 2: Συνδέσεις του αντικειμένου User	51
Πίνακας 3: Πεδία των συνδέσεων friendlists, friends και mutualfriends	53
Πίνακας 4: Πεδία των συνδέσεων like, books, movies, music και interests	53
Πίνακας 5: Πεδία του αντικειμένου Post.....	53
Πίνακας 6: Συνδέσεις του αντικειμένου Post.....	54
Πίνακας 7: Πεδία του αντικειμένου Comment	55
Πίνακας 8: Συνδέσεις του αντικειμένου Comment	55
Πίνακας 9: Απόδοση αλγορίθμου με ίσες τιμές παραμέτρων	71
Πίνακας 10: Απόδοση αλγορίθμου με $w_1=0.4$, $w_2=0.3$ και $w_3=0.3$	72
Πίνακας 11: Απόδοση αλγορίθμου με $w_1=0.3$, $w_2=0.4$ και $w_3=0.3$	73
Πίνακας 12: Απόδοση αλγορίθμου με $w_1=0.3$, $w_2=0.3$ και $w_3=0.4$	74
Πίνακας 13: Απόδοση αλγορίθμου με $w_1=0.5$, $w_2=0.3$ και $w_3=0.2$	75
Πίνακας 14: Απόδοση αλγορίθμου με $w_1=0.5$, $w_2=0.2$ και $w_3=0.3$	76
Πίνακας 15: Απόδοση αλγορίθμου με $w_1=0.3$, $w_2=0.5$ και $w_3=0.2$	77
Πίνακας 16: Απόδοση αλγορίθμου με $w_1=0.2$, $w_2=0.5$ και $w_3=0.3$	78
Πίνακας 17: Απόδοση αλγορίθμου με $w_1=0.3$, $w_2=0.2$ και $w_3=0.5$	79
Πίνακας 18: Απόδοση αλγορίθμου με $w_1=0.2$, $w_2=0.3$ και $w_3=0.5$	80

ΠΡΟΛΟΓΟΣ

Η παρούσα διπλωματική εργασία εκπονήθηκε στην Αθήνα κατά τη διάρκεια του ακαδημαϊκού έτους 2012-2013 στα πλαίσια φοίτησής μας στο διατμηματικό μεταπτυχιακό πρόγραμμα σπουδών “Οικονομική και Διοίκηση Τηλεπικοινωνιακών Δικτύων” του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών. Στόχος μας ήταν η εμβάθυνσή μας στον τομέα των συστημάτων συστάσεων, η μελέτη της εξέλιξής τους αλλά και η δημιουργία ενός νέου συστήματος ως μια εναλλακτική πρόταση στα συστήματα συστάσεων που υπάρχουν μέχρι σήμερα. Θα θέλαμε να ευχαριστήσουμε θερμά τον αναπληρωτή καθηγητή ΕΚΠΑ κύριο Ευστάθιο Χατζιευθυμιάδη για την ευκαιρία που μας έδωσε να επιλέξουμε το παρόν θέμα καθώς και τον υποψήφιο διδάκτορα ΕΚΠΑ κύριο Κωνσταντίνο Κολομβάτσο υπό την καθοδήγηση του οποίου φέραμε εις πέρας τη συγγραφή της εργασίας μας.

1. ΕΙΣΑΓΩΓΗ

Κατά τα τελευταία έτη, η δραματική αύξηση στη χρήση του Διαδικτύου και η ταχύτητα με την οποία εξαπλώνεται ο τεράστιος όγκος πληροφορίας που προέρχεται από αυτό έχουν οδηγήσει στο φαινόμενο της υπερπληροφόρησης (information overload). Μέσα σε αυτόν τον καταγίγισμό πληροφοριών οι χρήστες του Διαδικτύου δυσκολεύονται ολοένα και περισσότερο να ξεχωρίσουν το περιεχόμενο που τους ενδιαφέρει γρήγορα και αποτελεσματικά. Στην προσπάθεια δημιουργίας συστημάτων που θα επέλυαν αυτό το πρόβλημα, αναπτύχθηκαν τα συστήματα συστάσεων –ΣΣ (Recommender Systems) [11]. Τα συστήματα αυτά χρησιμοποιούνται σε διάφορους τομείς όπου είναι απαραίτητο το φιλτράρισμα της πληροφορίας. Λόγου χάρη, χρησιμοποιούνται σε εφαρμογές ψυχαγωγίας για προτάσεις ταινιών, μουσικής κλπ, σε διαδικτυακές υπηρεσίες για συστάσεις ταξιδιών και εστιατορίων, σε κοινωνικά μέσα για συστάσεις βίντεο, άρθρων κλπ. Τα ΣΣ είναι ιδιαίτερος σημαντικά σε εφαρμογές ηλεκτρονικού εμπορίου. Δεδομένου ότι η λήψη αποφάσεων των χρηστών για αγορά μέσα από μια μεγάλη ποσότητα διαθέσιμων προϊόντων στο ηλεκτρονικό εμπόριο αποτελεί φλέγον ζήτημα για τα ηλεκτρονικά καταστήματα, τα ΣΣ έχουν υιοθετηθεί ευρέως για να βοηθήσουν τους πελάτες να εντοπίσουν προϊόντα που τους ταιριάζουν και πιθανώς θα ήθελαν να αγοράσουν. Μπορεί ακόμη να βοηθήσουν χρήστες που έχουν ελλιπείς γνώσεις σχετικά με κάποιον τομέα προϊόντος. Στην ουσία, αυτά τα συστήματα εφαρμόζουν τεχνικές ανάλυσης δεδομένων σχετικά με τις προτιμήσεις κάθε πελάτη και την προηγούμενη συμπεριφορά του. Ως αποτέλεσμα, παρέχουν μια λίστα με προτεινόμενα προϊόντα για εκείνον στην επόμενη αλληλεπίδρασή του με το σύστημα τα οποία δεν είχε σκεφτεί να αγοράσει μέχρι εκείνη τη στιγμή.

Τα ΣΣ παρουσιάζουν πολλά ενδιαφέροντα πλεονεκτήματα. Πέρα από τις εξατομικευμένες προτάσεις αγοράς που κάνουν στους χρήστες τους μπορούν να προωθήσουν εξειδικευμένα προϊόντα που ενδεχομένως δεν έχουν μεγάλη ζήτηση και να ενισχύσουν την αλυσιδωτή πρόταση προϊόντων που σχετίζονται μεταξύ τους. Μπορούν ακόμη να δώσουν την ευκαιρία στους χρήστες να αξιολογήσουν τα προϊόντα εκφράζοντας δημοσίως την άποψή τους και κατ' επέκταση να βοηθήσουν άλλους χρήστες στη λήψη αποφάσεων.

Όσο περισσότερο ένας χρήστης χρησιμοποιεί μια ιστοσελίδα ηλεκτρονικού καταστήματος και αγοράζει αντικείμενα, τόσο περισσότερες πληροφορίες λαμβάνει το σύστημα για το χρήστη και συνεπώς βελτιώνονται οι συστάσεις που κάνει. Το γεγονός αυτό βοηθά στο να οικοδομηθεί μια σχέση προστιθέμενης αξίας μεταξύ της ιστοσελίδας και του χρήστη. Ένα καλό ΣΣ συμβάλλει στην ικανοποίηση του χρήστη καθώς βελτιώνει την εμπειρία του και αυξάνει την εμπιστοσύνη του στο σύστημα. Αυτό μεταφράζεται σε αύξηση των πωλήσεων και των κερδών για τις εταιρείες καθώς θα επιτυγχάνονται περισσότερες αγορές προϊόντων και έτσι ωθείται η ανταγωνιστικότητα μεταξύ των ηλεκτρονικών καταστημάτων. Για το λόγο αυτό, υπάρχει μεγάλη ερευνητική δραστηριότητα γύρω από την εύρεση καινοτόμων και αποτελεσματικών αλγορίθμων που θα δίνουν ακριβή αποτελέσματα. Τα συστήματα συστάσεων είναι μια ενεργή ερευνητική περιοχή στους τομείς της εξόρυξης δεδομένων και της μηχανικής μάθησης και αποτελεί κύριο θέμα συνεδρίων όπως το ACM RecSys, το SIGIR και το KDD.

Μια νέα τάση που διέπει τον Παγκόσμιο Ιστό την τελευταία δεκαετία είναι η μετατροπή των παθητικών χρηστών του Διαδικτύου σε ενεργούς χρήστες που οι ίδιοι παράγουν περιεχόμενο. Πλέον, έχουν δημιουργηθεί πλατφόρμες, που ενισχύουν την αλληλεπίδραση μεταξύ των χρηστών, τους δίνουν τη δυνατότητα να παρουσιάσουν τις απόψεις τους δημοσίως και να συνεισφέρουν στο περιεχόμενο του ιστοτόπου. Ιδιαίτερα δημοφιλή καθίστανται τα λεγόμενα **κοινωνικά δίκτυα** (social networks), δηλαδή

ιστοχώροι που προσομοιώνουν την κοινωνική λειτουργικότητα των χρηστών. Ένας χρήστης που έχει λογαριασμό σε κάποιο κοινωνικό δίκτυο μπορεί να αλληλεπιδράσει με όμοιους χρήστες, να επικοινωνήσει μαζί τους, να ανεβάσει φωτογραφίες, να δημοσιεύσει μια γνώμη, να σχολιάσει την άποψη ενός άλλου χρήστη, να διαθέσει δημοσίως πληροφορίες σχετικά με προτιμήσεις του, να συμμετέχει σε ομάδες, κ.λπ.

Το πλούσιο περιεχόμενο που παρέχεται μέσω των κοινωνικών δικτύων θα μπορούσε να αποβεί πολύ χρήσιμο για τη βελτίωση των ΣΣ ηλεκτρονικού εμπορίου. Στις περισσότερες περιπτώσεις, η κοινωνική επιρροή διαδραματίζει σπουδαίο ρόλο στο marketing των προϊόντων. Άτομα τα οποία συνδέονται με κάποιον κοινωνικό δεσμό, μπορούν να επηρεάσουν ο ένας τον άλλο σχετικά με κάποια λήψη απόφασης. Διαισθητικά, όταν κάποιος επιθυμεί να αγοράσει ένα προϊόν με το οποίο δεν είναι εξοικειωμένος, συχνά συμβουλευτεί τους φίλους ή τους γνωστούς του έστω, οι οποίοι είχαν ήδη κάποια εμπειρία με το εν λόγω προϊόν. Όταν κάποιος φίλος συστήσει ένα προϊόν, το άτομο τείνει να αποδεχθεί την πρότασή του, επειδή θεωρεί τη γνώμη του αξιόπιστη. Το γεγονός αυτό δίνει μια άλλη διάσταση στη δημιουργία ΣΣ. Οι παραδοσιακές εφαρμογές συστάσεων δε λαμβάνουν υπόψη τους τις ρητές κοινωνικές σχέσεις μεταξύ των χρηστών, μια πληροφορία που πλέον είναι διαθέσιμη μέσα από τα κοινωνικά δίκτυα. Εάν κάποιο σύστημα παράγει αποτελέσματα προσομοιώνοντας τον τρόπο με τον οποίο γίνονται οι συστάσεις στην πραγματική ζωή, από στόμα σε στόμα μεταξύ γνωστών, έχει ενδεχομένως μεγάλες πιθανότητες να βελτιωθεί από πλευράς απόδοσης. Αναφορικά με την ακρίβεια πρόβλεψης, οι συμπληρωματικές πληροφορίες σχετικά με τους χρήστες και τους φίλους τους που προέρχονται από τα κοινωνικά δίκτυα βελτιώνει την κατανόηση της συμπεριφοράς των χρηστών. Ως εκ τούτου, είναι δυνατή η μοντελοποίηση και η ερμηνεία των προτιμήσεων των χρηστών με μεγαλύτερη ακρίβεια, βελτιώνοντας έτσι την αξιοπιστία της πρόβλεψης. Επιπροσθέτως, έχοντας γνώση σχετική με τους φίλους του χρήστη μπορεί να προτείνει σε εκείνον προϊόντα που αρέσουν στους φίλους του σε περίπτωση που ο ίδιος δεν έχει αλληλεπιδράσει αρκετά με το σύστημα και κατά συνέπεια δεν μπορεί να βγάλει συμπέρασμα για τις προτιμήσεις του. Η οπτική αυτή μπορεί να έχει ισχύ εάν δύο άνθρωποι είναι πραγματικοί φίλοι οπότε είναι πιθανό να έχουν και αρκετά κοινά γνωρίσματα και παρόμοιες απόψεις.

Το ανωτέρω σκεπτικό μας παρακίνησε στο να σχεδιάσουμε ένα πρότυπο συστήματος που να μπορεί να εκμεταλλευτεί πληροφορίες από προφίλ των χρηστών του σε κάποιο κοινωνικό δίκτυο. Στην παραγωγή των τελικών συστάσεων θα διαδραματίζουν σημαντικό ρόλο οι κοινωνικές επαφές του χρήστη. Επειδή ωστόσο, υπάρχει μεγάλη πιθανότητα ο χρήστης να μην έχει πραγματική αλληλεπίδραση με τις επαφές του προφίλ του, αναπτύξαμε έναν αλγόριθμο βασισμένο στη θεωρία γράφων ο οποίος μεταξύ άλλων φιλτράρει τους κοινωνικούς δεσμούς και εντοπίζει τις επαφές εκείνες που ταιριάζουν περισσότερο με το χρήστη στον οποίο θέλουμε να κάνουμε συστάσεις.

Η εργασία μας απαρτίζεται από έξι κεφάλαια. Στο επόμενο κεφάλαιο, θα κάνουμε μια θεωρητική επισκόπηση των ΣΣ και στη συνέχεια θα περιγράψουμε τα κύρια χαρακτηριστικά του Κοινωνικού Ιστού. Στο τρίτο κεφάλαιο, θα παρουσιαστεί η αρχιτεκτονική ενός προτεινόμενου ΣΣ που χρησιμοποιεί πληροφορίες από κοινωνικά δίκτυα. Στο τέταρτο κεφάλαιο θα αναλυθεί ο αλγόριθμος που χρησιμοποιεί το σύστημα για να παράγει συστάσεις ενώ στο πέμπτο κεφάλαιο θα γίνει η αξιολόγηση του εν λόγω αλγορίθμου. Τέλος, στο έκτο κεφάλαιο θα παρουσιάσουμε τα τελικά μας συμπεράσματα.

2. ΣΥΣΤΗΜΑΤΑ ΣΥΣΤΑΣΕΩΝ

Τα **συστήματα συστάσεων –ΣΣ** (Recommender systems) [1, 2] είναι εφαρμογές που αξιοποιούνται κυρίως στις ιστοσελίδες του ηλεκτρονικού εμπορίου (e-commerce) για να προτείνουν ενδιαφέροντα και χρήσιμα προϊόντα αλλά και να παρέχουν πληροφορίες στους καταναλωτές προκειμένου αυτοί να αποφασίσουν ποια προϊόντα θα αγοράσουν. Με άλλα λόγια, ένα ΣΣ μπορεί να χρησιμοποιηθεί για να καθοδηγήσει έναν πελάτη να κάνει τη σωστή επιλογή στοχεύοντας στο φιλτράρισμα των ανεπιθύμητων αντικειμένων ή προβλέποντας τα ενδιαφέροντα προϊόντα αυτόματα εκ μέρους των χρηστών σύμφωνα με τις προσωπικές τους προτιμήσεις.

Τα ΣΣ χρησιμοποιούνται ευρέως σε πολλές εφαρμογές που προτείνουν προϊόντα, υπηρεσίες και πληροφορίες στοιχείων στους δυνητικούς καταναλωτές. Για παράδειγμα, ένα ευρύ φάσμα εταιρειών, όπως οι Amazon.com, Netflix.com, Half.com, CDNow, JC Penney, MovieLens.com, Drugstore.com, MovieFinder.com, Reel.com, eBay και η Procter & Gamble έχουν επεκτείνει με επιτυχία τα εμπορικά ΣΣ και έχουν αναφέρει αύξηση των πωλήσεων και μεγαλύτερη εμπιστοσύνη εκ μέρους των πελατών. Πολλές εταιρείες λογισμικού, όπως λόγω χάρη η Net Perceptions¹ παρέχουν αυτόνομες γενικές τεχνολογίες σύστασης [7].

2.1 Διαδικασία σύστασης

Γενικά, κάθε ΣΣ ακολουθεί μια συγκεκριμένη διαδικασία για την παραγωγή των συστάσεων. Αρχικά το σύστημα θα πρέπει να επεξεργαστεί κάποια δεδομένα εισόδου προτού να παράγει αποτελέσματα. Τρεις πιθανές πηγές πληροφορίας μπορούν να προσδιοριστούν ως είσοδος για τη διαδικασία σύστασης. Οι διαθέσιμες πηγές είναι:

- τα δεδομένα χρήστη,
- τα δεδομένα αντικειμένου και
- οι πιθανές αλληλεπιδράσεις ενός χρήστη με το αντικείμενο (αξιολογήσεις, αγορές κτλ).

Τα δημογραφικά στοιχεία για το χρήστη όπως το φύλο, η ηλικία κ.λπ. αλλά και οι γνώσεις σχετικά με τις προτιμήσεις του συνθέτουν ένα "*προφίλ χρήστη*". Παράλληλα, κάθε αντικείμενο περιγράφεται με τη βοήθεια χαρακτηριστικών γνωρισμάτων, τα οποία συμβάλλουν στη διαμόρφωση ενός προφίλ αντικειμένου, που ονομάζεται "*περιεχόμενο*". Για παράδειγμα ένα βιβλίο μπορεί να έχει ως χαρακτηριστικά γνωρίσματα τον τίτλο του, το συγγραφέα του, κ.λπ.

Η καταγραφή των δεδομένων του χρήστη μπορεί να γίνει είτε *άμεσα* (explicitly) είτε *έμμεσα* (implicitly). Καταγραφή δεδομένων άμεσα σημαίνει ότι ο χρήστης αλληλεπιδρά με το σύστημα ειδικά για το σκοπό της παροχής των απαραίτητων πληροφοριών σε αυτό. Για παράδειγμα, τα δημογραφικά χαρακτηριστικά του χρήστη καταγράφονται άμεσα όταν ζητούνται κατά τη διάρκεια της εγγραφής του σε ένα ηλεκτρονικό κατάστημα. Αντίθετα, καταγραφή δεδομένων έμμεσα σημαίνει ότι το σύστημα λαμβάνει τα δεδομένα ενώ ο χρήστης αλληλεπιδρά με αυτό για άλλο σκοπό. Λόγου χάρη, τα δεδομένα του χρήστη αποκτώνται έμμεσα όταν συγκεντρώνονται από τα συστήματα καταγραφής συναλλαγών (π.χ. κατά την πληρωμή ενός προϊόντος με τη χρήση

¹ <http://www.netperceptions.com.au/>

πιστωτικών καρτών) ή ακόμη μπορούν να συλλεχθούν με βάση τα δεδομένα της κάρτας SIM του κινητού τηλεφώνου.

Τα δεδομένα σχετικά με τα προϊόντα είναι εύκολο να ληφθούν με άμεσο τρόπο καθώς τα προϊόντα ως επί το πλείστον καταχωρούνται με όλα τα είδη των δεδομένων, των ετικετών, ή των χαρακτηριστικών τους. Για παράδειγμα, σε έναν ιστότοπο που αφορά ταινίες αναφέρεται το είδος της κάθε ταινίας, δηλαδή επιστημονικής φαντασίας, δράση ή κωμωδία. Το είδος της ταινίας αποτελεί παράδειγμα ενός στοιχείου του προϊόντος που παρέχεται από το/στο σύστημα. Τα δεδομένα πρέπει να παρέχονται από τον ιδιοκτήτη του ηλεκτρονικού καταστήματος ή οι χρήστες μπορούν να παρέχουν τα χαρακτηριστικά του προϊόντος δίνοντας τις κατάλληλες ετικέτες σε αυτό. Έμμεσα, τα στοιχεία μπορούν να εξαχθούν από έγγραφα χρησιμοποιώντας κατάλληλα εργαλεία.

Η αλληλεπίδραση χρήστη-αντικειμένου μπορεί και αυτή να πραγματοποιείται άμεσα ή έμμεσα. Η άμεση αλληλεπίδραση συνίσταται στην ιδέα ότι κάθε αντικείμενο μπορεί να εκτιμηθεί από τους χρήστες μέσω “αξιολογήσεων” (ratings), οι οποίες αντιπροσωπεύουν τη χρησιμότητα ενός αντικειμένου για ένα χρήστη. Κάθε αξιολόγηση εκπροσωπεί ένα χρήστη μέσω ενός ή περισσότερων από τις ακόλουθες επιλογές:

Διαδική (Binary): Η επιλογή αυτή βασίζεται στη λογική δύο δυνατών τιμών σχετικά με την προτίμηση του χρήστη (like / dislike).

Αριθμητική (Numeric): Μία αξιολόγηση μπορεί να δοθεί σε μια συγκεκριμένη κλίμακα. Αυτό είναι συχνά μια κλίμακα πέντε αστέρων (κλίμακα Likert [3, 4]), όπου ένα αστέρι αντιπροσωπεύει τη χαμηλότερη κατάταξη και τα πέντε αστέρια την υψηλότερη κατάταξη. Επίσης, είναι πιθανόν να υπάρχουν κλίμακες άνω των πέντε αστέρων. Η Amazon.com λόγω χάρη χρησιμοποιεί 1 αστέρι για την επιλογή “Το μισώ”, 2 αστέρια για την επιλογή “Δεν μου αρέσει”, 3 αστέρια για την επιλογή “Είναι εντάξει”, 4 αστέρια για την επιλογή “Μου αρέσει” και 5 αστέρια για την επιλογή “Το αγαπώ”.

Συμβολική (Symbolic): Η επιλογή σχετικά με την προτίμηση του χρήστη είναι μια εικόνα με την οποία ο αξιολογητής επευφημεί (π.χ. μια εικόνα “thumbs up”) ή αποδοκιμάζει το προϊόν (μια εικόνα “thumbs down”).

Κείμενο (Textual): Σε αυτήν την περίπτωση η εκτίμηση για ένα προϊόν μπορεί να γίνει μέσω ενός γραπτού κειμένου του αξιολογητή σε έναν ιστότοπο (π.χ. μια κριτική ή ένα σχόλιο). Αυτός ο τύπος αξιολόγησης είθισται να συναντάται σε ιστοσελίδες κοινωνικής δικτύωσης.

Ήχος (Audio): Η επιλογή αυτή αφορά την αξιολόγηση μέσω ενός ηχητικού μηνύματος π.χ. μια προφορική κριτική του προϊόντος.

Οπτική (Visual): Σε αυτού του είδους επιλογές μπορεί να χρησιμοποιηθούν πολυμεσικές εφαρμογές για την εξαγωγή της αξιολόγησης (π.χ. μια βίντεο-ηχογραφημένη κριτική του προϊόντος με τη χρήση ενός κινητού τηλεφώνου).

Υπάρχουν πολλές άλλες μέθοδοι για την απόκτηση άμεσων αξιολογήσεων, όπως το να ζητηθεί από το χρήστη να αναφέρει ρητά τις προτιμήσεις του ή να επιλέξει τα πιο αγαπημένα αντικείμενα από μια λίστα στοιχείων. Είναι δυνατό να χρησιμοποιηθούν επίσης υβριδικές τεχνικές που συνδυάζουν γενικές ερωτήσεις σχετικά με τις προτιμήσεις ακολουθούμενες από αξιολογήσεις των επιμέρους αντικειμένων.

Όλοι οι χρήστες δεν αξιολογούν όλα τα αντικείμενα που έχουν αγοράσει ή έχουν δει, επειδή απλά δε θέλουν να ξοδεύουν το χρόνο τους αξιολογώντας τα αντικείμενα ή δεν βλέπουν το λόγο για να το κάνουν. Ακόμη δεν είναι απαραίτητο ότι όλοι οι πελάτες που εγγράφονται στο ηλεκτρονικό κατάστημα είναι διατεθειμένοι να δώσουν όλες τις προσωπικές πληροφορίες τους. Για αυτούς τους χρήστες μια άλλη πηγή πληροφοριών

απαιτείται για να ξεπεραστεί η έλλειψη των αξιολογήσεων. Μια προσέγγιση σε αυτό το πρόβλημα είναι να χρησιμοποιηθούν έμμεσοι τρόποι αξιολόγησης μέσω της παρατήρησης της συμπεριφοράς του χρήστη. Ένα σενάριο για τη σύλληψη του προϊόντος με έμμεσο τρόπο μπορεί να είναι ένας καταναλωτής που αγοράζει ένα προϊόν, ή η επιστροφή και επισκευή ενός προϊόντος. Στην περίπτωση που ο χρήστης έχει αγοράσει πολλές φορές το ίδιο προϊόν, το σύστημα μπορεί να συμπεράνει ότι το εν λόγω προϊόν είναι αρεστό. Αντίθετα, αν ο χρήστης επιστρέψει την επόμενη ημέρα το προϊόν που είχε αγοράσει και αγοράσει ένα άλλο προϊόν της ίδιας κατηγορίας, το σύστημα θα μπορούσε να συνάγει ότι στο χρήστη δεν άρεσε το προϊόν που αγόρασε πρώτο. Ένα άλλο τυπικό σενάριο της έμμεσης συλλογής δεδομένων στο διαδίκτυο είναι η καταγραφή των δεδομένων όταν ο χρήστης κάνει περιήγηση στις περιγραφές των προϊόντων, όπως στο Amazon.com, ή σαρώνει προϊόντα στο ηλεκτρονικό κατάστημα με ένα κινητό τηλέφωνο στα πλαίσια μιας άλλης εφαρμογής. Το βασικό κίνητρο για τη χρήση έμμεσων αξιολογήσεων είναι ότι δεν υπάρχει κόστος για τον αξιολογητή όσον αφορά την εξέταση και την αξιολόγηση του στοιχείου. Ωστόσο, εάν η αξιολόγηση προκύπτει έμμεσα, μπορεί οι συστάσεις να μην είναι τόσο ακριβείς και αξιόπιστες. Για παράδειγμα, ο ίδιος υπολογιστής μπορεί να χρησιμοποιείται από περισσότερα του ενός άτομα. Όμως, το σύστημα μπορεί να θεωρεί ενδεχομένως ότι όλες οι διαφορετικές αλληλεπιδράσεις προέρχονται από τον ίδιο χρήστη και συνεπώς να καταλήγει σε λανθασμένες συστάσεις. Επιπλέον, ένα άλλο βασικό μειονέκτημα είναι η αδυναμία του συστήματος να εξάγει συμπεράσματα όταν δεν υπάρχουν ενέργειες από την πλευρά του χρήστη.

2.2 Μορφές της διαδικασίας σύστασης

Αφού συλλεχθούν όλα τα απαραίτητα δεδομένα εισόδου, η διαδικασία της σύστασης προϋποθέτει τη μαζική επεξεργασία τους και τη διάθεση του αποτελέσματος είτε προς όλους τους χρήστες είτε προς έναν επιλεγμένο αριθμό εν δυνάμει καταναλωτών. Εν γένει, υπάρχουν τέσσερις επιλογές πάνω στις οποίες μπορεί να βασίζεται μια διαδικασία σύστασης:

Ένας προς έναν: Αυτές είναι προσεγγίσεις όπου η σύσταση ενός συγκεκριμένου χρήστη απευθύνεται σε έναν άλλο χρήστη, συμπεριλαμβανομένου του ίδιου. Ειδικότερα, ο χρήστης που κάνει τη σύσταση γνωρίζει πολλά για έναν ορισμένο τύπο του προϊόντος και / ή πολλά για ένα συγκεκριμένο ομότιμο (peer) χρήστη οπότε συνιστά ένα συγκεκριμένο προϊόν σε αυτό το χρήστη. Αυτό το είδος σύστασης θα μπορούσε να παραλληλιστεί με το λεγόμενο word-of-mouth, όπου ένας καταναλωτής λέει προφορικά σε έναν άλλο καταναλωτή για την εμπειρία του σχετικά με ένα προϊόν. Μια προσέγγιση όπου τα στοιχεία του ιστορικού ενός χρήστη χρησιμοποιούνται για να προβλέψουν μια σύσταση για τον ίδιο τον χρήστη ονομάζεται αλγόριθμος φιλτραρίσματος με βάση το περιεχόμενο [3]. Είναι πιθανό ένας συγκεκριμένος χρήστης να κάνει μια σύσταση στον εαυτό του, ώστε να μην ξεχνά την επιλογή του προϊόντος που έχει κάνει και με την οποία είναι ικανοποιημένος. Τυπικά παραδείγματα στο διαδίκτυο είναι ηλεκτρονικές λίστες αγοράς και αυτόματα αποθηκευμένα ηλεκτρονικά καλάθια όταν αποσυνδεθεί ο χρήστης πριν από την αγορά, που χρησιμοποιούνται στο Amazon.com για συστάσεις.

Πολλοί προς πολλούς: Σε αυτές τις προσεγγίσεις, οι αξιολογήσεις πολλών απροσδιόριστων χρηστών μπορεί να συγκεντρώνονται και να υπόκεινται σε επεξεργασία (λόγου χάρη να υπολογίζεται ο μέσος όρος τους αν πρόκειται για αριθμητικές αξιολογήσεις) για ένα συγκεκριμένο προϊόν (το λεγόμενο φιλτράρισμα με βάση το αντικείμενο) και να διατίθενται σε πολλούς άλλους απροσδιόριστους χρήστες. Υπάρχει επίσης το ενδεχόμενο οι αξιολογήσεις των χρηστών να μην είναι αριθμητικά

μετρήσιμες (π.χ. κριτικές για το προϊόν σε κείμενο) και να διατίθενται ως έχουν για πολλούς χρήστες. Επειδή σε αυτήν την περίπτωση η σύσταση δεν έχει δημιουργηθεί ειδικά για ένα χρήστη, επαφίεται σε αυτόν να κρίνει αν αυτή ταιριάζει με τις ανάγκες του.

Πολλοί προς έναν: Σε αυτήν την περίπτωση, συστάσεις από πολλούς χρήστες αθροίζονται σε ένα συγκεκριμένο χρήστη. Η προσέγγιση αυτή βασίζεται στην ιδέα ότι οι άνθρωποι είναι ανεξάρτητα άτομα και τους αρέσουν διαφορετικά προϊόντα. Υπάρχουν αλγόριθμοι που εστιάζουν σε αυτό το γεγονός, λαμβάνοντας συστάσεις ειδικών χρηστών που έχουν ληφθεί έμμεσα και / ή άμεσα και απευθύνουν συστάσεις σε έναν άλλο χρήστη με παρόμοιο προφίλ. Η διαδικασία αυτή αναφέρεται συχνά ως συνεργατικό φιλτράρισμα [3].

Ένας προς πολλούς: Σε αυτήν την περίπτωση η σύσταση ενός χρήστη που έχει κάποια ιδιαίτερη σχέση με ένα προϊόν διατίθεται σε πολλούς ενδιαφερόμενους χρήστες. Ένα παράδειγμα είναι ένας εμπειρογνώμονας κρασιού που μοιράζεται την εμπειρία του για ένα ορισμένο κρασί με τους φίλους του μέσα από ένα κοινωνικό δίκτυο. Η εφαρμογή LivingSocial.com του Facebook είναι μέχρι στιγμής η πιο δημοφιλής αυτών των ΣΣ που βασίζονται στα κοινωνικά δίκτυα. Επιτρέπει στους χρήστες να επιλέξουν ρητά τα προϊόντα που τους αρέσουν, να τα κριτικάρουν ρητά και να μοιράζονται αυτά τα σχόλια με φίλους στις πλατφόρμες κοινωνικών δικτύων όπως το Facebook. Αλλά υπάρχουν επίσης εφαρμογές του Facebook, που συλλέγουν συστάσεις προϊόντων έμμεσα και τις διαθέτουν στους φίλους, οι λεγόμενες κοινωνικές διαφημίσεις. Για παράδειγμα, το Facebook έμμεσα συλλέγει συναλλαγές κάποιου χρήστη του με το δικτυακό τόπο εισιτηρίων “Fandago” και τις γνωστοποιεί σε άλλους χρήστες. Αυτή η γνωστοποίηση ουσιαστικά μπορεί να αποτελέσει ταυτόχρονα σύσταση για μια ταινία ή ένα θεατρικό έργο. Σε αυτή και παρόμοιες περιπτώσεις, η προστασία της ιδιωτικής ζωής είναι ένα σημαντικό ζήτημα. Η ιστορία δείχνει ότι οι καταναλωτές δε θέλουν να μοιραστούν συστάσεις προϊόντων που έχουν αποδοθεί στο όνομά τους με άλλους τυχαίους χρήστες.

Ακολουθώντας αυτό το θεωρητικό μοντέλο θα αναλυθούν εν συνεχεία οι βασικότερες τεχνολογίες που χρησιμοποιούν τα ΣΣ προκειμένου να προτείνουν στους χρήστες τα κατάλληλα προϊόντα ή υπηρεσίες.

2.3 Προσεγγίσεις σύστασης

Πολλές προσεγγίσεις σύστασης έχουν αναπτυχθεί, η απλούστερη από τις οποίες είναι η μη εξατομικευμένη προσέγγιση (non-personalized) [4]. Ωστόσο, στις περισσότερες περιπτώσεις οι συστάσεις απευθύνονται προς ένα συγκεκριμένο χρήστη, λαμβάνοντας υπόψη την ατομικότητά του. Ως εκ τούτου, εξατομικευμένα προϊόντα μπορεί να συνιστώνται σε εν δυνάμει πελάτες και κατά συνέπεια να επιτύχουν να μετατρέψουν τους ανθρώπους που απλά περιηγούνται στο Διαδίκτυο σε καταναλωτές. Οι σημαντικότερες προσεγγίσεις για τα εξατομικευμένα συστήματα είναι η προσέγγιση βασισμένη στο περιεχόμενο (content-based filtering - CB) και η προσέγγιση συνεργατικού ή κοινωνικού φιλτραρίσματος (collaborative filtering - CF or social filtering) [4]. Υπάρχουν και άλλες προσεγγίσεις που λειτουργούν συμπληρωματικά, όπως η βασισμένη στη γνώση προσέγγιση (knowledge-based filtering - KB) καθώς επίσης και η βασισμένη στα δημογραφικά στοιχεία προσέγγιση (Demography-based filtering - DB) [4].

2.3.1 Μη εξατομικευμένη προσέγγιση (non-personalized)

Στην περίπτωση αυτή οι συστάσεις είναι ίδιες για κάθε πελάτη και είτε επιλέγονται χειροκίνητα από το online κατάστημα, με βάση τη δημοτικότητα των αντικειμένων (μέσος όρος αξιολογήσεων, στοιχεία για τις πωλήσεις, συνολικές επισκέψεις κ.λπ) είτε αποτελούν τα N καλύτερα νέα προϊόντα του ηλεκτρονικού καταστήματος. Το πλεονέκτημα αυτής της μεθόδου είναι ότι είναι κατανοητή για τον ιδιοκτήτη του καταστήματος και τα δεδομένα είναι εύκολο να συλλεχθούν. Ωστόσο, οι συστάσεις σε αυτό το σύστημα είναι ίδιες για όλους τους χρήστες, δεν υπάρχει εξατομίκευση και επομένως είναι πιθανόν οι συστάσεις να μην αντικατοπτρίζουν τις ανάγκες του χρήστη.

2.3.2 Φιλτράρισμα βασισμένο στο περιεχόμενο (content-based filtering - CB)

Στις προσεγγίσεις φιλτραρίσματος βασισμένου στο περιεχόμενο, το σύστημα συστήνει αντικείμενα βασισμένο στη συσχέτιση μεταξύ του περιεχομένου των αντικειμένων (π.χ. λέξεις-κλειδιά που περιγράφουν τα αντικείμενα, όπως το είδος άλμπουμ, οι καλλιτέχνες, κλπ., για τα μουσικά κομμάτια) και των προτιμήσεων του χρήστη [22]. Η προσέγγιση σύστασης με βάση το περιεχόμενο, συνεπώς, έχει πρόσβαση σε ένα σύνολο χαρακτηριστικών που παρουσιάζουν τα προϊόντα και χρησιμοποιεί παρόμοιες τεχνικές με την ανάκτηση πληροφοριών. Ο χρήστης εκφράζει ανάγκες, οφέλη και περιορισμούς και έτσι το σύστημα κάνει σύγκριση μεταξύ του προφίλ του χρήστη και του περιεχομένου των αντικειμένων. Μέσα από αυτή τη σύγκριση αναγνωρίζει τις προτιμήσεις του χρήστη σε σχέση με αυτά τα χαρακτηριστικά και χρησιμοποιεί τις υπολογισμένες προτιμήσεις για να προτείνει αντικείμενα με παρόμοια γνωρίσματα. Η προσέγγιση αυτή βασίζεται στη λογική ότι εάν είναι εμφανές ότι ένας χρήστης έχει δείξει κάποια ιδιαίτερη προτίμηση για ένα σύνολο αντικειμένων με παρόμοιο περιεχόμενο, τότε ο εν λόγω χρήστης πιθανότατα ενδιαφέρεται και για οποιοδήποτε άλλο αντικείμενο που θα μπορούσε να υπάγεται σε αυτήν την κατηγορία αντικειμένων. Οι βασισμένες στο περιεχόμενο προσεγγίσεις μπορεί πιθανώς να εκμεταλλεύονται το ιστορικό προηγούμενων επερωτήσεων του χρήστη για να χτίσουν το προφίλ του ή το ιστορικό προηγούμενων αγορών. Τα NewsWeeder και InfoFinder αποτελούν παραδείγματα συστημάτων συστάσεων με βάση το περιεχόμενο.

