

Ανώτατο Τεχνολογικό Εκπαιδευτικό Ίδρυμα Πατρών (Παράρτημα Αμαλιάδας)
Σχολή Διοίκησης και Οικονομίας
Τμήμα Εφαρμογών Πληροφορικής στη Διοίκηση & στην Οικονομία

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΥΠΗΡΕΣΙΕΣ ΕΞΑΤΟΜΙΚΕΥΣΗΣ ΣΤΟ ΗΛΕΚΤΡΟΝΙΚΟ ΕΜΠΟΡΙΟ



Μελέτη
Κουφασίμη Κωνσταντίνα
Νικολόπουλος Νικόλαος

Επιβλέπων
Κουτσονίκος Ιωάννης

Πάτρα 2014

**ΥΠΗΡΕΣΙΕΣ ΕΞΑΤΟΜΙΚΕΥΣΗΣ ΣΤΟ ΗΛΕΚΤΡΟΝΙΚΟ
ΕΜΠΟΡΙΟ**

Περίληψη

Αντικείμενο μελέτης της παρούσας εργασίας θα αποτελέσει ο κλάδος υπηρεσιών εξατομίκευσης στο ηλεκτρονικό εμπόριο. Δεδομένου ότι η εξατομίκευση στο ηλεκτρονικό εμπόριο παίζει καθοριστικό ρόλο ώστε να δημιουργηθεί μια στενή σχέση με τον πελάτη η ανάπτυξη εφαρμογών που θα αντιμετωπίζουν τους πελάτες προσωπικά, θα καθορίσει και το ποιος θα επικρατήσει στην αγορά. Τα προβλήματα που καλείται ο ερευνητής να αντιμετωπίσει σχετικά με την εξατομίκευση είναι η αναγνώριση του πελάτη, η συγκέντρωση πληροφοριών γύρω από τον πελάτη, καθώς και η επεξεργασία στοιχείων για να κάνει μια υπηρεσία να φαίνεται προσωπικά σχεδιασμένη για ένα συγκεκριμένο πελάτη.

Η εργασία θα αναπτυχτεί σε τρία κεφάλαια. Στο πρώτο κεφάλαιο θα παρουσιαστούν τα γενικά χαρακτηριστικά του ηλεκτρονικού εμπορίου, όπως αυτά έχουν διαμορφωθεί με την υπάρχουσα τεχνολογία. Στο δεύτερο κεφάλαιο θα αναλυθεί η τεχνολογία των επιχειρησιακών συστημάτων EPR και των επιχειρηματικών διαδικασιών που υποστηρίζουν. Αντίστοιχα στο τρίτο κεφάλαιο θα γίνει η ανάλυση web εφαρμογών. Και στις δύο περιπτώσεις θα παρουσιαστούν παραδείγματα ηλεκτρονικών καταστημάτων που προσφέρουν υπηρεσίες εξατομίκευσης, ώστε να γίνουν απτές οι διαφορές μεταξύ των δύο τεχνολογιών. Στα συμπεράσματα της εργασίας θα παρουσιαστούν συγκριτικά αποτελέσματα.

Abstract

The subject of this paper will provide the services sector personalization in electronic commerce. Since personalization in e-commerce plays a key role in order to create a close relationship with the client to develop applications that will treat their customers personally, will determine who will prevail in the market. The problems required the researcher to face on the personalization is to identify the client to gather information about the customer, as well as data processing to make a service appear personally designed for a specific customer.

The work will thrive in three chapters. The first chapter will present the general characteristics of electronic commerce as they have been shaped by the existing technology. The second chapter will analyze the technology operational ERP systems and business processes they support. Respectively in the third chapter will be analyzed web applications. In both cases will be presented examples of online stores that offer personalization to make tangible differences between the two technologies. The conclusions of the work will be presented comparative results.

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΙΣΑΓΩΓΗ	11
i. Στόχος Μελέτης	12
ii. Μεθοδολογία Εργασίας.....	12
iii. Οι Έννοιες Πληροφορία και Δεδομένο	13
1.1 ΕΞΑΤΟΜΙΚΕΥΣΗ ΣΤΟ ΗΛΕΚΤΡΟΝΙΚΟ ΕΜΠΟΡΙΟ	15
1.1 Ορισμός Ηλεκτρονικού Εμπορίου.....	15
1.2 Ορισμός Εξατομίκευσης.....	16
1.3 Στόχος Εξατομίκευσης	17
1.4 Διαδικασία Εξατομίκευσης	18
1.4.1 Συλλογή Δεδομένων (Data Collection).....	18
1.4.2 Προεπεξεργασία δεδομένων (Data Preprocessing).....	19
1.4.3 Ανακάλυψη Προτύπων (Pattern Discovery)	21
1.4.4 Μεταεπεξεργασία Γνώσης (Knowledge Postprocessing)	27
1.5 Κατάλληλα Sites για Εξατομίκευση.....	27
1.6 Στοιχεία ενός Συστήματος Εξατομίκευσης	28
1.6.1 Διάλογος μεσω Statecharts.....	30
1.6.2 Σύστημα Διαχείρισης Πόρων	34
1.6.3 Συστήματα Συστάσεων (Recommender Systems).....	36
Συμπεράσματα Κεφαλαίου	39
2. ENTERPRISE RESOURCE PLANNING.....	41
2.1 Συστήματα Οργάνωσης Επιχειρήσεων.....	41
2.2 Εξελεγκτική Διαδικασία Συστημάτων Οργάνωσης	42
2.3 Πληροφοριακά Συστήματα.....	43
2.4 Προσφορά Συστήματος ERP στην Επιχείρηση.....	44
2.5 Λειτουργικότητα Συστήματος ERP.....	46
2.6 Χρόνος Υλοποίησης Εφαρμογής.....	51

2.7	Προσδοκώμενα Οφέλη από την Εφαρμογή του Συστήματος	51
2.8	Σχέση ERP με E-Commerce.....	52
2.9	Η Αρχιτεκτονική ενός ERP	56
2.9.1	Database Server.....	57
2.9.2	Application Server.....	58
2.9.3	Client.....	58
3.	ΣΥΝΕΡΓΑΤΙΚΗ ΔΙΗΘΗΣΗ ΔΕΔΟΜΕΝΩΝ	60
3.1	Φιλτράρισμα Πληροφοριών	60
3.2	Ορισμός Συνεργατικής Διήθησης Δεδομένων.....	63
3.3	Αναγκαιότητα Συστημάτων.....	64
3.4	Στόχος Συστήματος	65
3.5	Γενική Δομή Συστημάτων Collaborate Filtering.....	67
3.6	User-based Collaborative Filtering.....	68
3.5.1	Παράδειγμα Μελέτης	69
3.7	Item-based Collaborative Filtering.....	71
3.6.1	Παράδειγμα Μελέτης	73
3.8	Content Based Filtering	74
3.9	Αλγόριθμοι	75
3.9.1	Αλγόριθμοι Βασισμένοι σε Μνήμη.....	77
3.9.2	Αλγόριθμοι βασισμένοι σε μοντέλο.....	78
3.9.3	Σύγκριση Αλγορίθμων	79
3.10	Περιορισμοί Εφαρμογής.....	81
3.11	Αντιμετώπιση Προβλημάτων.....	83
4.	ΣΥΓΚΡΙΣΗ ΜΕΘΟΔΩΝ ΣΥΝΕΡΓΑΤΙΚΗΣ ΔΙΗΘΗΣΗΣ ΔΕΔΟΜΕΝΩΝ	84
4.1	Μελέτη Συγκεκριμένων Περιπτώσεων.....	84
4.2	Εφημερίδα Guardian.....	85
4.3	Cinematch - Ενοικιάσεις Ταινιών.....	88

4.4	Twitter – Follower of Follower	91
4.5	eBay	93
4.6	Amazon.....	95
4.7	YouTube	97
4.8	Μελέτη Ηλεκτρονικού Καταστήματος Εταιρείας Ιατρικών Μηχανημάτων	100
4.8.1	Εφαρμογή εισόδου εγγεγραμμένου χρήστη στο σύστημα (login)	100
4.8.2	Εφαρμογή εγγραφής νέου χρήστη στο σύστημα (signup)	101
4.8.3	Εφαρμογή πλοήγησης	102
4.8.4	Εφαρμογή online personalization.....	105
	ΣΥΜΠΕΡΑΣΜΑΤΑ.....	107
	Bibliography.....	109

Ευρετήριο Πινάκων

Πίνακας 1 Το ηλεκτρονικό εμπόριο μπορεί να οριστεί από τέσσερις διαφορετικές οπτικές	15
Πίνακας 2 Τεχνικές συλλογής και ανάλυσης δεδομένων	19
Πίνακας 3 Συστήματα φιλτραρίσματος.....	21
Πίνακας 4 Οι τέσσερις λειτουργίες της εξατομίκευσης.....	29
Πίνακας 5 Τύποι εισόδου συστήματος συστάσεων	36
Πίνακας 6 Βήματα σχεδίασης συστημάτων.....	38
Πίνακας 7 Οφέλη υιοθέτησης συστήματος ERP	44
Πίνακας 8 Ποσότητες παραγόμενων πληροφοριών σε παγκόσμιο επίπεδο (Μ. Ρήγκου, 2012)	64
Πίνακας 9 Διαγραμματική απεικόνιση του φαινομένου Information Overload όπου λόγω του όγκου των πληροφοριών οι χρήστες χάνουν συχνά το στόχο τους (Μ. Ρήγκου, 2012)	66
Πίνακας 10 Καταγραφή δεδομένων.....	69
Πίνακας 11 Δημιουργία ζευγών χρηστών.....	70
Πίνακας 12 Προσδιορισμός βαθμολόγησης προϊόντος από τον χρήστη	71
Πίνακας 13 Καταγραφή δεδομένων.....	73
Πίνακας 14 Παράδειγμα έλλειψης απαιτούμενων πληροφοριών για την λήψη πρόβλεψης	81

Ευρετήριο Εικόνων

Εικόνα 1 Σχέσεις μεταξύ τεχνολογιών ERP – Content - Based Filtering & Personalization	13
Εικόνα 2 Η διαδικασία συλλογής πληροφοριών του Facebook βασίζεται στην τεχνική explicit profiling.....	18
Εικόνα 3 n-Gram βασισμένο σε σημείο (α) και n-gram βασισμένο σε μονοπάτι (β).	26
Εικόνα 4 Λειτουργίες – στάδια εξατομίκευσης	29
Εικόνα 5 Η on-line φάση της διαδικασίας εξατομίκευσης	30
Εικόνα 6 Statechart (διάγραμμα κατάστασης) για το θέμα «painting» (Sven R., 2003)	32
Εικόνα 7 Statechart (διάγραμμα κατάστασης) για όλο τον διάλογο, (Sven R., 2003)	33
Εικόνα 8 Σε επίπεδο στήριξης λειτουργίας καταστημάτων ηλεκτρονικού εμπορίου το ERP λειτουργεί ως βοηθητικό παρασκηνιακό σύστημα εκτέλεσης συναλλαγών.....	35
Εικόνα 9 Εξέλιξη πληροφοριακών συστημάτων	42
Εικόνα 10 Βασικές λειτουργίες ενός πληροφοριακού συστήματος.....	43
Εικόνα 11 Αλληλεπίδραση υποσυστημάτων	46
Εικόνα 12 Σύνδεση συστήματος βάσεις δεδομένων με τις διάφορες εφαρμογές της επιχείρησης.....	50
Εικόνα 13 Η συνδρομή του ERP στην επίτευξη των εταιρικών στόχων (Π. Παναγιωτόπουλος, 2007).....	51
Εικόνα 14 Η δομή του συστήματος συνίσταται σε 4 στάδια	67
Εικόνα 15 Διάκριση συστημάτων συνεργατικής διήθησης	68
Εικόνα 16 Η δομή του συστήματος συνίσταται σε 4 στάδια	68
Εικόνα 17 Εφαρμογή εισόδου επισκέπτη	85
Εικόνα 18 Απόσπασμα αρχικής σελίδας.....	86
Εικόνα 19 Οι σχετικές προτάσεις της εφημερίδας, βάσει του προηγούμενου άρθρου που αναγνώστηκε	86

Εικόνα 20 Εφαρμογή πλοήγησης	87
Εικόνα 21 Αποτελέσματα Recommender System στο Netflix	89
Εικόνα 22 Κατάταξη των ταινιών βάσει της λίστας αναμονής που υπάρχει για αυτές. Για κάθε μια ταινία δίνεται η συνολική αξιολόγηση που έχει λάβει από το κοινό.....	90
Εικόνα 23 Απόσπασμα της αρχικής σελίδας του Twitter. Στην δεξιά στήλη υπάρχει το παράθυρο σύστασης προϊόν να ακολουθήσει ο χρήστης.....	91
Εικόνα 24 Το συνεργατικό φιλτράρισμα με βάση το προϊόν εφαρμόζεται ευρέως στο Amazon.com	96
Εικόνα 25 Παράδειγμα εξήγησης σύστασης που χρησιμοποιεί συνεργατικό φιλτράρισμα με βάση το προϊόν	96
Εικόνα 26 Παράδειγμα εξήγησης σύστασης που χρησιμοποιεί συνεργατικό φιλτράρισμα με βάση το προϊόν	97
Εικόνα 27 Εφαρμογή εισόδου επισκέπτη	100
Εικόνα 28 Εφαρμογή εγγραφής νέου χρήστη	101
Εικόνα 29 Εφαρμογή πλοήγησης.....	103
Εικόνα 30 Εφαρμογή on line personalization	105

ΕΙΣΑΓΩΓΗ

Η παρούσα μελέτη πραγματεύεται το ζήτημα των υπηρεσιών εξατομίκευσης στο ηλεκτρονικό εμπόριο. Η εργασία εστιάζει στην βελτιστοποίηση των στρατηγικών μάρκετινγκ που ακολουθούνται από τις επιχειρήσεις μέσω συστημάτων εξατομίκευσης των ηλεκτρονικών τους σελίδων. Στα πλαίσια δημιουργίας ανταγωνιστικών επιχειρήσεων σε παγκόσμια κλίμακα οι μικρομεσαίες επιχειρήσεις θα πρέπει να υιοθετήσουν σύγχρονα συστήματα μάρκετινγκ. Ο λόγος που καθιστά το θέμα μελέτης επίκαιρο είναι η ευκαιρία ανάπτυξης των επιχειρήσεων στον χώρο του ηλεκτρονικού επιχειρήν και οι υπηρεσίες εξατομίκευσης αποτελούν ένα σημαντικό εργαλείο που θα ενισχύσει τον ανταγωνισμό τους.

Για να αντισταθμιστούν οι επιπτώσεις της οικονομικής κρίσης, αλλά και για να αξιοποιηθούν νέες αναπτυξιακές δυνατότητες, οι επιχειρήσεις του κλάδου λιανικού εμπορίου, θα πρέπει να στραφούν στις τεχνολογίες πληροφορικής και επικοινωνιών ώστε να πετύχουν :

- Βελτίωση των εσωτερικών διαδικασιών, αυξάνοντας την παραγωγικότητα
- Ενίσχυση του κύκλου εργασιών του δεδομένου ότι απευθύνονται σε μεγαλύτερο πλήθος δυνητικών πελατών

Όσον αφορά το ελληνικό επιχειρείν οι οικονομικές κυρίως πιέσεις της αγοράς αναγκάζουν επιχειρήσεις του λιανικού εμπορίου να αναπτύξουν σχέδια για είσοδο στην ηλεκτρονική αγορά καθώς επίσης και στρατηγικές επέκτασης και επικράτησης εξαιτίας του ανταγωνισμού που έχει προκύψει από την παγκοσμιοποίηση.

Αναμφίβολα, η διοίκηση δεν μπορεί να λειτουργήσει στο σημερινό ανταγωνιστικό κλίμα χωρίς αξιόπιστη πληροφόρηση πάνω στην οποία θα στηριχτούν οι αποφάσεις της. Τα υπολογιστικά συστήματα αποτελούν την βάση μια επιχείρησης πάνω στην οποία θα χτίσει την στρατηγική προσέγγισης των πελατών της. Δεν νοείται στην σημερινή εποχή να επιχειρήσει κάποιος να ξεκινήσει τη λειτουργία ενός καταστήματος χωρίς προηγουμένως να έχει εξοπλιστεί με ένα καλό πληροφοριακό σύστημα, ηλεκτρονική προώθηση της επιχείρησης και στρατηγικές προώθησης της. Στην αγορά υπάρχει μια ποικιλία λύσεων που μπορεί να βοηθήσει τις επιχειρήσεις να ξεπεράσουν τις προκλήσεις που θέτει ο παγκόσμιος ανταγωνισμός. Στην παρούσα μελέτη εξετάζεται η εφαρμογή μιας λύσης που περιλαμβάνει τα εξής στάδια.

- Εφαρμογή συστήματος EPR για την διαχείριση όλων των τομέων της επιχείρησης (οικονομικός, παραγωγή, αποθήκη, κλπ)
- Κατασκευή ηλεκτρονικού καταστήματος για την προώθηση και ανάπτυξη της επιχείρησης πέραν από τα στενά όρια της ελληνικής αγοράς που ιδιαίτερα αυτή την εποχή πλήττεται από την έλλειψη ρευστότητας
- Χρήση στρατηγικών μεθόδων προώθησης ηλεκτρονικού προφίλ της εταιρείας και προσέγγισης του πελάτη (εξατομίκευση υπηρεσιών)

Η συντονισμένη εφαρμογή των ως άνω εργασιών, θα δώσουν ισχυρό προβάδισμα στην εταιρεία που θα το εφαρμόσει, έναντι του τοπικού ανταγωνισμού και θα την θέσει εντός «παιχνιδιού» σε παγκόσμιο επίπεδο. Ως μεθοδολογία εφαρμογής του εγχειρήματος αρχικώς θα πρέπει να εφαρμοστεί το πληροφοριακό σύστημα διαχείρισης της εταιρείας, ώστε να τοποθετηθούν τα θεμέλια μιας υγιούς ως προς το θέμα διοίκησης εταιρείας, δεδομένου ότι με την χρήση των πληροφοριακών συστημάτων ενισχύονται μια σειρά κατηγορίες ενεργειών που βοηθούν την επιχείρηση στην αποτελεσματικότερη λειτουργία της. Εν συνεχεία σαν ένα βήμα, θεωρείται η κατασκευή ενός ηλεκτρονικού καταστήματος που θα φέρει στοιχεία εξατομίκευσης.

i. Στόχος Μελέτης

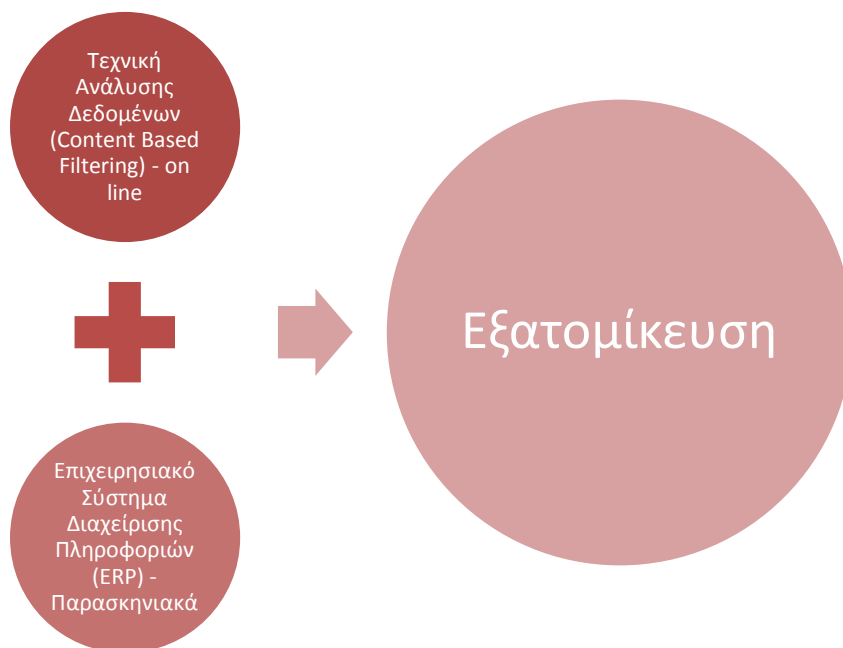
Ο κύριος στόχος της έρευνας σχετικά με την εξατομίκευση είναι ο προσδιορισμός των εργασιών που πρέπει να υλοποιούνται παρασκηνακά και on line ώστε να προκύπτουν σωστές προβλέψεις και γρήγορα αποτελέσματα για τον τελικό χρήστη.

ii. Μεθοδολογία Εργασίας

Η εργασία θα αναπτυχτεί σε τρία κεφάλαια. Στο πρώτο κεφάλαιο θα αναλυθεί η εξατομίκευση. Συγκεκριμένα θα παρουσιαστούν ο ορισμός, ο στόχος, η λειτουργία και οι περιορισμοί που θέτει η υπάρχουσα τεχνολογία για την ανάπτυξη εφαρμογών εξατομίκευσης. Επειδή για να υλοποιηθεί η εξατομίκευση θα πρέπει παρασκηνακά να λειτουργεί ένα σύστημα διαχείρισης πληροφοριών και σε on line επίπεδο απαιτείται μια τεχνολογία συλλογής και φιλτραρίσματος της πληροφορίας τα επόμενα δύο κεφάλαια εστιάζουν σε αυτές τις

εργασίες. Αναλυτικότερα στο δεύτερο κεφάλαιο της εργασίας αναλύεται η τεχνολογία των επιχειρησιακών συστημάτων ERP και των επιχειρηματικών διαδικασιών που υποστηρίζουν. Στην τρίτο κεφάλαιο μελετώνται τεχνικές συλλογής και ανάλυσης δεδομένων. Συγκεκριμένα αναλύονται τα Content-based Filtering και το Collaborative Filtering.

Στο σχήμα που ακολουθεί περιγράφεται η σχέση των τριών εννοιών – τεχνολογιών, διότι θα πρέπει να αποσαφηνιστεί πως η προσέγγιση της εξατομίκευσης σε αυτή την εργασία γίνεται από την ανάλυση των συστημάτων που την συστήνουν.



Εικόνα 1 Σχέσεις μεταξύ τεχνολογιών ERP – Content - Based Filtering & Personalization

iii. Οι Έννοιες Πληροφορία και Δεδομένο

Καθώς παρατηρείται οι έννοιες της πληροφορίας και των δεδομένων να συγχέονται μεταξύ τους γίνεται μια προσπάθεια αποσαφήνισης τους και διαχωρισμού τους.

Ως δεδομένα ορίζονται οποιαδήποτε ακατέργαστα γεγονότα. Πολλές επιχειρήσεις συλλέγουν μεγάλο όγκο δεδομένων, ωστόσο για να αποτελέσει ένα δεδομένο χρήσιμο για την επιχείρηση πρέπει να μετατραπεί σε πληροφορία διαφορετικά αν δεν μούν σε κάποιο πλαίσιο τείνουν να χαρακτηριστούν ως άχρηστα για την επιχείρηση. (Malaga, 2005)

Η διαφορά της πληροφορίας έναντι του δεδομένου έγκειται στο γεγονός ότι η πληροφορία είναι ένα αποτέλεσμα ενός επεξεργασμένου πρώην ακατέργαστου δεδομένου. Σύμφωνα με

τον Malaga (2005) οι πληροφορίες είναι ακατέργαστα δεδομένα μέσα σε ένα δεδομένο πλαίσιο. Στην προσπάθεια του να εξηγήσει με ποιο τρόπο μετατρέπεται ένα ακατέργαστο δεδομένο σε χρήσιμη πληροφορία παραθέτει ένα παράδειγμα. Λαμβάνει ως δεδομένα τη θερμοκρασία μιας περιοχής και το μετατρέπει σε πληροφορία βάζοντας το σε ένα συγκεκριμένο πλαίσιο. Σε ένα πλαίσιο παρατήρησης πως η χ θερμοκρασία που σημειώθηκε στην ψ περιοχή τον μήνα Ιούνιο στο μέσο της νύχτας θα μπορούσε να δώσει το συμπέρασμα ενός κύματος καύσωνα. (Τσόπογλου Σ., 2013)

1.1 ΕΞΑΤΟΜΙΚΕΥΣΗ ΣΤΟ ΗΛΕΚΤΡΟΝΙΚΟ ΕΜΠΟΡΙΟ

1.1 Ορισμός Ηλεκτρονικού Εμπορίου

Ως ηλεκτρονικό εμπόριο ορίζεται κάθε είδος εμπορικής συναλλαγής μεταξύ προσώπων (φυσικών και μη) που πραγματοποιείται με ηλεκτρονικά μέσα. Είναι η διάθεση και αγοραπωλησία προϊόντων ηλεκτρονικά, η διεκπεραίωση εμπορικών λειτουργιών και συναλλαγών χωρίς τη χρήση χαρτιού, συνήθως μέσω δικτύων ηλεκτρονικών υπολογιστών. Πρόκειται για την αγοραπωλησία αγαθών, πληροφοριών και υπηρεσιών μέσα από δίκτυα ηλεκτρονικών υπολογιστών.

Πίνακας 1 Το ηλεκτρονικό εμπόριο μπορεί να οριστεί από τέσσερις διαφορετικές οπτικές

Επιχειρήσεις	Ως εφαρμογή νέων τεχνολογιών προς την κατεύθυνση του αυτοματισμού των συναλλαγών και της ροής εργασιών.
Υπηρεσίες	Ως μηχανισμός που έχει στόχο να ικανοποιήσει την κοινή επιθυμία προμηθευτών και πελατών για καλύτερη ποιότητα υπηρεσιών, μεγαλύτερη ταχύτητα εκτέλεσης συναλλαγών και μικρότερο κόστος.
Απόσταση	Ως δυνατότητα αγοραπωλησίας προϊόντων και υπηρεσιών μέσω του Internet ανεξάρτητα από τη γεωγραφική απόσταση.
Επικοινωνία	Ως δυνατότητα παροχής πληροφοριών, προϊόντων ή υπηρεσιών, και πληρωμών μέσα από δίκτυα ηλεκτρονικών υπολογιστών.

1.2 Ορισμός Εξατομίκευσης

Η διαδικασία της εξατομίκευσης βρίσκει εφαρμογή στο ηλεκτρονικό εμπόριο μέσω προτάσεων για προϊόντα, εξατομικευμένη προώθηση, τιμολόγηση, διαμόρφωση της αρχικής σελίδας των ηλεκτρονικών καταστημάτων, κ.α. Ως εξατομίκευση ορίζεται

«Οποιαδήποτε ενέργεια που προσαρμόζει την πληροφορία ή τις υπηρεσίες που παρέχονται από ένα web site στη γνώση που κερδίζουμε από την πλοηγητική συμπεριφορά των χρηστών και τα ατομικά τους ενδιαφέροντα, σε συνδυασμό με το περιεχόμενο και τη δομή του web site.»

Ακόμα πιο συγκεκριμένα η έννοια της εξατομίκευσης για το ηλεκτρονικό εμπόριο¹ μπορεί να οριστεί ως

«Η διαδικασία της συγκέντρωσης και αποθήκευσης πληροφοριών αναφορικά με τους χρήστες ενός website, η ανάλυση των πληροφοριών αυτών και, με βάση την ανάλυση, η αποστολή σε κάθε χρήστη της σωστής πληροφορίας στο σωστό χρόνο.» (Μ. Ρήγκου, 2012)

Όλοι οι ορισμοί που συναντήθηκαν στη βιβλιογραφία, αντιλαμβάνονται την εξατομίκευση ως μια στενή σχέση μεταξύ του πελάτη και την προώθηση των προϊόντων, όχι όμως σε μαζικό επίπεδο αλλά σε σχέση ένας προς έναν.

