



ΑΛΕΞΑΝΔΡΕΙΟ ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΘΕΣΣΑΛΟΝΙΚΗΣ

ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



Συστήματα

Προτάσεων σε on-line Καταστήματα

Πτυχιακή Εργασία:

ΣΒΑΡΝΑΣ ΑΣΤΕΡΙΟΣ (ΑΜ:021968)

ΣΤΟΓΙΑΝΝΗΣ ΚΩΝΣΤΑΝΤΙΝΟΣ (ΑΜ:021915)

Επιβλέπων Καθηγητής: ΚΟΚΚΟΡΑΣ ΦΩΤΗΣ

ΘΕΣΣΑΛΟΝΙΚΗ - 2010

Πρόλογος

Η παρούσα πτυχιακή εργασία μελετά τα συστήματα συστάσεων (recommender systems) σε ηλεκτρονικά (on line) καταστήματα. Η αγορά σε ηλεκτρονικά καταστήματα γίνονται όλο και πιο δημοφιλείς και η πληθώρα των πληροφοριών και των προϊόντων που παρέχονται σε αυτά μπορεί μεν να υπερνικά την απόσταση και να εξοικονομεί χρόνο στον πελάτη, πλην όμως τον υπερφορτώνει πληροφοριακά καθώς του δίνει πρόσβαση σε πολλά παρεμφερή προϊόντα με πλήθος λεπτομερειών στις περιγραφές τους. Είναι πολύ συχνό το φαινόμενο του να μην μπορεί ο πελάτης να σχηματίσει καλή εικόνα της αγοράς πριν προβεί σε συναλλαγή. Το πρόβλημα αυτό επιχειρείται να λυθεί με την χρήση συστήματος συστάσεων που παρουσιάζει στον πελάτη παρόμοια σε χαρακτηριστικά προϊόντα ώστε να γνωρίζει ο πελάτης "τι άλλο παρόμοιο υπάρχει".

Αναπτύχθηκε μία εφαρμογή ηλεκτρονικού καταστήματος (με προϊόντα πληροφορικής) που ενσωματώνει σύστημα συστάσεων το οποίο εξάγει βαθμό ομοιότητας για το κάθε προϊόν, συγκριτικά με τα υπόλοιπα ίδιου τύπου προϊόντα που υπάρχουν στο ηλεκτρονικό κατάστημα και τα παρουσιάζει στον πελάτη ως εναλλακτικά, με σκοπό την διευκόλυνσή του στην αγορά.

Η παρούσα πτυχιακή εργασία εκπονήθηκε στο πλαίσιο του σχετικού μαθήματος του Τμήματος Πληροφορικής, της σχολής Τεχνολογικών Εφαρμογών του Αλεξάνδρειου Τεχνολογικού Ιδρύματος Θεσσαλονίκης, υπό την επίβλεψη του κ. Φώτη Κόκκορα (Επιστημονικός Συνεργάτης του ΤΕΙ).

Θα θέλαμε να ευχαριστήσουμε ιδιαίτερα τον επιβλέποντα καθηγητή μας για την καθοδήγησή του, τόσο κατά τη διάρκεια της ανάπτυξης της εφαρμογής όσο και κατά τη συγγραφή της πτυχιακής.

Στογιάννης Κ. – Σβάρνας Α.

18/03/2010

Περίληψη

Τα συστήματα συστάσεων (recommender systems) χρησιμοποιούνται για να λύσουν το πρόβλημα της πληθώρας των πληροφοριών (information overload problem) που αντιμετωπίζουν οι χρήστες των ηλεκτρονικών καταστημάτων, αποτελώντας την πιο δημοφιλή μορφή εξατομικευμένης πληροφόρησης. Για τη δημιουργία τους χρησιμοποιείται γνώση που παρέχεται από ειδικούς, από τα ίδια τα προϊόντα ή που εξάγεται από την παρακολούθηση της συμπεριφοράς των πελατών. Η γνώση αυτή χρησιμοποιείται προς διευκόλυνση του πελάτη στην επιλογή του αλλά και προς όφελος του καταστήματος που επιδιώκει περισσότερες πωλήσεις.

Στην παρούσα πτυχιακή εργασία παρουσιάζονται τα δομικά στοιχεία των συστημάτων συστάσεων, καθώς και η βασική τους ταξινόμηση. Επιπρόσθετα γίνεται εκτενής αναφορά στο λογισμικό που χρησιμοποιήθηκε για την ανάπτυξη του ηλεκτρονικού καταστήματος, αλλά και του υποσυστήματος συστάσεων που εμπεριέχετε σε αυτό. Αναλύεται εκτενώς η αρχιτεκτονική της εφαρμογής και ειδικότερα του υποσυστήματος συστάσεων καθώς είναι δομημένο σε μορφή αρθρώματος (module) που μπορεί με παραμετροποίηση να προσαρτηθεί και σε άλλα ηλεκτρονικά καταστήματα. Τέλος παρουσιάζεται η λειτουργικότητα του συστήματος που δημιουργήθηκε.

Το ηλεκτρονικό κατάστημα και το σύστημα συστάσεων αναπτύχθηκαν σε γλώσσα PHP με χρήση MySQL, σε περιβάλλον Uniform Server.

Abstract

Recommender systems are used to solve the information overload problem, which users of e-shopping are facing, making them the most popular form of personalization. For their creation is used knowledge provided from experts and also from the observation of customers behavior which guides them to the detection of product-services of their interest.

In the current thesis the structural elements of recommender systems are presented as well as their basic classification. Furthermore is mentioned the essential software that has been used for the formation of the electronic shop, which has been developed in the lines of the thesis, and of the recommender systems that are contained in it. Afterwards the design of the electronic shop application and its basic functions are extensively analyzed. Finally is presented the recommender system that has been developed, which uses the knowledge provided from the expert to accommodate the customer in his purchase. The electronic shop and the recommender system were developed in PHP language with the use of MySQL.

Περιεχόμενα

ΠΡΟΛΟΓΟΣ	I
ΠΕΡΙΛΗΨΗ	II
ABSTRACT	III
ΠΕΡΙΕΧΟΜΕΝΑ.....	V
1 ΕΙΣΑΓΩΓΗ	1
2 ΣΥΣΤΗΜΑΤΑ ΣΥΣΤΑΣΕΩΝ	5
2.1 ΕΙΣΑΓΩΓΗ ΣΤΑ ΣΥΣΤΗΜΑΤΑ ΣΥΣΤΑΣΕΩΝ	5
2.2 ΔΟΜΙΚΑ ΣΤΟΙΧΕΙΑ ΣΥΣΤΗΜΑΤΩΝ ΣΥΣΤΑΣΕΩΝ.....	8
2.2.1 Η Είσοδος.....	9
2.2.2 Η Έξοδος.....	11
2.2.3 Η Μέθοδος.....	14
2.3 ΤΡΟΠΟΙ ΥΠΟΒΟΛΗΣ ΣΥΣΤΑΣΕΩΝ	17
2.4 ΒΑΘΜΟΣ ΕΞΑΤΟΜΙΚΕΥΣΗΣ	18
2.5 ΤΑΞΙΝΟΜΗΣΗ ΣΥΣΤΗΜΑΤΩΝ ΣΥΣΤΑΣΕΩΝ	18
2.5.1 <i>Content-based approaches</i>	19
2.5.2 <i>Collaborative filtering approaches</i>	23
2.5.3 <i>Hybrid approaches</i>	27
2.6 ΠΕΡΙΟΡΙΣΜΟΙ ΚΑΙ ΠΡΟΒΛΗΜΑΤΑ ΣΕ ΣΥΣΤΑΣΕΙΣ	30
3 ΠΕΡΙΓΡΑΦΗ ΕΡΓΑΛΕΙΩΝ.....	35
3.1 Η ΓΛΩΣΣΑ PHP	35
3.2 ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ MYSQL	38
3.3 Ο ΔΙΑΚΟΜΙΣΤΗΣ ΙΣΤΟΥ APACHE.....	40
3.4 Ο UNIFORM SERVER.....	41
4 ΣΧΕΔΙΑΣΜΟΣ ΤΟΥ E-SHOP.....	43
4.1 ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ	43
4.2 ΟΡΓΑΝΩΣΗ ΣΧΗΜΑΤΟΣ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ.....	43

4.3	ΔΟΜΗ ΤΟΥ ΗΛΕΚΤΡΟΝΙΚΟΥ ΚΑΤΑΣΤΗΜΑΤΟΣ	46
5	ΛΕΙΤΟΥΡΓΙΕΣ ΤΟΥ E-SHOP	49
5.1	ΕΓΓΡΑΦΗ ΠΕΛΑΤΗ	49
5.2	ΕΙΣΟΔΟΣ ΠΕΛΑΤΗ	50
5.3	ΦΟΡΜΑ ΕΠΙΚΟΙΝΩΝΙΑΣ	52
5.4	ΒΑΘΜΟΛΟΓΗΣΗ ΠΡΟΪΟΝΤΟΣ	53
5.5	ΑΓΟΡΑ ΠΡΟΪΟΝΤΟΣ	53
5.6	ΟΙ ΚΥΡΙΕΣ ΣΕΛΙΔΕΣ ΤΟΥ E-SHOP	59
5.6.1	<i>Η αρχική σελίδα index.php.....</i>	<i>59</i>
5.6.2	<i>Η σελίδα με τη λίστα κάθε προϊόντος.....</i>	<i>60</i>
5.6.3	<i>Η σελίδα προϊόντος.....</i>	<i>61</i>
5.7	Η ΛΟΓΙΚΗ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ ΣΥΣΤΑΣΕΩΝ.....	63
5.8	ΔΗΜΙΟΥΡΓΙΑ ΣΥΣΤΗΜΑΤΟΣ ΣΥΣΤΑΣΕΩΝ	63
5.9	ΥΠΟΛΟΓΙΣΜΟΥ ΟΜΟΙΟΤΗΤΑΣ ΠΡΟΪΟΝΤΩΝ	67
5.10	ΑΠΕΙΚΟΝΙΣΗ ΟΜΟΙΟΤΗΤΑΣ	70
6	ΣΥΜΠΕΡΑΣΜΑΤΑ	72
	ΒΙΒΛΙΟΓΡΑΦΙΑ.....	75
	ΔΙΑΔΙΚΤΥΑΚΕΣ ΠΗΓΕΣ.....	78

1 Εισαγωγή

Με την ραγδαία ανάπτυξη των διαδικτυακών υπηρεσιών των τελευταίων ετών και την ευρύτατη καθιέρωση του ηλεκτρονικού εμπορίου, τα ηλεκτρονικά καταστήματα που αποτελούν ένα βασικό συστατικό του, γνωρίζουν μια διαρκώς διευρυνόμενη διάδοση και αποδοχή ως ένας εναλλακτικός τρόπος για την πραγματοποίηση αγορών από τους καταναλωτές.

Οι δυνατότητες που προσφέρει το Διαδίκτυο σε μια επιχείρηση είναι πολύ σημαντικές. Οι επιχειρήσεις που θα το κατανοήσουν αυτό νωρίς και θα ενσωματώσουν το “e” στην οργάνωση και τη λειτουργία τους είναι αυτές που θα αποκτήσουν συγκριτικό πλεονέκτημα στην ιδιαίτερα ανταγωνιστική ψηφιακή οικονομία. Μεταξύ των κυριότερων πλεονεκτημάτων περιλαμβάνονται: η παροχή πληροφοριών, προϊόντων και υπηρεσιών 24 ώρες το 24ώρο, 7 ημέρες την εβδομάδα, η παγκόσμια πρόσβαση ανεξαρτήτως γεωγραφικής περιοχής, η άμεση απόκριση, η απουσία χαρτιού, η μείωση του κόστους, οι εξατομικευμένες υπηρεσίες, κλπ.

Τα ηλεκτρονικά καταστήματα γενικά βασίζονται στο παράδειγμα λειτουργίας ενός φυσικού καταστήματος με το οποίο είναι εξοικειωμένοι όλοι οι καταναλωτές. Έτσι σε ένα οποιοδήποτε ηλεκτρονικό κατάστημα, ο πελάτης εισερχόμενος παίρνει ένα εικονικό «καλάθι αγορών» όπως ακριβώς θα έκανε μπαίνοντας σε κάποιο σουπερμάρκετ. Μέσα σε αυτό «τοποθετεί» τα προϊόντα που επιλέγει για να ψωνίσει πατώντας την επιλογή «Προσθήκη στο καλάθι» που υπάρχει σε κάθε διαθέσιμο προς πώληση προϊόν. Όταν ολοκληρώσει τις αγορές του, πηγαίνει στο «ταμείο» όπου ακολουθεί η διαδικασία της πληρωμής. Η παραλαβή των προϊόντων γίνεται μέσω κούριερ σε λίγες μέρες.

Η βασική ιδέα σχεδόν σε όλα τα ηλεκτρονικά καταστήματα είναι ίδια: θυμίζουν όσο το δυνατόν περισσότερο ένα φυσικό κατάστημα στην λογική και στον

τρόπο οργάνωσης, όσο είναι αυτό δυνατόν, για να εκμεταλλευτούν τις προσλαμβάνουσες παραστάσεις και την εξοικείωση των καταναλωτών με την διαδικασία των αγορών στα φυσικά καταστήματα ώστε να μπορούν αυτοί να λειτουργούν διαισθητικά και με την ελάχιστη προσπάθεια μέσα στο διαδικτυακό κατάστημα.

Από την άλλη πλευρά όμως, αντί για τα μερικές εκατοντάδες βιβλία που ο χρήστης μπορεί να βρει για παράδειγμα σε ένα παραδοσιακό βιβλιοπωλείο, έχει τώρα πρόσβαση σε χιλιάδες βιβλία που διαθέτουν τα online καταστήματα. Με την αύξηση όμως των προσφερόμενων επιλογών, αυξάνει παράλληλα και ο φόρτος που απαιτείται από το χρήστη για να επιλέξει τα προϊόντα που ταιριάζουν στις ανάγκες και στις απαιτήσεις του. Για να αντιμετωπισθεί αυτό το πρόβλημα της πληθώρας των πληροφοριών που είναι γνωστό και σαν information overload problem, εφαρμόζονται συνήθως τεχνικές εξατομίκευσης.

Τα συστήματα συστάσεων (recommendations systems, recommender systems ή RSs) αποτελούν ίσως την πιο δημοφιλή μορφή εξατομίκευσης και τείνουν να μετατραπούν στις μέρες μας σε σημαντικά επιχειρησιακά εργαλεία. Τα συστήματα αυτά χρησιμοποιούν γνώση που παρέχεται είτε από ειδικούς ή που εξάγεται από την παρακολούθηση της συμπεριφοράς των πελατών για να οδηγήσουν τους τελευταίους στον εντοπισμό των προϊόντων-υπηρεσιών που τους ενδιαφέρουν.

Όσον αφορά την υλοποίηση των ηλεκτρονικών καταστημάτων υπάρχουν πολλά διαθέσιμα software εργαλεία που ποικίλουν σε δυνατότητες και προσφερόμενα χαρακτηριστικά. Στην διαδικτυακή πραγματικότητα σήμερα έχει ξεχωρίσει μια τριάδα λογισμικού λόγω των προηγμένων χαρακτηριστικών της και των πολλών πλεονεκτημάτων της. Αυτή αποτελείται από την PHP, το MySQL και τον Apache server τα οποία συνδυαζόμενα αποτελούν την βασική επιλογή σε πάρα πολλές παρουσίες εμπορικών και μη διαδικτυακών τόπων.

Η PHP είναι η μηχανή που κρύβεται πίσω από τα εκατομμύρια των δυναμικών εφαρμογών του Ιστού. Το ευρύ σύνολο χαρακτηριστικών της, το προσιτό συντακτικό της, και η υποστήριξη στα πρότυπα των διαφόρων λειτουργικών συστημάτων και web servers την έχει κάνει μια ιδανική γλώσσα τόσο για γρήγορη ανάπτυξη εφαρμογών στον Παγκόσμιο Ιστό, όσο και για την συστηματική κατασκευή σύνθετων συστημάτων.

Ένα από τα πιο δυνατά και σημαντικά χαρακτηριστικά της PHP είναι η υποστήριξη που έχει για ένα μεγάλο σύνολο βάσεων δεδομένων. Η συγγραφή μιας σελίδας που υποστηρίζει βάσεις δεδομένων μπορεί να είναι εξαιρετικά απλή. Από τις πολλές υποστηριζόμενες βάσεις δεδομένων ξεχωρίζει το MySQL. Το MySQL είναι ένα πολύ γρήγορο, στιβαρό, σύστημα διαχείρισης σχεσιακής βάσης δεδομένων (RDBMS) που επιτρέπει την αποτελεσματική αποθήκευση, αναζήτηση, ταξινόμηση, και ανάκτηση δεδομένων. Ο MySQL server ελέγχει την πρόσβαση στα δεδομένα και διασφαλίζει ότι πολλοί χρήστες μπορούν να χρησιμοποιούν τις δυνατότητες του ταυτόχρονα, παρέχοντας γρήγορη πρόσβαση σε αυτές, και εξασφαλίζει ότι μόνο εξουσιοδοτημένοι χρήστες μπορούν να έχουν πρόσβαση. Ως εκ τούτου, ο MySQL server είναι ένας multi-user, multi-threaded server. Χρησιμοποιεί την SQL (Structured Query Language), την πρότυπη γλώσσα για ερωτήσεις στην βάση δεδομένων παγκοσμίως.

Όταν ξεκινάει κάποιος να δημιουργήσει έναν δικτυακό τόπο για ηλεκτρονικό εμπόριο, υπάρχουν πολλά διαφορετικά προϊόντα λογισμικού που θα μπορούσε να επιλέξει να χρησιμοποιήσει. Μερικές από αυτές τις επιλογές του θα εξαρτηθούν από άλλες. Παραδείγματος χάριν, όλα τα λειτουργικά συστήματα δεν θα τρέχουν σε όλα τα διαθέσιμα προϊόντα hardware, όλες οι scripting γλώσσες δεν θα μπορούν να συνδέονται με όλες τις βάσεις δεδομένων κ.λ.π. Με τον συνδυασμό PHP - MySQL δεν χρειάζεται να προσέξουμε ιδιαίτερα το hardware, το λειτουργικό σύστημα, ή το λογισμικό για τον Web server επειδή κάτι τέτοιο δεν είναι απαραίτητο. Ένα από τα ωραία χαρακτηριστικά γνωρίσματα της PHP είναι ότι είναι διαθέσιμη για τα Microsoft Windows, για πολλές εκδόσεις UNIX, και συνεργάζεται με οποιουδήποτε πλήρως - λειτουργικούς Web servers. Το MySQL είναι ομοίως ευπροσάρμοστο.

Ο Apache web server χρησιμοποιείται ευρύτερα από όλους τους άλλους web servers συνολικά πράγμα που αποδεικνύεται από τα στατιστικά στοιχεία. Σαφώς, το Apache Software Foundation πετυχαίνει το στόχο του “να προσφέρει έναν ασφαλή, αποδοτικό, και επεκτάσιμο server που διατηρεί τις HTTP υπηρεσίες του σε συμφωνία με τα HTTP πρότυπα.”

Ο Uniform Server συνδυάζει την PHP, τον MySQL server και τον Apache web server και μαζί με άλλα βοηθητικά προγράμματα είναι ένα ιδανικό πακέτο για την ανάπτυξη web εφαρμογών γενικά.

Η εφαρμογή που υλοποιείται, στην παρούσα πτυχιακή εργασία, με την χρήση των παραπάνω εργαλείων(PHP,MySQL,Apache server) έχει σκοπό την υλοποίηση ενός πλήρως λειτουργικού δικτυακού τόπου που να παρέχει όλες τις βασικές υπηρεσίες που χαρακτηρίζουν την λειτουργικότητα ενός διαδικτυακού ηλεκτρονικού καταστήματος. Επιπλέον, υλοποιείται ένα σύστημα συστάσεων το οποίο εισάγει την γνώση του ειδικού με σκοπό να οδηγηθεί ο πελάτης στην κατάλληλη αγορά.

2 Συστήματα Συστάσεων

Παρουσιάζονται τα βασικά δομικά στοιχεία ενός συστήματος συστάσεων (η είσοδος, η έξοδος, η μέθοδος παραγωγής των προτάσεων), ο τρόπος αποστολής των προτάσεων που αποτελεί κρίσιμη σχεδιαστική απόφαση των e-shops και η ταξινόμηση των συστημάτων συστάσεων σε τρεις βασικές κατηγορίες (φιλτράρισμα με βάση το περιεχόμενο, συνεργατικό φιλτράρισμα, υβριδικές προσεγγίσεις).

2.1 Εισαγωγή στα Συστήματα Συστάσεων

Τα ηλεκτρονικά καταστήματα (electronic shops, e-shops ή online shops) αποτελούν την κυριότερη εφαρμογή του ηλεκτρονικού εμπορίου (electronic commerce, ecommerce, e-εμπόριο ή EC), υλοποιούν το σενάριο επιχείρηση-προς-καταναλωτή (business-to-consumer ή B2C) και στοχεύουν στην προώθηση αλλά και στην πώληση προϊόντων και υπηρεσιών μέσω του Παγκοσμίου Ιστού Πληροφοριών (World Wide Web ή Web). Είναι γεγονός ότι τα τελευταία χρόνια ένας μεγάλος αριθμός από τέτοια καταστήματα λειτουργούν, ενώ καθημερινά νέα e-shops ανοίγουν τις διαδικτυακές τους πύλες. Άλλωστε, το e-εμπόριο έδωσε τη δυνατότητα στους χρήστες-πελάτες να επιλέξουν προϊόντα και υπηρεσίες μέσα από μια μεγάλη ποικιλία. Υπάρχουν παραδείγματα καταστημάτων τα οποία λειτουργούν μόνο μέσω ηλεκτρονικών καναλιών όπως το βιβλιοπωλείο Amazon (<http://www.amazon.com>), ενώ άλλα συνδυάζουν τη φυσική τους παρουσία με την ηλεκτρονική όπως η εταιρία Microsoft.

Είναι αναμφισβήτητο γεγονός ότι οι δυνατότητες που προσφέρει το Διαδίκτυο σε μια επιχείρηση είναι πολύ σημαντικές. Οι επιχειρήσεις που θα το κατανοήσουν αυτό νωρίς και θα ενσωματώσουν το “e” στην οργάνωση και τη λειτουργία τους είναι αυτές που θα αποκτήσουν συγκριτικό πλεονέκτημα στην ιδιαίτερα ανταγωνιστική ψηφιακή οικονομία. Μεταξύ των κυριότερων πλεονεκτημάτων περιλαμβάνονται: η παροχή πληροφοριών, προϊόντων και υπηρεσιών 24

ώρες το 24ώρο, 7 ημέρες την εβδομάδα, η παγκόσμια πρόσβαση ανεξαρτήτως γεωγραφικής περιοχής, η άμεση απόκριση, η απουσία χαρτιού, η μείωση του κόστους, οι εξατομικευμένες υπηρεσίες, κλπ. Ένα επιτυχημένο e-shop μπορεί να ενισχύσει τη δημόσια εικόνα της επιχείρησης, να αυτοματοποιήσει εσωτερικές και εξωτερικές διαδικασίες, να μειώσει το κόστος και το χρόνο εξυπηρέτησης και να την τροφοδοτεί συνεχώς με στοιχεία που αφορούν τόσο τους πελάτες όσο και τις συναλλαγές τους καθώς επίσης και με ερωτήσεις, σχόλια, παρατηρήσεις, συστάσεις, κλπ. Όλα αυτά αν χρησιμοποιηθούν σωστά από την επιχείρηση θα οδηγήσουν στη βελτίωση των παρεχόμενων υπηρεσιών και προϊόντων της.

Από την άλλη πλευρά όμως, αντί για τα μερικές εκατοντάδες βιβλία που ο χρήστης μπορεί να βρει για παράδειγμα σε ένα παραδοσιακό βιβλιοπωλείο, έχει τώρα πρόσβαση σε χιλιάδες βιβλία που διαθέτουν τα online καταστήματα. Με την αύξηση όμως των προσφερόμενων επιλογών, αυξάνει παράλληλα και ο φόρτος που απαιτείται από το χρήστη για να επιλέξει τα προϊόντα που ταιριάζουν στις ανάγκες και στις απαιτήσεις του. Για να αντιμετωπισθεί αυτό το πρόβλημα της πληθώρας των πληροφοριών που είναι γνωστό και σαν information overload problem, εφαρμόζονται συνήθως τεχνικές εξατομίκευσης (personalization). Ένας ορισμός της εξατομίκευσης είναι ο ακόλουθος: «εξατομίκευση είναι η διαδικασία προσαρμογής των σελίδων ενός web site στις προτιμήσεις και τα ιδιαίτερα χαρακτηριστικά κάθε χρήστη ξεχωριστά. Αποτελεί το μέσο που θα ικανοποιήσει τις ανάγκες ενός πελάτη πιο αποτελεσματικά, θα καταστήσει ευκολότερες και γρηγορότερες τις αλληλεπιδράσεις του και συνεπώς θα αυξήσει την ικανοποίησή του καθώς και την πιθανότητα να επισκεφτεί ξανά το site.» (DTV, 2005).