Οι προσεγγίσεις που βασίζονται στο περιεχόμενο εκμεταλλεύονται άμεσα τις πληροφορίες του προϊόντος με αποτέλεσμα το σύστημα να συνιστά στους πελάτες προϊόντα που ταιριάζουν με συγκεκριμένα ενδιαφέροντα ή προτιμήσεις τους, ενώ αυτά τα προϊόντα μπορεί να μην παρουσιάζουν ενδιαφέρον για άλλους. Έτσι, εκμεταλλεύονται τις αξιολογήσεις του ίδιου του χρήστη για να δημιουργήσουν το προφίλ του και δεν στηρίζονται σε ανατροφοδότηση από άλλους χρήστες. Επιπλέον, στη βασισμένη στο περιεχόμενο προσέγγιση, η μοντελοποίηση και η μέτρηση της ομοιότητας μεταξύ των προϊόντων μπορεί να πραγματοποιηθεί χωρίς σύνδεση (offline). Ένα άλλο πλεονέκτημα των συστημάτων αυτών είναι ότι χρησιμοποιούν πολύ συχνά εξηγήσεις σχετικά με το πώς προέκυψαν τα προϊόντα που προτείνουν, οπότε ενισχύεται η εμπιστοσύνη των χρηστών στο σύστημα και αυξάνονται οι πιθανότητες να δεχτεί ο χρήστης την πρόταση που του έγινε. Τέλος, τα βασισμένα στο περιεχόμενο συστήματα μπορούν να συστήσουν νέα προϊόντα που δεν έχουν ακόμα αξιολογηθεί από κανέναν χρήστη, καθώς συλλέγουν πληροφορία από την περιγραφή του προϊόντος που είναι δεδομένη εξ αρχής.

Από την άλλη πλευρά, τα συστήματα φιλτραρίσματος με βάση το περιεχόμενο επικρίνονται για κάποιες αδυναμίες:

1. Οι μέθοδοι ανάκτησης περιεχομένου μπορούν να εφαρμοστούν μόνο σε λίγα είδη περιεχομένου, όπως κείμενα και εικόνες, και τα εξαγόμενα χαρακτηριστικά μπορούν να συλλάβουν μόνο ορισμένες πτυχές του περιεχομένου. Αυτό συμβαίνει διότι πληροφορίες πολυμέσων, όπως εικόνες και ήχοι, δεν μπορούν να αναλυθούν από το σύστημα δεδομένου ότι το περιεχόμενο των πολυμέσων είναι δύσκολο να οριστεί.
2. Τα βασισμένα στο περιεχόμενο συστήματα συστάσεων παρέχουν συστάσεις απλώς με βάση τα προφίλ των χρηστών. Ως εκ τούτου, οι χρήστες δεν έχουν την ευκαιρία να εξερευνήσουν αντικείμενα που δεν είναι παρόμοια με τα είδη που περιλαμβάνονται στο προφίλ τους.
3. Τα βασισμένα στο περιεχόμενο συστήματα πάσχουν από το πρόβλημα αργής εκκίνησης για το νέο χρήστη (cold start new user problem). Πιο συγκεκριμένα, ένα τέτοιο σύστημα δε δύναται να καταλάβει τη γενική συμπεριφορά στις προτιμήσεις του χρήστη και να κάνει σωστές και ακριβείς συστάσεις, εάν δεν έχει συλλέξει αρκετά δεδομένα από τις αξιολογήσεις προϊόντων του χρήστη. Έτσι, το σύστημα δεν μπορεί να κάνει αξιόπιστες προτάσεις σε έναν καινούριο χρήστη που δεν έχει αλληλεπιδράσει ακόμα αρκετά με το σύστημα.
4. Όταν οι χρήστες έχουν αξιολογήσει πολλά αντικείμενα, οι προτιμήσεις τους στα διαμορφωμένα προφίλ χρήστη είναι δύσκολο να αλλάξουν. Αλλά επειδή στην πραγματικότητα οι προτιμήσεις εξελίσσονται, αυτό αποτελεί πρόβλημα. Το πρόβλημα αυτό ονομάζεται stability vs plasticity problem [8].
5. Αυτά τα συστήματα δεν εξερευνούν τους έμμεσους δεσμούς μεταξύ των χρηστών διότι περιορίζονται στις λέξεις-κλειδιά που χρησιμοποιούνται από τους χρήστες και τις περιγραφές των αντικειμένων.
6. Ένα βασισμένο στο περιεχόμενο σύστημα δεν μπορεί να κάνει καλές συστάσεις αν το περιεχόμενο δεν περιέχει αρκετή πληροφορία προκειμένου να μπορέσει να διαχωρίσει ποια προϊόντα αρέσουν στο χρήστη από εκείνα που δε του αρέσουν. Έτσι, η αυτόματη και η χειρωνακτική δήλωση των χαρακτηριστικών γνωρισμάτων μπορεί να μην αρκεί για να γίνει ο διαχωρισμός των προϊόντων, ο οποίος συμβάλει στην εξαγωγή του προφίλ του χρήστη, αλλά να απαιτείται ανάλυση περιεχομένου.
7. Επειδή ο υπολογισμός της ομοιότητας μεταξύ των αντικειμένων περιορίζεται στα χαρακτηριστικά τους, η ποιότητα των προϊόντων δεν μπορεί να αξιολογηθεί.

2.3.3 Συνεργατικό ή κοινωνικό φιλτράρισμα (collaborative filtering – CF or social filtering)

Οι συνεργατικές προσεγγίσεις (Collaborative-based) παρέχουν συστάσεις προϊόντων ή αυτόματες προβλέψεις (φιλτράρισμα) για τα ενδιαφέροντα ενός χρήστη, που βασίζονται στις απόψεις άλλων χρηστών που παρουσιάζουν παρόμοιο σκεπτικό ή συμπεριφορά (συνεργασία) με το χρήστη-στόχο. Επομένως, η συνεργατική σύσταση διαφέρει από τη σύσταση που βασίζεται στο περιεχόμενο στο ότι συνιστά στοιχεία που βασίζονται στο τι άρεσε σε παρόμοιους χρήστες αντί για τον ανεξάρτητο χρήστη και δεν απαιτεί εξωγενείς πληροφορίες ούτε για τους χρήστες ούτε για τα προϊόντα, όπως περιεχόμενο ή περιγραφές. Καθαρά συνεργατικά συστήματα, όπως το GroupLens [36] δε γνωρίζουν τίποτα σχετικά με τα ίδια τα αντικείμενα αλλά γνωρίζουν μόνο τι πιστεύουν οι χρήστες τους. Μια βασική παραδοχή της προσέγγισης του συνεργατικού φιλτραρίσματος είναι ότι εκείνοι που συμφώνησαν στο παρελθόν τείνουν να συμφωνήσουν και πάλι στο μέλλον.

Το συνεργατικό φιλτράρισμα αναφέρεται ως η πιο ευρέως υιοθετημένη και επιτυχημένη προσέγγιση σύστασης. Για παράδειγμα οι μεγάλοι εμπορικοί προμηθευτές, όπως η Amazon.com το χρησιμοποιούν, αν και με έναν απλουστευμένο τρόπο. Οι ερευνητές προχωρούν ενεργά τις τεχνολογίες συνεργατικού φιλτραρίσματος σε ποικίλες πλευρές περιλαμβάνοντας τη σχεδίαση αλγορίθμου, τη σχεδίαση της αλληλεπίδρασης ανθρώπου-υπολογιστή, την ανάλυση του κινήτρου του καταναλωτή και την προστασία της ιδιωτικότητας.

Το συνεργατικό φιλτράρισμα διαθέτει χαρακτηριστικά που αντιμετωπίζουν ορισμένες από τις αδυναμίες της βασισμένης στο περιεχόμενο προσέγγισης. Τυπικά, τα συστήματα συστάσεων συνεργατικού φιλτραρίσματος δε χρησιμοποιούν το πραγματικό περιεχόμενο των στοιχείων για σύσταση και δεν απαιτούν καμία γνώση σχετικά με τα χαρακτηριστικά του προϊόντος (domain independent). Οπότε, η φύση των προϊόντων δεν χρειάζεται να αναλυθεί, και έτσι τα συνιστώμενα προϊόντα είναι πιθανό να έχουν διαφορετικές ιδιότητες. Επιπλέον, επειδή οι προσεγγίσεις συνεργατικού φιλτραρίσματος χρησιμοποιούν συγκεκριμένες πληροφορίες από αξιολογήσεις χρηστών βλέπουν πέρα από τις προτιμήσεις του ανεξάρτητου χρήστη και είναι ισχυρές στην πρόταση νέων αντικειμένων που ο χρήστης ίσως να μη γνωρίζει. Τα παραπάνω συνηγορούν στο να δημιουργηθούν στους χρήστες δυναμικές ανάγκες και ενδιαφέροντα που ίσως να μην είχαν στον παρελθόν.

Εντούτοις, ένα μειονέκτημα των προσεγγίσεων συνεργατικού φιλτραρίσματος είναι ότι εφαρμόζονται μόνο όταν τα προϊόντα που πρόκειται να προταθούν είναι τυποποιημένα, δηλαδή πωλούνται με την ίδια μορφή σε πολλούς χρήστες και ένας χρήστης είναι πιθανό να αγοράσει πολλά αντικείμενα του ίδιου τύπου μέσα από την ίδια ιστοσελίδα. Τα είδη του προϊόντος που διαθέτουν αυτά τα χαρακτηριστικά μπορεί να είναι λόγου χάρη CD, βιβλία ή ταινίες. Αλλά υπάρχουν και άλλα προϊόντα (π.χ. φωτογραφικές μηχανές, αυτοκίνητα, υπολογιστές) που δεν επιδέχονται συνεργατικές τεχνικές σύστασης.

Τα συστήματα αυτά εμφανίζουν ακόμη μερικές αδυναμίες. Για παράδειγμα, αν ένα νέο αντικείμενο προστεθεί στη βάση δεδομένων, τότε δεν θα συνίσταται μέχρι κάποιος χρήστης να το αξιολογήσει. Αυτό είναι το λεγόμενο πρόβλημα αργής εκκίνησης για το νέο αντικείμενο (cold start new item problem). Η αξιολόγηση ενός νέου αντικειμένου θα μπορούσε να αποδειχθεί δύσκολη υπόθεση, καθώς ο χρήστης μπορεί να μην είναι σε θέση να βρει το αντικείμενο ώστε να το αξιολογήσει. Παράλληλα, εάν υπάρχει ένας μεγάλος αριθμός αντικειμένων και ένας μικρός αριθμός χρηστών τότε πολλά αντικείμενα θα μπορούσαν να καταλήξουν να μην έχουν αξιολογηθεί. Επομένως, οι τεχνικές συνεργατικού φιλτραρίσματος απαιτούν ένα τεράστιο ποσό ανάδρασης χρήστη (user feedback) πριν από την παραγωγή ικανοποιητικών συστάσεων και δεν είναι σε θέση να λάβουν υπόψη τις εξαρτώμενες από τη συνεδρία ανάγκες των χρηστών. Παράλληλα, τα συστήματα συνεργατικού φιλτραρίσματος υποφέρουν και αυτά από το stability vs plasticity πρόβλημα. Οι χρήστες με ασυνήθιστες προτιμήσεις μπορεί να έχουν επίσης προβλήματα με αυτόν τον τύπο συστήματος, καθώς είναι δυνατόν να μην υπάρχει κανείς αρκετά κοντά στις δικές τους προτιμήσεις ώστε να δώσει καλές συστάσεις. Το πρόβλημα αυτό καλείται grey sheep problem [4].

Μερικές ακόμα προκλήσεις που πρέπει να αντιμετωπίσουν τα συστήματα συνεργατικού φιλτραρίσματος είναι:

Συνωνυμία (Synonymy): Η συνωνυμία αναφέρεται στην τάση ενός αριθμού από ίδια ή πολύ παρόμοια στοιχεία να έχουν διαφορετικά ονόματα ή καταχωρήσεις. Δεδομένου ότι σε κάποιους αλγορίθμους τα περιεχόμενα των στοιχείων αγνοούνται παντελώς, μια πιθανή συσχέτιση μεταξύ των στοιχείων δε λαμβάνεται υπόψιν για συστάσεις και τα

στοιχεία αντιμετωπίζονται με διαφορετικό τρόπο. Έτσι, εφ' όσον δεν έχουν αξιολογηθεί νέα στοιχεία από κάποιο χρήστη, δεν συνιστώνται.

Shilling επιθέσεις: Σε ένα σύστημα σύστασης όπου ο καθένας μπορεί να δώσει αξιολογήσεις, οι χρήστες μπορεί να δώσουν πολλές θετικές αξιολογήσεις για τα δικά τους στοιχεία και αρνητικές αξιολογήσεις για τους ανταγωνιστές τους. Είναι συχνά αναγκαίο για τα συνεργατικά συστήματα φιλτραρίσματος να εισαγάγουν προφυλάξεις για την αποθάρρυνση τέτοιου είδους χειρισμών.

Διατήρηση απορρήτου: Καθώς η συνεργατική σύσταση ενεργοποιείται από τη συλλογή προτιμήσεων και τις πληροφορίες συμπεριφοράς από πολλούς χρήστες, η διατήρηση της ιδιωτικότητας των δεδομένων χρήστη που συλλέγονται είναι μια σημαντική ανησυχία. Έτσι, είναι επιθυμητή η ανάπτυξη αποτελεσματικών τεχνικών χωρίς άμεση χρησιμοποίηση ευαίσθητων προσωπικών δεδομένων των χρηστών.

Τα συστήματα συνεργατικού φιλτραρίσματος χωρίζονται σε δύο κατηγορίες, τα συστήματα με **βάση τη μνήμη (memory-based)** και με **βάση το μοντέλο (model-based)** [4].

2.3.4 Συνεργατικό φιλτράρισμα με βάση τη μνήμη (memory-based collaborative filtering)

Οι memory-based αλγόριθμοι συνεργατικού φιλτραρίσματος διατηρούν στη μνήμη όλες τις γνωστές αξιολογήσεις/προτιμήσεις και τις χρησιμοποιούν για να βρουν ομοιότητες ανάμεσα σε χρήστες ή αντικείμενα. Ειδικότερα, οι αλγόριθμοι αυτοί περιλαμβάνουν ένα δισδιάστατο πίνακα χρηστών/αντικειμένων στον οποίο αποθηκεύονται οι αξιολογήσεις που κάνει κάθε χρήστης για κάθε αντικείμενο. Οι memory-based προσεγγίσεις μπορεί να χρησιμοποιούν ολόκληρο ή ένα δείγμα του πίνακα της αξιολόγησης χρήστη-αντικειμένου για να εξάγουν προβλέψεις αξιολογήσεων σε νέα προϊόντα. Χαρακτηριστικά παραδείγματα των συστημάτων με βάση τη μνήμη είναι οι top-N συστάσεις με βάση τον χρήστη ή το προϊόν.

Στην κατηγορία αλγορίθμων με βάση τη μνήμη υπάγονται οι προσεγγίσεις με βάση το χρήστη (user-based collaborative filtering) και οι προσεγγίσεις με βάση το αντικείμενο (item-based collaborative filtering). Η πρώτη περίπτωση βασίζεται στη γειτνίαση, όπου επιλέγεται ένα υποσύνολο χρηστών με βάση την ομοιότητά τους με τον ενεργό χρήστη, και μια συναθροιστική συνάρτηση των αξιολογήσεών τους, συνήθως ο σταθμισμένος μέσος όρος, χρησιμοποιείται για να παράγει προβλέψεις για αυτόν το χρήστη σχετικά με τα αντικείμενα. Οι περισσότερες από αυτές τις προσεγγίσεις μπορούν να γενικευτούν από τον αλγόριθμο που συνοψίζεται στα ακόλουθα βήματα:

1. Ανάθεσε ένα βάρος σε όλους τους χρήστες σχετικά με την ομοιότητα με τον ενεργό χρήστη.
2. Επέλεξε k χρήστες που έχουν τη μεγαλύτερη ομοιότητα με τον ενεργό χρήστη (γείτονες).
3. Υπολόγισε μία πρόβλεψη από ένα σταθμισμένο συνδυασμό των αξιολογήσεων των επιλεγμένων γειτόνων.

Υπάρχουν διάφορες μέθοδοι για να υπολογιστεί η ομοιότητα μεταξύ των χρηστών. Το πιο συχνά χρησιμοποιούμενο μέτρο της ομοιότητας είναι ο συντελεστής συσχετισμού του Pearson ανάμεσα στις αξιολογήσεις δύο χρηστών. Ένας άλλος τρόπος είναι η ομοιότητα με βάση το συνημιτονικό διάνυσμα (cosine-based similarity). Η εν λόγω μέθοδος χρησιμοποιείται αρχικά για να υπολογίσει την ομοιότητα μεταξύ δύο εγγράφων αντιμετωπίζοντας κάθε έγγραφο ως ένα διάνυσμα των συχνοτήτων εμφάνισης μιας

λέξης και υπολογίζοντας το συνημίτονο της γωνίας που σχηματίζεται από τα διανύσματα συχνότητας. Υπάρχουν επίσης πολλά άλλα μέτρα ομοιότητας που χρησιμοποιούνται στη βιβλιογραφία, όπως η συσχέτιση Spearman rank, η συσχέτιση του Kendall, οι μέσες τετραγωνικές διαφορές, η εντροπία και η προσαρμοσμένη ομοιότητα συνημιτόνου [4].

Ένα βασικό πλεονέκτημα των user-based μεθόδων είναι ότι μπορούν να συστήσουν στο χρήστη διαφορετικού τύπου προϊόντα από αυτά που αξιολογεί. Αυτό συμβαίνει διότι η πρόβλεψη στηρίζεται στις αξιολογήσεις άλλων χρηστών, οι οποίοι είναι όμοιοι με τον χρήστη-στόχο στα κοινά προϊόντα που έχουν βαθμολογήσει, αλλά οι γείτονες μπορεί να έχουν ενδιαφέροντα σε άλλους τομείς πέρα από τα κοινά προϊόντα, οι οποίοι θα προταθούν στον χρήστη-στόχο.

Από την άλλη πλευρά, η user-based προσέγγιση δεν κλιμακώνεται καλά όταν ισχύει για εκατομμύρια χρήστες και αντικείμενα, λόγω της υπολογιστικής πολυπλοκότητας της αναζήτησης παρόμοιων χρηστών. Επιπλέον, λόγω του μεγάλου αριθμού των ειδών και την απροθυμία των χρηστών να αξιολογήσουν τα στοιχεία, συνήθως ο πίνακας χρήστη-αντικειμένου είναι αραιός (sparsity problem). Συνεπώς, το σύστημα δεν μπορεί να παρέχει συστάσεις για ορισμένους χρήστες και οι συστάσεις που δημιουργούνται δεν είναι ακριβείς.

Στο item-based συνεργατικό φιλτράρισμα συσχετίζονται τα αντικείμενα που αξιολογούνται από ένα χρήστη με παρόμοια αντικείμενα, αντί να συσχετίζονται παρόμοιοι χρήστες. Στην προσέγγιση αυτή, οι ομοιότητες μεταξύ δύο αντικειμένων υπολογίζονται και πάλι χρησιμοποιώντας το συσχετισμό του Pearson ενώ η αξιολόγηση για ένα αντικείμενο από ένα συγκεκριμένο χρήστη μπορεί να προβλεφθεί χρησιμοποιώντας έναν απλό σταθμισμένο μέσο.

Το συνεργατικό φιλτράρισμα με βάση τα αντικείμενα είναι πιο κλιμακώσιμο από την προσέγγιση με βάση το χρήστη, καθώς οι συσχετίσεις καθορίζονται μεταξύ ενός περιορισμένου αριθμού προϊόντων, αντί για ένα δυνητικά πολύ μεγάλο αριθμό χρηστών. Τα αντικείμενα επίσης κατηγοριοποιούνται εύκολα ενώ οι ενέργειες των χρηστών πρέπει να εξετάζονται και να αναλύονται. Επίσης επειδή ο αριθμός των αντικειμένων είναι από τη φύση του μικρότερος από τον αριθμό των χρηστών, η προσέγγιση με βάση τα αντικείμενα έχει μικρότερο πρόβλημα αραιών δεδομένων σε σχέση με την προσέγγιση με βάση το χρήστη.

2.3.5 Συνεργατικό φιλτράρισμα με βάση το μοντέλο (model-based collaborative filtering)

Τα συστήματα σύστασης με βάση το μοντέλο συμπεριλαμβάνουν την κατασκευή ενός μοντέλου προβλέψεων με βάση το σύνολο των δεδομένων των αξιολογήσεων που έχουν αποθηκευτεί στο σύστημα. Με άλλα λόγια, οι πληροφορίες πρέπει να προέρχονται από το σύνολο δεδομένων και μπορεί να χρησιμοποιηθεί ένα “μοντέλο” για να προβεί το σύστημα σε συστάσεις, χωρίς να χρειάζεται να χρησιμοποιηθεί το πλήρες σύνολο δεδομένων κάθε φορά. Για την κατασκευή του μοντέλου, χρησιμοποιούνται τεχνικές μάθησης (learning techniques) όπως τα bayesian δίκτυα, τα νευρωνικά δίκτυα, τα μοντέλα συσταδοποίησης (clustering models), η κρυφή σημασιολογική ανάλυση (latent semantic analysis), η μέγιστη εντροπία, οι μηχανές Boltzmann, οι μηχανές διανυσμάτων υποστήριξης (Support Vector Machines) ή η αποσύνθεση μοναδικής τιμής (Singular Value Decomposition). Για να είναι το μοντέλο ακριβές, πρέπει να είναι διαθέσιμη μια μεγάλη ποσότητα δεδομένων.

Η προσέγγιση αυτή προσφέρει δυνητικά τα οφέλη της τόσο στην ταχύτητα όσο και στην κλιμάκωση καθώς το μοντέλο έχει δημιουργηθεί πριν την online διαδικασία σύστασης, οπότε έχει υψηλότερη απόδοση από τη βασισμένη στη μνήμη προσέγγιση. Ανάλογα με την τεχνική μάθησης που χρησιμοποιείται για τη δημιουργία του μοντέλου, αυτή η προσέγγιση μπορεί να οδηγήσει σε υψηλότερη ακρίβεια σύστασης ενώ περιορίζεται το πρόβλημα αραιών δεδομένων (sparsity problem) με αποτέλεσμα να εφαρμόζεται εύκολα σε εφαρμογές που περιλαμβάνουν μεγάλα σύνολα δεδομένων. Ένα άλλο πλεονέκτημα είναι ότι μπορεί να δώσει μια διαισθητική εξήγηση των συστάσεων που κάνει και έτσι οι χρήστες εμπιστεύονται καλύτερα το σύστημα και αποδέχονται τις προτάσεις που τους γίνονται.

Εντούτοις, το κυριότερο μειονέκτημα της προσέγγισης με βάση το μοντέλο είναι ότι τα αποτελέσματα σύστασης δεν προσαρμόζονται αυτόματα στις αλλαγές δεδομένων. Αντ' αυτού, το μοντέλο πρέπει να ξαναδημιουργηθεί για να αντανakλά τα ενημερωμένα δεδομένα και η διαδικασία αυτή κοστίζει ακριβά. Επίσης, επειδή στα συστήματα αυτά χρησιμοποιούνται συχνά μέθοδοι μείωσης διαστάσεων, μπορεί να χαθεί χρήσιμη πληροφορία από τις μειώσεις.

2.3.6 Φιλτράρισμα βασισμένο στη γνώση (knowledge-based filtering - KB)

Οι βασισμένες στο περιεχόμενο προσεγγίσεις αλλά και οι προσεγγίσεις συνεργατικού φιλτραρίσματος απαιτούν τη συσσώρευση του ιστορικού των επιλογών ή των προτιμήσεων προϊόντων που χτίζεται με το χρόνο. Πιθανότατα δεν είναι κατάλληλες για τη σύσταση των αντικειμένων που δεν αγοράζονται συχνά όπως τα προϊόντα/υπηρεσίες κινητής. Επίσης, έχουν προβλήματα αργής εκκίνησης. Ένας νέος χρήστης δε μπορεί να λάβει συστάσεις από τέτοιου είδους συστήματα. Ένα νέο στοιχείο δεν μπορεί να συστηθεί στα καθαρά συστήματα συνεργατικού φιλτραρίσματος. Η τεχνολογία σύστασης που βασίζεται στη γνώση αποφεύγει αυτά τα μειονεκτήματα, χρησιμοποιώντας τη γνώση σχετικά με τους χρήστες και τα προϊόντα και εξάγοντας διαπιστώσεις όσον αφορά ποια προϊόντα πληρούν τις απαιτήσεις του χρήστη. Η προσέγγιση αυτή, επομένως, εξαρτάται από συμπεράσματα για τις ανάγκες και τις προτιμήσεις του χρήστη και όχι από το ποσοστό εμπορεύματος των πελατειακών δεδομένων για να κάνει συστάσεις. Αυτή η τεχνολογία σύστασης χωρίζεται σε τρεις κατηγορίες: την ανακάλυψη γνώσεων (knowledge) σε βάσεις δεδομένων, την αιτιολογία βάσει περίπτωσης (case based reasoning – CBR) και το συμπερασμό γνώσης. Η προσέγγιση CBR, για παράδειγμα, είναι μια μεθοδολογία που αντιμετωπίζει ένα νέο πρόβλημα ανακτώντας μια παλαιότερη παρόμοια περίπτωση, που έχει ήδη επιλυθεί, και στη συνέχεια επαναχρησιμοποιεί αυτήν την περίπτωση για την επίλυση του σημερινού προβλήματος.

Το πλεονέκτημα της προσέγγισης που βασίζεται στη γνώση είναι ότι το σύστημα δεν χρειάζεται να αποθηκεύσει οποιαδήποτε πληροφορία σχετικά με τον χρήστη για μεγάλο χρονικό διάστημα. Έτσι, κάθε χρήστης που χρησιμοποιεί το σύστημα παρέχει σε αυτό την προτίμησή του και μπορεί να λάβει αμέσως τις συστάσεις του. Αν οι προτιμήσεις του πελάτη αλλάξουν είναι εύκολο να γίνει αναπροσαρμογή. Το μειονέκτημα είναι ότι το σύστημα δεν μαθαίνει από τις προτιμήσεις των χρηστών και κάθε φορά που κάποιος το χρησιμοποιεί θα πρέπει να δηλώσει την προτίμησή του. Προκειμένου να δοθούν σωστές συστάσεις απαιτείται το πεδίο γνώσης. Για παράδειγμα, εάν κάποιος επιλέξει ένα σύνολο προτιμήσεων για ένα προϊόν, το σύστημα πρέπει να καταλάβει ποια προϊόντα θα ταιριάζουν με τις προτιμήσεις καλύτερα, αυτό όμως απαιτεί κάποια μηχανική γνώση.

2.3.7 Φιλτράρισμα βασισμένο στα δημογραφικά στοιχεία (Demography-based filtering - DF)

Το φιλτράρισμα βάσει δημογραφικών στοιχείων (Demography-based filtering - DF) βασίζεται στις συστάσεις δημογραφικών τεχνικών. Στόχος αυτής της προσέγγισης είναι η κατηγοριοποίηση των χρηστών με βάση προσωπικά χαρακτηριστικά (ηλικία, φύλο, μορφωτικό επίπεδο, καταγωγή, κοινωνική τάξη κλπ.) και η δημιουργία αντίστοιχων συστάσεων. Οι δημογραφικές πληροφορίες μπορούν να χρησιμοποιηθούν για τον προσδιορισμό του είδους των χρηστών που τους αρέσουν παρόμοια αντικείμενα ή υπηρεσίες. Το βασικό στοιχείο του φιλτραρίσματος που βασίζεται στα δημογραφικά στοιχεία είναι ότι δημιουργεί κατηγορίες χρηστών που έχουν παρόμοια δημογραφικά χαρακτηριστικά και παρακολουθεί τη συνολική αγοραστική συμπεριφορά και τις προτιμήσεις των χρηστών σε αυτές τις κατηγορίες. Οι συστάσεις για ένα νέο χρήστη παράγονται βρίσκοντας αρχικά σε ποια κατηγορία ανήκει αυτός και στη συνέχεια εφαρμόζοντας τις συνολικές αγοραστικές προτιμήσεις των προηγούμενων χρηστών σε αυτήν την κατηγορία.

Το πλεονέκτημα της δημογραφικής προσέγγισης είναι ότι δε χρησιμοποιούνται οι αξιολογήσεις χρήστη-αντικειμένου και έτσι οι νέοι χρήστες μπορούν να πάρουν συστάσεις πριν αξιολογήσουν οποιοδήποτε αντικείμενο. Επίσης, δεν είναι απαραίτητες οι γνώσεις σχετικά με τα αντικείμενα και τα χαρακτηριστικά τους. Το μειονέκτημα της δημογραφικής προσέγγισης είναι ότι η συγκέντρωση των απαραίτητων δημογραφικών δεδομένων οδηγεί σε ζητήματα προστασίας της ιδιωτικότητας. Επιπρόσθετα, η δημογραφική κατηγοριοποίηση είναι πάρα πολύ αργή για εξατομικευμένες συστάσεις. Οι γενικεύσεις που δημιουργούνται από την κατηγοριοποίηση είναι συχνά ψευδείς, ειδικά όταν προέρχονται από αντικείμενα κουλτούρας όπως είναι τα βιβλία, η μουσική ή οι ταινίες. Η δημογραφική προσέγγιση υποφέρει και αυτή από το grey sheep πρόβλημα [4]. Πελάτες με διαφορετικές απόψεις ή με ασυνήθιστες προτιμήσεις καταλήγουν σε ένα χαμηλό συντελεστή συσχέτισης με άλλους πελάτες. Είναι, συνεπώς, πολύ δύσκολο να βρεθούν συστάσεις για αυτούς και προκαλούνται επίσης περίεργες συστάσεις για τους συσχετιζόμενους χρήστες. Τέλος, μια άλλη πρόκληση είναι η δυσκολία να αλλαχθεί ένα προφίλ χρήστη σε περίπτωση που αλλάξουν οι προτιμήσεις του (stability vs plasticity problem).

2.3.8 Υβριδικό Φιλτράρισμα (hybrid filtering)

Μια ξεχωριστή κατηγορία συστημάτων συστάσεων είναι η προσέγγιση υβριδικού φιλτραρίσματος (hybrid filtering) που αποτελεί ένα συνδυασμό των προηγούμενων προσεγγίσεων και έχει ως στόχο να ξεπεράσει τους περιορισμούς που αυτές εμφανίζουν. Ο συνδυασμός δύο ή περισσότερων τεχνικών σύστασης έχει ως αποτέλεσμα τη βελτιστοποίηση του τελικού συστήματος και την εμφάνιση σε αυτό λιγότερων από τις αδυναμίες που παρουσιάζει κάθε ξεχωριστό σύστημα. Για παράδειγμα, ο συνδυασμός του συνεργατικού φιλτραρίσματος με τη δημογραφική προσέγγιση αποτελεί μια λύση στο πρόβλημα της αργής εκκίνησης για το νέο χρήστη. Οι πιο δημοφιλείς υβριδικές προσεγγίσεις είναι αυτές που συνδυάζουν προσεγγίσεις με βάση το περιεχόμενο και προσεγγίσεις συνεργατικού φιλτραρίσματος. Οι τελευταίες μάλιστα λύνουν το πρόβλημα της αργής εκκίνησης για το νέο αντικείμενο. Υπάρχουν διάφορες στρατηγικές με τις οποίες η υβριδοποίηση μπορεί να επιτευχθεί και σε γενικές γραμμές κατατάσσονται σε επτά κατηγορίες [22]:

1. **Σταθμισμένη υβριδοποίηση (Weighted hybridization):** Οι αξιολογήσεις των διαφόρων τεχνικών σύστασης συνδυάζονται για να παράγουν μια μοναδική σύσταση.
2. **Μεταγωγή (Switching):** Το σύστημα εναλλάσσει τις τεχνικές σύστασης ανάλογα με την τρέχουσα κατάσταση.
3. **Μικτή υβριδοποίηση (Mixed hybridization):** Οι συστάσεις από αρκετά διαφορετικά συστήματα σύστασης παρουσιάζονται ταυτόχρονα.
4. **Συνδυασμός χαρακτηριστικών (Feature combination):** Τα χαρακτηριστικά από διαφορετικές πηγές δεδομένων σύστασης συνδυάζονται και χρησιμοποιούνται μαζί σε έναν ενιαίο αλγόριθμο σύστασης.
5. **Διαδοχική υβριδοποίηση (Cascade hybridization):** Η τεχνική αυτή περιλαμβάνει μια σταδιακή διαδικασία κατά την οποία πρώτα χρησιμοποιείται μια τεχνική σύστασης για να παραχθεί μια κατάταξη υποψηφίων συστάσεων και έπειτα μια δεύτερη τεχνική βελτιώνει τις συστάσεις που δίνονται από την πρώτη τεχνική.
6. **Αύξηση χαρακτηριστικών (Feature augmentation):** Μια τεχνική σύστασης χρησιμοποιείται για να υπολογιστεί ένα χαρακτηριστικό (ή ένα σύνολο χαρακτηριστικών) το οποίο στη συνέχεια χρησιμοποιείται ως είσοδος σε μια άλλη τεχνική.
7. **Υβριδοποίηση μετα-επιπέδου (Meta-level hybridization):** Μια τεχνική σύστασης παράγει ένα μοντέλο το οποίο έπειτα χρησιμοποιείται ως είσοδος για μια άλλη τεχνική.

2.4 Δημιουργία συστάσεων με χρήση του Κοινωνικού Ιστού

Όπως είδαμε τα συστήματα συστάσεων είναι στην πλειοψηφία τους εξατομικευμένα συστήματα και παρουσιάζουν την πληροφορία με τέτοιο τρόπο ώστε οι χρήστες να βλέπουν τα αντικείμενα που σχετίζονται περισσότερο με αυτούς. Το συνεργατικό Φιλτράρισμα είναι η κυρίαρχη τεχνική [16] για συστήματα συστάσεων. Ωστόσο, παρατηρεί συσχετίσεις μεταξύ των χρηστών, χωρίς να προσπαθεί να κατανοήσει για ποιο λόγο έγιναν αυτοί οι συσχετισμοί, δηλαδή δεν είναι σε θέση να διακρίνει ποιοι από τους λεγόμενους γείτονες του χρήστη είναι φίλοι του ή άγνωστοι ομότιμοι με παρόμοιες προτιμήσεις. Αυτό δεν αξιοποιεί πληροφορίες πλαισίου που θα μπορούσαν, για παράδειγμα, να εξηγήσουν γιατί ένας χρήστης μπορεί να εμφανίζει παρόμοια χαρακτηριστικά με έναν άλλο χρήστη σε ορισμένες μόνο περιπτώσεις ενώ σε άλλες να παρουσιάζει διαφορές.

Προηγούμενη ερευνητική δραστηριότητα [17] έχει δείξει ότι οι άνθρωποι έχουν την τάση να τους αρέσουν στοιχεία που αρέσουν και στους φίλους τους και έλκονται από τις δραστηριότητες των άλλων στον κοινωνικό τους περίγυρο. Οπότε, θα μπορούσαν να θεωρηθούν ως πιο αξιόπιστες πηγές σύστασης οι γείτονες που όχι μόνο εμφανίζουν παρόμοια στοιχεία με τον εν λόγω χρήστη αλλά συνδέονται με αυτόν με κάποιους κοινωνικούς δεσμούς.

Για να κατανοηθούν καλύτερα η ομοιότητα συμπεριφορών και η κοινωνική επιρροή από φίλους, πρέπει να αναπτυχθούν τεχνικές εξόρυξης δεδομένων που λαμβάνουν πληροφορίες ύπαρξης κοινωνικών δεσμών μαζί με τα δεδομένα συμπεριφοράς του χρήστη. Η ποσοτική μέτρηση της ομοιότητας στη συμπεριφορά και της κοινωνικής επιρροής μεταξύ φίλων μπορεί να παρέχει τη θεωρητική βάση για την καταγραφή των κοινωνικών συμπεριφορών στη διαδικασία λήψης αποφάσεων και ως εκ τούτου είναι

πολύ χρήσιμη για συνεργατικές συστάσεις. Κατά τις πρώτες ημέρες του Διαδικτύου, ο εντοπισμός των στενών φίλων ενός χρήστη ήταν δύσκολος [18]. Πλέον όμως η δημιουργία ιστοτόπων κοινωνικής δικτύωσης καθιστά εύκολη την ενσωμάτωση πληροφοριών κοινωνικού δικτύου κατά τη δημιουργία συστάσεων.