Σε γενικές γραμμές, η προσωποποίηση αφορά τη δημιουργία μιας σχέσης που θα εμπνέει εμπιστοσύνη στον εκάστοτε «πελάτη» (Manber U., 2000), (Mobasher B., 2000). Στη συνέχεια, μέσα από αυτή τη σχέση και κάτω από τις υπάρχουσες συνθήκες, επιχειρείται να κατανοηθούν και να ικανοποιηθούν οι επιθυμίες του «πελάτη». Συγκεκριμένα στον χώρο του Διαδικτύου, η έννοια της προσωποποίησης περιλαμβάνει την μοντελοποίηση διαδικτυακών αντικειμένων (προϊόντα ή ιστοσελίδες) και υποκειμένων (χρήστες), την κατηγοριοποίηση τους, την εύρεση ομοιοτήτων ανάμεσα τους και τον καθορισμό των απαραίτητων ενεργειών που τελικά θα προταθούν. Η προσωποποίηση στο ηλεκτρονικό εμπόριο (Manber U., 2000) έχει ως σκοπό να αναγνωρίσει τις προτιμήσεις του υποψήφιου αγοραστή και να εντοπίσει τις

¹ Πέραν του ηλεκτρονικού εμπορίου άλλα πεδία εφαρμογής της τεχνολογίας είναι οι πληροφοριακές πύλες, τα περιβάλλοντα ηλεκτρονικής μάθησης και οι μηχανές αναζήτησης (Μ. Ρήγκου, 2012)

πιθανές του ελλείψεις ώστε να του παρουσιάσει τελικά εκείνα τα προϊόντα που είναι πιθανότερο να αγοράσει. Κατά ανάλογο τρόπο, η προσωποποίηση στην ηλεκτρονική διακυβέρνηση θέτει ως στόχο την εξατομικευμένη εξυπηρέτηση του χρήστη στις δημόσιες υπηρεσίες με βάση τα χαρακτηριστικά του (π.χ. οικογενειακή κατάσταση, εισόδημα κλπ) και το νομοθετικό πλαίσιο που διέπει τις υπηρεσίες που συνθέτουν την ηλεκτρονική διακυβέρνηση. Όπως προκύπτει από τον τρόπο λειτουργίας των συστημάτων συστάσεων, ο τρόπος λειτουργίας των υπηρεσιών της ηλεκτρονικής διακυβέρνησης είναι ανάλογος με τον τρόπο λειτουργίας των Συστημάτων Συστάσεων γνώσης, αφού περιλαμβάνουν τον διάλογο του χρήστη με το σύστημα και αποσκοπούν στην ανάκτηση συγκεκριμένων πληροφοριών που ενδιαφέρουν τον χρήστη.

1.3 Στόχος Εξατομίκευσης

Το μέλλον βρίσκεται στην έννοια της ηλεκτρονικής επιχείρησης (σε αντίθεση με το "απλό" μοντέλο του ηλεκτρονικού καταστήματος) όπου οι συναλλαγές και οι διεργασίες της επιχείρησης μεταβάλλονται ώστε να γίνονται με ηλεκτρονικά μέσα. Μια εκ των στρατηγικών μεθόδων μάρκετινγκ αποτελεί η εξατομίκευση. Οι υπηρεσίες εξατομίκευσης στο ηλεκτρονικό εμπόριο είναι ένας τομέας που συνδυάζει την επιστήμη του μάρκετινγκ και του προγραμματισμού και παρουσιάζει μεγάλη ανάπτυξη την τελευταία πενταετία.

Η εξατομίκευση εστιάζει στην ικανοποίηση των ειδικών – προσωπικών αναγκών του κάθε ξεχωριστού πελάτη. Το ηλεκτρονικό εμπόριο είναι μια τεχνική βοήθεια συναλλακτικών δραστηριοτήτων, (αγορά – πώληση προϊόντων και υπηρεσιών).² Μεταξύ των δύο εννοιών υπάρχει ένας κοινός τομέας – η επιτυχημένη προσέγγιση του πελάτη. (Turban, 2000)

Η χρήση υπηρεσιών εξατομίκευσης σε ένα ηλεκτρονικό κατάστημα στοχεύει στην παρουσίαση

- Στο χρήστη περιεχόμενο (διαφήμιση, πληροφορία) που τον ενδιαφέρει
- Το περιεχόμενο αυτό τον ενδιαφέρει τόσο ώστε το session να διαρκέσει 1 click περισσότερο
- Όσες περισσότερες φορές κάνει click ο χρήστης τόσο περισσότερο διαρκεί το μέσο session

² Να σημειωθεί πως η έννοια του ηλεκτρονικού εμπορίου περιλαμβάνει επίσης και τις μηχανές αναζήτησης και τις σελίδες σύγκρισης υπηρεσιών και προϊόντων.

- Τα μεγαλύτερα sessions υπονοούν πιο ικανοποιημένους πελάτες. Οι ικανοποιημένοι πελάτες εξυπηρετούν τους επιχειρηματικούς στόχους.

1.4 Διαδικασία Εξατομίκευσης

Η διαδικασία εξατομίκευσης αποτελείται από τα ακόλουθα στάδια:

- Συλλογή Δεδομένων (Data Collection)
- Προεπεξεργασία δεδομένων (Data Preprocessing)
- Ανακάλυψη Προτύπων (Pattern Discovery)
- Μεταεπεξεργασία Γνώσης (Knowledge Postprocessing)

1.4.1 Συλλογή Δεδομένων (Data Collection)

Αρχικώς οι εξατομικευμένες υπηρεσίες εφαρμόστηκαν στο διαδίκτυο για να απασχολήσουν όσο το δυνατόν περισσότερο τον χρήστη στο δίκτυο. Πλέον στοχεύουν στην ταχύτερη και ακριβέστερη παράδοση των πληροφοριών στον χρήστη με στόχο το κέρδος. Σε πρώτο στάδιο υλοποιείται η συλλογή δεδομένων. Η συλλογή των δεδομένων πραγματοποιείται από τους εξυπηρετητές (servers), μέσω π.χ. των server log files, των cookies, κ.ά., τους πελάτες (clients) που συνδέονται σε αυτούς, μέσω π.χ. πρακτόρων (agents) ενσωματωμένων στις ιστοσελίδες, είτε ενδιάμεσες πηγές, όπως ενδιάμεσοι εξυπηρετητές, προγραμμάτων παρακολούθησης δικτύου (π.χ. packet sniffers).

The screenshot shows a Facebook profile completion interface. At the top, it says 'Ολοκληρώστε το προφίλ σας: Τοποθεσίες · 2/2' with a progress bar. Below this, the question 'Σε ποια πόλη ζείτε;' is displayed. Three location suggestions are listed, each with a radio button: 'Brussels, Belgium' (with a note about Sarah J. Phoenix and 13 friends), 'Melbourne, Victoria, Australia' (with a note about Suzie Nguyen and Greg Polymerou), and 'Istanbul, Turkey' (with a note about Selma Özkan, Filiz Ivriz, and Deger Cil). At the bottom, there is a text input field for 'Προσθέστε τον τόπο διαμονής', a 'Φίλοι' dropdown menu, and a 'Παράβλεψη' button.

Εικόνα 2 Η διαδικασία συλλογής πληροφοριών του social Network Facebook βασίζεται στην τεχνική explicit profiling

Οι πλέον διαδεδομένες τεχνικές είναι οι Explicit profiling, Implicit Profiling, Legacy data. Στον πίνακα που ακολουθεί παρουσιάζονται τα χαρακτηριστικά της κάθε μίας.

Πίνακας 2 Τεχνικές συλλογής και ανάλυσης δεδομένων

Explicit profiling	Ο χρήστης καλείται να δηλώσει ρητά πληροφορίες για τον ίδιο και τις ανάγκες/προτιμήσεις του (π.χ. να συμπληρώσει κάποιο ερωτηματολόγιο). Το πλεονέκτημα αυτής της μεθόδου είναι ότι αφήνει τους χρήστες να υπαγορεύσουν μόνοι τους τί τους ενδιαφέρει.
Implicit profiling	Καταγράφεται η συμπεριφορά του επισκέπτη (διαφανής διαδικασία). Συνήθως αυτό που καταγράφεται και αναλύεται είναι το browsing pattern (cookies) και το buying pattern (site database)
Legacy data	Πιστωτικές πληροφορίες (στοιχεία κίνησης πιστωτικών καρτών) ή προηγούμενες αγορές. Για παλιότερους πελάτες τα legacy data αποτελούν την πιο πλούσια (και αξιόπιστη) πηγή πληροφοριών για την κατασκευή προφίλ

1.4.2 Προεπεξεργασία δεδομένων (Data Preprocessing)

Τα δεδομένα τα οποία έχουν συγκεντρωθεί στο προηγούμενο στάδιο μπορεί να περιέχουν πλεονάζουσες και περιττές πληροφορίες. Για αυτό το λόγο πρέπει να υποστούν κατάλληλη επεξεργασία και να κωδικοποιηθούν για να χρησιμοποιηθούν στο επόμενο στάδιο. Η προεπεξεργασία -όπως ονομάζεται- των δεδομένων περιλαμβάνει τα στάδια του φιλτραρίσματος των δεδομένων (data filtering), της αναγνώρισης του χρήστη που επισκέπτεται την ιστοσελίδα (user identification) και της αναγνώρισης της συνόδου του χρήστη (user session identification), δηλαδή ένα πεπερασμένο σύνολο σελίδων που επισκέφτηκε ο χρήστης κατά τη διάρκεια μιας επίσκεψης σε μια ιστοσελίδα.

Τα πλέον σημαντικά συστήματα φιλτραρίσματος και ανάλυσης είναι τα collaborative filtering systems και τα content-based filtering. Το φιλτράρισμα συνεργασίας (collaborative filtering) είναι ίσως η πιο διαδεδομένη και αποδοτική μέθοδος σύστασης στο διαδίκτυο. Δεν είναι τυχαίο ότι ιστοσελίδες που βασίζονται στην σύσταση (όπως οι Amazon, iTunes, Netflix, LastFm, Delicious, Pandora) χρησιμοποιούν την μέθοδο του φιλτραρίσματος συνεργασίας. Τα δεδομένα που επεξεργάζεται αυτή η μέθοδος είναι κατά κύριο λόγο οι αξιολογήσεις αντικειμένων από τον χρήστη. Η αξιολόγηση μπορεί να γίνεται με άμεσο ή με έμμεσο τρόπο. Άμεση αξιολόγηση είναι το «Rating», δηλαδή η βαθμολόγηση του αντικειμένου από τον χρήστη σε μια κλίμακα βαθμολόγησης. Η έμμεση αξιολόγηση τώρα βασίζεται κυρίως πάνω στις κινήσεις του χρήστη, κάποιες κινήσεις μπορεί να δείχνουν ότι κάτι τον έχει τραβήξει το ενδιαφέρον. Δηλαδή με άλλα λόγια όταν ο χρήστης κάνει κλικ πάνω σε ένα αντικείμενο, μετά την σύντομη περιγραφή που διάβασε στον κατάλογο των αντικειμένων, αυτό δείχνει ότι το συγκεκριμένο αντικείμενο του τράβηξε την προσοχή οπότε, θεωρητικά, το αξιολογεί καλώς. Επίσης, η έμμεση αξιολόγηση του αντικειμένου αυξάνεται αν ο χρήστης παραμείνει μεγάλο χρονικό διάστημα στην σελίδα αυτού του αντικειμένου ή αν ανεβάσει το link του αντικειμένου σε κάποιο χώρο κοινωνικής δικτύωσης. Πάντως, ανάμεσα στις δύο μεθόδους αξιολόγησης, πιο αποτελεσματική είναι η άμεση λόγω του ότι δείχνει ξεκάθαρα τις προτιμήσεις του χρήστη.

Η σύσταση βασισμένη στο περιεχόμενο (content-based filtering) είναι άλλη μια μέθοδος φιλτραρίσματος των δεδομένων. Όπως και στο CF, έτσι και εδώ, βασικός παράγοντας για την σύσταση είναι οι αξιολογήσεις του χρήστη στο αντικείμενο με άμεσο ή έμμεσο τρόπο. Η διαφορά είναι ότι ενώ στο φιλτράρισμα συνεργασίας βασιζόμαστε στην υπόθεση ότι “άνθρωποι με παρόμοιες εκτιμήσεις έχουν και παρόμοια γούστα”. Στο φιλτράρισμα περιεχομένου γίνεται η υπόθεση ότι “αντικείμενα με παρόμοια χαρακτηριστικά θα εκτιμηθούν με παρόμοιο τρόπο”. Με άλλα λόγια, σε αυτήν την μέθοδο οι συγκρίσεις δεν γίνονται ανάμεσα σε χρήστες αλλά ανάμεσα σε αντικείμενα. Για να συμβεί αυτό, πρέπει η ιστοσελίδα να διαθέτει λεπτομερείς πληροφορίες-χαρακτηριστικά για κάθε αντικείμενο που εμπορεύεται και το οποίο θέλει να προτείνει στον χρήστη. Αυτό είναι απαραίτητο γιατί την στιγμή που ο χρήστης θα αξιολογήσει ένα αντικείμενο με καλή βαθμολογία, το σύστημα θα πρέπει να ερευνήσει για αντικείμενα με παρόμοια χαρακτηριστικά ώστε να τα προτείνει. Τα χαρακτηριστικά του κάθε αντικειμένου δεν έχουν την ίδια βαρύτητα στην πραγματοποίηση της σύστασης. Αυτό σημαίνει ότι μια απλή ομοιότητα ανάμεσα σε δυο αντικείμενα δεν είναι πάντα λόγος για να πραγματοποιηθεί σύσταση.

Στον πίνακα που ακολουθεί παρουσιάζονται τα χαρακτηριστικά της κάθε μίας. (Δ. Παρούτσας, 2013)

Πίνακας 3 Συστήματα φιλτραρίσματος

Collaborative filtering systems	<p>Το collaborative filtering απαιτεί μια μεγάλη βάση πελατών ώστε να είναι δυνατό να εντοπιστεί για κάθε χρήστη μια ομάδα ομοϊδεατών. Όσο ο αριθμός των χρηστών είναι μικρός, η ποιότητα και η ακρίβεια των υπολογισμών είναι χαμηλή (cold-start)</p> <p>Κάθε φορά που εισάγεται ένα νέο προϊόν για το οποίο όπως είναι φυσικό δεν υπάρχουν βαθμολογήσεις από χρήστες, είναι δύσκολο να καταλήξει να προταθεί.</p>
Content-based filtering	<p>Το content-based filtering είναι κατάλληλο για τις περιπτώσεις που τα προϊόντα μπορούν εύκολα να αναλυθούν από ένα υπολογιστικό σύστημα, η επιλογή (ή η αξιολόγηση) ενός προϊόντος δεν επηρεάζεται σημαντικά από υποκειμενικούς παράγοντες.</p>

1.4.3 Ανακάλυψη Προτύπων (Pattern Discovery)

Σε αυτό το στάδιο εφαρμόζονται στα επεξεργασμένα δεδομένα που έχουν προκύψει από το προηγούμενο στάδιο οι τεχνικές εξόρυξης γνώσης. Οι τεχνικές που χρησιμοποιούνται είναι η συσταδοποίηση (clustering), η ταξινόμηση (classification), η ανακάλυψη συσχετίσεων (association discovery) και η ανακάλυψη ακολουθιακών προτύπων (sequential pattern discovery). Αποτέλεσμα της διαδικασίας είναι η κατασκευή προτύπων που αντανακλούν τα συμπεριφορά και τα ενδιαφέροντα των χρηστών της ιστοσελίδας.

Στην περιγραφή της διαδικασίας του online personalization χρησιμοποιούνται οι έννοιες του Support και του Confidence, οι οποίες αποτελούν μεγέθη που εκφράζουν πιθανότητες και οι ορισμοί τους δίνονται παρακάτω:

Support: Κατά την αναζήτηση δημοφιλών navigation patterns/subpatterns στις πλοηγήσεις των χρηστών της πύλης, εισάγουμε ένα ελάχιστο support που πρέπει να εμφανίζει κάθε pattern. Το support υποδηλώνει την πιθανότητα εμφάνισης του κάθε subpattern σαν τμήμα ενός άλλου pattern μεγαλύτερου μήκους και παίρνει ξεχωριστή τιμή για κάθε ομάδα χρηστών. Δηλαδή, για κάθε navigation pattern S, αν P είναι το σύνολο των navigation patterns που έχει πραγματοποιήσει ένα group χρηστών, το support του S για το συγκεκριμένο group είναι:

$$\sigma(S) = \frac{\{\text{nav. pattern } p \in P : S \text{ is subsequence of } p\}}{P}$$

Confidence: Για τη δημιουργία μίας συντόμευσης κατά την πλοήγηση του χρήστη, εξετάζεται ποια δημοφιλή navigation patterns μήκους k+1, όπου k το μήκος του navigation window w, περιέχουν το παράθυρο πλοήγησης και μία επόμενη θεματική κατηγορία X. Η κατηγορία X εμφανίζεται ως συντόμευση στο χρήστη μόνο εάν το confidence που της αντιστοιχεί είναι μεγαλύτερο από ένα κάτω όριο. Το confidence ορίζεται ως η πιθανότητα (εκφρασμένη ως support) εμφάνισης του pattern wo{X} προς την πιθανότητα εμφάνισης του w στα αποθηκευμένα δημοφιλή patterns. Δηλαδή:

$$\alpha(X) = \frac{\sigma(wo\{X\})}{\sigma(w)}$$

Σειριακά πρότυπα είναι εκείνες οι ακολουθίες αντικειμένων που εντοπίζονται συχνά σε ένα επαρκώς μεγάλο ποσοστό από συναλλαγές. Τα σειριακά πρότυπα (sequential patterns) που εντοπίζονται σε δεδομένα χρήσης web αναπαριστούν τα μονοπάτια από ιστοσελίδες που ακολουθούνται συχνά από τους χρήστες, με τη σειρά με την οποία

‘περπατούν’ από σελίδα σε σελίδα (Mobasher, 2004). Μια ακολουθία $\langle s_1, s_2, \dots, s_n \rangle$ συμβαίνει σε μια συναλλαγή $t = \langle p_1, p_2, \dots, p_m \rangle$, όπου $n \leq m$, αν υπάρχουν n θετικοί ακέραιοι a_i $1 \leq i \leq n$ και $s_i = p_{i+a_i}$, $\forall i$.

Επιπλέον, ορίζεται ότι το $\langle cs_1, cs_2, \dots, cs_n \rangle$ είναι μια συνεχής ακολουθία στην t αν υπάρχει ένας ακέραιος b, τέτοιος ώστε $0 \leq b \leq m - n$, και $\forall i = 1 + cs_i$ έως n. Σε ένα συνεχές σειριακό πρότυπο, κάθε ζευγάρι συνεχόμενων στοιχείων s_i και s_{i+1} πρέπει να εμφανίζεται συνεχόμενα σε μια συναλλαγή t που θεωρείται ότι υποστηρίζει το πρότυπο, ενώ ένα σειριακό πρότυπο μπορεί να αναπαριστά μη-συνεχόμενες συχνές ακολουθίες στο σύνολο των συναλλαγών.

Με δεδομένο ένα σύνολο συναλλαγών T και ένα σύνολο $\{S_1, S_2, \dots, S_n\}$ από σειριακά (ή συνεχόμενα σειριακά, αντίστοιχα) πρότυπα που εμφανίζονται συχνά στο T , το *support* καθενός από τα S_i υπολογίζεται ως εξής:

$$\sigma(S_i) = \frac{|\{t \in T : S_i \text{ μια (συνεχής) υποακολουθία του } t\}|}{|T|}$$

Το confidence του κανόνα $X \rightarrow Y$, όπου τα X και Y είναι συνεχή σειριακά πρότυπα ορίζεται ως:

$$a(X \Rightarrow Y) = \frac{\sigma(X \circ Y)}{\sigma(X)}$$

όπου το \circ είναι ο τελεστής παράθεσης ή συνένωσης (concatenation operator). Αξίζει να σημειωθεί ότι οι τιμές κατωφλίου για το support τόσο των σειριακών προτύπων, όσο και των συνεχών σειριακών προτύπων, ικανοποιούν το κριτήριο προς τα κάτω κλειστότητας (*downward closure*), δηλαδή, εάν μια (συνεχής) ακολουθία στοιχείων S έχει μια οποιαδήποτε υποακολουθία που δεν ικανοποιεί το κριτήριο ελάχιστου support, τότε και η S δεν έχει το απαιτούμενο ελάχιστο support.

Για τον εντοπισμό σειριακών και συνεχών σειριακών προτύπων μπορεί να χρησιμοποιηθεί ο αλγόριθμος Apriori (Agrawal et al., 1993) που τυπικά χρησιμοποιείται για τον εντοπισμό frequent itemsets (σύνολα στοιχείων που εμφανίζονται μαζί με μεγάλη συχνότητα) και την εξόρυξη κανόνων συσχέτισης. Αυτό επιτυγχάνεται συνήθως μεταβάλλοντας τον τρόπο υπολογισμού του support, ώστε να βασίζεται στη συχνότητα εμφάνισης υποακολουθιών στοιχείων, αντί για υποσύνολα στοιχείων (Agrawal & Srikant, 1995), ώστε να λαμβάνεται υπόψη και ο παράγοντας της χρονικής αλληλουχίας εμφάνισης.

Η αναπαράσταση των συναλλαγών στο web με τη μορφή ακολουθιών από pageviews, επιτρέπει τον εντοπισμό και την ανάλυση σειριακών προτύπων πλοήγησης χρησιμοποιώντας ένα πλήθος από μοντέλα που έχουν μελετηθεί διεξοδικά σε συναφείς επιστημονικές περιοχές. Ένας όρος ισοδύναμος του όρου ακολουθία³, που προέρχεται από την περιοχή της στατιστικής επεξεργασίας φυσικής γλώσσας (statistical natural language processing) είναι το n gram. Με τον όρο n -gram (n -γραμμα) ορίζουμε μία ακολουθία από n οντότητες με κοινά χαρακτηριστικά και ιδιότητες. Στο πεδίο του web mining, τα n -grams αναπαριστούν αιτήσεις

³ Τυπικά, ο όρος ακολουθία χρησιμοποιείται για να εκφράσει ένα διατεταγμένο σύνολο από πράξεις, που στο χώρο που εξετάζουμε είναι αιτήσεις εξυπηρέτησης.

χρηστών για πόρους στο web (ιστοσελίδες ή άλλου τύπου αρχεία), που παράγονται στη χρονική διάρκεια ενός συγκεκριμένου user session. Για παράδειγμα, το ταξινομημένο ζευγάρι (A, B) αποτελεί ένα 2-gram, που δηλώνει το γεγονός ότι από ένα συγκεκριμένο χρήστη και στη διάρκεια της ίδιας συνόδου, ζητήθηκε πρώτο το pageview A και στη συνέχεια το B. Συνεπώς, ένα n -gram παρέχει ένα μοντέλο αναπαράστασης n συνεχόμενων αιτήσεων που προέρχονται από ένα συγκεκριμένο χρήστη κατά τη διάρκεια μιας συνόδου. Οι Frias-Martinez και Karamcheti (2002) ορίζουν ότι για το πεδίο του web mining, ένα n -gram μιας συνόδου S_i είναι οποιοδήποτε υποσύνολο από n συνεχόμενα URLs (στην ουσία, pageviews) στο S_i . Η πρόβλεψη που βασίζεται σε n -grams, πραγματοποιείται μέσω της αντιστοίχισης (matching) ενός μονοπατιού πλοήγησης –ή μέρους αυτού- με το πρόθεμα ενός από τα ήδη καταγεγραμμένα n -grams.

Στη γενική περίπτωση διακρίνονται δύο τύποι μοντέλων n -gram (Li, 2001; Su et al., 2000):

- μοντέλα βασισμένα σε σημείο (point-based)
- μοντέλα βασισμένα σε μονοπάτι (path-based)

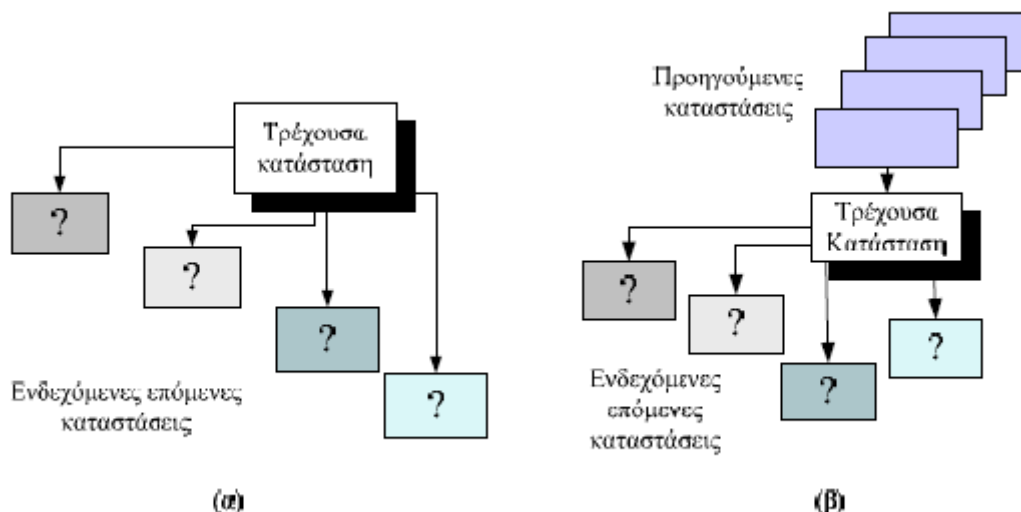
Τα μοντέλα που βασίζονται σε σημείο, χρησιμοποιούν n grams μοναδιαίου μήκους (1-gram) για την κατασκευή του πλαισίου πρόβλεψης, με βάση ένα σύνολο από ενέργειες ταξινομημένες χρονικά. Με τη χρήση των συγκεκριμένων μοντέλων, η πρόβλεψη της επόμενης αίτησης καθορίζεται αυστηρά από την τελευταία καταγεγραμμένη αίτηση που απέστειλε ο συγκεκριμένος χρήστη. Λόγω όμως του ότι τα μοντέλα αυτά βασίζονται σε αρκετά περιορισμένη πληροφορία, πολύ συχνά οδηγούν σε εσφαλμένα συμπεράσματα (και ανακριβείς προβλέψεις). Χαρακτηριστικά αναφέρεται ότι τα μοντέλα που βασίζονται σε σημείο δεν έχουν υπερβεί σε ακρίβεια το 30%. Τα n -grams που βασίζονται σε μονοπάτια λαμβάνουν υπόψη τους για την πρόβλεψη τις n πιο πρόσφατες αιτήσεις που έχουν σταλεί από ένα συγκεκριμένο χρήστη, πετυχαίνοντας καλύτερα αποτελέσματα από άποψη ακρίβειας. Το μειονέκτημα των μοντέλων που βασίζονται σε μονοπάτι είναι η περιορισμένη εφαρμοσιμότητα (applicability), καθώς οι ακολουθίες μεγάλου μήκους επαναλαμβάνονται με πολύ χαμηλότερη συχνότητα.