Τα συστήματα συστάσεων (recommendations systems, recommender systems ή RSs) αποτελούν ίσως την πιο δημοφιλή μορφή εξατομίκευσης και τείνουν να μετατραπούν στις μέρες μας σε σημαντικά επιχειρησιακά εργαλεία (Μαρκέλλου, 2003). Αρχικά, εμφανίστηκαν στα μέσα της δεκαετίας του '90 και χρησιμοποιούνταν από λίγα web sites. Σήμερα δεν νοούνται εφαρμογές e-εμπορίου που να

μην ενσωματώνουν τεχνολογίες συστημάτων συστάσεων προκειμένου να βοηθήσουν τους πελάτες στις αγορές τους όπως:

- προτάσεις για βιβλία, CDs και άλλα προϊόντα στο Amazon.com
- προτάσεις για κινηματογραφικές ταινίες στο MovieLens
- ειδήσεις στο VERSIFI Technologies¹

ενώ ακόμα και εταιρίες λογισμικού όπως η Microsoft ενσωμάτωσαν recommendation χαρακτηριστικά στα προϊόντα τους π.χ. στον Commerce Server. Με αυτό τον τρόπο, τα e-shops μπορούν να βελτιώσουν την απόδοσή τους καλύπτοντας τις ανάγκες και τις προτιμήσεις κάθε πελάτη ξεχωριστά, να αυξήσουν την ικανοποίησή του (satisfaction), να εδραιώσουν την εμπιστοσύνη (e-trust) και την e-πίστη του (e-loyalty) και να καθιερώσουν ένα-προς-ένα σχέσεις (1-1 relationships) (Markellou et al., 2005).

Τα συστήματα αυτά χρησιμοποιούν γνώση που παρέχεται είτε από ειδικούς ή που εξάγεται από την παρακολούθηση της συμπεριφοράς των πελατών για να οδηγήσουν τους τελευταίους στον εντοπισμό των προϊόντων-υπηρεσιών που τους ενδιαφέρουν. Οι προτάσεις μπορεί να βασίζονται: στις μεγαλύτερες συνολικές πωλήσεις του site, στα δημογραφικά δεδομένα του πελάτη, στην ανάλυση της προηγούμενης αγοραστικής συμπεριφοράς του σαν μια πρόβλεψη της μελλοντικής συμπεριφοράς του, σε δεδομένα από κοινότητες χρηστών (user communities), κλπ. Οι τύποι των προτάσεων περιλαμβάνουν: εξατομικευμένες προσφορές, τιμές, προϊόντα ή υπηρεσίες, προσθήκη ή παράλειψη θεματικών ενοτήτων ή παραγράφων, ταξινόμηση, κρύψιμο, εμφάνιση, προσθήκη, παράληψη ή έμφαση συνδέσμων, προαιρετικές εξηγήσεις ή αναλυτικές πληροφορίες, σύνοψη απόψεων μιας κοινότητας, παροχή κριτικών μιας κοινότητας, κλπ. (Μαρκέλλου, 2005). Η όλη διαδικασία μπορεί να υλοποιηθεί χειροκίνητα π.χ. συνδυασμός προϊόντων (cross-selling) ή αυτόματα από τον ηλεκτρονικό υπολογιστή π.χ. συσχέτιση πελατών με παρόμοια αγοραστική συμπεριφορά.

Ένα σύστημα συστάσεων ενισχύει τις πωλήσεις ενός καταστήματος e-εμπορίου με τους ακόλουθους τρόπους: μετατρέπει τους επισκέπτες σε αγοραστές, αυξάνει το cross-selling και το up-selling και «χτίζει» την εμπιστοσύνη και την πίστη των πελατών. Σε έναν κόσμο όπου οι ανταγωνιστές είναι μόνο ένα

«κλικ» μακριά, το να κερδιθεί η εμπιστοσύνη των πελατών αποτελεί μια σημαντική επιχειρηματική στρατηγική. Τα συστήματα συστάσεων βελτιώνουν την εμπιστοσύνη αφού δημιουργούν μια ιδιαίτερη σχέση μεταξύ του πελάτη και του site. Τα sites που επενδύουν στο να μάθουν τους πελάτες τους, χρησιμοποιούν αυτά τα συστήματα ώστε να παρουσιάσουν interfaces που ταιριάζουν στις ανάγκες των χρηστών τους. Οι πελάτες από τη μεριά τους ανταποδίδουν αυτές τις ευκολίες με το να ξαναγυρίσουν στα sites που τους ικανοποιούν σε μεγαλύτερο βαθμό. Όσο περισσότερο ο πελάτης χρησιμοποιεί το σύστημα (το οποίο μαθαίνει τι αυτός χρειάζεται), τόσο μεγαλύτερη είναι και η εμπιστοσύνη του σε αυτό το site. Ακόμη, δίνοντας τη δυνατότητα δημιουργίας σχέσεων μεταξύ των πελατών π.χ. μέσω μιας κοινότητας χρηστών με κοινά χαρακτηριστικά και ενδιαφέροντα, αυξάνεται η πιθανότητα οι πελάτες να επιστρέψουν σε αυτό το site αφού το προτείνουν και άλλοι χρήστες με τους οποίους έχουν επικοινωνία.

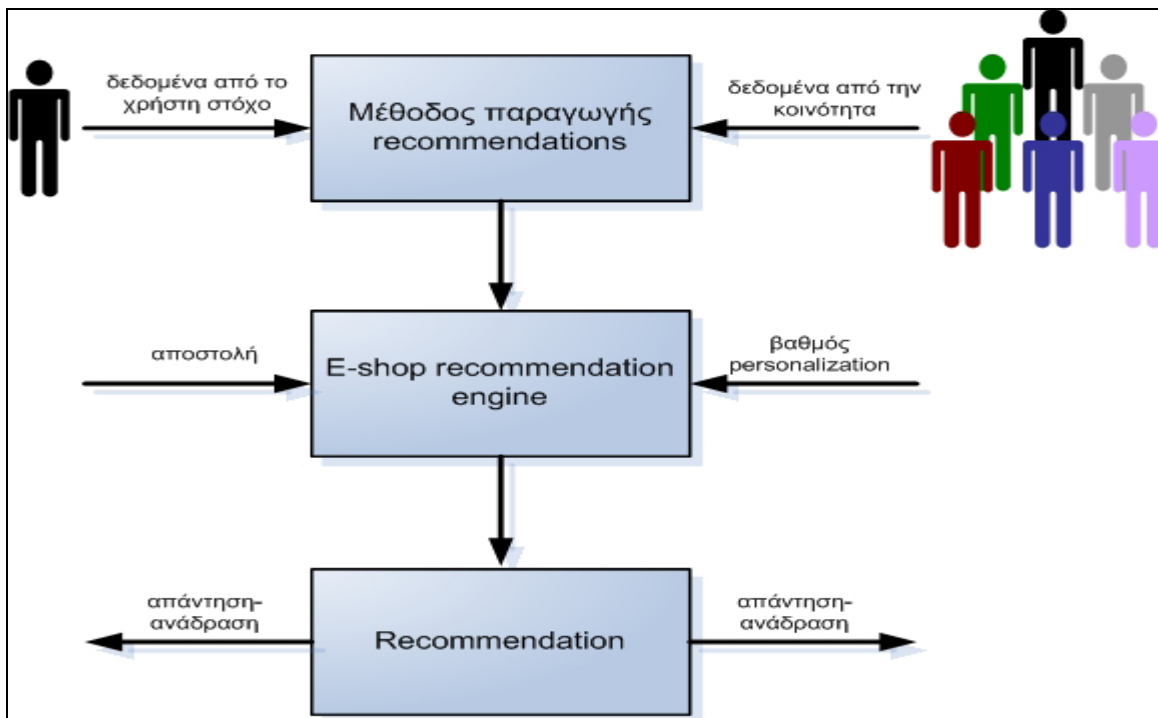
Γενικά, τα συστήματα συστάσεων μπορούν να ταξινομηθούν σε τρεις βασικές κατηγορίες: content based filtering (φιλτράρισμα με βάση το περιεχόμενο), collaborative filtering (συνεργατικό φιλτράρισμα) και hybrid approaches (υβριδικές προσεγγίσεις). Το content based filtering λαμβάνει υπόψη του τις παραμέτρους του χρήστη με τις πληροφορίες των προϊόντων (μεταδεδομένα, λέξεις κλειδιά, κλπ.), ενώ το collaborative filtering (CF) την «ομοιότητα» των χρηστών προκειμένου να παράγουν συστάσεις. Οι δύο αυτές μέθοδοι μπορούν να συνδυαστούν σε υβριδικές προσεγγίσεις.

2.2 Δομικά Στοιχεία Συστημάτων Συστάσεων

Τα βασικά δομικά στοιχεία ενός συστήματος συστάσεων είναι η είσοδος, η έξοδος και η μέθοδος παραγωγής των προτάσεων. Επιπρόσθετα, μπορούμε να θεωρήσουμε τον τρόπο αποστολής των προτάσεων και το βαθμό εξατομίκευσης. Αυτά τα στοιχεία δεν είναι ανεξάρτητα μεταξύ τους αφού για παράδειγμα συγκεκριμένες σχεδιαστικές επιλογές απαιτούν συγκεκριμένες εισόδους ή συγκεκριμένες έξοδοι μπορούν να παραχθούν μόνο με κάποιες και όχι με όλες τις μεθόδους προτάσεων.

2.2.1 Η Είσοδος

Ένα σύστημα συστάσεων δέχεται σαν είσοδο μια συλλογή από δεδομένα που μπορεί να περιλαμβάνουν μεταξύ άλλων τις προτιμήσεις των πελατών, τα χαρακτηριστικά των προϊόντων, τις συσχετίσεις των συναλλαγών, κλπ. ώστε να παράγει εξόδους-προτάσεις. Τα δεδομένα αυτά καλύπτουν ένα μεγάλο εύρος και μπορούν να διακριθούν σε σχέση με την προέλευσή τους σε α) δεδομένα για τον πελάτη στον οποίο θέλουμε να προτείνουμε προϊόντα και β) γενικά δεδομένα από τις κοινότητες ή τη συγκεκριμένη κοινότητα που ανήκει ο πελάτης(εικόνα 1).



Εικόνα 1: Τα δομικά στοιχεία ενός συστήματος συστάσεων

Προσανατολισμένα στον πελάτη δεδομένα εισόδου

Τα δεδομένα αυτά εισάγονται στο σύστημα συστάσεων ώστε να παραχθούν προσανατολισμένες στον πελάτη προτάσεις. Ένα σύστημα που δεν χρησιμοποιεί πληροφορίες για κάθε πελάτη ξεχωριστά, μπορεί να εμφανίσει μόνο γενικές προτάσεις. Οι κυριότεροι τύποι δεδομένων αφορούν:

- Υπονοούμενη πλοήγηση (implicit navigation). Τα δεδομένα σε αυτή την περίπτωση εξάγονται από την πλοηγητική συμπεριφορά του πελάτη χωρίς αυτός να είναι ενήμερος για την παρουσία του συστήμα-

- Σαφής πλοήγηση (explicit navigation). Σε αντίθεση με την υπονοούμενη, στη σαφή πλοήγηση οι πελάτες εισάγουν τα δεδομένα με σκοπό να πληροφορήσουν το σύστημα για τις προτιμήσεις τους. Ο πιο απλός τρόπος είναι με τη χρήση ερωτηματολογίων ή φορμών που συγκεντρώνουν δημογραφικά ή άλλα στοιχεία.
- Λέξεις κλειδιά (keywords) και χαρακτηριστικά των προϊόντων (attributes). Συχνά τα δεδομένα από τον πελάτη δεν περιορίζονται σε μια απλή κατηγορία ή κάποιο προϊόν που τον ενδιαφέρει. Μπορεί επίσης να χρησιμοποιηθούν λέξεις κλειδιά ή χαρακτηριστικά των προϊόντων είτε με σαφή τρόπο μέσω π.χ. μιας μηχανής αναζήτησης, είτε με υπονοούμενο τρόπο π.χ. από τα προϊόντα που ο πελάτης βλέπει την τρέχουσα στιγμή. Σε κάθε περίπτωση, αυτά τα δεδομένα μεταφράζονται σαν είσοδος που υποδεικνύει τα τρέχοντα ενδιαφέροντα του πελάτη.
- Βαθμολόγηση (rating). Οι πελάτες παρέχουν τις πιο χρήσιμες και σαφείς πληροφορίες στο σύστημα συστάσεων όταν αξιολογούν τα προϊόντα που έχουν ήδη αγοράσει. Η αξιολόγηση μπορεί να γίνει με βαθμούς π.χ. «βαθμολόγησε ένα προϊόν σε μια κλίμακα από 1-5» ή με απλή ερώτηση της μορφής «σου άρεσε αυτό;».
- Ιστορικό συναλλαγών (transactions history). Το ιστορικό μπορεί να χρησιμοποιηθεί σαν μια μορφή υπονοούμενης βαθμολόγησης. Με βάση αυτό μπορούν να συσχετισθούν προϊόντα και να εξαχθούν συμπεράσματα για τις προτιμήσεις των πελατών.

Δεδομένα από τις κοινότητες

Τα δεδομένα αυτά περιλαμβάνουν μια ποικιλία πληροφοριών για τον τρόπο που τα διάφορα μέλη της κοινότητας ή ολόκληρη η κοινότητα αντιλαμβάνεται τα προϊόντα. Δεδομένα εισόδου που αντανακλούν τις απόψεις μιας κοινότητας είναι εκείνα τα χαρακτηριστικά των προϊόντων (item attributes) στα οποία αποδίδονται ετικέτες ή κατηγορίες π.χ. το είδος ενός βιβλίου ή μιας ταινίας. Ομοίως, η

δημοτικότητα ενός προϊόντος (item popularity) μπορεί να εκφράζει τις συνολικές πωλήσεις ή μια λίστα με best-sellers. Με τον ίδιο τρόπο που χρησιμοποιείται το ιστορικό συναλλαγών για να παραχθούν προτάσεις για έναν πελάτη, μπορεί να χρησιμοποιηθεί και το ιστορικό συναλλαγών μιας κοινότητας (community transactions history). Τα παραπάνω μπορούν να συνδυασθούν για να παραχθεί η λίστα των καλύτερων προϊόντων ενός site ή να εξαχθούν συμπεράσματα σχετικά με την κίνηση των πωλήσεων ή τις ομοιότητες/διαφορές των προϊόντων. Παράλληλα, πολλά sites ενθαρρύνουν τα σχόλια κειμένου (text comments) από τους πελάτες τους, τα οποία αν και θεωρούνται πολύ χρήσιμα, εντούτοις απαιτούν αρκετό χρόνο επεξεργασίας, αφού κάποιος πρέπει να διαβάσει μια παράγραφο για να καταλάβει αν το σχόλιο είναι θετικό ή αρνητικό. Για να διευκολυνθεί αυτή η διαδικασία τα περισσότερα συστήματα συστάσεων παρέχουν δυνατότητες βαθμολόγησης (rating) των προϊόντων.

2.2.2 Η Έξοδος

Η έξοδος ενός συστήματος συστάσεων εξαρτάται από τον τύπο, την ποσότητα και την ποιότητα των δεδομένων που ανέλυσε. Ο πιο κοινός τύπος εξόδου είναι οι προτάσεις (suggestions). Αυτές συνήθως έχουν τη μορφή «δοκίμασε αυτό» ή «αυτό» στη σελίδα που ο χρήστης πλοηγείται. Ένας δεύτερος τύπος αφορά σε προβλέψεις (predictions) για τη βαθμολόγηση που θα έδινε ο πελάτης σε ένα προϊόν. Μπορούν να παρουσιάζονται σαν μια μεμονωμένη πρόταση ή σαν μια λίστα προτάσεων ή να εμφανίζονται στη σελίδα παρουσίασης του προϊόντος. Όταν οι κοινότητες είναι μικρές ή τα μέλη γνωρίζονται μεταξύ τους είναι χρήσιμο να παρουσιάζονται οι βαθμολογίες (ratings) συγκεκριμένων μελών ώστε κάθε χρήστης να μπορεί να εξάγει τα δικά του συμπεράσματα για τη σημασία της πρότασης. Αυτή η τεχνική είναι χρήσιμη όταν ο πελάτης μπορεί να επιλέξει γνωστά μέλη της κοινότητας ή όταν οι βαθμολογίες συνοδεύονται και από κριτικές (reviews). Οι κριτικές είναι παραδείγματα προτάσεων που περιέχουν εκτιμήσεις τις οποίες δεν μπορεί να κατανοήσει πλήρως ένα υπολογιστικό σύστημα. Πράγματι, είναι δύσκολο να διακριθούν τα θετικά ή τα αρνητικά σχόλια για ένα προϊόν και για αυτό το λόγο συχνά συνοδεύονται και από αριθμητικές βαθμολογίες.

Μερικά αντιπροσωπευτικά παραδείγματα εξόδων από το σύστημα συστάσεων του Amazon.com (<http://www.amazon.com>) είναι τα ακόλουθα:

- Customers who bought. Η επιλογή αυτή εμφανίζεται μέσα στη σελίδα παρουσίασης του βιβλίου και προτείνει βιβλία που αγοράζονται συχνά από τους πελάτες που έχουν αγοράσει το επιλεγμένο βιβλίο (εικόνα 2).

Customers Who Bought This Item Also Bought

Item	Author	Price	Reviews
	Deception Point by Dan Brown	\$12.48	707
	Digital Fortress: A Thriller by Dan Brown	\$9.99	956
	Holy Blood, Holy Grail by Michael Baigent	\$7.99	483
	The Da Vinci Code (Two-Disc Extended Cut + BD Liv... DVD ~ Tom Hanks)	\$23.99	619

Εικόνα 2: Customers who bought

- Your recommendations. Το Amazon.com ενθαρρύνει την άμεση ανάδραση (feedback) από τους πελάτες για τα βιβλία που έχουν διαβάσει. Οι πελάτες βαθμολογούν τα βιβλία με μια κλίμακα από 1-5 από «δεν μου αρέσει καθόλου» έως «μου άρεσε πάρα πολύ» (εικόνα 3).

Rate This Item to Improve Your Recommendations

I own it ★★★★★ Rate this item

Εικόνα 3: Προτρέπει τον χρήστη να αξιολογήσει το προϊόν

- Editorial Reviews και Customer Reviews. Οι πελάτες μπορούν να διαβάσουν σκέψεις και κριτικές της εκδοτικής εταιρίας, από εφημερίδες και περιοδικά σχετικά του είδους. Επίσης, υπάρχουν αναφορές πελατών οι οποίες βαθμολογούνται για το κατά πόσο ήταν βοηθητικές (εικόνα 4).

Customer Reviews

3,932 Reviews

5 star:		(1,601)	Average Customer Review ★★★★☆ (3,932 customer reviews)
4 star:		(659)	
3 star:		(542)	
2 star:		(414)	
1 star:		(716)	

Most Helpful Customer Reviews

246 of 268 people found the following review helpful:

★★★★☆ **Just Read It, DON'T Base Your Life On It!**, October 18, 2003
By [Janet](#) (outside Portland) - [See all my reviews](#)
This review is from: [The Da Vinci Code \(Hardcover\)](#)

An excellent read, but it's truly SAD to think that some readers assume that Dan Brown's contrived history is factual and would even base their spiritual beliefs on a book of fiction. Just read some of the other reviews to see what I'm talking about. It reminds of the guy who watched too many episodes of Highlander and decided he was an immortal! (I'm not making this up.)

One reader compared Da Vinci Code to James BeauSeigneur's Christ Clone Trilogy and suggested that like BeauSeigneur, Brown should footnote all the factual material. While BeauSeigneur and Brown have a similar style and both deal with controversial religious topics, BeauSeigneur can footnote the facts in his fiction BECAUSE THEY ARE FACTS. Brown's "facts" cannot be footnoted because they are a fictitious as the rest of the book.

[Comments \(3\)](#) | [Permalink](#) | Was this review helpful to you? ([Report this](#))

Editorial Reviews

Amazon.com Review
With *The Da Vinci Code*, Dan Brown masterfully concocts an intelligent and lucid thriller that marries the gusto of an international murder mystery with a collection of fascinating esoterica culled from 2,000 years of Western history.

A murder in the silent after-hour halls of the Louvre museum reveals a sinister plot to uncover a secret that has been protected by a clandestine society since the days of Christ. The victim is a high-ranking agent of this ancient society who, in the moments before his death, manages to leave gruesome clues at the scene that only his granddaughter, noted cryptographer Sophie Neveu, and Robert Langdon, a famed symbologist, can untangle. The duo become both suspects and detectives searching for not only Neveu's grandfather's murderer but also the stunning secret of the ages he was charged to protect. Mere steps ahead of the authorities and the deadly competition, the mystery leads Neveu and Langdon on a breathless flight through France, England, and history itself. Brown (*Angels and Demons*) has created a page-turning thriller that also provides an amazing interpretation of Western history. Brown's hero and heroine embark on a lofty and intriguing exploration of some of Western culture's greatest mysteries--from the nature of the Mona Lisa's smile to the secret of the Holy Grail. Though some will quibble with the veracity of Brown's conjectures, therein lies the fun. *The Da Vinci Code* is an enthralling read that provides rich food for thought. --Jeremy Pugh --This text refers to the [Hardcover](#) edition.

Εικόνα 4: Βαθμολόγηση αναφορών πελατών

- **What Do Customers Ultimately Buy After Viewing This Item.** Ο πελάτης μπορεί να πληροφορηθεί τι αγόρασαν οι άλλοι πελάτες αφότου είδαν αυτό το αντικείμενο. Επίσης υπάρχει η επιλογή να δουν «όμοια» αντικείμενα(εικόνα 5).

What Do Customers Ultimately Buy After Viewing This Item?

	76% buy the item featured on this page: The Da Vinci Code ★★★★★ (3,932)
	12% buy Angels & Demons - Movie Tie-In: A Novel ★★★★★ (2,347) \$10.88
	9% buy The Lost Symbol \$16.17
	2% buy Deception Point ★★★★★ (708) \$12.48

[Explore similar items](#)

Εικόνα 5: What Do Customers Ultimately Buy After Viewing This Item

2.2.3 Η Μέθοδος

Στη συνέχεια παρουσιάζονται οι πιο γνωστές μέθοδοι που χρησιμοποιούν τα συστήματα συστάσεων. Κάθε μια αντιπροσωπεύει μια οικογένεια αλγορίθμων και τεχνικών, ενώ πρέπει να σημειωθεί ότι είναι δυνατός ο συνδυασμός περισσότερων της μιας μεθόδου για την παραγωγή των προτάσεων.

- **Raw retrieval.** Στη μέθοδο αυτή παρέχεται στους πελάτες ένα interface αναζήτησης μέσω του οποίου μπορούν να κάνουν ερωτήσεις στη βάση δεδομένων των προϊόντων. Στη συνέχεια, το σύστημα συστάσεων με απλή συντακτική διαδικασία επιστρέφει στον πελάτη ό,τι ζήτησε. Για παράδειγμα, όταν ο πελάτης ζητά από ένα online βιβλιοπωλείο ένα βιβλίο του Dan Brown, τότε το σύστημα επιστρέφει μια λίστα με όλα τα βιβλία του, δίνοντας τη δυνατότητα στον πελάτη να βρει ίσως κάποιο βιβλίο του συγγραφέα που δεν γνώριζε.
- **Manual selection.** Ορισμένα συστήματα συστάσεων παράγουν προτάσεις με βάση δεδομένα που έχουν συλλεχθεί χειρωνακτικά από συντάκτες, καλλιτέχνες, κριτικούς ή άλλους ειδικούς. Αυτές οι προτάσεις αντανakλούν τις προτιμήσεις, τα ενδιαφέροντα, τους υποκειμενικούς λόγους και σκοπούς των δημιουργών τους και μπορεί συχνά να συνοδεύονται από σχόλια κειμένου. Για παράδειγμα, αν ο χρήστης επιλέξει ένα συγκεκριμένο είδος ταινιών π.χ. θρίλερ, τότε εμφανίζεται μια λίστα ταινιών τις οποίες κάποιος κριτικός θεωρεί σαν τα καλύτερα θρίλερ όλων των εποχών. Η μέθοδος δεν κάνει υπολογιστική επεξεργασία, απλώς αναπαράγει ό,τι θα εμφανιζόταν σε μια στήλη περιοδικού κριτικής ταινιών.
- **Statistical summaries.** Αρκετά συστήματα συστάσεων παρέχουν στατιστικές αναλύσεις των απόψεων μιας κοινότητας. Αυτά τα στοιχεία μπορεί να περιλαμβάνουν μετρήσεις δημοτικότητας μέσα στην κοινότητα (π.χ. ποσοστό ανθρώπων που τους αρέσει ή αγόρασαν ένα συγκεκριμένο προϊόν) ή συνοπτικές βαθμολογίες (π.χ. ο αριθμός των ατόμων που προτείνουν ένα προϊόν ή η μέση βαθμολογία για ένα

- **Attribute-based technologies.** Αυτά τα συστήματα συστάσεων βασίζονται στα συντακτικά χαρακτηριστικά των προϊόντων. Αν και στην πιο απλή μορφή τους εκφυλίζονται στη μέθοδο raw retrieval, όσα συστήματα χρησιμοποιούν αυτό το μοντέλο προχωράνε ένα βήμα παραπέρα από το απλό ερώτημα. Για παράδειγμα, ένα πελάτης πλοηγείται στο τμήμα «κλασική μουσική» ενός e-shop και έχει στο καλάθι του διάφορα CDs ειδικής προσφοράς. Τότε μπορεί να λάβει μια πρόταση για έκπτωση σε κάποιο CD κλασικής μουσικής. Άλλα συστήματα για να παράγουν προτάσεις χρησιμοποιούν τα προφίλ των χρηστών που υποδεικνύουν τις προτιμήσεις τους π.χ. μπορεί ένας πελάτης να αγοράζει μόνο CDs που έχουν έκπτωση, ενώ κάποιος άλλος ποτέ δεν αγοράζει μουσική της δεκαετίας του '70.
- **Item-to-item correlation.** Πολλά συστήματα συστάσεων αναγνωρίζουν τα προϊόντα που εμφανίζονται πιο συχνά μαζί με τα προϊόντα για τα οποία ο χρήστης έχει εκφράσει κάποιο ενδιαφέρον. Η συσχέτιση μπορεί να βασίζεται σε δεδομένα αγορών, προτιμήσεων από κοινούς πελάτες, ή άλλες μετρικές. Στην πιο απλή μορφή της, μπορεί να χρησιμοποιηθεί για να αναγνωρίσει ταιριαστά είδη για ένα συγκεκριμένο προϊόν όπως άλλου είδους ρουχισμός που αγοράζεται συνήθως μαζί με ένα παντελόνι π.χ. πουκάμισο ή πουλόβερ. Πιο προηγμένα συστήματα συσχετίζουν ένα ολόκληρο σύνολο από προϊόντα π.χ. αυτά που έχει ο χρήστης στο καλάθι του για να παράγουν τις κατάλληλες προτάσεις. Συνήθως, χρησιμοποιούνται οι τρέχουσες πωλήσεις ή άλλα τρέχοντα ενδιαφέροντα αντί για τη μακροπρόθεσμη ιστορία του πελάτη γεγονός που τις κάνει πολύ χρήσιμες για προτάσεις δώρων.