Οι ιστοχώροι κοινωνικής δικτύωσης είναι online κοινότητες (άτομα, οργανώσεις ή άλλες κοινωνικές οντότητες) [19] που συνδέονται με ένα σύνολο κοινωνικών σχέσεων, όπως η φιλία, η συνεργασία ή η ανταλλαγή πληροφοριών σε ποικίλα περιβάλλοντα, λόγου χάρη, όσον αφορά την ψυχαγωγία, τη θρησκεία, τις γνωριμίες, ή τις επιχειρήσεις. Οι χρήστες στα κοινωνικά δίκτυα δημιουργούν φωτογραφίες, video, και άλλα αντικείμενα, συνεργάζονται με άλλους χρήστες, κοινωνικοποιούνται με τους φίλους τους και μοιράζονται τις απόψεις τους online. Υπάρχουν και άλλοι τρόποι δημιουργίας κοινωνικών ομάδων στο Διαδίκτυο μεταξύ των οποίων περιλαμβάνονται ιστοσελίδες που υποστηρίζουν φόρουμ (Internet Forums), κοινωνικά blogs, microblogging, wikis, podcasts και social bookmarking. Όλες αυτές οι τεχνολογίες που υποστηρίζουν την επικοινωνία των κοινωνικών ομάδων αποτελούν τα κοινωνικά μέσα ενημέρωσης (Social Media). Τα κοινωνικά μέσα ενημέρωσης απολαμβάνουν μεγάλη επιτυχία τα τελευταία χρόνια με εκατομμύρια χρήστες να επισκέπτονται ιστοσελίδες όπως το Facebook για κοινωνική δικτύωση, το Wordpress για blogging, το Twitter για micro-blogging, το Flickr και το YouTube για το διαμοιρασμό φωτογραφιών και βίντεο, αντίστοιχα, το Digg για την ανάγνωση κοινωνικών ειδήσεων και το Delicious για κοινωνικό bookmarking.

Η εμφάνιση του Twitter και του Facebook σηματοδότησε μια νέα φάση στην εξέλιξη του Διαδικτύου, το οποίο χαρακτηρίζεται από ροές ενημερώσεων (updates) και ειδήσεων. Εκατομμύρια χρήστες οι οποίοι μοιράζονται τις δραστηριότητές τους με τους φίλους και τους οπαδούς τους (followers) καταλήγουν να αντανakλούν καυτά θέματα και τάσεις και επιτρέπουν τη διάδοση ειδήσεων σε πραγματικό χρόνο. Αυτή η εξέλιξη συχνά αναφέρεται ως ιστός πραγματικού χρόνου (real-time web) [20].

Το Twitter επιτρέπει στους χρήστες να δημοσιεύσουν σύντομα μηνύματα (που συχνά αποκαλούνται "ενημερώσεις κατάστασης" ή "tweets"), περιγράφοντας τις δραστηριότητες και τις απόψεις τους ή υποδεικνύοντας ενδιαφέρον περιεχόμενο. Η ροή Twitter είναι ομοιογενής, με την έννοια ότι αποτελείται αποκλειστικά από ενημερώσεις κατάστασης (status updates). Αντίθετα, οι χαρακτηριστικές ροές δραστηριότητας είναι ετερογενείς και δημοσιεύουν διάφορα είδη ειδήσεων και δραστηριοτήτων μέσα σε ένα δικτυακό τόπο κοινωνικής δικτύωσης (SNS). Η πιο γνωστή ροή δραστηριότητας είναι οι Ενημερώσεις (News Feed) του Facebook, της οποίας η εισαγωγή σηματοδότησε μια μεγάλη αλλαγή στον ιστοχώρο. Οι ενημερώσεις καταλαμβάνουν το κεντρικό τμήμα της αρχικής σελίδας κάθε χρήστη του Facebook, δείχνοντας τις πρόσφατες δραστηριότητες των φίλων, συμπεριλαμβανομένων των ενημερώσεων κατάστασης (status updates), τις προσθήκες φίλων, τη συμμετοχή σε ομάδα (join group), τη σελίδα "προτίμησης" (page liking), τις αλλαγές στο προφίλ, την κοινή χρήση φωτογραφιών ή ετικετοποίηση (tagging), και άλλα. Μετά το Facebook, άλλα κορυφαία site κοινωνικής δικτύωσης, όπως το MySpace και το LinkedIn έχουν προσθέσει τις δικές τους ροές δραστηριότητας. Επιπλέον, οι υπηρεσίες τρίτων όπως το FriendFeed άρχισαν να συγκεντρώνουν τις ροές δραστηριότητας από τους διάφορους ιστοχώρους κοινωνικών μέσων ενημέρωσης. Πρόσφατα έχει δημιουργηθεί η μορφή των ροών δραστηριότητας για τη δημοσίευση των κοινωνικών δραστηριοτήτων γύρω από το Διαδίκτυο, το οποίο έχει υιοθετηθεί από κορυφαίους χώρους, όπως το Facebook και το MySpace.

Εν κατακλείδι, ο κοινωνικός ιστός (social Web) έχει προσθέσει νέες διαστάσεις στον τρόπο με τον οποίο οι άνθρωποι αλληλεπιδρούν στο Διαδίκτυο, δίνοντας έμφαση στο παραγόμενο περιεχόμενο από τον ίδιο το χρήστη. Ο πολύπλοκος χώρος των πληροφοριών που παράγεται από τον κοινωνικό ιστό προσφέρει εκτός των άλλων ένα

πλούσιο και δυναμικό περιβάλλον για να ανακαλύψουν οι χρήστες νέο περιεχόμενο. Λόγω αυτής της ραγδαίας ανάπτυξης του περιεχομένου, υπάρχει και σε αυτήν την περίπτωση η ανάγκη για τη δημιουργία εξατομικευμένων συστημάτων συστάσεων για την αντιμετώπιση του πλεονάσματος πληροφορίας και τη σύσταση μόνο χρήσιμου περιεχομένου σε χρήστες. Όπως και στο ηλεκτρονικό εμπόριο για τη σύσταση αντικειμένων στο χρήστη (π.χ. ταινίες, βιβλία, ξενοδοχεία κ.λπ.), ο στόχος ενός εξατομικευμένου συστήματος συστάσεων είναι η προσαρμογή του περιεχομένου που βασίζεται στα χαρακτηριστικά των μεμονωμένων χρηστών. Κοινωνικά μέσα ενημέρωσης και εξατομικευμένα συστήματα συστάσεων μπορούν να επωφεληθούν το ένα από το άλλο: από τη μία πλευρά, τα κοινωνικά μέσα ενημέρωσης εισάγουν νέους τύπους δημόσιων δεδομένων και μεταδεδομένων, όπως αξιολογήσεις, ετικέτες (tags), σχόλια, και άμεσες κοινωνικές σχέσεις των ανθρώπων, τα οποία μπορούν να χρησιμοποιηθούν για να ενισχυθούν οι συστάσεις που παράγονται από ένα σύστημα. Από την άλλη πλευρά, οι τεχνολογίες συστάσεων μπορούν να διαδραματίσουν σημαντικό ρόλο στην επιτυχία των εφαρμογών κοινωνικής δικτύωσης και του κοινωνικού ιστού ως σύνολο, εξασφαλίζοντας ότι κάθε χρήστης παρουσιάζεται με το πιο ελκυστικό και σχετικό περιεχόμενο, σε προσωπικό επίπεδο.

2.4.1 Βασικές οντότητες του Κοινωνικού Ιστού

Μέχρι στιγμής αναφέρθηκε ότι τα ΣΣ εκτελούν τη διαδικασία σύστασης έχοντας σαν είσοδο τα δεδομένα χρήστη, τα δεδομένα αντικειμένου και τις αξιολογήσεις των χρηστών. Στην περίπτωση των κοινωνικών δικτύων, τα δεδομένα αυτά εμπλουτίζονται με επιπλέον στοιχεία τα οποία διαμορφώνονται από τα ιδιαίτερα χαρακτηριστικά του κοινωνικού ιστού. Ο κοινωνικός ιστός μπορεί να θεωρηθεί ως η συλλογή των online αλληλεπιδράσεων μεταξύ των χρηστών οι οποίες μάλιστα μπορούν να πάρουν πολλές μορφές. Μερικές αλληλεπιδράσεις είναι μεταξύ των ατόμων ενώ άλλες περιστρέφονται γύρω από το περιεχόμενο ή το πώς το περιεχόμενο αξιολογείται. Έτσι, στα κοινωνικά δίκτυα διακρίνονται και πάλι οι χρήστες, τα αντικείμενα και οι αξιολογήσεις ως οι τρεις βασικές οντότητες.

Ο κοινωνικός ιστός μπορεί να θεωρηθεί ως ένας γράφος, στον οποίο οι χρήστες συνδέονται μεταξύ τους, είτε άμεσα μέσω κοινωνικών συνδέσμων (links) διαφόρων ειδών είτε έμμεσα μέσω συνδέσεων στο περιεχόμενο. Οι χρήστες μπορεί να σχηματίζουν ομάδες ανθρώπων με "ισχυρούς" δεσμούς μεταξύ των μελών της κάθε ομάδας, και "ασθενείς" δεσμούς που συνδέουν τους ανθρώπους σε ομάδες. Ενώ οι ισχυροί δεσμοί συνήθως αποτελούνται από τους στενούς φίλους, οι ασθενείς δεσμοί αποτελούνται από γνωστούς ή συναδέλφους εξ αποστάσεως. Η παραπάνω λογική συνιστά την υπόθεση της δύναμης-των-αδύναμων-δεσμών (strength-of-weak-ties) στα κοινωνικά δίκτυα [21] σύμφωνα με την οποία οι ασθενείς δεσμοί είναι χρήσιμοι για τη διάδοση των πληροφοριών, την επιρροή και την οικονομική κινητικότητα, επειδή οι ασθενείς δεσμοί βοηθούν να συνδεθούν διαφορετικές ομάδες ανθρώπων μεταξύ τους. Επίσης, στην υπόθεση αυτή υποβόσκει το φαινόμενο της ομοφιλίας (homophily) το οποίο αναφέρει ότι οι άνθρωποι με παρόμοια γνωρίσματα όσον αφορά την ηλικία, την εθνικότητα, τη γεωγραφική θέση, την οικονομική κατάσταση κλπ, έχουν την τάση να ομαδοποιούνται. Η ομοφιλία οδηγεί σε ισχυρούς δεσμούς μεταξύ των ανθρώπων με παρόμοια γνωρίσματα και οι ασθενείς δεσμοί συνδέουν διαφορετικές ομάδες μαζί.

Οι αλληλεπιδράσεις των χρηστών μπορεί να είναι είτε μονόδρομες είτε αμφίδρομες. Ένα παράδειγμα μονόδρομης αλληλεπίδρασης είναι μία συνδρομή στο Twitter, όπου ο χρήστης που ακολουθείται δεν έχει τον έλεγχο για το ποιος είναι ο follower. Αντίστοιχα, ένα παράδειγμα αμφίδρομης αλληλεπίδρασης αποτελεί η δημοφιλής σχέση φιλίας

(“friend” relationship) στο Facebook, όπου συστήνονται ως φίλοι άτομα με επικαλυπτόμενους κοινωνικούς κύκλους. Μάλιστα, η δημιουργία κοινωνικών συνδέσεων αποτελεί μία βασική λειτουργία σύστασης στον κοινωνικό ιστό. Τέλος, οι αλληλεπιδράσεις μεταξύ των χρηστών μπορεί να περιλαμβάνουν ομάδες διαφορετικού μεγέθους. Μπορεί λόγω χάρη ένας χρήστης να δημοσιεύει περιεχόμενο σε μία ομάδα στενών μόνο φίλων ή ακόμα και σε ολόκληρο τον κόσμο.

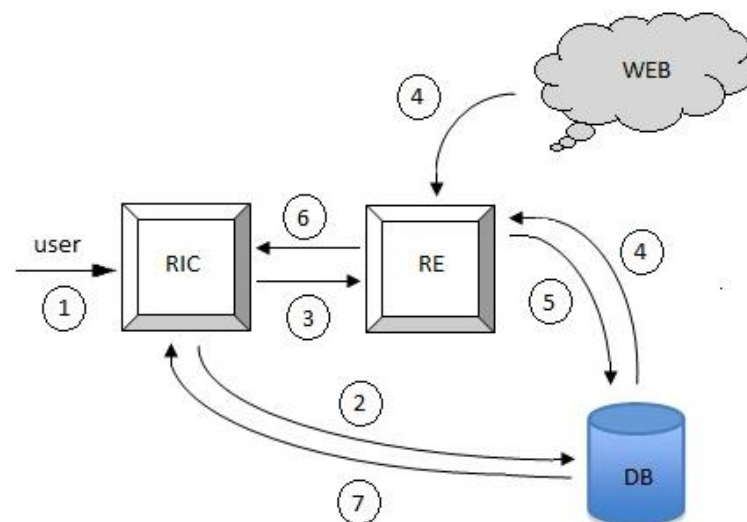
Ο κοινωνικός ιστός παρουσιάζει έναν αχανή πίνακα αντικειμένων που υπόκεινται σε σύσταση. Ένας λόγος που οι άνθρωποι έρχονται στους κοινωνικούς ιστοχώρους είναι για να πληροφορηθούν σχετικά με νέα αντικείμενα που παρουσιάζουν τα ΣΣ, όπως ταινίες, μουσική, εστιατόρια κ.λπ. Ωστόσο, στον κοινωνικό ιστό, θα υπάρχουν συχνά πολλές διαδρομές για την εύρεση ενός συγκεκριμένου στοιχείου. Παραδείγματος χάριν, ένα μουσικό κομμάτι μπορεί να προσπελαστεί μέσω του καλλιτέχνη, μέσω μιας ετικέτας (tag), μέσα από ένα άλμπουμ, μέσα από λίστα αγαπημένων ενός φίλου και άλλα.

Οι χρήστες στις κοινωνικές ιστοσελίδες μπορούν να εκφράζουν τις απόψεις τους για τα περιεχόμενα του ιστοχώρου με ποικίλους τρόπους. Μπορούν να βάλουν ετικέτες (tag), να κάνουν δημοσιεύσεις (posts), tweet ή like, να αξιολογήσουν ή να γράψουν σχόλια. Αυτές οι αλληλεπιδράσεις μπορούν να διαβαστούν, να απαντηθούν ή να προωθηθούν σε άλλους χρήστες. Στις περισσότερες περιπτώσεις, οι αξιολογήσεις έχουν μορφή κειμένου και η απόσπαση πληροφορίας από αυτές απαιτεί ανάλυση του περιεχομένου. Λόγου χάρη, η ετικέτα είναι μια λέξη-κλειδί που συνδέεται ή αποδίδεται σε κομμάτια πληροφορίας όπως εικόνες, άρθρα, αρχεία βίντεο κτλ. Μία ετικέτα μπορεί να θεωρηθεί ως μία εκτίμηση που κάνει ο χρήστης για ένα αντικείμενο. Επιπρόσθετα, ένας χρήστης συχνά εφαρμόζει πολλές ετικέτες όταν σχολιάζει ένα αντικείμενο. Επειδή οι ετικέτες δημιουργούνται από ένα πλήθος ατόμων για πολλούς λόγους χωρίς να υπάρχει κάποιο προκαθορισμένο λεξιλόγιο, θα πρέπει αυτές να ερμηνεύονται σε σχέση με τα άτομα που τις δημιούργησαν. Για παράδειγμα, για ένα χρήστη που του αρέσουν τα αυτοκίνητα η ετικέτα “jaguar” πιθανώς υπονοεί τη γνωστή μάρκα αυτοκινήτων Jaguar, ενώ το “jaguar” για έναν άλλο χρήστη ίσως είναι μία ετικέτα που χρησιμοποιείται για τα αιλουροειδή. Αντί, λοιπόν, να υπάρχει μια μονοδιάστατη περιγραφή που αντιπροσωπεύει το περιεχόμενο ενός αντικειμένου, εμφανίζεται ένα σύνολο εννοιών, προσφέροντας καθεμιά τη σκοπιά ενός διαφορετικού χρήστη.

Εν τέλει, τα κοινωνικά μέσα στο σύνολό τους χρησιμοποιούν συστήματα συστάσεων προκειμένου να αυξήσουν τη δημοτικότητά τους και ανακυκλώνουν αποκλειστικά δικό τους περιεχόμενο το οποίο το φιλτράρουν και το παρέχουν στους δικούς τους χρήστες. Λίγη προσπάθεια έχει γίνει μέχρι στιγμής όσον αφορά την απόσπαση πληροφοριών πλαισίου ιδιαίτερα από τα κοινωνικά δίκτυα οι οποίες θα μπορούσαν να χρησιμοποιηθούν ως είσοδος σε ένα σύστημα συστάσεων άλλου τομέα, λόγω χάρη ηλεκτρονικού εμπορίου. Στα επόμενα κεφάλαια θα παρουσιάσουμε ένα σύστημα ηλεκτρονικού εμπορίου που θα εκμεταλλεύεται τις ιδιαιτερότητες του κοινωνικού ιστού με σκοπό τη δημιουργία ακόμη πιο εξατομικευμένων συστάσεων.

3. ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΣΥΣΤΗΜΑΤΟΣ

Στο κεφάλαιο αυτό θα γίνει μία αναλυτική περιγραφή του ΣΣ που έχει υλοποιηθεί στα πλαίσια της εργασίας. Το σύστημα αυτό απαρτίζεται από μία διεπαφή σύστασης (Recommender Interface Component - RIC), μία μηχανή συστάσεων (Recommender Engine - RE) και μία βάση δεδομένων (Database – DB). Για τη λειτουργία του συστήματος θεωρούμε μια εικονική αγορά από προϊόντα όπου ένα υποσύνολο αυτών θα συσταθεί στους χρήστες του συστήματος. Η εικονική αγορά αποτελείται αποκλειστικά από βιβλία, μουσικά άλμπουμ και ταινίες. Οι χρήστες εγγράφονται/συνδέονται στο σύστημα και καταχωρούν πληροφορίες ενώ παράλληλα διαθέτουν ένα λογαριασμό σε κάποιο κοινωνικό δίκτυο. Εν συνεχεία, εκτελείται ένας αλγόριθμος σύστασης ο οποίος συνδυάζει διαθέσιμες πληροφορίες από το σύστημα και το λογαριασμό στο κοινωνικό δίκτυο και παράγει αποτελέσματα για προτεινόμενα προϊόντα στο χρήστη. Τα προϊόντα εμφανίζονται στη οθόνη με αναγραφή όλων των χαρακτηριστικών τους. Σημειώνεται ότι το σύστημα έχει τη μορφή ενός απλοποιημένου ηλεκτρονικού καταστήματος (e-shop) και συνεπώς εμφανίζονται μαζί με τα χαρακτηριστικά των προϊόντων πληροφορίες σχετικά με την τιμή τους, τον προμηθευτή και τα έξοδα αποστολής τους. Δεν υπάρχει ωστόσο κάποια λειτουργία online αγοράς ή αναζήτησης κάποιου προϊόντος. Ολόκληρη η διαδικασία που ακολουθείται κατά τη λειτουργία του συστήματος συνοψίζεται στο Σχήμα 1:



Σχήμα 1: Λειτουργία του συστήματος

Όπως φαίνεται στο Σχήμα 1, αρχικά ο χρήστης συνδέεται στο RIC και δίνει πληροφορίες οι οποίες καταχωρούνται στη βάση δεδομένων. Ο χρήστης στη συνέχεια επιλέγει αλγόριθμο και το RIC καλεί τη RE. Η RE συλλέγει πληροφορίες από τη βάση δεδομένων και τα κοινωνικά δίκτυα στο Web. Στη συνέχεια, καταχωρεί τα αποτελέσματα σύστασης στη βάση δεδομένων και ανακατευθύνεται πίσω στο RIC προκειμένου να εμφανιστούν τα αποτελέσματα στην οθόνη. Τέλος, το RIC διαβάζει τα αποτελέσματα από τη βάση δεδομένων και τα εμφανίζει στο χρήστη. Ο κώδικας που υλοποιεί το παραπάνω σύστημα έχει δημιουργηθεί στην πλατφόρμα Netbeans 7.2.1 με server τον Apache Tomcat.

3.1 Διεπαφή σύστασης (Recommender Interface Component - RIC)

Το υποσύστημα αυτό αποτελεί τον ιστοχώρο (website) όπου πραγματώνονται οι online συστάσεις και είναι υλοποιημένο σε JSP. Προσφέρει μια γραφική διεπαφή, φιλική στο χρήστη, η οποία υλοποιεί όλες τις απαραίτητες ενέργειες που μπορεί να κάνει εκείνος. Συνεπώς, προσφέρει τρόπους για την εισαγωγή δεδομένων χρήστη μέσα από φόρμες καταχώρησης στοιχείων. Επίσης, είναι υπεύθυνο για την προώθηση των αποτελεσμάτων στο χρήστη – στόχο έπειτα από την εκτέλεση του αλγορίθμου σύστασης. Ειδικότερα οι λειτουργίες που επιτελούνται περιγράφονται ακολούθως:

Αρχική σελίδα εισόδου: Αυτή αποτελεί την πρώτη οθόνη του συστήματος (βλ. Εικόνα 1) όπου παρέχονται η δυνατότητα σύνδεσης των ήδη εγγεγραμμένων χρηστών αλλά και η επιλογή εγγραφής των νέων χρηστών. Επιπρόσθετα, παρέχεται έλεγχος λογαριασμού σε περίπτωση που ο χρήστης δώσει λανθασμένα ή ψευδή στοιχεία, εμφανίζοντας μήνυμα αναφοράς σφάλματος στην οθόνη.

Καλώς ήρθατε στο σύστημα συστάσεων

Όνομα χρήστη:

Κωδικός Πρόσβασης:

Διατήρησε τη σύνδεση μου

Σύνδεση

Νέος χρήστης: [Εγγραφή εδώ](#)

Εικόνα 1: Αρχική σελίδα εισόδου

Εγγραφή: Η λειτουργία εγγραφής παρέχει μία φόρμα συμπλήρωσης προσωπικών δεδομένων (βλ. Εικόνα 2) μεταξύ των οποίων περιλαμβάνονται το ονοματεπώνυμο του χρήστη, η ημερομηνία γέννησης, το φύλο, η περιοχή, το επάγγελμα, η ηλεκτρονική διεύθυνση (email) και στοιχεία λογαριασμού όπως όνομα χρήστη (username), κωδικό πρόσβασης (password) και επιβεβαίωση κωδικού. Η εγγραφή ολοκληρώνεται μόνο εάν έχουν πραγματοποιηθεί επιτυχώς έλεγχοι που αφορούν τα στοιχεία που καταχωρούνται. Ειδικότερα, ο χρήστης θα πρέπει να έχει συμπληρώσει υποχρεωτικά όλα τα πεδία της φόρμας, να έχει δώσει μια λογική ημερομηνία γέννησης, να έχει μοναδικό όνομα χρήστη και ο κωδικός πρόσβασής του να είναι τουλάχιστον έξι χαρακτήρες και να συμπίπτει με την επιβεβαίωση του κωδικού του. Η μη επιτυχής ολοκλήρωση της εγγραφής εμφανίζει μήνυμα λάθους στην οθόνη που διευκρινίζει τι θα πρέπει να διορθωθεί.

Δημιουργία νέου λογαριασμού

Όνομα:	<input type="text"/>
Επώνυμο:	<input type="text"/>
Ημερομηνία γέννησης:	Ημέρα <input type="text"/> Μήνας <input type="text"/> Έτος <input type="text"/>
Φύλο:	<input type="radio"/> Άντρας <input type="radio"/> Γυναίκα
Περιοχή:	Περιοχή <input type="text"/>
Επάγγελμα:	Επάγγελμα <input type="text"/>
Email:	<input type="text"/>
Όνομα χρήστη:	<input type="text"/>
Κωδικός πρόσβασης:	<input type="text"/>
Επαλήθευση κωδικού:	<input type="text"/>
<input type="button" value="Αποδοχή"/> <input type="button" value="Ακύρωση"/>	

Εικόνα 2: Φόρμα καταχώρησης στοιχείων

Η συμπλήρωση της φόρμας για την καταχώρηση στοιχείων ακολουθείται από την εισαγωγή προτιμήσεων του χρήστη (βλ. Εικόνα 3). Οι προτιμήσεις αυτές αφορούν χαρακτηριστικά των προϊόντων που υπάρχουν στη βάση δεδομένων του συστήματος. Αρχικά, ο χρήστης επιλέγει την κατηγορία προϊόντος για την οποία επιθυμεί να εισάγει τις προτιμήσεις του. Η επιλογή «Τέλος» σηματοδοτεί τη λήξη της διαδικασίας εγγραφής.

Επιλέξτε μία από τις παρακάτω κατηγορίες:

- Βιβλία
- Ταινίες
- Μουσική

Εικόνα 3: Επιλογή κατηγορίας προϊόντος

Εν συνεχεία ο χρήστης έχει τη δυνατότητα να επιλέξει μέσα από μία λίστα χαρακτηριστικών που του παρέχονται χωρίς να υπάρχει περιορισμός στον αριθμό των επιλογών του (βλ. Εικόνες 4-6).

Δήλωση Προτιμήσεων

Κατηγορία: Βιβλία

1. **Τι είδη βιβλίων σας αρέσουν;**

- Μυθιστόρημα
- Βιογραφία
- Επιστημονικό
- Ιστορικό
- Λογοτεχνικό
- Φαντασίας
- Ψυχολογικό
- Παιδικό
- κόμικ
- Μαγειρικής
- Θρησκευτικό
- Εγκυκλοπαίδειες
- Ιατρικό
- Ποιητική συλλογή

2. **Ποιοι συγγραφείς σας αρέσουν;**

Charles Dickens (Καρλ Ντίκενς)

Edgar Allan Poe ^

Frances Hodgson Burnett

J.K Rowling v

Jorge Bucay (Χόρχε Μπουκάι)

3. **Ποιες εκδόσεις προτιμάτε;**

Άγκυρα ^

Κάκτος

Καστανιώτη

Κέδρος v

Λιβάνη

Εικόνα 4: Καταχώρηση προτιμήσεων για βιβλία

Δήλωση Προτιμήσεων

Κατηγορία: Μουσική

1. **Τι μουσική ακούτε;**

- Rock
- Pop
- Jazz
- Κλασική
- Hip Hop
- Reggae
- Rap
- Metal
- Latin
- Ethnic
- House
- Blues
- R&B
- Soul
- Alternative
- Country
- Έντεχνη

2. **Ποιοι ξένοι τραγουδιστές/συγκροτήματα σας αρέσουν;**

Adele

Black Sabbath ^

Blind Guardian

Bob Marley v

Bob Dylan

3. **Ποιοι Έλληνες τραγουδιστές/συγκροτήματα σας αρέσουν;**

Άννα Βίση

Γιάννης Κότσιρας ^

Δέσποινα Βανδή

Έλενα Παπαρίζου (Helena Paparizou) v

Μαρία Κάλλας

Εικόνα 5: Καταχώρηση προτιμήσεων για μουσική

Δήλωση Προτιμήσεων

Κατηγορία: Ταινίες

1. **Τι είδη ταινιών σας αρέσουν;**

- Κωμωδία
- Αισθηματική
- Δράμα
- Θρίλερ
- Περιπέτεια
- Φαντασίας
- Επιστημονικής Φαντασίας
- Μιούζικαλ
- Βιογραφία
- Ντοκιμαντέρ
- Κινούμενα σχέδια

2. **Ποιοι σκηνοθέτες σας αρέσουν;**

Catherine Hardwicke (Κάθριν Χάρντγουικ) ▲

Darren Aronofsky

Garry Marshall

James Cameron (Τζέιμς Κάμερον) ▼

Martin Scorsese (Μάρτιν Σκορσέζε)

3. **Ποιους ηθοποιούς προτιμάτε;**

Leonardo di Caprio ▲

Kate Winslet

Γιώργος Χωραφάς

Christopher Lee ▼

Owen Wilson

Εικόνα 6: Καταχώρηση προτιμήσεων για ταινίες





Με το πέρας της εγγραφής, εκκινείται μία συνεδρία, ο χρήστης αποκτά ένα μοναδικό αναγνωριστικό και παραπέμπεται στη σελίδα συνδεδεμένου χρήστη.

Σελίδα συνδεδεμένου χρήστη: Αυτή αποτελεί μία εξατομικευμένη σελίδα (βλ. Εικόνα 7) στην οποία παρουσιάζονται τα αποτελέσματα του αλγορίθμου σύστασης. Διαθέτει φίλτρα αποτελεσμάτων μεταξύ των οποίων περιλαμβάνονται η δυνατότητα επιλογής συγκεκριμένης κατηγορίας προϊόντων, η δυνατότητα καταχώρησης συγκεκριμένου εύρους τιμών και τέλος η δυνατότητα επιλογής αλγορίθμου σύστασης. Σημειώνεται ότι η γραφική διεπαφή είναι κατασκευασμένη ώστε να περιλαμβάνει και τη χρήση άλλων γνωστών αλγορίθμων σύστασης, που δεν έχουν υλοποιηθεί στα πλαίσια της διπλωματικής μας εργασίας καθώς δεν αποτελούν αντικείμενο της. Τέλος, παρέχεται μενού επιλογών για τις υπόλοιπες λειτουργίες του συστήματος.



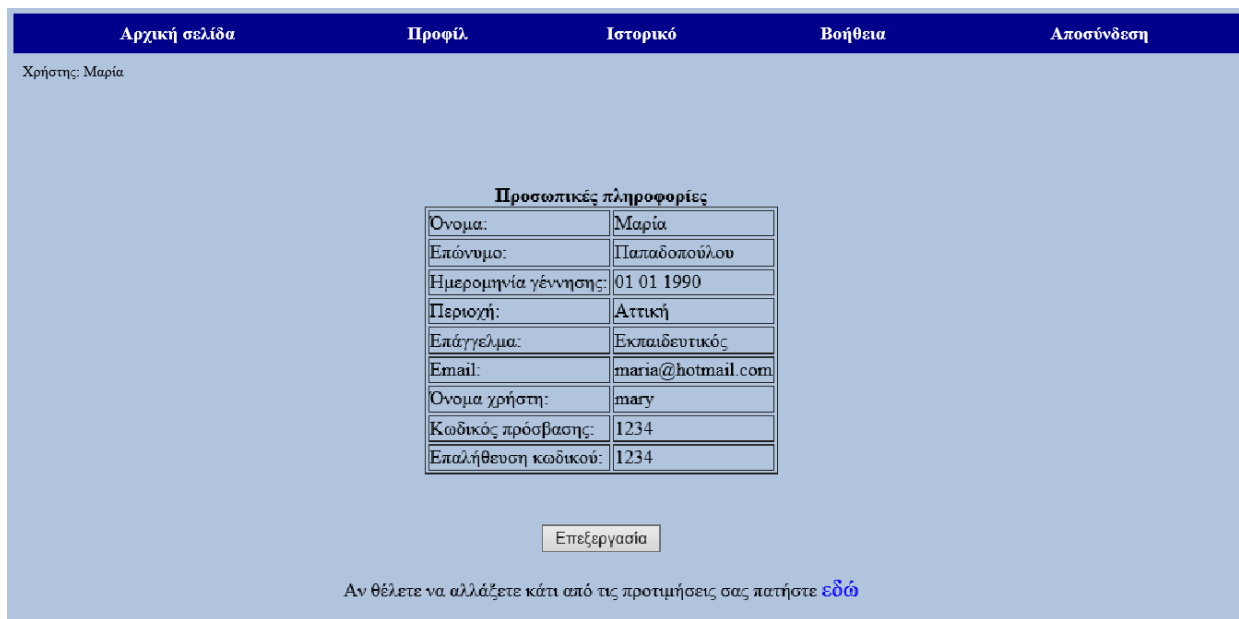
Εικόνα 7: Σελίδα συνδεδεμένου χρήστη

Όταν ο χρήστης επιλέξει τον αλγόριθμο Social-Network based Collaborative Filtering εμφανίζονται κάτω ακριβώς από τα φίλτρα αποτελεσμάτων κατάλληλα προϊόντα για εκείνον μαζί με τη λεπτομερή περιγραφή των χαρακτηριστικών τους (βλ. Εικόνα 8). Η επιλογή αλγορίθμου είναι υποχρεωτική προκειμένου να παρουσιαστούν αποτελέσματα στην οθόνη. Αντιθέτως, η χρήση των φίλτρων είναι προαιρετική.

Αρχική σελίδα	Προφίλ	Ιστορικό	Βοήθεια	Αποσύνδεση
<p>Χρήστης: Μαρία</p> <p>Κατηγορία</p> <p><input type="checkbox"/> Βιβλία <input type="checkbox"/> Ταινίες <input type="checkbox"/> Μουσική</p> <p>Περιοχή Τιμών</p> <p>Από: € <input type="text"/> Έως: € <input type="text"/></p> <p>Αλγόριθμος</p> <p><input checked="" type="radio"/> Social-Network Based Collaborative Filtering <input type="radio"/> User Based Collaborative filtering <input type="radio"/> Item Based Collaborative filtering <input type="radio"/> Clustering <input type="radio"/> Knowledge Based filtering <input type="radio"/> Demography Based filtering <input type="radio"/> Content Based filtering</p> <p><input type="button" value="Καταχώρηση"/></p>				
	<p>Album: Loud (2010)</p> <p>Καλλιτέχνης: Rihanna Είδος: Hip Hop Δισκογραφική εταιρεία: Def Jam Recordings Κατάσταση: Καινούριο</p>	<p>Πληροφορίες Παράδοσης:</p> <p>Προμηθευτής: Public.gr Διεύθυνση: Καραγεώργη Σερβίας 1 Πλ. Συντάγματος, Αθήνα Ημερομηνία Παράδοσης: 2013-02-12</p>		<p>Τιμή: 9.99 € +0 € έξοδα αποστολής</p>
	<p>Album: 21(Limited Edition) (2011)</p> <p>Καλλιτέχνης: Adele Είδος: Pop Δισκογραφική εταιρεία: XL/Beggars Group Κατάσταση: Καινούριο</p>	<p>Πληροφορίες Παράδοσης:</p> <p>Προμηθευτής: Public.gr Διεύθυνση: Καραγεώργη Σερβίας 1 Πλ. Συντάγματος, Αθήνα Ημερομηνία Παράδοσης: 2013-02-12</p>		<p>Τιμή: 17.98 € +0 € έξοδα αποστολής</p>
	<p>Album: The Collection (2006)</p> <p>Καλλιτέχνης: Deep Purple Είδος: Rock Δισκογραφική εταιρεία: SONY MUSIC Κατάσταση: Καινούριο</p>	<p>Πληροφορίες Παράδοσης:</p> <p>Προμηθευτής: Public.gr Διεύθυνση: Καραγεώργη Σερβίας 1 Πλ. Συντάγματος, Αθήνα Ημερομηνία Παράδοσης: 2013-02-12</p>		<p>Τιμή: 6.99 € +0 € έξοδα αποστολής</p>
	<p>Album: Santana III (1998)</p> <p>Καλλιτέχνης: Carlos Santana Είδος: Rock Δισκογραφική εταιρεία: COL Κατάσταση: Καινούριο</p>	<p>Πληροφορίες Παράδοσης:</p> <p>Προμηθευτής: Public Διεύθυνση: Καραγεώργη Σερβίας 1 Πλ. Συντάγματος, Αθήνα Ημερομηνία Παράδοσης: 2013-02-12</p>		<p>Τιμή: 9.99 € +0 € έξοδα αποστολής</p>

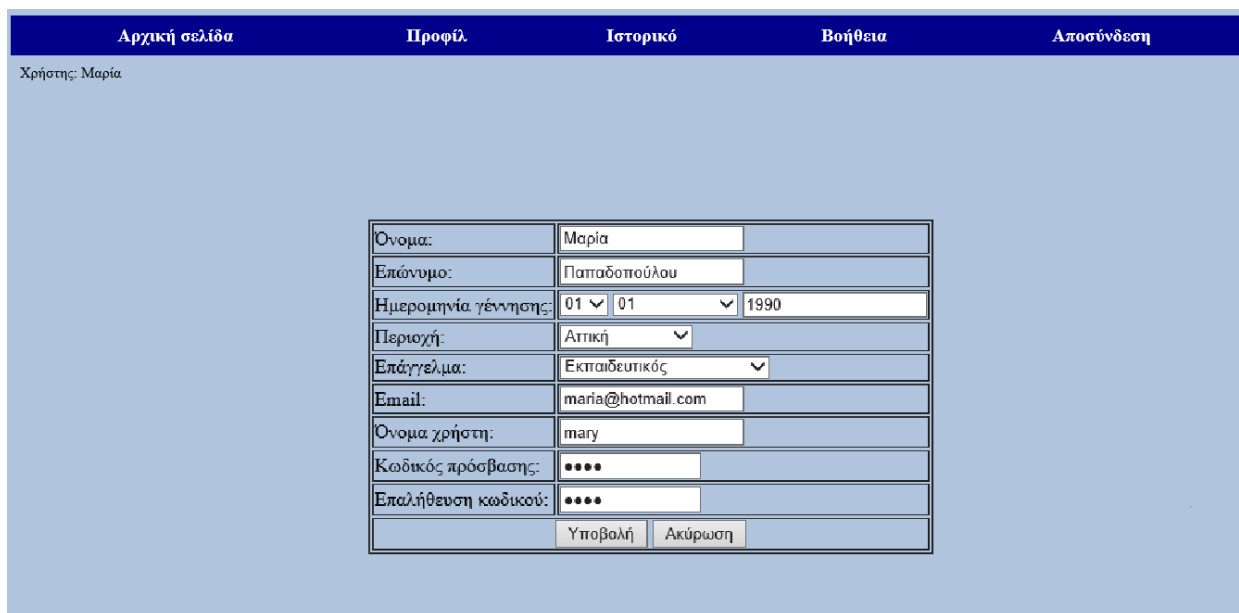
Εικόνα 8: Εμφάνιση αποτελεσμάτων

Επεξεργασία προφίλ: Εδώ δίνεται η δυνατότητα στο χρήστη να τροποποιήσει τα προσωπικά του στοιχεία καθώς και τις προτιμήσεις που καταχώρησε στο σύστημα κατά την εγγραφή του (βλ. Εικόνες 9-10). Επιλέγοντας “*Προφίλ*” από το μενού επιλογών εμφανίζονται σε πρώτο στάδιο οι προσωπικές πληροφορίες του χρήστη.