Για την αναπαράσταση των n -grams, έχει προταθεί η χρήση επιπρόσθετων ειδικών χαρακτήρων (wildcard), ώστε να είναι υπάρχει δυνατότητα γενίκευσης των προτύπων περιήγησης. Πιο συγκεκριμένα, οι Spilioroulou et al. (1999), εισήγαγαν την έννοια της ακολουθίας g , η οποία προσδιορίζει ένα διάνυσμα γεγονότων (αιτήσεων ιστοσελίδων) και

ειδικών χαρακτήρων, όπου για παράδειγμα, με τη μορφή A^*D , αναπαριστώνται όλες τις ακολουθίες που περιέχουν τα γεγονότα A και D , με τη συγκεκριμένη σειρά εμφάνισης, ενώ το $A^*[1;2]D$, αναπαριστά όλες τις ακολουθίες στις οποίες εμφανίζονται ένα ή δύο το πολύ γεγονότα μεταξύ των A και D . Με αυτόν τον τρόπο αναπαράστασης είναι δυνατό να περιγραφούν και τα *frequent itemsets* (συχνά εμφανιζόμενα σύνολα αντικειμένων), που στο πεδίο του web mining εντοπίζονται με βάση μια συγκεκριμένη τιμή κατωφλίου στη συχνότητα εμφάνισης σε ένα δεδομένο αρχικό σύνολο (υπό)συνόλων (support threshold).

Στη σχετική βιβλιογραφία έχουν καταγραφεί πλήθος περιπτώσεων μεθόδων πρόβλεψης βασιμμένων σε μοντέλα n -gram, με ικανοποιητικά επίπεδα ακρίβειας αλλά και εφαρμοσιμότητας. Οι Frias-Martinez & Karamcheti (2002), πρότειναν το μοντέλο

κατασκευής ακολουθιακών κανόνων συσχέτισης (sequential association rules ή SAR), με στόχο τον προσδιορισμό της σειράς με την οποία προσπελούνται οι σελίδες ενός site, αλλά και τη χρονική διάρκεια παραμονής των χρηστών σε κάθε μία από αυτές. Από τους κανόνες συσχέτισης αυτού του είδους προκύπτει πληροφορία σχετική με την απόσταση μεταξύ των σελίδων που προηγούνται της πιο πρόσφατης αίτησης και αυτών που ακολουθούν και υπολογίζεται με βάση τον αριθμό των clicks του χρήστη κατά την περιήγησή του από τη μία ιστοσελίδα στην άλλη. Η έννοια της απόστασης επιτρέπει τον προσδιορισμό του χρονικού διαστήματος κατά το οποίο αναμένεται να προσπελαστεί μια σελίδα, πληροφορία ιδιαίτερα χρήσιμη τόσο για εφαρμογές prefetching, όσο και για συστήματα παραγωγής συστάσεων. Ανάλογα με το μοντέλο που προτείνεται, η πρόβλεψη περιλαμβάνει εκτός από το αρχείο που αναμένεται να ζητηθεί, και ένα δείκτη της εκτιμώμενης απόστασης του αρχείου αυτού από την τελευταία καταγεγραμμένη αίτηση του συγκεκριμένου χρήστη. Σημειώνεται ότι και σε αυτή την περίπτωση, λαμβάνονται υπόψη μόνο οι κανόνες συσχέτισης που πληρούν συγκεκριμένους περιορισμούς σχετικά με τις τιμές των παραμέτρων support και confidence.



Εικόνα 3 n-Gram βασισμένο σε σημείο (α) και n-gram βασισμένο σε μονοπάτι (β)

Οι Su et al. (2000) προτείνουν αλγόριθμους οι οποίοι βασίζονται σε n -grams, με στόχο την πρόβλεψη της επόμενης αίτησης ενός χρήστη βάσει των περιηγήσεων που έχουν καταγραφεί στο αρχείο log. Επιπλέον προτείνουν τη διαγραφή των συνόδων μήκους μικρότερο ή ίσο του 3 από το σύνολο των δεδομένων, θεωρώντας ότι τέτοιου μήκους ακολουθίες είναι εξαιρετικά δύσκολο να οδηγήσουν σε πρόβλεψη με αποδεκτά επίπεδα ακρίβειας. Στη μεθοδολογία που προτείνουν, χρησιμοποιούν n -gram μήκους 2 και 3, καθώς από πειραματικές μετρήσεις που παρουσιάζουν προκύπτει ότι για μεγαλύτερο μήκος n -gram το κόστος σε εφαρμοσιμότητα είναι δυσανάλογο σε σχέση με το κέρδος σε ακρίβεια. Για κάθε πρόβλεψη, ελέγχεται αρχικά το σύνολο των 3-gram και, σε περίπτωση όπου δεν υπάρχει κάποιο αντίστοιχο με τις τελευταίες τρεις αιτήσεις του χρήστη, ελέγχεται το σύνολο των 2-gram έχοντας πλέον ως κριτήριο σύγκρισης τις τελευταίες δύο αιτήσεις του χρήστη. Η ακρίβεια που επιτυγχάνει αυτό το υβριδικό μοντέλο είναι ελαφρά χειρότερη από αυτήν που επιτυγχάνει ένα μοντέλο το οποίο βασίζεται σε 4-gram, ενώ η εφαρμοσιμότητα είναι σαφώς μεγαλύτερη. Η συγκεκριμένη προσέγγιση υιοθετήθηκε ως ένα βαθμό και στο προτεινόμενο σχήμα πρόβλεψης, αλλά επεκτάθηκε ώστε να υποστηρίζει πλήρη κάλυψη (Oikonomopoulou et al., 2004; Oikonomopoulou et al., 2005).

1.4.4 Μεταεπεξεργασία Γνώσης (Knowledge Postprocessing)

Σε αυτό το στάδιο η γνώση που έχει προκύψει σχετικά με τη συμπεριφορά των χρηστών της πύλης αξιολογείται από το διαχειριστή της πύλης, ώστε να ληφθούν αποφάσεις για την τροποποίηση της δομής ή του περιεχομένου της πύλης. Η εξατομίκευση μπορεί να γίνει σε δύο επίπεδα:

Σε ατομικό επίπεδο (προσαρμογή της πύλης στα ενδιαφέροντα του κάθε χρήστη ξεχωριστά).

Σε επίπεδο ομάδων χρηστών με κοινά ενδιαφέροντα.

Αν και η πρώτη περίπτωση μπορεί να έχει καλύτερα αποτελέσματα για τις πλοηγήσεις του κάθε χρήστη ξεχωριστά, είναι αρκετά δύσκολο να υλοποιηθεί, καθώς θα πρέπει να δημιουργηθούν εκατομμύρια διαφορετικές μορφές παρουσίασης της πύλης, κάτι που σημαίνει τεράστιο όγκο δεδομένων και σημαντική απώλεια στην ταχύτητα εμφάνισης και λειτουργίας της. Αντίθετα, όταν η εξατομίκευση γίνεται σε επίπεδο ομάδων χρηστών, η ακρίβεια στην προσαρμογή της πύλης στις προτιμήσεις του κάθε χρήστη είναι μειωμένη, ωστόσο η λειτουργία της δεν επιβαρύνεται σημαντικά. (Ellsworth J. & Ellsworth W., 1997)

Η εξατομίκευση μπορεί να αφορά λειτουργίες που επηρεάζουν σε πραγματικό χρόνο (κατά τη διάρκεια της πλοήγησης των χρηστών) τη μορφή της πύλης για κάθε ομάδα χρηστών, αλλά και στατικές επεμβάσεις που πραγματοποιούνται από το διαχειριστή της, μετά από τη συλλογή και ανάλυση των πλοηγήσεων.

1.5 Κατάλληλα Sites για Εξατομίκευση

Η εφαρμογή συστημάτων εξατομίκευσης δεν ενδείκνυται για κάθε ηλεκτρονική σελίδα. Οι κατηγορίες σελίδων που επωφελούνται από την χρήση της είναι:

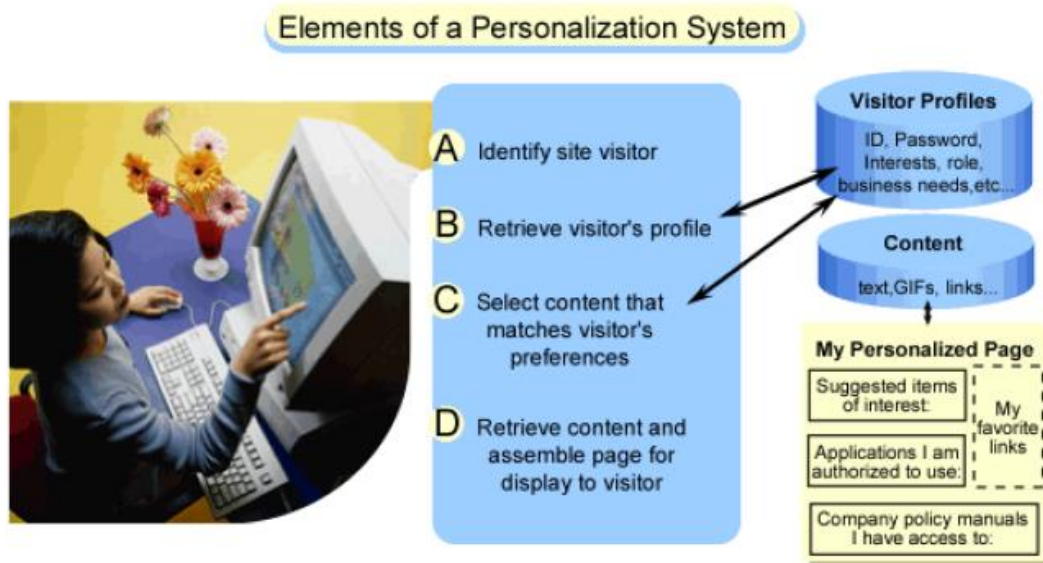
- περιεχομένου με μεγάλο όγκο πληροφοριών (όπως πληροφορίες τεχνολογικής, οικονομικής ή ενημερωτικής φύσης)
- Πώλησης μεγάλου αριθμού ομοειδών προϊόντων (ταινίες, βιβλία, CDs) ή προϊόντα ίδιου είδους αλλά από διαφορετικές
- Εταιρίες που θέλουν να συλλέξουν και να χρησιμοποιήσουν τη γνώμη των χρηστών για τα προϊόντα

- Περιεχομένου που είναι εστιασμένα, έχουν απλή δομή και περιέχουν περιορισμένο και συγκεκριμένο περιεχόμενο (π.χ. το site ενός τοπικού αεροδρομίου)
- Ηλεκτρονικού εμπορίου με λίγα προϊόντα (π.χ. με καφέ από μια συγκεκριμένη μάρκα, ή ρούχα ενός συγκεκριμένου σχεδιαστή)

1.6 Στοιχεία ενός Συστήματος Εξατομίκευσης

Η εξατομίκευση ξεκινάει μετά τον προσδιορισμό του χρήστη, (π.χ. από πληροφορίες που ζητούνται από το σύστημα κατά την είσοδο «login» στην σελίδα). Η διαδικασία που ακολουθείται είναι ευαίσθητη, καθώς μπορεί να προκύψουν λάθος συμπεράσματα και απαιτούν αρκετές πληροφορίες που αν αφορούν τις συνήθειες του πελάτη. Υπάρχουν πλήθος διαδικασιών όπου οι χειριστές των ηλεκτρονικών καταστημάτων μπορούν να αντλήσουν πληροφορίες σχετικά με τις συνήθειες του πελάτη :

- Με αλληλεπίδραση με το Δίκτυο - Ροή επιλογών
- Από τις συναλλαγές – αγοραστικές προτιμήσεις - Είτε ρητά με ερώτηση
- Ανάλογα με τις προτιμήσεις τους
- Βασιζόμενη σε ηλεκτρονικά ερωτηματολόγια
- Χρησιμοποιώντας τη συχνότητα και την ώρα επίσκεψης στην σελίδα



Εικόνα 4 Λειτουργίες – στάδια εξατομίκευσης

Πίνακας 4 Οι τέσσερις λειτουργίες της εξατομίκευσης

Απομνημόνευση	Καταγράφει και αποθηκεύει πληροφορίες για το χρήστη και τις χρησιμοποιεί για να ανακαλέσει τη συμπεριφορά του Χαιρετισμός χρήστη Bookmarking Εξατομικευμένα δικαιώματα πρόσβασης
Καθοδήγηση	βοηθά το χρήστη να εντοπίσει γρήγορα την πληροφορία που χρειάζεται και του παρέχει εναλλακτικές επιλογές πλοήγησης Σύσταση συνδέσμων, εκπαίδευση χρήστη
Παραμετροποίηση	Παρέχει παραλλαγές μιας ιστοσελίδας (από άποψη περιεχομένου, δομής, και εμφάνισης), ώστε να ανταποκρίνεται στις γνώσεις, τις προτιμήσεις και τα ενδιαφέροντα κάθε χρήστη. Εξατομικευμένη διάρθρωση (layout) Προσαρμογή περιεχομένου ή συνδέσμων

Υποστήριξη διεκπεραίωσης εργασιών

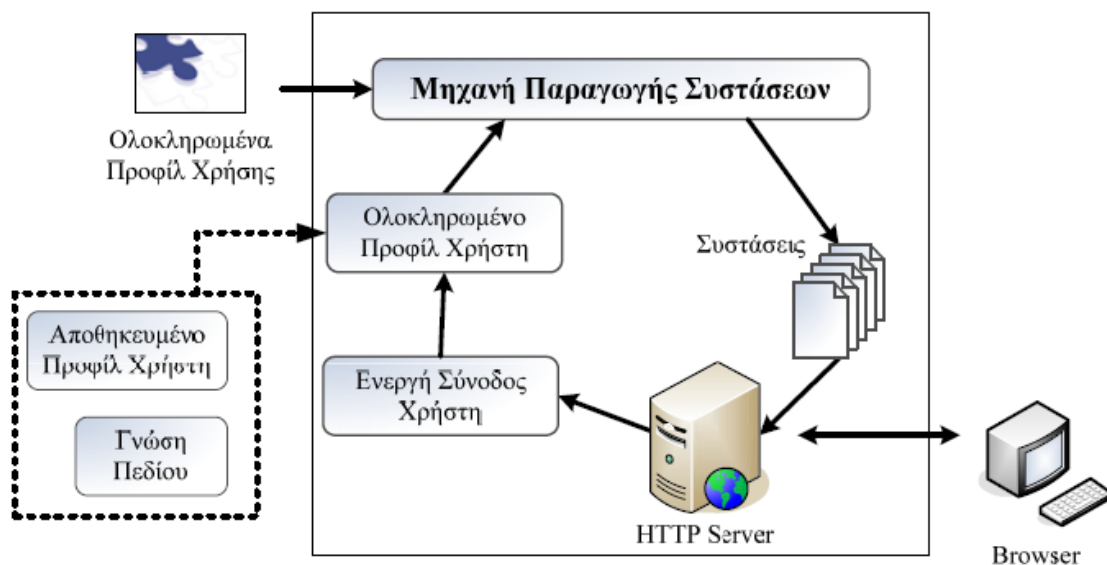
Εξατομικευμένο σχήμα τιμολόγησης

Εξατομικευμένη διαφοροποίηση προϊόντων

Εκτέλεση μιας σειράς πράξεων από το σύστημα για λογαριασμό του χρήστη

Εκπροσωπεί το χρήστη σε online δημοπρασίες

Συμπληρώνει/ βελτιώνει τις ερωτήσεις που υποβάλει ο χρήστης σε μια μηχανή αναζήτησης



Εικόνα 5 Η on-line φάση της διαδικασίας εξατομίκευσης

1.6.1 Διάλογος μεσω Statecharts

Λόγω της σύνθετης και συχνά μεταβαλλόμενης αγοράς των προϊόντων, η συντήρηση ενός ηλεκτρονικού συστήματος συστάσεων είναι μία διαδικασία που απαιτεί χρόνο και χρήμα, δεδομένου ότι η διεπαφή ανθρώπου-μηχανής πρέπει να προσαρμόζεται στο πρότυπο του προϊόντος όποτε αλλάζει. Σε αυτό το στάδιο παρουσιάζονται τα statecharts (διαγράμματα κατάστασης) ως μια προσέγγιση για την παραγωγή ενός λεπτομερούς διαλόγου από ένα δεδομένο μοντέλο προϊόντος, τα οποία εξασφαλίζουν ότι οι αλλαγές στο μοντέλο του προϊόντος αντιπροσωπεύονται άμεσα και στο διάλογο. (Sven R., 2003)

Η μοντελοποίηση του statechart λαμβάνει υπόψη ότι ένα προϊόν συνήθως αποτελείται από διαφορετικά *συστατικά* που είναι κατά ένα μεγάλο μέρος ανεξάρτητα το ένα από το άλλο.

Κάθε συστατικό ενός τμήματος του προϊόντος έχει ένα πλήθος διακριτών *γνωρισμάτων*, τα οποία συνεισφέρουν στην απόφαση υπέρ ή κατά ενός συγκεκριμένου συστατικού. Το εύρος τιμών $dom(f)$ ενός γνωρίσματος αναπαριστάται στο μοντέλο ως ένα πεπερασμένο σύνολο τιμών (στην πράξη, τα συνεχή γνωρίσματα όπως η "τιμή" μπορούν να διακριτοποιηθούν εύκολα).

Έστω ότι ένα αυτοκίνητο αποτελείται από τρία συστατικά που είναι ανεξάρτητα μεταξύ τους, από την άποψη της λήψης απόφασης από τον πελάτη: το σώμα του οχήματος, η μηχανή του, και το χρώμα. Εκτός από έναν προσδιορισμό/ όνομα, το σώμα έχει τα ακόλουθα χαρακτηριστικά γνωρίσματα: Ένα πρότυπο κατηγορίας όπως "Sedan", "Roadster", "SUV" κ.λπ., μια τιμή που διαιρείται σε κατηγορίες, την ιπποδύναμή του, τον τύπο καυσίμων ή την κατανάλωση καυσίμων. Το «χρώμα» έχει έναν τύπο ("φυσικό" ή "μεταλλικό"), φυσικά ένα χρώμα και, πάλι, μια τιμή.

Ο διάλογος συστάσεων στο παράδειγμα αποτελείται από τρία θέματα: "το σώμα του οχήματος", "τη μηχανή" και "το χρώμα". Το θέμα "χρώμα" θα αποτελείται από τρεις ερωτήσεις, που εξετάζουν "τον τύπο", "το χρώμα" και "την τιμή", αντίστοιχα.

Καθώς ο χρήστης δηλώνει τις προτιμήσεις του σχετικά με τις πιθανές τιμές των γνωρισμάτων κάθε συστατικού, τα συστατικά μέρη του προϊόντος μπορούν να συνδυαστούν με έναν ανεξάρτητο τρόπο για να συντεθεί το τελικό προϊόν που θα προταθεί στον χρήστη. Όποτε με βάση αυτή την προϋπόθεση, είναι κατάλληλο να αντιπροσωπευθούν τα τμήματα του προϊόντος ως χωριστά θέματα, που θα είναι δομημένα ως ένας αριθμός ερωτήσεων αποσπώντας από το χρήστη τις προτιμήσεις για ένα συγκεκριμένο γνώρισμα.

Αρχικά, έχουμε μία αρχική κατάσταση, μία τελική κατάσταση και μία κατάσταση για κάθε ερώτηση. Το σύστημα παραμένει σε μία από τις καταστάσεις ερώτησης έως ότου απαντήσει ο χρήστης στην ερώτηση. Η απάντηση του χρήστη οδηγεί στο γεγονός (συμβάν) της απάντησης (**answer event**), το οποίο έχει ως αποτέλεσμα την αποθήκευση των προτιμήσεων που έχει δώσει ο χρήστης στην απάντησή του/της σε μία βάση δεδομένων. Στη συνέχεια, το **answer event** οδηγεί στην κατάσταση **chosenext question**, στην οποία μπορεί είτε να επιλέξει μία άλλη ερώτηση από το ίδιο θέμα, η οποία θα αποτελεί την επόμενη ερώτηση, είτε να προκαλέσει μια μετάβαση στην τελική κατάσταση, υποδεικνύοντας την αλλαγή του θέματος.

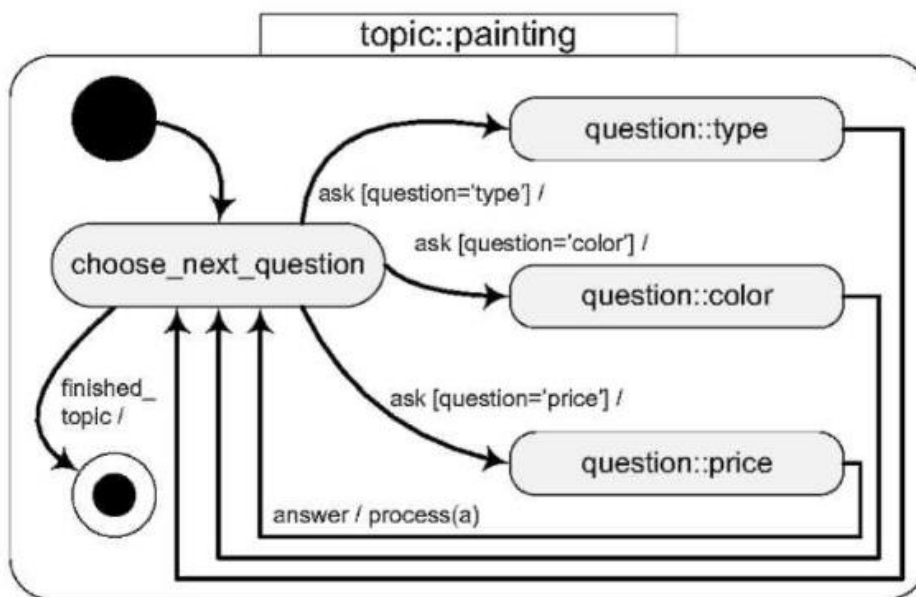
Στην κατάσταση **chosenext question**, εφαρμόζεται το πιθανολογικό συμπέρασμα (το οποίο βασίζεται στα Bayesian δίκτυα) για την αλλαγή θέματος και την εύρεση της καταλληλότερης ερώτησης που θα ερωτηθεί αμέσως μετά. Η χρήση του πιθανολογικού συμπεράσματος επιτρέπει τη συνάθροιση της πιθανότητας δύο ή περισσότερων παραμέτρων για την εύρεση

της πιθανότητας αναγκαιότητας αλλαγής του θέματος. Επιπλέον, οι ερωτήσεις μπορούν να αλλάξουν προτεραιότητα βασιζόμενες στις προηγούμενες απαντήσεις, οι οποίες μπορούν να μοντελοποιηθούν σε ένα Bayesian δίκτυο.

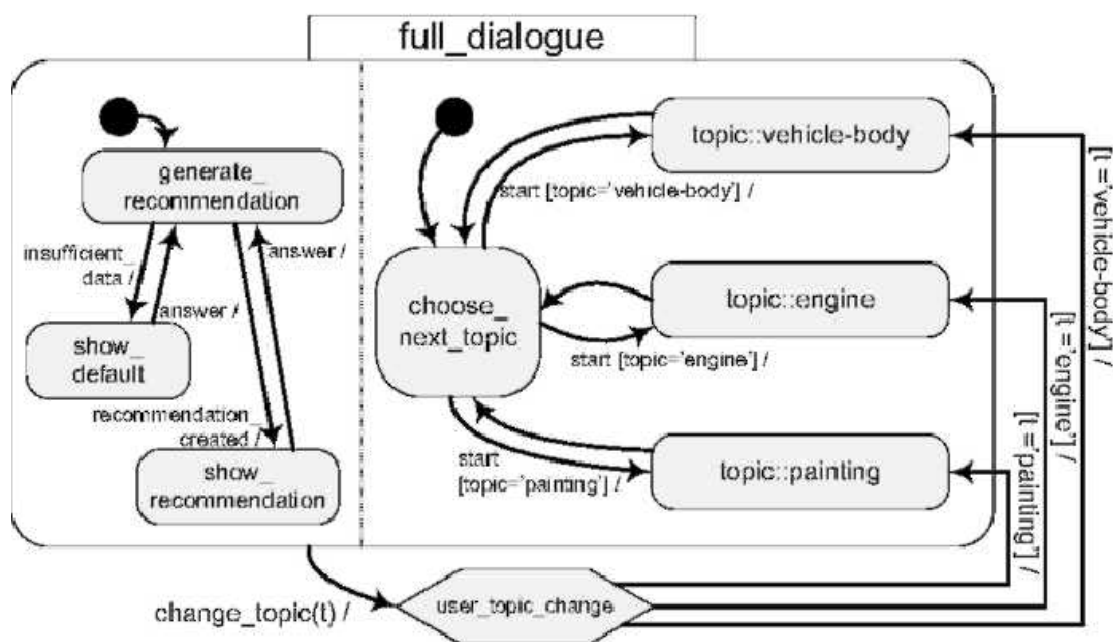
Στο σημείο αυτό, είναι σημαντικό να αναλυθεί η διαδικασία που ακολουθείται σε ένα statechart για την παραγωγή της σύστασης στον χρήστη.

Λόγω της υπόθεσης της ανεξαρτησίας των τμημάτων του προϊόντος, οι συστάσεις μπορούν να δημιουργηθούν ανεξάρτητα για κάθε συστατικό του προϊόντος. Ορίζουμε μια σύσταση ως ένα ταξινομημένο υποσύνολο όλων των διαθέσιμων στοιχείων του καταλόγου προϊόντος για ένα δεδομένο τμήμα του προϊόντος. Για να δημιουργήσουμε συστάσεις, μετασχηματίζουμε το σύνολο προτιμήσεων του χρήστη σε εντολές της SQL. Πιο γενικά, στις προτιμήσεις που ενδιαφέρουν τον χρήστη εκχωρούνται θετικά βάρη, στις προτιμήσεις που δεν τον ενδιαφέρουν εκχωρούνται μηδενικά βάρη ενώ οι άσχετες προτιμήσεις λαμβάνουν αρνητικά βάρη. Αυτό οδηγεί σε ένα διατεταγμένο σύνολο αποτελεσμάτων το οποίο έχει ως πρώτο στοιχείο το καταλληλότερο προϊόν.

Στην εικόνα 5 αναπαριστώνται τα θέματα που μοντελοποιήθηκαν σε έναν ολοκληρωμένο διάλογο στην εικόνα 4.



Εικόνα 6 Statechart (διάγραμμα κατάστασης) για το θέμα «painting» (Sven R., 2003)



Εικόνα 7 Statechart (διάγραμμα κατάστασης) για όλο τον διάλογο, (Sven R., 2003)

Στην κατάσταση *generate_recommendation* (παραγωγή σύστασης), οι τρέχουσες διαθέσιμες προτιμήσεις του χρήστη εφαρμόζονται στον κατάλογο προϊόντων για να ληφθεί η τρέχουσα σύσταση. Η μηχανή σύστασης μπορεί να έχει έξοδο είτε το γεγονός *recommendation_created* (δημιουργημένη σύσταση), που υποδηλώνει μια έγκυρη σύσταση και προκαλεί μια μετάβαση στην κατάσταση *show_recommendation* (εμφάνιση σύστασης), το οποίο οδηγεί το σύστημα συστάσεων στην εμφάνιση των προτεινόμενων προϊόντων στην διεπαφή του χρήστη.

Εναλλακτικά, η έξοδος θα είναι το γεγονός *insufficient_data* (ανεπαρκή δεδομένα), το οποίο σημαίνει ότι δεν υπήρξαν αρκετές προτιμήσεις για να δημιουργήσουν μια σύσταση (δηλ., καμία προτίμηση, αμέσως μετά το ξεκίνημα). Το σύστημα καταστάσεων μεταπηδά στο *show_default* (εμφάνιση προεπιλεγμένου) όπου η διεπαφή του χρήστη εμφανίζει κάποιο είδος «προεπιλεγμένης» σύστασης (π.χ., τρέχουσες ειδικές προσφορές).

Η κατάσταση *choose_next_topic*, είναι υπεύθυνη για τον καθορισμό του επόμενου κατάλληλου θέματος. Τότε αρχίζει μια αλλαγή θέματος με μια μετάβαση σε μια από τις τελικές καταστάσεις του θέματος. Αυτό "επιστρέφει" τον έλεγχο στην κατάσταση *choose_next_topic*, η οποία καλεί έπειτα κάποια άλλη από τις καταστάσεις του θέματος, λαμβάνοντας υπόψη ποια από τα θέματα δεν έχουν ακόμα τελειώσει.