- User-to-user correlation. Πολλά συστήματα συστάσεων προτείνουν προϊόντα σε έναν πελάτη με βάση τη συσχέτισή του με άλλους πελάτες που έχουν αγοράσει προϊόντα από το eshop. Αυτή η τεχνολογία είναι γνωστή και σαν collaborative filtering ή social filtering και προέρχεται από την τεχνική information filtering που χρησιμοποιεί απόψεις ομάδων για να προτείνει προϊόντα σε διάφορα άτομα. Το My CDNOW είναι ένα σύστημα που χρησιμοποιεί user-to-user correlation για να αναγνωρίσει μια κοινότητα από πελάτες που τείνουν να έχουν και να τους αρέσουν τα ίδια σύνολα από CDs. Η τεχνική στηρίζεται στην αρχή ότι εάν διάφορα μέλη της κοινότητας που ανήκει ο πελάτης έχουν και τους αρέσει το τελευταίο άλμπουμ του Sting τότε είναι αρκετά πιθανό να αρέσει και σε αυτόν.

Ένα σημαντικό θέμα που θα πρέπει να προσεχθεί ιδιαίτερα σχετικά με τη μέθοδο recommendation είναι αν ο υπολογισμός θα γίνει εξολοκλήρου online, όταν δηλαδή ο πελάτης αλληλεπιδρά με το e-shop, ή μέρη του να υπολογίζονται offline, κυρίως για λόγους απόδοσης. Οι online προτάσεις προτιμώνται γιατί απαντούν απευθείας στο χρήστη. Οι raw retrieval, manual selection, statistical summaries, attribute-based μέθοδοι εκτελούν απλούς υπολογισμούς που συνήθως γίνονται στη διάρκεια της αλληλεπίδρασης του πελάτη με το site. Οι item-to-item correlation και user-to-user correlation τεχνικές είναι περισσότερο πολύπλοκες υπολογιστικά και συχνά απαιτούν μια offline επεξεργασία η οποία θα προετοιμάσει το μοντέλο που θα εκτελεσθεί online. Μια πρόκληση στο σχεδιασμό του μοντέλου είναι να διασφαλισθεί ότι το online σύστημα θα παράγει αποτελέσματα που θα ανταποκρίνονται όσο το δυνατόν περισσότερο στην αλληλεπιδραστική είσοδο του χρήστη.

2.3 Τρόποι Υποβολής Συστάσεων

Ο τρόπος υποβολής/αποστολής των προτάσεων στους πελάτες αποτελεί μια κρίσιμη σχεδιαστική απόφαση των e-shops. Οι push τεχνολογίες έχουν το πλεονέκτημα ότι φτάνουν στον πελάτη όταν αυτός δεν αλληλεπιδρά με το κατάστημα. Το e-mail είναι η πιο διαδεδομένη μέθοδος αυτής της κατηγορίας. Η αποστολή προτάσεων και προσφορών μπορεί να οδηγήσει έναν πελάτη στο κατάστημα. Οι σημερινές τεχνολογίες επιτρέπουν στους πελάτες να κάνουν «κλικ» σε ένα link μέσα στο κείμενο του e-mail και να μεταφερθούν κατευθείαν στο προϊόν. Οι εφαρμογές που χρησιμοποιούν pull τεχνολογίες επιτρέπουν στον πελάτη να ελέγξει πότε θα του παρουσιασθούν οι προτάσεις. Ο πελάτης ενημερώνεται για το χρόνο που είναι ενεργές οι προτάσεις (π.χ. εμφανίζοντας ένα link) αλλά δεν παρουσιάζονται σε αυτόν πριν ο ίδιος να τις ζητήσει.

Σήμερα, οι pull τεχνολογίες είναι μια επιλογή σχεδίασης για εφαρμογές όπου οι τύποι των recommendations είναι περιφερειακοί (π.χ. η λίστα των top 10 προτάσεων για δώρα) αντί να είναι ενσωματωμένοι στην εφαρμογή. Οι passive τεχνολογίες παρουσιάζουν τα recommendations στο φυσικό περιβάλλον του e-shop π.χ. προτάσεις που σχετίζονται με το τρέχον προϊόν, προτάσεις για προϊόντα που σχετίζονται με το θέμα ενός κειμένου, προτάσεις σχετικές με το περιεχόμενο της εξερεύνησης του χρήστη, κλπ. Έχουν το πλεονέκτημα ότι φτάνουν στον πελάτη όταν ο ίδιος είναι ήδη ανοικτός στην ιδέα να λάβει προτάσεις. Τα e-shops χρησιμοποιούν αυτή την τεχνολογία σαν ένα μέρος της διαδικασίας παραγγελίας, προτείνοντας καλύτερες επιλογές αποστολής των προϊόντων, ή την ώρα που ο πελάτης ολοκληρώνει την παραγγελία του γεγονός που την κάνει πιο αποδοτική.

Γενικά, οι μέθοδοι αποστολής αλλάζουν με την πάροδο του χρόνου. Οι πρώιμες εφαρμογές συστήματα συστάσεων εστίαζαν σε push και pull τεχνολογίες γιατί η επιθυμία να δείξουν ενδιαφέρον στους πελάτες ήταν σημαντική. Πρόσφατα, οι εφαρμογές κινούνται προς την κατεύθυνση των passive και push τεχνολογιών, με την πρώτη να εφαρμόζεται στα web sites και τη δεύτερη να χρησιμοποιείται για να φέρει τους πελάτες πίσω σε αυτά.

2.4 Βαθμός Εξατομίκευσης

Τα συστήματα συστάσεων μπορούν να παράγουν προτάσεις με διαφορετικό βαθμό εξατομίκευσης (personalization degree). Όταν το σύστημα προτείνει όμοια recommendations σε κάθε πελάτη τότε η εφαρμογή θεωρείται ότι είναι μη-εξατομικευμένη (non-personalized). Τα συγκεκριμένα recommendations μπορεί να βασίζονται σε επιλογή με το χέρι, στατιστικά στοιχεία, ή άλλες τεχνικές. Οι μεγαλύτερες πωλήσεις ενός προϊόντος, οι επιλογές των κριτικών, οι μέσοι όροι, τα ακατέργαστα σχόλια, κλπ. παρουσιάζουν τις ίδιες προτάσεις σε όλους τους πελάτες του e-shop. Τα συστήματα συστάσεων που χρησιμοποιούν δεδομένα των πελατών για να προσαρμόσουν τις προτάσεις στα ενδιαφέροντά τους παρέχουν εφήμερη εξατομίκευση (ephemeral personalization). Στην περίπτωση αυτή, οι προτάσεις ανταποκρίνονται στην πλοήγηση του πελάτη και στις επιλογές του. Αυτά τα συστήματα βασίζονται συνήθως σε item-to-item correlation, attribute-based recommendation ή και τα δυο μαζί. Τέλος, οι εφαρμογές που χρησιμοποιούν επίμονη εξατομίκευση (persistent personalization) μπορούν να δημιουργήσουν προτάσεις που είναι διαφορετικές για διαφορετικούς πελάτες ακόμα και αν αυτοί κοιτάνε τα ίδια προϊόντα. Συνήθως χρησιμοποιούν item-to-item correlation, user-to-user correlation ή attribute-based recommendation. Απαιτούν από τους πελάτες να συντηρούν μόνιμες ταυτότητες (profiles) αλλά τους επιβραβεύουν με μεγαλύτερο βαθμό εξατομίκευσης.

2.5 Ταξινόμηση Συστημάτων Συστάσεων

Όπως έχουμε ήδη αναφέρει τα συστήματα συστάσεων μπορούν να ταξινομηθούν σε τρεις βασικές κατηγορίες: content-based filtering (φιλτράρισμα με βάση το περιεχόμενο), collaborative filtering (συνεργατικό φιλτράρισμα) και hybrid approaches (υβριδικές προσεγγίσεις). Τα τελευταία χρόνια σημαντική έρευνα έχει γίνει στην περιοχή. Η εικόνα 6 παρουσιάζει μια ταξινόμηση των τεχνικών RSs καθώς και αντιπροσωπευτικά ερευνητικά παραδείγματα (Adomavicius & Tuzhilin, 2005).

Κατηγορίες τεχνικών RSs	Ευριστικές (heuristic-based)	Μοντέλου (model-based)
Φιλτράρισμα με βάση το περιεχόμενο ή content-based filtering	<p>Τεχνικές:</p> <ul style="list-style-type: none"> • TF-IDF (Information Retrieval) • Clustering <p>Αντιπροσωπευτικά παραδείγματα:</p> <ul style="list-style-type: none"> • Lang, 1995 • Balabanovic & Shoham, 1997 • Pazzani & Billsus, 1997 	<p>Τεχνικές:</p> <ul style="list-style-type: none"> • Bayesian classifiers • Clustering • Decision trees • Artificial neural networks <p>Αντιπροσωπευτικά παραδείγματα:</p> <ul style="list-style-type: none"> • Pazzani & Billsus, 1997 • Mooney et al., 1998 • Mooney & Roy, 1999 • Billsus & Pazzani, 1999, 2000
Συνεργαστικό φιλτράρισμα ή collaborative filtering	<p>Τεχνικές:</p> <ul style="list-style-type: none"> • Nearest neighbor (cosine, correlation) • Clustering • Graph theory <p>Αντιπροσωπευτικά παραδείγματα:</p> <ul style="list-style-type: none"> • Resnick et al., 1994 • Hill et al., 1995 • Shardanand & Maes, 1995 • Breese et al., 1998 • Nakamura & Abe, 1998 • Aggarwal et al., 1999 • Delgado & Ishii, 1999 • Pennock & Horwitz, 1999 • Sarwar et al., 2001 	<p>Τεχνικές:</p> <ul style="list-style-type: none"> • Bayesian networks • Clustering • Artificial neural networks • Linear regression • Probabilistic models <p>Αντιπροσωπευτικά παραδείγματα:</p> <ul style="list-style-type: none"> • Billsus & Pazzani, 1998 • Breese et al., 1998 • Ungar & Foster, 1998 • Chien & George, 1999 • Getoor & Sahami, 1999 • Pennock & Horwitz, 1999 • Goldberg et al., 2001 • Kumar et al., 2001 • Pavlov & Pennock, 2002 • Shani et al., 2002 • Yu et al., 2002, 2004 • Hofmann 2003, 2004 • Marlin, 2003 • Si & Jin, 2003
Υβριδικές προσεγγίσεις ή hybrid approaches	<p>Τεχνικές:</p> <ul style="list-style-type: none"> • Linear combination of predicted ratings • Various voting schemes • Incorporating one component as a part of the heuristic for the other <p>Αντιπροσωπευτικά παραδείγματα:</p> <ul style="list-style-type: none"> • Balabanovic & Shoham, 1997 • Craypool et al., 1999 • Good et al., 1999 • Pazzani, 1999 • Billsus & Pazzani, 2000 • Tran & Cohen, 2000 • Melville et al., 2002 	<p>Τεχνικές:</p> <ul style="list-style-type: none"> • Incorporating one component as a part of the model for the other • Building one unifying model <p>Αντιπροσωπευτικά παραδείγματα:</p> <ul style="list-style-type: none"> • Basu et al., 1998 • Condliff et al., 1999 • Soboroff & Nicholas, 1999 • Ansari et al., 2000 • Popescul et al., 2001 • Schein et al., 2002

Εικόνα 6: Ταξινόμηση των τεχνικών των συστημάτων συστάσεων

2.5.1 Content-based approaches

Η content-based προσέγγιση για την παραγωγή recommendations έχει τις ρίζες της στις επιστημονικές περιοχές της Ανάκτησης Πληροφοριών (Information Retrieval) και του Φιλτραρίσματος Πληροφοριών (Information Filtering) και υιοθετεί τεχνικές που έχουν μελετηθεί εκτενώς στο παρελθόν. Η βελτίωση των παραδο-

σιακών τεχνικών προέρχεται από τη χρήση των προφίλ των χρηστών που περιέχουν πληροφορίες για την ταυτότητά τους, τις προτιμήσεις, τα ενδιαφέροντα και τις ανάγκες τους. Οι πληροφορίες αυτές μπορούν να εξαχθούν από τους χρήστες ρητά π.χ. μέσω ερωτηματολογίων, ή υπονοούμενα μέσω της παρακολούθησης της συμπεριφοράς και των συναλλαγών τους κατά τη διάρκεια του χρόνου. Οι αλγόριθμοι αυτοί συνήθως χρησιμοποιούνται όταν οι συστάσεις αφορούν σε ιστοσελίδες, δημοσιεύσεις, αστεία, ειδήσεις, κλπ. Το σύστημα συλλέγει πληροφορίες για τις προτιμήσεις των χρηστών είτε μέσω της διαδικασίας εγγραφής στο site, είτε μέσω της εκτίμησης των εγγράφων. Οι συστάσεις παράγονται στη συνέχεια λαμβάνοντας υπόψη το περιεχόμενο των εγγράφων και φιλτράρονται ώστε να επιτευχθεί η καλύτερη αντιστοίχιση με τις προτιμήσεις του χρήστη όπως αυτές έχουν καταγραφεί στο προφίλ του. Για παράδειγμα, ένα σύστημα συστάσεων κινηματογραφικών ταινιών προκειμένου να παράγει προτάσεις για ταινίες m_i για το χρήστη u , προσπαθεί να βρει μεταξύ των ταινιών m_j που ο u έχει βαθμολογήσει υψηλά στο παρελθόν τις ομοιότητες (π.χ. ηθοποιοί, σκηνοθέτες, είδος ταινιών, κλπ.). Έτσι, μόνο οι ταινίες που έχουν υψηλό βαθμό ομοιότητας με τις προτιμήσεις του χρήστη προτείνονται. Έστω $Content(m)$ είναι το προφίλ ενός αντικειμένου m , δηλαδή ένα σύνολο από χαρακτηριστικά που αντιπροσωπεύουν το περιεχόμενο του m και το χαρακτηρίζουν. Το περιεχόμενο αυτό συνήθως περιγράφεται με λέξεις-κλειδιά (keywords). Για παράδειγμα, το σύστημα FAB (Balabanovic & Shoham, 1997) προτείνει Web σελίδες στους χρήστες του και ορίζει τα περιεχόμενα της σελίδας με βάση τις 100 πιο σημαντικές λέξεις της. Η «σημαντικότητα» της λέξης k_i στο έγγραφο d_j καθορίζεται με μια μετρική βάρους w_{ij} που μπορεί να οριστεί με διάφορους τρόπους. Μια από τις πιο γνωστές μετρικές από την περιοχή του Information Retrieval είναι η TF – IDF (term frequency/inverse document frequency) που ορίζεται ως εξής: υποθέτουμε ότι N είναι ο συνολικός αριθμός των documents που μπορούν να προταθούν στους χρήστες και το keyword k_j εμφανίζεται σε n_j από αυτούς. Σαν f_{ij} ορίζουμε τον αριθμό εμφάνισης (συχνότητα) του keyword k_i στο document d_j . Τότε η TF_{ij} (term frequency) του keyword k_i στο document d_j ορίζεται σαν:

$$TF_{ij} = \frac{f_{ij}}{\max_z f_{zj}}$$

όπου το \max υπολογίζεται σε όλες τις συχνότητες f_{zj} όλων των k_z που εμφανίζονται σε ένα d_j . Εντούτοις, τα keywords που εμφανίζονται σε πολλά έγγραφα δεν μπορούν πάντα να διακρίνουν αν ένα έγγραφο είναι σχετικό ή όχι. Έτσι, η μετρική IDF_i (inverse document frequency) χρησιμοποιείται σε συνδυασμό με την TF_{ij} και ορίζεται για το keyword k_i σαν:

$$IDF_i = \log \frac{N}{n_i}$$

Τότε, το TF-IDF βάρος w_{ij} του keyword k_i στο document d_j δίνεται από τη σχέση:

$$w_{ij} = TF_{ij} \times IDF_i$$

και το περιεχόμενο (content) του document d_j ορίζεται σαν:
 $Content(d_j) = (w_{1j}, \dots, w_{kj})$

Έστω $ContentBasedProfile(u)$ είναι το προφίλ του χρήστη u που περιλαμβάνει τα χαρακτηριστικά, τα ενδιαφέροντα και τις προτιμήσεις του. Αυτό το προφίλ δημιουργείται από την επεξεργασία του περιεχομένου των αντικειμένων που έχει δει και βαθμολογήσει και βασίζεται με τεχνικές keywords ανάλυσης. Για παράδειγμα, μπορεί να ορισθεί σαν ο πίνακας των βαρών (w_{u1}, \dots, w_{uk}) όπου κάθε βάρος w_{ui} δηλώνει τη σημασία του keyword k_i για το χρήστη u και μπορεί να εκτιμηθεί από τις ατομικές βαθμολογίες με διάφορες τεχνικές π.χ. μέσος όρος. Αυτό που πρέπει να υπολογισθεί στη συνέχεια είναι το:

$$score(ContentBasedProfile(u), Contents(m))$$

Τόσο το $ContentBasedProfile(u)$ του χρήστη u όσο και το $Contents(m)$ του εγγράφου m μπορούν να αναπαρασταθούν σαν $TF - IDF$ πίνακες από βάρη \overline{w}_u και \overline{w}_m αντίστοιχα. Το $score$ μπορεί να είναι η μετρική cosine similarity:

$$\cos(\overline{w}_u, \overline{w}_m) = \frac{\overline{w}_u \times \overline{w}_m}{|\overline{w}_u|^2 \times |\overline{w}_m|^2} = \frac{\sum_{i=1}^K w_{iu} \times w_{im}}{\sqrt{\sum_{i=1}^K w_{iu}^2} \times \sqrt{\sum_{i=1}^K w_{im}^2}}$$

όπου K είναι ο συνολικός αριθμός των keywords στο σύστημα. Για παράδειγμα, αν ο χρήστης u διαβάζει πολλά online άρθρα για Βιοπληροφορική, τότε το σύστημα συστάσεων θα του προτείνει και άλλα άρθρα Βιοπληροφορικής. Αυτό συμβαίνει γιατί τα άρθρα αυτά περιέχουν τους σχετικούς όρους π.χ. γωνιδίωμα,

ακολουθία, κλπ. περισσότερο από ότι τα άρθρα με άλλα θεματικά. Οπότε το ContentBasedProfile(u) όπως παριστάνεται από το $\overline{w_u}$ θα αναπαριστά του όρους k_i με υψηλά βάρη w_{ui} . Το σύστημα συστάσεων χρησιμοποιώντας την cosine ή κάποια άλλη μετρική θα δώσει υψηλό score σε αυτά τα άρθρα m που έχουν υψηλά βαθμολογημένους τους όρους Βιοπληροφορικής στο $\overline{w_m}$ και χαμηλό score στα υπόλοιπα άρθρα που οι όροι αυτοί είναι χαμηλά βαθμολογημένοι.

Άλλες τεχνικές της κατηγορίας που έχουν επίσης χρησιμοποιηθεί είναι οι Bayesian Classifiers (Ταξινομητές) (Pazzani & Billsus, 1997), οι Machine Learning τεχνικές συμπεριλαμβανομένων και τεχνικών Clustering (Συσταδοποίησης), Decision Trees (Δέντρων Απόφασης) και Artificial Neural Networks (Τεχνητών Νευρωνικών Δικτύων) (Pazzani & Billsus, 1997). Αυτές οι τεχνικές διαφέρουν στο γεγονός ότι υπολογίζουν τις προβλέψεις όχι με ευρυστικό τρόπο π.χ. το μέτρο ομοιότητας συνημίτονου (cosine similarity measure), αλλά βασισμένες σε ένα μοντέλο που εκπαιδεύεται από συλλεγμένα στοιχεία χρησιμοποιώντας στατιστικές και Machine Learning τεχνικές. Για παράδειγμα, με βάση ένα σύνολο ιστοσελίδων που εκτιμήθηκαν ως «σχετικές» ή «άσχετες» από το χρήστη (Pazzani & Billsus, 1997) χρησιμοποιούμε τον naïve Bayesian Classifier που κατηγοριοποιεί τις αταξινομητες ιστοσελίδες. Συγκεκριμένα, ο ταξινομητής υπολογίζει την πιθανότητα ότι μια σελίδα p_j ανήκει σε μια συγκεκριμένη κλάση C_i (π.χ. σχετική ή άσχετη) δεδομένου ενός συνόλου από keywords k_{1j}, \dots, k_{nj} της σελίδας: $P(C_i | k_{1j} \& \dots \& k_{nj})$. Χρησιμοποιώντας την υπόθεση ότι τα keywords είναι ανεξάρτητα μεταξύ τους, τότε η προηγούμενη πιθανότητα είναι ανάλογη του $P(C_i) \prod_x P(k_{xj} | C_i)$. Πειραματικά αποτελέσματα έδειξαν ότι ο naïve Bayesian Classifier παρουσιάζει υψηλή ακρίβεια στην ταξινόμηση. Επιπλέον, οι $P(k_{xj} | C_i)$ και $P(C_i)$ μπορούν να υπολογιστούν από δεδομένα εκπαίδευσης. Έτσι για κάθε σελίδα p_j , η πιθανότητα $P(C_i | k_{1j} \& \dots \& k_{nj})$ υπολογίζεται για κάθε κλάση C_i και η σελίδα p_j τοποθετείται στην κλάση C_i που έχει τη μεγαλύτερη πιθανότητα.

2.5.2 Collaborative filtering approaches

Σε αντίθεση με τις content-based τεχνικές, τα collaborative filtering συστήματα προσπαθούν να προβλέψουν τη χρησιμότητα των αντικειμένων για ένα συγκεκριμένο χρήστη λαμβάνοντας υπόψη τα αντικείμενα που έχουν βαθμολογηθεί προηγουμένως από άλλους χρήστες. Αυτό σημαίνει ότι η εκτίμηση του αντικειμένου m για το χρήστη u υπολογίζεται με βάση τις εκτιμήσεις που δίνονται στο m από εκείνους τους χρήστες u_j που είναι «παρόμοιοι» με τον u . Για παράδειγμα ένα CF-Recommendation System προκειμένου να προτείνει ταινίες στο χρήστη u προσπαθεί να βρει άλλους χρήστες με παρόμοιες προτιμήσεις στις ταινίες (έχουν εκτιμήσει τις ίδιες ταινίες με παρόμοιο τρόπο). Στη συνέχεια, θα προτείνει μόνο εκείνες τις ταινίες που άρεσαν περισσότερο στους «γείτονες» του u . Σύμφωνα με τους Breese et al., (1998), τα CF-Recommendation Systems χωρίζονται σε δυο κατηγορίες: στα memory-based (ή heuristic-based) και στα model-based.

2.5.2.1 Memory-based τεχνικές

Οι memory-based αλγόριθμοι αποτελούν heuristics που κάνουν προβλέψεις βασισμένοι σε μια ολόκληρη συλλογή από προηγούμενα βαθμολογημένα αντικείμενα από τους χρήστες. Η τιμή μιας άγνωστης βαθμολογίας r_{um} για το χρήστη u και το αντικείμενο m συνήθως υπολογίζεται σαν η συνάθροιση των βαθμολογιών κάποιων άλλων χρηστών (συνήθως των N περισσότερο όμοιων) για το ίδιο αντικείμενο m : $r_{um} = \text{aggr}(r_{u'm})$ όπου το \bar{U} είναι το σύνολο των N χρηστών που είναι πιο όμοιοι με τον u και έχουν βαθμολογήσει το αντικείμενο m (το N μπορεί να κυμαίνεται από το 1 έως τον αριθμό όλων των χρηστών).