Εικόνα 9: Εμφάνιση προσωπικών στοιχείων

Για να τροποποιήσει τις προσωπικές του πληροφορίες, ο χρήστης πρέπει να πατήσει το κουμπί “Επεξεργασία” και εν συνεχεία να υποβάλλει τα νέα του στοιχεία.







Εικόνα 10: Επεξεργασία προσωπικών στοιχείων

Αν ο χρήστης ακολουθήσει το σχετικό σύνδεσμο για την τροποποίηση των προτιμήσεών του, θα ξεκινήσει μια διαδικασία όμοια με αυτήν που εκτελείται κατά την εγγραφή για την καταχώρησή τους στο σύστημα.

Ιστορικό συστάσεων: Η επιλογή αυτή (βλ. Εικόνα 11) παρουσιάζει στην οθόνη όλα τα διαφορετικά προϊόντα που έχουν συσταθεί στο χρήστη σε παλιότερες συνδέσεις του. Η

Χρήση Κοινωνικών Δικτύων ως Βάση ενός Συστήματος Συστάσεων

εμφάνιση του προϊόντος συνοδεύεται από τα κυριότερα χαρακτηριστικά του, τον αλγόριθμο που χρησιμοποιήθηκε για τη σύσταση και την ημερομηνία που πραγματοποιήθηκε η σύσταση.

Αρχική σελίδα	Προφίλ	Ιστορικό	Βοήθεια	Αποσύνδεση
Χρήστης: Μαρία				
Σας έχουν προταθεί τα παρακάτω προϊόντα:				
	Album: Loud Καλλιτέχνης: Rihanna	Προτάθηκε στις 21/3/2013 με τον αλγόριθμο Social network based CF		
	Album: 21 (Limited Edition) Καλλιτέχνης: Adele	Προτάθηκε στις 21/3/2013 με τον αλγόριθμο Social network based CF		
	Album: The Collection Καλλιτέχνης: Deep Purple	Προτάθηκε στις 21/3/2013 με τον αλγόριθμο Social network based CF		
	Album: Santana III Καλλιτέχνης: Carlos Santana	Προτάθηκε στις 21/3/2013 με τον αλγόριθμο Social network based CF		

Εικόνα 11: Ιστορικό συστάσεων

Εγχειρίδιο βοήθειας: Πρόκειται για ένα απλοποιημένο εγχειρίδιο με τη μορφή ερωτήσεων - απαντήσεων σχετικά με διάφορα θέματα που αφορούν τη λειτουργία του ιστοχώρου (βλ. Εικόνα 12).

Αρχική σελίδα	Προφίλ	Ιστορικό	Βοήθεια	Αποσύνδεση
Χρήστης: Μαρία				
<ul style="list-style-type: none"> Πώς μπορώ να επεξεργαστώ τα προσωπικά μου στοιχεία; Από το μενού επιλογών μεταβείτε στο σύνδεσμο Προφίλ, και αυτομάτως θα εμφανιστούν τα προσωπικά σας στοιχεία και τα στοιχεία λογαριασμού που έχετε καταχωρήσει στο σύστημα. Για οποιαδήποτε τροποποίηση, πατήστε το κουμπί Επεξεργασία και υποβάλλετε εκ νέου τα στοιχεία σας συμπληρώνοντας τη φόρμα που θα εμφανιστεί. Αφού συνδεθείτε με το σύστημα, πατήστε το σύνδεσμο Προφίλ. Για να αλλάξετε τα στοιχεία σας πατήστε το σύνδεσμο Επεξεργασία και συμπληρώστε την αντίστοιχη φόρμα. Πώς μπορώ να αλλάξω τις προτιμήσεις μου; Από το μενού επιλογών μεταβείτε στο σύνδεσμο Προφίλ, και έπειτα κάντε κλικ στον αντίστοιχο σύνδεσμο που εμφανίζεται για την τροποποίηση των προτιμήσεων κάτω από το κουμπί "Επεξεργασία". Επιλέξτε την κατηγορία προϊόντων για την οποία θέλετε να επεξεργαστείτε τις προτιμήσεις σας και πατήστε Συνέχεια. Για οποιαδήποτε αλλαγή, πρέπει να επαναληφτηθούν οι προτιμήσεις σας από την αρχή. Οι παλιές προτιμήσεις δε διατηρούνται. Αν δεν επιθυμείτε τελικά να αλλάξετε τις προτιμήσεις σας, αλλά έχετε ήδη μεταβεί στην αντίστοιχη σελίδα πατήστε το κουμπί Άκυρο. Με την ολοκλήρωση εισαγωγής νέων προτιμήσεων ανακατευθύνεστε στη σελίδα επιλογής κατηγορίας προϊόντος. Για τον τερματισμό της διαδικασίας πατήστε το κουμπί Τέλος. Πώς μπορώ να πάρω αποτελέσματα συστάσεων; Από την αρχική σελίδα επιλέξτε αλγόριθμο και πατήστε το κουμπί Καταχώρηση. Μπορείτε προαιρετικά να επιλέξετε ένα ή περισσότερα φίλτρα για να εμφανιστούν αποτελέσματα συγκεκριμένης κατηγορίας προϊόντων ή καθορισμένης τιμής. Για την επιλογή τιμής πρέπει να εισαχθεί στη φόρμα ένα λογικό εύρος τιμών. Μπορεί εναλλακτικά να εισαχθεί η ίδια τιμή και στα δύο πεδία ώστε να εμφανιστούν τα αποτελέσματα μίας και μόνο συγκεκριμένης τιμής αντί για ένα διάστημα τιμών. Πώς λειτουργεί ο αλγόριθμος Social-Network Based Collaborative filtering; Ο αλγόριθμος αυτός χρησιμοποιεί τα προσωπικά στοιχεία και τις προτιμήσεις σας που έχετε εισάγει στο σύστημα ενώ παράλληλα αντλεί πληροφορίες από το Facebook. Συνεπώς για να εμφανιστούν αποτελέσματα θα πρέπει υποχρεωτικά να έχετε έναν προσωπικό λογαριασμό και τουλάχιστον ένα φίλο στο Facebook. Επισκερθείτε τη σελίδα μας στο Facebook πατώντας εδώ για να δείτε τα προϊόντα μας και να τα βαθμολογήσετε εσείς και οι φίλοι σας. Μπορώ να επιλέξω άλλον αλγόριθμο εκτός από το Social-Network Based Collaborative filtering; Προσωρινά ο αλγόριθμος Social-Network Based Collaborative filtering είναι ο μόνος ο οποίος εμφανίζει αποτελέσματα. Οι υπόλοιποι αλγόριθμοι είναι υπό κατασκευή. Τι είναι το ιστορικό; Στο ιστορικό εμφανίζονται όλα τα προϊόντα που σας έχει προτείνει το σύστημα στο παρελθόν αναφέροντας την ημερομηνία που πραγματοποιήθηκε η σύσταση καθώς και τον αλγόριθμο που χρησιμοποιήθηκε. 				

Εικόνα 12: Εγχειρίδιο βοήθειας

Αποσύνδεση: Η αποσύνδεση έχει ως κύριο ρόλο την έξοδο του χρήστη από τη σελίδα συνδεδεμένου χρήστη, τον τερματισμό της συνεδρίας και την παραπομπή του στην αρχική σελίδα εισόδου.

3.2 Βάση Δεδομένων (DataBase - DB)

Η βάση δεδομένων περιλαμβάνει όλες τις απαραίτητες πληροφορίες για τις οντότητες που χρησιμοποιούνται στο σύστημα και τις μεταξύ τους συσχετίσεις. Ο σχεδιασμός της βασίζεται στη δημιουργία ενός επεκταμένου μοντέλου οντοτήτων-συσχετίσεων (Enhanced Entity-Relationship model) με top-down μεθοδολογία. Σύμφωνα με το μοντέλο αυτό διακρίνονται οι ακόλουθες τρεις οντότητες:

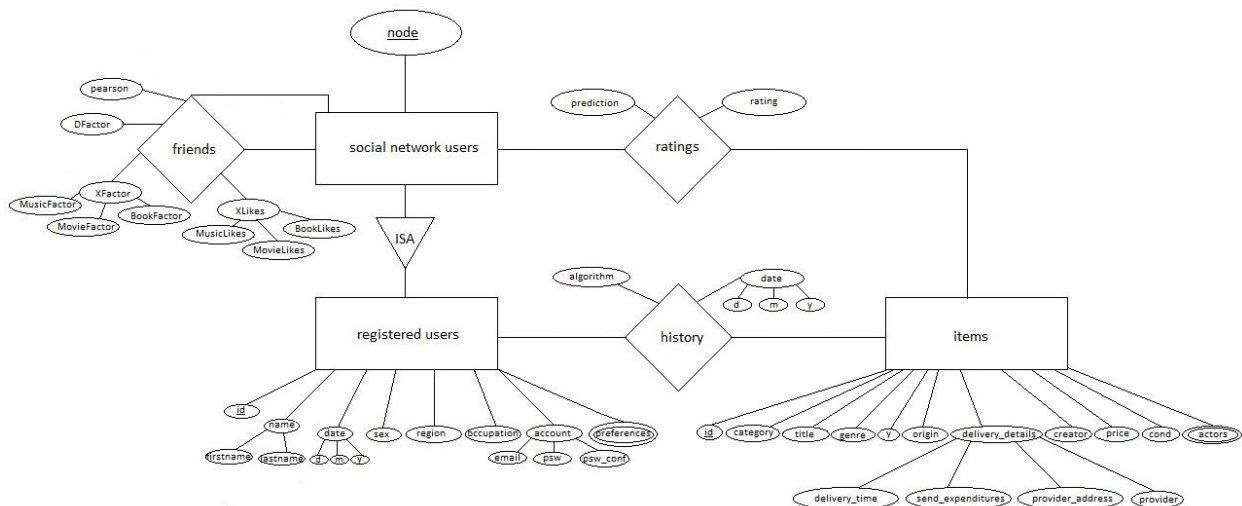
- Χρήστες που είναι εγγεγραμμένοι στο σύστημα:** Πρόκειται για μια οντότητα η οποία αντιπροσωπεύει ένα σύνολο ανθρώπων που αποκτούν λογαριασμό στο RIC και διαμορφώνουν ένα προσωπικό προφίλ. Το προφίλ απαρτίζεται από δημογραφικά στοιχεία, προτιμήσεις του χρήστη για χαρακτηριστικά προϊόντων και δεδομένα για τη διαχείριση του λογαριασμού του. Οι προτιμήσεις αποτελούν ένα πλειότιμο γνώρισμα καθώς μπορούν να πάρουν ένα σύνολο από διαφορετικές τιμές.
- Χρήστες κοινωνικών δικτύων:** Η οντότητα αυτή αντικατοπτρίζει ανθρώπους που έχουν λογαριασμό σε κάποιο κοινωνικό δίκτυο. Αυτοί βαθμολογούν αντικείμενα μέσω του λογαριασμού τους, έχουν προσωπικές πληροφορίες που είναι αναρτημένες στο προφίλ κοινωνικού δικτύου που διαθέτουν και συνάπτουν κοινωνικούς δεσμούς με άλλους χρήστες που είναι εγγεγραμμένοι στο ίδιο κοινωνικό δίκτυο. Στην προσέγγιση που έχουμε υιοθετήσει για το σχεδιασμό του

συστήματος θεωρούμε ότι όλοι οι χρήστες του συστήματός μας είναι ταυτόχρονα και χρήστες κοινωνικών δικτύων καθώς σε αντίθετη περίπτωση ο προτεινόμενος αλγόριθμος δεν έχει ισχύ. Το αντίστροφο δεν ισχύει απαραίτητα. Κατόπιν αυτού, οι χρήστες κοινωνικών δικτύων αποτελούν μια υπερ-οντότητα ενώ οι χρήστες που είναι εγγεγραμμένοι στο σύστημα είναι μια οντότητα χαμηλότερου επιπέδου που κληρονομεί όλα τα γνώρισμα και τις συσχετίσεις της υπερ-οντότητας

- Αντικείμενα του συστήματος:** Η εν λόγω οντότητα αναφέρεται στα προϊόντα της εικονικής αγοράς πάνω στην οποία δομείται το σύστημα. Από αυτό το σύνολο αντικειμένων, ένας αριθμός προϊόντων θα συσταθούν στους χρήστες ως αποτέλεσμα του αλγορίθμου σύστασης. Τα προϊόντα που αποτελούν ταινίες διαθέτουν ένα επιπλέον πλειότιμο γνώρισμα το οποίο είναι οι ηθοποιοί που πρωταγωνιστούν σε αυτές.

Το μοντέλο περιλαμβάνει επίσης τρεις συσχετίσεις μεταξύ των οντοτήτων. Αρχικά, κάθε χρήστης κοινωνικού δικτύου συνδέεται με ένα προϊόν μέσω της αριθμητικής αξιολόγησης (βαθμολογίας) που δίνει σε αυτό. Η αξιολόγηση μπορεί είτε να έχει δοθεί από το χρήστη και να λαμβάνεται ως δεδομένο είτε να έχει υπολογιστεί από τον αλγόριθμο σύστασης. Επίσης, υπάρχει μια αναδρομική συσχέτιση μεταξύ χρηστών κοινωνικού δικτύου που υποδηλώνει τους κοινωνικούς δεσμούς (φιλίες) μεταξύ τους. Τέλος, οι χρήστες του συστήματος συσχετίζονται με τα προϊόντα μέσω του ιστορικού συστάσεων. Το ιστορικό συστάσεων διατηρεί όλα τα προϊόντα που το σύστημα έχει προτείνει στο χρήστη στο παρελθόν.

Το μοντέλο οντοτήτων-συσχετίσεων που περιγράφει την ανωτέρω βάση δεδομένων απεικονίζεται στο Σχήμα 2:



Σχήμα 2: Διάγραμμα οντοτήτων-συσχετίσεων

Η βάση δεδομένων περιλαμβάνει πίνακες σχετικά με τους χρήστες που είναι εγγεγραμμένοι στο σύστημα, τα προϊόντα της εικονικής αγοράς, καθώς και πίνακες με συνδυαστικές πληροφορίες. Η οντότητα των χρηστών κοινωνικών δικτύων δεν αναπαρίσταται. Δημιουργούνται επίσης νέες σχέσεις για την αναπαράσταση των πλειοτίμων γνωρισμάτων των οντοτήτων. Ειδικότερα, οι πίνακες που συνθέτουν τη βάση είναι οι εξής:

- ***user_details (id, firstname, lastname, d, m, y, sex, region, occupation, email, username, psw, psw_conf, node)***

Ο πίνακας αυτός αφορά τους χρήστες που είναι εγγεγραμμένοι στο σύστημα και διατηρεί τις προσωπικές πληροφορίες για κάθε έναν από αυτούς. Συγκεκριμένα, οι πληροφορίες αφορούν τα δεδομένα που καταχωρεί ο χρήστης στη φόρμα εγγραφής και το αναγνωριστικό που του δίνεται από το σύστημα (id). Το αναγνωριστικό αυτό αποτελεί και το πρωτεύον κλειδί του πίνακα. Σύνθετα γνωρίσματα όπως το ονοματεπώνυμο και η ημερομηνία γέννησης αποσυντίθενται στα απλά γνωρίσματα που τα απαρτίζουν. Στον user_details υπάρχει ένα επιπλέον πεδίο με όνομα node που λαμβάνει τιμή σε μεταγενέστερο στάδιο και αφορά μια μοναδική συμβολοσειρά που αντιπροσωπεύει τον κόμβο του χρήστη σε ένα κοινωνικό δίκτυο (Social Network). Το πεδίο node είναι ένα γνώρισμα που κληρονομείται από την υπερ-οντότητα των χρηστών κοινωνικών δικτύων και η λειτουργία του θα αναλυθεί σε επόμενη ενότητα.

- ***user_preferences (id, userid, preference, category, attribute)***

Εδώ αποθηκεύονται οι προτιμήσεις κάθε χρήστη με βάση το αναγνωριστικό του. Καθώς οι προτιμήσεις αποτελούν πλειότιμο γνώρισμα, ο πίνακας περιλαμβάνει ένα πεδίο (userid) που ταυτίζεται με το πρωτεύον κλειδί του user_details. Η προτίμηση λαμβάνεται σαν ένα σύνθετο γνώρισμα το οποίο αποτελείται από τρία απλά πεδία, τα preference, category και attribute. Το πεδίο preference δηλώνει το χαρακτηριστικό του αντικείμενου που προτιμά ο χρήστης (π.χ. μυθιστόρημα, κωμωδία, Rock κ.λπ.), το πεδίο category δείχνει την κατηγορία στην οποία ανήκει το αντικείμενο (π.χ. Βιβλίο, Ταινία, Μουσική) και το πεδίο attribute αναφέρει την ιδιότητα του χαρακτηριστικού (π.χ. είδος βιβλίου, ηθοποιός, τραγουδιστής κ.λπ.). Το πεδίο id είναι ένα πρωτεύον κλειδί αυτόματης αρίθμησης.

- ***bmm_details (id, category, title, genre, y, origin, creator, price, delivery_time, cond, send_expenditures, provider, provider_address)***

Ο πίνακας bmm_details αποθηκεύει τα στοιχεία των αντικειμένων της εικονικής αγοράς. Μεταξύ των στοιχείων περιλαμβάνονται χαρακτηριστικά όπως είδος αντικείμενου (μυθιστόρημα, θρίλερ κ.λπ.), κατηγορία στην οποία ανήκει (μουσική, βιβλίο, ταινία), τιμή, έτος κυκλοφορίας και λοιπά καθώς και πληροφορίες παράδοσης. Το πεδίο origin αποκτά διαφορετικό χαρακτήρα ανάλογα με την κατηγορία του προϊόντος. Στα βιβλία αφορά τις εκδόσεις τους, στη μουσική τη δισκογραφική εταιρεία ενώ στις ταινίες την εταιρεία παραγωγής. Ομοίως το πεδίο creator αντιπροσωπεύει το συγγραφέα στα βιβλία, τον τραγουδιστή ή το μουσικό συγκρότημα στα άλμπουμ και το σκηνοθέτη στις ταινίες. Επίσης, σε κάθε αντικείμενο αντιστοιχεί ένα μοναδικό αναγνωριστικό (id), το οποίο και αποτελεί πρωτεύον κλειδί του πίνακα αυτού.

- **actors (actorid, name, movieid)**

Εδώ βρίσκονται τα ονοματεπώνυμα των ηθοποιών και η αντιστοίχισή τους με τις ταινίες του πίνακα `bmm_details` στις οποίες πρωταγωνιστούν. Το πεδίο `actorid` είναι το αναγνωριστικό κάθε ηθοποιού.

- **user_ratings (id, node, itemid, rating, prediction)**

Στον εν λόγω πίνακα εισάγονται οι αξιολογήσεις που δίνουν χρήστες κοινωνικών δικτύων, συμπεριλαμβανομένων και των χρηστών του συστήματος, στα αντικείμενα που διατηρεί η βάση. Κάποια πεδία του πίνακα αυτού συμπληρώνονται από τις πληροφορίες που αποσπώνται από τα κοινωνικά δίκτυα. Το πεδίο `node` είναι ο κόμβος του χρήστη στο κοινωνικό δίκτυο και το πεδίο `itemid` ταυτίζεται με το αναγνωριστικό προϊόντος. Στο πεδίο `rating` αποθηκεύεται η αξιολόγηση του χρήστη για ένα αντικείμενο. Σε κάθε εκτέλεση του αλγορίθμου σύστασης ο `user_ratings` συμπληρώνεται με αξιολογήσεις που προβλέπει το σύστημα για το χρήστη – στόχο για τα αντικείμενα που δεν έχει ακόμη αξιολογήσει. Η διάκριση σχετικά με το αν η αξιολόγηση έχει διαβαστεί από κοινωνικό δίκτυο ή έχει υπολογιστεί από τον αλγόριθμο γίνεται με το πεδίο `prediction` το οποίο λαμβάνει τιμές 0 και 1 αντίστοιχα. Το πεδίο `id` είναι ένα πρωτεύον κλειδί αυτόματης αρίθμησης.

- **friends (id, node1, node2, weight, Dfactor, MusFactor, MovFactor, BFactor, MusLikes, Blikes, MovLikes)**

Ο πίνακας `friends` κρατά τους άμεσους κοινωνικούς δεσμούς μεταξύ των χρηστών όπως αυτοί λαμβάνονται από το κοινωνικό δίκτυο. Όπως θα αναλυθεί παρακάτω οι κοινωνικοί δεσμοί θα αποτελέσουν για τον αλγόριθμο ένα πρώτο σύνολο γειτόνων (`neighbours`) του χρήστη-στόχου. Παράλληλα, για κάθε δεσμό αποθηκεύονται μετρικές που υπολογίζει ο αλγόριθμος. Οι μετρικές αυτές υποδηλώνουν σε τι ποσοστό ένας χρήστης ταιριάζει με ένα γείτονά του. Ουσιαστικά, κάθε εγγραφή του πίνακα `friends` αναπαριστά μια σταθμισμένη (`weighted`) ακμή σε έναν υπογράφο του κοινωνικού δικτύου. Το πεδίο `id` είναι ένα πρωτεύον κλειδί αυτόματης αρίθμησης.

- **history (id, userid, itemid, d, m, y, algorithm)**

Στον πίνακα αυτόν συλλέγονται οι πληροφορίες που αφορούν το ιστορικό των συστάσεων. Περιλαμβάνει πεδία για την αποθήκευση της ημερομηνίας και του αλγορίθμου σύστασης. Η ημερομηνία αποτελεί σύνθετο γνώρισμα που διασπάται σε ημέρα (`d`), μήνα (`m`) και έτος (`y`). Επιπλέον τα γνωρίσματα `userid` και `itemid` ταυτίζονται με τα πεδία κλειδιά των πινάκων `user_details` και `bmm_details` αντίστοιχα. Το πεδίο `id` είναι ένα πρωτεύον κλειδί αυτόματης αρίθμησης.

3.3 Μηχανή συστάσεων (Recommender Engine - RE)

Η μηχανή συστάσεων είναι ένα βασικό στοιχείο του συνολικού συστήματος που συνδέεται με τη διεπαφή σύστασης (RIC) και τη βάση δεδομένων ενώ παράλληλα αντλεί και πληροφορίες από κοινωνικά δίκτυα. Είναι το υποσύστημα που εκτελεί τον αλγόριθμο σύστασης, παράγει αποτελέσματα και τα επιστρέφει πίσω στο RIC ώστε να εμφανιστούν στο χρήστη. Η RE, όπως συμβαίνει και στα περισσότερα συστήματα συστάσεων, δέχεται σαν δεδομένα εισόδου ένα προφίλ χρήστη, ένα προφίλ αντικειμένων και την ανάδραση χρήστη προκειμένου να προκύψουν οι συστάσεις.

Το προφίλ χρήστη διαμορφώνεται τόσο από ρητά (explicit) δεδομένα όσο και από έμμεσα (implicit). Τα ρητά δεδομένα προκύπτουν από τα δημογραφικά στοιχεία και τις προτιμήσεις που έχει εισάγει ο χρήστης κατά την εγγραφή του στο RIC. Τα έμμεσα δεδομένα αφορούν κυρίως πληροφορίες που εξάγονται από προφίλ χρηστών σε κοινωνικά δίκτυα. Το προφίλ αντικειμένων λαμβάνεται με άμεσο τρόπο από τον πίνακα `bmm_details` αφού τα προϊόντα έχουν καταχωρηθεί απευθείας με όλα τα χαρακτηριστικά τους. Η ανάδραση χρήστη πραγματοποιείται από τις αξιολογήσεις των αντικειμένων που δίνουν οι χρήστες σε κοινωνικά δίκτυα. Στα πλαίσια της παρούσας διπλωματικής οι αξιολογήσεις αφορούν αριθμητικές βαθμολογίες που κυμαίνονται από το 1 έως το 5 όπου το 1 αντιπροσωπεύει πολύ κακό προϊόν κατά την άποψη του χρήστη, το 2 αντιστοιχεί σε κακό προϊόν, το 3 σε μέτριο, το 4 σε καλό και το 5 σε πολύ καλό.

Για την υλοποίηση της μηχανής συστάσεων αντλήθηκαν πληροφορίες από μόνο ένα κοινωνικό δίκτυο, το Facebook. Το Facebook επιλέχθηκε ενδεικτικά έναντι άλλων κοινωνικών δικτύων επειδή διαθέτει ένα ευρύ φάσμα πλεονεκτημάτων. Καταρχάς, αποτελεί ένα από τα πιο δημοφιλή κοινωνικά δίκτυα παγκοσμίως και απαρτίζεται από αμφίδρομες κοινωνικές σχέσεις. Επιπρόσθετα, τα προφίλ των χρηστών του είναι εμπλουτισμένα με μια πληθώρα προσωπικών πληροφοριών. Όσον αφορά τον προγραμματισμό εφαρμογών υπάρχουν αρκετά εγχειρίδια και εργαλεία που είναι διαθέσιμα για προγραμματιστές στον ιστότοπο του Facebook ώστε να μπορεί να κανείς να δημιουργήσει κοινωνικές εφαρμογές. Τέλος, παρέχει και τη δυνατότητα δημιουργίας σελίδων και από απλούς χρήστες. Μια τέτοια σελίδα δημιουργήθηκε ώστε να αναρτώνται εκεί οι αξιολογήσεις των χρηστών του συστήματός μας. Στο επόμενο κεφάλαιο αναλύεται εκτενώς ολόκληρη η διαδικασία που εκτελείται για την παραγωγή των συστάσεων συμπεριλαμβάνοντας την πλήρη επεξήγηση του αλγορίθμου σύστασης αλλά και την ανάκτηση και συμμετοχή σε αυτόν δεδομένων από το Facebook.

4. ΜΗΧΑΝΙΣΜΟΣ ΣΥΣΤΑΣΕΩΝ

4.1 Η χρήση του Facebook

Το Facebook είναι μια διαδικτυακή (online) υπηρεσία κοινωνικής δικτύωσης (SNS) που ιδρύθηκε το Φεβρουάριο του 2004 από τον Mark Zuckerberg. Το όνομά του προέρχεται από το βιβλίο που δίνεται από κάποια Πανεπιστήμια των Ηνωμένων Πολιτειών στους σπουδαστές τους κατά την έναρξη του ακαδημαϊκού έτους για να τους βοηθήσουν να γνωριστούν μεταξύ τους. Αν και αρχικά η ιστοσελίδα αποτελούσε ένα δίκτυο για φοιτητές κολλεγίων που ήθελαν να μένουν διασυνδεδεμένοι ανταλλάζοντας πληροφορίες και βρίσκοντας νέους φίλους, στη συνέχεια επεκτάθηκε σε ένα παγκόσμιο δίκτυο. Από το Σεπτέμβριο του 2012, το Facebook έχει πάνω από ένα δισεκατομμύριο ενεργούς χρήστες [27]. Η πρόσβαση στον ιστότοπο είναι δωρεάν για τα μέλη και τα έσοδα αντλούνται από διαφημίσεις [30].

Οι χρήστες μετά την εγγραφή τους στο δικτυακό τόπο δημιουργούν ένα προσωπικό προφίλ με φωτογραφίες, λίστες προσωπικών ενδιαφερόντων, στοιχεία επικοινωνίας, και άλλες προσωπικές πληροφορίες. Μπορούν επίσης να προσθέσουν άλλους χρήστες ως φίλους (friends) στις κοινωνικές επαφές τους και να τους κατηγοριοποιήσουν προαιρετικά σε λίστες, όπως «Οι άνθρωποι από την εργασία» ή «Στενοί φίλοι». Χρησιμοποιώντας τις δυνατότητες αναζήτησης, είναι δυνατό να εντοπίσουν άλλα μέλη του Facebook στέλνοντας μια πρόσκληση, ή μπορούν να προσκαλούν κόσμο να συμμετάσχει στο Facebook [30]. Τα μέλη επικοινωνούν με τις επαφές τους μέσω ιδιωτικών ή δημόσιων μηνυμάτων, μέσω γραπτής συνομιλίας (chat) σε πραγματικό χρόνο ή ακόμη και μέσω βιντεοκλήσεων.

Οι χρήστες του Facebook μπορούν να ενταχθούν σε ομάδες χρηστών (Groups) κοινού ενδιαφέροντος, που διοργανώνονται από το χώρο εργασίας, το σχολείο, το κολέγιο, ή άλλους τομείς, επιτρέποντας στα μέλη να βρουν ο ένας τον άλλο και να αλληλεπιδρούν μεταξύ τους. Κάθε κοινότητα ανθρώπων μπορεί να δημιουργήσει μια τέτοια ομάδα οι διαχειριστές της οποίας ενδέχεται να αποδεχθούν όλα τα μέλη ή να απορρίψουν τις αιτήσεις ένταξης σε αυτήν με βάση τα κριτήρια της ομάδας. Άλλα μέλη δημιουργούν μια σελίδα Facebook (Facebook Page) που είναι δομημένη γύρω από ένα συγκεκριμένο θέμα. Σε πολλές περιπτώσεις, οι σελίδες δημιουργούνται από επιχειρήσεις για την προώθηση των προϊόντων και των εμπορικών σημάτων τους. Όλα τα μέλη σε αυτήν την περίπτωση γίνονται δεκτά ως οπαδοί. Σε αντίθεση με το Facebook προφίλ, το οποίο είναι συνήθως μια αμφίδρομη αλληλεπίδραση με τους φίλους, μια σελίδα είναι πρωτίστως μια μονόδρομη μετάδοση από το διαχειριστή της προς τα μέλη της. Οι σελίδες έχουν επίσης στατιστικά στοιχεία, με βάση τα οποία οι διαχειριστές καταλαβαίνουν πώς αλληλεπιδρούν οι άλλοι χρήστες με τη σελίδα. Ακόμη, δίνεται η δυνατότητα στους χρήστες να δημιουργήσουν ειδικές σελίδες που ονομάζονται εκδηλώσεις (Events). Αυτές επιτρέπουν στα μέλη να δημοσιεύσουν μια εκδήλωση και διάφορες πληροφορίες σχετικά με αυτήν, να προσκαλούν κόσμο και να παρακολουθούν ποιοι σχεδιάζουν να παραστούν.

Στο προσωπικό προφίλ του κάθε μέλους, υπάρχουν πολλά βασικά στοιχεία του δικτύου. Το πιο δημοφιλές είναι το χρονολόγιο (Timeline), δηλαδή μια περιοχή που παραθέτει τις πιο αντιπροσωπευτικές φωτογραφίες, ανακοινώσεις και εμπειρίες του χρήστη με χρονολογική σειρά. Το Facebook παρέχει ένα χαρακτηριστικό microblogging το οποίο δίνει στα μέλη τη δυνατότητα να κάνουν δημοσιεύσεις (posts) στο χρονολόγιό τους, στο χρονολόγιο άλλου χρήστη είτε σε κάποια σελίδα. Μια δημοσίευση μπορεί να περιλαμβάνει γραπτό κείμενο, συνδέσμους, φωτογραφίες, βίντεο, τοποθεσίες με συνοδεία χάρτη ή προσθήκη ετικετών (tags) σε φίλους. Όταν μάλιστα η δημοσίευση ενός χρήστη γίνεται από τον ίδιο στο δικό του χρονολόγιο, αποτελεί μια ενημέρωση

κατάστασης (status update). Ένα άλλο δημοφιλές στοιχείο είναι το εικονικό Άλμπουμ Φωτογραφιών. Οι φωτογραφίες μπορούν να φορτωθούν από την επιφάνεια εργασίας ή απευθείας από μια κάμερα κινητού τηλεφώνου. Δεν υπάρχει περιορισμός στην ποσότητα, αλλά αφαιρούνται ακατάλληλες εικόνες ή εικόνες με πνευματικά δικαιώματα. Μια διαδραστική λειτουργία album καθιστά εφικτό να προσδιοριστούν τα άτομα της φωτογραφίας με ετικέτες (tags) τόσο από τον ίδιο το χρήστη που ανέβασε τη φωτογραφία όσο και από τους φίλους του. Τα μέλη μπορούν να αλληλεπιδρούν μεταξύ τους εκφράζοντας σχόλια (comments) για κάποιο στοιχείο του δικτύου, λόγω χάρη για μια δημοσίευση ή για κάποια φωτογραφία. Μία επιπλέον δημοφιλή λειτουργία είναι το κουμπί like το οποίο επιτρέπει στους χρήστες να εκφράσουν την εκτίμησή τους για περιεχόμενο, όπως είναι οι δημοσιεύσεις, τα σχόλια, οι φωτογραφίες, οι σελίδες Facebook κτλ. Όλες οι δραστηριότητες και αλληλεπιδράσεις των μελών δημοσιεύονται στις ενημερώσεις (News Feed), ενώ πολύ ενεργά μέλη βλέπουν τις ενημερώσεις σε πραγματικό χρόνο σε μια στήλη ενημέρωσης (Ticker) που υπάρχει στο προφίλ τους [30].

Για λόγους προστασίας της ιδιωτικής ζωής, το Facebook επιτρέπει στους χρήστες να καθορίσουν τις δικές τους ρυθμίσεις απορρήτου και να επιλέξουν ποιος μπορεί να δει συγκεκριμένα μέρη του προφίλ τους. Ειδικότερα, ένα μέλος μπορεί να κάνει όλες τις επικοινωνίες του ορατές σε όλους, μπορεί να τις καταστήσει ορατές μόνο σε συγκεκριμένες επαφές ή μπορεί να τις κρατήσει όλες ιδιωτικές. Τα μέλη μπορούν να επιλέξουν εάν πρέπει ή όχι να είναι δυνατή η αναζήτησή τους, να αποφασίσουν ποια μέρη του προφίλ τους είναι δημόσια, να αποφασίσουν τι δεν πρέπει να δημοσιευθεί στις ενημερώσεις τους και να καθορίσουν ακριβώς ποιος μπορεί να δει τις δημοσιεύσεις ή τα άλμπουμ φωτογραφιών τους. Οι χρήστες του Facebook από προεπιλογή μπορούν να δουν μόνο τα προφίλ των επιβεβαιωμένων φίλων τους.

Ένα άλλο χαρακτηριστικό του Facebook, το οποίο το καθιστά διαφορετικό από άλλους ιστοχώρους κοινωνικής δικτύωσης, είναι η δυνατότητα να προσθέσει κάποιος χρήστης επιπλέον εφαρμογές στο προφίλ του δηλαδή μικρά προγράμματα που αναπτύχθηκαν ειδικά για Facebook προφίλ [29]. Μερικά παραδείγματα περιλαμβάνουν ενημερωτικές εφαρμογές όπως δελτία καιρού, εφαρμογές με ημερολόγια που υπενθυμίζουν σε ένα χρήστη τα γενέθλια των φίλων του καθώς και εκατοντάδες εφαρμογές παιχνιδιών. Το Μάιο του 2007, το Facebook εισήγαγε τη δική του πλατφόρμα εφαρμογών για τρίτους προγραμματιστές προκειμένου να τους επιτρέψει να δημιουργήσουν εφαρμογές που, αφού εγκριθούν, θα μπορούσαν να διανεμηθούν μέσα από την κοινότητα του Facebook [28].