Το συντακτικό στοιχείο της σύνταξης του διαγράμματος καταστάσεων (statechart) που χρησιμοποιείται εδώ καλείται AND-state, και εισάγει τον *παραλληλισμό* στην εκτέλεση των statecharts. Καθώς η μηχανή κατάστασης (state machine) βρίσκεται στη "σύνθετη" κατάσταση του *πλήρους διάλογου*, βρίσκεται ταυτόχρονα σε μια από τις υπο-καταστάσεις του κάθε συστατικού κατάστασης. Κάθε μέρος της σύνθετης κατάστασης αντιδρά στα γεγονότα και εκτελεί τις μεταβάσεις ανεξάρτητα από το άλλο μέρος. Ακόμη, ο παραλληλισμός παρατηρείται και από το γεγονός, ότι και το αριστερό τμήμα από το διάγραμμα κατάστασης της σύστασης αντιδρούν στο γεγονός *answer*.

Ο χρήστης έχει την άδεια να αλλάξει ο ίδιος το θέμα. Θα πρέπει να εξεταστούν δύο "επίπεδα":

- η αλλαγή ερώτησης μέσα στο τρέχον θέμα
- αλλαγή του θέματος από τον χρήστη

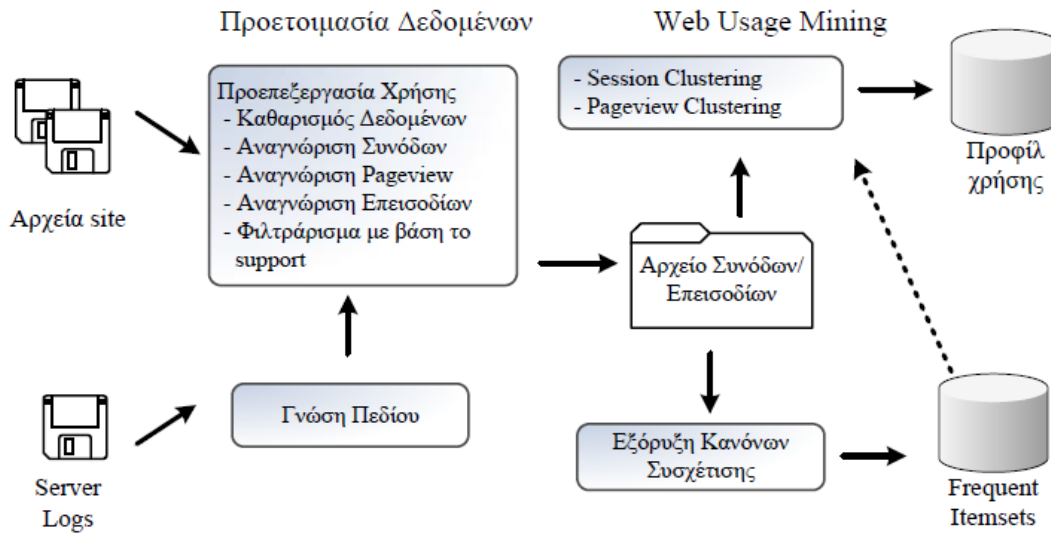
Για αυτόν τον λόγο, εισάγουμε άλλα δύο γεγονότα: το *change_question(q)* και το *change_topic(t)*. Και τα δύο γεγονότα προκαλούν μεταβάσεις στις καταστάσεις απόφασης, και «αποστέλλονται» (dispatch) στην κατάλληλη ερώτηση ή θέμα, αντίστοιχα.

1.6.2 Σύστημα Διαχείρισης Πόρων

Ο Προγραμματισμός Επιχειρησιακών Πόρων υποστηρίζεται από συστήματα λογισμικού που *διευκολύνουν την ροή της πληροφορίας* ανάμεσα σε όλες τις λειτουργίες ενός οργανισμού, (π.χ. την κατασκευή, τα logistics, τα χρηματοοικονομικά και τους ανθρώπινους πόρους). Πρόκειται δηλαδή, για ένα σύστημα το οποίο καλύπτει ολόκληρο το εύρος του οργανισμού. Είναι μία βάση δεδομένων, η οποία καλύπτει όλο το φάσμα των επιχειρησιακών δεδομένων του οργανισμού, λειτουργεί πάνω σε μία κοινή πλατφόρμα, αλληλεπιδρά με ένα ολοκληρωμένο πακέτο εφαρμογών, ενοποιώντας με τον τρόπο αυτό όλες τις εργασίες του οργανισμού σε ένα και μοναδικό υπολογιστικό περιβάλλον. Στην ιδανική περίπτωση, ο στόχος ενός συστήματος ERP είναι να παρέχει τη δυνατότητα εισαγωγής στοιχείων σε αυτό μία και μόνο μία φορά.

Τα σύγχρονα μοντέλα λειτουργίας και διοίκησης απαιτούν οι διάφορες επιχειρησιακές διαδικασίες να ολοκληρώνονται σε περισσότερα από ένα τμήματα της επιχείρησης τα οποία τώρα υποχρεώνονται να εργάζονται με άμεση, συνεχή και αρμονική συνεργασία μεταξύ τους.

Αυτός ο επιμερισμός των διαδικασιών στα διάφορα τμήματα έχει σαν αποτέλεσμα η συλλογή των αναγκαίων πληροφοριών που σχετίζονται με μία συγκεκριμένη διαδικασία να γίνεται σε διαφορετικά σημεία μέσα στην επιχείρηση τα οποία ενδεχομένως να βρίσκονται και σε διαφορετικές γεωγραφικές περιοχές.



Εικόνα 8 Σε επίπεδο στήριξης λειτουργίας καταστημάτων ηλεκτρονικού εμπορίου το ERP λειτουργεί ως βοηθητικό παρασκηνακό σύστημα εκτέλεσης συναλλαγών

Ο Προγραμματισμός των Επιχειρηματικών Πόρων [Enterprise Resource Planning (ERP)] είναι μία συστηματική μέθοδος για την δυναμική εξισορρόπηση αλλά και την βέλτιστη αξιοποίηση των πόρων ενός οργανισμού. Όταν χρησιμοποιείται με αποτελεσματικό τρόπο, μπορεί να φέρει αξιοσημείωτα αποτελέσματα όσον αφορά στην ανάπτυξη του οργανισμού, στην κερδοφορία του, στην ανάπτυξη καινοτομικών προϊόντων και υπηρεσιών και στην διαμόρφωση νέων πλαισίων επιχειρησιακής λειτουργίας. (Χ. Χριστάκου, 2012)

1.6.3 Συστήματα Συστάσεων (Recommender Systems)

Ως *Recommender System* ορίζεται το σύστημα που στοχεύει να προτείνει πληροφορίες ή κοινωνικά στοιχεία που πιθανότατα ενδιαφέρουν το χρήστη. Αυτό που ουσιαστικά κάνει ένα τέτοιο σύστημα είναι να προσπαθεί να μαντέψει πόσο θα άρεσε στο χρήστη μια πληροφορία ή ένα κοινωνικό στοιχείο το οποίο ο ίδιος δεν έχει βαθμολογήσει στο παρελθόν. Για να το πετύχει αυτή συγκρίνει το προφίλ του χρήστη με κάποια χαρακτηριστικά. Η λειτουργία τους βασίζεται στην ιστορία της αλληλεπίδρασης του χρήστη αλλά και στην ομοιότητα του χρήστη με άλλους χρήστες. Το πώς το σύστημα θα μάθει τις προτιμήσεις του χρήστη μπορεί να γίνει με πολλούς διαφορετικούς τρόπους. Τα Recommender Systems αποτελούν μία χρήσιμη εναλλακτική των αλγορίθμων αναζήτησης εφόσον βοηθούν τους χρήστες να ανακαλύψουν αντικείμενα που πιθανώς δε θα έβρισκαν μόνοι τους ή θα έβρισκαν ξοδεύοντας πολύτιμο χρόνο.

Οι δύο βασικές έννοιες που εμφανίζονται σε όλα τα Συστήματα Συστάσεων είναι ο *χρήστης (user)*, ο οποίος αναφέρεται και ως πελάτης και το *στοιχείο (item)*, το οποίο αναφέρεται και ως προϊόν. Ο *χρήστης* είναι το άτομο που χρησιμοποιεί το σύστημα συστάσεων παρέχοντας την γνώμη του για ποικίλα στοιχεία και λαμβάνει συστάσεις για νέα στοιχεία από το σύστημα. Η είσοδος ενός Συστήματος Συστάσεων εξαρτάται από τον τύπο του αλγορίθμου φιλτραρίσματος. Γενικά, η είσοδος ανήκει σε μια από τις παρακάτω κατηγορίες:

Πίνακας 5 Τύποι εισόδου συστήματος συστάσεων

Αξιολογήσεις	Εκφράζουν την γνώμη των χρηστών για τα στοιχεία. Οι αξιολογήσεις παρέχονται από τον χρήστη και ακολουθούν μια συγκεκριμένη αριθμητική κλίμακα (για παράδειγμα από 1-ασχημα έως 5-τέλεια). Οι αξιολογήσεις μπορούν να ληφθούν από το ιστορικό αγοράς του χρήστη, τις επισκέψεις συνδέσμων, την αναζήτηση συνηθειών, ή άλλου είδους πληροφοριών πρόσβασης.
Δημογραφικά Δεδομένα	Αναφέρονται ως πληροφορίες όπως η ηλικία, το φύλο και η εκπαίδευση των χρηστών. Αυτά του είδους τα δεδομένα, είναι δύσκολο να συλλεχθούν και λαμβάνονται από τον χρήστη.
Δεδομένα περιεχομένου	Βασίζονται στην λεκτική ανάλυση (textual analysis) που γίνεται στα αρχεία σχετικά με τα στοιχεία που αξιολογήθηκαν από τον χρήστη. Τα γνωρίσματα που εξάγονται από αυτή την ανάλυση χρησιμοποιούνται ως

	είσοδος στον αλγόριθμο φιλτραρίσματος για να συνάγουν το προφίλ του χρήστη.
Έξοδος Συστήματος	Μπορεί να είναι μία πρόβλεψη ή μια σύσταση.
Η πρόβλεψη	Εκφράζεται σαν μια αριθμητική τιμή, που εκφράζει την άποψη του χρήστη που χρησιμοποιεί το σύστημα για ένα στοιχείο. Αυτή η τιμή θα πρέπει απαραίτητα να είναι στην ίδια αριθμητική κλίμακα (για παράδειγμα από 1-άσχημα έως 5-τέλεια).
Η σύσταση	Εκφράζεται ως μια λίστα N στοιχείων, την οποία ο ενεργός χρήστης επέλεξε ότι του αρέσει περισσότερο. Η συνηθισμένη προσέγγιση σε αυτή την περίπτωση απαιτεί αυτή η λίστα να περιλαμβάνει μόνο στοιχεία που ο ενεργός χρήστης δεν έχει αγοράσει, δει ή αξιολογήσει.
Οι απόψεις των χρηστών	Δηλώνονται σε μορφή <i>αξιολόγησης με βαθμολογία</i> . Συγκεκριμένα, η αξιολόγηση του u χρήστη για το στοιχείο ij , όπου $j = 1 \dots n$, δηλώνεται ως r_{ij} , όπου κάθε αξιολόγηση είναι είτε ένας πραγματικός αριθμός που ανήκει στη αριθμητική κλίμακα είτε \perp , το σύμβολο για "καμία αξιολόγηση". Όλες αυτές οι διαθέσιμες αξιολογήσεις συλλέγονται σε έναν πίνακα <i>χρήστη-στοιχείων</i> . Οι προτεινόμενοι αλγόριθμοι φιλτραρίσματος υιοθετούν διάφορες τεχνικές είτε στις <i>γραμμές</i> , που αντιστοιχούν στις αξιολογήσεις ενός μόνο χρήστη για τα διαφορετικά στοιχεία, είτε στις <i>στήλες</i> , που αντιστοιχούν στις αξιολογήσεις των διαφορετικών χρηστών για ένα μόνο στοιχείο, του <i>πίνακα χρήστη-στοιχείων</i> .

Τα Συστήματα Συστάσεων (Recommender systems) παράγουν εξατομικευμένες συστάσεις για στοιχεία από ένα μεγάλο διάστημα πιθανών επιλογών (Montaner M., 2003). Για να το κάνουν αυτό, το Σύστημα Συστάσεων χρειάζεται ένα μοντέλο με τις προτιμήσεις του χρήστη σε έναν τομέα που έχει επιλέξει ότι τον ενδιαφέρει, και μια μηχανή συστάσεων που χρησιμοποιεί το μοντέλο προτιμήσεων του χρήστη για να προτείνει ένα υποσύνολο των στοιχείων του τομέα. Η σχεδίαση των συστημάτων που βασίζονται στη μοντελοποίηση χρήστη στηρίζονται σε τρία βήματα:

Πίνακας 6 Βήματα σχεδίασης συστημάτων

Αρχικοποίηση	αποτελείται από την συλλογή των προτιμήσεων του χρήστη προκειμένου να κατασκευαστεί ένα σωστό και επαρκές μοντέλο προτιμήσεων. (Burke R., 2002)
Αναβάθμιση	είναι μια πιο μακροπρόθεσμη συντήρηση του μοντέλου προτίμησης προκειμένου να τηρηθούν οι προτιμήσεις του χρήστη. Περιλαμβάνει την παρακολούθηση των στοιχείων που αγόρασε, των αξιολογήσεων του χρήστη για στοιχεία ή ιδιότητες, και άλλα. Ένας κοινός τρόπος για να ενημερωθεί το μοντέλο είναι η ανατροφοδότηση χρηστών στα στοιχεία που προτείνονται.
Μοντελοποίηση	χρήση του μοντέλου του χρήστη για την παραγωγή προσαρμοστικής λειτουργικότητας (Auriol E., 1994).

Μετά την αρχικοποίηση του, το μοντέλο προτίμησης χρήστη τίθεται σε χρήση από μια μηχανή που βασίζεται σε έναν ή στον συνδυασμό περισσότερων αλγορίθμων συστάσεων.

Συμπεράσματα Κεφαλαίου

Οι πύλες καταλόγου είναι πληροφοριακά συστήματα που έχουν υλοποιηθεί με κύριο γνώμονα την αποδοτική και σαφή οργάνωση των πληροφοριών σε θεματικές κατηγορίες, ώστε χρήστες με διαφορετικά ενδιαφέροντα να έχουν εξίσου εύκολη πρόσβαση στη γνώση που αναζητούν.

Οι πύλες καταλόγου είναι υπό μία έννοια απαραίτητες – ειδικά στο χώρο του διαδικτύου -, λόγω του ραγδαία αυξανόμενου όγκου πληροφοριών, οι οποίες χωρίς την κατάλληλη οργάνωση είναι δύσκολα προσβάσιμες άρα και εκμεταλλεύσιμες

Για παράδειγμα, αν ένα μεγάλο πλήθος χρηστών δείχνει μεγάλο ενδιαφέρον για τις Τέχνες (ή για κάποια τέχνη συγκεκριμένα), θα τους διευκόλυνε να βρίσκεται η κατηγορία “Τέχνες” σε περίοπτη θέση στην κεντρική σελίδα της πύλης ή να εμφανίζονται συντομεύσεις για δημοφιλείς υποκατηγορίες της. Τέτοιες αλλαγές προσαρμοσμένες στην ετερογένεια των ενδιαφερόντων των χρηστών της πύλης συνιστούν τη λειτουργία της εξατομίκευσης.

Μια σημαντική πτυχή της εξατομίκευσης και των Συστημάτων Συστάσεων είναι η ποιότητα των συστάσεων για τους χρήστες που τα χρησιμοποιούν. Οι χρήστες χρειάζονται συστάσεις που μπορούν να εμπιστευθούν και θα τους βοηθήσουν να βρουν τα προϊόντα που τους αρέσουν. Εάν ένας πελάτης εμπιστεύεται ένα Σύστημα Συστάσεων, και αγοράζοντας ένα προϊόν ανακαλύψει ότι το προϊόν δεν του αρέσει, τότε είναι απίθανο να χρησιμοποιήσει ξανά το σύστημα συστάσεων. Τα Συστήματα Συστάσεων έχουν δύο χαρακτηριστικούς τύπους λαθών: τα *λανθασμένα αρνητικά (false negatives)*, που αντιπροσωπεύουν τα προϊόντα που δεν προτείνονται, αν και ο πελάτης θα τα ήθελε, και τα *λανθασμένα θετικά (false positives)*, τα οποία αντιπροσωπεύουν τα προϊόντα που προτείνονται, αν και δεν αρέσουν στον πελάτη. Στον τομέα του ηλεκτρονικού

εμπορίου τα σημαντικότερα λάθη που θα πρέπει να αποφεύγονται είναι τα *λανθασμένα θετικά (false positives)*, δεδομένου ότι αυτά τα λάθη οδηγούν σε θυμωμένους πελάτες, και δεδομένου ότι υπάρχουν πολλά προϊόντα στους ιστοχώρους ηλεκτρονικού εμπορίου που θα ήθελε να αγοράσει ένας πελάτης, δεν υπάρχει κανένας λόγος να διακινδυνεύσουμε προτείνοντας ένα προϊόν που δεν επιθυμεί.

Στην παρούσα εργασία, όπου επικεντρωνόμαστε στην μελέτη του τομέα της ηλεκτρονικής διακυβέρνησης, θα πρέπει να αποφεύγονται τόσο τα *λανθασμένα αρνητικά* όσο και τα *λανθασμένα θετικά*, και γενικά η πιθανότητα σφάλματος να είναι μηδενική, αφού η ενημέρωση του χρήστη με λανθασμένες πληροφορίες οδηγεί σε προβλήματα για τον ίδιο αλλά και την υπηρεσία.

Για τους ως άνω λόγους το σύστημα θα πρέπει να στηρίζεται από ένα ολοκληρωμένο σύστημα ERP (Enterprise Resource Planning) η εφαρμογή του οποίου θα διασφαλίζει την ποιότητα των παρεχόμενων υπηρεσιών.

2. ENTERPRISE RESOURCE PLANNING

2.1 Συστήματα Οργάνωσης Επιχειρήσεων

Το Enterprise Resource Planning αποτελεί μια εφαρμογή οργάνωσης της εταιρείας και στην περίπτωση της εξατομίκευσης υπηρεσιών ηλεκτρονικού εμπορίου λειτουργεί ως ένα βοηθητικό σύστημα υποστήριξης ανάπτυξης των μένων υπηρεσιών. Με την εφαρμογή του η εκάστοτε εταιρεία, ασχέτως του κλάδου στον οποίο δραστηριοποιείται, προσβλέπει στην απλοποίηση των διαδικασιών ανταλλαγής πληροφοριών μεταξύ των διαφόρων τμημάτων που την στελεχώνουν. Χαρακτηριστικά στον ορισμό που δίνεται στο άρθρο του (T. Wailgum, 2013), σχετικά με το ERP, υποστηρίζεται πως όλες οι πληροφορίες μια εταιρείας μπορούν να συγκεντρωθούν σε ένα και μόνο υπολογιστή.

Η κατασκευής ενός προγράμματος που θα είναι σε θέση αν επιτηρεί και να ελέγχει όλους τους τομείς της επιχείρησης αδιαμφισβήτητα είναι υψηλός στόχος. Για να κατανοηθεί η πολυπλοκότητα του συστήματος που το πρόγραμμα καλείται να απλουστεύσει θα πρέπει να αναλογιστούμε πως το τμήμα πωλήσεων, το τμήμα αποθήκευσης και το τμήμα συντήρησης μιας υποθετικής επιχείρησης, διαθέτουν ξεχωριστές βάσεις δεδομένων και μηχανογράφηση που η συγκέντρωσή τους σε μια ενιαία βάση δεδομένων είναι τουλάχιστον σύνθετη διαδικασία.

Πλέον είμαστε σε θέση λόγω της προόδου της τεχνολογίας έχει βοηθήσει στην βελτίωση των τεχνικών αποθήκευσης και ανάλυσης δεδομένων. Με τη βοήθεια της σύγχρονης τεχνολογίας πληροφοριακών συστημάτων, οι εταιρείες είναι σε θέση να συναρμολογούν πληθώρα αποσπασματικών πληροφοριών. (Φ. Μιχελινάκης, 2011)

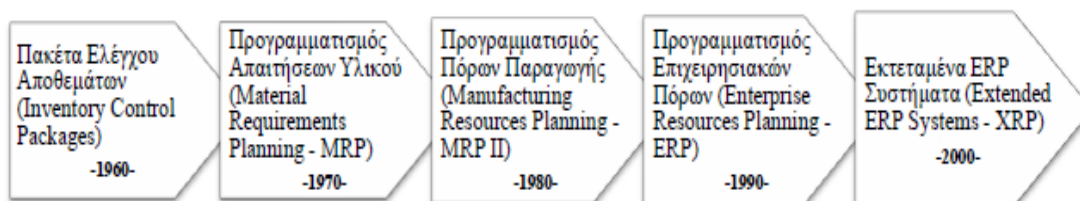
2.2 Εξελεγκτική Διαδικασία Συστημάτων Οργάνωσης

Τη δεκαετία του 1960, αναπτύχθηκε σε διεθνές επίπεδο η μηχανογραφημένη υποστήριξη των πολύπλοκων λειτουργιών των μεγάλων επιχειρήσεων της εποχής. Στα πλαίσια αυτής της ανάπτυξης παρουσιάστηκαν εξειδικευμένα πληροφορικά πακέτα που υποβοηθούσαν βασικές διαδικασίες της οικονομικής διαχείρισης, όπως είναι η λογιστική και η μισθοδοσία, με δευτερογενές αποτέλεσμα την διευκόλυνση της εφαρμογής αναλυτικών μεθόδων ελέγχου άλλων τομέων (π.χ. αποθέματα).

Την περίοδο 1968 - 1973 εμφανίστηκαν τα συστήματα MRP (Material Requirements Planning), τα οποία παρουσίασαν κάποιο βαθμό ολοκλήρωσης καθώς μετέφραζαν το Βασικό Πλάνο Παραγωγής (Master Production Schedule) των τελικών προϊόντων σε χρονικά καταναμημένες απαιτήσεις παραγωγής υποσυναρμολογημάτων και συστατικών, και σε απαιτήσεις προμήθειας πρώτων υλών. Με την εμφάνιση του MRP-II (Manufacturing Resources Planning) στα τέλη της δεκαετίας του 1970, το σύστημα MRP συνέδεσε μεταξύ τους τα κυκλώματα προγραμματισμού παραγωγής, του ελέγχου παραγωγής και της κοστολόγησης, και των προμηθειών.

Τη δεκαετία του 80' ξεκινά μια ερευνητική προσπάθεια για επιχειρηματική ολοκλήρωση (enterprise integration), η οποία χρησιμοποιεί ως τεχνολογικό υπόβαθρο τις βάσεις δεδομένων (databases) και προσπαθεί να ενοποιήσει τις βασικές επιχειρηματικές διαδικασίες με βασική προτεραιότητα το κύκλωμα οικονομικής διαχείρισης και το κύκλωμα παραγωγής.

Αποτέλεσμα της προσπάθειας αυτής είναι η εμφάνιση των συστημάτων Enterprise Resources Planning (Προγραμματισμός Επιχειρηματικών Πόρων) στα τέλη της δεκαετίας του 1980, τα οποία ολοκληρώνουν, πέραν του κυκλώματος Οικονομικής Διαχείρισης και Παραγωγής, και άλλες βασικές επιχειρηματικές διαδικασίες, όπως τη Διαχείριση Ανθρώπινων Πόρων, το κύκλωμα Πωλήσεων κ.λπ. Τα συστήματα ERP είναι λοιπόν ολοκληρωμένα συστήματα πληροφορικής, τα οποία καλύπτουν όλες τις λειτουργικές περιοχές μιας επιχείρησης, ώστε να ικανοποιηθούν οι στόχοι της, ενοποιώντας όλες τις διαδικασίες της.



Εικόνα 9 Εξέλιξη πληροφοριακών συστημάτων

2.3 Πληροφοριακά Συστήματα

Σύμφωνα με τους Laudon & Laudon 2009

«Ένα πληροφοριακό σύστημα τεχνικά ορίζεται ως ένα σύνολο αλληλοσχετιζόμενων στοιχείων τα οποία συλλέγουν, επεξεργάζονται, αποθηκεύουν και διανέμουν πληροφορίες που υποστηρίζουν τη λήψη αποφάσεων και τον έλεγχο σε έναν οργανισμό.»



Εικόνα 10 Βασικές λειτουργίες ενός πληροφοριακού συστήματος

Για να μπορέσει μια επιχείρηση να παίρνει αποφάσεις, να ελέγχει τις λειτουργίες, να αναλύει τα προβλήματα και να δημιουργεί νέα προϊόντα ή υπηρεσίες, θα πρέπει ένα πληροφοριακό σύστημα να δραστηριοποιείται σε τέσσερις βασικές διαδικασίες :

- Είσοδο - Input
- Επεξεργασία - Processing
- Έξοδος - Output
- Αναπληρόρηση - Feedback

Ο ρόλος της εισόδου σε ένα πληροφοριακό σύστημα είναι να συλλέγει πρωτογενή δεδομένα μέσα από τον οργανισμό ή ακόμα από το εξωτερικό περιβάλλον του έχοντας ως σκοπό την επεξεργασία τους μέσα σε ένα πληροφοριακό σύστημα

Η επεξεργασία σε ένα πληροφοριακό σύστημα αποτελεί τη μετατροπή της πληροφορίας, πιο συγκεκριμένα την κατάλληλη διαχείριση και ανάλυση της πρωτογενούς εισόδου σε μια πιο κατανοητή μορφή για τους ανθρώπους.

Η έξοδος είναι υπεύθυνη για τη διανομή επεξεργασμένων πληροφοριών στους ανθρώπους που είτε θα τις χρησιμοποιήσουν είτε θα χρησιμοποιηθούν σε άλλες δραστηριότητες.

Αναπληρόρηση πρόκειται ουσιαστικά για την έξοδο που επιστρέφει στα κατάλληλα μέρη του οργανισμού ώστε να τα βοηθήσει να αξιολογήσουν ή να διορθώσουν την είσοδο.

2.4 Προσφορά Συστήματος ERP στην Επιχείρηση

Ένα σύστημα πληροφοριών μπορεί να οριστεί τεχνικά ως ένα σύνολο αλληλοσχετιζόμενων στοιχείων, τα οποία συλλέγουν, επεξεργάζονται αποθηκεύουν και διανέμουν πληροφορίες που υποστηρίζουν τη λήψη αποφάσεων και τον έλεγχο σε ένα οργανισμό. Επιπλέον της υποστήριξης στη λήψη αποφάσεων, στο συντονισμό και στον έλεγχο τα συστήματα πληροφοριών μπορούν να βοηθήσουν τα στελέχη και το προσωπικό στην ανάλυση προβλημάτων, στην απεικόνιση σύνθετων θεμάτων και στη δημιουργία νέων προϊόντων.

Σύμφωνα με τον (T. Wailgum, 2013)⁴ υπάρχουν πέντε κύριοι παράγοντες για τους οποίους οι εταιρείες, αναλαμβάνουν να υιοθετήσουν ένα σύστημα ERP.