Μερικά παραδείγματα της aggregation συνάρτησης είναι:

$$r_{um} = \frac{1}{N} \sum_{u' \in \bar{U}} r_{u'm} \quad (\alpha),$$

$$r_{um} = k \sum_{u' \in \bar{U}} \text{sim}(u, u') \times r_{u'm} \quad (\beta),$$

$$r_{um} = \bar{r}_u + k \sum_{u' \in \bar{U}} \text{sim}(u, u') \times (r_{u'm} - \bar{r}_{u'}) \quad (\gamma)$$

όπου ο πολλαπλασιαστής k λειτουργεί σαν παράγοντας κανονικοποίησης και είναι:

$$k = \frac{1}{|\sum_{u',v'} sim(u,u')|}$$

ενώ η μέση εκτίμηση \bar{r}_u του χρήστη u στην (γ) δίνεται από τον τύπο:

$$\bar{r}_u = \left(\frac{1}{|M_u|}\right) \sum_{m \in M_u} r_{um}$$

όπου $M_u = \{m \in M \mid r_{um} \neq \emptyset\}$ ($= \emptyset$ σημαίνει ότι το αντικείμενο m δεν έχει βαθμολογηθεί από το χρήστη u). Στην πιο απλή περίπτωση η aggregation είναι ο μέσος όρος (α), ενώ σε πιο προχωρημένες περιπτώσεις είναι το σταθμισμένο (weighted) άθροισμα (β). Η συνάρτηση ομοιότητας μεταξύ των χρηστών u και u' $sim(u,u')$ είναι μια μετρική απόστασης. Στις περισσότερες προσεγγίσεις για τον υπολογισμό της similarity, η ομοιότητα μεταξύ δύο χρηστών είναι βασισμένη στις εκτιμήσεις των αντικειμένων και από τους δύο χρήστες. Οι πιο δημοφιλείς μετρικές είναι η correlation και η cosine-based. Έστω M_{XY} είναι το σύνολο όλων των αντικειμένων που και οι δύο χρήστες X και Y έχουν εκτιμήσει δηλαδή $M_{XY} = \{m \in M \mid r_{Xm} \neq \emptyset \mid r_{Ym} \neq \emptyset\}$. Το M_{XY} είναι συνήθως ένα ενδιάμεσο αποτέλεσμα στη διαδικασία για τον υπολογισμό των «κοντινότερων γειτόνων» του χρήστη X και υπολογίζεται συχνά σαν η τομή των συνόλων M_X και M_Y . Εντούτοις, μερικές μέθοδοι, όπως η graph-theoretic μπορούν να καθορίσουν τους κοντινότερους γείτονες του X χωρίς τον υπολογισμό του M_{XY} για όλους τους χρήστες Y . Στην περίπτωση της correlation-based, ο συντελεστής Pearson χρησιμοποιείται για να μετρήσει την ομοιότητα (Resnick et al., 1994).

$$sim(X,Y) = \frac{\sum_{m \in M_{XY}} (r_{Xm} - \bar{r}_X) \times (r_{Ym} - \bar{r}_Y)}{\sqrt{\sum_{m \in M_{XY}} (r_{Xm} - \bar{r}_X)^2} \sqrt{\sum_{m \in M_{XY}} (r_{Ym} - \bar{r}_Y)^2}}$$

Στην περίπτωση της cosine-based, οι δύο χρήστες X και Y αντιμετωπίζονται σαν δύο διανύσματα στο μ -διαστάσεων χώρο όπου $\mu = |M_{XY}|$. Η ομοιότητα μεταξύ των δύο διανυσμάτων μπορεί να μετρηθεί με τον υπολογισμό του συνημίτονου της γωνίας μεταξύ τους:

$$sim(X,Y) = \cos(\vec{X}, \vec{Y}) = \frac{\vec{X} \times \vec{Y}}{|\vec{X}| \times |\vec{Y}|} = \frac{\sum_{m \in M_{XY}} r_{Xm} r_{Ym}}{\sqrt{\sum_{m \in M_{XY}} r_{Xm}^2} \sqrt{\sum_{m \in M_{XY}} r_{Ym}^2}}$$

Μια άλλη προσέγγιση στη μέτρηση της ομοιότητας μεταξύ των χρηστών είναι η mean squared difference μετρική (Shardanand & Maes, 1995). Διαφορετικά συστήματα συστάσεων υιοθετούν διαφορετικές μεθόδους προκειμένου να υπολογίσουν την ομοιότητα των χρηστών και να παράγουν εκτιμήσεις όσο το δυνατόν αποτελεσματικότερα. Μια μεθοδολογία είναι να υπολογιστούν όλες οι ομοιότητες $sim(X, Y)$ των χρηστών (συμπεριλαμβανομένου και του υπολογισμού M_{xx}) εκ των προτέρων και να τους υπολογίζει εκ νέου μόνο όταν απαιτείται (δεδομένου ότι δεν συμβαίνουν δραματικές αλλαγές σε μικρό χρονικό διάστημα). Κατόπιν, όποτε ο χρήστης ζητά μια σύσταση, οι εκτιμήσεις μπορούν να υπολογιστούν αποτελεσματικά χρησιμοποιώντας τις προϋπολογισμένες ομοιότητες. Πρέπει να σημειωθεί σε αυτό το σημείο ότι οι content-based και οι CF προσεγγίσεις χρησιμοποιούν τη μετρική cosine από την περιοχή του Information Retrieval. Στην πρώτη όμως περίπτωση χρησιμοποιείται για να μετρήσει την ομοιότητα μεταξύ των διανυσμάτων των $TF - IDF$ βαρών, ενώ στη δεύτερη την ομοιότητα μεταξύ των διανυσμάτων των πραγματικών εκτιμήσεων των χρηστών.

Τροποποιήσεις των παραπάνω έχουν προταθεί σε πολλές εργασίες με στόχο τη βελτίωση της απόδοσης π.χ. default voting, inverse user frequency, case amplification (Breese et al., 1998) και weighted-majority prediction (Delgado & Ishii, 1999). Ενώ οι παραπάνω τεχνικές παραδοσιακά έχουν χρησιμοποιηθεί για να υπολογίσουν τις ομοιότητες μεταξύ των χρηστών, οι Sarwar et al., (2001) πρότειναν να υπολογιστούν οι ομοιότητες μεταξύ των αντικειμένων και να ληφθούν οι εκτιμήσεις από αυτές. Αυτή η ιδέα επεκτάθηκε περαιτέρω στην εργασία των Deshpande & Karypis (2004) για την παραγωγή των N καλύτερων προτάσεων. Τα εμπειρικά αποτελέσματά τους δείχνουν ότι έχουν καλύτερη υπολογιστική απόδοση από τις παραδοσιακές μεθόδους.

2.5.2.2 Model-based τεχνικές

Οι model-based μέθοδοι βασίζονται αρχικά σε ένα σύνολο δεδομένων για να εκπαιδεύσουν ένα μοντέλο, το οποίο χρησιμοποιούν στη συνέχεια για να κάνουν τις προβλέψεις. Οι άγνωστες εκτιμήσεις υπολογίζονται ως εξής:

$$r_{um} = E(r_{um}) = \sum_{i=0}^n i \times \Pr(r_{um} = i | r_{um}, m' \in M_u)$$

Οι τιμές εκτίμησης είναι ακέραιοι αριθμοί μεταξύ 0 και n και η έκφραση πιθανότητας είναι η πιθανότητα ότι ο χρήστης u θα κάνει μια ιδιαίτερη εκτίμηση για το στοιχείο m δεδομένων των προηγούμενων εκτιμήσεων του. Για να υπολογισθεί αυτή η εκτίμηση προτείνονται δύο εναλλακτικά πιθανοτικά μοντέλα: το Cluster model και τα Bayesian δίκτυα. Στην πρώτη περίπτωση «ομοιόδεατες» χρήστες ομαδοποιούνται σε clusters (ομάδες). Λαμβάνοντας υπόψη την ιδιότητα μέλους κάθε ομάδας, οι εκτιμήσεις του χρήστη υποθέτουμε ότι ήταν ανεξάρτητες, δηλαδή η δομή του μοντέλου είναι αυτή του naïve Bayesian. Ο αριθμός των ομάδων και οι παράμετροι του μοντέλου εκπαιδεύονται από τα δεδομένα εκπαίδευσης. Το δεύτερο μοντέλο αντιπροσωπεύει κάθε αντικείμενο στην περιοχή σαν κόμβο σε ένα Bayesian δίκτυο, όπου οι καταστάσεις κάθε κόμβου αντιστοιχούν σε πιθανές εκτιμήσεις για κάθε αντικείμενο. Τόσο η δομή του δικτύου όσο και οι υπό συνθήκη πιθανότητες εκπαιδεύονται από τα δεδομένα. Ένας περιορισμός αυτής της προσέγγισης είναι ότι κάθε χρήστης μπορεί να τοποθετηθεί σε μια μόνο ομάδα, ενώ τα συστήματα συστάσεων μπορούν να ωφεληθούν από τη δυνατότητα να τοποθετούνται οι χρήστες σε διαφορετικές ομάδες. Σε ένα σύστημα συστάσεων που προτείνει βιβλία, ένας χρήστης μπορεί να ενδιαφερθεί για το θέμα «Προγραμματισμός» για εργασιακούς λόγους και για το τελείως διαφορετικό θέμα «Κηπουρική» για χόμπι.

Επιπλέον, οι Billsus & Pazzani, (1998) προτείνουν μια μέθοδο CF στην περιοχή του Machine Learning όπου διάφορες τεχνικές όπως τα Artificial Neural Networks μαζί με τεχνικές εξαγωγής χαρακτηριστικών γνωρισμάτων μπορούν να χρησιμοποιηθούν. Σε ένα άλλο στατιστικό μοντέλο συγκρίθηκαν διαφορετικοί αλγόριθμοι όπως ο K-means Clustering και ο Gibbs sampling. Άλλες μέθοδοι είναι το Bayesian model, το πιθανοτικό relational model, η linear regression τεχνική και το maximum entropy model. Πρόσφατα έχει γίνει προσπάθεια να μοντελοποιηθεί η διαδικασία παραγωγής προτάσεων χρησιμοποιώντας πιο σύνθετα πιθανοτικά πρότυπα. Οι Shani et al., (2002) θεωρούν αυτή τη διαδικασία σαν ένα ακολουθιακό πρόβλημα απόφασης και χρησιμοποιούν Markov διαδι-

κασίες (γνωστή πιθανοτική τεχνική για τη μοντελοποίηση διαδοχικών αποφάσεων) για την παραγωγή των συστάσεων. Στην εργασία (Pennock & Horvitz, 1999) συνδυάζονται memory-based και model-based προσεγγίσεις και αποδεικνύεται εμπειρικά ότι οδηγούν σε καλύτερες προτάσεις. Οι Yu et al., (2002) παρουσιάζουν μια διαφορετική προσέγγιση για τη βελτίωση της απόδοσης των CF αλγορίθμων, όπου το σύνολο δεδομένων που περιλαμβάνει τις καθορισμένες από το χρήστη εκτιμήσεις έχει επιλεγεί προσεκτικά με βάση διάφορες τεχνικές που αποκλείουν το θόρυβο, τον πλεονασμό και εκμεταλλεύονται την αραιότητα των εκτιμήσεων. Τα εμπειρικά αποτελέσματα έδειξαν βελτίωση της ακρίβειας και της αποδοτικότητας των model-based CF αλγορίθμων.

2.5.3 Hybrid approaches

Αρκετά συστήματα συστάσεων χρησιμοποιούν την υβριδική προσέγγιση συνδυάζοντας collaborative και content-based τεχνικές με στόχο να αποφύγουν τους περιορισμούς που ενέχει κάθε μία από αυτές (Balabanovic & Shoham, 1997). Οι διαφορετικοί τρόποι που μπορούν να συνδυαστούν οι δύο τεχνικές είναι οι ακόλουθοι:

- Υλοποίηση των τεχνικών ξεχωριστά και συνδυασμός των προβλέψεων.
- Ενσωμάτωση content-based χαρακτηριστικών σε μια collaborative προσέγγιση.
- Ενσωμάτωση collaborative χαρακτηριστικών σε μια content-based προσέγγιση.
- Δημιουργία ενός γενικού μοντέλου το οποίο ενσωματώνει και content-based και collaborative χαρακτηριστικά.

Συνδυασμός ξεχωριστών συστημάτων συστάσεων

Στην προσέγγιση αυτή υλοποιούνται ξεχωριστά ένα collaborative και ένα content-based σύστημα συστάσεων. Στη συνέχεια μπορούν να εφαρμοστούν δύο σενάρια. Πρώτον μπορούν να συνδυαστούν τα αποτελέσματα (εκτιμήσεις) που λαμβάνονται από τα μεμονωμένα συστήματα συστάσεων σε μια τελική λίστα. Εναλλακτικά, μπορούμε να χρησιμοποιήσουμε κάθε δεδομένη στιγμή ένα

από τα συστήματα συστάσεων, επιλέγοντας εκείνο που δίνει τα καλύτερα αποτελέσματα με βάση κάποια μετρικής «ποιότητας». Για παράδειγμα, το σύστημα DailyLearner (Billsus & Pazzani, 2000) επιλέγει το σύστημα συστάσεων που μπορεί να δώσει τη σύσταση με το υψηλότερο επίπεδο εμπιστοσύνης (confidence), ενώ το (Tran & Cohen, 2000) επιλέγει αυτό το σύστημα συστάσεων του οποίου οι προτάσεις είναι πιο σύμφωνες με τις προηγούμενες εκτιμήσεις του χρήστη.

Ενσωμάτωση content-based χαρακτηριστικών σε collaborative μοντέλο

Πολλά υβριδικά συστήματα συστάσεων όπως στο σύστημα Fab (Balabanovic & Shoham, 1997) και σε αυτό που παρουσιάζεται στην εργασία (Pazzani, 1999) και αναφέρεται σαν «συνεργασία μέσω περιεχομένου» ο συνδυασμός των δύο τεχνικών γίνεται μέσω (α) της δημιουργίας και της διατήρησης των προφίλ των χρηστών με βάση την ανάλυση περιεχομένου και κάνοντας χρήση μεθόδων ανάκτησης πληροφοριών και άλλων content-based τεχνικών και (β) συγκρίνοντας άμεσα τα προφίλ που κατασκευάζονται ώστε να ευρεθούν όμοιοι χρήστες και να παραχθούν οι προτάσεις. Αυτό σημαίνει ότι στους χρήστες μπορούν να προταθούν αντικείμενα τα οποία είτε έχουν υψηλή βαθμολόγηση σε σχέση με το προφίλ του χρήστη, είτε έχουν βαθμολογηθεί υψηλά από έναν άλλο χρήστη με παρόμοιο προφίλ (Balabanovic & Shoham, 1997). Επιπλέον, επιτρέπει να ξεπεραστούν ορισμένα προβλήματα που οφείλονται στην αραιότητα των δεδομένων των CF-Recommender Systems αφού πολλά ζευγάρια χρηστών έχουν ένα σημαντικό αριθμό από κοινά βαθμολογημένα αντικείμενα. Οι (Good et al., (1999) πρότειναν μια παρόμοια προσέγγιση που χρησιμοποιεί μια ποικιλία από διαφορετικούς filterbots (ειδικοί content-based πράκτορες) που λειτουργούν σαν επιπλέον συμμετέχοντες στη CF κοινότητα. Σαν αποτέλεσμα, οι χρήστες των οποίων οι βαθμολογίες συμφωνούν με τις εκτιμήσεις των filterbots μπορούν να λάβουν καλύτερες προτάσεις. Ομοίως, οι (Melville et al., 2002) χρησιμοποιούν την CF προσέγγιση όπου το παραδοσιακό διάνυσμα με τις εκτιμήσεις των χρηστών εμπλουτίζεται με πρόσθετες εκτιμήσεις που υπολογίζονται χρησιμοποιώντας έναν content-based predictor.

Ενσωμάτωση collaborative χαρακτηριστικών σε content-based μοντέλο

Η πιο δημοφιλής προσέγγιση αυτής της κατηγορίας είναι η χρήση της dimensionality reduction τεχνικής σε μια ομάδα από content-based προφίλ. Οι (Soboroff & Nicholas 1999) χρησιμοποιούν latent semantic indexing (LSI) έτσι ώστε να δημιουργηθεί μια collaborative άποψη από μια συλλογή από προφίλ χρηστών, τα οποία αναπαρίστανται σαν διανύσματα όρων οδηγώντας σε βελτίωση της απόδοσης σε σύγκριση με τις content-based τεχνικές.

Γενικό μοντέλο με content-based και collaborative χαρακτηριστικά

Πολλοί ερευνητές ακολουθούν αυτή την προσέγγιση τα τελευταία χρόνια. Για παράδειγμα, οι (Basu et al., 1998) προτείνουν τη χρήση content-based και collaborative χαρακτηριστικών (π.χ. ηλικία ή φύλο των χρηστών ή είδος των ταινιών) σε ένα ενιαίο βασισμένο σε κανόνες (rule-based) ταξινομητή (classifier). Οι (Popescul et al., 2001) προτείνουν μια πιθανοτική μέθοδο που συνδυάζει collaborative και content-based συστάσεις και βασίζεται στη latent semantic analysis (LSA). Μια άλλη προσέγγιση προτείνεται από τους (Condliff et al., 1999), όπου Bayesian regression μοντέλα χρησιμοποιούνται και τα οποία υιοθετούν μεθόδους Markov chain Monte Carlo για παραμετρική εκτίμηση και πρόβλεψη. Ειδικότερα χρησιμοποιείται το προφίλ των χρηστών και των αντικειμένων σε ένα ενιαίο στατιστικό μοντέλο που υπολογίζει τις άγνωστες εκτιμήσεις r_{uj} για το χρήστη u και το αντικείμενο m : $r_{uj} = x_{uj}\mu + z_u + w_j\lambda_j + \theta_{uj}$, με $\theta_{uj} \sim N(0, \sigma^2)$, $\lambda_j \sim N(0, \Lambda)$, $r_j \sim N(0, \Gamma)$ αφορούν στις τυχαίες μεταβλητές που λαμβάνουν υπόψη το θόρυβο επίδρασης, καθώς και απαρατήρητες πηγές ετερογένειας των χρηστών και ετερογένειας των αντικειμένων αντίστοιχα, ενώ τα $u = 1, \dots, U$ και $m = 1, \dots, M$ αντιπροσωπεύουν τους χρήστες και τα αντικείμενα. Επίσης, x_{uj} είναι μια μήτρα που περιέχει τα χαρακτηριστικά των χρηστών και των αντικειμένων, z_u είναι ένα διάνυσμα με τα χαρακτηριστικά των χρηστών και w_j είναι ένα διάνυσμα με τα χαρακτηριστικά των αντικειμένων. Οι άγνωστες παράμετροι αυτού του μοντέλου μ , σ^2 , Λ και Γ υπολογίζονται από τα στοιχεία των ήδη γνωστών εκτιμήσεων χρησιμοποιώντας τις μεθόδους Markov chain Monte Carlo. Στην ουσία, χρησιμοποιούν τα χαρακτηριστικά των χρηστών $\{z_u\}$ (ένα τμήμα του προφίλ των χρηστών), τα χαρακτηριστικά των αντικειμένων $\{w_j\}$ (ένα

τμήμα του προφίλ των αντικειμένων) και τις αλληλεπιδράσεις τους $\{x_{ij}\}$ για να υπολογίσουν την εκτίμηση για ένα αντικείμενο.

Τα υβριδικά συστήματα συστάσεων μπορούν επίσης να βελτιωθούν με χρήση τεχνικών βασισμένων στη γνώση όπως case-based reasoning, προκειμένου να αυξηθεί η ακρίβεια των συστάσεων και να ξεπεραστούν οι περιορισμοί των παραδοσιακών συστημάτων συστάσεων π.χ. νέος χρήστης, νέα αντικείμενα. Το knowledge-based σύστημα συστάσεων Entree χρησιμοποιεί γνώση για τα εστιατόρια, τις κουζίνες και τα φαγητά (π.χ. «τα θαλασσινά» δεν είναι «λαχανικά») για να προτείνει εστιατόρια στους χρήστες του. Το κύριο μειονέκτημα των βασισμένων στη γνώση εφαρμογών είναι η ανάγκη για την απόκτηση αυτής της γνώσης, που αποτελεί άλλωστε και περιορισμό πολλών εφαρμογών τεχνητής νοημοσύνης. Εντούτοις, τα βασισμένα στη γνώση συστήματα συστάσεων έχουν αναπτυχθεί για περιοχές εφαρμογής όπου η γνώση των περιοχών είναι εύκολα διαθέσιμη σε κάποια δομημένη και αναγνώσιμη από τους υπολογιστές μορφή π.χ. οντολογία. Τα συστήματα Quickstep και Foxtrot χρησιμοποιούν την οντολογία θέματος ερευνητικών εργασιών για να προτείνουν online ερευνητικά άρθρα. Τέλος, πολλές εργασίες όπως (Balabanovic & Shoham, 1997), (Melville et al., 2002), (Pazzani, 1999), συγκρίνουν εμπειρικά την απόδοση των υβριδικών προσεγγίσεων με τις content-based και collaborative μεθόδους και αποδεικνύουν ότι οι πρώτες μπορούν να οδηγήσουν σε καλύτερες προτάσεις.

2.6 Περιορισμοί και Προβλήματα σε Συστάσεις

Οι τεχνικές content-based και collaborative-based που χρησιμοποιούνται συνήθως για την παραγωγή των συστάσεων αντιμετωπίζουν σημαντικά προβλήματα και περιορισμούς που σχετίζονται με:

- Ελλιπή δεδομένα. Όταν τα δεδομένα που χρησιμοποιεί το σύστημα συστάσεων είναι ελλιπή, τότε είναι αναγκαίο να γίνουν μια σειρά από υποθέσεις πριν την εφαρμογή των τεχνικών εξόρυξης γνώσης. Αυτό έχει σαν αποτέλεσμα ορισμένα από τα πρότυπα (patterns) που παράγονται να μην είναι κατάλληλα ή σωστά.

- **Περιορισμένα δεδομένα.** Το πρόβλημα αυτό σχετίζεται τόσο με τους χρήστες όσο και με τα αντικείμενα. Για να είναι επιτυχημένο ένα σύστημα συστάσεων απαιτείται σημαντικός αριθμός χρηστών να έχουν βαθμολογήσει σημαντικό αριθμό αντικειμένων. Ένα παράδειγμα που φανερώνει το πρόβλημα είναι όταν πολλές ταινίες έχουν βαθμολογηθεί από λίγους χρήστες, τότε αυτές θα προτείνονται σπάνια, ακόμα και αν έχουν λάβει υψηλές βαθμολογίες.
- **Ανακριβή δεδομένα.** Το πρόβλημα αυτό σχετίζεται με τον ίδιο το χρήστη και τα δεδομένα που εισάγει στο σύστημα συστάσεων είτε με ρητό είτε με υπονοούμενο τρόπο. Στην πρώτη περίπτωση έχουμε το παράδειγμα του χρήστη που συμπληρώνει τη φόρμα εγγραφής με ψεύτικα ή ανακριβή στοιχεία που οδηγούν σε αποτυχημένες μελλοντικές συστάσεις. Στη δεύτερη περίπτωση μπορεί το πρόβλημα να δημιουργείται χωρίς την υπαιτιότητα του χρήστη, αν και το αποτέλεσμα θα είναι το ίδιο, δηλαδή «άχρηστα» recommendations. Αυτό συμβαίνει όταν ο χρήστης χαθεί μέσα στο site, οπότε τα click streams που καταγράφονται στο αρχείο log είναι πλασματικά. Αυτό γίνεται περισσότερο προβληματικό όταν το web site είναι άστοχα σχεδιασμένο και οι χρήστες επισκέπτονται αυτόκλητες σελίδες κάνοντας τες να φαίνονται δημοφιλείς.
- **Νέες σελίδες.** Το πρόβλημα προκύπτει όταν νέες σελίδες που παρουσιάζουν νέα αντικείμενα προστίθενται στο web site. Επειδή αυτές τις σελίδες δεν τις έχουν ακόμα επισκεφτεί οι χρήστες και φυσικά δεν έχουν βαθμολογήσει τα προϊόντα τους, το σύστημα συστάσεων δεν τα συνιστά, ακόμα και στην περίπτωση που αυτά σχετίζονται με το προφίλ του χρήστη και πιθανότατα θα τα έβρισκε ενδιαφέροντα. Παράλληλα, όσο περισσότερο μια σελίδα προτείνεται, τόσο περισσότερο την επισκέπτονται οι χρήστες αυξάνοντας την πιθανότητα να είναι υποψήφια για μελλοντικές συστάσεις.