4.1.1 Facebook Graph API

Το κοινωνικό δίκτυο του Facebook δεν είναι ένας απλός πίνακας δεδομένων αλλά έχει τη μορφή ενός γράφου όπως αυτός ορίζεται στην επιστήμη της Πληροφορικής. Όλα τα αντικείμενα που το απαρτίζουν, μεταξύ των οποίων και οι χρήστες του, αποτελούν ένα σύνολο κόμβων που συνδέονται μεταξύ τους. Το Facebook μάλιστα χρησιμοποιεί επισήμως τον όρο κοινωνικό γράφημα (Social Graph) για να περιγράψει το κοινωνικό του δίκτυο. Ο όρος διαδόθηκε στο συνέδριο του Facebook στις 24 Μαΐου του 2007, στη συνέχεια όμως επεκτάθηκε για να αναφερθεί σε ένα κοινωνιόγραμμα που απεικονίζει προσωπικές σχέσεις όλων των χρηστών του Διαδικτύου.

Ένα σημαντικό ζήτημα που τέθηκε αφορά το γεγονός ότι το κοινωνικό γράφημα του Facebook ανήκει στην εταιρεία και δεν διαμοιράζεται με άλλες υπηρεσίες, γεγονός που της δίνει ένα σημαντικό πλεονέκτημα έναντι των ανταγωνιστών της. Αυτό παράλληλα παρεμποδίζει τη φορητότητα και την ιδιοκτησία των προσωπικών πληροφοριών ενός

ατόμου έτσι ώστε να μπορεί να τις χρησιμοποιήσει και σε άλλες υπηρεσίες. Κάθε κοινωνικό δίκτυο απαιτεί εκ νέου δημιουργία λογαριασμού, προσωπικού προφίλ και επαφών. Η Google προσπάθησε να προσφέρει μια λύση σε αυτό το πρόβλημα με τη δημιουργία του Social Graph API, που κυκλοφόρησε τον Ιανουάριο του 2008, το οποίο επέτρεπε σε ιστοσελίδες να δηλώσουν αυτοματοποιημένα τις κοινωνικές συνδέσεις μεταξύ τους. Με αυτόν τον τρόπο έκανε τις πληροφορίες σχετικά με τις συνδέσεις μεταξύ των ανθρώπων στο Διαδίκτυο εύκολα διαθέσιμες και χρήσιμες για τους προγραμματιστές σχηματίζοντας μια φορητή online ταυτότητα για κάθε άτομο. Το Social Graph API αποσύρθηκε το 2012 ενώ το Facebook εισάγει το δικό του Graph API στο συνέδριο του 2010 προκειμένου να διαθέσει τη χρήση του γραφήματός του σε προγραμματιστές [32].

Το API αυτό παρουσιάζει μια απλή, συνεκτική εικόνα του κοινωνικού γράφου του Facebook, αντιπροσωπεύοντας ομοιόμορφα όλα τα αντικείμενα στο γράφο και τις συνδέσεις μεταξύ τους. Έτσι επιτρέπει σε ιστοσελίδες να αντλούν πληροφορίες και σχετικά με άλλα αντικείμενα εκτός από χρήστες, όπως φωτογραφίες, εκδηλώσεις, σελίδες, κ.λπ. Αποτελεί μάλιστα το βασικό μηχανισμό που χρησιμοποιείται από τα εργαλεία του Facebook για iOS, Android, PHP και JavaScript ενώ παράλληλα δύναται να χρησιμοποιηθεί από εργαλεία τρίτων που διατίθενται για άλλες γλώσσες. Υπάρχουν δύο επιπλέον APIs, τα οποία επίσης χρησιμοποιούνται στο Facebook για πρόσβαση στο γράφημα: το FQL και το API REST Legacy. Αυτά τα APIs περιέχουν παρόμοια και επικαλυπτόμενη λειτουργικότητα, αλλά έχουν αντικατασταθεί. Το Graph API είναι πλέον ο κύριος τρόπος για να λάβει κανείς δεδομένα από τον κοινωνικό γράφο του Facebook και συνεπώς η μελέτη της λειτουργίας του είναι ιδιαίτερα σημαντική.

4.1.2 Πρόσβαση δεδομένων

Για λόγους προστασίας της ιδιωτικής ζωής, η επιτυχής χρήση του Graph API απαιτεί μια σειρά διαδικασιών προκειμένου να δοθεί στους προγραμματιστές πρόσβαση στα αντικείμενα και τις συνδέσεις του γράφου. Κατ' αρχάς, το Graph API καθώς και πολλές από τις πιο προηγμένες λειτουργίες που αφορούν APIs του Facebook απαιτούν οι χρήστες να είναι συνδεδεμένοι με μια υπηρεσία που καλείται "σύνδεση Facebook" (Facebook Login). Η υπηρεσία αυτή χρησιμοποιείται για τη σύνδεση χρηστών σε μια εφαρμογή ή έναν ιστοχώρο που ενσωματώνει στοιχεία και χρήστες του Facebook. Έχει πολλά χαρακτηριστικά ασφαλείας για την προστασία προσωπικών πληροφοριών, επιτρέποντας στους χρήστες να ελέγχουν τι μοιράζονται και στους προγραμματιστές να ζητήσουν ασφαλή πρόσβαση στις πληροφορίες αυτές. Επιπλέον χρησιμοποιεί το πρωτόκολλο OAuth2.0 για αυθεντικοποίηση (authentication) και εξουσιοδότηση (authorization).

Μετά τη σύνδεση του χρήστη, η εφαρμογή έχει αυτομάτως πρόσβαση στο δημόσιο προφίλ του και τη λίστα φίλων του (public profile and friend list). Από προεπιλογή, το δημόσιο προφίλ περιλαμβάνει το ονοματεπώνυμό του, το όνομα χρήστη στο Facebook, το φύλο, την ηλικιακή ομάδα, το URL του προφίλ του, τη γλώσσα, τη χώρα, την ταυτότητα (id) του κόμβου του στον κοινωνικό γράφο και άλλες δημόσιες πληροφορίες. Για να αποκτήσει πρόσβαση σε περισσότερες πληροφορίες σχετικά με τους χρήστες, η εφαρμογή θα πρέπει να ζητήσει πρόσθετα δικαιώματα (permissions) από αυτούς. Το είδος των ζητούμενων δικαιωμάτων εξαρτάται από τη φύση των δεδομένων στα οποία χρειάζεται να αποκτήσει πρόσβαση. Στο πλαίσιο αυτό, το Facebook υλοποιεί ένα ευρύ σύνολο ειδικότερων δικαιωμάτων τα οποία εμπίπτουν στις ακόλουθες ομάδες:

Δικαιώματα Email (Email permissions): Χρησιμοποιούνται για πρόσβαση στην ηλεκτρονική διεύθυνση ενός ατόμου. Η ηλεκτρονική διεύθυνση αποτελεί ευαίσθητο και

προστατευμένο δεδομένο και για το λόγο αυτό η πρόσβαση σε αυτήν πρέπει να ζητηθεί από την εφαρμογή ρητά και θα επιτευχθεί μόνο εάν χορηγηθεί τελικά από το χρήστη.

Εκτεταμένες Ιδιότητες Προφίλ (Extended Profile Properties): Παρέχουν πρόσβαση σε περισσότερα δεδομένα προφίλ και πιο ευαίσθητες πληροφορίες. Χρησιμοποιούνται μόνο για ανάγνωση. Δεδομένα που υπάγονται σε αυτήν την κατηγορία είναι οι φωτογραφίες του χρήστη, τα ενδιαφέροντά του, τα γενέθλιά του, το θρήσκευμά του κ.λπ. Τα δικαιώματα αυτά διακρίνονται σε δικαιώματα δεδομένων χρήστη (user data permissions) και δικαιώματα δεδομένων φίλου (friend's data permissions). Τα πρώτα αφορούν δεδομένα σχετικά με το συνδεδεμένο χρήστη ενώ τα δεύτερα αναφέρονται σε δεδομένα σχετικά με τους φίλους του χρήστη και έχουν την ίδια μορφή με τα δικαιώματα δεδομένων χρήστη.

Εκτεταμένα Δικαιώματα (Extended Permissions): Δίνουν βαθύτερη πρόσβαση σε ευαίσθητες πληροφορίες ενός ατόμου και τη δυνατότητα δημοσίευσης και διαγραφής δεδομένων. Τα δικαιώματα αυτά είναι προαιρετικά δηλαδή ένα άτομο μπορεί να αρνηθεί ένα εκτεταμένο δικαίωμα. Αλλά ακόμη και αν το άτομο τα έχει χορηγήσει στην εφαρμογή, μπορεί να αλλάξει αυτήν την απόφαση αργότερα στις ρυθμίσεις απορρήτου του. Παραδείγματα δικαιωμάτων αυτής της κατηγορίας είναι η πρόσβαση στα εισερχόμενα μηνύματα του χρήστη (read_mailbox), και η δημοσίευση περιεχομένου και σχολίων στη ροή του χρήστη και των φίλων του (publish_actions).

Δικαιώματα ανοιχτού γραφήματος (Open Graph Permissions): Το ανοιχτό γράφημα (Open Graph) είναι ένα εργαλείο που επιτρέπει στις εφαρμογές να αφηγούνται ιστορίες και γεγονότα στο Facebook μέσα από ένα API. Κατ' επέκταση, τα δικαιώματα ανοιχτού γραφήματος παρέχουν πρόσβαση στα Open Graph APIs και δίνουν τη δυνατότητα στην εφαρμογή να δημοσιεύει στο ανοιχτό γράφημα (Open Graph) διάφορες δραστηριότητες καθώς και να ανακαλεί δραστηριότητες που έχουν δημοσιευθεί από άλλες εφαρμογές. Το ανοιχτό γράφημα είναι μία από τις βασικές έννοιες του κοινωνικού δικτύου και εισήχθη από το Facebook το 2010. Πλέον είναι διαθέσιμο για εφαρμογές τρίτων αποτελώντας ένα ισχυρό εργαλείο για μεγάλα εμπορικά σήματα (brands). Για παράδειγμα, η μουσική εφαρμογή Spotify χρησιμοποιεί την τεχνολογία του ανοιχτού γραφήματος για να δημοσιεύσει ό,τι ακούει ο συνδεδεμένος σε αυτήν χρήστης στις ενημερώσεις (news feed) των φίλων του [26].

Δικαιώματα Σελίδας (Page Permissions): Αφορούν δικαιώματα που σχετίζονται με τις σελίδες του Facebook που διαχειρίζεται ένας χρήστης. Είναι προαιρετικά όπως και τα εκτεταμένα δικαιώματα και μπορούν να αφαιρεθούν αργότερα από το χρήστη αφού τα έχει χορηγήσει στην εφαρμογή.

Σημειώνεται ότι επειδή το Facebook δίνει τη δυνατότητα στους χρήστες του να ελέγχουν τι θα επιλέξουν να μοιραστούν με τις εφαρμογές, ακόμα κι αν μια εφαρμογή ζητήσει κάποια δικαιώματα, υπάρχει ενδεχόμενο να μην είναι σε θέση να δει τα δεδομένα που ζητά. Ακόμη και για μια απλή είσοδο σε μια ιστοσελίδα, ένα πρόσωπο πρέπει να επιβεβαιώσει ότι θα επιτρέψει στην εφαρμογή να έχει πρόσβαση στις πληροφορίες του. Εάν ο χρήστης δεχτεί τελικά να χορηγήσει στην εφαρμογή τα δικαιώματα που εκείνη ζητά, ξεκινά μια ροή εφαρμογής που ονομάζεται εξουσιοδότηση (authorization). Η ροή αυτή απαιτεί τη συμμετοχή του πελάτη (mobile ή web), των διακομιστών αν υπάρχουν και των διακομιστών του Facebook. Κατά τη φάση της εξουσιοδότησης τα δικαιώματα λαμβάνονται ως ένα σύνολο συμβολοσειρών διαχωρισμένων με κόμμα.

Στο τέλος της διαδικασίας εξουσιοδότησης δημιουργείται ένα διακριτικό πρόσβασης (Access Token). Το Access Token είναι μια τυχαία συμβολοσειρά που παρέχει προσωρινή, ασφαλή πρόσβαση στα Facebook APIs. Τα διακριτικά πρόσβασης είναι συνήθως χρονικά περιορισμένα και περιλαμβάνουν πληροφορίες σχετικά με το πότε

λήγουν καθώς και ποια εφαρμογή τα παρήγαγε. Όλα τα Access Token δημιουργούνται με αυθεντικοποίηση και διαδικασίες έγκρισης OAuth 2.0. Υπάρχουν τρεις τύποι διακριτικών πρόσβασης. Ο κυριότερος τύπος token είναι τα διακριτικά πρόσβασης χρήστη. Παράγονται στην ροή εισόδου (login flow) όταν ο χρήστης παραχωρήσει δικαιώματα στην εφαρμογή και μπορεί να χρησιμοποιηθούν για να εκτελεστούν API κλήσεις για λογαριασμό του, ανάλογα με αυτά τα δικαιώματα. Για παράδειγμα, με ένα διακριτικό πρόσβασης χρήστη μπορεί να ανακτηθεί η λίστα των φίλων ενός χρήστη ή οι δημοσιεύσεις του. Αυτά τα tokens έχουν μια σύντομη περίοδο ισχύος από προεπιλογή (συνήθως μία με δύο ώρες), αλλά είναι δυνατόν να ανταλλαχθούν για μακροβιότερα tokens και είναι μοναδικά για κάθε χρήστη και για κάθε εφαρμογή.

Υπάρχουν επίσης τα διακριτικά πρόσβασης σελίδας και τα διακριτικά πρόσβασης εφαρμογής. Αυτοί οι δύο τύποι tokens δεν χρησιμοποιούνται σε μια ευρεία ποικιλία από εφαρμογές, αλλά είναι διαθέσιμοι, έτσι ώστε να μπορούν να χτιστούν εφαρμογές διαχείρισης σελίδας ή να παραχωρήσουν πρόσβαση στα δεδομένα μιας εφαρμογής αντίστοιχα. Ειδικότερα, το διακριτικό πρόσβασης σελίδας χρησιμοποιείται για την πραγματοποίηση κλήσεων API για λογαριασμό μιας σελίδας. Λόγου χάρη μπορεί να χρησιμοποιηθεί για τη δημοσίευση μιας ενημέρωσης κατάστασης (status update) σε μια σελίδα αντί στο χρονοδιάγραμμα του χρήστη (timeline). Για να δημιουργηθεί ένα διακριτικό πρόσβασης σε αυτήν την περίπτωση, ο χρήστης που είναι ο διαχειριστής σελίδας πρέπει να χορηγήσει ένα δικαίωμα σελίδας που ονομάζεται *manage_pages*. Τα διακριτικά πρόσβασης σελίδας είναι μοναδικά σε κάθε σελίδα, διαχειριστή (admin) και εφαρμογή. Τέλος, ένα διακριτικό πρόσβασης εφαρμογής χρησιμεύει για να τροποποιήσει τις ρυθμίσεις της εφαρμογής, να δημιουργήσει και να διαχειριστεί δοκιμαστικούς χρήστες (test users)² ή να διαβάσει στατιστικά δεδομένα για τη χρήση και την κυκλοφορία των εφαρμογών. Μπορεί επίσης να χρησιμοποιηθεί για τη δημοσίευση ή τη διαγραφή περιεχομένου για λογαριασμό ενός χρήστη, ο οποίος έδωσε αντίστοιχα δικαιώματα στην εφαρμογή. Για παράδειγμα, εάν ένας χρήστης έχει χορηγήσει στην εφαρμογή το δικαίωμα *publish_actions*, η εφαρμογή μπορεί να χρησιμοποιήσει ένα διακριτικό πρόσβασης εφαρμογής για να δημοσιεύσει μια ενημέρωση κατάστασης για λογαριασμό του. Τα διακριτικά πρόσβασης εφαρμογής είναι μοναδικά για κάθε εφαρμογή και εν γένει δε λήγουν εκτός αν ακυρωθεί κάποια ρύθμιση της εφαρμογής.

4.1.3 Κλήσεις στο Graph API (Graph API calls)

Το Graph API είναι ένα απλό (χαμηλού επιπέδου) HTTP-based API που μπορεί να χρησιμοποιηθεί για αναζήτηση δεδομένων στο γράφο του Facebook, δημοσίευση νέων ιστοριών, ή οποιοδήποτε άλλη ενέργεια μπορεί να χρειάζεται να εκτελέσει μια εφαρμογή. Το Graph API οδηγείται από HTTP αιτήματα στον host **graph.facebook.com** και οι HTTP μέθοδοι αντιστοιχίζονται σε συγκεκριμένες λειτουργίες που μπορούν να διεξαχθούν στο γράφημα. Η μέθοδος GET χρησιμοποιείται για ανάκτηση των πληροφοριών των κόμβων, η μέθοδος POST για τροποποίηση και προσθήκη κόμβων ενώ η μέθοδος DELETE για κατάργηση κόμβων.

Κάθε κόμβος στο γράφο αποτελεί ένα αντικείμενο (object) που περιλαμβάνει διάφορα πεδία (fields). Για να πραγματοποιηθεί η προσπέλαση των κόμβων τίθεται σε αυτούς ένα ερώτημα. Είναι δυνατόν επίσης να τεθούν ερωτήματα και στις συνδέσεις μεταξύ των

² Ο δοκιμαστικός χρήστης είναι ένας ειδικός λογαριασμός χρήστη, αόρατος στους "κανονικούς" χρήστες, ο οποίος δημιουργείται για λογαριασμό μιας εφαρμογής για δοκιμαστικούς σκοπούς. Κάθε εφαρμογή μπορεί να δημιουργήσει μέχρι και 2000 δοκιμαστικούς χρήστες.

κόμβων. Οτιδήποτε υπάρχει στο γράφο του facebook έχει ένα μοναδικό αναγνωριστικό (id) και συνήθως μέσω αυτού γίνεται η προσπέλασή του. Εξαιρέση αποτελούν οι χρήστες και οι σελίδες που μπορούν να προσεγγιστούν και με το όνομά τους. Η πρόσβαση σε συγκεκριμένα πεδία του αντικειμένου γίνεται με ένα query string της μορφής ?fields=FIELD1, FIELD2,.....,FIELDN όπου τα FIELD1, FIELD2,.....,FIELDN είναι τα ονόματα διαφόρων πεδίων. Λόγου χάρη, έστω ένας χρήστης με αναγνωριστικό USER_ID για τον οποίο θέλουμε να έχουμε πρόσβαση στο δημόσιο προφίλ του και να ανακτήσουμε τα στοιχεία του. Το URL θα έχει τη μορφή:

https://graph.facebook.com/USER_ID

και θα εκτελεστεί το HTTP αίτημα:

GET /USER_ID HTTP/1.1

Host: graph.facebook.com

Connection: close

HTTP/1.1 200 OK

Content-Type: text/javascript; charset=UTF-8

Expires: Sat, 01 Jan 2000 00:00:00 GMT

Date: Thu, 11 Oct 2012 17:28:02 GMT

Connection: close

Content-Length: 48

το οποίο θα επιστρέψει όλα τα διαθέσιμα πεδία για το εν λόγω αντικείμενο (όνομα, επίθετο, φύλο, κ.λπ.). Αν θέλαμε να διαβάσουμε μόνο το όνομά του, το URL θα έπαιρνε τη μορφή:

https://graph.facebook.com/USER_ID?fields=name

όπου *name* καλείται το πεδίο στο οποίο αποθηκεύεται το όνομα του χρήστη.

Ο τρόπος να επιτευχθεί η πρόσβαση σε μια σύνδεση είναι να δημιουργηθεί ένα μονοπάτι (path) με την προσθήκη της σύνδεσης μετά από το αντικείμενο που είναι η πηγή της σύνδεσης ή ανάμεσα στα αντικείμενα που συνδέονται. Λόγου χάρη, έστω ότι ο χρήστης με αναγνωριστικό USER_ID συνδέεται με ένα σύνολο από φίλους. Η ανάγνωση της λίστας φίλων του συγκεκριμένου χρήστη απαιτεί τη φόρτωση του μονοπατιού /USER_ID/friends όπου friends είναι το όνομα της σύνδεσης που αντιπροσωπεύει τους φίλους. Το αντίστοιχο URL θα έχει τη μορφή:

https://graph.facebook.com/USER_ID/friends

Στις παραπάνω περιπτώσεις, τα αιτήματα ήταν τύπου GET. Επειδή η μέθοδος GET είναι προεπιλεγμένη, δεν είναι απαραίτητο να συμπεριλαμβάνεται στη σύνταξη του

URL. Στην περίπτωση όμως των POST και DELETE θα πρέπει η μέθοδος να αναφέρεται ρητά. Εάν για παράδειγμα κάποια εφαρμογή θέλει να κάνει, για λογαριασμό του χρήστη, like σε μια δημοσίευση (post) με αναγνωριστικό POST_ID, το αντίστοιχο URL θα είναι:

https://graph.facebook.com/POST_ID/likes?method=POST

Στην προκειμένη περίπτωση χρησιμοποιείται η μέθοδος POST εφόσον εκτελείται μια ενέργεια τροποποίησης.

Ωστόσο, με κάποια από τα ανωτέρω URL σε αυτήν τη μορφή δε θα επιστραφεί ακόμη κάποιο αποτέλεσμα. Συνήθως, λόγω των ελέγχων προστασίας προσωπικών δεδομένων που διαθέτει το Facebook, κάθε αίτημα στο Graph API θα απαιτήσει να περαστεί κατά μήκος του ένα διακριτικό πρόσβασης. Ως εκ τούτου, σε ορισμένα από τα παραπάνω παραδείγματα, προκειμένου να είναι επιτυχείς οι κλήσεις στο API θα πρέπει να τεθεί επιπλέον και το διακριτικό πρόσβασης στο τέλος του αιτήματος. Για παράδειγμα, η ανάκτηση των φίλων ενός χρήστη θα προκύψει τελικά από το URL:

https://graph.facebook.com/USER_ID/friends?access_token=ACCESS_TOKEN

όπου ACCESS_TOKEN είναι ένα έγκυρο διακριτικό πρόσβασης χρήστη. Ομοίως για να κάνει η εφαρμογή like σε μια δημοσίευση για λογαριασμό του χρήστη το URL θα πρέπει να πάρει τη μορφή:

https://graph.facebook.com/POST_ID/likes?method=POST&access_token=ACCESS_TOKEN

Το διακριτικό πρόσβασης σε κάθε περίπτωση θα πρέπει να έχει προκύψει από την χορήγηση των κατάλληλων δικαιωμάτων. Λόγου χάρη στην προηγούμενη περίπτωση για να κάνει like στη δημοσίευση η εφαρμογή, θα πρέπει να της έχει χορηγηθεί το εκτεταμένο δικαίωμα “publish_stream”. Εάν η εφαρμογή δεν έχει λάβει από το χρήστη τα κατάλληλα δικαιώματα, κάποια δεδομένα πιθανόν να λείπουν από τα αποτελέσματα μιας κλήσης. Λόγου χάρη, όσον αφορά το URL ***https://graph.facebook.com/USER_ID***, λανθασμένο διακριτικό πρόσβασης ή παντελής απουσία του κατά την κλήση για ανάκτηση των πληροφοριών ενός χρήστη θα επιστρέψει ορισμένα μόνο πεδία ενώ κάποια άλλα δε θα εμφανίζονται καθόλου. Υπάρχουν συνεπώς πεδία τα οποία απαιτούν την προσθήκη διακριτικού πρόσβασης ενώ άλλα εμφανίζονται και χωρίς αυτό. Για παράδειγμα, το όνομα του χρήστη δεν απαιτεί διακριτικό πρόσβασης οπότε σε αυτήν την περίπτωση το URL ***https://graph.facebook.com/USER_ID?fields=name*** θα επιστρέψει κανονικά το επιθυμητό αποτέλεσμα. Σημειώνεται επίσης ότι όλα τα αντικείμενα που μπορεί να εμφανιστούν στα αποτελέσματα αναπαρίστανται σε JSON μορφοποίηση.

Εν συνεχεία αναφέρονται πληροφορίες σχετικά με κάποια βασικά αντικείμενα του γράφου του Facebook συμπεριλαμβανομένων των πεδίων τους, των συνδέσεών τους και των δικαιωμάτων που απαιτούν ώστε να επιτευχθεί η πρόσβαση σε αυτά. Ένα πολύ σημαντικό αντικείμενο του κοινωνικού γραφήματος είναι το αντικείμενο User το οποίο

αντιπροσωπεύει το χρήστη του Facebook. Μερικά από τα βασικότερα πεδία του αντικειμένου User είναι τα ακόλουθα (βλ. Πίνακα 1):

Πίνακας 1: Πεδία του αντικειμένου User

Όνομα	Περιγραφή	Δικαιώματα	Τύπος επιστροφής
id	Το Facebook αναγνωριστικό του χρήστη	Δεν απαιτείται access_token	String
name	Το πλήρες όνομα του χρήστη	Δεν απαιτείται access_token	String
first_name	Το όνομα του χρήστη	Δεν απαιτείται access_token	String
middle_name	Το ενδιάμεσο όνομα του χρήστη	Δεν απαιτείται access_token	String
last_name	Το επώνυμο του χρήστη	Δεν απαιτείται access_token	String
gender	Το φύλο του χρήστη: άνδρας ή γυναίκα	Δεν απαιτείται access_token	String
age_range	Περιοχή ηλικίας του χρήστη. Επιστρέφεται μόνο εάν ζητηθεί ειδικά ως παράμετρος του URL	Απαιτεί access_token	Το αντικείμενο περιέχει min και max για την περιοχή ηλικίας. Πιθανές ηλικιακές ομάδες είναι 13-17, 18-20 και 21 +. Το max δεν ορίζεται αν η ηλικία είναι 21 +
work	Μία λίστα του ιστορικού εργασίας του χρήστη	user_work_history ή friends_work_history	Πίνακας αντικειμένων που περιέχει τον εργοδότη (employer), την τοποθεσία (location), τη

			θέση (position), την ημερομηνία έναρξης (start_date) και την ημερομηνία λήξης (end)
birthday	Τα γενέθλια του χρήστη	user_birthday ή friends_birthday	String ημερομηνίας στη MM/HH/EEEE μορφή

Το αντικείμενο User έχει μεταξύ άλλων τις ακόλουθες συνδέσεις (βλ. Πίνακα 2):

Πίνακας 2: Συνδέσεις του αντικειμένου User

Όνομα	Περιγραφή	Δικαιώματα	Τύπος επιστροφής
books	Τα βιβλία που αναφέρονται στο προφίλ του χρήστη	user_likes ή friends_likes	Πίνακας αντικειμένων που περιέχει σαν πεδία το αναγνωριστικό του βιβλίου (id), το όνομα (name), την κατηγορία (category) και το χρόνο δημιουργίας του (create_time)
friendlists	Η λίστα φίλων του χρήστη	read_friendlists	Πίνακας αντικειμένων που περιέχει τα πεδία id και name της λίστας φίλων
friends	Οι φίλοι του χρήστη	Κάθε έγκυρο access_token του τρέχοντος session χρήστη	Πίνακας των αντικειμένων που περιέχει τα πεδία id και name των φίλων
groups	Οι ομάδες στις οποίες ανήκει ο χρήστης	user_groups ή friends_groups	Πίνακας αντικειμένων με τα πεδία version (παλιά-0 ή νέα ομάδα-1), name, id και bookmark_order
interests	Τα ενδιαφέροντα	user_interests ή	Πίνακας αντικειμένων με τα

	που αναφέρονται στο προφίλ του χρήστη	friends_interests	πεδία id, name, category και create_time του ενδιαφέροντος
likes	Όλες οι σελίδες που άρεσαν στο χρήστη	user_likes ή friends_likes	Πίνακας αντικειμένων με τα πεδία id, name, category και create_time του like
movies	Οι ταινίες που αναφέρονται στο προφίλ του χρήστη	user_likes ή friends_likes	Πίνακας αντικειμένων με τα πεδία id, name, category και create_time της ταινίας
music	Η μουσική που αναφέρεται στο προφίλ του χρήστη	user_likes ή friends_likes	Πίνακας αντικειμένων με τα πεδία id, name, category και create_time της μουσικής
mutualfriends	Οι κοινοί φίλοι μεταξύ δύο χρηστών	Κάθε έγκυρο access_token του τρέχοντος session χρήστη	Πίνακας των αντικειμένων που περιέχει τα πεδία id και name των φίλων
permissions	Τα δικαιώματα που έχει χορηγήσει ο χρήστης στην εφαρμογή	κανένα	Πίνακας που περιέχει ένα μόνο αντικείμενο που έχει ως κλειδιά τα ονόματα δικαιωμάτων και ως τιμές τις τιμές των δικαιωμάτων
posts	Οι δημοσιεύσεις του χρήστη	κάθε έγκυρο access_token ή read_stream για την προβολή ιδιωτικών δημοσιεύσεων	Πίνακας από Post αντικείμενα

Ειδικότερα, οι συνδέσεις friendlists, friends και mutualfriends επιστρέφουν αντικείμενα με τα ακόλουθα πεδία (βλ. Πίνακα 3):

Πίνακας 3: Πεδία των συνδέσεων friendlists, friends και mutualfriends

Όνομα	Περιγραφή	Τύπος επιστροφής
Id	Facebook ID του φίλου	String
Name	Όνομα του φίλου του τρέχοντα χρήστη	String

ενώ οι συνδέσεις like, books, movies, music και interests επιστρέφουν αντικείμενα με τα εξής πεδία (βλ. Πίνακα 4):

Πίνακας 4: Πεδία των συνδέσεων like, books, movies, music και interests

Όνομα	Περιγραφή	Τύπος επιστροφής
Id	Facebook ID της σελίδας	String
Name	Όνομα της σελίδας	String
Category	Κατηγορία της σελίδας, για παράδειγμα 'Website', 'Product/Service', κτλ.	String
created_time	ISO-8601 datetime που δείχνει πότε συνδέθηκε ο χρήστης στη σελίδα.	String

Μια δημοσίευση αναπαρίσταται από το αντικείμενο Post. Μερικά από τα πεδία του αντικειμένου Post είναι (βλ. Πίνακα 5):

Πίνακας 5: Πεδία του αντικειμένου Post

Όνομα	Περιγραφή	Δικαιώματα	Τύπος επιστροφής
id	Το αναγνωριστικό της δημοσίευσης	Απαιτεί access_token	String
from	Πληροφορίες για το χρήστη που δημοσίευσε το μήνυμα	Απαιτεί access_token	Αντικείμενο που περιέχει το όνομα και το Facebook id του χρήστη που δημοσίευσε το

			μήνυμα
to	Τα προφίλ που αναφέρονται η στοχεύουν στο συγκεκριμένο post	Απαιτεί access_token	Πίνακας αντικειμένων, καθένα με το όνομα και το Facebook id του χρήστη
message	Το μήνυμα	Απαιτεί access_token	String
comments	Τα σχόλια για αυτό το post	read_stream	Δομή που περιέχει ένα αντικείμενο δεδομένων που με τη σειρά του περιέχει ένα πίνακα αντικειμένων, το καθένα με τα πεδία id, from, message, και created_time για κάθε σχόλιο
object_id	Το Facebook αναγνωριστικό του αντικειμένου για το ανέβασμα μιας φωτογραφίας ή ενός βίντεο	read_stream	αριθμός

Μερικές συνδέσεις του αντικειμένου Post είναι οι παρακάτω (βλ. Πίνακα 6):

Πίνακας 6: Συνδέσεις του αντικειμένου Post

Όνομα	Περιγραφή	Δικαιώματα	Τύπος επιστροφής
comments	Όλα τα σχόλια σε αυτό το post	Κάθε έγκυρο access_token	πίνακας αντικειμένων που περιλαμβάνει τα πεδία id, from, to message και created_time.
likes	Τα likes στο	Κάθε έγκυρο	πίνακας αντικειμένων που

	συγκεκριμένο ποστ	access_token	περιλαμβάνει τα πεδία id και name
--	----------------------	--------------	--------------------------------------

Τέλος, ένα σχόλιο αναπαρίσταται από το αντικείμενο Comment. Μερικά από τα πεδία του Comment είναι τα παρακάτω (βλ. Πίνακα 7):

Πίνακας 7: Πεδία του αντικειμένου Comment

Όνομα	Περιγραφή	Δικαιώματα	Τύπος επιστροφής
id	Το Facebook ID του comment	generic access_token	string
from	Ο χρήστης που έγραψε το comment	generic access_token	object που περιέχει το id και το όνομα (name) του χρήστη που το δημιούργησε
message	Το κείμενο του comment	generic access_token	string
parent	Αν το comment είναι απάντηση, το πεδίο αυτό επιστρέφει τον parent του comment, αλλιώς δεν επιστρέφει τίποτα	generic access_token	reference
comment_count	Ο αριθμός από απαντήσεις (replies) σε αυτό το comment, αν υπάρχουν	generic access_token	integer

Το αντικείμενο Comment έχει επίσης τις παρακάτω συνδέσεις (βλ. Πίνακα 8):

Πίνακας 8: Συνδέσεις του αντικειμένου Comment

Όνομα	Περιγραφή	Δικαιώματα	Τύπος επιστροφής
comments	Όλα τα comments που είναι	Οποιοδήποτε έγκυρο	Πίνακας από αντικείμενα comment

	απαντήσεις σε αυτό	access_token	
likes	Όλα τα likes σε αυτό το comment	Οποιοδήποτε έγκυρο access_token	Πίνακας από αντικείμενα με τα πεδία id και name του χρήστη που έκανε like στο comment.

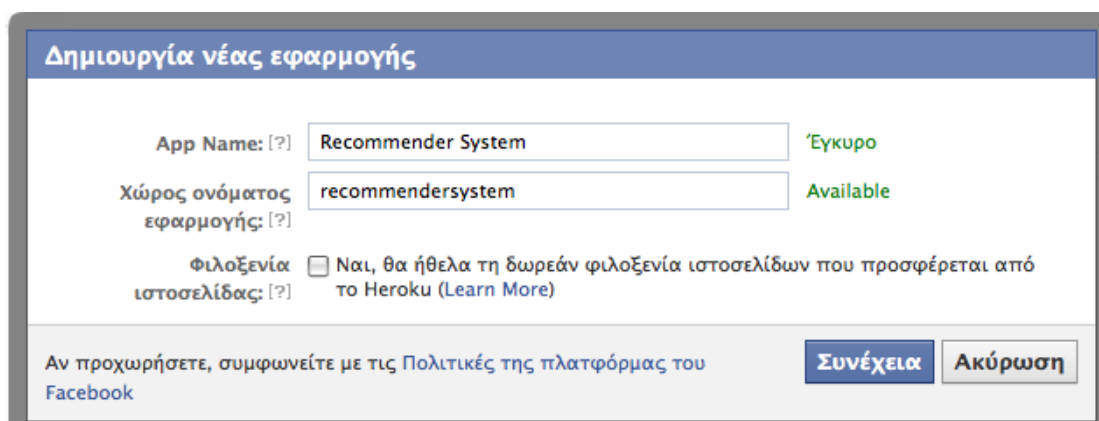
4.1.4 Υλοποίηση του Graph API σε Java

Το Facebook Graph API είναι ένα εργαλείο που μπορεί να χρησιμοποιηθεί τόσο σε εφαρμογές που εκτελούνται σε κάποια ιστοσελίδα του Facebook (Canvas Page) όσο και σε ανεξάρτητους ιστοχώρους (websites) που απλώς ενσωματώνουν δεδομένα και λειτουργίες του Facebook. Είναι δυνατόν επίσης να χρησιμοποιηθεί και για εφαρμογές κινητών συσκευών με λειτουργικό iOS ή Android. Σε κάθε περίπτωση, πριν από τις κλήσεις στο Graph API θα πρέπει να υλοποιηθεί η ροή σύνδεσης ώστε στο τέλος της διαδικασίας εξουσιοδότησης να δημιουργηθεί το κατάλληλο διακριτικό πρόσβασης. Το Facebook παρέχει διαφορετικές ροές σύνδεσης ανάλογα με τα projects που πρέπει να υλοποιηθούν και τις συσκευές στις οποίες απευθύνονται.