Πίνακας 7 Οφέλη υιοθέτησης συστήματος ERP

Ενσωματωμένες Οικονομικές Πληροφορίες	Δεδομένου ότι ο διευθύνων σύμβουλος προσπαθεί να κατανοήσει τη συνολική απόδοση της εταιρείας, μπορεί να βρει πολλές διαφορετικές εκδοχές της αλήθειας. Στο τομέα των οικονομικών έχει τη δική του άποψη για το σύνολο των εσόδων, των πωλήσεων και οι διαφορετικές επιχειρηματικές μονάδες μπορεί να έχουν τη δική τους εκδοχή του πόσο συνέβαλε στα έσοδα. Το ERP
--	---

⁴ http://www.cio.com/article/40323/ERP_Definition_and_Solutions?page=2#long

**Ενσωματωμένες
Πληροφορίες
Παραγγελιών Πελάτη**

δημιουργεί μια ενιαία εκδοχή της αλήθειας που δεν μπορεί να αμφισβητηθεί, διότι ο καθένας χρησιμοποιεί το ίδιο σύστημα.

Τα ERP συστήματα μπορούν να παρακολουθούν όλη την διάρκεια της παραγγελίας από τη στιγμή που ένας εκπρόσωπος εξυπηρέτησης πελατών την λαμβάνει μέχρι τη φόρτωση των εμπορευμάτων και την αποστολή του τιμολογίου. Από την αρχειοθέτηση αυτών των πληροφοριών σε ένα σύστημα λογισμικού και όχι διάσπαρτα ανάμεσα σε πολλά διαφορετικά συστήματα που δεν μπορούν να επικοινωνήσουν μεταξύ τους, οι εταιρείες μπορούν να παρακολουθούν τις παραγγελίες πιο εύκολα

**Ταχύτερη και
Ασφαλέστερη Διαδικασία
Παραγωγής**

Εταιρείες που έχουν προέλθει από συγχωνεύσεις και εξαγορές, συχνά διαπιστώνουν ότι οι πολλαπλές επιχειρηματικές μονάδες σε ολόκληρη την εταιρεία κάνει την ίδια εργασία χρησιμοποιώντας διαφορετικές μεθόδους και συστήματα ηλεκτρονικών υπολογιστών. Τα ERP συστήματα εφαρμόζουν τυποποιημένες μεθόδους χρησιμοποιώντας ένα ενιαίο, ολοκληρωμένο σύστημα ηλεκτρονικού υπολογιστή το οποίο μπορεί να εξοικονομήσει χρόνο, να αυξήσει την παραγωγικότητα και να μειώσει τον αριθμό κεφάλι.

Μείωση Αποθεμάτων

Ένα σύστημα ERP βοηθά τη ροή της διαδικασίας παραγωγής και βελτιώνει την ορατότητα της διαδικασίας εκτέλεσης παραγγελιών στο εσωτερικό της εταιρείας. Αυτό μπορεί να οδηγήσει σε μείωση αποθεμάτων και μπορεί να βοηθήσει στην ταχύτερη παράδοση των παραγγελιών, μειώνοντας το τελικό καλό απόθεμα στις αποθήκες και τα φορτωτικά έγγραφα. Για να βελτιώσει πραγματικά τη ροή της εφοδιαστικής αλυσίδας, θα πρέπει να είναι αντίστοιχα καλά οργανωμένη η προμηθευτική αλυσίδα.

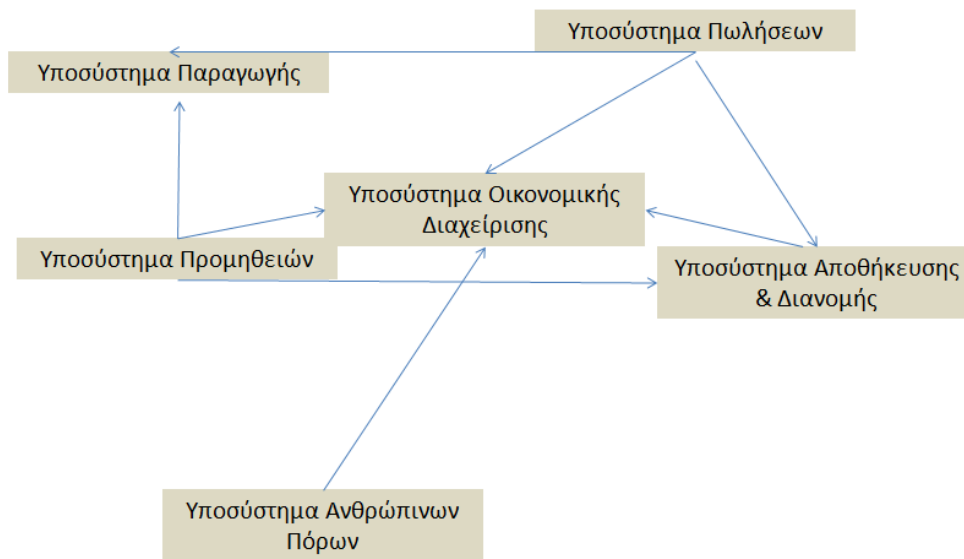
Διαχείριση Προσωπικού

Ειδικά σε εταιρείες με πολλαπλές επιχειρηματικές μονάδες, το τμήμα διαχείρισης προσωπικού δεν μπορεί να έχει μια ενιαία, απλή μέθοδος για την παρακολούθηση του χρόνου των εργαζομένων και την επικοινωνία μαζί τους για τα οφέλη και τις υπηρεσίες. Ένα σύστημα ERP μπορεί να διορθώσει αυτό.

2.5 Λειτουργικότητα Συστήματος ERP

Τα συστήματα ERP είναι δομημένα σε "λειτουργικά υποσυστήματα" (functional modules). Για κάθε τύπο επιχείρησης, ασχέτως του τομέα στον οποίο δραστηριοποιείται διακρίνονται τα εξής λειτουργικά συστήματα.

- Το υποσύστημα Οικονομικής Διαχείρισης
- Το υποσύστημα των Πωλήσεων
- Το υποσύστημα των Προμηθειών
- Το υποσύστημα Ανθρώπινων Πόρων
- Το υποσύστημα της Παραγωγής
- Το υποσύστημα Αποθήκευσης - Διανομής



Εικόνα 11 Αλληλεπίδραση υποσυστημάτων

Το υποσύστημα **Οικονομικής Διαχείρισης** είναι η καρδιά του ERP, και ανταλλάσσει πληροφορίες με όλα τα υπόλοιπα υποσυστήματα. Βασικές διαδικασίες της Οικονομικής Διαχείρισης περιλαμβάνουν (Λ. Στάμκου, 2011)

- τη Γενική Λογιστική (General Ledger)
- την Αναλυτική Λογιστική (Analytical Ledger)
- τη Διαχείριση Παγίων (Asset Management)
- τις Οικονομικές Καταστάσεις (Financial Statements)

- τους Εισπρακτέους Λογαριασμούς (Accounts Receivable)
- τους Πληρωτέους Λογαριασμούς (Accounts Payable)
- τη Διαχείριση Διαθεσίμων (Treasury Management)

Το **υποσύστημα Πωλήσεων** ανταλλάσσει πληροφορίες κυρίως με τα υποσυστήματα Οικονομικής Διαχείρισης, Αποθήκευσης και Διανομής, και Παραγωγής. Οι βασικές λειτουργίες του υποσυστήματος Πωλήσεων - Marketing περιλαμβάνουν

- την Παραγγελιοληψία (Order Entry),
- την Τιμολόγηση (Invoicing),
- τη Διαχείριση Συμβολαίων (Sales Contracts),
- το Μητρώο Πελατών (Customer Table),
- τα Αξιόγραφα, Open Items,
- και Στατιστικά Πωλήσεων.
- Εξυπηρέτηση Πελατών (Customer Service)
- Προβλέψεις Ζήτησης (Forecasting),
- Ηλεκτρονικό Εμπόριο μέσω Internet (Electronic Commerce)

Το **υποσύστημα Προμηθειών** ανταλλάσσει πληροφορίες κυρίως με τα υποσυστήματα Οικονομικής Διαχείρισης, Αποθήκευσης και Διανομής, και Παραγωγής. Οι βασικές λειτουργίες του υποσυστήματος Προμηθειών περιλαμβάνουν

- τον Έλεγχο και Διαχείριση Αιτήσεων Αγοράς (Purchase Inquiries Control & Management)
- τη Διαχείριση Εντολών Αγοράς (Purchase Orders Management)
- τον Έλεγχο Παραλαβών (Receipt Control)
- την Αξιολόγηση Προμηθευτών (Supplier Evaluation)
- τη Διαχείριση Συμβάσεων (Contract Management)

Το **υποσύστημα Αποθήκευσης - Διανομής** ανταλλάσσει πληροφορίες με τα υποσυστήματα Οικονομικής Διαχείρισης, Πωλήσεων - Marketing, Προμηθειών και Παραγωγής. Οι βασικές λειτουργίες του υποσυστήματος Αποθήκευσης - Διανομής περιλαμβάνουν

- τη Διαχείριση Αποθεμάτων (Inventory Control)

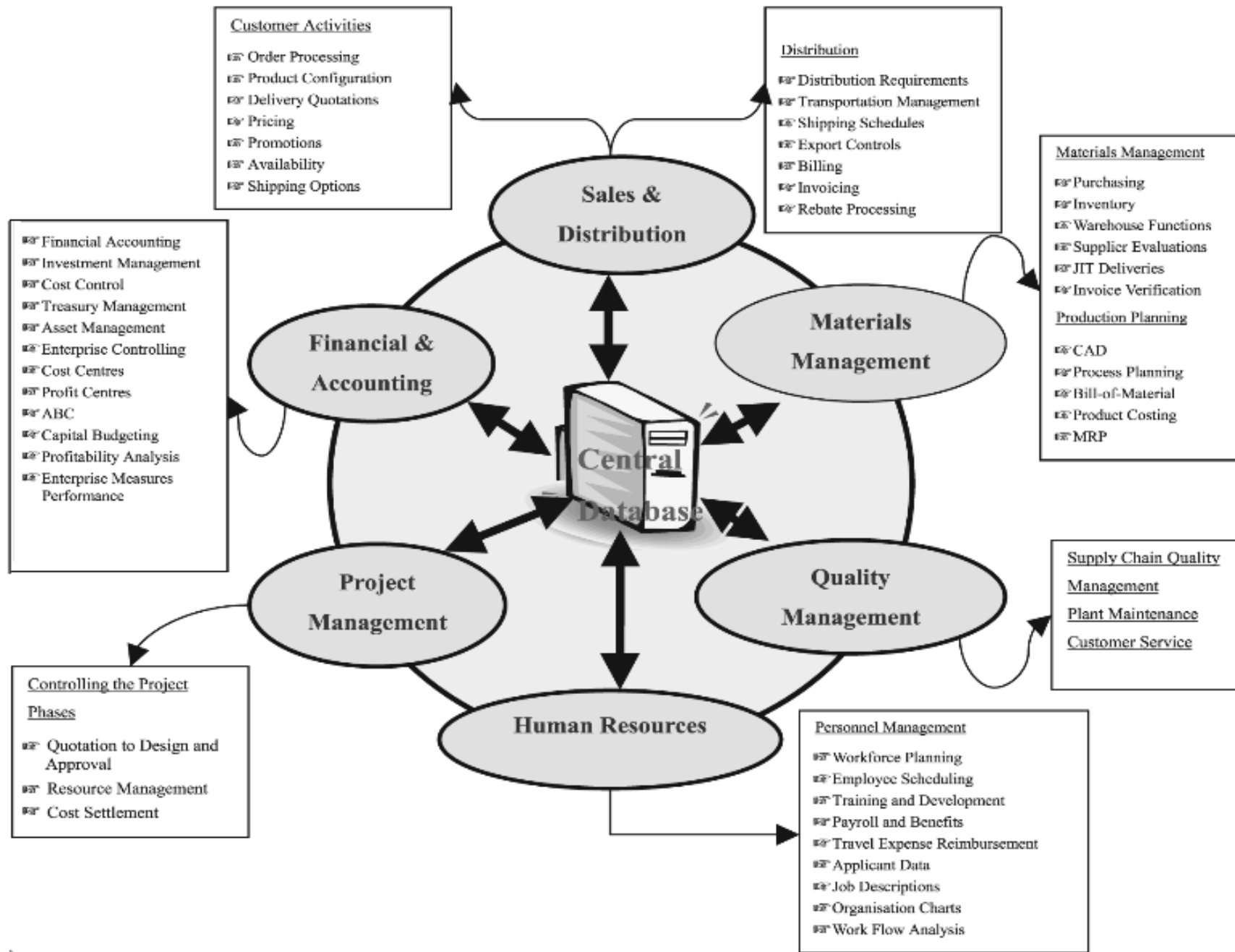
- τον Προγραμματισμό Απαιτήσεων Διανομής (Distribution Requirement Planning). τη Διαχείριση Αποθηκών (Warehouse Management)
- τη Διαχείριση Στόλου Φορτηγών (Fleet Management).

Το σύστημα Πόρων ανταλλάσσει πληροφορίες κυρίως με το υποσύστημα Οικονομικής Διαχείρισης. Οι βασικές λειτουργίες που καλύπτει το υποσύστημα Ανθρώπινων Πόρων περιλαμβάνουν

- τον Προγραμματισμό Προσωπικού (Personnel Planning),
- τη Μισθοδοσία (Payroll),
- την Αξιολόγηση Προσωπικού (Personnel Evaluation)
- τα Εξοδολόγια (Personnel Expenses),

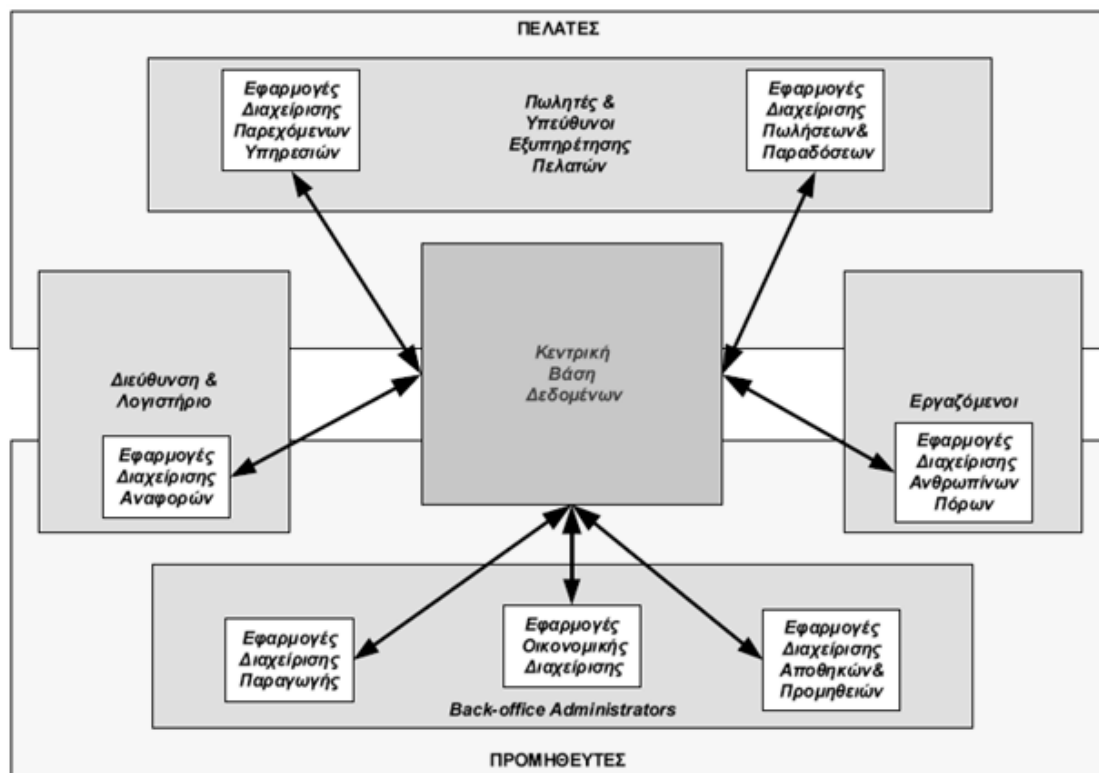
Το **υποσύστημα Παραγωγής** ανταλλάσσει πληροφορίες με τα υποσυστήματα Οικονομικής Διαχείρισης, Πωλήσεων - Marketing, Προμηθειών και Αποθήκευσης - Διανομής. Οι βασικές λειτουργίες που καλύπτει το υποσύστημα Παραγωγής περιλαμβάνουν τον

- Προγραμματισμό Απαιτήσεων Δυναμικότητας (Capacity Requirements Planning)
- το Μακροπρόθεσμο Προγραμματισμό Παραγωγής (Master Prod. Scheduling)
- τον Προγραμματισμό Απαιτήσεων Υλικών (Material Requirements Planning)
- τον Έλεγχο Παραγωγής (Shop Floor Control)
- την Κοστολόγηση Παραγωγής (Cost Accounting)



Τα λειτουργικά αυτά υποσυστήματα υποστηρίζονται από τη βάση δεδομένων του συστήματος, στην οποία κάθε στοιχείο αντιπροσωπεύεται μια και μοναδική φορά. Η βάση δεδομένων αποτελεί το πληροφοριακό μοντέλο της ολοκληρωμένης γνώσης της επιχείρησης. Η υλοποίηση των ως άνω υποστηρίζεται από μια αρχιτεκτονική τριών επιπέδων (3tier). Ο ρόλος κάθε επιπέδου είναι ο εξής :

- Η βάση δεδομένων αποτελεί το μέσο αποθήκευσης των δεδομένων, εξασφαλίζοντας την ακεραιότητα τους και υλοποιώντας τμήμα των λειτουργιών του συστήματος.
- Ο εξυπηρετητής εκτελεί το κύριο τμήμα των λειτουργιών, εξασφαλίζει την ακεραιότητα και την λειτουργία της βάσης δεδομένων και την επικοινωνία με αυτήν.
- Ο πελάτης ασχολείται με την εισαγωγή, εμφάνιση και εκτύπωση των πληροφοριών από και προς τον τελικό χρήστη



Εικόνα 12 Σύνδεση συστήματος βάσεις δεδομένων με τις διάφορες εφαρμογές της επιχείρησης

2.6 Χρόνος Υλοποίησης Εφαρμογής

Η εφαρμογή του συστήματος χρειάζεται ένα μεγάλο χρονικό διάστημα ώστε να αποδώσει αποτελέσματα. Οι μικρές εταιρείες που εγκαθιστούν σύστημα ERP θα πρέπει να περιμένουν απτά αποτελέσματα μετά από έξι μήνες εφαρμογής στου. Το διάστημα των έξι μηνών είναι πραγματικά πολύ σύντομο και είναι λογικό καθώς μια μικρομεσαία εταιρεία δεν διαθέτει πολλά τμήματα. Στις περιπτώσεις μεγάλων εταιρειών και οργανισμών το σύστημα απαιτεί από ένα έως τρία χρόνια για την υλοποίηση του. Η επιτυχημένη εφαρμογή του απαιτεί τον επαναπροσδιορισμό του τρόπου εργασίας του προσωπικού των στόχων και της μεθοδολογίας που θέλει να εφαρμόσει η εταιρεία. Σύμφωνα με τον (T. Wailgum, 2013) σημασία δεν έχει πόσο χρόνο θα διαρκέσει η τελική υλοποίηση του αλλά η κατανόηση της σημαντικότητας του από το προσωπικό της εταιρείας.



Εικόνα 13 Η συνδρομή του ERP στην επίτευξη των εταιρικών στόχων (Π. Παναγιωτόπουλος, 2007)

2.7 Προσδοκώμενα Οφέλη από την Εφαρμογή του Συστήματος

Με την εφαρμογή του συστήματος μια εταιρεία προσπαθεί να καλύψει τις κυριότερες διοικητικές τις ανάγκες. Αυτές αναφέρονται στην βιβλιογραφία ως εξής :

- Ανάγκη αποδοτικότερης διαχείρισης πόρων
- Ανάγκη μείωσης κόστους λειτουργίας
- Ανάγκης τήρησης μικρών αποθεμάτων πρώτων υλών
- Ανάγκη συντονισμού όλων των τμημάτων της
- Ανάγκη μείωσης λαθών
- Ανάγκη καλύτερης διαχείρισης προβλημάτων

Για την κάλυψη των αναγκών που αναφέρθηκαν το σύστημα θα πρέπει να πληρεί τις εξής προδιαγραφές:

- Ευέλικτη δομή που να επιτρέπει τις παραμετρικές αλλαγές (ιδιαίτερα χρήσιμο για τα συστήματα εξατομίκευσης)
- Αρθρωτή και ανοιχτή δομή ώστε να δουλεύει σε όσο το δυνατόν περισσότερες πλατφόρμες
- Υποστήριξη παραγωγής διαφορετικών προϊόντων. Κυρίως αναφέρεται ως προς την ροή εργασιών που απαιτεί η παραγωγή κάθε προϊόντος.
- Εφαρμογή της στρατηγικής just in time. Η στρατηγική JIT σύμφωνα με τον ορισμό που τις αποδίδει η Wikipedia⁵ προσπαθεί να επιστρέψει την επένδυση μέσω της μείωσης (carrying costs)⁶ των εξόδων αποθήκευσης, χρόνου κατάληψης της αποθήκης και μεταφοράς
- Υποστήριξη on-line διασύνδεσης με άλλα συστήματα, ιδιαίτερα στην περίπτωση ανάπτυξης ηλεκτρονικού καταστήματος που μας ενδιαφέρει

2.8 Σχέση ERP με E-Commerce

Τα συστήματα ERP αρχικώς δεν είχαν σχεδιάσει στα πρότυπα του ηλεκτρονικού εμπορίου. Θεωρούνταν πως οι μοναδικοί ενδιαφερόμενοι για την χρήση των συστημάτων θα ήταν οι υπεύθυνοι των επιχειρήσεων που τα χρησιμοποιούσαν για την ορθολογική διοίκηση των εταιρειών τους. Στην συνέχεια διαπιστώθηκε πως οι πελάτες και οι προμηθευτές απαιτούσαν πρόσβαση στις ίδιες πληροφορίες, πληροφορίες όπως κατάσταση και στάδιο παραγγελίας – με μόνη διαφορά ότι ήθελαν την πληροφορία απλουστευμένη χωρίς τις δευτερεύουσες που πλαισιώνουν την ζητούμενη.

Ένα από τα σημαντικότερα προβλήματα που αντιμετωπίζει η σύζευξη των δύο τεχνολογιών είναι ότι το διαδίκτυο δεν σταματάει ποτέ ενώ τα συστήματα ERP λόγω της πολυπλοκότητας τους απαιτούν αρκετή προεργασία για την διατήρηση της

⁵ http://en.wikipedia.org/wiki/Just_in_time_%28business%29

⁶ http://en.wikipedia.org/wiki/Carrying_cost

αποτελεσματικότητας τους. Στις περιπτώσεις όπου το σύστημα ERP είναι απευθείας συνδεδεμένο με την ηλεκτρονική σελίδα, θα πρέπει να «κατεβεί» για λίγες ώρες η σελίδα ώστε να γίνει συντήρηση του ERP. Όμως οι πιο έμπειροι σχεδιαστές δεν συνδέουν το σύστημα με τη σελίδα άμεσα, αλλά μέσω υπερσυνδέσεων, ώστε η συντήρηση του πρώτου να μην επηρεάζει την λειτουργία της δεύτερης.

Για την αντιμετώπιση του προβλήματος οι εταιρείες υλοποίησαν μια εναλλακτική εφαρμογή. Δημιουργήσανε ένα υποπρόγραμμα το οποίο λαμβάνει τις πληροφορίες από το ERP και τις μετατρέπει σε μορφή που να μπορούν να «διαβαστούν» από την εφαρμογή που τρέχει την ηλεκτρονική σελίδα. (. Markus Jessenitschnig and Markus Zanker, 2013)

Στα πλαίσια της έρευνας εξετάστηκαν 4 εταιρείες που δραστηριοποιούνται και στον τομέα σχεδιασμού συστημάτων ERP. Οι εν λόγω εταιρείες είναι οι εξής τέσσερις

Altec – Ολοκληρωμένο λογισμικό πακέτο εφαρμογών Atlantis για μικρομεσαίες επιχειρήσεις της εταιρείας Altec. (Τσόπογλου Σ., 2013)

Entersoft - Ολοκληρωμένο λογισμικό πακέτο εφαρμογών Entersoft για μικρομεσαίες επιχειρήσεις της εταιρείας Entersoft. Βασίζεται εξολοκλήρου στην πλατφόρμα .NET της Microsoft.

Singular - Ολοκληρωμένο λογισμικό πακέτο εφαρμογών Enterprise για μικρομεσαίες επιχειρήσεις της εταιρείας SingularLogic. Όσον αφορά τα τεχνικά χαρακτηριστικά το Enterprise ERP αποτελείται από ένα ενιαίο περιβάλλον όπου τα υποσυστήματα επικοινωνούν αυτόματα μεταξύ τους χωρίς την προσθήκη ετερογενών εφαρμογών.