- Νέοι χρήστες. Όταν ένας νέος χρήστης επισκέπτεται το web site, επειδή το ιστορικό του είναι κενό είναι δύσκολο να συσχετισθεί με άλλους χρήστες. Επίσης, σε πολλές περιπτώσεις πρέπει να βαθμολογήσει ένα σημαντικό αριθμό αντικειμένων για να μπορέσει το σύστημα συστάσεων να κατανοήσει τις προτιμήσεις του και να του παρουσιάσει ένα αξιόπιστο σύνολο από συστάσεις.
- «Cold start problem». Το πρόβλημα αυτό έχει ονομαστεί και σαν «day one» λόγω του γεγονότος ότι την πρώτη μέρα της υπηρεσίας του, το σύστημα δεν έχει διαθέσιμα δεδομένα για να ξεκινήσει να παράγει προτάσεις. Ιδιαίτερα οι CF τεχνικές παρουσιάζουν φτωχά αποτελέσματα για τα νέα προϊόντα π.χ. προϊόντα που δεν έχουν βαθμολογηθεί ακόμα από κάποιον πελάτη. Από την άλλη μεριά όσο ο αριθμός των χρηστών που επισκέπτονται το e-shop μεγαλώνει, τόσο πληθαίνουν οι συγκρίσεις που πρέπει να γίνουν μεταξύ όλων των χρηστών, γεγονός που επηρεάζει σοβαρά το scalability του συστήματος συστάσεων.
- «Banana problem». Τα συστήματα συστάσεων που εξάγουν συσχετίσεις από τις συναλλαγές των χρηστών είναι ιδιαίτερα «ευαίσθητα» σε σχέση με τη συχνότητα των βαθμολογιών συγκεκριμένων αντικειμένων. Για παράδειγμα, εφόσον οι μπανάνες αγοράζονται πολύ συχνά από τους πελάτες των super-markets, το σύστημα συστάσεων που χρησιμοποιεί δεδομένα από το καλάθι αγορών τους για να εξάγει συστάσεις θα προτείνει πάντα τις μπανάνες. Ένα αντίστροφο πρόβλημα είναι όταν προτείνεται ένα προϊόν που δεν αγοράζεται συχνά και μόνο ένα κάθε φορά (π.χ. αυτοκίνητο). Επίσης, είναι δυνατόν να αναγνωριστεί κάποιος χρήστης που έχει ασυνήθιστες προτιμήσεις, με αποτέλεσμα να είναι δύσκολο να βρεθούν «παρόμοιοι» πελάτες ώστε να παραχθούν συστάσεις.
- Υποκειμενικότητα των βαθμολογήσεων. Το πρόβλημα αυτό προκύπτει από το γεγονός ότι οι χρήστες βαθμολογούν τα προϊόντα με βά-

- Μονοδιάστατες προτάσεις. Όταν ένα σύστημα συστάσεων προτείνει αντικείμενα που έχουν υψηλή βαθμολογία λαμβάνοντας υπόψη μόνο το προφίλ του χρήστη, τότε ποτέ δεν θα του προτείνει αντικείμενα που είναι τελείως διαφορετικά από αυτά που ο χρήστης έχει ήδη βαθμολογήσει. Αυτό σημαίνει για παράδειγμα ότι ένας χρήστης που δεν έχει εμπειρία στην ελληνική κουζίνα, δε θα λάβει ποτέ recommendation για το καλύτερο ελληνικό εστιατόριο στην πόλη. Σε ειδικές περιπτώσεις το σύστημα δεν θα προτείνει αντικείμενα που είναι πολύ όμοια με αυτά που έχει δει ήδη ο χρήστης π.χ. διαφορετικά άρθρα που όμως περιγράφουν το ίδιο γεγονός. Ιδανικά, ο χρήστης πρέπει να λαμβάνει σειρά από επιλογές και όχι ένα ομοιογενές σύνολο από εναλλακτικές προτάσεις. Για παράδειγμα, δεν είναι απαραίτητα καλή ιδέα να προτείνονται σε ένα χρήστη όλες τις ταινίες του Woody Allen, ενώ αυτός έδειξε προτίμηση για μια μόνο από αυτές.

Τα τελευταία χρόνια η έρευνα στην περιοχή έχει στραφεί στο συνδυασμό τεχνικών και αλγορίθμων από διάφορες επιστημονικές περιοχές (π.χ. Web Usage Mining) και έχει οδηγήσει στα hybrid συστήματα συστάσεων τα οποία ως ένα βαθμό αντιμετωπίζουν τα προβλήματα και τους περιορισμούς των πιο παραδοσιακών προσεγγίσεων. Εντούτοις, αυτά τα συστήματα και οι εφαρμογές, όπως άλλωστε και τα παραδοσιακά, δεν λαμβάνουν υπόψη τους τη σημασιολογική γνώση (semantic knowledge) για την περιοχή που εφαρμόζονται. Χωρίς αυτή τη γνώση, τα συστήματα συστάσεων δεν έχουν τη δυνατότητα να προτείνουν διαφορετικούς τύπους πολύπλοκων αντικειμένων με βάση τα υποκείμενα χαρακτηριστικά και τις ιδιότητές τους. Πράγματι, η ολοκλήρωση σημασιολογικής γνώσης αποτελεί την πρόκληση για την επόμενη γενιά των συστημάτων συστάσεων.

3 Περιγραφή Εργαλείων

Στο παρόν κεφάλαιο περιγράφεται το λογισμικό που χρησιμοποιήθηκε για την ανάπτυξη του ηλεκτρονικού καταστήματος και την υλοποίηση του συστήματος συστάσεων.

3.1 Η γλώσσα PHP

Η PHP ανήκει σε μια κατηγορία γλωσσών που είναι γνωστές σαν middleware. Αυτές οι γλώσσες συνεργάζονται στενά με τον Web server για να ερμηνεύσουν τα αιτήματα που προέρχονται από το World Wide Web, επεξεργάζονται αυτά τα αιτήματα, αλληλεπιδρούν με άλλα προγράμματα στον εξυπηρετητή για να εκπληρώσουν τα αιτήματα, και έπειτα υποδεικνύουν στον Web server τι ακριβώς να αποστείλει στον φυλλομετρητή του πελάτη.

Αυτοί που αναπτύσσουν εφαρμογές για το Web, πραγματοποιούν τον μεγαλύτερο όγκο της εργασίας τους στο middleware. Με λίγη τύχη ο Web server, μπορεί να τεθεί σε λειτουργία χωρίς μεγάλη προσπάθεια, και αφότου τεθεί σε λειτουργία, δεν θα χρειαστεί ιδιαίτερες επεμβάσεις. Αντίθετα καθώς αναπτύσσονται οι εφαρμογές, θα χρειαστεί πολύς χρόνος για να γραφτεί ο κώδικας που κάνει την εφαρμογή να δουλεύει σωστά. Εκτός από την PHP, υπάρχουν διάφορες γλώσσες που επιτελούν παρόμοιες λειτουργίες. Μερικές από τις δημοφιλέστερες επιλογές είναι η ASP, η Perl, και η Cold Fusion.

Η PHP είναι μια γλώσσα συγγραφής σεναρίων στην πλευρά του διακομιστή (server-side scripting language). Μέσα σε μια HTML σελίδα, μπορεί να ενσωματωθεί PHP κώδικας που θα εκτελεσθεί κάθε φορά που η σελίδα δέχεται επίσκεψη. Ο PHP κώδικας ερμηνεύεται στον Web server και παράγεται HTML ή άλλη έξοδος που θα δει ο επισκέπτης. Η ιδέα της PHP συνελήφθη μέσα στο 1994 και ήταν αρχικά η εργασία ενός ατόμου, του Rasmus Lerdorf. Υιοθετήθηκε και από άλλους ταλαντούχους ανθρώπους και έχει περάσει από πολλές σημαντικές ε-

πανεγγραφές για να καταλήξει στο σημερινό ώριμο προϊόν που κυκλοφορεί σήμερα.

Μπορούμε να φανταστούμε την PHP σαν ένα πρόσθετο (plug-in) για τον Web server που του δίνει τη δυνατότητα να κάνει κάτι περισσότερο από το να στέλνει απλά και μόνο απλές ιστοσελίδες (Web pages) όταν τις ζητάνε οι φυλλομετρητές. Με εγκατεστημένη την PHP, ο Web server θα είναι σε θέση να διαβάσει ένα νέο είδος αρχείων, τα αποκαλούμενα PHP scripts, τα οποία μπορούν να πραγματοποιούν εργασίες όπως είναι η ανάκτηση πληροφοριών της τελευταίας στιγμής (retrieve up-to-the-minute information) από μια βάση δεδομένων και καταχώρησή τους σε μια ιστοσελίδα πριν αυτή αποσταλεί στον φυλλομετρητή που τη ζήτησε.

Η PHP είναι ένα προϊόν ανοικτού κώδικα και επομένως οι χρήστες έχουν πρόσβαση στον πηγαίο κώδικα της. Μπορούν να τον χρησιμοποιήσουν, να τον αλλάξουν, και το αναδιανείμουν χωρίς δαπάνη. Η PHP αρχικά σήμαινε Personal Home Page, αλλά αλλάχτηκε και τώρα σημαίνει PHP Hypertext Preprocessor.

Μερικοί από τους κύριους ανταγωνιστές της PHP είναι η Perl, η Microsoft Active Server Pages (ASP), η Java Server Pages (JSP), και η Allaire Cold Fusion. Σε σύγκριση με αυτά τα προϊόντα, η PHP έχει πολλά "δυνατά" σημεία. Ενδεικτικά αναφέρουμε τα ακόλουθα:

- Υψηλή απόδοση: η PHP είναι πολύ αποδοτική. Χρησιμοποιώντας έναν μοναδικό ανέξοδο server, μπορούν να εξυπηρετηθούν εκατομμύρια προσβάσεις την ημέρα. Συγκριτικές μετρήσεις επιδόσεων δημοσιευμένες από την Zend Technologies (<http://www.zend.com>) εμφανίζουν την PHP να ξεπερνάει τους ανταγωνιστές της.
- Διεπαφές σε πολλά διαφορετικά συστήματα βάσεων δεδομένων: η PHP έχει διαθέσιμες εγγενείς συνδέσεις σε πολλά συστήματα βάσεων δεδομένων. Εκτός από την MySQL, μπορεί άμεσα να συνδεθεί στην PostgreSQL, στην mSQL, στον Oracle, καθώς και στα dbm, filePro, Hyperwave, Informix, InterBase, και Sybase, μεταξύ των άλλων. Χρησιμοποιώντας το Open Database Connectivity Standard (ODBC), μπορεί

- Ενσωματωμένες βιβλιοθήκες για πολλές συνήθεις εργασίες Web: Επειδή η PHP σχεδιάστηκε για χρήση στο Web, διαθέτει πολλές ενσωματωμένες λειτουργίες για την εκτέλεση πολλών χρήσιμων εργασιών που είναι σχετικές με το Web. Μπορούν να παραχθούν GIF εικόνες on-the-fly, να γίνει σύνδεση με άλλες υπηρεσίες δικτύου, να σταλεί ηλεκτρονικό ταχυδρομείο, να γίνει εργασία με cookies, και να παραχθούν PDF έγγραφα, όλα με μερικές γραμμές κώδικα.
- Χαμηλό κόστος: Η PHP είναι δωρεάν. Η πιο πρόσφατη έκδοση είναι διαθέσιμη για download οποιαδήποτε στιγμή από το <http://www.php.net> χωρίς αντίτιμο.
- Ευκολία εκμάθησης και χρήσης: Η σύνταξη της PHP είναι βασισμένη σε άλλες γλώσσες προγραμματισμού, πρώτιστα στην C και στην Perl. Εάν ξέρει κάποιος ήδη C ή Perl, ή μία γλώσσα που μοιάζει με την C όπως την C++ ή την Java, θα είναι παραγωγική η χρησιμοποίηση της PHP σχεδόν αμέσως.
- Μεταφερσιμότητα: η PHP είναι διαθέσιμη για πολλά διαφορετικά λειτουργικά συστήματα. Μπορεί κανείς να γράψει PHP προγράμματα στα δωρεάν παρόμοια με Unix λειτουργικά συστήματα όπως το Linux και το FreeBSD, στις εμπορικές Unix εκδόσεις όπως το Solaris και το IRIX, ή στις διαφορετικές εκδόσεις των Microsoft Windows. Ο κώδικάς σας θα λειτουργήσει συνήθως χωρίς τροποποίηση σε ένα διαφορετικό σύστημα που τρέχει PHP.
- Διαθεσιμότητα του πηγαίου κώδικα: Έχει κανείς πρόσβαση στον πηγαίο κώδικα της PHP. Αντίθετα από τα εμπορικά, προϊόντα με κλειστό κώδικα, εάν υπάρχει κάτι που θέλει κανείς να τροποποιήσει ή να προσθέσει στη γλώσσα, είναι ελεύθερος να το κάνει. Δεν χρειάζεται να περιμένει πότε ο κατασκευαστής θα δημιουργήσει patches. Δεν χρειάζεται να ανησυχεί για το εάν ο κατασκευαστής θα χρεοκοπήσει ή θα αποφασίσει να σταματήσει να υποστηρίζει ένα προϊόν.

3.2 Βάση Δεδομένων MySQL

Τα σχεσιακά συστήματα διαχείρισης βάσεων δεδομένων (ΣΔΒΔ - RDBMS) παρέχουν έναν άριστο τρόπο για να αποθηκεύσουν και να παρέχουν πρόσβαση σε σύνθετες πληροφορίες. Προϋπήρχαν από το Web, το Linux, και τα Windows, και έτσι υπάρχουν πολλά ΣΔΒΔ για να επιλέξει κανείς. Όλες οι σημαντικές βάσεις δεδομένων χρησιμοποιούν την Structured Query Language (SQL).

Μερικά από τα δημοφιλέστερα από τα εμπορικά ΣΔΒΔ είναι τα Oracle, Sybase, Informix, το SQL Server της Microsoft, και το db2 της IBM. Εκτός από την MySQL, τώρα υπάρχουν δύο σημαντικές σχεσιακές βάσεις δεδομένων ανοικτού πηγαίου κώδικα. Η Postgres είναι η σημαντικότερη εναλλακτική λύση έναντι της MySQL και η Interbase της Borland διαθέτει άδεια ανοικτού κώδικα και παρέχεται για δωρεάν χρήση.

Η MySQL είναι ένα πολύ γρήγορο, εύρωστο, σχεσιακό σύστημα διαχείρισης βάσεων δεδομένων (RDBMS). Μια βάση δεδομένων επιτρέπει την αποτελεσματική αποθήκευση, αναζήτηση, ταξινόμηση, και ανάκτηση των δεδομένων. Ο MySQL server ελέγχει την πρόσβαση στα δεδομένα για να διασφαλίσει ότι πολλοί χρήστες μπορούν να συνεργαστούν με αυτόν ταυτόχρονα, παρέχει γρήγορη πρόσβαση σε αυτά, και διασφαλίζει ότι μόνο εξουσιοδοτημένοι χρήστες μπορούν να έχουν πρόσβαση. Σε γενικές γραμμές είναι ένα πακέτο λογισμικού που είναι πολύ καλό στην οργάνωση και τη διαχείριση μεγάλων ποσοτήτων πληροφοριών. Ως εκ τούτου, ο MySQL είναι ένας server πολλών χρηστών (multi-user), και πολλαπλών διεργασιών (multi-threaded) με υλοποίηση client/server.

Χρησιμοποιεί την SQL (Structured Query Language), την τυποποιημένη γλώσσα παγκοσμίως για την διατύπωση ερωτημάτων σε βάσεις δεδομένων. Η MySQL είναι δημόσια διαθέσιμη από το 1996, αλλά έχει ένα ιστορικό ανάπτυξης που ξεκινά από το 1979. Η MySQL είναι διαθέσιμη με άδεια ανοικτού λογισμικού, αλλά και εμπορικές άδειες είναι επίσης διαθέσιμες εάν είναι απαραίτητο. Η ταχύτητά της, η ευρωστία της, και ευκολία της χρήσης της την καθιστούν μια

δημοφιλή βάση δεδομένων για ιστοχώρους με δυναμικό περιεχόμενο. Η PHP, η C, η C++, η Java, και η Perl έχουν άμεση πρόσβαση στην MySQL.

Με τη χρήση της MySQL είναι εύκολη η πρόσβαση στις πληροφορίες χρησιμοποιώντας μια γλώσσα συγγραφής σεναρίων στην πλευρά του διακομιστή (server-side scripting language), όπως είναι η PHP. Έτσι ο κώδικας που αναπτύσσεται θα μπορεί να εφαρμοσθεί σ' έναν διακομιστή που βασίζεται στα Windows ή στο Unix.

Μερικοί από τους κύριους ανταγωνιστές της MySQL είναι η PostgreSQL, ο Microsoft SQL Server, και ο Oracle. Η MySQL έχει πολλά δυνατά σημεία, συμπεριλαμβανομένης της υψηλής απόδοσης, του χαμηλού κόστους, της ευκολίας στην διαμόρφωση και στην εκμάθηση, της μεταφερισιμότητας, και της διαθεσιμότητας του πηγαίου κώδικα. Πιο αναλυτικά:

- Απόδοση: Η MySQL είναι αναντίρρητα γρήγορη. Μπορεί κανείς να δει τις συγκριτικές μετρήσεις επιδόσεων στην σελίδα <http://web.mysql.com/benchmark.html>. Πολλές από αυτές τις συγκριτικές μετρήσεις επιδόσεων παρουσιάζουν την MySQL να είναι πολλές τάξεις μεγέθους γρηγορότερη σε σχέση με τον ανταγωνισμό.
- Χαμηλό κόστος: Η MySQL είναι διαθέσιμη με μηδενικό κόστος, με άδεια ανοικτού λογισμικού, ή με χαμηλό κόστος με εμπορική άδεια αν είναι απαραίτητο για την εφαρμογή.
- Ευκολία χρήσης: Οι περισσότερες σύγχρονες βάσεις δεδομένων χρησιμοποιούν SQL. Εάν κάποιος έχει χρησιμοποιήσει άλλο RDBMS, δεν πρέπει να έχει κανένα πρόβλημα προσαρμογής. Η MySQL εγκαθίσταται επίσης ευκολότερα από πολλά παρόμοια προϊόντα.
- Μεταφερισιμότητα: Η MySQL μπορεί να χρησιμοποιηθεί σε πολλά διαφορετικά συστήματα UNIX όπως και κάτω από τα Microsoft Windows.
- Πηγαίος κώδικας: Όπως με την PHP, μπορεί κανείς να πάρει και να τροποποιήσει τον πηγαίο κώδικα για την MySQL.

3.3 Ο Διακομιστής Ιστού Apache

Εξυπηρετητής ή διακομιστής (web server) είναι υλικό ή/και λογισμικό που αναλαμβάνει την παροχή διάφορων υπηρεσιών, «εξυπηρετώντας» πελάτες (clients). Όταν ένας υπολογιστής εκτελεί κυρίως προγράμματα εξυπηρετητές συνεχόμενα, 24ώρες την ημέρα τότε μπορούμε να αναφερθούμε σε όλον τον υπολογιστή σαν διακομιστής αφού αυτή είναι η κύρια λειτουργία του. Σαν πελάτη μπορούμε να θεωρήσουμε είτε λογισμικό που μπορεί να μιλήσει και να υποβάλει αιτήματα στον εξυπηρετητή είτε τον υπολογιστή στον οποίο εκτελείται το λογισμικό πελάτη.

Η επικοινωνία πελάτη-εξυπηρετητή γίνεται μέσω ενός τοπικού δικτύου η ακόμα και στο Παγκόσμιο Ιστό. Σαν συσκευή είναι συνήθως ένας υπολογιστής που διαφέρει ως προς την σύνθεση του από άλλους κοινούς υπολογιστές μιας και οι δυνατότητες του είναι σαφώς αναβαθμισμένες. Κύρια χαρακτηριστικά ενός εξυπηρετητή είναι οι επεξεργαστές που υποστηρίζει και που χρησιμοποιεί για την επεξεργασία των πολλαπλών δεδομένων που δέχεται, οι γρήγοροι και μεγάλης χωρητικότητας σκληροί δίσκοι αλλά και οι ταχύτητες μνήμης που υποστηρίζει. Συνήθως συνοδεύεται με σύστημα διπλής τροφοδοσίας (dual power supply) και από συσκευή παροχής αδιάλειπτης ενέργειας (UPS) για μεγαλύτερη αξιοπιστία και σιγουριά στις παρεχόμενες υπηρεσίες του.

Ο Apache είναι ο πιο επιτυχημένος Web εξυπηρετητής εφόσον χρησιμοποιείται στην μεγάλη πλειοψηφία των ιστοχώρων. Ανήκει στον χώρο του λογισμικού ανοικτού κώδικα και η οργάνωση του project αποτελείται από μια βασική ομάδα υπεύθυνων για την ανάπτυξη οι οποίοι αναλαμβάνουν εκ περιτροπής την διεύθυνση του project. Αυτοί είναι υπεύθυνοι για την ανάπτυξη τμημάτων του λογισμικού έχοντας συμφωνήσει σε έναν αριθμό από διάφορα βασικά πρότυπα που θα χρησιμοποιούνται.

Κάθε βασικό μέλος έχει επίσης υπό την επίβλεψη του μια βασική ομάδα προγραμματιστών του που εργάζεται για ένα module ολόκληρου του συστήματος. Τελικά υπάρχει ένας μεγάλος αριθμός άλλων ανθρώπων που παρέχουν

βοήθεια στην προσπάθεια με διαφορετικούς βαθμούς συμμετοχής (προγραμματισμό, τεκμηρίωση, αποσφαλμάτωση, κλπ.).

Η πρώτη του έκδοση, γνωστή ως NCSA HTTPd, δημιουργήθηκε από τον Robert McCool και κυκλοφόρησε το 1993. Θεωρείται ότι έπαιξε σημαντικό ρόλο στην αρχική επέκταση του ιστού. Ήταν η πρώτη βιώσιμη εναλλακτική επιλογή που παρουσιάστηκε απέναντι στον εξυπηρετητή http της εταιρείας Netscape και από τότε έχει εξελιχθεί στο σημείο να ανταγωνίζεται άλλους εξυπηρετητές βασισμένους στο Unix σε λειτουργικότητα και απόδοση. Από τον Απρίλη του 1996 ο Apache είναι ο πιο δημοφιλής http εξυπηρετητής σε χρήση. Τον Σεπτέμβριο του 2009 ο Apache εξυπηρετούσε το 54.48% όλων των ιστοσελίδων και το 66% από το ένα εκατομμύριο δημοφιλέστερες.

3.4 Ο Uniform Server

Με δεδομένη την ανάγκη δημιουργίας μια μεταφέρσιμης (portable) εφαρμογής που να μπορεί να λειτουργήσει ακόμη και από ένα CD ώστε να μπορεί εύκολα να επιδεικνύεται, επιλέχθηκε η χρήση του Uniform Server.

Ο Uniform εξυπηρετητής είναι ένα πακέτο WAMP που μας επιτρέπει να τρέξουμε έναν εξυπηρετητή σε κάθε υπολογιστή που διαθέτει MS Windows λειτουργικό. Το ακρώνυμο WAMP αποτελείται από τα αρχικά του λειτουργικού Windows και τα αρχικά των κυριότερων μερών του πακέτου: Apache, MySQL και PHP (ή Perl ή Python). Είναι μικρό και φορητό για κατέβασμα ή μεταφορά και μπορεί επίσης να χρησιμοποιηθεί ή να εγκατασταθεί σαν ένας production/live εξυπηρετητής. Το πακέτο περιλαμβάνει το πρόγραμμα phpMyAdmin το οποίο παρέχει ένα γραφικό περιβάλλον για τον διαχειριστή της MySQL βάσης δεδομένων, δίνοντάς του την δυνατότητα να δημιουργήσει, να τροποποιήσει ή να διαγράψει βάσεις δεδομένων, πίνακες, παιδιά ή εγγραφές, χρησιμοποιώντας SQL εντολές ή διοικώντας χρήστες και άδειες.

4 Σχεδιασμός του e-Shop

Σε αυτό το κεφάλαιο εξηγείται και παρουσιάζεται το σχήμα της Βάσης Δεδομένων και η δομή του ηλεκτρονικού καταστήματος.