Για τη δημιουργία Web εφαρμογών υπάρχουν δύο ειδών ροές σύνδεσης. Η πιο εύκολη ροή σύνδεσης είναι η σύνδεση που βασίζεται στο φυλλομετρητή (browser-based). Για εφαρμογές που η εν λόγω ροή σύνδεσης δεν είναι κατάλληλη υπάρχει και η ροή σύνδεσης που χρησιμοποιεί εξ' ολοκλήρου server-side κώδικα. Και στις δύο περιπτώσεις διαπιστώνεται αν ο χρήστης είναι συνδεδεμένος στο Facebook με το λογαριασμό του και αν έχει εξουσιοδοτήσει την εφαρμογή χορηγώντας της τα κατάλληλα δικαιώματα. Εάν ο χρήστης είναι συνδεδεμένος με το Facebook αλλά δεν έχει εξουσιοδοτήσει την εφαρμογή, ανακατευθύνεται σε ένα παράθυρο διαλόγου (dialog)³ το οποίο εμφανίζει τα δικαιώματα που πρέπει να χορηγηθούν στην εφαρμογή και ζητά την έγκριση του χρήστη. Εάν ο χρήστης δεν είναι συνδεδεμένος ανακατευθύνεται σε ένα παράθυρο διαλόγου που εμφανίζει φόρμες καταχώρησης του ονόματος χρήστη και του κωδικού πρόσβασης ώστε να πραγματοποιηθεί η σύνδεση. Εν συνεχεία, ελέγχεται αν έχει πραγματοποιηθεί η εξουσιοδότηση. Η διαφορά ανάμεσα στις δύο ροές έγκειται στον τρόπο με τον οποίο παράγεται το διακριτικό πρόσβασης. Εάν ο χρήστης είναι συνδεδεμένος και ταυτόχρονα έχει χορηγήσει δικαιώματα στην εφαρμογή πραγματοποιείται μια διαδικασία για να εξασφαλιστεί ότι το άτομο που εξουσιοδότησε την εφαρμογή είναι το ίδιο με εκείνο που ξεκίνησε τη ροή σύνδεσης. Στην πρώτη περίπτωση, η ροή εκτελεί αυτόματα τους δικούς της ελέγχους ασφαλείας και συνεπώς δεν απαιτείται κάποια χειροκίνητη χρήση των διακριτικών πρόσβασης από τον προγραμματιστή της εφαρμογής. Αντίθετα η δεύτερη περίπτωση περιλαμβάνει συγγραφή κώδικα για την επιβεβαίωση της ταυτότητας του χρήστη. Στην εν λόγω ροή, οι χρήστες ανακατευθύνονται στο URL της εφαρμογής με παράμετρο έναν oauth κωδικό (oauth code) ο οποίος παράγεται από το ίδιο το Facebook. Στη συνέχεια, η εφαρμογή ανταλλάσσει αυτόν τον κωδικό με ένα διακριτικό πρόσβασης.

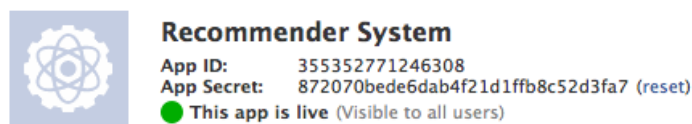
³ Το παράθυρο διαλόγου είναι μια απλή διεπαφή που επιτρέπει σε μια εφαρμογή να αλληλεπιδρά με ένα χρήστη. Πρόκειται για ένα πλαίσιο που εμφανίζει κάποιο μήνυμα στην οθόνη του χρήστη σχετικά με μια ενέργεια (λόγου χάρι την εξουσιοδότηση μιας εφαρμογής, την αποστολή ενός αιτήματος φιλίας ή ενός μηνύματος σε έναν άλλο χρήστη κ.λπ.). Ανάλογα με την περίπτωση υπάρχουν διάφορες επιλογές αποδοχής ή απόρριψης του περιεχομένου ενός παραθύρου διαλόγου.

Οποιαδήποτε ροή σύνδεσης και αν χρησιμοποιηθεί, είναι απαραίτητη η δημιουργία μιας Facebook εφαρμογής και η ανάθεση των κατάλληλων ρυθμίσεων σε αυτήν. Η Facebook εφαρμογή δημιουργείται επιλέγοντας την καρτέλα Apps η οποία βρίσκεται στο μενού επιλογών του ιστοχώρου <https://developers.facebook.com>. Προκειμένου να χρησιμοποιήσει ένας χρήστης αυτήν την επιλογή θα πρέπει να εγγραφεί στο Facebook ως προγραμματιστής (developer). Για την υλοποίηση του συστήματος που περιγράφεται στην παρούσα διπλωματική δημιουργήθηκε μια εφαρμογή με όνομα Recommender System (βλ. Εικόνα 13).



Εικόνα 13: Δημιουργία εφαρμογής

Η επιτυχής δημιουργία της εφαρμογής θα έχει ως αποτέλεσμα την ανάκτηση δύο κωδικών με ονόματα App ID και App Secret αντίστοιχα όπως φαίνεται στην Εικόνα 14:



Εικόνα 14: Κωδικοί εφαρμογής

Το App ID είναι ένας μοναδικός ακεραίος αριθμός (αναγνωριστικό) για κάθε εφαρμογή και χρησιμοποιείται για την πραγματοποίηση κλήσεων στο γράφημα του Facebook αλλά και σε ορισμένα προγραμματιστικά εργαλεία του. Το App Secret είναι συμβολοσειρά που συνδέεται με ένα App ID. Χρησιμεύει στην αποκωδικοποίηση κρυπτογραφημένων μηνυμάτων από το Facebook (τα μηνύματα κρυπτογραφούνται προκειμένου να προστατευτούν τα ιδιωτικά δεδομένα των χρηστών από κλοπή) και έχει ως στόχο να εξασφαλίσει την αξιόπιστη χρήση της εφαρμογής. Μπορεί να δημιουργήσει διακριτικά πρόσβασης με εύκολο τρόπο σε ορισμένες ροές σύνδεσης.

Μαζί με τους κωδικούς εφαρμογών εμφανίζονται και διάφορες επιλογές ρυθμίσεων. Εάν πρόκειται για μια Facebook εφαρμογή θα πρέπει να καθοριστούν οι κατάλληλες ρυθμίσεις στο πεδίο "App on Facebook" των βασικών ρυθμίσεων ενώ αν πρόκειται για δικτυακό τόπο με χρήση σύνδεσης Facebook θα πρέπει να συμπληρωθεί το πλήρες URL του ιστοχώρου. Ανάλογες επιλογές ρυθμίσεων υπάρχουν και για εφαρμογές κινητών συσκευών.

Στα πλαίσια της εργασίας υλοποιήθηκε ένας ανεξάρτητος ιστοχώρος ο οποίος ανακτά το αναγνωριστικό (id) του χρήστη στο Facebook, τη λίστα των φίλων του, τις δημοσιεύσεις και τα σχόλιά του σε μια ειδική Facebook σελίδα δικής μας κατασκευής καθώς και τα like που έχει κάνει όσον αφορά προτιμήσεις του σε βιβλία, μουσική και ταινίες. Η εφαρμογή Recommender System ζητά από το χρήστη τα δικαιώματα user_likes και friends_likes προκειμένου να έχει στη διάθεσή της ένα κατάλληλο διακριτικό πρόσβασης χρήστη. Για την υλοποίηση των API κλήσεων χρησιμοποιήθηκε η ανεπίσημη JAVA βιβλιοθήκη facebook4j. Το facebook 4j περιλαμβάνει αμιγώς java κώδικα και κατόπιν αυτού χρησιμοποιήθηκε για λόγους συμβατότητας με το υπόλοιπο σύστημα που είναι υλοποιημένο σε jsp.

Για τη χρήση του facebook4j πρέπει αρχικά να αποκτηθεί ένα Facebook στιγμιότυπο (Facebook instance). Το στιγμιότυπο αυτό αρχικοποιείται με την κλάση ConfigurationBuilder και εν συνεχεία δημιουργείται με τη μέθοδο getInstance() της κλάσης FacebookFactory ως ακολούθως:

```
ConfigurationBuilder cb = new ConfigurationBuilder();
cb.setDebugEnabled(true)
.setOAuthAppId("*****")
.setOAuthAppSecret("*****")
.setOAuthAccessToken("*****")
.setOAuthPermissions("user_likes, friends_likes");
FacebookFactory ff = new FacebookFactory(cb.build());
Facebook facebook = ff.getInstance();
```

Ο παραπάνω κώδικας αναθέτει στο Facebook στιγμιότυπο τα App ID και App Secret της εφαρμογής, τα δικαιώματα που ζητούνται από το χρήστη σαν συμβολοσειρές διαχωρισμένες με κόμμα και ένα έγκυρο διακριτικό πρόσβασης. Το διακριτικό πρόσβασης μπορεί είτε να ανατεθεί στην παρούσα φάση αν υπάρχει διαθέσιμο από την πλατφόρμα του Facebook κάποιο έγκυρο διακριτικό πρόσβασης είτε μπορεί να παραλειφθεί από το στιγμιότυπο και να προκύψει από την υλοποίηση της ροής σύνδεσης. Το Facebook στιγμιότυπο καταστρέφεται με τη μέθοδο shutdown() η οποία απελευθερώνει ταυτόχρονα τους πόρους που διατίθενται.

Στη συνέχεια περιγράφονται συνοπτικά οι λειτουργίες μερικών συναρτήσεων της facebook4j βιβλιοθήκης που χρησιμοποιούνται για την υλοποίηση του Graph API.

getId(): Η μέθοδος αυτή επιστρέφει το μοναδικό αναγνωριστικό (id) ενός αντικειμένου στο κοινωνικό γράφημα του Facebook. Το αναγνωριστικό που επιστρέφεται είναι μια συμβολοσειρά. Όταν πρόκειται για το αντικείμενο “χρήστης”, η μέθοδος αυτή μπορεί εσωτερικά να καλέσει το μονοπάτι “/me”. Η getId() χρησιμοποιείται αρχικά για την ανάκτηση του αναγνωριστικού του τρέχοντος χρήστη στο Facebook και με αυτό ενημερώνεται το πεδίο node του πίνακα user_details. Υπενθυμίζεται ότι το πεδίο node μένει κενό κατά την εγγραφή ενός χρήστη στο σύστημα και παίρνει τιμή μόνο εάν επιλέξει τον αλγόριθμο σύστασης που υλοποιείται στην παρούσα διπλωματική. Άλλες χρήσεις της μεθόδου είναι η ανάκτηση του αναγνωριστικού μιας δημοσίευσης ή ενός αντικειμένου like.

getFriends(): Επιστρέφει μια λίστα με τους φίλους του τρέχοντος χρήστη ή FacebookException όταν η υπηρεσία ή το δίκτυο του Facebook δεν είναι διαθέσιμα. Η μέθοδος χρησιμοποιείται για την ενημέρωση του πίνακα friends. Στα πεδία node1 και node2 του πίνακα αποθηκεύονται το αναγνωριστικό του τρέχοντα χρήστη και το

αναγνωριστικό του φίλου του. Για κάθε σχέση φιλίας μεταξύ δύο χρηστών δημιουργούνται δύο εγγραφές, μία στην οποία το αναγνωριστικό του τρέχοντος χρήστη βρίσκεται στο πεδίο `node1` και άλλη μία στην οποία το αναγνωριστικό αυτό θα βρίσκεται στο πεδίο `node2`. Η προσέγγιση αυτή διατηρεί κάθε φορά την κατεύθυνση του κοινωνικού δεσμού και διασπά την αμφίδρομη σχέση φιλίας σε δύο αντίθετες μονόδρομες σχέσεις έτσι ώστε και τα δύο πεδία να περιλαμβάνουν ανά πάσα στιγμή όλους τους χρήστες. Ο πίνακας `friends` ενημερώνεται σε κάθε εκτέλεση με τους τρέχοντες φίλους του χρήστη. Με αυτόν τον τρόπο, εάν ο χρήστης έχει μπλοκάρει ορισμένες επαφές του, ο πίνακας δε θα τις περιλαμβάνει πλέον.

getPosts(): Επιστρέφει μια λίστα με τις δημοσιεύσεις (`posts`) του τρέχοντος αντικειμένου ή `FacebookException` όταν η υπηρεσία ή το δίκτυο του Facebook δεν είναι διαθέσιμα. Η μέθοδος μπορεί να κληθεί είτε χωρίς ορίσματα είτε με διάφορες παραμέτρους όπως το αναγνωριστικό του αντικειμένου, διάφορα άλλα πεδία του, κάποια σχετική ημερομηνία, κ.λπ.

getPostComments(): Επιστρέφει μια λίστα με όλα τα σχόλια (`comments`) μιας δημοσίευσης ή `FacebookException` όταν η υπηρεσία ή το δίκτυο του Facebook δεν είναι διαθέσιμα. Ομοίως, η μέθοδος μπορεί να κληθεί με ή χωρίς ορίσματα. Η ανάκτηση του χρήστη που έκανε το σχόλιο γίνεται με την κλήση της μεθόδου `getFrom()` σε ένα σχόλιο ενώ η ανάκτηση του μηνύματός του γίνεται με τη μέθοδο `getMessage()`.

getUserLikes(): Επιστρέφει μια λίστα με όλες τις σελίδες στις οποίες ο τρέχων χρήστης έχει κάνει `like` ή `FacebookException` όταν η υπηρεσία ή το δίκτυο του Facebook δεν είναι διαθέσιμα. Και εδώ η μέθοδος μπορεί να κληθεί με ή χωρίς ορίσματα. Η ανάκτηση της κατηγορίας στην οποία ανήκει το `like` αντικείμενο γίνεται με την κλήση της μεθόδου `getCategory()`.

getOAuthAuthorizationURL(String callbackURL): Επιστρέφει το URL για ένα παράθυρο διαλόγου. Δέχεται σαν όρισμα το URL που περιλαμβάνει τον κώδικα για την υλοποίηση της ροής σύνδεσης.

getOAuthAccessToken(String oauthCode): Ανταλλάζει τον `oauth` κωδικό του Facebook με ένα έγκυρο διακριτικό πρόσβασης. Επιστρέφει ένα διακριτικό πρόσβασης χρήστη ή `FacebookException` όταν η υπηρεσία ή το δίκτυο του Facebook δεν είναι διαθέσιμα ή εάν ο χρήστης δεν έχει εξουσιοδοτήσει την εφαρμογή.

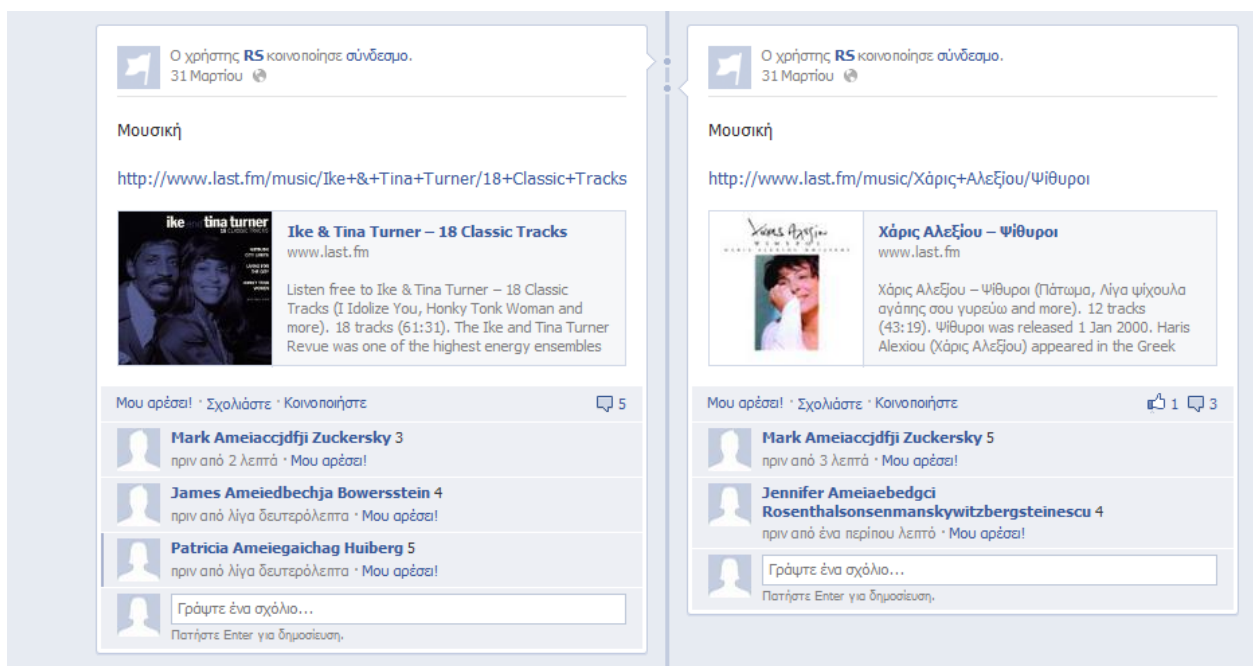
4.1.5 Αξιολογήσεις χρηστών

Προκειμένου το σύστημά μας να περιλαμβάνει μια υποδομή ανάδρασης χρήστη, δημιουργήσαμε μια σελίδα στο Facebook (βλ. Εικόνα 15) στην οποία αναρτώνται οι αριθμητικές αξιολογήσεις που δίνουν οι χρήστες μας για τα προϊόντα μιας εικονικής αγοράς. Η σελίδα βρίσκεται στο URL:

<https://www.facebook.com/pages/RS/464696976933635>

Σε αυτόν τον ιστότοπο, έχει αναρτηθεί μια δημοσίευση για κάθε προϊόν. Το αναγνωριστικό κάθε δημοσίευσης αποθηκεύεται σαν αναγνωριστικό προϊόντος στο πεδίο `id` του πίνακα `bmm_details`. Οι χρήστες βαθμολογούν ένα προϊόν γράφοντας έναν ακέραιο αριθμό σαν σχόλιο κάτω από την αντίστοιχη δημοσίευση. Ο αριθμός αυτός και γενικότερα το μήνυμα κάθε σχολίου αποθηκεύεται στο πεδίο `message` του σχολίου. Ο αλγόριθμος σύστασης διαβάσει διαδοχικά όλες τις δημοσιεύσεις της σελίδας και ελέγχει εάν το αναγνωριστικό της δημοσίευσης βρίσκεται αποθηκευμένο σε κάποια εγγραφή του πίνακα `bmm_details`. Εάν ισχύει κάτι τέτοιο, διαβάζονται τα σχόλια της δημοσίευσης ενώ σε αντίθετη περίπτωση η δημοσίευση αγνοείται. Ο έλεγχος αυτός γίνεται καθώς η

σελίδα διαθέτει και δημοσιεύσεις με οδηγίες οπότε πρέπει να γίνεται διάκριση ανάμεσα στις δημοσιεύσεις που αναπαριστούν προϊόντα από τις υπόλοιπες. Αναφορικά με τα σχόλια, ο αλγόριθμος ελέγχει αν είναι ακέραιοι αριθμοί από το 1 έως το 5. Σε περίπτωση που το σχόλιο είναι αποδεκτό, συμπληρώνεται μια εγγραφή στον πίνακα `user_ratings`. Στο πεδίο `node` του πίνακα τοποθετείται το αναγνωριστικό του χρήστη που έκανε το σχόλιο, στο πεδίο `itemid` τοποθετείται το αναγνωριστικό της αντίστοιχης δημοσίευσης, στο πεδίο `rating` αποθηκεύεται το σχόλιο και στο πεδίο `prediction` τίθεται η τιμή 0 που υποδηλώνει ότι πρόκειται για αξιολόγηση που λαμβάνεται σαν δεδομένο εισόδου στον αλγόριθμο. Το σχόλιο αγνοείται σε περίπτωση που δεν έχει την επιθυμητή μορφή.



Εικόνα 15: Η σελίδα αξιολογήσεων

4.2 Περιγραφή του αλγορίθμου

Όπως έχει αναφερθεί, το υποσύστημα RE είναι υπεύθυνο για την εκτέλεση του αλγορίθμου σύστασης. Για την εκπόνηση της παρούσας διπλωματικής εργασίας χρησιμοποιήθηκε ο αλγόριθμος που υλοποιεί την τεχνική του συνεργατικού φιλτραρίσματος (CF) και ειδικότερα μιας παραλλαγής του φιλτραρίσματος με βάση το χρήστη (user-based). Οι διαφοροποιήσεις έγκεινται στο γεγονός ότι στον υπολογισμό των βαρών ανάμεσα στους χρήστες προσμετρούνται εκτός από τις αριθμητικές αξιολογήσεις (ratings), μετρικές που προκύπτουν από τεχνικές ομοιότητας για τα χαρακτηριστικά των χρηστών. Η βασική διαφορά είναι ότι αντί να αναθέσουμε ένα βάρος σε όλους τους χρήστες του συστήματος προκειμένου να βρεθούν ομοιότητες με το χρήστη-στόχο και έτσι να προκύψουν τελικά οι γείτονές του, θεωρούμε εξαρχής ότι γείτονές του αποτελούν όλες οι επαφές του (φίλοι) στο Facebook. Ο ισχυρισμός αυτός βασίζεται στο γεγονός ότι ανάμεσα σε δύο ανθρώπους που είναι φίλοι υπάρχει μεγάλη πιθανότητα να εντοπιστούν εκ των πραγμάτων κοινά γνωρίσματα. Από το πλήθος των φίλων επιλέγουμε ένα υποσύνολο αυτού που ταιριάζουν περισσότερο με το χρήστη στόχο. Το εν λόγω υποσύνολο θα αποτελέσει τους τελικούς συστήνοντες (recommenders) από τους οποίους θα προκύψει το αποτέλεσμα του αλγορίθμου.

4.2.1 Ομοιότητα χρηστών

Προκειμένου να υπολογίσουμε το βαθμό στον οποίο δύο φίλοι εμφανίζουν ομοιότητες μεταξύ τους, ορίζουμε τέσσερις μετρικές για κάθε κοινωνικό δεσμό που διαμορφώνουν το συνολικό του βάρος. Όλες οι μετρικές αφού υπολογιστούν αποθηκεύονται στα αντίστοιχα πεδία του πίνακα friends. Αυτές είναι οι εξής:

Συντελεστής συσχετισμού Pearson (Pearson Correlation coefficient): Αυτός ο συντελεστής χρησιμοποιείται ευρέως στη στατιστική και δείχνει το βαθμό στον οποίο δύο μεταβλητές συσχετίζονται γραμμικά μεταξύ τους. Όσον αφορά τη χρήση του στον αλγόριθμο σύστασης, αποτελεί ένα μέτρο συσχέτισης ως προς τις αξιολογήσεις που δίνουν δύο φίλοι στα ίδια αντικείμενα και δίνεται από την εξίσωση (1):

$$w(u,v) = \frac{\sum_{i \in I} (R_{u,i} - \bar{R}_u)(R_{v,i} - \bar{R}_v)}{\sqrt{\sum_{i \in I} (R_{u,i} - \bar{R}_u)^2 \sum_{i \in I} (R_{v,i} - \bar{R}_v)^2}} \quad (1)$$

όπου $R_{u,i}$ είναι η αξιολόγηση του χρήστη στόχου u στο αντικείμενο i , $R_{v,i}$ είναι η αξιολόγηση του γείτονα του χρήστη στόχου u στο αντικείμενο i , \bar{R}_u και \bar{R}_v είναι οι μέσες αξιολογήσεις του χρήστη-στόχου u και κάθε γείτονα του v αντίστοιχα και $w(u,v)$ είναι η συσχέτιση (βάρος) μεταξύ των χρηστών u και v . Σημειώνεται επίσης ότι $u \in U$ όπου U είναι το σύνολο όλων των χρηστών και αξιολογητών, $v \in V$ όπου V είναι το σύνολο των γειτόνων και $V \subseteq U$. Επίσης, $i \in I$ όπου I το σύνολο των αντικειμένων που διατίθενται στο σύστημα τα οποία έχουν αξιολογηθεί και από τους δύο χρήστες u,v .

Από τον ανωτέρω μαθηματικό τύπο προκύπτουν τιμές στο διάστημα $[-1,1]$. Προκειμένου να αποκλεισθούν οι αρνητικές τιμές καθώς θέλουμε όλες οι μετρικές που θα υπολογιστούν στο σύστημα να έχουν το ίδιο θετικό διάστημα τιμών, κάνουμε το

μετασχηματισμό $\frac{1+w(u,v)}{2}$ ώστε ο συντελεστής συσχέτισης να λαμβάνει τιμές στο διάστημα $[0,1]$.

Ένας συσχετισμός με τιμή 1 υποδηλώνει ότι οι δύο φίλοι έχουν πλήρη ταύτιση ενώ η τιμή μηδέν συνεπάγεται πλήρη αναντιστοιχία. Αυτό σημαίνει ότι στην πρώτη περίπτωση ο φίλος αποτελεί έναν τέλειο συστήνοντα (recommender) ενώ στη δεύτερη η γνώση των αξιολογήσεων του εν λόγω γείτονα δε δίνει καμία απολύτως πληροφορία για το χρήστη στόχο.

Συντελεστής συσχέτισης δημογραφικών χαρακτηριστικών (DFactor): Αυτή η μετρική δείχνει σε τι βαθμό δύο χρήστες ταιριάζουν ως προς τα δημογραφικά τους στοιχεία. Οι πληροφορίες που λαμβάνονται υπόψη για τη διαμόρφωση αυτής της μετρικής είναι η ηλικία, το φύλο, η περιοχή διαμονής και το επάγγελμα που ασκούν οι χρήστες. Ως χρήστες με ίδια ηλικία θεωρούνται εκείνοι που έχουν διαφορά έως και μία πενταετία. Εάν δεν υπάρχουν πληροφορίες για τα στοιχεία των χρηστών είτε οι χρήστες δεν ταιριάζουν καθόλου ο DFactor παίρνει την τιμή 0 ενώ αντίστοιχα τιμή 1 υποδηλώνει και πάλι πλήρη ταύτιση. Το βάρος που προκύπτει από τα δημογραφικά στοιχεία υπολογίζεται από το μέσο όρο του πλήθους των κοινών δημογραφικών μεταξύ του χρήστη στόχου και του φίλου του δηλαδή, σύμφωνα με την εξίσωση (2), θα ισχύει:

$$DFactor = \frac{\#κοινων_δημογραφικων}{\#συνολικων_δημογραφικων} \quad (2)$$

Συντελεστές συσχέτισης προτιμήσεων: Σε αυτήν την περίπτωση ορίζονται μετρικές για κάθε κατηγορία προϊόντος οι οποίες δείχνουν σε τι βαθμό ταιριάζουν οι χρήστες ως προς τις προτιμήσεις τους σε συγκεκριμένα χαρακτηριστικά αντικειμένων ανά κατηγορία. Εφόσον το σύστημα απευθύνεται σε μια εικονική αγορά που αποτελείται από τρεις κατηγορίες προϊόντων (βιβλία, ταινίες και μουσικά άλμπουμ), ορίζονται αντίστοιχα οι εξής τρεις μετρικές:

- **MusicFactor:** δείχνει σε τι βαθμό ταιριάζουν οι χρήστες ως προς τις μουσικές τους προτιμήσεις. Τα χαρακτηριστικά αντικειμένου που διαμορφώνουν την τιμή αυτής της μετρικής είναι το είδος της μουσικής που αρέσει στο χρήστη (πχ κλασική, rock, pop κ.λπ.) και οι αγαπημένοι του τραγουδιστές (Έλληνες ή ξένοι).
- **BookFactor:** δείχνει σε τι βαθμό ταιριάζουν οι χρήστες ως προς τις προτιμήσεις τους σε χαρακτηριστικά βιβλίων. Τα χαρακτηριστικά που λαμβάνονται υπόψη στον υπολογισμό της μετρικής είναι το είδος του βιβλίου (πχ μυθιστόρημα, ιστορικό κ.λπ.), οι αγαπημένοι συγγραφείς του χρήστη καθώς και οι εκδόσεις που τον ενδιαφέρουν.
- **MovieFactor:** δείχνει σε τι βαθμό ταιριάζουν οι χρήστες ως προς τις προτιμήσεις τους σε χαρακτηριστικά ταινιών. Στην περίπτωση αυτή, η τιμή της μετρικής προκύπτει από τις προτιμήσεις ως προς το είδος της ταινίας (πχ κωμωδία, δράμα κ.λπ.) και τους αγαπημένους ηθοποιούς και σκηνοθέτες του χρήστη.

Η τιμή κάθε μίας από τις παραπάνω μετρικές προκύπτει με την ακόλουθη διαδικασία: αρχικά υπολογίζεται το πλήθος των κοινών προτιμήσεων μεταξύ του χρήστη στόχου και του γείτονά του για την κατηγορία προϊόντος που αντιστοιχεί στην κάθε μετρική. Εν συνεχεία υπολογίζεται το πλήθος των συνολικών προτιμήσεων της κατηγορίας για το χρήστη στόχο και το γείτονά του και διαιρούμε τον αριθμό των κοινών προτιμήσεων με το μικρότερο πλήθος συνολικών προτιμήσεων των δύο, δηλαδή δίνεται από την εξίσωση (3):

$$XFactor = \frac{\#κοινων_προτιμησεων_X_{(u,v)}}{\min(\#προτιμησεων_X_u, \#προτιμησεων_X_v)} \quad (3)$$

όπου X η κατηγορία προϊόντος από το σύνολο {Music, Movie, Book}.

Συντελεστές συσχέτισης likes: Ο συντελεστής αυτός υπολογίζει ομοιότητες χρηστών ως προς τις σελίδες Facebook στις οποίες κάνουν like οι χρήστες. Και σε αυτήν την περίπτωση ορίζονται διαφορετικές μετρικές ανάλογα με την κατηγορία στην οποία ανήκει το συγκεκριμένο like. Προκειμένου να υπάρχει αντιστοιχία με το συντελεστή συσχέτισης προτιμήσεων, ορίζονται αντίστοιχα οι εξής τρεις μετρικές:

- **MusicLikes:** δείχνει σε τι βαθμό ταιριάζουν οι χρήστες ως προς τα like σε σελίδες μουσικής. Τα like που διαμορφώνουν την τιμή αυτής της μετρικής πρέπει να έχουν στο πεδίο Category τις τιμές musician/band, song ή album.
- **BookLikes:** δείχνει σε τι βαθμό ταιριάζουν οι χρήστες ως προς τα like σε σελίδες βιβλίων. Τα like που λαμβάνονται υπόψη στον υπολογισμό της μετρικής είναι αυτά που έχουν στο πεδίο Category τις τιμές book ή author.
- **MovieLikes:** δείχνει σε τι βαθμό ταιριάζουν οι χρήστες ως προς τα like τους σε σελίδες ταινιών. Στην περίπτωση αυτή λαμβάνονται υπόψη τα like που έχουν στο πεδίο Category μία από τις τιμές movie, actor/director.

Η τιμή για κάθε μία από τις παραπάνω μετρικές προκύπτει με την ίδια διαδικασία που ακολουθείται και για τον υπολογισμό των μετρικών συσχέτισης προτιμήσεων και δίνεται από την εξίσωση (4):

$$XLikes = \frac{\#κοινων_Likes_X_{(u,v)}}{\min(\#Likes_X_u, \#Likes_X_v)} \quad (4)$$

Αφού βρεθούν οι τιμές όλων των ανωτέρω μετρικών, διαμορφώνεται ένα βάρος ανά κατηγορία προϊόντος για κάθε κοινωνικό δεσμό που δείχνει συνολικά τη συσχέτιση μεταξύ των ατόμων που τον απαρτίζουν. Τα τρία βάρη δίνονται από τους ακόλουθους τύπους (5), (6) και (7):

$$Z_1 = w_1 \cdot PearsonResult + w_2 \cdot DFactor + w_3 \cdot MusicFactor + w_4 \cdot MusicLikes \quad (5)$$

$$Z_2 = w_1 \cdot PearsonResult + w_2 \cdot DFactor + w_3 \cdot BookFactor + w_4 \cdot BookLikes \quad (6)$$

$$Z_3 = w_1 \cdot PearsonResult + w_2 \cdot DFactor + w_3 \cdot MovieFactor + w_4 \cdot MovieLikes \quad (7)$$

όπου $\sum_{i=1}^4 w_i = 1$. Τα w_1 , w_2 , w_3 και w_4 είναι παράμετροι που δείχνουν σε τι ποσοστό συμμετέχει η κάθε μετρική στο συνολικό βάρος και οι τιμές τους διαμορφώνονται ανάλογα με το πού θέλουμε να δώσουμε έμφαση.

Η παραπάνω προσέγγιση υιοθετήθηκε με το σκεπτικό ότι δύο φίλοι ενδέχεται να ταιριάζουν απόλυτα στις προτιμήσεις τους όσον αφορά μια συγκεκριμένη κατηγορία προϊόντος αλλά να μην υπάρχει συσχέτισμός σε κάποια άλλη κατηγορία. Μπορεί δηλαδή να ταιριάζουν λόγου χάρη στα είδη μουσικής που ακούν αλλά να μην ταιριάζουν στα είδη ταινιών που επιθυμούν να παρακολουθούν. Ορίζοντας διαφορετικά βάρη ανά κατηγορία επιδιώκεται να παρέχονται από το σύστημα πιο εύστοχες συστάσεις.

4.2.2 Τελικός υπολογισμός συστάσεων

Μοντελοποιούμε το πρόβλημα σύστασης ορίζοντας ένα γράφο $G=(U,E,W)$ όπου U είναι το σύνολο των κόμβων, εν προκειμένω οι χρήστες που απαρτίζουν τον κοινωνικό γράφο, E είναι το σύνολο των ακμών που αντικατοπτρίζουν τις κοινωνικές σχέσεις (φιλίες) μεταξύ των χρηστών και W είναι το βάρος κάθε κοινωνικού δεσμού. Σημειώνεται ότι ο γράφος διαμορφώνεται σταδιακά με κάθε διαδοχική εκτέλεση του αλγορίθμου. Κάθε διαφορετικός χρήστης που εγγράφεται στο σύστημα προσθέτει επιπλέον ακμές στο γράφημα από τις πληροφορίες που λαμβάνει το σύστημα σχετικά με τους κοινωνικούς δεσμούς του. Ο σχεδιασμός του συστήματος έχει γίνει με την παραδοχή ότι ο γράφος που σχηματίζεται είναι συνδεδεμένος και ενώνει όλους τους χρήστες του συστήματος έστω και έμμεσα αντικατοπτρίζοντας μια ευρεία παρέα φίλων. Καθώς έχουν οριστεί τρία βάρη, αντίστοιχα θα δημιουργηθούν τελικά τρεις τέτοιοι γράφοι G_i με $i=1,2,3$ δηλαδή ένας γράφος για κάθε κατηγορία προϊόντος. Για κάθε επιμέρους γράφο, υπολογίζουμε ένα Δέντρο Επικάλυψης Ελαχίστου Κόστους – ΔΕΕΚ (Minimum Spanning Tree) από το οποίο θα προκύψουν τελικά οι συστήνοντες (recommenders). Για το σκοπό αυτό, πάνω στις ακμές κάθε γράφου ανατίθεται το $1/Z_i$ όπου Z_i το αντίστοιχο βάρος ανάλογα με την κατηγορία. Όσο πιο μικρό είναι το $1/Z_i$ τόσο μεγαλύτερη θα είναι η ομοιότητα ανάμεσα στους χρήστες που συνδέει η ακμή. Εφόσον $Z_i \in (0,1]$ συνεπάγεται ότι $1/Z_i \geq 1$ και επομένως ο σταθμισμένος γράφος θα

περιλαμβάνει μόνο θετικά βάρη. Για $Z_i = 0$, το $1/Z_i \rightarrow \infty$ οπότε σε αυτήν την περίπτωση αναθέτουμε στο $1/Z_i$ μια πολύ μεγάλη τιμή.

Για την υλοποίηση του ΔΕΕΚ επιλέχθηκε ο αλγόριθμος Prim πλησιέστερου γείτονα (nearest neighbour) [37]. Ο εν λόγω αλγόριθμος έχει πολυπλοκότητα $O(n^2)$. Συγκριτικά με τον αλγόριθμο Kruskal που επίσης χρησιμοποιείται συχνά για υπολογισμούς ΔΕΕΚ, παρουσιάζει καλύτερη απόδοση σε πυκνούς γράφους. Ειδικότερα ο αλγόριθμος Kruskal έχει πολυπλοκότητα $O(m \log n)$. Για πυκνούς γράφους ισχύει ότι $m = O(n^2)$ οπότε η πολυπλοκότητα του Kruskal θα είναι $O(n^2 \log n) > O(n^2)$ για $n \geq 2$. Επιπρόσθετα, ο αλγόριθμος Prim εμφανίζει το πλεονέκτημα ότι σε κάθε βήμα του παράγει ένα δένδρο ενώ αντιθέτως ο Kruskal παράγει δάσος από δένδρα. Συνεπώς, σε περίπτωση πρόωρης διακοπής του αλγορίθμου σύστασης, ο Prim θα δώσει σαν αποτέλεσμα ένα υποτυπώδες δένδρο ενώ ο Kruskal θα εμφανίσει σφάλμα μη συνδεδεμένου γράφου οδηγώντας στη διακοπή της εκτέλεσης του αλγορίθμου σύστασης.