Microsoft - Ολοκληρωμένο λογισμικό πακέτο εφαρμογών Navision της εταιρείας Microsoft. Το λογισμικό δίνει την δυνατότητα της μηχανογράφησης όλων των διαδικασιών της επιχείρησης. (Σάββας Ι. & Μαυρέλλης Ν., 2005)

Χαρακτηριστικά	Atlantis Altec	Entersoft Entersoft	Enterprise Singular	Navision Microsoft
Υποσυστήματα και Λειτουργίες	<p>Λογιστική</p> <p>Χρηματοοικονομικά</p> <p>Αποθέματα</p> <p>Πωλήσεις και Διανομές</p> <p>Αγορές</p> <p>CRM</p> <p>Παραγωγή</p> <p>Service</p> <p>MRP</p> <p>Business Intelligence</p> <p>Μισθοδοσία & HRM</p> <p>Enterprise Web Modules</p> <p>Tools</p> <p>Auto</p>	<p>Οικονομική Διαχείριση</p> <p>Γενική Λογιστική</p> <p>Διεθνή Λογιστικά Πρότυπα</p> <p>Προϋπολογισμός και Έλεγχος</p> <p>Χρηματοοικονομική Διαχείριση</p> <p>Διαχείριση Εισπρακτέων και Πληρωτέων</p> <p>Διαχείριση Αποθεμάτων και Αποθηκών</p> <p>Διαχειριση Πωλήσεων</p> <p>Διαχείριση Αγορών και Προμηθειών</p> <p>Διαχείριση Παραγωγής</p> <p>Διαχείριση Έργων</p>	<p>Παρακολούθηση συναλλασόμενων</p> <p>Διαχείριση αποθήκης</p> <p>Παραγγελιοληψία – πωλήσεις</p> <p>Παραγγελιοδοσία – αγορές</p> <p>Χρηματοοικονομική διαχείριση (Εισπράξεις Πληρωμές Αξιόγραφα)</p> <p>Βιβλία Α' Β' Γ' κατηγορίας,</p>	<p>Οικονομική Διαχείριση</p> <p>Πωλήσεις και Μάρκετινγκ</p> <p>Διαχείριση Εφοδιαστικής Αλυσίδας</p> <p>Παραγωγή</p> <p>Διαχείριση Τεχνικής Υποστήριξης</p> <p>Επιχειρηματική Ευφυΐα και Εκτυπώσεις</p> <p>Διχειριση Έργων</p> <p>Διαχειριση Ανθρώπινων Πόρων</p>
Ενιαίο Λογισμικό για όλες τις Επιχειρήσεις	<p>Με την επιλογή [ΕΤΑΙΡΕΙΕΣ] ο διαχειριστής του προγράμματος ελέγχει τα στοιχεία των εταιρειών που παρακολουθεί με το ATLANTIS. Κάθε εταιρεία έχει ξεχωριστά δεδομένα που κρατούνται σαν διαφορετικές οντότητες από το σύστημα.</p>			

	Εξαίρεση αποτελούν ορισμένα υπερεταιρικά στοιχεία που είναι κοινά σε όλες τις εταιρείες			
Δυνατότητα Παραμετροποίησης του Συστήματος	Παραμετροποιείται και προσαρμόζεται στα δεδομένα κάθε επιχείρησης	Παραμετροποιείται και προσαρμόζεται στα δεδομένα κάθε επιχείρησης	Παραμετροποιείται και προσαρμόζεται στα δεδομένα κάθε επιχείρησης	Παραμετροποιείται και προσαρμόζεται στα δεδομένα κάθε επιχείρησης
Συνολικό Κόστος	Το κόστος κυμαίνεται από 8.000 έως 30.000 με διάφορα επιπλέον κόστη π.χ. πολλαπλοί χρήστες κτλ.	Η τιμή του κάθε module ξεκινά από 1.000 μέχρι 5.000	Ξεκινά από 20000 και μπορεί να φτάσει μέχρι 50.000 ή 60.000	Από 37.000 έως 112.260
Δυνατότητα Σύνδεσης με άλλες Εφαρμογές (π.χ. Ηλεκτρονικό Εμπόριο)	Εξάγει στο διαδίκτυο και αυτοματοποιεί όλες τις συναλλαγές μεταξύ των συνεργαζομένων επιχειρήσεων, προσφέροντας ασφάλεια και επιτάχυνση των διεπιχειρησιακών διαδικασιών, ανεξάρτητα από το μοντέλο λειτουργίας (franchising, ιδιόκτητα δίκτυα καταστημάτων αντιπροσώπων κλπ)	Μέσω των XML documents αλλά και μέσω των XML WEB Services επιτρέπεται η on-line αποστολή και λήψη μηνυμάτων παραστατικών και άλλων πληροφοριών με άλλα πληροφοριακά συστήματα αλλά και η διεκπεραίωση διαιτητικών διαδικασιών	Υπάρχει το υποσύστημα Enterprise e-order που μπορεί να πραγματοποιήσει την παρακολούθηση των παραγγελιών μέσω ιντερνετ.	Συνδέει τον κάθε εργαζόμενο της επιχείρησης με πελάτες προμηθευτές συνεργάτες μέσω του internet, οποιαδήποτε στιγμή, οπουδήποτε
Τεχνολογία ERP	Χρησιμοποιεί την τεχνολογία ROADS, BASISM, ENO STO RAD Studio XE, με αρχιτεκτονική multitier, client server	Το User Interface εκμεταλλεύεται τις δυνατότητες που παρέχει το Microsoft NET για την ανάπτυξη Windows εφαρμογών	Αξιοποιεί όλες τις νέες τεχνολογίες στον τομέα της πληροφορικής, υποστηρίζοντας την αρχιτεκτονική client-server 3tier, το σχεσιακό σύστημα διαχείρισης βάσης δεδομένων ORACLE διαθέτει γραφικό περιβάλλον εργασίας	Χρησιμοποιεί την πολυετή τεχνολογία της Microsoft και τα Microsoft Productivity Tools

2.9 Η Αρχιτεκτονική ενός ERP

Ένα ERP σύστημα μπορεί να βασίζεται σε Three Tier αρχιτεκτονική και να λειτουργεί σε δίκτυο υπολογιστών με χρήση του πρωτοκόλλου επικοινωνίας TCP/IP. Το σύστημα συνήθως αποτελείται από τρία διακριτά επίπεδα: (Z. I. Magabe, 2006)

- Database Server
- Application Server
- Client

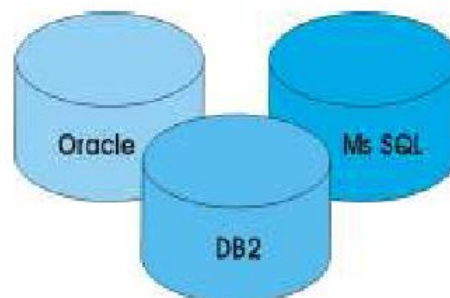
Εφαρμογές που έχουν βασιστεί στην αρχιτεκτονική 3-tier αποκαλούνται συχνά server-centric, διότι επιτρέπουν στα κομμάτια της εφαρμογής (components) που ανήκουν στο business tier να τρέχουν σε ειδικούς servers που είναι εντελώς ανεξάρτητοι από το user interface ή την υλοποίηση των βάσεων δεδομένων που ανήκουν στο data tier. Αυτή η απεξάρτηση της λογικής της εφαρμογής από το presentation tier και το data tier επιφέρει πολλά πλεονεκτήματα. Μερικά από αυτά τα πλεονεκτήματα συνοψίζονται στον παρακάτω πίνακα.

Υποστήριξη πολλαπλών προγραμματιστικών γλωσσών	Τα components μιας εφαρμογής μπορούν να γραφτούν άνετα σε διαφορετικές προγραμματιστικές γλώσσες. Θα μπορούσε, παραδείγματος χάριν, η λογική της εφαρμογής να υλοποιηθεί σε C++, ενώ το user interface θα μπορούσε να υλοποιηθεί σε Visual Basic.
Κεντριοποιημένα Components	Τα components όλων των εφαρμογών μπορούν να τρέχουν σε κεντρικούς διακομιστές (servers), έτσι ώστε να γίνεται πολύ απλούστερη η διαδικασία συντήρησης ή αλλαγής τους.
Load Balancing	Τα components μιας εφαρμογής μπορούν να μοιραστούν σε πολλαπλούς servers στο δίκτυο, επιτρέποντας έτσι την εξισσορόπιση του φόρτου εργασίας του κάθε server.
Καλύτερη πρόσβαση στα δεδομένα	Περιορισμοί που πιθανόν να υπήρχαν στην σύνδεση πολλαπλών χρηστών στην βάση δεδομένων εκμηδενίζονται, καθώς η βάση

	<p>‘μιλάει’ πλέον μόνο με κάποια συγκεκριμένα components.</p> <p>Επίσης, οι οδηγοί και η σύνδεση για την βάση δεδομένων δεν χρειάζεται πλέον να βρίσκονται τοπικά στο μηχάνημα του χρήστη.</p> <p>Οι πόροι του συστήματος χρησιμοποιούνται πολύ καλύτερα, καθώς στην αρχιτεκτονική 3-tier οι συνδέσεις προς την βάση δεδομένων γίνονται μόνο όταν χρειάζονται και απελευθερώνονται αμέσως μόλις εκπληρώσουν τον σκοπό τους.</p>
Βελτιωμένη ασφάλεια	<p>Components που ανήκουν στο Business Tier μπορούν πλέον να διασφαλιστούν χρησιμοποιώντας κάποια κεντροποιημένη υποδομή. Η πρόσβαση στο κάθε component μπορεί να δοθεί ή να αρνηθεί ξεχωριστά, διευκολύνοντας το ρόλο των Administrators.</p>
Απλοποιημένη πρόσβαση σε εξωτερικούς πόρους	<p>Η πρόσβαση σε εξωτερικές εφαρμογές όπως είναι εφαρμογές που τρέχουν πάνω σε mainframes που συλλέγουν δεδομένα απλοποιείται σημαντικά. Ένας gateway server π.χ., θεωρείται απλά άλλο ένα component που χρησιμοποιεί η εφαρμογή.</p>

2.9.1 Database Server

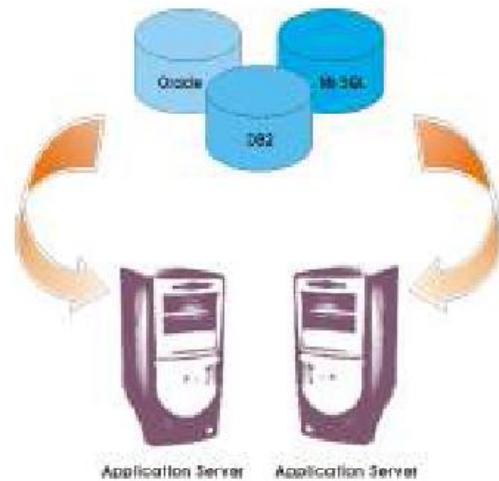
Αποτελώντας το βασικότερο επίπεδο του συστήματος, ο Database Server παρέχει όλες τις απαραίτητες λειτουργίες για την αποθήκευση, ανάκτηση, ενημέρωση και συντήρηση των δεδομένων του συστήματος καθώς επίσης και όλους τους απαραίτητους μηχανισμούς για την ακεραιότητα των δεδομένων. (Altec Software, 2008)



2.9.2 Application Server

Αποτελεί το κύριο τμήμα του λογισμικού, στο οποίο εκτελούνται οι περισσότερες λειτουργίες, εκτός εκείνων που σχετίζονται με τη διαμόρφωση των οθονών εργασίας. Υπάρχει δυνατότητα εγκατάστασης περισσότερων του ενός Application Server σε διαφορετικά μηχανήματα, αξιοποιώντας, με τον τρόπο αυτό, οποιαδήποτε διαθέσιμη υπολογιστική ισχύ και εξασφαλίζοντας εξαιρετικά αποτελέσματα ανταπόκρισης, αξιοπιστίας και επεκτασιμότητας.

Με την κατανομή των Application Servers σε ανεξάρτητα μηχανήματα, επιτυγχάνεται αποσυμφόρηση του συνολικού φόρτου του συστήματος, αφού κάθε Application Server είναι σε θέση να υποστηρίξει ένα υποσύνολο του συνολικού αριθμού των χρηστών. (Altec Software, 2008)



2.9.3 Client

Το τρίτο επίπεδο του λογισμικού αποτελεί τη επαφή του χρήστη με το σύστημα (User Interface). Στο επίπεδο αυτό, πραγματοποιείται η διαχείριση των Οθονών Εργασίας (User Screens) καθώς επίσης και η μορφοποίηση των δεδομένων που εμφανίζονται. Η επικοινωνία του Client με τον Application ή τους Application Servers πραγματοποιείται κάνοντας χρήση ενός μόνο πακέτου δεδομένων κάθε φορά.

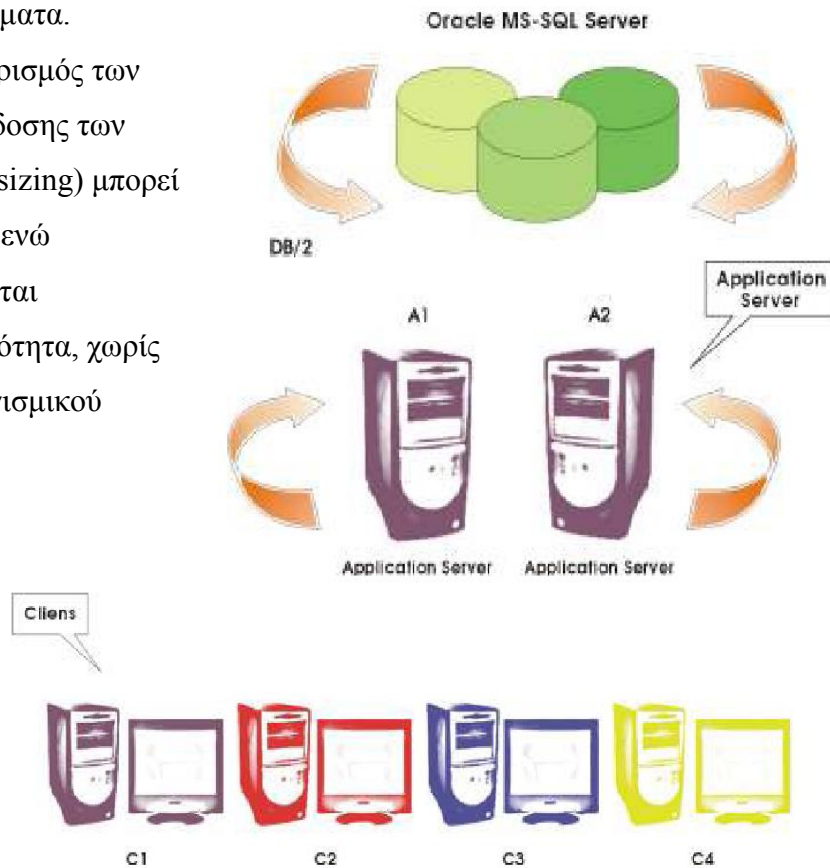
Έτσι, επιτυγχάνεται ο βέλτιστος χρόνος απόκρισης μεταξύ του Client και του Application Server, δεδομένου ότι τα δυο αυτά επίπεδα μπορούν να λειτουργήσουν πάνω σε μια τηλεπικοινωνιακή γραμμή (Leased Line, Dialup, Internet Connection, EDI), εξασφαλίζοντας έτσι μικρούς χρόνους απόκρισης σε όλο το σύστημα. Η αρχιτεκτονική Client -Server τριών επιπέδων (Three Tier) έχει διεθνώς αποδειχθεί ως η πλέον κατάλληλη για δικτυακές εγκαταστάσεις, σε αντίθεση με την αρχιτεκτονική Client -Server δύο επιπέδων (Two Tier). (Altec Software, 2008)

Η συγκρότηση του συστήματος σε τρία επίπεδα εξασφαλίζει την ελαχιστοποίηση της επιβάρυνσης του δικτύου λόγω μεταφοράς μεγάλου όγκου δεδομένων π.χ. η εκτέλεση

ενός Query (αναζήτηση) για την ανάκτηση μερικών εγγραφών από έναν πίνακα με δεκάδες χιλιάδες εγγραφές γίνεται στο διακομιστή εφαρμογής (Application Server), από τον οποίο μεταφέρεται στο χρήστη μόνο το αποτέλεσμα

Τη δυνατότητα διαχωρισμού του διακομιστή δεδομένων (Data Server) από το διακομιστή ή τους διακομιστές εφαρμογής (Application Server) ώστε να εκτελούνται σε διαφορετικά μηχανήματα.

Κατά συνέπεια, ο καθορισμός των κρίσιμων μεγεθών απόδοσης των αντίστοιχων μηχανών (sizing) μπορεί να γίνεται ανεξάρτητα, ενώ παράλληλα εξασφαλίζεται απεριόριστη επεκτασιμότητα, χωρίς ανακατασκευή, του λογισμικού



3. ΣΥΝΕΡΓΑΤΙΚΗ ΔΙΗΘΗΣΗ ΔΕΔΟΜΕΝΩΝ

3.1 Φιλτράρισμα Πληροφοριών

Τα συστήματα ανάκτησης πληροφοριών επιτρέπουν στους χρήστες να εκφράσουν ερωτήσεις για να επιλέξουν στοιχεία που ταιριάζουν με ένα συγκεκριμένο θέμα και παράλληλα επιτρέπουν στις εταιρείες να συλλέξουν πληροφορίες σχετικά με το προφίλ του κάθε χρήστη. Οι τεχνικές ανάκτησης πληροφοριών ωστόσο δεν είναι αρκετά χρήσιμες στην πραγματική διαδικασία σύστασης, δεδομένου ότι δεν συλλαμβάνουν καμία πληροφορία για τις προτιμήσεις των χρηστών πέραν της συγκεκριμένης ερώτησης. Για την αποτελεσματικότερη προσέγγιση των πελατών εφαρμόζονται μέθοδοι φιλτραρίσματος των πληροφοριών, τα χαρακτηριστικά των οποίων παρουσιάζονται στον επόμενο πίνακα.

Δημογραφικό Φιλτράρισμα	Οι δημογραφικές προσεγγίσεις φιλτραρίσματος χρησιμοποιούν τις περιγραφές των χρηστών για να μάθουν τη σχέση μεταξύ ενός στοιχείου και του τύπου ανθρώπων που τους αρέσει. Τα προφίλ χρηστών δημιουργούνται με την ταξινόμηση των χρηστών σε στερεοτυπικές περιγραφές, που αντιπροσωπεύουν τα χαρακτηριστικά γνωρίσματα των κατηγοριών των χρηστών. Τα προσωπικά στοιχεία του χρήστη είναι απαραίτητα και χρησιμοποιούνται για την ταξινόμηση. Οι ταξινομήσεις χρησιμοποιούνται ως γενικοί χαρακτηρισμοί για τους χρήστες και τα ενδιαφέροντά τους. Συνήθως, τα προσωπικά στοιχεία του χρήστη λαμβάνονται κατά την αίτηση εγγραφής στο σύστημα (φύλο, επάγγελμα, ηλικία, τρόπος ζωής κ.α.). (X. Χριστάκου, 2012)
Φιλτράρισμα Περιεχομένου	Το φιλτράρισμα περιεχομένου επιλέγει τις σωστές πληροφορίες για τους χρήστες με τη σύγκριση της αναπαράστασης της πληροφορίας αναζήτησης με την αναπαράσταση του περιεχομένου των παραμέτρων χρήστη που εκφράζει τα ενδιαφέροντά του (profile). Το φιλτράρισμα περιεχομένου πληροφοριών έχει αποδειχθεί

	<p>αποτελεσματικό στην εντόπιση κειμενικών στοιχείων σχετικών με ένα θέμα χρησιμοποιώντας τεχνικές, όπως Boolean ερωτήσεις, ερωτήσεις διανυσματικού χώρου, το πιθανολογικό πρότυπο, τα νευρωνικά δίκτυα και το μοντέλο ασαφών συνόλων.</p>
<p>Φιλτράρισμα με βάση την Γνώση</p>	<p>Τα συστήματα αυτού του είδους στηρίζονται για τις συστάσεις που κάνουν σε συγκεκριμένη γνώση η οποία καθορίζει κατά πόσο τα χαρακτηριστικά ενός προϊόντος ανταποκρίνονται στις ανάγκες και τα ενδιαφέροντα του χρήστη, δηλαδή αν το προϊόν θα είναι χρήσιμο στον χρήστη ή όχι. Το σύστημα συγκεντρώνει τα αιτήματα του χρήστη και προτείνει και εξηγεί τις συστάσεις που βρίσκει ως λύση. Η συνάρτηση ομοιότητας στα Συστήματα Προτάσεων με βάση την γνώση εκτιμά πόσο οι ανάγκες του χρήστη συσχετίζονται με τις συστάσεις και έτσι τελικά δείχνει την χρησιμότητα της σύστασης για τον ενδιαφερόμενο.</p>
<p>Συνεργατικό Φιλτράρισμα (Διήθηση) Δεδομένων</p>	<p>Στην συνεργατική προσέγγιση αντί να συστηθούν στοιχεία επειδή είναι παρόμοια με τα στοιχεία που ένας χρήστης επιδοκίμασε στο παρελθόν, συστήνονται στοιχεία που άλλοι χρήστες με γειτονικό προφίλ έχουν συμπαθήσει. Αντί δηλαδή να υπολογιστεί η ομοιότητα των προϊόντων, υπολογίζεται η ομοιότητα των πελατών. Χαρακτηριστικά, για κάθε χρήστη βρίσκεται ένα σύνολο «πλησιέστερων χρηστών γειτόνων» με των οποίων τις μέχρι τώρα εκτιμήσεις υπάρχει ο ισχυρότερος συσχετισμός. Τα αποτελέσματα για τα άγνωστα στοιχεία προβλέπονται με βάση συνδυασμό αποτελεσμάτων που είναι γνωστά από τους «πλησιέστερους γείτονες».</p>
<p>Υβριδικό Φιλτράρισμα</p>	<p>Η κατηγορία αυτή συστημάτων χρησιμοποιεί ένα συνδυασμό των μεθόδων που αναφέραμε παραπάνω, εκμεταλλευόμενα τα προτερήματα τις μίας τεχνικής για να καλύψουν τα μειονεκτήματα της άλλης. Υπάρχουν πολλοί διαφορετικοί τρόποι με τους οποίους συνδυάζονται δύο ή και περισσότερες τεχνικές συστημάτων προτάσεων για να δημιουργηθεί ένα υβριδικό σύστημα. Στόχος του συνδυασμού διαφορετικών μεθόδων είναι η βελτίωση της απόδοσής</p>

Σχετικά όμως με τις δύο πρώτες κατηγορίες φιλτραρίσματος εντοπίζονται κάποια σημαντικά μειονεκτήματα, τα όποια φαίνεται να υπερπηδήσει το συνδυαστικό φιλτράρισμα. Συγκεκριμένα το δημογραφικό σύστημα φιλτραρίσματος έχει δύο βασικά μειονεκτήματα:

- Είναι βασισμένο σε μια γενίκευση των ενδιαφερόντων του χρήστη, έτσι ώστε το σύστημα συστήνει τα ίδια στοιχεία στους ανθρώπους με παρόμοια δημογραφικά προφίλ. Δεδομένου ότι κάθε χρήστης είναι διαφορετικός, αυτές οι συστάσεις αποδεικνύονται πάρα πολύ γενικές.
- Οι δημογραφικές προσεγγίσεις δεν παρέχουν οποιαδήποτε μεμονωμένη προσαρμογή στις αλλαγές ενδιαφέροντος. Τα ενδιαφέροντα του χρήστη τείνουν να αλλάζουν με το πέρασμα του χρόνου, και έτσι οι παράμετροι του χρήστη πρέπει να προσαρμόζονται στην αλλαγή.

Αντίστοιχα για το φιλτράρισμα περιεχομένου ενώ ορθώς είναι βασισμένο στις αντικειμενικές πληροφορίες του προϊόντος πολλές φορές, η επιλογή κάποιου στοιχείου βασίζεται σε ένα μεγάλο ποσοστό στις υποκειμενικές ιδιότητες του στοιχείου. Για παράδειγμα στα έγγραφα κειμένων οι αντιπροσωπεύσεις συλλαμβάνουν μόνο ορισμένες πτυχές του περιεχομένου, ενώ υπάρχουν πολλές άλλες που θα επηρέαζαν την εμπειρία ενός χρήστη. Για ιστοσελίδες, παραδείγματος χάριν, οι τεχνικές φιλτραρίσματος περιχεόμενου αγνοούν εντελώς τις αισθητικές ιδιότητες, και τους παράγοντες δικτύων όπως ο χρόνος φόρτωσης. Ακόμα, είναι δύσκολο να προέρχονται οι συστάσεις από ένα ευρύ φάσμα θεμάτων, επειδή όλες οι πληροφορίες επιλέγονται και συστήνονται βασισμένες στο περιεχόμενο.

3.2 Ορισμός Συνεργατικής Διήθησης Δεδομένων

Με τον όρο collaborative filtering, ο οποίος στα Ελληνικά αποδίδεται ως «συνεργατική διήθηση δεδομένων», περιγράφεται η διαδικασία της απόρριψης ή αποδοχής κάποιων δεδομένων σε σχέση με κάποια άλλα. Η διαδικασία πραγματοποιείται μέσω υπολογιστή και με χρήση τεχνικών που απαιτούν την συνεργασία παραγόντων όπως είναι τα αποθηκευτικά μέσα, οι απόψεις των χρηστών, οι πηγές πληροφόρησης κλπ.

Ο όρος αναφέρεται κυρίως στα δεδομένα χρηστών του διαδικτύου, καθώς σ' αυτό υπάρχουν υπερβολικά μεγάλες συγκεντρώσεις πηγών πληροφορίας.

Για να γίνει περισσότερο κατανοητή η σημασία του όρου, πρέπει να σημειωθεί ότι στην καθημερινή ζωή, οι άνθρωποι προκειμένου να κάνουν μια επιλογή, βασίζονται σε συστάσεις ή προτροπές άλλων ανθρώπων μέσω του προφορικού λόγου, συστατικών επιστολών, ειδησεογραφικών αναφορών από τα μέσα ενημέρωσης, γενικών ερευνών, τουριστικών οδηγών και ούτω καθεξής. Έτσι, έχουν αναπτυχθεί παρόμοια ηλεκτρονικά συστήματα συστάσεων, τα οποία προτείνουν την καταλληλότερη για το χρήστη πληροφορία. Τα ηλεκτρονικά συστήματα συστάσεων, ενισχύουν και αυξάνουν αυτή τη φυσική κοινωνική διαδικασία βοηθώντας τους ανθρώπους να διακρίνουν ανάμεσα στα διαθέσιμα βιβλία, άρθρα, ιστοσελίδες, ταινίες, μουσική, εστιατόρια ή λίστες ανεκδότητων ώστε να επιλέξουν τις πιο ενδιαφέρουσες και αξιόλογες για τους ίδιους πληροφορίες.

Ουσιαστικά, λοιπόν, πρόκειται για την μέθοδο παραγωγής μιας αυτόματης πρόβλεψης (φιλτράρισμα, διήθηση) προς όφελος του χρήστη με συλλογή πληροφοριών για τις προτιμήσεις άλλων χρηστών (συνεργασία).

Να σημειωθεί ότι παρ' όλο που αυτές οι προβλέψεις απευθύνονται στον συγκεκριμένο χρήστη χρησιμοποιούν πληροφορίες προερχόμενες από πολλούς άλλους. Αυτό διαφέρει από την πιο απλή προσέγγιση της παροχής μιας μέσης βαθμολογίας για το αντικείμενο ενδιαφέροντος, όπως για παράδειγμα με βάση τον αριθμό των θετικών ψήφων που παρέχονται σε ένα ερωτηματολόγιο.

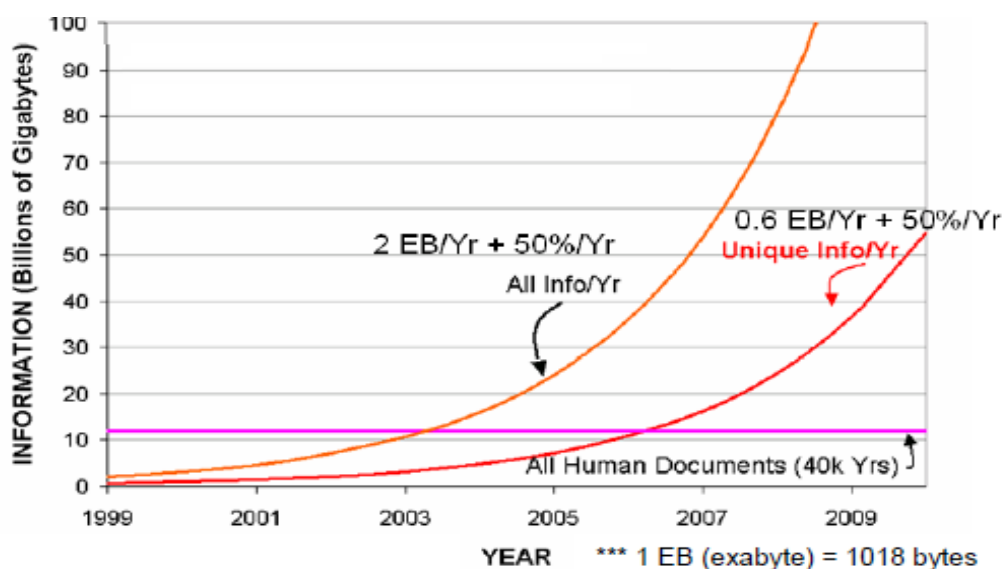
Ο όρος "collaborating filtering" επινοήθηκε από τους προγραμματιστές ενός από πρώτα συστήματα συστάσεων, του Tapestry και έκτοτε έχει υιοθετηθεί ευρέως, ανεξαρτήτως του ότι οι "συστήνοντες" δεν "συνεργάζονται" στην πραγματικότητα ούτε μεταξύ τους ούτε με τους αποδέκτες. Από την άλλη, ενώ τα αποτελέσματα μπορεί να υποδεικνύουν εξαιρετικά ενδιαφέρουσες επιλογές για το χρήστη, μπορεί

όμως να περιέχουν και προτάσεις που θα έπρεπε τελικά να φιλτράρονται και να μην εμφανίζονται. (Δ. Παρούτσας, 2013)

3.3 Αναγκαιότητα Συστημάτων

Πλέον τα συστήματα καταγραφής και η διαχείριση δεδομένων από ηλεκτρονικές εφαρμογές, καλούνται να εφαρμόσουν νέες τεχνικές καθώς τα δεδομένα αυξάνονται ραγδαία και σε πολλές περιπτώσεις καταρρίπτονται καθιερωμένες μέθοδοι επεξεργασίας τους. Παράλληλα η αυξημένη χρήση των web υπηρεσιών δημιούργησε την ανάγκη εμφάνισης προσωποποιημένων συστάσεων στους χρήστες. Αποτέλεσμα αυτής της εξέλιξης είναι η αναζήτηση νέων μεθόδων διαχείρισης του μεγάλου όγκου πληροφοριών στα πλαίσια παροχής εξατομικευμένων υπηρεσιών στους χρήστες.