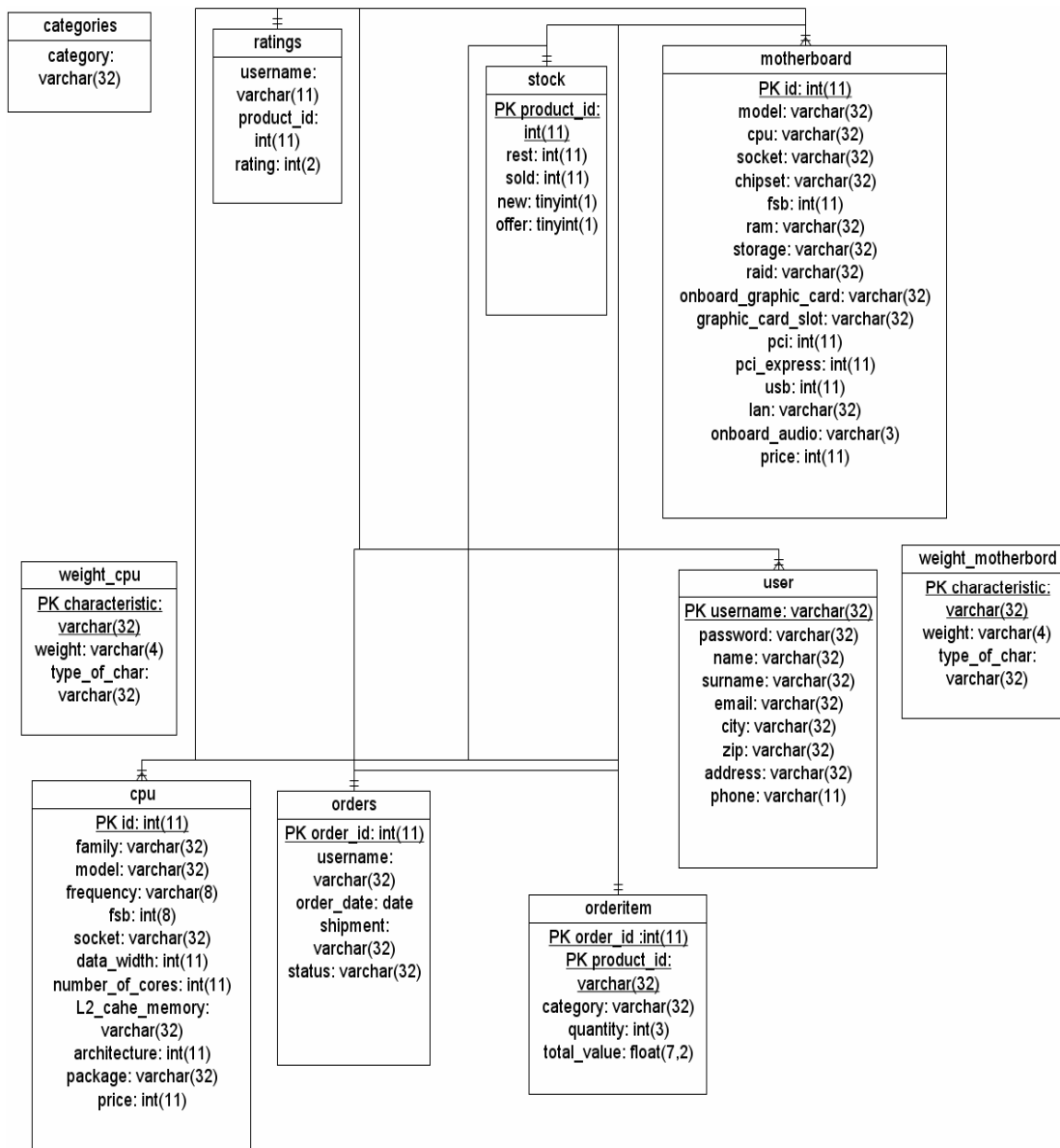
4.1 Βάση Δεδομένων

Η βάση δεδομένων υλοποιήθηκε με την χρήση του εργαλείου phpMyAdmin και χρησιμοποιήθηκε το σχεσιακό σύστημα διαχείρισης βάσεων δεδομένων MySQL όπως αναφέρεται στο κεφάλαιο 3. Παρακάτω παρουσιάζεται η οργάνωση της βάσης δεδομένων του ηλεκτρονικού καταστήματος.

4.2 Οργάνωση Σχήματος Βάσης Δεδομένων

Η βάση δεδομένων που υλοποιεί το ηλεκτρονικό κατάστημα αποτελείται από 8 πίνακες (εικόνα 7). Οι πίνακες αυτοί είναι:

- user: Κρατάει τα στοιχεία κάθε χρήστη.
- categories: Αποθηκεύει τα είδη των προϊόντων.
- motherboard: Το προϊόν motherboard.
- cpu: Το προϊόν cpu.
- ratings: Οι εγγεγραμμένοι χρήστες βαθμολογούν τα προϊόντα.
- orders: Καταχωρούνται οι παραγγελίες που έχει πραγματοποιήσει ο κάθε εγγεγραμμένος χρήστης.
- orderitem: Περιέχει τα προϊόντα ανά παραγγελία.
- stock: Υλοποιεί την αποθήκη του ηλεκτρονικού καταστήματος.



Εικόνα 7: Σχεσιακό διάγραμμα βάσης δεδομένων του ηλεκτρονικού καταστήματος

Στον πίνακα user καταχωρούνται όλες οι πληροφορίες που αναφέρονται στα στοιχεία των εγγεγραμμένων πελατών του ηλεκτρονικού καταστήματος. Η ολοκλήρωση της παραγγελίας προϋποθέτει την εγγραφή του πελάτη. Με την χρήση του username και του password ο χρήστης μπορεί να κάνει Login στο ηλεκτρονικό κατάστημα. Το πεδίο username αποτελεί το κύριο κλειδί και είναι τύπου varchar και έχει μήκος 32 χαρακτήρες. Το πεδίο password αναφέραμε ότι απαιτείται για την είσοδο του χρήστη και είναι τύπου varchar με μέγιστο μήκος 32 χαρακτήρες. Ακόμη τα πεδία name, surname, email, city, zip, address, phone είναι απαραίτητα για την επικοινωνία και την σωστή αποστολή της παραγγελίας στον πελάτη.

Ο πίνακας `categories` έχει ένα πεδίο, το `category`, που αναφέρεται στο είδος του προϊόντος και είναι βοηθητικός για την σύνδεση του κώδικα PHP της εφαρμογής με τα MySQL ερωτήματα. Το πεδίο `category` είναι τύπου `varchar` με μήκος 32 χαρακτήρες. Το πεδίο δανείζεται στον πίνακα `orderitem` που αναφέρεται παρακάτω.

Ο πίνακας `motherboard` έχει κύριο κλειδί το πεδίο `id` τύπου `int` με μήκος 11 χαρακτήρες. Το πεδίο `id` αυξάνεται κατά 1 όταν εισάγεται νέο προϊόν, με την επιλογή `auto_increment` από το `menu` του εργαλείου `phpMyAdmin`. Ο πίνακας έχει όλα τα απαραίτητα πεδία για την περιγραφή του προϊόντος, όπως το πεδίο `model` που περιέχει το μοντέλο της μητρικής, το πεδίο `cpu` που αναφέρει με ποιους επεξεργαστές είναι συμβατή η μητρική, το πεδίο `price` που αναφέρει την τιμή του προϊόντος. Τα υπόλοιπα πεδία με τα τεχνικά χαρακτηριστικά είναι `socket`, `fsb`, `ram`, `storage`, `raid`, `onboard_graphic_card`, `graphic_card_slot`, `pci`, `pci_express`, `usb`, `lan`, `onboard_audio`.

Ο πίνακας `cpu` είναι αντίστοιχος με τον πίνακα `motherboard`, έχοντας κύριο κλειδί το πεδίο `id` τύπου `int` με μήκος 11 χαρακτήρες και το οποίο αυξάνει αυτόματα με την εισαγωγή νέου προϊόντος. Βασικά πεδία είναι: το `model` που αναφέρει το μοντέλο του επεξεργαστή, το `frequency` που αναφέρει την συχνότητα λειτουργίας του επεξεργαστή και το πεδίο `price` με την τιμή του. Τα υπόλοιπα πεδία με τα τεχνικά χαρακτηριστικά είναι `family`, `fsb`, `socket`, `data_width`, `number_of_cores`, `L2_cache_memory`, `architecture`, `package`.

Στον πίνακα `ratings` οι χρήστες που είναι εγγεγραμμένοι βαθμολογούν τα προϊόντα. Ο πίνακας έχει το πεδίο `username` τύπου `varchar` με μήκος 11 χαρακτήρες, το πεδίο `product_id` το οποίο αναφέρεται στο `id` των προϊόντων και κρατάει το προϊόν που βαθμολόγησε ο χρήστης, είναι τύπου `int` με μήκος 11 χαρακτήρες. Το τελευταίο πεδίο είναι το `rating` όπου αποθηκεύεται η βαθμολογία του χρήστη, ο οποίος δίνει τιμές από 1 έως 10, και είναι τύπου `int` με μήκος 2 χαρακτήρες.

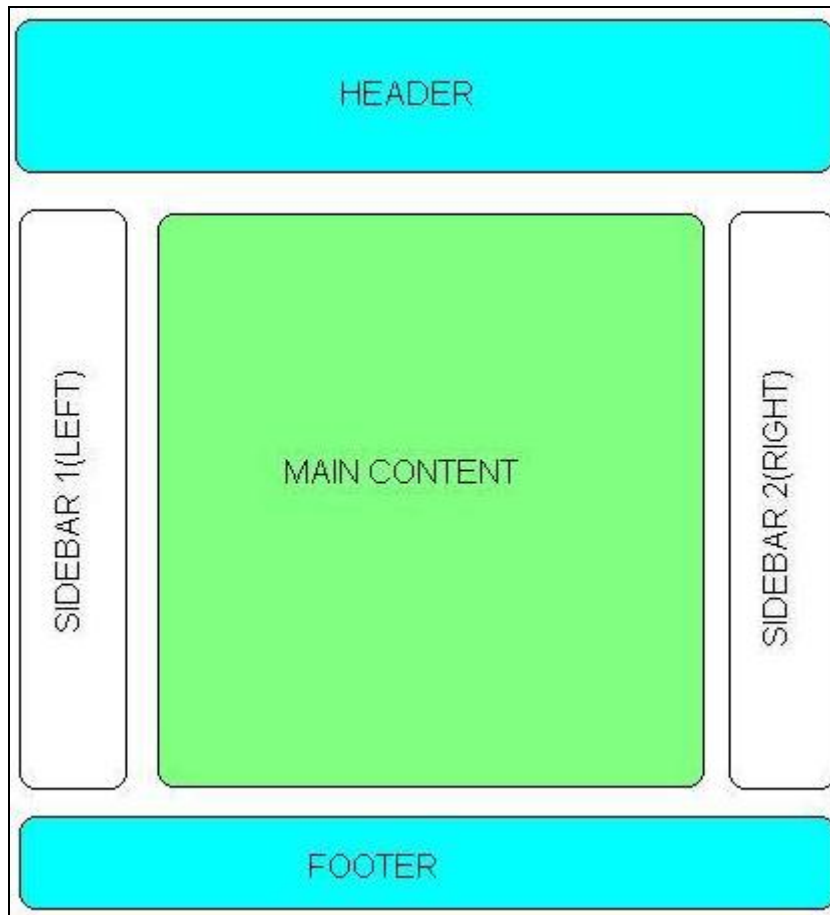
Ο πίνακας `orders` έχει σαν κύριο κλειδί το πεδίο `order_id` τύπου `int` με μήκος 11 χαρακτήρες, το οποίο αυξάνεται αυτόματα κατά 1 με την πραγματοποίηση νέας παραγγελίας. Ο πίνακας έχει το πεδίο `username` για να ελέγχει ποιός έκανε την παραγγελία, το πεδίο `order_date` τύπου `date` που συγκρατεί πότε έγινε η παραγγελία, το πεδίο `shipment` που αναφέρεται στον τρόπο αποστολής, και το πεδίο `status` που περιέχει την κατάσταση της προόδου της παραγγελίας.

Ο πίνακας `orderitem` έχει σαν κύρια κλειδιά το `order_id` που είναι ίδιο με το `order_id` του πίνακα `orders` και το `product_id` που περιέχει το συγκεκριμένο προϊόν που έχει παραγγελθεί και είναι παρμένο από τους πίνακες των προϊόντων (`cpu`, `motherboard`). Ο πίνακας έχει το πεδίο `category` δανεισμένο από τον πίνακα `categories`, έχει το πεδίο `quantity` τύπου `int` με μήκος 3 χαρακτήρες, το οποίο κρατάει την ποσότητα του προϊόντος που παραγγέλθηκε. Ακόμη, το πεδίο `total_value` τύπου `float` που κρατάει την αξία της ποσότητας του προϊόντος.

Ο πίνακας `stock` που είναι η αποθήκη του ηλεκτρονικού καταστήματος. Έχει κύριο κλειδί το `product_id` που είναι δανεισμένο από τους πίνακες των προϊόντων. Το πεδίο `rest` τύπου `int` με μήκος 11 χαρακτήρες που ενημερώνεται για το πόσα τεμάχια του προϊόντος είναι διαθέσιμα στην αποθήκη. Το πεδίο `sold` τύπου `int` με μήκος 11 χαρακτήρες που κρατάει πόσα τεμάχια πουλήθηκαν. Ακόμη τα πεδία `new` και `offer` που δείχνουν να το προϊόν είναι καινούργιο ή προσφορά αντίστοιχα.

4.3 Δομή του Ηλεκτρονικού Καταστήματος

Το σύστημα έχει δομηθεί με τέτοιο τρόπο, ώστε να είναι πολύ εύκολο να γίνει οποιαδήποτε αλλαγή χρειαστεί. Αυτό πετυχαίνεται με την χρήση `template` το οποίο αποτελείται από πέντε πλαίσια, όπως φαίνεται και παρακάτω (Εικόνα 8).



Εικόνα 8: Η δομή του ηλεκτρονικού καταστήματος

Ουσιαστικά το μόνο πλαίσιο που αλλάζει είναι το πλαίσιο περιεχομένου main-Content. Το πλαίσιο header περιέχει την γραμμή περιήγησης του ηλεκτρονικού καταστήματος top_menu.php(εικόνα 9).



Εικόνα 9: Η γραμμή περιήγησης

Η αριστερή στήλη sidebar1 είναι υπεύθυνη για την εμφάνιση λίστας με τα είδη προϊόντων(products.php) και για την είσοδο του χρήστη στο σύστημα(login.php). Η δεξιά στήλη sidebar2 περιέχει το καλάθι αγορών(my_basket.php) και το κάτω πλαίσιο footer αναγράφει τον κατασκευαστή της εφαρμογής.

5 Λειτουργίες του e-shop

Παρουσιάζονται αναλυτικά οι σημαντικότερες λειτουργίες του ηλεκτρονικού καταστήματος (εγγραφή πελάτη, σύνδεση πελάτη, φόρμα επικοινωνίας, βαθμολόγηση προϊόντος, αγορά προϊόντος) και οι σελίδες του ηλεκτρονικού καταστήματος (αρχική σελίδα, λίστα προϊόντων, σελίδα περιγραφής του προϊόντος). Περιγράφεται η δημιουργία του συστήματος συστάσεων από τον ειδικό που εισάγει την εμπειρία και τις γνώσεις του. Ακολουθεί η επεξήγηση του τρόπου υπολογισμού της ομοιότητας ανάμεσα στα προϊόντα με την χρήση της ευκλείδειας απόστασης και τέλος η απεικόνιση των χαρακτηριστικών των προϊόντων με τη χρήση του εργαλείου “Google Chart Tools”.

5.1 Εγγραφή Πελάτη

Για να εγγραφεί κάποιος στο ηλεκτρονικό κατάστημα, υπάρχει στην αριστερή στήλη ο υπερσύνδεσμος «Sign up». Όταν κάνει κλικ καλείται η `sign_up.php` και εμφανίζεται μια φόρμα όπου πρέπει να δηλωθούν τα απαιτούμενα στοιχεία για την εγγραφή του πελάτη στο σύστημα(εικόνα 10).

Home	Products	Contact																												
Fill in the sign-up form to register. The fields with (*) are required.																														
PRODUCTS cpu motherboard Login Form Username: <input type="text"/> Password: <input type="password"/> <input type="button" value="Login"/> Sign up	<table border="1"> <tr><td>* Username:</td><td><input type="text"/></td></tr> <tr><td>* Password:</td><td><input type="password"/></td></tr> <tr><td>* Re-enter password:</td><td><input type="password"/></td></tr> <tr><td>* Name:</td><td><input type="text"/></td></tr> <tr><td>* Surname:</td><td><input type="text"/></td></tr> <tr><td>* E-mail address:</td><td><input type="text"/></td></tr> <tr><td>Address:</td><td><input type="text"/></td></tr> <tr><td>Zip code:</td><td><input type="text"/></td></tr> <tr><td>City:</td><td><input type="text"/></td></tr> <tr><td>Phone number:</td><td><input type="text"/></td></tr> <tr><td colspan="2"><input type="button" value="Confirm"/></td></tr> </table>	* Username:	<input type="text"/>	* Password:	<input type="password"/>	* Re-enter password:	<input type="password"/>	* Name:	<input type="text"/>	* Surname:	<input type="text"/>	* E-mail address:	<input type="text"/>	Address:	<input type="text"/>	Zip code:	<input type="text"/>	City:	<input type="text"/>	Phone number:	<input type="text"/>	<input type="button" value="Confirm"/>		My Basket <table border="1"> <thead> <tr> <th>Item</th> <th>Price</th> <th>Quantity</th> </tr> </thead> <tbody> <tr> <td colspan="3">Empty Basket</td> </tr> </tbody> </table>	Item	Price	Quantity	Empty Basket		
* Username:	<input type="text"/>																													
* Password:	<input type="password"/>																													
* Re-enter password:	<input type="password"/>																													
* Name:	<input type="text"/>																													
* Surname:	<input type="text"/>																													
* E-mail address:	<input type="text"/>																													
Address:	<input type="text"/>																													
Zip code:	<input type="text"/>																													
City:	<input type="text"/>																													
Phone number:	<input type="text"/>																													
<input type="button" value="Confirm"/>																														
Item	Price	Quantity																												
Empty Basket																														

Εικόνα 10: Η φόρμα εγγραφής

Τέτοια στοιχεία είναι το όνομα χρήστη, ο κωδικός, το ονοματεπώνυμο, η ηλεκτρονική διεύθυνση, η διεύθυνση και το τηλέφωνο του πελάτη. Εκτός από τα δυο τελευταία, τα άλλα είναι απαραίτητα να συμπληρωθούν. Το όνομα χρήστη(username) χρησιμοποιείται σαν κύριο κλειδί στη βάση δεδομένων, οπότε αν υπάρχει ήδη ζητείται από τον πελάτη να εισάγει κάποιο άλλο. Κάνοντας κλικ στο «Confirm» καλείται η checksignup.php όπου γίνονται οι απαραίτητοι έλεγχοι. Οι βασικοί είναι: ο έλεγχος για το username όπως αναφέραμε, ο κωδικός να αποτελείται από τουλάχιστον τρεις χαρακτήρες και να έχουν εισαχθεί όλα τα απαραίτητα στοιχεία. Εφόσον όλα είναι σωστά εμφανίζεται μήνυμα καλωσορίσματος και μετά από δυο δευτερόλεπτα ο χρήστης ανακατευθύνεται στην τελευταία σελίδα που έβλεπε.

5.2 Είσοδος Πελάτη

Για να συνδεθεί κάποιος στο ηλεκτρονικό κατάστημα, πρέπει να έχει κάνει εγγραφή. Αφού γίνει αυτό, μπορεί να συνδεθεί εισάγοντας το όνομα χρήστη και τον κωδικό στην φόρμα που βρίσκεται στην αριστερή στήλη του ηλεκτρονικού καταστήματος(εικόνα 11).

The image shows a web page with a header 'PRODUCTS' and a sidebar menu containing 'cpu' and 'motherboard'. The main content area features a 'Login Form' with the following elements: a 'Username:' label followed by a text input field, a 'Password:' label followed by a text input field, a 'Login' button, and a 'Sign up' link.

Εικόνα 11: Η φόρμα εισόδου(Login)

Κάνοντας κλικ στο «Login» καλείται η `checklogin.php` όπου ελέγχεται η ορθότητα του ονόματος χρήστη και του κωδικού. Σε περίπτωση λάθους καλείται η `login_failure.php` που του ζητάει να προσπαθήσει πάλι ή να κάνει εγγραφή. Εφόσον τα στοιχεία είναι σωστά καλείται η `login_success.php` όπου επαναφέρει τον χρήστη στην σελίδα που βρισκόταν και δημιουργεί ένα `session` ονόματος `logged` με τιμή `true`. Το `session` είναι απαραίτητο για τον έλεγχο αν ο πελάτης παραμένει συνδεδεμένος στο ηλεκτρονικό κατάστημα. Με την επιτυχημένη σύνδεση καλωσορίζεται ο πελάτης και έχει την δυνατότητα να αποσυνδεθεί επιλέγοντας τον υπερσύνδεσμο «Log out»(εικόνα 12) ο οποίος καταστρέφει τα `session username`, `password` και `logged` με αποτέλεσμα την αποσύνδεση. Ο πελάτης παραμένει στην σελίδα που ήταν αλλά αποσυνδεδεμένος.

The image shows the same web page as in Figure 11, but the user is now logged in. The sidebar menu still contains 'cpu' and 'motherboard'. The main content area now displays 'Welcome kostas' and a 'Log out' link.

Εικόνα 12: Ο πελάτης είναι συνδεδεμένος

5.3 Φόρμα Επικοινωνίας

Για την καλύτερη επικοινωνία του πελάτη με την επιχείρηση, υπάρχει μια φόρμα επικοινωνίας όπου μπορεί να γράψει κάποιο μήνυμα(εικόνα 13) .Ο υπερσύνδεσμος «Contact» βρίσκεται στην επάνω λεζάντα του ηλεκτρονικού καταστήματος.



The image shows a web form titled "Contact Form". It contains four input fields: "Subject" (a single-line text box), "Detail" (a large multi-line text area), "Name" (a single-line text box), and "Email" (a single-line text box). Below the fields is a "Submit" button.

Εικόνα 13: Φόρμα επικοινωνίας

Ο πελάτης πρέπει να γράψει το θέμα, το μήνυμα, το όνομα και το email του. Επιλέγοντας το «Submit» αποστέλλεται email με τα παραπάνω στην επιχείρηση.

```
272 <?php
273 // Contact subject
274 $subject =$_POST["subject"];
275 // Details
276 $message=$_POST["detail"];
277 $name=$_POST["name"];
278 // Mail of sender
279 $mail_from=$_POST["customer_mail"];
280 // From
281 $header='from: '.$name.' <'.$mail_from.'>'.phpversion();
282
283 // Enter your email address
284 $to ='konstogi@stogian.x10hosting.com';
285
286 $send_contact=mail($to,$subject,$message,$header);
```

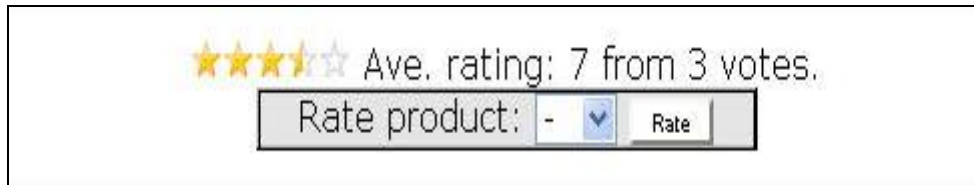
Εικόνα 14: Ο κώδικας για την αποστολή email μέσω της φόρμας επικοινωνίας

Η γλώσσα PHP έχει την λειτουργία mail() η οποία μας επιτρέπει εφόσον υπάρχει mail server να υλοποιήσουμε την παραπάνω λειτουργία. Ο κώδικας παρουσιάζεται στην παρακάτω εικόνα(εικόνα 14).

Εφόσον σταλεί σωστά το email εμφανίζεται μήνυμα ότι η επιχείρηση έλαβε τις πληροφορίες επικοινωνίας του πελάτη.

5.4 Βαθμολόγηση Προϊόντος

Ο πελάτης που έχει εγγραφεί και είναι συνδεδεμένος μπορεί να βαθμολογήσει τα προϊόντα του ηλεκτρονικού καταστήματος. Στη σελίδα όπου εμφανίζονται τα χαρακτηριστικά ενός προϊόντος, περιλαμβάνεται και η φόρμα βαθμολόγησης προϊόντος rate.php (εικόνα 15).



Εικόνα 15: Η φόρμα βαθμολόγησης προϊόντος

Ο πελάτης επιλέγει μια τιμή από 1 έως 10 και πατώντας το «rate» βαθμολογεί το προϊόν. Με το κλικ στο rate καλείται η inser_rating.php η οποία καταχωρεί στη βάση δεδομένων την τιμή. Κάθε πελάτης δίνει μία βαθμολογία στο προϊόν και η τιμή της μπορεί να αλλάξει. Στη βάση δεδομένων υπάρχει ο πίνακας ratings ο οποίος κρατάει την βαθμολογία ενός πελάτη, τον πελάτη και τον κωδικό του προϊόντος. Έτσι για να βρούμε το μέσο όρο της βαθμολογίας διαιρούμε το άθροισμα των βαθμολογιών με το πλήθος των πελατών που βαθμολόγησαν το προϊόν.

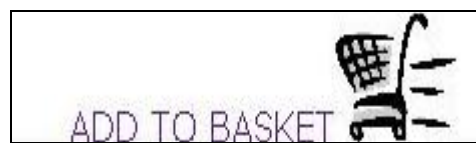
5.5 Αγορά Προϊόντος

Η αγορά ενός ή περισσότερων προϊόντων είναι διαδικασία εύκολη και ολοκληρώνεται σε λίγα μόλις βήματα. Ο χρήστης επιλέγει το προϊόν που θέλει να αγοράσει και το προσθέτει στο καλάθι αγορών. Υπάρχει η δυνατότητα να προστεθούν και άλλα προϊόντα στο καλάθι αγορών, οπότε επαναλαμβάνεται η ίδια διαδικασία τόσες φορές όσα είναι τα προϊόντα που θέλει να προσθέσει. Ο πελάτης μπορεί να επιλέξει πόσα τεμάχια θέλει από κάθε προϊόν. Όταν συμπληρω-

θεί το καλάθι με τα επιθυμητά προϊόντα, επιλέγει να συνεχίσει προς την ολοκλήρωση της παραγγελίας όπου εμφανίζονται τα στοιχεία του τα οποία μπορεί να αλλάξει αν θέλει. Με την επιλογή της ολοκλήρωσης της παραγγελίας, καταγράφεται η παραγγελία στη βάση δεδομένων και του στέλνεται σε email.