Ο αλγόριθμος ολοκληρώνεται προβλέποντας τις αξιολογήσεις που θα έδινε ο χρήστης στόχος στα αντικείμενα που δεν έχει αξιολογήσει έως την τρέχουσα εκτέλεση. Η πρόβλεψη της αξιολόγησης του ενεργού χρήστη u για το αντικείμενο i προκύπτει από ένα σταθμισμένο συνδυασμό των αξιολογήσεων (ratings) των επιλεγμένων γειτόνων-φίλων σύμφωνα με την εξίσωση (8):

$$p_{u,i} = \bar{R}_u + \frac{\sum_{k \in K} (R_{k,i} - \bar{R}_k) \cdot w_{u,k}}{\sum_{k \in K} w_{u,k}} \quad (8)$$

όπου $w_{u,v}$ είναι η ομοιότητα μεταξύ των χρηστών u και v δηλαδή εν προκειμένω το αντίστοιχο $1/Z_i$ της ακμής του γράφου που συνδέει τους δύο χρήστες και K είναι το σύνολο των πιο όμοιων χρηστών οι οποίοι έχουν προκύψει από το ΔΕΕΚ ενώ ισχύει $K \subseteq V$. Τα αντικείμενα που τελικά προτείνονται στον ενεργό χρήστη είναι εκείνα που έχουν προβλεπόμενη αξιολόγηση μεγαλύτερη από 3.

Παρακάτω παρατίθεται ο αλγόριθμος, έστω SNB_algorithm (Social-Network Based algorithm) σε μορφή ψευδοκώδικα (βλ. Αλγόριθμος 1). Για την εκτέλεσή του υποθέτουμε ότι υπάρχει ένας χρήστης-στόχος u με ένα σύνολο φίλων $F = \{f_1, f_2, \dots, f_v\}$ καθώς και ένα σύνολο αντικειμένων $I = \{i_1, i_2, \dots, i_v\}$. Ο αλγόριθμος διαβάζει επιπλέον τα δεδομένα ενός γράφου $G = (U, V, W)$ ο οποίος μέχρι την πρώτη εκτέλεση είναι κενός και ανανεώνεται με τα δεδομένα των χρηστών σε κάθε επόμενη εκτέλεση. Τέλος, θεωρούμε ότι υπάρχει ένα σύνολο $S = \{s_1, s_2, \dots, s_v\}$ από χρήστες κοινωνικού δικτύου. Ο SNB_algorithm διαβάζει τις αξιολογήσεις r κάθε τέτοιου χρήστη για κάθε αντικείμενο και τις αποθηκεύει. Για ευκολία, παραλείπεται από τον ψευδοκώδικα ο διαχωρισμός των αντικειμένων σε κατηγορίες και η δημιουργία των τριών ΔΕΕΚ.

SNB_algorithm

Input: $G=(U,E,W)$, I , S , u , $r_{s,i}$

Output: $r_{u,i}$

$W_u = \emptyset$, $E_u = \emptyset$, $V_u = \emptyset$, $T(U,V,W) = \emptyset$, $sum_of_w = 0$, $weighted_sum = 0$

for every s_u in S do

$ratings[s_u, i] \leftarrow r_{s_u, i}$;

end for

$F \leftarrow getFriends(u)$;

$V_u \leftarrow u \cup F$;

for every f_j in F do

$Z = w_1 \bullet PearsonResult(u, f_j) + w_2 \bullet DFactor(u, f_j) + w_3 \bullet XFactor(u, f_j) + w_4 \bullet XLikes(u, f_j)$;

$E_u \leftarrow E_u \cup (u, f_j)$;

$W_u \leftarrow W_u \cup \frac{1}{Z}$;

end for

$U \leftarrow U \cup V_u$;

$E \leftarrow E \cup E_u$;

$W \leftarrow W \cup W_u$;

$T(U, E, W) \leftarrow prim\{G(U, E, W)\}$;

for every i in I that is not rated by u do

for every t in $T(U, V, W)$ do

if $t = \{(u, f_j, w_{u, f_j})\}$ or $t = \{(f_j, u, w_{u, f_j})\}$

$neighbour \leftarrow f_j$;

$sum_of_w \leftarrow sum_of_w + w_{u, neighbour}$;

$weighted_sum \leftarrow weighted_sum + (ratings[neighbour, i] - \overline{r_{neighbour}}) \bullet w_{u, neighbour}$;

end if

end for

$r_{u, i} = r_u + \frac{weighted_sum}{sum_of_w}$;

end for

Αλγόριθμος 1: SNB_algorithm

5. ΑΞΙΟΛΟΓΗΣΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ

Στο κεφάλαιο αυτό παρουσιάζονται ενδεικτικές εκτελέσεις του αλγορίθμου σύστασης για την εξαγωγή προβλέψεων βαθμολογιών από χρήστες σε διάφορα αντικείμενα. Οι εκτελέσεις διεξάγονται στα πλαίσια πειραμάτων στα οποία διαμορφώνουμε κατάλληλα ένα προκαθορισμένο σύνολο δεδομένων (dataset) επιδιώκοντας να προσομοιώσουμε τη συμπεριφορά των χρηστών που αλληλεπιδρούν με το σύστημα. Εν συνεχεία, οι προβλέψεις αξιολογούνται χρησιμοποιώντας συνήθεις μετρικές απόδοσης των συστημάτων συνεργατικού φιλτραρίσματος.

5.1 Σύνολο δεδομένων

Για την πραγματοποίηση των πειραμάτων, επιλέξαμε να χρησιμοποιήσουμε σύνολα δεδομένων από τον ιστότοπο MovieLens [34]. Το MovieLens αποτελεί ένα σύστημα συστάσεων το οποίο προτείνει ταινίες στους χρήστες του βάσει των προτιμήσεών τους χρησιμοποιώντας τον αλγόριθμο συνεργατικού φιλτραρίσματος. Οι χρήστες καλούνται να αξιολογήσουν τις ταινίες δίνοντας μια βαθμολογία από το 1 έως το 5 και παροτρύνονται να αξιολογήσουν όσο το δυνατόν περισσότερες ώστε να γίνουν πιο ακριβείς οι προτάσεις του συστήματος. Ταυτόχρονα με τις συστάσεις ταινιών, το MovieLens παρέχει πληροφορίες σχετικά με τις ταινίες που διαθέτει όπως είναι οι ηθοποιοί που πρωταγωνιστούν σε αυτές και οι σκηνοθέτες τους. Δημιουργός του συστήματος αυτού είναι το GroupLens, ένα ερευνητικό εργαστήριο του τμήματος επιστήμης υπολογιστών του πανεπιστημίου της Μινεσότα. Το GroupLens [36] συλλέγει σε διάφορες χρονικές περιόδους τα σύνολα δεδομένων του MovieLens και τα διαθέτει δημόσια για χρήση. Στα πλαίσια της εργασίας χρησιμοποιήθηκαν σύνολα δεδομένων που περιελάμβαναν βαθμολογίες από 1000 χρήστες σε 1700 ταινίες, χαρακτηριστικά των ταινιών όπως το είδος τους, οι ηθοποιοί και οι σκηνοθέτες τους καθώς και δεδομένα σχετικά με τα δημογραφικά χαρακτηριστικά των χρηστών όπως η ηλικία, το φύλο τους και το επάγγελμά τους. Οι χρήστες είναι ανώνυμοι και η ταυτότητά τους προσδιορίζεται από ένα μοναδικό αναγνωριστικό χρήστη (userId).

Το σύνολο δεδομένων του MovieLens εμφανίζει το μειονέκτημα ότι δεν παρέχει πληροφορίες σχετικά με κοινωνικές συνδέσεις χρηστών οι οποίες είναι απαραίτητες για την εκτέλεση του αλγορίθμου. Για το σκοπό αυτό, δημιουργούμε από την πλατφόρμα του Facebook σαράντα (40) δοκιμαστικούς χρήστες (test users) της εφαρμογής μας καθώς και τυχαίες φιλίες μεταξύ αυτών (βλ. Εικόνα 16). Για κάθε έναν από αυτούς, η πλατφόρμα του Facebook παρέχει ένα κατάλληλο διακριτικό πρόσβασης. Εν συνεχεία, αντιστοιχούμε κάθε δοκιμαστικό χρήστη με ένα αναγνωριστικό χρήστη (UserId) από το σύνολο δεδομένων του MovieLens.

Αναφορικά με τις προτιμήσεις χρηστών που εισάγονται στο σύστημα, δε χρησιμοποιούμε κάποιο έτοιμο διαθέσιμο σύνολο δεδομένων αλλά τις εξάγουμε υποθετικά για τον καθένα εντοπίζοντας κοινά χαρακτηριστικά ανάμεσα στις ταινίες που ο χρήστης έχει αξιολογήσει με μεγάλο βαθμό.

Test Users				Δημιουργία · See All Roles
<input type="button" value="Επιλογή όλων"/> <input type="button" value="Select None"/> <input type="button" value="Enable Ticker"/> <input type="button" value="Make Friends"/> <input type="button" value="Add to Other Apps"/> <input type="button" value="Αφαιρέστε"/>				
<input type="checkbox"/>	Name	ID	Email	
<input type="checkbox"/>	Mark Ameiaccjdfji Zuckersky	100005913304609	mark_yfstxid_zuckersky@tfnw.net	See Less Show Token · Switch To · Επεξεργασία
<input type="checkbox"/>	Jennifer Ameiaebdgcj Rosenthalsen...	100005915254739	ajxfckk_rosenthalsenmanskywitzbergsteinescu_1368709132@tfnw.net	Δείτε περισσότερα
<input type="checkbox"/>	James Ameiedbechja Bowersstein	100005954253801	ophtpkj_bowersstein_1368709131@tfnw.net	Δείτε περισσότερα
<input type="checkbox"/>	Patricia Ameiegaichag Huiberg	100005957193817	patricia_svimgmj_huiberg@tfnw.net	Δείτε περισσότερα
<input type="checkbox"/>	Elizabeth Amehiichdgcj Bowersberg	100005899384705	uhzuvlc_bowersberg_1368709131@tfnw.net	Δείτε περισσότερα
<input type="checkbox"/>	Dorothy Ameiddhfcgjf Goldmanwitz	100005944863676	dorothy_rsiactg_goldmanwitz@tfnw.net	Δείτε περισσότερα
<input type="checkbox"/>	John Ameibjcbdfjg Carrierosteimanesc...	100005920324607	qjwnpd_carrierosteimanescuwitzbergsonsensky_1368709129@tfnw.net	Δείτε περισσότερα
<input type="checkbox"/>	Nancy Ameieieccgjf Goldmanman	100005955933706	hkzytuo_goldmanman_1368709128@tfnw.net	Δείτε περισσότερα
<input type="checkbox"/>	David Ameiabjadfgg Bowerswitz	100005912014677	pudzrnp_bowerswitz_1368709127@tfnw.net	Δείτε περισσότερα

Εικόνα 16: Δημιουργία δοκιμαστικών χρηστών

5.2 Μετρικές απόδοσης

Η αξιολόγηση ενός αλγορίθμου συνεργατικού φιλτραρίσματος απαιτεί τη χρήση ανάλογων μετρικών απόδοσης που καθορίζουν την ποιότητα του συστήματος. Μια συνήθης πρακτική αξιολόγησης περιλαμβάνει τη μέτρηση της ικανότητας πρόβλεψης του συστήματος. Οι μετρικές που σχετίζονται με αυτή τη διαδικασία μετρούν εμπειρικά κατά πόσο διαφέρουν οι αξιολογήσεις των αντικειμένων που παράγει ένα σύστημα συστάσεων για ένα χρήστη σε σχέση με τις πραγματικές του αξιολογήσεις στα αντικείμενα αυτά. Η πιο συνηθισμένη μετρική που χρησιμοποιείται για την αξιολόγηση της ακρίβειας πρόβλεψης είναι το μέσο απόλυτο σφάλμα (Mean Absolute Error -MAE) η οποία υπολογίζει το μέσο όρο της απόλυτης τιμής της διαφοράς ανάμεσα στις προβλεπόμενες και τις πραγματικές αξιολογήσεις όπως φαίνεται από την εξίσωση (9):

$$MAE = \frac{\sum_{i,j} |p_{i,j} - r_{i,j}|}{n} \quad (9)$$

όπου n είναι ο αριθμός των προβλεπόμενων αξιολογήσεων, $p_{i,j}$ είναι η προβλεπόμενη αξιολόγηση του χρήστη i στο αντικείμενο j και $r_{i,j}$ είναι η πραγματική αξιολόγηση. Σημειώνεται ότι όσο μικρότερο είναι το MAE τόσο μεγαλύτερη ακρίβεια έχει ο αλγόριθμος στα αποτελέσματα που παρουσιάζει.

Μια δεύτερη μετρική ακρίβειας πρόβλεψης είναι η ρίζα του μέσου τετραγωνικού σφάλματος (Root Mean Squared Error, -RMSE) η οποία έγινε ιδιαίτερα δημοφιλής αφότου χρησιμοποιήθηκε από την online εταιρεία ενοικίασης ταινιών Netflix για τη μέτρηση της απόδοσης σε συστάσεις ταινιών. Το RMSE δίνει την τετραγωνική ρίζα της

μέσης τιμής της διαφοράς υψωμένης στο τετράγωνο κάθε ζεύγους προβλεπόμενης και πραγματικής αξιολόγησης και δίνεται από την εξίσωση (10):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i,j} (p_{i,j} - r_{i,j})^2} \quad (10)$$

όπου ομοίως n είναι ο αριθμός των προβλεπόμενων αξιολογήσεων, $p_{i,j}$ είναι η προβλεπόμενη αξιολόγηση του χρήστη i στο αντικείμενο j και $r_{i,j}$ είναι η πραγματική αξιολόγηση. Το RMSE ενισχύει τις εισφορές των απόλυτων σφαλμάτων μεταξύ των προβλέψεων και των πραγματικών τιμών δίνοντας μεγαλύτερη έμφαση σε μεγάλα σφάλματα. Λόγου χάρη, ένα σφάλμα μιας μονάδας αυξάνει το άθροισμα των τετραγωνικών σφαλμάτων κατά ένα όμως ένα σφάλμα των δύο μονάδων αυξάνει το άθροισμα κατά τέσσερα. Και στην περίπτωση του RMSE όσο μικρότερη είναι η τιμή της μετρικής τόσο καλύτερες είναι οι προβλεπόμενες τιμές από τον αλγόριθμο.

5.3 Μοντέλο αξιολόγησης

Για τις ανάγκες αξιολόγησης του αλγορίθμου, δημιουργήθηκε ένα απλοποιημένο σύστημα που χρησιμοποιεί το προαναφερθέν σύνολο δεδομένων ως είσοδο. Το σύστημα αυτό προσομοιώνει μια διαδικασία αξιολόγησης και παραβλέπει αρκετές από τις λειτουργίες της διεπαφής όπως είναι η εγγραφή, η καταχώρηση και επεξεργασία προτιμήσεων, το ιστορικό, κ.λπ. Το μοντέλο περιλαμβάνει μια βάση η οποία αποτελείται από τους εξής πέντε πίνακες:

test_details (id, age, sex, region, occupation, username, psw, node): Σε αυτόν τον πίνακα αποθηκεύονται τα δημογραφικά στοιχεία του χρήστη (ηλικία, φύλο, επάγγελμα) όπως διατίθενται από το σύνολο δεδομένων του MovieLens. Η περιοχή δε λαμβάνεται υπόψη από τον αλγόριθμο αφού δεν υπάρχει κάποια σχετική πληροφορία διαθέσιμη. Επίσης σε κάθε χρήστη ανατίθεται ένα όνομα χρήστη (username) και ένα συνθηματικό (psw) ώστε να μπορεί να συνδεθεί στο σύστημα. Το πεδίο node περιλαμβάνει το αναγνωριστικό του αντίστοιχου δοκιμαστικού χρήστη στο Facebook. Τέλος, στο πεδίο id αποθηκεύεται το αναγνωριστικό του χρήστη στο σύστημα.

test_preferences (id, userid, preference, attribute): Ο πίνακας αυτός έχει ακριβώς την ίδια λειτουργικότητα με τον πίνακα user_preferences, δηλαδή αποθηκεύει τις προτιμήσεις των χρηστών με αναγνωριστικό userid για συγκεκριμένα χαρακτηριστικά προϊόντων. Η διαφορά έγκειται στο γεγονός ότι εδώ παραλείπεται το πεδίο «category» καθώς το σύνολο δεδομένων μας περιλαμβάνει αποκλειστικά αξιολογήσεις για ταινίες.

test_friends (id, node1, node2, weight, DFactor, MovFactor): Εδώ αποθηκεύονται οι φιλίες μεταξύ κάθε ζεύγους δοκιμαστικών χρηστών. Και σε αυτήν την περίπτωση έχουν παραλειφθεί οι μετρικές που αφορούν συσχετίσεις χρηστών σε βιβλία ή μουσική.

movielens_ratings (id, node, itemid, rating, real_rating, prediction): Ο εν λόγω πίνακας αποθηκεύει τις αξιολογήσεις των χρηστών όπως αυτές διαβάζονται από το σύνολο δεδομένων του MovieLens. Στην περίπτωση πρόβλεψης αποθηκεύεται και η πραγματική αξιολόγηση (real_rating) του χρήστη.

matching (node, movielensid): Ο πίνακας matching αντιστοιχεί ένα αναγνωριστικό δοκιμαστικού χρήστη στο Facebook (node) με ένα αναγνωριστικό χρήστη στο MovieLens.

Το σύστημα αξιολόγησης απαρτίζεται ακόμη από 4 επιμέρους προγράμματα που εκτελούνται διαδοχικά για την περάτωση της συνολικής διαδικασίας. Το πρώτο πρόγραμμα (Initialization) συμπληρώνει τους πίνακες test_details (εκτός από το πεδίο

node), movielens_ratings και matching. Είναι υπεύθυνο δηλαδή για την ανάκτηση των δοκιμαστικών χρηστών της εφαρμογής, την αντιστοίχισή τους με χρήστες από το MovieLens, την παραγωγή κωδικών πρόσβασης των χρηστών στο σύστημα και την απόσπαση πληροφοριών για αξιολογήσεις και δημογραφικά στοιχεία των χρηστών. Από τις διαθέσιμες αξιολογήσεις των χρηστών, ένα μέρος αποτελεί το σύνολο εκπαίδευσης (training set) και ένα άλλο μέρος αποτελεί το σύνολο δοκιμής (test set). Το σύνολο εκπαίδευσης περιλαμβάνει τις αξιολογήσεις όλων των χρηστών που θα χρησιμοποιηθούν ως δεδομένα εισόδου στον αλγόριθμο. Το σύνολο δοκιμής περιλαμβάνει ταινίες για τις οποίες ο αλγόριθμος θα αγνοήσει προσωρινά τις αξιολογήσεις τους και θα κάνει προβλέψεις. Στην παρούσα φάση, στον πίνακα movielens_ratings αποθηκεύεται μόνο το σύνολο εκπαίδευσης από το πρόγραμμα. Το Facebook στιγμιότυπο δημιουργήθηκε χρησιμοποιώντας ένα διακριτικό πρόσβασης εφαρμογής που παρέχεται από την πλατφόρμα του Facebook.

Το δεύτερο πρόγραμμα (FindPreferences) είναι υπεύθυνο για την παραγωγή προτιμήσεων των χρηστών και τη συμπλήρωση του πίνακα test_preferences. Ειδικότερα, για κάθε χρήστη αναζητά τις ταινίες στις οποίες έχει δώσει υψηλή βαθμολογία. Από το σύνολο των ταινιών εντοπίζονται κοινοί σκηνοθέτες ή ηθοποιοί ενώ παράλληλα υπολογίζονται πόσες ταινίες αξιολογήθηκαν από κάθε είδος (πχ κωμωδία, δράμα, περιπέτεια, κ.λπ). Στον πίνακα αποθηκεύονται σαν προτιμήσεις τα πέντε δημοφιλέστερα είδη ταινιών του χρήστη με την προϋπόθεση ότι ο αριθμός των ταινιών που ανήκουν στο εν λόγω είδος είναι διψήφιος, οι ηθοποιοί που πρωταγωνιστούν ταυτόχρονα σε περισσότερες από τρεις ταινίες καθώς και οι σκηνοθέτες που έχουν σκηνοθετήσει τουλάχιστον τρεις ταινίες από το σύνολο των ταινιών που έχει αξιολογήσει με μεγάλο βαθμό ο χρήστης.

Το τρίτο πρόγραμμα (SocialData) προσομοιώνει τη διαδοχική αλληλεπίδραση των χρηστών με το σύστημα. Μπορεί να εκτελεστεί περισσότερες από μία φορές ανάλογα με τον αριθμό των χρηστών που επιθυμούμε να εγγράψουμε στο σύστημα ανά χρονική στιγμή. Αν υποθέσουμε λόγω χάρη ότι την τρέχουσα χρονική στιγμή έχουν εισαχθεί πέντε χρήστες στο σύστημα, το πρόγραμμα για καθέναν από αυτούς διαδοχικά θα ανακτήσει τους φίλους του στο Facebook και θα τους αποθηκεύσει στη βάση δεδομένων. Θα υπολογίσει επίσης και θα αποθηκεύσει τις αντίστοιχες μετρικές συσχέτισής του με αυτούς. Το πρόγραμμα SocialData ενημερώνει το πεδίο node του πίνακα test_details και όλα τα πεδία του πίνακα test_friends. Σημειώνεται ότι ο υπολογισμός των μετρικών συσχέτισης αφορά μόνο σε ταινίες ενώ έχουν αφαιρεθεί οι διαδικασίες υπολογισμού όσον αφορά μετρικές για βιβλία και μουσική. Επιπροσθέτως έχει αφαιρεθεί η μετρική συσχέτισης των χρηστών ως προς τα like καθώς αυτή η λειτουργία δεν υποστηρίζεται από τους δοκιμαστικούς χρήστες.

Το τέταρτο και τελευταίο πρόγραμμα (Evaluation) εκτελεί τον αλγόριθμο σύστασης για έναν χρήστη-στόχο της επιλογής μας και παράγει αποτελέσματα. Το πρόγραμμα περιλαμβάνει μια απλοποιημένη διεπαφή που εμφανίζει τη σελίδα συνδεδεμένου χρήστη όπως αυτή περιγράφηκε στο τρίτο κεφάλαιο, χωρίς όμως τη λειτουργία εγγραφής. Στα πεδία της φόρμας καταχωρούμε το όνομα χρήστη και τον κωδικό πρόσβασης του χρήστη-στόχου που έχουν παραχθεί από το πρόγραμμα Initialization. Η καταχώρηση των στοιχείων οδηγεί απευθείας στην εκτέλεση του αλγορίθμου σύστασης όπως περιγράφηκε στο τέταρτο κεφάλαιο με κάποιες παραλλαγές ώστε να γίνει προσαρμογή με το διαθέσιμο σύνολο δεδομένων. Έχουν αφαιρεθεί και σε αυτήν την περίπτωση όλες οι διαδικασίες υπολογισμού μετρικών για βιβλία και μουσική αλλά και των αντίστοιχων ΔΕΕΚ καθώς και η μετρική συσχέτισης των χρηστών ως προς τα like. Σημειώνεται ότι η εκτέλεση του αλγορίθμου για το χρήστη-στόχο προϋποθέτει τη χρήση του διακριτικού πρόσβασης ειδικά για το δοκιμαστικό χρήστη το οποίο και

ανατίθεται απευθείας στο Facebook σιγμιότυπο. Αναφορικά με τις αξιολογήσεις, ο αλγόριθμος αποθηκεύει στον πίνακα *monielens_ratings* το σύνολο δοκιμής για το χρήστη-στόχο. Για καθένα από τα αντικείμενα του συνόλου δοκιμής υπολογίζεται μια υποτιθέμενη προβλεπόμενη αξιολόγηση του χρήστη ενώ παράλληλα αποθηκεύεται και η πραγματική αξιολόγηση ώστε να υπολογιστεί στη συνέχεια η ακρίβεια προβλέψεων του αλγορίθμου. Ο τερματισμός του προγράμματος συνοδεύεται από την εμφάνιση στην οθόνη αποτελεσμάτων όπως είναι τα στοιχεία του χρήστη, οι φίλοι του, οι τελικοί του γείτονες, οι αξιολογήσεις που προβλέπει ο αλγόριθμος καθώς και οι μετρικές ακρίβειας πρόβλεψης MAE και RMSE.

5.4 Πειραματικά αποτελέσματα

Στην ενότητα αυτή παρουσιάζονται διάφορα σενάρια εκτέλεσης του αλγορίθμου. Σε όλες τις περιπτώσεις χρησιμοποιείται ο ίδιος χρήστης-στόχος και καταγράφονται τα αποτελέσματα πρόβλεψης αξιολογήσεων για εκείνον ανά πέντε εγγραφές χρηστών. Από το σύνολο των σαράντα δοκιμαστικών χρηστών εγγράφουμε στο σύστημα τους τριάντα. Οι χρήστες εγγράφονται ο ένας μετά τον άλλο και το εφοδιάζουν με τις απαραίτητες πληροφορίες ενώ όσοι δεν έχουν εγγραφεί μέχρι την τρέχουσα εκτέλεση του αλγορίθμου συμπεριλαμβανομένων και των υπολοίπων δέκα δοκιμαστικών χρηστών που αφήσαμε συνεισφέρουν μόνο μέσω των αξιολογήσεών τους στο τελικό αποτέλεσμα. Όπως περιγράφηκε στο τέταρτο κεφάλαιο, το συνολικό βάρος που ορίζεται για κάθε κοινωνικό δεσμό παραμετροποιείται από επιμέρους βάρη που αθροίζουν στη μονάδα και καθορίζουν σε ποια μετρική επιθυμούμε να δοθεί έμφαση. Κατόπιν προσαρμογής με το σύνολο δεδομένων μας, το μοντέλο αξιολόγησης θα πρέπει να υπολογίσει το βάρος που δίνεται από τη εξίσωση (11):

$$Z = w_1 \cdot \text{PearsonResult} + w_2 \cdot \text{DFactor} + w_3 \cdot \text{MovieFactor} \quad (11)$$

Θα παρουσιάσουμε τρία βασικά σενάρια σχετικά με τις τιμές των παραμέτρων w_1 , w_2 και w_3 και έπειτα για καθένα από αυτά θα αξιολογήσουμε τα αποτελέσματα που προκύπτουν.

Αρχικά θεωρούμε ότι όλες οι μετρικές συσχέτισης συνεισφέρουν στο συνολικό βάρος κατά το ίδιο ακριβώς ποσοστό, οπότε θα ισχύει η σχέση (12):

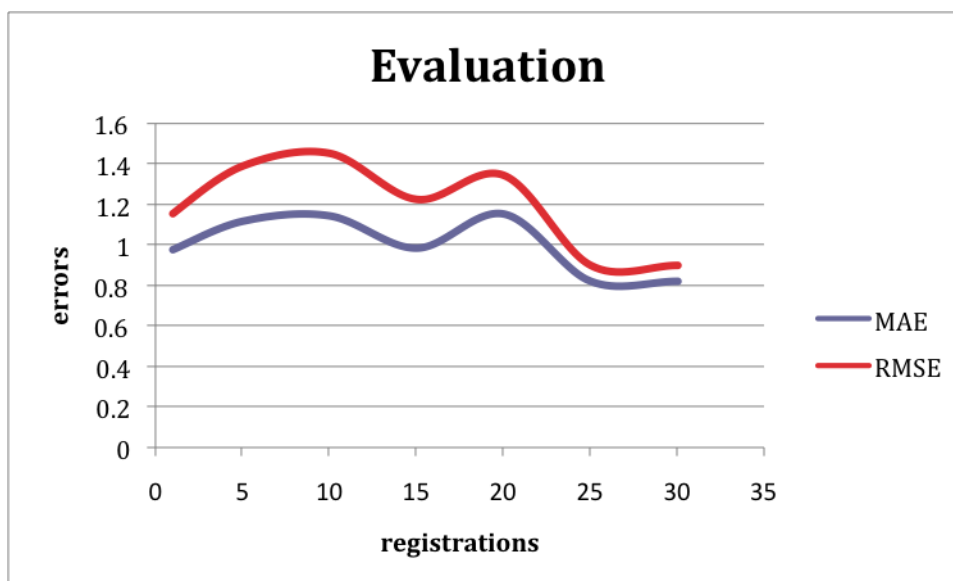
$$w_1 = w_2 = w_3 = \frac{1}{3} \quad (12)$$

Τα αποτελέσματα των μετρικών απόδοσης για κάθε πεντάδα εγγραφών των χρηστών παρουσιάζονται στον Πίνακα 9:

Πίνακας 9: Απόδοση αλγορίθμου με ίσες τιμές παραμέτρων

Εγγραφές	MAE	RMSE
1	0,97	1,15
5	1,11	1,39
10	1,14	1,45
15	0,98	1,22
20	1,15	1,34
25	0,82	0,9
30	0,82	0,9

Από τις τιμές που λαμβάνουν οι μετρικές απόδοσης προκύπτει ότι το σύστημα εμφανίζει μια αστάθεια με αυξομειώσεις σφαλμάτων μέχρι και τις πρώτες είκοσι εγγραφές ενώ από εκεί και πέρα η απόδοσή του βελτιώνεται και σταθεροποιείται (βλ. Σχήμα 3).



Σχήμα 3: Διάγραμμα μετρικών απόδοσης για ίσες τιμές παραμέτρων

Στη συνέχεια αυξάνουμε ελαφρώς το βάρος για έναν από τους παράγοντες PearsonResult, Dfactor και MovieFactor και θέτουμε ίσες τιμές στους άλλους δύο παράγοντες. Καταρχάς, θα δώσουμε έμφαση στην τιμή του PearsonResult, οπότε θα ισχύει η σχέση (13):

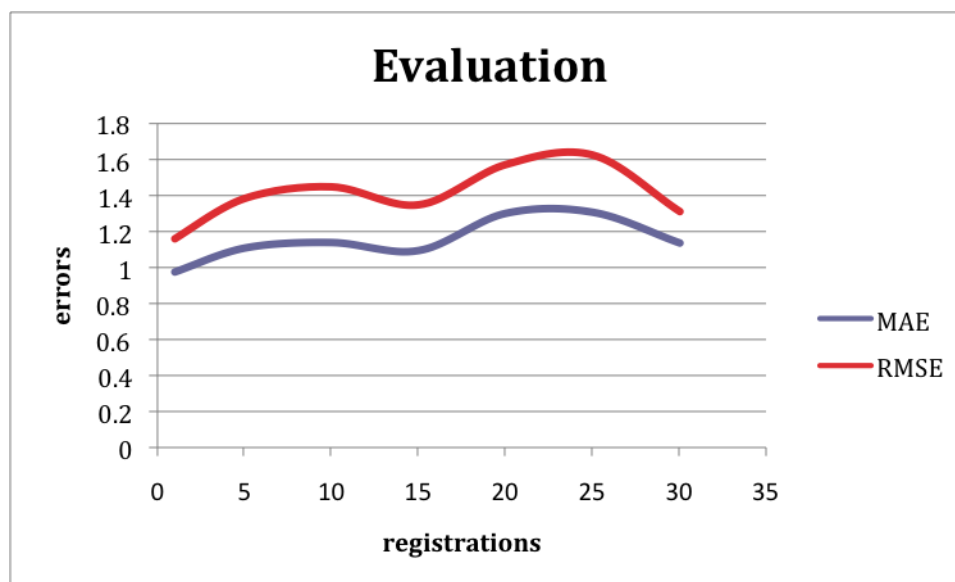
$$w_1 = 0.4 \quad w_2 = 0.3 \quad w_3 = 0.3 \quad (13)$$

Οι ανωτέρω παράμετροι διαμορφώνουν τα αποτελέσματα σχετικά με τη μέτρηση σφαλμάτων ως εξής (βλ. Πίνακα 10):

Πίνακας 10: Απόδοση αλγορίθμου με $w_1=0.4$, $w_2=0.3$ και $w_3=0.3$

Εγγραφές	MAE	RMSE
1	0,97	1,15
5	1,11	1,38
10	1,14	1,44
15	1,09	1,34
20	1,3	1,57
25	1,3	1,62
30	1,13	1,31

Τα σφάλματα σε αυτήν την περίπτωση αυξομειώνονται καθ' όλη τη διάρκεια εκτέλεσης του πειράματος και η μεγαλύτερη άνοδος τους διαπιστώνεται από τις είκοσι μέχρι τις εικοσιπέντε εγγραφές χρηστών. Τα σφάλματα αρχίζουν και πάλι να φθίνουν στις τριάντα εγγραφές προσεγγίζοντας τα αρχικά τους επίπεδα (βλ. Σχήμα 4).



Σχήμα 4: Διάγραμμα μετρικών απόδοσης για $w_1=0.4$, $w_2=0.3$ και $w_3=0.3$

Αν δοθεί έμφαση στα δημογραφικά χαρακτηριστικά των χρηστών, οι τιμές των παραμέτρων θα είναι οι ακόλουθες, σύμφωνα με τη σχέση (14):

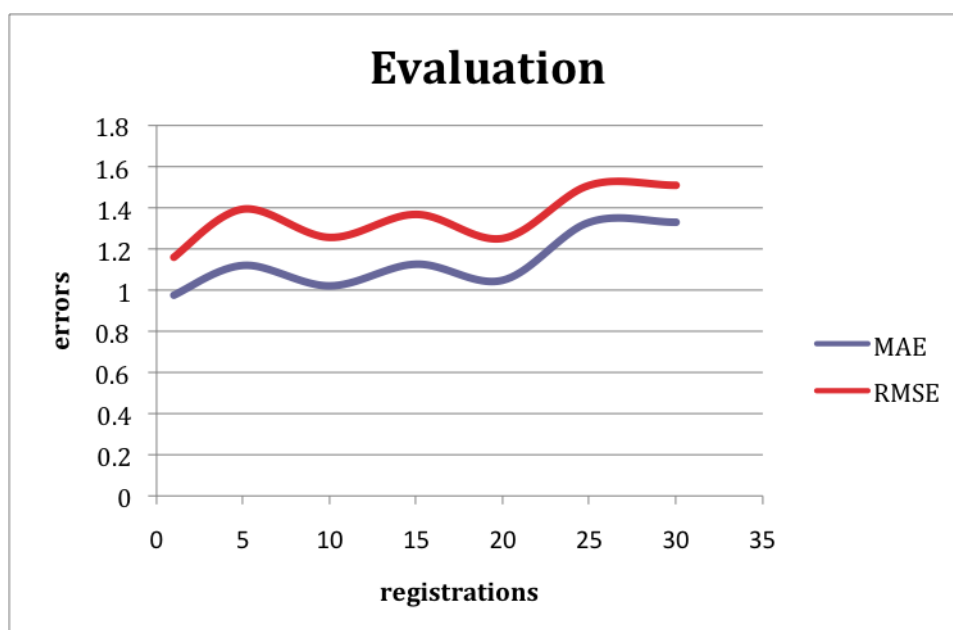
$$w_1 = 0.3 \quad w_2 = 0.4 \quad w_3 = 0.3 \quad (14)$$

και οι μετρικές απόδοσης διαμορφώνονται ως εξής (βλ. Πίνακα 11):

Πίνακας 11: Απόδοση αλγορίθμου με $w_1=0.3$, $w_2=0.4$ και $w_3=0.3$

Εγγραφές	MAE	RMSE
1	0,97	1,15
5	1,12	1,39
10	1,02	1,25
15	1,12	1,36
20	1,04	1,25
25	1,33	1,5
30	1,33	1,5

Στην προκειμένη περίπτωση, τα σφάλματα παρουσιάζουν μια κυματοειδή μορφή και σταθεροποιούνται μετά από εικοσιπέντε εγγραφές σε ένα υψηλότερο επίπεδο σε σχέση με την εκκίνηση του συστήματος (βλ. Σχήμα 5).