Πίνακας 8 Ποσότητες παραγόμενων πληροφοριών σε παγκόσμιο επίπεδο (Μ. Ρήγκου, 2012)



Το Collaborative Filtering είναι μια τεχνική που χρησιμοποιείται κυρίως από συστήματα που κάνουν στον χρήστη προτάσεις επιλογών. Συστήματα που έχουν επεξεργαστεί τις επιλογές και τις συνήθειες των προηγούμενων χρηστών και με τη μορφή προτιμήσεων ή βαθμολογήσεων που δίνουν οι χρήστες και μέσω μιας μηχανής, προσεγγίζουν τις προτιμήσεις των νέων χρηστών. Λόγω του ότι στηρίζεται στην εμπειρία που αποκτά από τη συμπεριφορά των χρηστών, συνεχώς βελτιώνονται τα αποτελέσματα του. Το σύστημα ταιριάζει τις συμπεριφορές των χρηστών και μόλις εντοπίσει κοινά στοιχεία των νέων χρηστών με τους παλιότερους κάνει σύσταση για το πιθανώς ο νέος πελάτης χρειάζεται. (Χ. Χριστάκου, 2012)

Η πολυπλοκότητα του συστήματος έγκειται στο γεγονός ότι συγκεντρώνει φαινομενικά ασύνδετες πληροφορίες και από τον συνδυασμό τους πιθανολογεί το αποτέλεσμα. Αν βρει άλλους χρήστες με τις ίδιες καταναλωτικές συμπεριφορές τότε ελέγχει τι δεν καταναλώθηκε και γίνεται σύσταση.

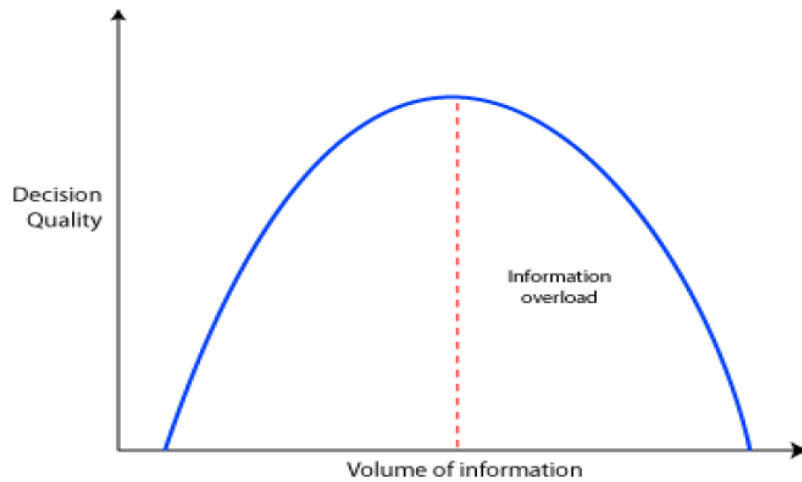
Επίσης χαρακτηριστικό στοιχείο του συστήματος είναι ότι όσο περισσότερο το χρησιμοποιεί ο ίδιος χρήστης τόσο περισσότερο πιο εξατομικευμένη γίνεται η σελίδα προς αυτόν. Σύμφωνα με τα ως άνω προκύπτει πως η εφαρμογή ισχυροποιείται όταν :

- Οι χρήστες που την χρησιμοποιούν γίνονται περισσότεροι
- Το σύστημα αποκτάει εμπειρία για τον εκάστοτε χρήστη

Στα προβλήματα που αντιμετωπίζει η εφαρμογή τοποθετείται ότι δεν μπορεί αυτομάτως και εξ αρχής να κατηγοριοποιήσει τον χρήστη. Χρειάζεται να αποκτήσει εμπειρία μέσα από την χρήση της σελίδας από τον χρήστη. Πέραν αυτού όμως οι επιλογές του νέου χρήστη μπορεί να προκαλέσουν επιπλοκές στο σύστημα και να αποπροσανατολίσουν την κατεύθυνση του συστήματος.

3.4 Στόχος Συστήματος

Από την εφαρμογή της τεχνολογίας CF προσδοκάτε η δημιουργία μιας μηχανής αναζήτησης η οποία θα φιλτράρει τα δεδομένα και θα αποδίδει στον τελικό χρήστη προσωποποιούμενες προτάσεις. Δηλαδή στοχεύει από ένα τεράστιο όγκο πληροφοριών να είναι σε θέση να επιλέξει μόνο τις επιθυμητές από τον χρήστη. Το πρόβλημα που καλείται να λύσει η εφαρμογή είναι υπαρκτό καθώς εκατομμύρια χρήστες επισκέπτονται συγκεκριμένα sites με χιλιάδες προϊόντα (Amazon), ή με εκατομμύρια αναρτήσεις (Google News). Το πρόβλημα γιγαντώνεται από την συνεχή ροή νέων πληροφοριών που εισάγονται, είτε από την πλευρά των χρηστών είτε από των προσφερόμενων υπηρεσιών, (προϊόντα, ειδήσεις, κλπ)

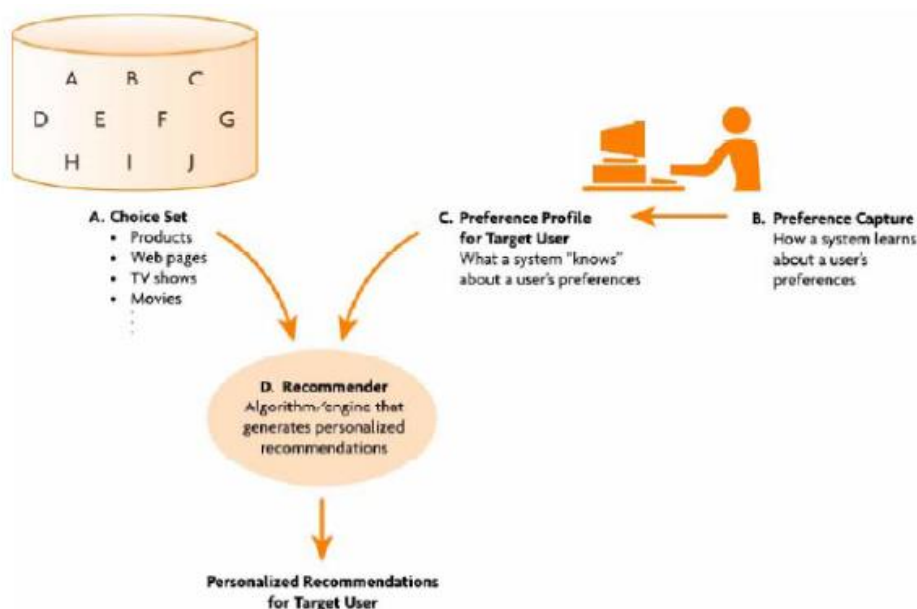


Πίνακας 9 Διαγραμματική απεικόνιση του φαινομένου Information Overload όπου λόγω του όγκου των πληροφοριών οι χρήστες χάνουν συχνά το στόχο τους (Μ. Ρήγκου, 2012)

Στην εποχή που διανύουμε η οποία χαρακτηρίζεται από την έκρηξη των πληροφοριών η τεχνική το CF μπορεί να αποδειχθεί πολύ χρήσιμη, καθώς ο αριθμός των αντικειμένων σε μία μόνο κατηγορία (μουσική, ταινίες, βιβλία, ειδήσεις, ιστοσελίδες) έχει γίνει τόσο μεγάλος, ώστε ένα άτομο δεν δυνατό να τον προσπελάσει, προκειμένου να επιλέξει αυτά που τον ενδιαφέρουν. Αν η τεχνική στηριζόταν μόνο σε ένα σύστημα βαθμολόγησης το οποίο εντοπίζει το μέσο όρο για όλους τους χρήστες τότε θα αγνοούσε τις απαιτήσεις του συγκεκριμένου χρήστη, και θα ήταν ιδιαίτερα φτωχή σε περιπτώσεις όπου υπάρχει μεγάλη διακύμανση ενδιαφέροντος, όπως για παράδειγμα η πρόταση για συγκεκριμένο είδος μουσικής.

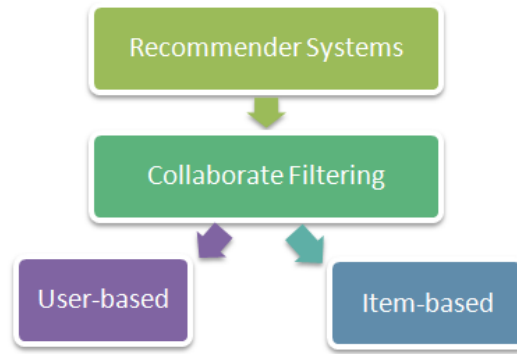
3.5 Γενική Δομή Συστημάτων Collaborate Filtering

Οι τεχνικές CF χρησιμοποιούν βάσεις δεδομένων που προκύπτουν από τις επιλογές των χρηστών, ώστε να γίνουν προβλέψεις για προϊόντα που πιθανώς να χρειαστούν νέοι πελάτες με το ίδιο προφίλ. Σε ένα τυπικό σενάριο υπάρχει μια λίστα με χ χρήστες ($\chi_1, \chi_2, \dots, \chi_n$) και μια λίστα με ψ προϊόντα ($\psi_1, \psi_2, \dots, \psi_n$) και κάθε χρήστης (χ_n) έχει μια λίστα με προϊόντα (Ψ_{χ_n}), τα οποία ο χρήστης έχει βαθμολογήσει. Η βαθμολόγηση μπορεί να είναι της κλίμακας από 1 έως 5 είτε μέσω ενδείξεων από τις οποίες ο χρήστης έχει να επιλέξει. Εντούτοις είναι πολύ πιθανό να μην υπάρχουν πάντα δεδομένα αξιολογήσεων αλλά αντίθετα να υπάρχουν διάφορα δυαδικά στοιχεία. (π.χ αν ένα αντικείμενο αγοράστηκε ή όχι.)



Εικόνα 14 Η δομή του συστήματος συνίσταται σε 4 στάδια

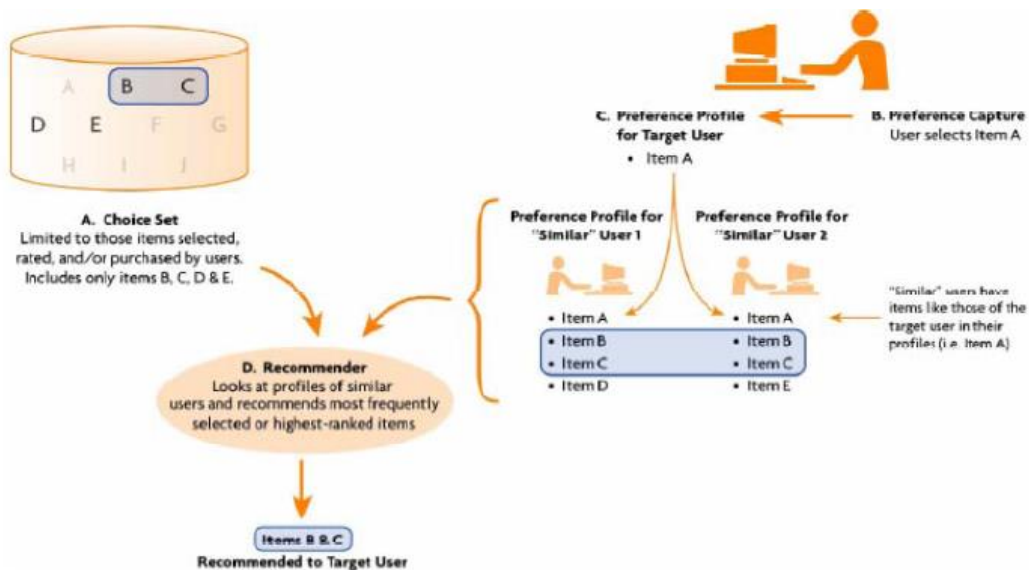
Στην συνέχεια η συνδυαστική σύγκριση της βαθμολόγησης του κάθε χρήστη για το κάθε προϊόν δίνει στο σύστημα την δυνατότητα να κάνει προβλέψεις για ένα καινούργιο χρήστη ο οποίος έχει δώσει στο σύστημα πληροφορίες σχετικά με τις προτιμήσεις του. Κατά μια έννοια το κλικ το κάθε χρήστη πάνω σε μια επιλογή (άρθρο, προϊόν, κλπ) αποτελεί μια ψήφο. Τα συστήματα Collaborate Filtering διακρίνονται σε Item-based και User-based.



Εικόνα 15 Διάκριση συστημάτων συνεργατικής διήθησης

3.6 User-based Collaborative Filtering

Ο όρος στα Ελληνικά αποδίδεται ως «συνεργατική διήθηση με βάση τους χρήστες». Στην κατηγορία αυτή ελέγχεται ποιοι χρήστες μοιράζονται τα ίδια πρότυπα αξιολόγησης με το χρήστη για τον οποίο προορίζεται η πρόβλεψη. Στη συνέχεια χρησιμοποιούνται αυτές οι αξιολογήσεις για τη δημιουργία μιας πρόβλεψης που απευθύνεται στο συγκεκριμένο άτομο.



Εικόνα 16 Η δομή του συστήματος συνίσταται σε 4 στάδια

3.5.1 Παράδειγμα Μελέτης

Ας θεωρήσουμε μια βάση δεδομένων 5 χρηστών με ιστορικό 4 προϊόντων. Εισάγεται ένα νέο προϊόν το οποίο έχουν χρησιμοποιήσει οι 4 εξ αυτών και επιχειρείται να γίνει πρόβλεψη για τον 5^ο χρήστη. Τα στοιχεία δίνονται στον ακόλουθο πίνακα

Πίνακας 10 Καταγραφή δεδομένων

	Προϊόν 1	Προϊόν 2	Προϊόν 3	Προϊόν 4	Προϊόν 5
Χρήστης 5	5	3	4	4	?
Χρήστης 4	3	1	2	3	3
Χρήστης 3	4	3	4	3	5
Χρήστης 2	3	3	1	5	4
Χρήστης 1	1	5	5	2	1

Τα ερωτήματα που δημιουργούνται είναι

- Πως θα μετρηθεί η ομοιότητα μεταξύ των χρηστών;
- Πόσους και ποιούς χρήστες θα πρέπει να συγκρίνουμε;
- Πως θα παράξουμε συμπεράσματα από την σύγκριση;

Για τον υπολογισμό της ομοιότητας των δεδομένων χρησιμοποιείται ο δείκτης ομοιότητας Jaccard ο οποίο λαμβάνει τιμές από 0 έως 1

$$sim(a, b) = \frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}}$$

Όπου

a, b : Οι συγκρινόμενοι χρήστες. Πάντα τοποθετούνται ζευγάρια μεταξύ αυτού που προσπαθεί να προσεγγιστεί και αυτού που ήδη γνωρίζουμε

$r_{a,p}$: βαθμολόγηση για το προϊόν a από τον χρήστη p

P : ο αριθμός των προϊόντων που βαθμολογήθηκαν και από τους δύο χρήστες

Για το παράδειγμα μελέτης προέκυψαν οι εξής τιμές

Πίνακας 11 Δημιουργία ζευγών χρηστών

Ζευγάρια Χρηστών	Sim
Χρήστες 5 - 4	0,85
Χρήστες 5 - 3	0,00
Χρήστες 5 - 2	0,70
Χρήστες 5 - 1	0,79

Από τα 4 ζευγάρια θα επιλεγθούν τα δύο, αυτά με την μεγαλύτερη τιμή - ομοιότητα. Προκύπτουν τα ζευγάρια χρηστών 5-4 και 5-1 Στην συνέχεια γίνεται πρόβλεψη βάσει των αποτελεσμάτων των δύο πιο κοντινών χρηστών όπως προέκυψε από το sim.

Δημιουργείται ο παρακάτω πίνακας :

Πίνακας 12 Προσδιορισμός βαθμολόγησης προϊόντος από τον χρήστη

Ζευγάρια Χρηστών	Μέσος όρος Βαθμολόγησης	Βαθμολόγηση για το Προϊόν 5
Χρήστης 4	2,25	3
Χρήστης 5	4	?
Χρήστης 1	3,25	1

Και τέλος πραγματοποιείται η εξής πράξη : $(2,25-3)+(3,25-1)/2 = 0,75$

Οπότε η πρόβλεψη για τον χρήστη 5 είναι $4-0,75 = 3,25$. Από τα ως άνω γίνεται κατανοητό πως είναι ο εφικτός ο προσδιορισμός της τάσης των καταναλωτών, ειδικά όταν η βάση δεδομένων περιέχει μεγάλο αριθμό εγγεγραμμένων.

3.7 Item-based Collaborative Filtering

Εκτός από τη μέθοδο «συνεργατική διήθηση με βάση τους χρήστες» υπάρχει και η μέθοδος «συνεργατική διήθηση με βάση το αντικείμενο» η οποία διαδόθηκε από την εταιρεία Amazon όπου τα διάφορα αντικείμενα βαθμολογούνται από τους χρήστες και στη συνέχεια με βάση τον μέσο όρο των αξιολογήσεων δημιουργούνται οι προβλέψεις.

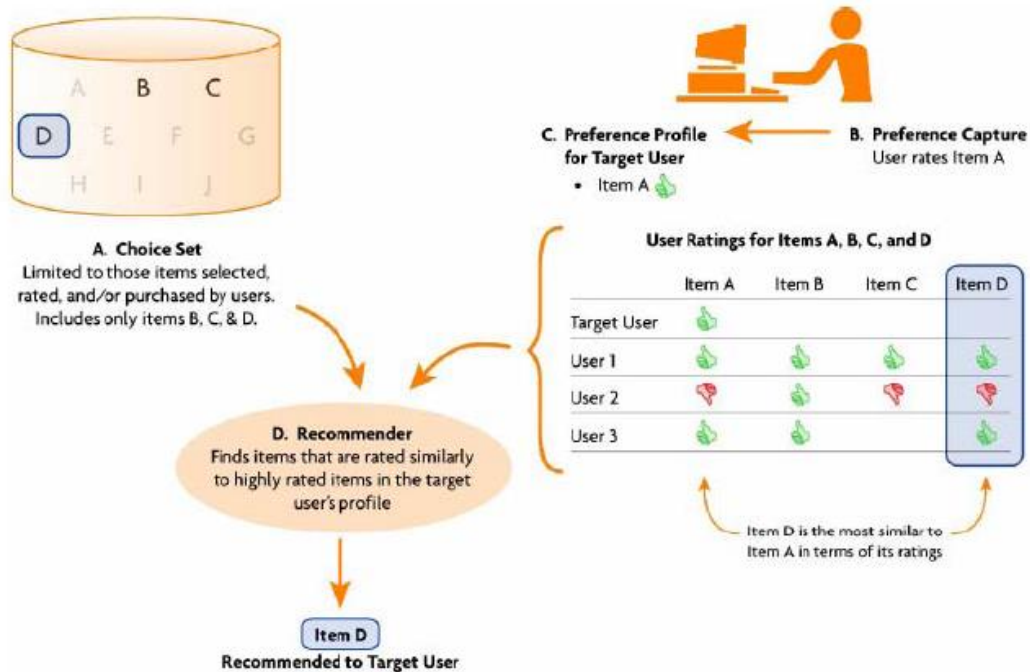
Ο τρόπος λειτουργίας είναι ο εξής: Πρώτα δημιουργείται ένας πίνακας που καθορίζει τις σχέσεις μεταξύ ζευγών αντικειμένων. Στη συνέχεια χρησιμοποιώντας αυτή τη "μήτρα" συνάγονται συμπεράσματα για τις προτιμήσεις του συγκεκριμένου χρήστη.

Η πρόβλεψη βασίζεται σε ένα πίνακα που καθορίζει τις σχέσεις μεταξύ ζευγών αντικειμένων. Στη συνέχεια χρησιμοποιώντας αυτή τη "μήτρα" συνάγονται συμπεράσματα για τις προτιμήσεις του συγκεκριμένου χρήστη.

Η διαδικασία αυτή είναι η απλούστερη μορφή συνεργατικής διήθησης δεδομένων που βασίζεται σε πραγματικές αξιολογήσεις από πραγματικούς χρήστες. Η απλότητά της

καθιστά ιδιαίτερα εύκολη την αποτελεσματική εφαρμογή της, ενώ η ακρίβειά της είναι συχνά στο ίδιο επίπεδο με τους πιο περίπλοκους και ακριβούς υπολογιστικούς αλγορίθμους. Έτσι όταν υπάρχουν διαθέσιμες αξιολογήσεις ενός αντικειμένου, όπως για παράδειγμα όταν δίνεται στους ανθρώπους η ευκαιρία να βαθμολογήσουν ένα προϊόν (π.χ από 1 ως 5), η συνεργατική διήθηση στοχεύει στο να προβλέψει την αξιολόγηση ενός ατόμου, με βάση κάποιες παλιότερες αξιολογήσεις του ή μια (μεγάλη) βάση δεδομένων με αξιολογήσεις άλλων χρηστών.

Εντούτοις είναι πολύ πιθανό να μην υπάρχουν πάντα δεδομένα αξιολογήσεων αλλά αντίθετα να υπάρχουν διάφορα δυαδικά στοιχεία. (π.χ αν ένα αντικείμενο αγοράστηκε ή όχι.). Στην περίπτωση αυτή οι αλγόριθμοι με βάση την βαθμολόγηση δεν ισχύουν.



3.6.1 Παράδειγμα Μελέτης

Ας θεωρήσουμε μια βάση δεδομένων 3 χρηστών με ιστορικό 3 προϊόντων. Καθένας από τους χρήστες αγόρασε ή δεν αγόρασε καθένα από τα προϊόντα. Τα στοιχεία δίνονται στον ακόλουθο πίνακα.

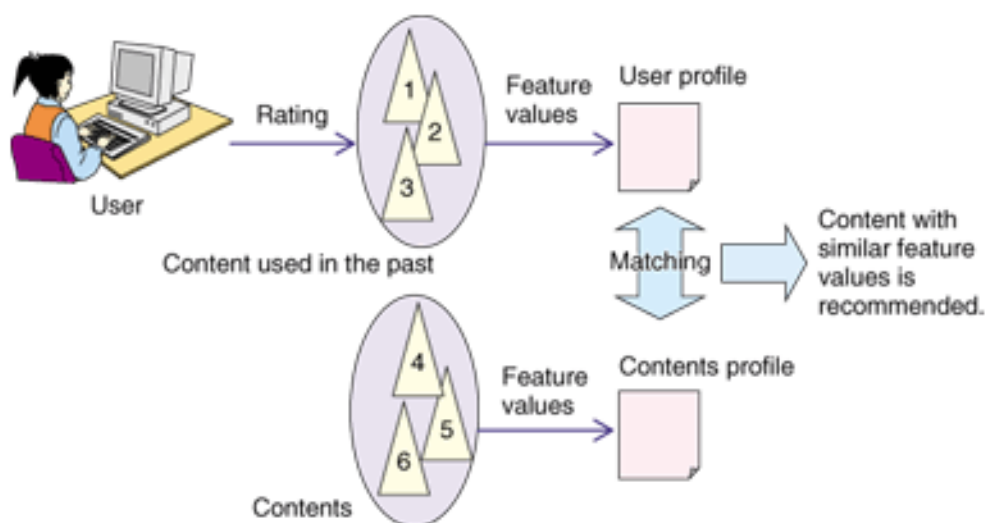
Πίνακας 13 Καταγραφή δεδομένων

	Προϊόν 1	Προϊόν 2	Προϊόν 3
Χρήστης 1	Το αγόρασε	Το αγόρασε	Δεν το αγόρασε
Χρήστης 2	Δεν το αγόρασε	Το αγόρασε	Δεν το αγόρασε
Χρήστης 3	Το αγόρασε	Δεν το αγόρασε	Το αγόρασε

Ως εκ τούτου, στον χρήστη που επιλέγει το αντικείμενο 1 θα προταθεί το αντικείμενο 3, (αφού ο Χρήστης 1 που αγόρασε το 1 αγόρασε επίσης και το 3), στον χρήστη που επιλέγει το αντικείμενο 2 θα προταθεί επίσης το 3 (αφού ο Χρήστης 2 αγόρασε επίσης τα ίδια βιβλία) και, τέλος, ένας χρήστης που επιλέγει το αντικείμενο 3, θα λάβει το αντικείμενο 1 (και, στη συνέχεια, το αντικείμενο 2) ως πρόταση, αφού από τους δύο που αγόρασαν το αντικείμενο 3 ο πρώτος αγόρασε το 1 και ο δεύτερος το 2. Αυτός ο πίνακας στην πραγματικότητα αποτελείται από χιλιάδες χρήστες οι οποίοι δημιουργούν για την περίπτωση των 3 προϊόντων (3!) 6 διαφορετικούς συνδυασμούς. Το σύστημα για κάθε νέο χρήστη δημιουργεί προβλέψεις του τι θέλει να αγοράσει με βάση την ταύτιση του σε επιλογές από κάποιον από τους προηγούμενους χρήστες.

3.8 Content Based Filtering

Τα συστήματα βασισμένα στο περιεχόμενο (content-based recommendation systems), βασίζονται στον εντοπισμό ομοιοτήτων ανάμεσα στα αντικείμενα του ενδιαφέροντος (όπως οι ιστοσελίδες) και στα προφίλ των χρηστών. Τα Συστήματα Συστάσεων που βασίζονται στο περιεχόμενο προτείνουν ένα στοιχείο στον χρήστη βασιζόμενα στην περιγραφή του στοιχείου και στο προφίλ των ενδιαφερόντων του χρήστη. Τα προφίλ των χρηστών μπορούν να έχουν προκύψει είτε από πληροφορίες που έχει δώσει ο ίδιος ο χρήστης, σε μορφή βαθμολόγησης ή like/don't like (explicit data) είτε από πληροφορίες που έχουν συλλεχθεί έμμεσα και τον αφορούν, το οποίο μπορεί να προκύπτει από το ιστορικό, την συμπεριφορά και τις προτιμήσεις του χρήστη (implicit data). Τα Συστήματα Συστάσεων βασισμένα στο περιεχόμενο μπορούν να χρησιμοποιηθούν σε ποικίλους τομείς από την σύσταση ιστοσελίδων, άρθρων ειδήσεων, εστιατορίων, τηλεοπτικών προγραμμάτων, και των στοιχείων για πώληση.



Το προφίλ συχνά δημιουργείται και ενημερώνεται αυτόματα μέσω της ανατροφοδότησης των επιθυμητών στοιχείων που έχουν παρουσιαστεί στο χρήστη.