Ας περιγράψουμε αναλυτικά την λειτουργία της παραγγελίας επειδή είναι η πιο βασική και σημαντική λειτουργία του ηλεκτρονικού καταστήματος.

Ο πελάτης βρίσκεται στη σελίδα του προϊόντος και πατάει την εικόνα με το καλάθι



Εικόνα 16: Η εικόνα για προσθήκη στο καλάθι αγορών

αγορών(εικόνα 16). Τότε καλείται η `check_basket.php` η οποία εισάγει το προϊόν στο καλάθι αγορών(εικόνα 17).

My Basket		
Item	Price	Quantity
E5300	65	<input type="text" value="1"/>
P5Q EM117		<input type="text" value="1"/>
P5QL SE	72	<input type="text" value="1"/>
Total	<input type="button" value="new"/>	254
Proceed to Order		

Εικόνα 17: Καλάθι αγορών γεμάτο

Δημιουργείται ένα session ονόματος `cart` που κρατάει το `id` κάθε προϊόντος που είναι στο καλάθι. Αφού το `id` είναι κύριο κλειδί στον πίνακα των προϊόντων, εκτελείται ένα SQL ερώτημα για να εμφανίσει το όνομα του προϊόντος και την τιμή του όπως φαίνεται στην εικόνα 17. Ακόμη δημιουργείται ένα session `quantity` που κρατάει τα τεμάχια κάθε προϊόντος στο καλάθι. Ο πελάτης μπορεί να δώσει την τιμή των τεμαχίων στο πεδίο `quantity` και πατώντας `new` ενημερώνεται το session `quantity` και το συνολικό κόστος της παραγγελίας(`Total`).

Εφόσον ο πελάτης είναι συνδεδεμένος, εμφανίζεται η επιλογή «Proceed to Order» όπως φαίνεται στη παραπάνω εικόνα. Επιλέγοντάς το, οδηγείται στην σελίδα my_order.php(εικόνα 18).

The screenshot shows a web interface with a navigation bar (Home, Products, Contact) and a sidebar (PRODUCTS, cpu, motherboard, Welcome konstogi, Log out). The main content area is divided into three sections:

- Order Data:** A table with fields for Name, Surname, City, Zip, Address, and Phone Number, all filled with the customer's information. A "Change My Info" link is at the bottom.
- My Order:** A table with columns Item, Price, and Quantity. It lists items E5300, P5Q EM, and P5QL SE, along with a Total of 254. A "new" button is next to the total, and a "Complete Order" link is at the bottom.
- My Basket:** A table with columns Item, Price, and Quantity. It lists the same items as the My Order table. A "new" button is next to the Total of 254, and a "Proceed to Order" link is at the bottom.

Εικόνα 18: Η σελίδα my_order.php

Σε αυτή την σελίδα εμφανίζονται τα στοιχεία του πελάτη που είναι απαραίτητα για την αποστολή της παραγγελίας. Εδώ ο πελάτης αν επιλέξει το «Change my info» πηγαίνει στη σελίδα my_info.php όπου μπορεί να τροποποιήσει τα στοιχεία του(εικόνα 19).

The screenshot shows a web interface similar to the previous one, but with a form for updating customer information. The form has the following fields:

- Name : kostas
- Surname : stogiannis
- City : thessalonikh
- Zip : 54323
- Address : imvrou 10
- Phone Number : 2310612987

At the bottom of the form, there is an "Update" button and a "Back to my order" link. The "My Basket" section on the right is identical to the previous screenshot.

Εικόνα 19: Φόρμα με τις πληροφορίες του πελάτη

Συνεχίζοντας για την σελίδα my_order.php, βλέπουμε την παραγγελία του πελάτη όπου επιλέγοντας το «complete Order» καλείται η complete_order.php. Η complete_order.php εισάγει την παραγγελία στον πίνακα orders της βάσης δεδομένων. Ο πίνακας orders έχει το πεδίο order_id το οποίο είναι κύριο κλειδί και αυξάνει εξορισμού κατά ένα(auto_increment), το πεδίο username που δείχνει ποιος έκανε την παραγγελία, το πεδίο order_date που κρατάει πότε έγινε η παραγγελία και τα πεδία shipment και status που δείχνουν τον τρόπο αποστολής και την κατάσταση της παραγγελίας(εικόνα 20).

			order_id	username	order_date	shipment	status
<input type="checkbox"/>			171	konstogi	2009-09-22		
<input type="checkbox"/>			170	kostas	2009-09-22		
<input type="checkbox"/>			169	kostas	2009-09-22		
<input type="checkbox"/>			168	kostas	2009-09-22		
<input type="checkbox"/>			167	kostas	2009-09-22		
<input type="checkbox"/>			166	konstogi	2009-09-22		
<input type="checkbox"/>			165	konstogi	2009-09-22		
<input type="checkbox"/>			164	kostas	2009-09-20		
<input type="checkbox"/>			172	kostas	2009-09-25		
<input type="checkbox"/>			173	kostas	2009-09-25		
<input type="checkbox"/>			174	kostas	2009-09-27		
<input type="checkbox"/>			175	kostas	2009-09-27		
<input type="checkbox"/>			176	kostas	2009-09-27		
<input type="checkbox"/>			177	kidfgrk	2009-09-27		
<input type="checkbox"/>			178	kidfgrk	2009-09-27		
<input type="checkbox"/>			179	kostas	2009-09-27		
<input type="checkbox"/>			180	kidfgrk	2009-09-27		
<input type="checkbox"/>			181	kostas	2009-09-27		
<input type="checkbox"/>			188	kalitexnis	2009-09-28		
<input type="checkbox"/>			184	ipapagia	2009-09-27		
<input type="checkbox"/>			185	ipapagia	2009-09-27		
<input type="checkbox"/>			186	ipapagia	2009-09-27		
<input type="checkbox"/>			187	ipapagia	2009-09-27		
<input type="checkbox"/>			189	kostas	2009-09-29		
<input type="checkbox"/>			190	kostas	2009-09-29		
<input type="checkbox"/>			191	kostas	2009-10-01		

Εικόνα 20: Μέρος του πίνακα orders από το phpMyAdmin

Ακόμη υπάρχει ο πίνακας orderitem που κρατάει το πεδίο order_id το οποίο είναι το ίδιο με το order_id του πίνακα orders, το πεδίο product_id που αναφέρεται στο id του προϊόντος που παραγγέλθηκε, το πεδίο category που αναφέρει την κατηγορία του προϊόντος της παραγγελίας, το πεδίο quantity για το πλήθος των τεμαχίων που παραγγέλθηκαν, το πεδίο total_value που κρατάει το συνολικό κόστος του προϊόντος(εικόνα 21).

			order_id	product_id	category	quantity	total_value
<input type="checkbox"/>			179	1008	cpu	1	185.00
<input type="checkbox"/>			176	2001	motherboard	3	137.31
<input type="checkbox"/>			176	2002	motherboard	2	270.00
<input type="checkbox"/>			175	1006	cpu	3	448.50
<input type="checkbox"/>			174	2006	motherboard	2	210.00
<input type="checkbox"/>			171	2008	motherboard	1	81.00
<input type="checkbox"/>			171	1007	cpu	1	159.00
<input type="checkbox"/>			171	1001	cpu	1	38.20
<input type="checkbox"/>			170	1002	cpu	5	225.00
<input type="checkbox"/>			170	1004	cpu	2	130.00
<input type="checkbox"/>			170	2009	motherboard	1	117.20
<input type="checkbox"/>			169	2010	cpu	3	250.50
<input type="checkbox"/>			169	1010	motherboard	2	503.80
<input type="checkbox"/>			168	2010	cpu	5	417.50
<input type="checkbox"/>			168	1010	motherboard	1	251.90
<input type="checkbox"/>			167	2010	cpu	3	250.50
<input type="checkbox"/>			167	1010	motherboard	1	251.90
<input type="checkbox"/>			166	2003	cpu	3	375.00
<input type="checkbox"/>			165	1008	motherboard	1	185.00
<input type="checkbox"/>			164	1002	cpu	1	45.00
<input type="checkbox"/>			164	2007	cpu	1	72.40
<input type="checkbox"/>			164	1009	motherboard	1	199.00
<input type="checkbox"/>			181	2005	motherboard	1	138.00
<input type="checkbox"/>			183	2001	motherboard	22	1006.94
<input type="checkbox"/>			188	2003	motherboard	2	250.00

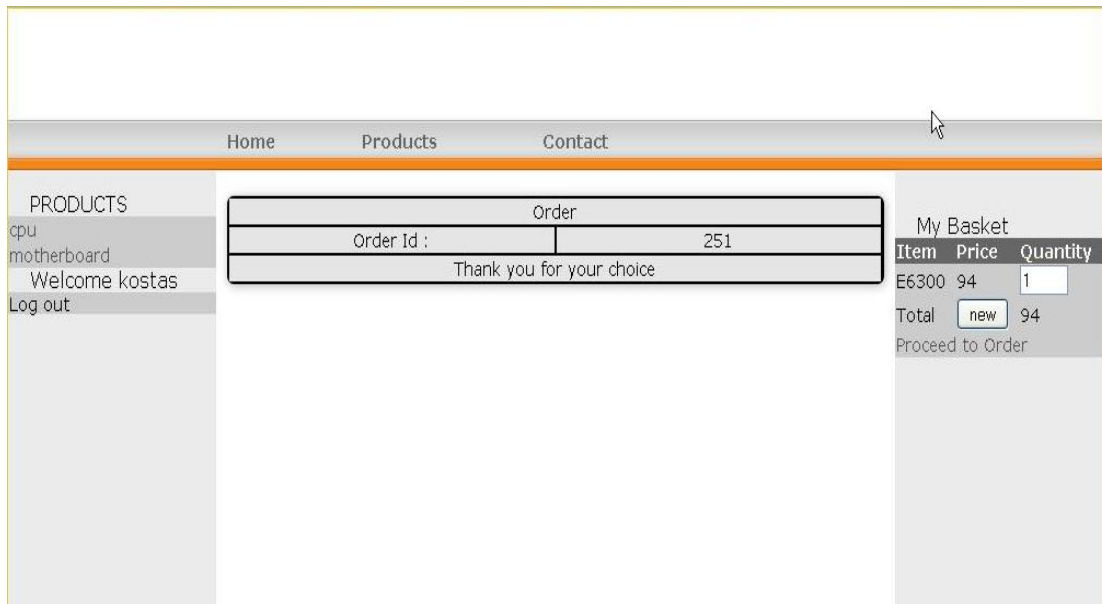
Εικόνα 21: Μέρος του πίνακα orderitem από το phpMyAdmin

Επίσης ενημερώνεται ο πίνακας stock που είναι η αποθήκη του ηλεκτρονικού καταστήματος. Αφαιρούνται οι ποσότητες των προϊόντων που παραγγέλθηκαν και ενημερώνεται το πεδίο με την ποσότητα των τεμαχίων κάθε προϊόντος που πουλήθηκε(εικόνα 22).

			product_id	rest	sold	new	offer
<input type="checkbox"/>			2001	59	36	0	0
<input type="checkbox"/>			2002	63	37	1	0
<input type="checkbox"/>			2003	96	4	0	0
<input type="checkbox"/>			2004	82	18	0	1
<input type="checkbox"/>			2005	64	36	0	0
<input type="checkbox"/>			2006	94	4	1	0
<input type="checkbox"/>			2007	71	29	0	1
<input type="checkbox"/>			2008	78	22	0	0
<input type="checkbox"/>			2009	95	5	0	1
<input type="checkbox"/>			2010	95	5	0	0
<input type="checkbox"/>			1001	45	55	0	0
<input type="checkbox"/>			1002	97	3	1	0
<input type="checkbox"/>			1003	100	0	0	1
<input type="checkbox"/>			1004	89	11	0	0
<input type="checkbox"/>			1005	98	2	1	0
<input type="checkbox"/>			1006	83	17	0	1
<input type="checkbox"/>			1007	93	7	0	0
<input type="checkbox"/>			1008	98	2	0	0
<input type="checkbox"/>			1009	96	4	1	0
<input type="checkbox"/>			1010	91	9	0	0

Εικόνα 22: Η αποθήκη του ηλεκτρονικού καταστήματος από το phpMyAdmin

Τέλος, η complete_order.php αποστέλλει email στον πελάτη την παραγγελία και ανακατευθύνει τον πελάτη στη σελίδα completed_order.php στην οποία εμφανίζεται μήνυμα ολοκλήρωσης της παραγγελίας με τον κωδικό παραγγελίας(εικόνα 23).



Εικόνα 23: Η σελίδα completed_order.php εμφανίζει τον κωδικό της παραγγελίας

5.6 Οι Κύριες Σελίδες του e-Shop

Αφού περιγράφηκαν οι λειτουργίες του ηλεκτρονικού καταστήματος, παρακάτω παρουσιάζονται οι σελίδες του. Όπως αναφέρθηκε στο κεφάλαιο 4, το μόνο κομμάτι που διαφέρει από σελίδα σε σελίδα είναι το πλαίσιο mainContent το οποίο είναι στη μέση. Τα υπόλοιπα τέσσερα μέρη παραμένουν ίδια.

5.6.1 Η αρχική σελίδα index.php

Η αρχική σελίδα index.php εμφανίζει στο mainContent τέσσερεις πίνακες με προτάσεις για προϊόντα. Ο πρώτος εμφανίζει τα πέντε πιο αγορασμένα προϊόντα, ο δεύτερος τα πέντε προϊόντα με την μεγαλύτερη βαθμολογία από τους πελάτες, ο τρίτος εμφανίζει καινούργια προϊόντα και ο τελευταίος της προσφορές του ηλεκτρονικού καταστήματος(εικόνα 24).

Home Products Contact






PRODUCTS






cpu
motherboard
Login Form
Username:






Password:






Login
Sign up

My Basket
Item Price Quantity
Empty Basket

Best Sellers				
				
E1400 38 €	LP DK X48 T2RS 135 €	P5N73 AM 46 €	LP JR GF9400 T2RS 138 €	P5QL SE 72 €

Best Rated				
				
E1500 45 €	P5Q EM 117 €	E1400 38 €	E6300 94 €	P45T A 105 €

New Products				
				
LP DK X48 T2RS 135 €	P45T A 105 €	E1500 45 €	E6300 94 €	Q9550 199 €

Offers				
				
UT P45 T3RS 175 €	P5QL SE 72 €	P5Q EM 117 €	E5200 57 €	E8400 149 €

Εικόνα 24: Η αρχική σελίδα του ηλεκτρονικού καταστήματος(index.php)

5.6.2 Η σελίδα με τη λίστα κάθε προϊόντος

Η σελίδα εμφανίζει μία λίστα με τα προϊόντα κάθε είδους. Η λίστα είναι ταξινομημένη σύμφωνα με την τιμή των προϊόντων και παρουσιάζει το μοντέλο και τον κωδικό του καθενός(εικόνα 25).



Εικόνα 25: Η σελίδα με τη λίστα των επεξεργαστών

Ο πελάτης μπορεί να επιλέξει συγκεκριμένο προϊόν κάνοντας κλικ στη μεσαία στήλη Model όπου κάθε πεδίο είναι υπερσύνδεσμος για το κάθε προϊόν.

5.6.3 Η σελίδα προϊόντος

Η σελίδα εμφανίζει την εικόνα του προϊόντος που επέλεξε να δει ο πελάτης, υπάρχει ο πίνακας με τα χαρακτηριστικά του προϊόντος, ο πελάτης μπορεί να βαθμολογήσει το προϊόν εφόσον είναι συνδεδεμένος και να πληροφορηθεί την βαθμολογία που του έχει δοθεί. Ακόμη παρουσιάζονται τρεις προτάσεις παρόμοιων προϊόντων σύμφωνα με το σύστημα προτάσεων που υλοποιήθηκε στην εργασία και παρουσιάζεται στο επόμενο κεφάλαιο (εικόνα 26).

Home Products Contact

PRODUCTS

cpu
motherboard
Login Form
Username:

Password:

Sign up

My Basket
Item Price Quantity
Empty Basket



ADD TO BASKET 

id	2005
model	LP JR GF9400 T2R5
cpu	Intel Core 2 Duo/Quad Core
socket	s775
chipset	Nvidia 9400
fsb	1333
ram	4 x DDR2 667/800/1066
storage	1 x ATA133/6 x SATAII
raid	SATA Raid 0, 1, 0+1, 5, JBOD
onboard_graphic_card	nVIDIA 9400
graphic_card_slot	PCI Express 16x
pci	1
pci_express	1
usb	4
lan	1000
onboard_audio	7.1
price	138

★☆☆☆☆ Ave. rating: 4 from 4 votes.
Rate product:

Nearest Products

 P45T A 105 €	 LP DK X48 T2R5 135 €	 P5Q PRO 125 €
--	--	--

Why

Created by A.S & K.S

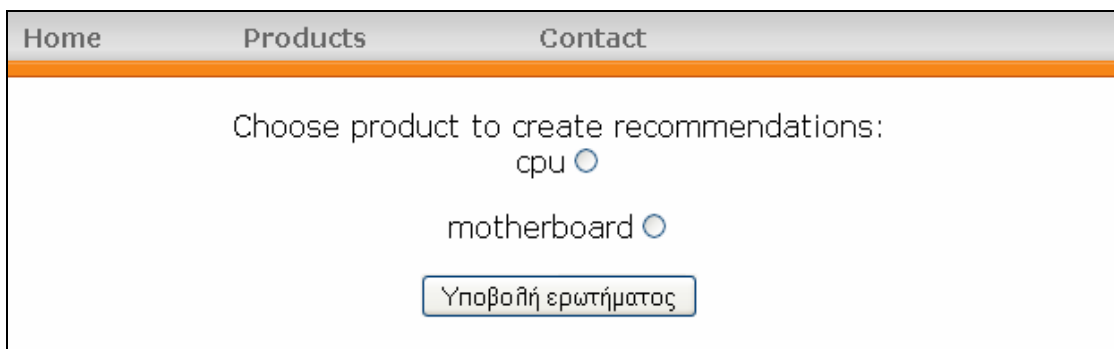
Εικόνα 26: Η σελίδα με προϊόντος

5.7 Η Λογική του Συστήματος Συστάσεων

Το σύστημα συστάσεων που υλοποιήθηκε στην παρούσα εργασία χρησιμοποιεί την γνώση που παρέχεται από τον ειδικό για να διευκολύνει τον πελάτη να επιλέξει το κατάλληλο προϊόν. Το ηλεκτρονικό κατάστημα που αναπτύχθηκε περιλαμβάνει προϊόντα πληροφορικής. Λόγο της φύσης των προϊόντων υπάρχει η δυνατότητα ο ειδικός να εισάγει κριτήρια ομοιότητας ανάμεσα στα προϊόντα. Με αυτόν τον τρόπο το σύστημα συστάσεων προτείνει στον πελάτη όμοια προϊόντα με αυτό που βλέπει εκείνη την στιγμή, βοηθώντας τον στην καλύτερη δυνατή επιλογή.

5.8 Δημιουργία Συστήματος Συστάσεων

Η υλοποίηση του συστήματος συστάσεων γίνεται με την βοήθεια κατάλληλου μενού. Ο ειδικός αρχικά επιλέγει την κατηγορία του προϊόντος που θα αναπτύξει το σύστημα συστάσεων απ την σελίδα `create_rs.php`(εικόνα 27).



Home	Products	Contact
Choose product to create recommendations:		
cpu <input type="radio"/>		
motherboard <input type="radio"/>		
<input type="button" value="Υποβολή ερωτήματος"/>		

Εικόνα 27: Επιλογή κατηγορίας προϊόντος για δημιουργία συστήματος συστάσεων

Επιλέγοντας κατηγορία πηγαίνει στην σελίδα `begin_rs.php` όπου επιλέγει τα χαρακτηριστικά των προϊόντων σύμφωνα με τα οποία θα υπολογιστεί η ομοιότητα τους(εικόνα 28).

Home	Products	Contact
characteristics of motherboard id <input type="checkbox"/> model <input type="checkbox"/> cpu <input checked="" type="checkbox"/> socket <input type="checkbox"/> chipset <input type="checkbox"/> fsb <input checked="" type="checkbox"/> ram <input checked="" type="checkbox"/> storage <input type="checkbox"/> raid <input checked="" type="checkbox"/> onboard_graphic_card <input type="checkbox"/> graphic_card_slot <input checked="" type="checkbox"/> pci <input checked="" type="checkbox"/> pci_express <input checked="" type="checkbox"/> usb <input checked="" type="checkbox"/> lan <input checked="" type="checkbox"/> onboard_audio <input type="checkbox"/> price <input checked="" type="checkbox"/> <input type="button" value="Υποβολή ερωτήματος"/>		

Εικόνα 28: Επιλογή χαρακτηριστικών που θα λυφθούν υπόψη για το σύστημα συστάσεων

Στην σελίδα `type_of_char_rs.php` ο ειδικός επιλέγει αν το κάθε χαρακτηριστικό είναι ποσοτικό ή ποιοτικό (εικόνα 29).

Home	Products	Contact
Choose type of characteristics: fsb:quantitative <input checked="" type="radio"/> qualitative <input type="radio"/> ram:quantitative <input checked="" type="radio"/> qualitative <input type="radio"/> storage:quantitative <input type="radio"/> qualitative <input checked="" type="radio"/> raid:quantitative <input type="radio"/> qualitative <input checked="" type="radio"/> pci:quantitative <input type="radio"/> qualitative <input checked="" type="radio"/> pci_express:quantitative <input type="radio"/> qualitative <input checked="" type="radio"/> usb:quantitative <input type="radio"/> qualitative <input checked="" type="radio"/> lan:quantitative <input type="radio"/> qualitative <input checked="" type="radio"/> price:quantitative <input type="radio"/> qualitative <input checked="" type="radio"/> <input type="button" value="Υποβολή ερωτήματος"/>		

Εικόνα 29: Κατηγοριοποίηση χαρακτηριστικών ως ποσοτικών ή ποιοτικών

Στη σελίδα `weight_of_rs.php` δίνεται το βάρος με το οποίο θα ληφθεί υπόψη η τιμή του κάθε χαρακτηριστικού για τον υπολογισμό της ομοιότητας των προϊόντων (εικόνα 30).

Home	Products	Contact
fsb		
Please give weight: 0.9 ▼		
ram		
Please give weight: 0.7 ▼		
storage		
Please give weight: 0.3 ▼		
raid		
Please give weight: 0.2 ▼		
pci		
Please give weight: 0.3 ▼		
pci_express		
Please give weight: 0.3 ▼		
usb		
Please give weight: 0.3 ▼		
lan		
Please give weight: 0.3 ▼		
price		
Please give weight: 1 ▼		
<input type="button" value="Υποβολή ερωτήματος"/>		

Εικόνα 30: Ο ειδικός δίνει βάρη στα χαρακτηριστικά

Στην επόμενη σελίδα, την `order_qualitive.php`, εμφανίζονται τα βάρη και ταξινομούνται οι τιμές των χαρακτηριστικών που δόθηκαν ως ποιοτικά (Εικόνα 30). Στα ποιοτικά χαρακτηριστικά εμφανίζονται όλες οι τιμές που παίρνουν και δίπλα δίνεται με επιλογή από drop-down menu η σειρά των τιμών.

fsb
The weight of fsb is 0.9.

ram
The weight of ram is 0.7.
Please give order!

2 x DDR2 667/800 1 ▾

4 x DDR2 667/800/1066 4 ▾

2 x DDR2 667/800/1066 2 ▾

4 x DDR2 667/800/1066/1200 5 ▾

4 x DDR2 667/800 3 ▾

4 x DDR3 1066/1333 6 ▾

pci
The weight of pci is 0.3.
Please give order!

2- ▾

1- ▾

3- ▾

pci_express
The weight of pci_express is 0.2.
Please give order!

1- ▾

2- ▾

4- ▾

3- ▾

usb
The weight of usb is 0.2.
Please give order!

4- ▾

2- ▾

lan
The weight of lan is 0.3.
Please give order!

10/100- ▾

1000- ▾

price
The weight of price is 1.