Σχήμα 5: Διάγραμμα μετρικών απόδοσης για $w_1=0.3$, $w_2=0.4$ και $w_3=0.3$

Δίνοντας έμφαση στις προτιμήσεις των χρηστών, οι παράμετροι θα είναι όπως φαίνεται στη σχέση (15):

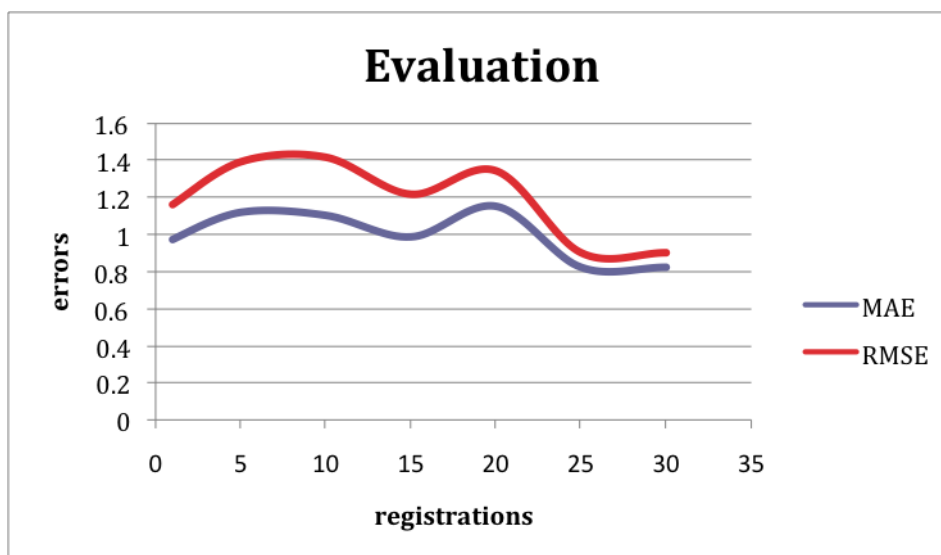
$$w_1 = 0.3 \quad w_2 = 0.3 \quad w_3 = 0.4 \quad (15)$$

και κατ' επέκταση οι μετρικές απόδοσης παίρνουν τις παρακάτω τιμές (βλ. Πίνακα 12):

Πίνακας 12: Απόδοση αλγορίθμου με $w_1=0.3$, $w_2=0.3$ και $w_3=0.4$

Εγγραφές	MAE	RMSE
1	0,97	1,15
5	1,12	1,39
10	1,10	1,41
15	0,98	1,21
20	1,15	1,34
25	0,82	0,9
30	0,82	0,9

Εδώ τα αποτελέσματα παρουσιάζουν μια παρόμοια εικόνα με την πρώτη περίπτωση εκτέλεσης όπου όλες οι παράμετροι έχουν την ίδια τιμή. Τα σφάλματα φθίνουν σημαντικά και σταθεροποιούνται μετά από είκοσι εγγραφές (βλ. Σχήμα 6).



Σχήμα 6: Διάγραμμα μετρικών απόδοσης για $w_1=0.3$, $w_2=0.3$ και $w_3=0.4$

Πραγματοποιούμε ακόμη μια τελευταία σειρά πειραμάτων παραμετροποίησης θέτοντας διαφορετική συνεισφορά κάθε μετρικής συσχέτισης στο συνολικό βάρος του κοινωνικού δεσμού. Θα εξετάσουμε καταρχάς τις περιπτώσεις κατά τις οποίες το PearsonResult συμβάλλει περισσότερο στη διαμόρφωση του τελικού αποτελέσματος. Ένα πρώτο σενάριο αποδίδει στις παραμέτρους τις ακόλουθες τιμές της σχέσης (16):

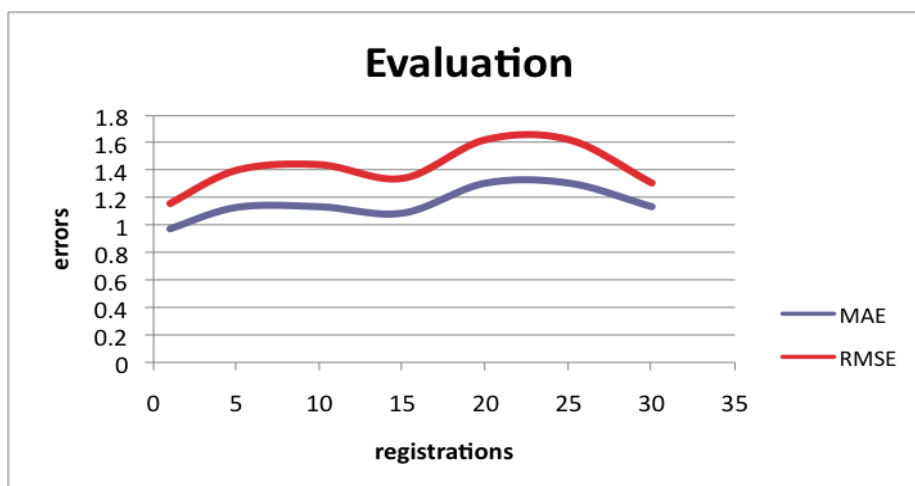
$$w_1 = 0.5 \quad w_2 = 0.3 \quad w_3 = 0.2 \quad (16)$$

οπότε προκύπτουν και οι ανάλογες μετρικές απόδοσης (βλ. Πίνακα 13):

Πίνακας 13: Απόδοση αλγορίθμου με $w_1=0.5$, $w_2=0.3$ και $w_3=0.2$

Εγγραφές	MAE	RMSE
1	0,97	1,15
5	1,13	1,4
10	1,13	1,44
15	1,09	1,34
20	1,3	1,62
25	1,3	1,62
30	1,13	1,31

Τα αποτελέσματα παρουσιάζουν ανάλογη εικόνα με το σύνολο παραμέτρων $w_1=0.4$, $w_2=0.3$, $w_3=0.3$. Τα σφάλματα κορυφώνονται ανάμεσα στις είκοσι με εικοσιπέντε εγγραφές και φθίνουν πάλι στη συνέχεια προσεγγίζοντας τα αρχικά τους επίπεδα (βλ. Σχήμα 7).



Σχήμα 7: Διάγραμμα μετρικών απόδοσης για $w_1=0.5$, $w_2=0.3$ και $w_3=0.2$

Εάν αντί των δημογραφικών στοιχείων αναθέσουμε στις προτιμήσεις την αμέσως μεγαλύτερη τιμή παραμέτρου μετά το PearsonResult, τότε θα ισχύει η σχέση (17):

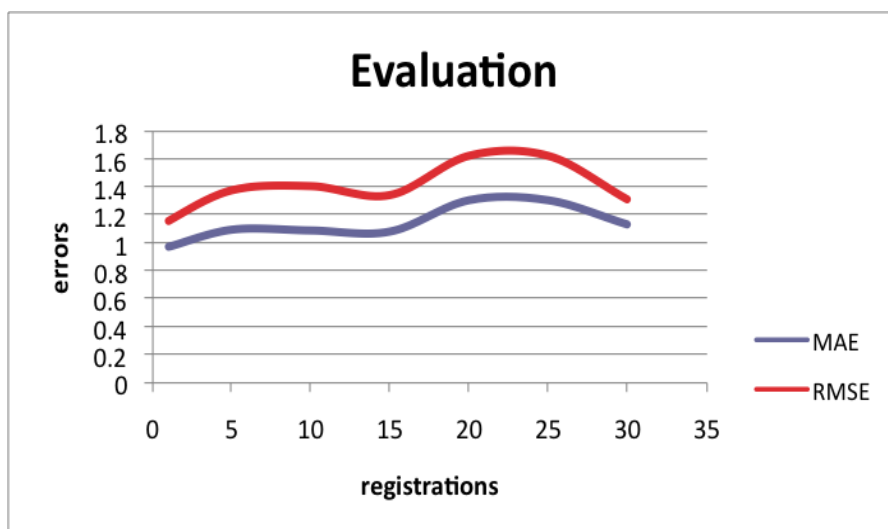
$$w_1 = 0.5 \quad w_2 = 0.2 \quad w_3 = 0.3 \quad (17)$$

και οι μετρικές απόδοσης θα είναι (βλ. Πίνακα 14):

Πίνακας 14: Απόδοση αλγορίθμου με $w_1=0.5$, $w_2=0.2$ και $w_3=0.3$

Εγγραφές	MAE	RMSE
1	0,97	1,15
5	1,1	1,37
10	1,09	1,4
15	1,08	1,34
20	1,3	1,62
25	1,3	1,62
30	1,13	1,31

Τα αποτελέσματα δε διαφέρουν σημαντικά σε σχέση με την αμέσως προηγούμενη περίπτωση όπως φαίνεται και από το Σχήμα 8.



Σχήμα 8: Διάγραμμα μετρικών απόδοσης για $w_1=0.5$, $w_2=0.2$ και $w_3=0.3$

Θα δώσουμε τώρα μεγαλύτερο συντελεστή στα δημογραφικά στοιχεία και αμέσως μετά στο PearsonResult. Οι παράμετροι θα πάρουν τις τιμές σύμφωνα με τη σχέση (18):

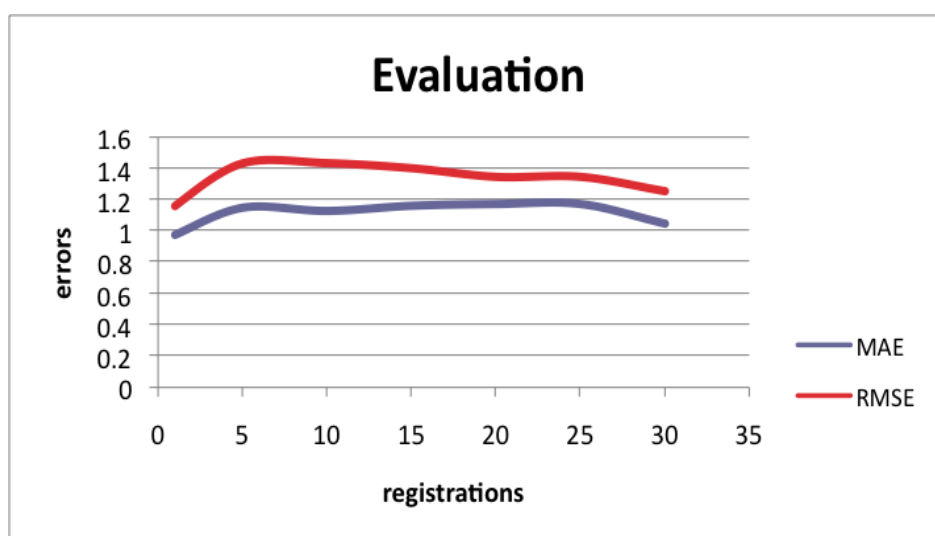
$$w_1 = 0.3 \quad w_2 = 0.5 \quad w_3 = 0.2 \quad (18)$$

και κατ' επέκταση οι μετρικές απόδοσης θα είναι (βλ. Πίνακα 15):

Πίνακας 15: Απόδοση αλγορίθμου με $w_1=0.3$, $w_2=0.5$ και $w_3=0.2$

Εγγραφές	MAE	RMSE
1	0,97	1,15
5	1,14	1,43
10	1,12	1,43
15	1,16	1,4
20	1,17	1,34
25	1,17	1,34
30	1,04	1,25

Εδώ τα σφάλματα ακολουθούν μια ανοδική πορεία από τις δεκαπέντε εγγραφές και έπειτα, φθίνουν όμως στη συνέχεια φτάνοντας σε ένα χαμηλότερο επίπεδο σε σχέση με τις προηγούμενες περιπτώσεις (βλ. Σχήμα 9).



Σχήμα 9: Διάγραμμα μετρικών απόδοσης για $w_1=0.3$, $w_2=0.5$ και $w_3=0.2$

Εάν αναθέσουμε χαμηλότερη τιμή παραμέτρου στο PearsonResult και αυξήσουμε αντίστοιχα τη συνεισφορά του συντελεστή προτιμήσεων, θα έχουμε από τη σχέση (19):

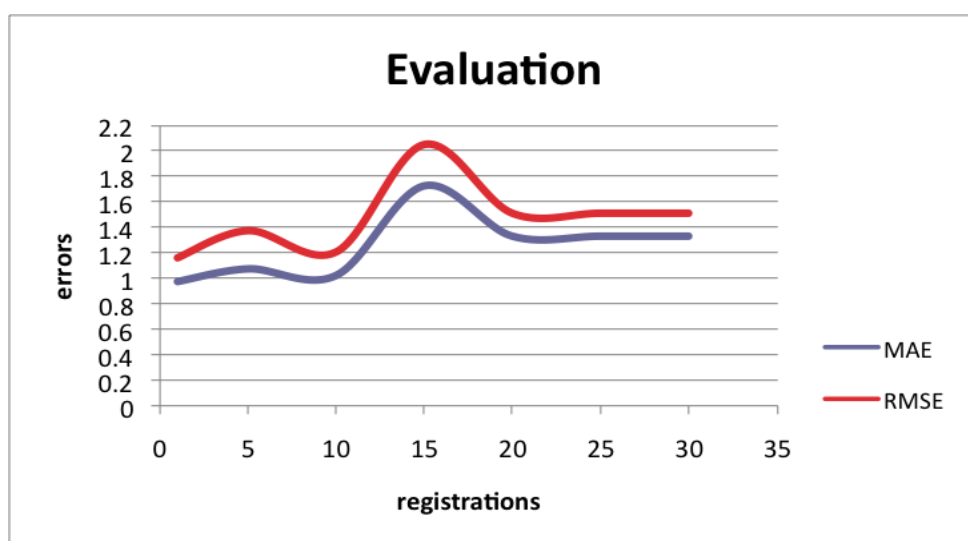
$$w_1 = 0.2 \quad w_2 = 0.5 \quad w_3 = 0.3 \quad (19)$$

οπότε προκύπτουν και οι ακόλουθες τιμές αναφορικά με τις μετρικές απόδοσης (βλ. Πίνακα 16):

Πίνακας 16: Απόδοση αλγορίθμου με $w_1=0.2$, $w_2=0.5$ και $w_3=0.3$

Εγγραφές	MAE	RMSE
1	0,97	1,15
5	1,07	1,37
10	1,02	1,2
15	1,72	2,05
20	1,33	1,51
25	1,33	1,51
30	1,33	1,51

Στην προκειμένη περίπτωση, η απόδοση του συστήματος έχει ήδη αρχίσει να σταθεροποιείται από τις είκοσι εγγραφές (βλ. Σχήμα 10).



Σχήμα 10: Διάγραμμα μετρικών απόδοσης για $w_1=0.2$, $w_2=0.5$ και $w_3=0.3$

Κλείνοντας, θα εξετάσουμε τις περιπτώσεις όπου οι προτιμήσεις των χρηστών διαδραματίζουν τον ισχυρότερο ρόλο στη διαμόρφωση του συνολικού βάρους του κοινωνικού δεσμού. Δίνοντας στο PearsonResult τη μεγαλύτερη έμφαση αμέσως μετά τις προτιμήσεις, προκύπτουν οι παρακάτω τιμές των παραμέτρων σύμφωνα με τη σχέση (20):

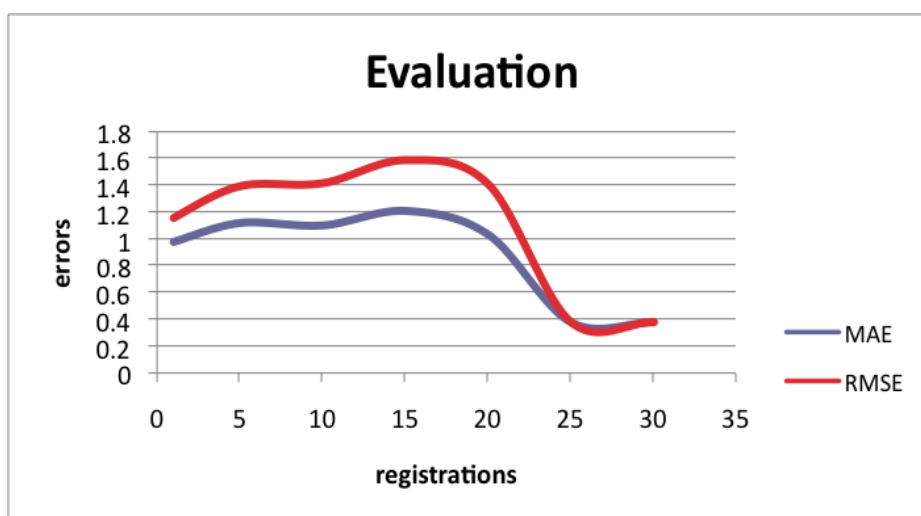
$$w_1 = 0.3 \quad w_2 = 0.2 \quad w_3 = 0.5 \quad (20)$$

Οι μετρικές απόδοσης σε αυτήν την περίπτωση θα είναι (βλ. Πίνακα 17):

Πίνακας 17: Απόδοση αλγορίθμου με $w_1=0.3$, $w_2=0.2$ και $w_3=0.5$

Εγγραφές	MAE	RMSE
1	0,97	1,15
5	1,11	1,39
10	1,1	1,41
15	1,2	1,58
20	1,03	1,4
25	0,37	0,37
30	0,37	0,37

Παρατηρούμε ότι το σύστημα βελτιώνεται αισθητά από τις εικοσιπέντε εγγραφές και έπειτα προσεγγίζοντας μια πολύ χαμηλή τιμή όσον αφορά το MAE αλλά και το RMSE όπως φαίνεται και από το Σχήμα 11.



Σχήμα 11: Διάγραμμα μετρικών απόδοσης για $w_1=0.3$, $w_2=0.2$ και $w_3=0.5$

Εάν δώσουμε στα δημογραφικά στοιχεία μεγαλύτερη έμφαση σε σχέση με το PearsonResult, θα αναθέσουμε στις παραμέτρους τις τιμές της σχέσης (21)

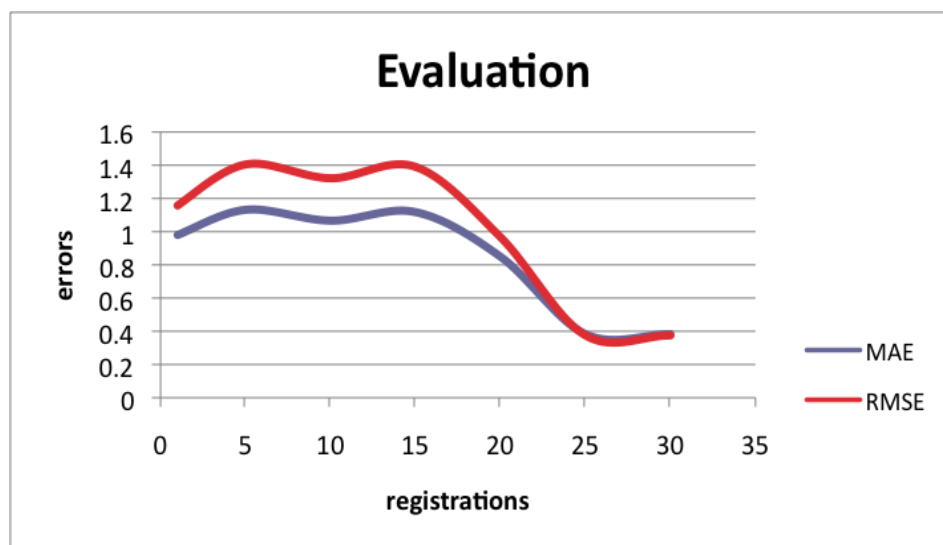
$$w_1 = 0.2 \quad w_2 = 0.3 \quad w_3 = 0.5 \quad (21)$$

και θα προκύψουν τα ακόλουθα αποτελέσματα απόδοσης (βλ. Πίνακα 18):

Πίνακας 18: Απόδοση αλγορίθμου με $w_1=0.2$, $w_2=0.3$ και $w_3=0.5$

Εγγραφές	MAE	RMSE
1	0,97	1,15
5	1,13	1,4
10	1,06	1,32
15	1,11	1,39
20	0,84	0,96
25	0,37	0,37
30	0,37	0,37

Το σύστημα με αυτήν την παραμετροποίηση παρουσιάζει την καλύτερη απόδοση στο δείγμα των τριάντα εγγεγραμμένων χρηστών που επιλέξαμε. Τα σφάλματα έχουν ήδη αρχίσει να φθίνουν από τις είκοσι εγγραφές και έπειτα, προσεγγίζοντας πολύ χαμηλές τιμές των μετρικών απόδοσης (βλ. Σχήμα 12).

Σχήμα 12: Διάγραμμα μετρικών απόδοσης για $w_1=0.2$, $w_2=0.3$ και $w_3=0.5$

Εν τέλει, από τα πειράματα που εκτελέσαμε, το βέλτιστο σενάριο για το χρήστη-στόχο που επιλέξαμε δίνεται από τις τιμές της σχέσης (21). Σύμφωνα με αυτό, το σφάλμα ελαχιστοποιείται όταν υπάρχουν σημαντικές πληροφορίες σχετικά με τις προτιμήσεις των χρηστών και σε δεύτερο στάδιο σχετικά με τα δημογραφικά τους στοιχεία. Πιθανώς η ομοιότητα των χρηστών είναι εκείνη που επηρεάζει τον τρόπο με τον οποίο οι χρήστες κατ' επέκταση αποδίδουν αξιολογήσεις στα αντικείμενα.

6. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παρούσα διπλωματική εργασία αναπτύξαμε ένα σύστημα συστάσεων που χρησιμοποιεί πληροφορίες από κοινωνικά δίκτυα και τεχνικές από τη θεωρία γράφων. Η υλοποίηση αλγορίθμων προς αυτήν την κατεύθυνση εμφανίζει σημαντικά πλεονεκτήματα γεγονός που καθιστά πιθανή τη στροφή των ερευνητών μελλοντικά προς την περαιτέρω εξέλιξή τους. Τα συστήματα που βασίζονται σε κοινωνικά δίκτυα μπορούν να μειώνουν την άμεση αλληλεπίδραση των χρηστών τους με αυτά καθώς δύνανται να αποσπούν πολλές πληροφορίες έμμεσα από τα προφίλ κοινωνικού δικτύου των χρηστών. Επιπλέον μπορούν να ερμηνεύσουν για ποιο λόγο οι χρήστες τους μπορεί να ταιριάζουν στις απόψεις σε ορισμένα αντικείμενα ενώ σε άλλα όχι. Αυτό συμβαίνει γιατί γνωρίζουν πληροφορίες για τους χρήστες τους, τις προτιμήσεις τους και τις αλληλεπιδράσεις τους με σελίδες του κοινωνικού ιστού. Αντιθέτως τα συνήθη συστήματα συνεργατικού φιλτραρίσματος έχουν επίγνωση μόνο των αξιολογήσεων των χρηστών.

Το σύστημα που περιγράφηκε στα πλαίσια της διπλωματικής εργασίας χρησιμοποιεί τους φίλους του χρήστη από το Facebook ως αρχικούς γείτονές του. Η εκδοχή αυτή εμφανίζει την προδιαγραφή του συστήματος, να παράγει αποτελέσματα σύστασης ακόμη και αν ο χρήστης δεν έχει κάνει καθόλου αξιολογήσεις. Εν γένει, στα συστήματα συστάσεων συνεργατικού φιλτραρίσματος ένας νέος χρήστης στο σύστημα θα πρέπει να κάνει επαρκείς αξιολογήσεις προκειμένου να βρεθούν οι γείτονές του και κατ'επέκταση να προκύψουν αποτελέσματα σύστασης. Το σύστημά μας αδυνατεί να εξάγει προβλέψεις αξιολογήσεων εάν ο χρήστης δεν έχει καθόλου φίλους. Στην πραγματικότητα όμως ένα τέτοιο φαινόμενο είναι σπάνιο καθώς από τη φύση τους οι ιστοχώροι κοινωνικής δικτύωσης έχουν δημιουργηθεί πρωτίστως για το σκοπό της σύναψης κοινωνικών δεσμών μεταξύ των μελών τους. Εάν ο χρήστης έχει φίλους που έχουν κάνει αξιολογήσεις σε αντικείμενα αλλά ο ίδιος δεν έχει κάνει καμία απολύτως αξιολόγηση το σύστημα έχει τη δυνατότητα να δώσει αποτελέσματα βασισμένο στις αξιολογήσεις των φίλων του. Κάθε χρήστης που εγγράφεται στο σύστημα από προεπιλογή έχει υποχρεωτικά καταχωρήσει τουλάχιστον τα δημογραφικά του στοιχεία προκειμένου να ολοκληρωθεί επιτυχώς η διαδικασία της εγγραφής. Το στοιχείο αυτό μαζί με την ενδεχόμενη γνώση του συστήματος σχετικά με τα likes των χρηστών στο Facebook μπορεί να χρησιμοποιηθεί ώστε από τους φίλους του να επιλεγούν εκείνοι που του ταιριάζουν περισσότερο και ύστερα από την εύρεση του ΔΕΕΚ να προκύψουν οι τελικοί συστήνοντες. Κατάλληλη βελτιστοποίηση προς αυτήν την κατεύθυνση θα μπορούσε να εξομαλύνει το πρόβλημα αργής εκκίνησης νέου χρήστη (cold-start new user problem). Μια άλλη προσέγγιση στο ζήτημα της αργής εκκίνησης είναι η εύρεση ενός αλγορίθμου ομαδοποίησης (clustering) των χρηστών βάσει της συμπεριφοράς τους στην αξιολόγηση αντικειμένων και των γενικότερων χαρακτηριστικών τους.

Μια σημαντική πτυχή του περιγραφόμενου συστήματος σύστασης είναι η εύρεση μέσω του ΔΕΕΚ εκείνων των φίλων που ταιριάζουν περισσότερο με το χρήστη-στόχο. Με αυτήν τη μέθοδο επιδιώκουμε να φιλτράρουμε τις επαφές του χρήστη καθώς στα κοινωνικά δίκτυα υπάρχει ένα ποσοστό επαφών σε κάθε προφίλ που δεν αποτελούν πραγματικούς φίλους του χρήστη αλλά τυχαίες επαφές με μηδενική αλληλεπίδραση. Τεχνικές υπολογισμού της πραγματικής αλληλεπίδρασης ανάμεσα σε ένα χρήστη και τις επαφές του όπως, λόγου χάρη, η μέτρηση του ποσοστού ανταλλαγής δημοσιεύσεων ημερησίως μεταξύ των φίλων στα χρονολόγιά τους θα μπορούσε να ωφελήσει την εύρεση καλύτερων γειτόνων για το χρήστη-στόχο. Όσο μεγαλύτερη είναι η πραγματική αλληλεπίδραση μεταξύ δύο φίλων στο κοινωνικό δίκτυο τόσο αυξάνεται η πιθανότητα να ταιριάζουν οι εν λόγω χρήστες. Η διαμόρφωση του βάρους μεταξύ τους θα μπορούσε επιπλέον να διαμορφώνεται και από άλλα στοιχεία που ομαδοποιούν τους

χρήστες όπως είναι οι ομάδες (groups), ένα κοινό ίδρυμα (πχ σχολείο, πανεπιστήμιο κ.λπ.), η συμμετοχή σε κοινές εκδηλώσεις και άλλα. Θα ήταν ενδιαφέρον επιπροσθέτως να λαμβάνεται υπόψη και ένα ποσοστό επίδρασης των φίλων των επαφών του χρήστη (mutual friends) καθώς αυτοί με τη σειρά τους επηρεάζουν τις επαφές του χρήστη-στόχου άρα έμμεσα και τον ίδιο.

Κλείνοντας, επισημαίνεται ότι τα κοινωνικά μέσα είναι χώρος συζητήσεων και έκφρασης απόψεων. Συστήματα που βασίζονται σε κοινωνικά δίκτυα μπορούν να δομηθούν λαμβάνοντας ως είσοδο κάποιο είδος αξιολόγησης σε μορφή κειμένου όπως λόγου χάρη μία ετικέτα (tag) ή μία γραπτή κριτική. Το γεγονός αυτό ανοίγει νέους δρόμους στην ανάπτυξη συστημάτων συστάσεων που μπορούν να επικεντρωθούν σε τεχνικές ανάλυσης κειμένου (text mining).

ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ

Ξενόγλωσσος όρος	Ελληνικός Όρος
Recommender Systems	Συστήματα Συστάσεων
Information overload	Υπερπληροφόρηση
e-commerce	Ηλεκτρονικό εμπόριο
e-shop	Ηλεκτρονικό κατάστημα
Rating	Αξιολόγηση
Content-based filtering	Φιλτράρισμα βασισμένο στο περιεχόμενο
Collaborative filtering	Συνεργατικό φιλτράρισμα
Memory-based filtering	Φιλτράρισμα με βάση τη μνήμη
Model-based filtering	Φιλτράρισμα με βάση το μοντέλο
User-based filtering	Φιλτράρισμα με βάση το χρήστη
Item-based filtering	Φιλτράρισμα με βάση το αντικείμενο
Knowledge-based filtering	Φιλτράρισμα βασισμένο στη γνώση
Demography-based filtering	Φιλτράρισμα βασισμένο στα δημογραφικά
Cold start new user	Αργή εκκίνηση για το νέο χρήστη
Cold start new item	Αργή εκκίνηση για το νέο αντικείμενο
Cosine-based similarity	Ομοιότητα με βάση το συνημιτονικό διάνυσμα
Data sparsity	Αραιά δεδομένα
Learning techniques	Τεχνικές μάθησης
Clustering models	Μοντέλα ομαδοποίησης
Latent semantic analysis	Κρυφή σημασιολογική ανάλυση
Support vector machines	Μηχανές διανυσμάτων υποστήριξης
Singular value decomposition	Αποσύνθεση μοναδικής τιμής
Hybrid filtering	Υβριδικό φιλτράρισμα
Enhanced entity-relationship model	Επεκταμένο μοντέλο οντοτήτων-συσχετίσεων
Social networks	Κοινωνικά δίκτυα
Social media	Κοινωνικά μέσα ενημέρωσης
Social web	Κοινωνικός ιστός
Social graph	Κοινωνικό γράφημα
Open Graph	Ανοιχτό γράφημα
Authorization	Εξουσιοδότηση
Authentication	Αυθεντικοποίηση
Login flow	Ροή εισόδου
Facebook login	Σύνδεση Facebook
Permissions	Δικαιώματα
Access token	Διακριτικό πρόσβασης
Timeline	Χρονολόγιο
NewsFeed	Ενημερώσεις
Ticker	Στήλη ενημέρωσης
Status update	Ενημέρωση κατάστασης
Facebook page	Σελίδα Facebook
Group	Ομάδα χρηστών
Event	Εκδήλωση
Chat	Συνομιλία
Post	Δημοσίευση
Comment	Σχόλιο

Χρήση Κοινωνικών Δικτύων ως Βάση ενός Συστήματος Συστάσεων

Tag	Ετικέτα
User	Χρήστης
Friends	Φίλοι
Mutual friends	Κοινοί φίλοι
Friendlist	Λίστα φίλων
Test user	Δοκιμαστικός χρήστης
Dialog	Παράθυρο διαλόγου
Recommenders	Συστήνοντες
Pearson Correlation Coefficient	Συντελεστής συσχέτισης Pearson
Minimum cost spanning tree	Δέντρο επικάλυψης ελαχίστου κόστους
Dataset	Σύνολο δεδομένων
Training set	Σύνολο εκπαίδευσης
Test set	Σύνολο δοκιμής
Evaluation metrics	Μετρικές απόδοσης
Mean absolute error	Μέσο απόλυτο σφάλμα
Root mean squared error	Τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος
Text mining	Ανάλυση κειμένου

ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ

ΣΣ	Συστήματα Συστάσεων
ACM RecSys	Association for Computing Machinery Recommender System
SIGIR	Special Interest Group on Information Retrieval
KDD	Knowledge Discovery in Databases
CB	Content-Based
CF	Collaborative Filtering
KB	Knowledge-Based
DB	Demography-Based
CBR	Case Based Reasoning
SNS	Social Network Site
RIC	Recommender Interface Component
RE	Recommender Engine
DB	DataBase
JSP	Java Server Pages
SNB	Social Network Based
API	Application Programming Interface
FQL	Facebook Query Language
PHP	Personal Home Page/ Hypertext PreProcessor
HTTP	HyperText Transfer Protocol
URL	Uniform Resource Locator
ΔΕΕΚ	Δέντρο Επικάλυψης Ελαχίστου Κόστους
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
ΕΚΠΑ	Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

ΑΝΑΦΟΡΕΣ

- [1] Francesco Ricci, Adriano Venturini, Dario Cavada, Nader Mirzadeh, Dennis Blaas, and Marisa Nones, "Product Recommendation with Interactive Query Management and Twofold Similarity", eCommerce and Tourism Research Laboratory ITC-irst via Sommarive 18, 38050 Povo, Italy.
- [2] Dianshuang Wu, Jie Lu, Guangquan Zhang, Hua Lin, "A fuzzy matching based recommendation approach for mobile products/services", 10th International Conference on Intelligent Systems Design and Applications, 2010.
- [3] Felix von Reischach, Florian Michahelles, Albrecht Schmidt, "The Design Space of Ubiquitous Product Recommendation Systems".
- [4] Manisha Hiralall, "Recommender systems for e-shops", Vrije Universiteit, Amsterdam, 2011.
- [5] João Carlos D. Lima, Cristiano C. Rocha, Iara Augustin, Mário A. R. Dantas, "A Context-Aware Recommendation System to Behavioral Based Authentication in Mobile and Pervasive Environments", IFIP Ninth International Conference on Embedded and Ubiquitous Computing, 2011.
- [6] Khaled W. Sadeddin, Alexander Serenko, James Hayes, "Online shopping bots for electronic commerce: The comparison of functionality and performance", Faculty of Business Administration, Lakehead University, Canada.
- [7] Zan Huang, Daniel Zeng and Hsinchun Chen, "A Comparative Study of Recommendation Algorithms in E-Commerce Applications".
- [8] J. Ben Schafer, Joseph A. Konstan, John Riedl, "E-Commerce Recommendation Applications", GroupLens Research Project, Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455.
- [9] Ming CHEN, "Research on Recommender Technology in E-commerce Recommendation System", 2nd International Conference on Education Technology and Computer (ICETC), 2010.
- [10] Eoghan Parle and Aaron Quigley, "Proximo, Location-Aware Collaborative Recommender", UCD School of Computer Science & Informatics, University College Dublin, Ireland.
- [11] CYRUS SHAHABI, YI-SHIN CHEN, "An Adaptive Recommendation System without Explicit Acquisition of User Relevance Feedback", Kluwer Academic Publishers, Netherlands, 2003.
- [12] Sung-Shun Weng, Mei-Ju Liu, "Personalized Product Recommendation in E-Commerce", IEEE International Conference on e-Technology, e-Commerce and e-Service, 2004.
- [13] Robin Burke, Jonathan Gemmell, Andreas Hotho, Robert Jöschke, "Recommendation in the Social Web", Association for the Advancement of Artificial Intelligence, 2011.
- [14] Ido Guy, Inbal Ronen, Ariel Raviv, "Personalized Activity Streams: Sifting through the «River of News»".
- [15] Aaditeshwar Seth, "Design of a Social Network Based Recommender System for Participatory Media Content", School of Computer Science, University of Waterloo, ON, Canada.
- [16] Massa, P. and Avesani, P., Trust-aware recommender systems. in, (2007), ACM, 24.
- [17] Lerman, K. Social networks and social information filtering on digg. *Arxiv preprint cs/0612046.N*.
- [18] Liu, F. and Lee, H.J. Use of social network information to enhance collaborative filtering performance. *Expert Systems with Applications*.
- [19] Shekarpour, S. and Katebi, S.D. Modeling and Evaluation of Trust with an Extension In Semantic Web. *Web Semantics: Science, Services and Agents on the World Wide Web*.
- [20] Garcia Esparza, S., O'Mahony, M. P., & Smyth, B. 2010. On the real-time web as a source of recommendation knowledge. *Proc. RecSys '10*, 305-308.
- [21] M. Granovetter, "The Strength of Weak Ties," American Journal of Sociology, Vol. 78, No. 6, 1973.
- [22] R. Burke, Hybrid Web Recommender Systems: Survey and Experiments, *User Modeling and User-Adapted Interaction*, Volume 12, Issue 4, 2002, p. 331–370.
- [23] Thomas Hess, "Recommender Engines Seminar paper", RWTH Aachen University, University of Bonn, February 1, 2009 Available at <http://www.slideshare.net/antiraum/recommender-engines-seminar-paper>
- [24] Wikipedia, "Collaborative Filtering", Available at http://en.wikipedia.org/wiki/Collaborative_filtering
- [25] Xiaoyuan Su, Taghi M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques", Available at <http://www.hindawi.com/journals/aai/2009/421425/>
- [26] Justin Lafferty, "What can Facebook 's Open Graph Do", Available at http://allfacebook.com/what-can-facebooks-open-graph-do_b99921
- [27] Wikipedia, "Facebook", Available at <http://en.wikipedia.org/wiki/Facebook>
- [28] WhatIs.com, "Definition of Facebook", Available at <http://whatis.techtarget.com/definition/Facebook>
- [29] TechTerms.com, "Definition of facebook", Available at <http://www.techterms.com/definition/facebook>
- [30] Pcmag.com Encyclopedia, "Definition of Facebook", available at <http://www.pcmag.com/encyclopedia/term/57226/facebook>
- [31] WebGoals, "facebook Glossary", Available at <http://webgoals.co/facebook-glossary/>

[32] Wikipedia, “Social Graph”, Available at http://en.wikipedia.org/wiki/Social_graph

[33] Facebook Developers, Available at <https://developers.facebook.com/>

[34] MovieLens official site, Available at <http://en.wikipedia.org/wiki/MovieLens>

[35] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, John T. Riedl, “Evaluating Collaborative Filtering Recommender Systems”, Oregon State University, University of Minnesota

[36] GroupLens Research, Available at <http://www.grouplens.org/>

[37] Ζησιμόπουλος Βασίλειος, “Αλγόριθμοι και πολυπλοκότητα”, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, 2008