Εικόνα 31: Ταξινόμηση των ποιοτικών χαρακτηριστικών

Το σύστημα προτάσεων ολοκληρώνετε στην σελίδα complete_rs.php. Εδώ δημιουργείται στη βάση ένας πίνακας που κρατάει τα βάρη και τους τύπους των χαρακτηριστικών, όπως δόθηκαν από τον ειδικό (εικόνα 32).

			characteristic	weight	type_of_char
<input type="checkbox"/>			price	1	quantitative
<input type="checkbox"/>			fsb	0.9	quantitative
<input type="checkbox"/>			ram	0.7	quantitative
<input type="checkbox"/>			pci	0.3	qualitative
<input type="checkbox"/>			pci_express	0.2	qualitative
<input type="checkbox"/>			usb	0.2	qualitative
<input type="checkbox"/>			lan	0.3	qualitative

Εικόνα 32: Πίνακας με τα βάρη και τον τύπο κάθε χαρακτηριστικού

Στη συνέχεια δημιουργούνται πίνακες με τις τιμές των χαρακτηριστικών και τις κανονικοποιημένες τιμές. Οι κανονικοποιημένες τιμές έχουν εύρος 0 έως 10 και υπολογίζονται ως εξής:

$$\text{range} = (\max(\text{price_of_characteristic}) - \min(\text{price_of_characteristic})) / 10$$

$$((\text{current}(\text{p_o_c}) - \min(\text{price_of_characteristic})) / \text{range}) * \text{weight}$$

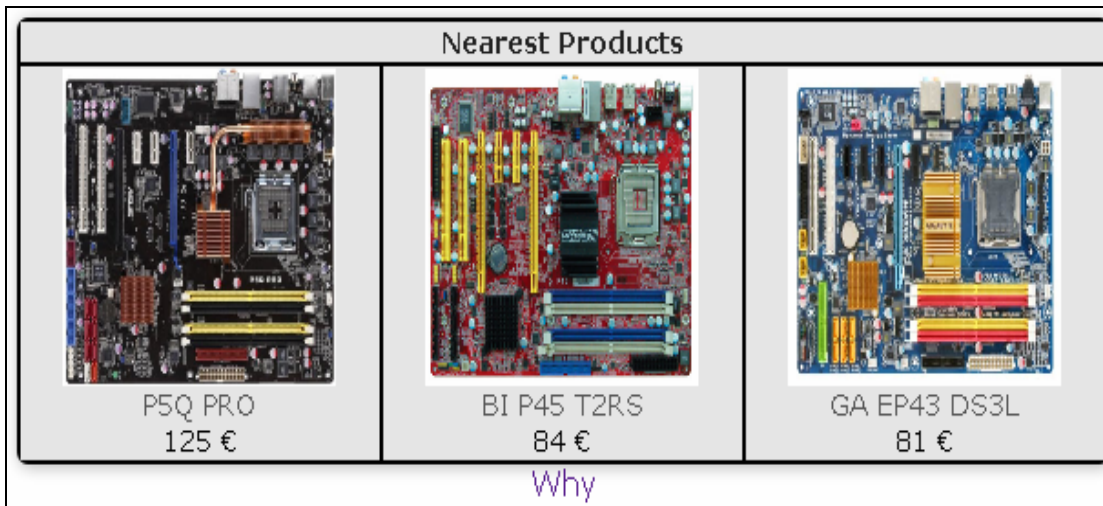
όπου $\max(\text{price_of_characteristic})$ η μέγιστη τιμή ενός χαρακτηριστικού του προϊόντος, αντίστοιχα $\min(\text{price_of_characteristic})$ η ελάχιστη τιμή, $\text{current}(\text{price_of_characteristic})$ η τρέχουσα τιμή του χαρακτηριστικού, weight το βάρος που αντιστοιχεί σε κάθε χαρακτηριστικό και range το εύρος τιμών του. Παράδειγμα στην εικόνα 33 βλέπουμε τον πίνακα `motherboard_ram` με 2 στήλες. Την στήλη `ram` όπου κρατούνται οι διακριτές τιμές του χαρακτηριστικού και την στήλη `rsram` όπου έχουμε τις κανονικοποιημένες τιμές. Βλέπουμε ότι επειδή το βάρος είναι 0.7 η μεγαλύτερη κανονικοποιημένη τιμή της `rsram` είναι η 7 $((6-1)/0.5)*0.7$.

			ram	rsram
<input type="checkbox"/>			2 x DDR2 667/800	0
<input type="checkbox"/>			4 x DDR2 667/800/1066	4.2
<input type="checkbox"/>			2 x DDR2 667/800/1066	1.4
<input type="checkbox"/>			4 x DDR2 667/800/1066/1200	5.6
<input type="checkbox"/>			4 x DDR2 667/800	2.8
<input type="checkbox"/>			4 x DDR3 1066/1333	7

Εικόνα 33: Πίνακας με τις διακριτές τιμές του χαρακτηριστικού `ram` και τις κανονικοποιημένες

5.9 Υπολογισμού Ομοιότητας Προϊόντων

Εφόσον είδαμε τον τρόπο με τον οποίο ο ειδικός εισάγει την γνώση και την εμπειρία του, μπορούμε τώρα να υπολογίσουμε την ομοιότητα των προϊόντων. Όταν ο πελάτης κατά την περιήγησή του βρίσκεται στην σελίδα `sel.php`, σελίδα όπου εμφανίζονται τα χαρακτηριστικά του προϊόντος, καλείται η συνάρτηση `eykl_ar.php` όπου γίνεται ο υπολογισμός της ομοιότητας του τρέχοντος προϊόντος με τα υπόλοιπα προϊόντα ίδιας κατηγορίας. Σαν αποτέλεσμα επιστρέφει τα τρία πιο κοντινά(εικόνα 34).



Εικόνα 34: Πρόταση τριών όμοιων προϊόντων με αυτό που βλέπει ο χρήστης

Η συνάρτηση δουλεύει ως εξής:

1. Δέχεται ως όρισμα τον κωδικό και την κατηγορία του προϊόντος(`eykl_ap($product, $id)`).
2. Εκτελεί SQL ερώτημα όπου επιστρέφονται οι πίνακες με τις κανονικοποιημένες τιμές(εικόνα 35).

```

7  $sql_n="show tables from project_1 like '";
8  $sql_n.="$product";
9  $sql_n.="_?";
10
11 $names=mysql_query($sql_n); //epistrefei tous kanonikopoihmenous pinakes
12 $i=0;
13 while($row = mysql_fetch_array($names))
14 {
15     $str_name[$i]=$row[0]; //o str_name krataei ta onomata twm pinakwn pou exoun tis kanonikopoih
16     $i++;
17 }

```

Εικόνα 35: ερώτημα επιστρέφει τα ονόματα των πινάκων με τις κανονικοποιημένες τιμές

3. Υπολογίζει την ευκλείδεια απόσταση ανάμεσα στα χαρακτηριστικά του τρέχοντος προϊόντος με τα υπόλοιπα(εικόνα 36).

```

19 for($j=0;$j<count($str_name);$j++){ //υπολογίζει την ευκλείδεια απόσταση του προϊόντος που βλέπουμε με τα υπόλοιπα
20
21
22 $rs_pedio=show_pedia($str_name[$j]); //επιστρέφει τα ονόματα των προϊόντων του κάθε πίνακα
23
24 $sql_normal_trexon= mysql_query("SELECT $rs_pedio[1] FROM $str_name[$j] , $product Where $str_name[$j].$rs_pedio[0]=$product
$product.id=$id");
25 $row = mysql_fetch_array($sql_normal_trexon);
26
27 $id_product = mysql_query("SELECT id FROM $product where $product.id!=$id");
28
29 $i=0;
30 while($row3 = mysql_fetch_array($id_product))
31 {
32
33 $temp_id=$row3['id'];
34
35 $sql_normal[$i]=mysql_query("SELECT $rs_pedio[1] FROM $str_name[$j] , $product Where $str_name[$j].$rs_pedio[0]=$product.
$product.id=$temp_id");
36
37 $row2 = mysql_fetch_array($sql_normal[$i]);
38 $char[$j]=$row[$rs_pedio[1]];
39 $apostasi[1][$i]= square($row[$rs_pedio[1]]-$row2[$rs_pedio[1]]);
40 $apostasi[0][$i]=$temp_id;
41
42
43
44 $i++;
45 $min_id_length=$i;
46 } //kleinei h while
47 } //kleinei h for

```

Εικόνα 36: Υπολογισμός ευκλείδειας απόστασης

Η ευκλείδεια απόσταση των χαρακτηριστικών μεταξύ του τρέχοντος προϊόντος και των υπολοίπων υπολογίζεται ως το άθροισμα του τετραγώνου της διαφοράς των κανονικοποιημένων τιμών κάθε χαρακτηριστικού, όπως φαίνεται στην εικόνα 36 γραμμή 39. Υπολογίζονται όλες οι αποστάσεις μεταξύ του τρέχοντος προϊόντος και των υπολοίπων, και ταξινομούνται οι κωδικοί των προϊόντων σύμφωνα με την μεγαλύτερη ομοιότητα μεταξύ του τρέχοντος προϊόντος. Δηλαδή από τα προϊόντα που έχουν μικρή ευκλείδεια απόσταση με το τρέχον, προς αυτά με μεγάλη ευκλείδεια απόσταση (εικόνα 37).

```

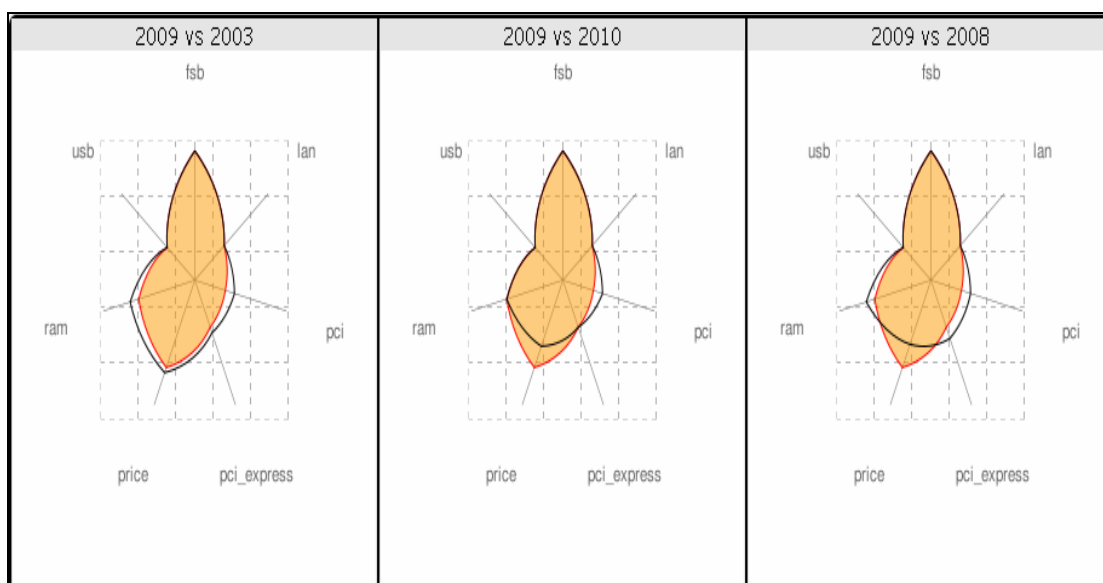
49 for($i=0;$i<$min_id_length;$i++){// o pinakas min_id krataei ta id se shra apo to nearest sto pio makrino
50     $min_ap=array_keys($apostasi[1],min($apostasi[1]));
51     $min_id[$i]=$apostasi[0][$min_ap[0]];
52     unset($apostasi[1][$min_ap[0]]);
53 }

```

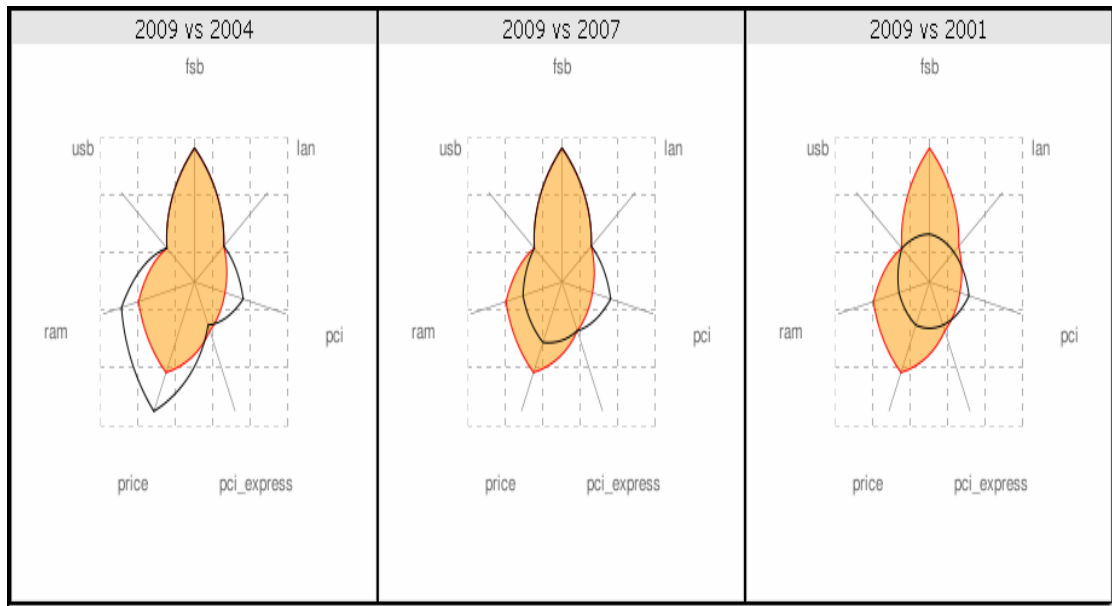
Εικόνα 37: Ταξινόμηση από το ομοιότερο προς το πιο ανόμοιο

5.10 Απεικόνιση Ομοιότητας

Για να γίνει πιο κατανοητή η ομοιότητα μεταξύ των αντικειμένων χρησιμοποιήθηκε το εργαλείο της Google “Google Chart Tools”. Συγκεκριμένα το διάγραμμα τύπου ραντάρ (Radar Charts). Το “Google Chart Tools” δίνει τη δυνατότητα εισαγωγής σχεδιαγραμμάτων σε κάθε σελίδα του ιστού. Το διάγραμμα παράγεται δυναμικά με την χρήση των μεταβλητών της PHP. Επιλέχτηκε το διάγραμμα τύπου ραντάρ γιατί δίνει την καλύτερη εντύπωση για την απεικόνιση της ομοιότητας των προϊόντων. Κάθε άξονας αναφέρεται σε ένα χαρακτηριστικό των προϊόντων και οι τιμές είναι κανονικοποιημένες. Σε κάθε διάγραμμα επιτρέπεται να αναπαρίσταται οποιοδήποτε πλήθος διαφορετικών μεταβλητών, στην περίπτωση μας χαρακτηριστικών προϊόντων πληροφορικής. Παρακάτω απεικονίζονται τα τρία πιο όμοια και τα τρία πιο ανόμοια προϊόντα με αυτό που βλέπει ο χρήστης το οποίο περιέχει το χρώμα(εικόνα 38,39).



Εικόνα 38: Απεικόνιση των τριών πιο όμοιων προϊόντων με το προϊόν με κωδικό 2009



Εικόνα 39: Απεικόνιση των τριών πιο ανόμοιων προϊόντων με το προϊόν με κωδικό 2009

Τα διαγράμματα εμφανίζονται κάνοντας κλικ στην υπερσύνδεση “Why” όπως φαίνεται στην εικόνα 34. Η σελίδα είναι η why_nearest.php.

6 Συμπεράσματα

Το μεγάλο πρόβλημα που προκύπτει από την ανάπτυξη του ηλεκτρονικού εμπορίου είναι το γνωστό σαν πρόβλημα πληροφοριακής υπερφόρτωσης (information overload). Ο πελάτης έχει να διαλέξει ανάμεσα σε ένα μεγάλο πλήθος προϊόντων κάτι που τον προτρέπει να αποκτήσει καλή εικόνα της αγοράς πριν προβεί σε συναλλαγές. Με την αύξηση όμως των προσφερόμενων επιλογών, αυξάνει παράλληλα και ο φόρτος που απαιτείται για την επιλογή των προϊόντων που ταιριάζουν στις ανάγκες και στις απαιτήσεις του χρήστη. Τα συστήματα συστάσεων έρχονται να λύσουν αυτό το πρόβλημα. Χρησιμοποιούν γνώση που παρέχεται είτε από ειδικούς ή που εξάγεται από την παρακολούθηση της συμπεριφοράς των πελατών για να οδηγήσουν τους τελευταίους στον εντοπισμό των προϊόντων-υπηρεσιών που τους ενδιαφέρουν.

Στη παρούσα πτυχιακή εργασία μελετήθηκε η βιβλιογραφία για την ανάπτυξη και την χρήση των συστημάτων συστάσεων. Παρουσιάστηκαν οι σημαντικότερες μέθοδοι, οι χρήσεις τους και τα δομικά συστατικά τους. Το ηλεκτρονικό κατάστημα που αναπτύχθηκε περιέχει κάποια απλοϊκά συστήματα συστάσεων στην αρχική του σελίδα όπου προτείνονται τα προϊόντα με την καλύτερη βαθμολογία (best rated) από τους πελάτες και τα προϊόντα με τις περισσότερες πωλήσεις (best sellers). Επιπλέον, το βασικό σύστημα συστάσεων που αναπτύχθηκε προσομοιάζει ικανοποιητικά την συμβουλή και την εμπειρία του ειδικού. Θα μπορούσε να χρησιμοποιηθεί σε υβριδικές μορφές, σε συνδυασμό με τεχνικές βάσει περιεχομένου και τεχνικές συνεργατικού φιλτραρίσματος, ιδιαίτερα για να επιλύσει τα προβλήματά που εμφανίζουν στα αρχικά στάδια λειτουργίας ενός ηλεκτρονικού καταστήματος. Βέβαια, θα πρέπει πάντα να αναφέρονται σε δεδομένα με ικανοποιητικό αριθμό ποσοτικών και ποιοτικών μετρίσιμων χαρακτηριστικών ώστε να μπορεί να λειτουργήσει.

Το ηλεκτρονικό κατάστημα είναι εύχρηστο και φιλικό προς τον τελικό χρήστη. Η διαχείριση και η παραμετροποίηση του είναι αρκετά εύκολη χάρη και στις απεριόριστες δυνατότητες που δίνει η γλώσσα PHP. Βέβαια θα μπορούσε να βελτιωθεί αρκετά περισσότερο με τη βελτίωση της εμφάνισης και την εισαγωγή περαιτέρω λειτουργιών όπως πληρωμή με πιστωτική κάρτα, εισαγωγή δικλίδων ασφαλείας κ.ά.

Ο συνδυασμός λογισμικού Apache-PHP-MySQL αποτελεί ένα πολύ καλό εργαλείο που παρέχει πολύ καλά αποτελέσματα σε ικανοποιητικό χρόνο. Η λειτουργικότητα της εφαρμογής ελέγχεται άμεσα και η δημιουργία και η διαχείριση της βάσης δεδομένων είναι πολύ εύκολη και ομαλή με αποτέλεσμα τις πολύ καλές επιδόσεις σε παραγωγικότητα.

Η διαδικασία ανάπτυξης της εφαρμογής αν και χρονοβόρα και επίπονη κυρίως λόγω έλλειψης εμπειρίας, υπήρξε αρκετά ενδιαφέρουσα και πρόσφερε τις βάσεις για την απόκτηση γνώσεων και δεξιοτήτων πάνω σε θέματα ηλεκτρονικού εμπορίου που δεν είναι εφικτά στο πλαίσιο ενός εξαμηνιαίου μαθήματος. Τα προβλήματα και οι δυσκολίες που παρουσιάστηκαν αντιμετωπίστηκαν με την χρήση πληροφοριών που ήταν δημοσιευμένες σε διαδικτυακούς τόπους όπως forums και διαδικτυακά tutorials, κάτι που δείχνει και το ενδιαφέρον που υπάρχει για υποδομές ηλεκτρονικού εμπορίου.

Βιβλιογραφία

Adomavicius, G., & Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: a Survey of the State-of-the-art and Possible Extensions. *IEEE Transactions on Knowledge Management and Data Engineering*, 17(6): 734-749.

Balabanovic, M., & Shoham, Y. (1997). Fab: Content-Based, Collaborative Recommendation. *Communication of ACM*, 40(3): 66-72.

Basu, C., Hirsh, H., & Cohen, W. (1998). Recommendation as Classification: Using Social and Content-Based Information in Recommendation. *Recommender Systems Workshop, Technical Report WS-98-08*, AAAI Press.

Billsus, D., & Pazzani, M. (1998). Learning Collaborative Information Filters. *Proceedings of the International Conference on Machine Learning*.

Billsus, D., & Pazzani, M. (2000). User Modeling for Adaptive News Access. *User Modeling and User-Adapted Interaction*, 10(2-3): 147-180.

Breese, J.S., Heckerman, D., & Kadie, C. (1998). Empirical Analysis of Predictive Algorithms for Collaborative Filtering. *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*.

Condliff, M., Lewis, D., Madigan, D., & Posse, C. (1999). Bayesian Mixed-Effects Models for Recommender Systems. *Proceedings of the ACM SIGIR'99 Workshop on Recommender Systems: Algorithms and Evaluation*.

Delgado, J., & Ishii, N. (1999). Memory-Based Weighted-Majority Prediction for Recommender Systems. Proceedings of the ACM SIGIR'99 Workshop on Recommender Systems: Algorithms and Evaluation.

Deshpande, M., & Karypis, G. (2004). Item-Based Top-N Recommendation Algorithms. ACM Transactions of Information Systems, 22(1): 143-177.

DTV (2005). Digital Dictionary. Retrieved from:
http://www.iptv.org/digital/dictionary_internet.cfm.

Good, N., Schafer, J.B., Konstan, J.A., Borchers, A., Sarwar, B., Herlocker, J.L., & Riedl, J. (1999). Combining Collaborative Filtering with Personal Agents for Better Recommendations. Proceedings of the Conference of American Association of Artificial Intelligence (AAAI'99), 439-446.

Markellou, P., Rigou, M., & Sirmakessis, S. (2005). An Online Consumer Purchase Decision Cycle. In Clarke, I., & Flaherty, Th. (Eds.), Advances in Electronic Marketing, Idea Group Publishing Inc., 287-299.

Melville, P., Mooney, R.J., & Nagarajan, R. (2002). Content-Boosted Collaborative Filtering for Improved Recommendations. Proceedings of the 18th National Conference on Artificial Intelligence.

Pazzani, M. (1999). A Framework for Collaborative, Content-Based, and Demographic Filtering. Artificial Intelligence Review, 393-408.

Pazzani, M., & Billsus, D. (1997). Learning and Revising User Profiles: The Identification of Interesting Web Sites, Machine Learning Journal, 27: 313-331.

Pennock, D.M., & Horvitz, E. (1999). Collaborative Filtering by Personality Diagnosis: A Hybrid Memory And Model-Based Approach. Proceedings of the In-

ternational Joint Conference on Artificial Intelligence Workshop: Machine Learning for Information Filtering.

Popescul, A., Ungar, L.H., Pennock, D.M., & Lawrence, S. (2001). Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments. Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence.

Resnick, P., Iakovou, N., Sushak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An Open Architecture for Collaborative Filtering of Netnews. Proceedings of the Computer Supported Cooperative Work Conference.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-Based Collaborative Filtering Recommendation Algorithms. Proceedings of the 10th International WWW Conference.

Shani, G., Brafman, R., & Heckerman, D. (2002). An MDP-Based Recommender System. Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence.

Soboroff, I., & Nicholas, C. (1999). Combining Content and Collaboration in Text Filtering. Proceedings of the International Joint Conference on Artificial Intelligence Workshop: Machine Learning for Information Filtering.

Shardanand, U, & Maes, P. (1995). Social Information Filtering: Algorithms for Automating 'Word of Mouth'. Proceedings of the Conference on Human Factors in Computing Systems.

Tran, T., & Cohen, R. (2000). Hybrid Recommender Systems for Electronic Commerce. Proceedings of the Knowledge-Based Electronic Markets, AAAI Workshop, Technical Report WS-00-04, AAAI Press.

Yu, K., Xu, X., Tao, J., Ester, M., & Kriegel, H.-P. (2002). Instance Selection Techniques for Memory-Based Collaborative Filtering. Proceedings of the 2nd SIAM International Conference on Data Mining (SDM'02).

Μαρκέλλου, Π. (2005). Ηλεκτρονικό Κατάστημα: Σχεδιασμός και Λειτουργίες. Κεφάλαιο 3 στο βιβλίο e-Επιχειρηματικότητα: από τη Θεωρία στην Πράξη, Εκδόσεις Ελληνικά Γράμματα – Ε.Α.Ι.Τ.Υ. CTI press, Σειρά: Τεχνολογία Πληροφορικής και Επικοινωνιών, pp. 105-144.

Μαρκέλλου, Π. (2005). Τεχνικές και Συστήματα Διαχείρισης Γνώσης στο Διαδίκτυο, 97-124.

Διαδικτυακές πηγές

Amazon.com, <http://www.amazon.com/>

BSD Licenses, <http://www.bsdnewsletter.com/bsd/license.html>

Css, <http://www.w3schools.com/css/>

Google Chart Tools, <http://code.google.com/intl/el-GR/apis/charttools/>

Html, <http://www.w3schools.com/html/default.asp>

LAMP, <http://www.lamphowto.com/>

MySQL, <http://www.mysql.com/>

PHP, <http://www.php.net/>

The Apache software foundation, <http://www.apache.org/>

Uniform Server, <http://www.uniformserver.com/>

WAMP, <http://www.wampserver.com/en/>