Τα Συστήματα Συστάσεων βασισμένα στο περιεχόμενο αναλύουν τις περιγραφές των στοιχείων για να προσδιορίσουν τα στοιχεία που παρουσιάζουν ενδιαφέρον για το χρήστη. Το προφίλ με τα ενδιαφέροντα του χρήστη χρησιμοποιείται από τα περισσότερα Συστήματα Συστάσεων και μπορεί να αποτελείται από διαφορετικούς τύπους πληροφορίας. Οι δύο βασικοί τύποι πληροφορίας είναι:

1. Το μοντέλο προτιμήσεων του χρήστη, δηλαδή μια περιγραφή των ειδών των στοιχείων που ενδιαφέρουν το χρήστη. Υπάρχουν πολλές πιθανές εναλλακτικές αναπαραστάσεις αυτής της περιγραφής, αλλά μια κοινή αναπαράσταση είναι μια λειτουργία που για οποιοδήποτε στοιχείο προβλέπει την πιθανότητα ο χρήστης να ενδιαφέρεται για εκείνο το στοιχείο.
2. Το ιστορικό των αλληλεπιδράσεων του χρήστη με το Σύστημα Συστάσεων. Αυτό μπορεί να περιλαμβάνει την αποθήκευση των στοιχείων που ένας χρήστης έχει δει μαζί με άλλες πληροφορίες που αφορούν την αλληλεπίδραση του χρήστη, (π.χ., εάν ο χρήστης έχει ξανά αγοράσει το στοιχείο ή μια αξιολόγηση που μπορεί να έχει κάνει ο χρήστης για το στοιχείο).

3.9 Αλγόριθμοι

Στην τεχνολογία της συνεργατικής διήθησης δεδομένων «Collaborative Filtering» διακρίνονται δύο μοντέλα (όπως αναλύθηκαν παραπάνω) και εφαρμόζονται αντίστοιχα δύο τύποι αλγορίθμων, οι Memory-based και οι Model-based.

Στα CF χρησιμοποιούνται συνήθως δύο τρόποι προσέγγισης για την σύσταση την βασισμένη στην μνήμη «Memory-based» και βασισμένη στο μοντέλο «Model based». Στην «Memory-based» προσέγγιση το σύστημα επεξεργάζεται όλα τα αντικείμενα που έχει αξιολογήσει ο χρήστης για να μπορέσει να τον καταχωρίσει σε ομάδα κοινών ενδιαφερόντων. Το πρόβλημα είναι ότι τις περισσότερες φορές πρέπει να προσπελάσει μεγάλο όγκο δεδομένων. Αυτό έχει σαν αποτέλεσμα έχει να είναι αδύνατο μερικές φορές να πραγματοποιηθεί σύσταση σε πραγματικό χρόνο. Στην βασισμένη στο μοντέλο «Model-based» προσέγγιση δημιουργούμε ένα πρότυπο για τα δεδομένα που χρειαζόμαστε για την εκάστοτε σύσταση. Με άλλα λόγια, χρησιμοποιούμε μόνο τις αξιολογήσεις των χρηστών που μας χρειάζονται, όχι το σύνολο τους. Αυτή η προσέγγιση προσφέρει την ζητούμενη ταχύτητα και εξατομίκευση σε μια σύσταση.

Πλεονεκτήματα: Τα πλεονεκτήματα του φιλτραρίσματος συνεργασίας είναι αρκετά. Πρώτον, το σύστημα δεν χρειάζεται να έχει καμία γνώση πάνω στο αντικείμενο για

το οποίο πραγματοποιείται η σύσταση. Δεύτερον, μέσα από την μεθοδο του φιλτραρίσματος μπορούν να πραγματοποιηθούν ασυνήθιστες συστάσεις αλλά, ταυτόχρονα, άκρως ουσιαστικές. Τρίτον, αν ακολουθήσουμε την «Model-based» προσέγγιση θα έχουμε χαμηλές απαιτήσεις σε μνήμη και ελάχιστο χρόνο επεξεργασίας δεδομένων.

Μειονεκτήματα: Τα μειονεκτήματα είναι κυρίως δύο. Το πρώτο είναι ότι η ποσότητα των δεδομένων που επεξεργάζεται επηρεάζει την ποιότητα της σύστασης που λαμβάνει ο χρήστης. Το δεύτερο πρόβλημα που υπάρχει είναι το λεγόμενο «cold start problem». Δηλαδή, είναι δύσκολο να πραγματοποιηθεί μια σύσταση σε έναν νέο χρήστη που δεν έχει αξιολογήσει τίποτα ακόμα και για το νέο αντικείμενο είναι δύσκολο να συσταθεί αφού δεν έχει αξιολογηθεί από κανέναν χρήστη.

Όσον αφορά την μέθοδο «συνεργατική διήθηση με βάση τους χρήστες» αυτή αντιστοιχεί με τον αλγόριθμο Memory-based όπου γίνονται προβλέψεις μέσω της αξιολόγησης των χρηστών βασιζόμενοι σε στοιχεία προηγούμενων αξιολογήσεων που οι ίδιοι οι χρήστες έχουν κάνει. Η πρόβλεψη υπολογίζεται ως ένας μέσος όρος αξιολογήσεων του κάθε χρήστη αλλά και των χρηστών που το προφίλ τους έχει ομοιότητες. Για την υλοποίηση της πρόβλεψης δημιουργούνται πίνακες ομοιότητας ζευγών χρηστών οι οποίοι λειτουργούν παρασκηνακά. Οι συγκριτικοί πίνακες λειτουργούν με διάδικο σύστημα με 1 ή 0 ανάλογα με την αξιολόγηση ή όχι ενός αντικειμένου από τον χρήστη. Παράλληλα έχει ορισθεί ένα κατώτατο όριο ομοιοτήτων για να αποφασιστεί αν υπάρχει τελικώς ομοιότητα μεταξύ των χρηστών του ζεύγους.

3.9.1 Αλγόριθμοι Βασισμένοι σε Μνήμη

Όσον αφορά τους αλγόριθμους που είναι βασισμένοι σε μνήμη θα πρέπει να αναφερθεί πως το σύστημα διατηρεί στην μνήμη όλες τις γνωστές βαθμολογίες/προτιμήσεις και τις χρησιμοποιεί για να βρει ομοιότητες ανάμεσα σε χρήστες ή αντικείμενα. Δύο χρήστες μοιάζουν όταν ενδιαφέρονται για παρόμοια πράγματα ενώ δύο αντικείμενα μοιάζουν όταν ένα σύνολο χρηστών τα αντιμετωπίζουν με παρόμοια αρέσκεια.

Στην περίπτωση που θέλουμε να βασιστούμε στις ομοιότητες χρηστών (user-based collaborative filtering), αναπαριστούμε τους χρήστες ως διανύσματα στο χώρο των αντικειμένων και υπολογίζουμε την ομοιότητα τους με βάση την απόσταση αυτών των διανυσμάτων. Όταν θέλουμε να εκτιμήσουμε την προτίμηση ενός χρήστη για ένα άγνωστο αντικείμενο συγκεντρώνουμε τις προτιμήσεις των N κοντινότερων χρηστών που έχουν εκφράσει την προτίμηση τους για το αντικείμενο. Εκτιμούμε την προτίμηση του χρήστη εφαρμόζοντας μια συναθροιστική συνάρτηση, συνήθως τον σταθμισμένο μέσο όρο, πάνω στις τιμές που συγκεντρώσαμε.

Φορμαλιστικά, έστω $s(u_i, u_j)$ η συνάρτηση που υπολογίζει την ομοιότητα ανάμεσα στους χρήστες u_i και u_j , $S \subseteq A$ το σύνολο των N όμοιων χρηστών, $r_{u,I}$ η προτίμηση του χρήστη u για το αντικείμενο I . Η γενική μορφή της συνάρτησης για τον υπολογισμό της εκτιμώμενης προτίμησης θα είναι $E(u,I) = \text{aggr}(S)$ και αν η συναθροιστική συνάρτηση είναι ο μέσος όρος η συναθροιστική συ

$$E(u, I) = \frac{\sum_{u_o \in S} s(u, u_o) * r_{u_o, I}}{\sum_{u_o \in S} s(u, u_o)}$$

Αντίστοιχα όταν βασιζόμαστε στις ομοιότητες αντικειμένων (item-based collaborative filtering) αναπαριστούμε τα αντικείμενα ως διανύσματα στον χώρο των χρηστών και υπολογίζουμε τις αποστάσεις των διανυσμάτων. Όταν θέλουμε να εκτιμήσουμε την προτίμηση ενός χρήστη για ένα άγνωστο αντικείμενο βρίσκουμε τα N κοντινότερα αντικείμενα τα οποία έχει βαθμολογήσει ο χρήστης και υπολογίζουμε μέσω μιας συνάρτησης συνήθως με σταθμισμένο μέσο όρο, την τιμή βάσει των άλλων προτιμήσεων του. Φορμαλιστικά έστω $s(I_i, I_j)$ η ομοιότητα ανάμεσα στα αντικείμενα I_i και I_j και $S \subseteq B$ το σύνολο των N όμοιων αντικειμένων. Η εκτίμηση

της προτίμησης είναι $E(u, I) = aggr(SI)$ και αν υποθέσουμε ότι χρησιμοποιούμε τον σταθμισμένο μέσο όρο

$$E(u, I) = \frac{\sum_{I_o \in SI} s(I, I_o) * r_{u, I_o}}{\sum_{I_o \in SI} s(I, I_o)}$$

3.9.2 Αλγόριθμοι βασισμένοι σε μοντέλο

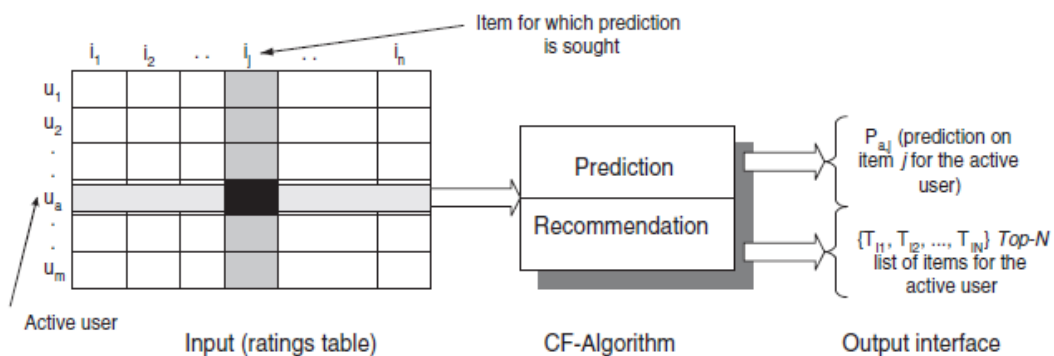
Τα περισσότερα συστήματα συνεργατικών τεχνικών βασίζονται σε αλγορίθμους μνήμης καθώς υλοποιούνται εύκολα. Παρουσιάζουν όμως πρόβλημα κλιμάκωσης και είναι ευαίσθητα στην εισαγωγή ψεύτικων προφίλ (profile injection). Σε αυτού του τύπου τις επιθέσεις γίνεται προσπάθεια εισαγωγής ψεύτικων προφίλ χρηστών ώστε μέσω των προτιμήσεων να ευνοηθούν αντικείμενα που στην πραγματικότητα δεν θα ήταν το ίδιο δημοφιλή. Οι αλγόριθμοι που βασίζονται σε μοντέλο χρησιμοποιούν τα δεδομένα των προτιμήσεων/βαθμολογιών ως σύνολο εκπαίδευσης αλγορίθμων μηχανικής μάθησης για να παράξουν μοντέλα πρόβλεψης βαθμολογιών. Η επεξεργασία γίνεται σε μη πραγματικό χρόνο και έτσι μειώνεται το πρόβλημα κλιμάκωσης καθώς μπορεί να γίνει επεξεργασία μεγάλου αριθμού από προφίλ καθώς δεν απαιτείται απάντηση σε πραγματικό χρόνο. Η δυνατότητα της επεξεργασίας στο παρασκήνιο επιτρέπει την εκτέλεση πολύπλοκων αλγορίθμων αντιμετώπισης του θορύβου που προκαλούν οι επιθέσεις ψεύτικων προφίλ.

Τα μοντέλα πολλές φορές τείνουν να έχουν λιγότερο ακριβή αποτελέσματα από τους αλγορίθμους μνήμης (O'Conner & Herlocker 1999). Για αυτό οι σχεδιαστές συστημάτων συστάσεων θα πρέπει να αναλογιστούν την σχέση κλιμάκωσης προς απόδοση πριν αποφασίσουν ποιες τεχνικές θα χρησιμοποιήσουν.

Δεδομένου ότι η πλειονότητα των χρηστών παρουσιάζουν πολλά ενδιαφέροντα βρίσκονται αντίστοιχα ταξινομημένοι σε διαφορετικές και περισσότερες της μιας κατηγορίες χρηστών. Για να αντιμετωπιστεί αυτή η πολυπλοκότητα χρησιμοποιείται η μέθοδος ομαδοποίησης MinHash. Η μέθοδος MinHashing είναι μια μέθοδος ομαδοποίησης πιθανοτήτων. Αναθέτει ένα ζεύγος χρηστών της ίδιας κατηγορίας, με πιθανότητα ανάλογη της επικάλυψης των αντικειμένων που οι δύο χρήστες έχουν επιλέξει. Ο κάθε χρήστης αντιπροσωπεύεται από μια λίστα αντικειμένων, τα οποία

αποτελούν το ιστορικό της αναζήτησης του. Η ομοιότητα μεταξύ των δύο χρηστών καθορίζεται από την επικάλυψη (κινά αντικείμενα) μεταξύ των λιστών του καθενός. Η ομοιότητα αποδίδεται από τον συντελεστή Jaccard που αναλύθηκε σε προηγούμενο κεφάλαιο. Ωστόσο η πρόβλεψη όταν πρόκειται για πραγματικό χρόνο δεν είναι πάντα εφικτή. Προκύπτουν για αυτό τεχνικές που περιορίζουν την αναζήτηση, όπως για παράδειγμα η δημιουργία ενός hash table, για να ανακαλύψουμε χρήστες που έχουν τουλάχιστον μια κοινή ψήφο- επιλογή.

Η βασική ιδέα του MinHashing είναι να υπολογίζεται μια ενιαία τιμή για κάθε χρήστη ώστε να μειώνεται ο όγκος των συσχετίσεων που μπορούν να προκύψουν. Επομένως μπορούμε να σκεφτούμε το MinHashing ως ένα αλγόριθμο που μπορεί να καταδείξει την πιθανότητα ομοιότητας, κάνοντας το με το να αντιστοιχεί σε μια θέση για κάθε χρήστη.



3.9.3 Σύγκριση Αλγόριθμων

Όπως προκύπτει από την έως τώρα ανάλυση ο παραδοσιακός αλγόριθμος collaborative filtering κάνει ελάχιστο ή καθόλου offline υπολογισμός, και ο online υπολογισμός εξαρτάται από τον αριθμό των πελατών και των αντικείμενων των καταλόγων. Ο αλγόριθμος είναι μη πρακτικός σε μεγάλα σύνολα στοιχείων, εκτός αν χρησιμοποιεί τη μείωση διαστατικότητας, δειγματοληψία, ή χωρισμός –εκ των οποίων όλα μειώνουν την ποιότητα σύστασης.

Τα βασισμένα στην αναζήτηση πρότυπα κατασκευάζουν τη λέξη κλειδί, την κατηγορία, και τις ενδείξεις των συντακτών offline, αλλά αποτυγχάνουν να παρέχουν συστάσεις με ενδιαφέροντες, επιλεγμένους τίτλους. Κάνουν επίσης κακές εκτιμήσεις για τους πελάτες με πολυάριθμες αγορές και εκτιμήσεις.

Το κλειδί για την εξελιξιμότητα και την απόδοση του ο αλγόριθμος item-to-item collaborative filtering είναι ότι δημιουργεί τους ακριβούς πίνακες για παρόμοια στοιχεία offline. Τα online συστατικά του αλγορίθμου -που ανατρέχει σε παρόμοια στοιχεία για τις αγορές και τις εκτιμήσεις του χρήστη- κάνουν εκτιμήσεις ανεξάρτητα από το μέγεθος καταλόγων ή το συνολικό αριθμό από τους πελάτες, εξαρτάται μόνο από το μέγεθος των τίτλων που ο χρήστης έχει αγοράσει ή έχει εκτιμήσει. Κατά συνέπεια, ο αλγόριθμος είναι γρήγορος ακόμη και για τα εξαιρετικά μεγάλα σύνολα στοιχείων. Επειδή ο αλγόριθμος συστήνει ιδιαίτερα υψηλά συσχετισμένα παρόμοια στοιχεία, η ποιότητα σύστασης είναι άριστη. Αντίθετα από τον παραδοσιακό αλγόριθμο collaborative filtering, ο αλγόριθμος αποδίδει επίσης καλά με περιορισμένα στοιχεία χρηστών, παράγοντας υψηλής ποιότητας συστάσεις με βάση μόνο δύο ή τρία στοιχεία.

Οι αλγόριθμοι σύστασης παρέχουν μια αποτελεσματική μορφή οροθετημένου μάρκετινγκ με τη δημιουργία μιας εξατομικευμένης εμπειρίας αγορών για κάθε πελάτη. Για τους μεγάλους λιανοπωλητές όπως το Amazon.com, ένας καλός αλγόριθμος σύστασης είναι εξελικτικός αναφορικά με μια πολύ μεγάλη βάση πελατών και έναν μεγάλο κατάλογο προϊόντων, απαιτεί μόνο κλάσματα δευτερολέπτου επεξεργασίας για να παραχθούν οι online συστάσεις, είναι σε θέση να αντιδράσει αμέσως στις αλλαγές στα στοιχεία ενός χρήστη, και κάνει αναγκαστικές συστάσεις για όλους τους χρήστε ανεξάρτητα από τον αριθμό αγορών και εκτιμήσεων. Αντίθετα με άλλους αλγόριθμους, ο αλγόριθμος item-to-item collaborative filtering είναι ικανός να αντιμετωπίσει αυτήν την πρόκληση.

Στο μέλλον, αναμένουμε τη λιανική βιομηχανία να εφαρμόσει ευρύτερα τους αλγορίθμους σύστασης για οροθετημένο μάρκετινγκ, και online και offline. Ενώ οι επιχειρήσεις ηλεκτρονικού εμπορίου έχουν τα ευκολότερα οχήματα για την εξατομίκευση, τα αυξανόμενα ποσοστά μετατροπής της τεχνολογίας όπως αυτά συγκρίνονται με τις παραδοσιακές προσεγγίσεις ευρείας κλίμακας, θα καταστήσουν επίσης αναγκαίο για τους offline λιανοπωλητές τη χρήση αυτού στις ταχυδρομικές αποστολές, δελτία, και άλλες μορφές επικοινωνίας πελατών.

3.10 Περιορισμοί Εφαρμογής

Οι αλγόριθμοι που χρησιμοποιούνται για συμβουλευτικές υπηρεσίες στο ηλεκτρονικό εμπόριο συχνά δρουν σε ένα ανταγωνιστικό περιβάλλον, ειδικά για μεγάλα ηλεκτρονικά καταστήματα, όπως το eBay και το Amazon. Συχνά ένα συμβουλευτικό σύστημα προωθεί γρήγορα και ακριβή αποτελέσματα προτάσεις προσελκύουν το ενδιαφέρον των πελατών και προσφέρουν κέρδη στις εταιρείες. Για τα συστήματα CF η απόδοση έγκυρων αποτελεσμάτων εξαρτάται από το πόσο καλά οργανωμένα εισάγονται οι πληροφορίες το οποίο είναι άλλωστε και το χαρακτηριστικό της λειτουργίας τους – η εισαγωγή δεδομένων

Υπάρχουν αρκετοί περιορισμοί στις βάσεις μνήμης των τεχνικών CF. Χαρακτηριστικά αναφέρεται πως είναι αναξιόπιστες όταν ο χρήστης δεν έχει συμπληρώσει όλα τα απαιτούμενα πεδία με αποτέλεσμα να μην μπορεί να υπάρξει ταύτιση των δεδομένων ώστε να γίνει κάποια πρόβλεψη.

Πίνακας 14 Παράδειγμα έλλειψης απαιτούμενων πληροφοριών για την λήψη πρόβλεψης

(a)

Alice: (like) Shrek, Snow White, (dislike) Superman
Bob: (like) Snow White, Superman, (dislike) spiderman
Chris: (like) spiderman, (dislike) Snow white
Tony: (like) Shrek, (dislike) Spiderman

(b)

	Shrek	Snow White	Spider-man	Super-man
Alice	Like	Like		Dislike
Bob		Like	Dislike	Like
Chris		Dislike	Like	
Tony	Like		Dislike	?

Αραιή Αναπαράσταση Δεδομένων –Data Sparsity : Στην πραγματικότητα πολλά διαφημιστικά προωθητικά συστήματα χρησιμοποιούνται για να βελτιστοποιήσουν την διαχείριση μεγάλων ποσοτήτων προϊόντων. Οι πίνακες στοιχείων των χρηστών χρησιμοποιούνται για συνδυαστικό φιλτράρισμα το οποίο στις περιπτώσεις που οι χρήστες δεν συμπληρώνουν ολοκληρωμένα αμφισβητείται η αξιοπιστία του.

Η ελλιπής συμπλήρωση πινάκων στοιχείων από τους χρήστες είναι σύνηθες φαινόμενο. Παρατηρείται κυρίως όταν οι πληροφορίες ζητούνται να εισαχθούν στην αρχή όταν ακόμα η σχέση μεταξύ του χρήστη και του προγράμματος είναι κρύα.

Επίσης τα νέα προϊόντα δεν μπορούν να εκτιμηθούν μέχρι κάποιος χρήστης τα βαθμολογήσει και συνήθως οι για αυτό το λόγο στα πρώτα στάδια το σύστημα δεν μπορεί να κάνει σωστές προβλέψεις λόγω έλλειψης ιστορικού του προϊόντος.

Κάλυψη – Coverage : Μπορεί να οριστεί ως για ποια προϊόντα μπορεί το σύστημα να κάνει προβλέψεις στους υποψήφιους πελάτες. Η μειωμένη κάλυψη παρουσιάζεται όταν οι χρήστες βαθμολογούν μόνο ένα μικρό μέρος των προϊόντων, με αποτέλεσμα οι αλγόριθμοι να μην μπορούν να καταλήξουν σε ασφαλή συμπεράσματα.

Μεταβλητότητα Γειτονικά Δεδομένων – Neighbor Transitivity : Το πρόβλημα παρουσιάζεται από την ύπαρξη βάσεων δεδομένων με λίγες πληροφορίες που προκαλείται όταν οι χρήστες δεν έχουν εισάγει όλα τα δεδομένα που απαιτούνται. Ως αποτέλεσμα το σύστημα μπορεί να ταυτίσει μόνο μερικές από τις απαντήσεις με αυτές των άλλων χρηστών και δεν μπορεί να καταλήξει σε συμπεράσματα. Αυτό επηρεάζει κυρίως συστήματα που συγκρίνουν τα δείγματα σε ζευγάρια – και ενώ δηλαδή έχει δημιουργηθεί ένα τέτοιο ζευγάρι δύο πελατών – η έλλειψη στοιχείων για το ένα δεν μπορεί να τελεσφορήσει αποτελέσματα για το άλλο.

Όπως διαπιστώνεται η βασική περιοριστική παράμετρος στην επίτευξη πρόβλεψης είναι ο ανθρώπινος παράγοντας και συγκεκριμένα η έλλειψη πνεύματος συνεργασίας με το σύστημα.

Οι ευφυείς βοηθοί στους ιστοχώρους, ιδιαίτερα σε ιστοχώρους ηλεκτρονικού εμπορίου, γίνονται όλο και περισσότερο κοινοί. Ένα βασικό μειονέκτημα είναι ότι, πολύ συχνά αυτοί οι οδηγοί απόφασης αποτελούνται από έναν τεράστιο κατάλογο ερωτήσεων, οι οποίες πρέπει να απαντηθούν προκειμένου να βρεθεί το καταλληλότερο στοιχείο. Αυτό έχει ως αποτέλεσμα να δημιουργούνται διάλογοι μεγάλης διάρκειας, οι οποίοι οδηγούν στην μείωση του ενδιαφέροντος του χρήστη καθώς δεν τον ενθαρρύνουν στην χρήση αυτού του είδους των ιστοχώρων. Επομένως, στόχος είναι η ύπαρξη ενός ευφυή βοηθού, ο οποίος θα υποβάλλει τον *ελάχιστο αριθμό* ερωτήσεων στην κατάσταση διαλόγου, ελαχιστοποιώντας ταυτόχρονα και τον αριθμό των κύκλων αίτημα-απάντησης. Η μείωση του μήκους του διαλόγου μπορεί να επιτευχθεί μέσω της παραγωγής *επικεντρωμένων (focused)* ερωτήσεων στον χρήστη. (Aha D., 1995)

3.11 Αντιμετώπιση Προβλημάτων

Για την άμβλυνση των προβλημάτων που δημιουργούνται από την έλλειψη πληροφοριών έχουν προταθεί αρκετές τεχνικές αντιμετώπισης. Τεχνικές μείωσης της διάστασης του φαινομένου όπως η SVD Singular Value Decomposition η οποία δεν λαμβάνει υπόψη τα ερωτηματολόγια που παρουσιάζουν ελλείψεις, ώστε να μειώσουν τα φαινόμενα αναντιστοιχίας. Η μέθοδος της SVD στηρίζεται στο ότι είναι προτιμότερο να μην μπορούν να γίνουν προβλέψεις για μεγαλύτερο χρονικό διάστημα λόγω έλλειψης ιστορικού από το να δίνονται εσφαλμένες προτάσεις στους καταναλωτές. Το πρόβλημα της μεθόδου είναι ότι καθώς απορρίπτει τα ελλιπή βιογραφικά, υπάρχει ο κίνδυνος να χαθούν και σημαντικές πληροφορίες.

Μια άλλη τεχνική η (LSI) Latent Semantic Indexing (LSI) χρησιμοποιεί την αρχή ανάκτησης πληροφοριών. Η μέθοδος αξιοποιεί όλα τα ερωτηματολόγια που συμπληρώθηκαν ανεξαρτήτου της πυκνότητας των δεδομένων και από τις ομοιότητες που εντοπίζει μεταξύ του καταναλωτικού προφίλ των χρηστών προσπαθεί να κάνει προβλέψεις. (X. Χριστάκου, 2012)

4. ΣΥΓΚΡΙΣΗ ΜΕΘΟΔΩΝ ΣΥΝΕΡΓΑΤΙΚΗΣ ΔΙΗΘΗΣΗΣ ΔΕΔΟΜΕΝΩΝ

4.1 Μελέτη Συγκεκριμένων Περιπτώσεων

Θεωρείται ως δεδομένο πως το ηλεκτρονικό εμπόριο αποτελεί το μέλλον και την εξέλιξη του παραδοσιακού εμπορίου. Για να μπορέσει μια εταιρεία να γίνει ανταγωνιστική στο ηλεκτρονικό εμπόριο θα πρέπει να εφαρμόσει μεθόδους προσέλκυσης του πελάτη και να μπορέσει να προβλέψει τις ανάγκες του. Για αυτό τον λόγο έχουν αναπτυχθεί λογάριθμοι προσδιορισμού των αναγκών του καταναλωτή και αντίστοιχα συστήματα συστάσεων. Η υιοθέτηση μηχανισμών προσέλκυσης νέων πελατών θα πρέπει να γίνεται παράλληλα με εφαρμογή συστημάτων διαχείρισης της εταιρείας, (συστήματα ERP που αναπτύχθηκαν στο 2^ο κεφάλαιο). Σε αντίθετη περίπτωση, όταν δηλαδή η εταιρεία αναπτύσσεται μονοδιάστατα, χωρίς την βελτίωση της οργάνωσης της το αποτέλεσμα είναι να εκτίθεται δημοσίως. (Σιωμίκος Γ., 2002), (Φούκη Ι, 2013)

Για την κατανόηση της σημαντικότητας αυτής της διττής ανάπτυξης μιας εταιρείας παρουσιάζονται και αναλύονται κάποια αρκετά διαδεδομένα παραδείγματα ιστοσελίδων. Συγκεκριμένα παρουσιάζονται οι εξής ιστότοποι (H. Marmanis, 2008)

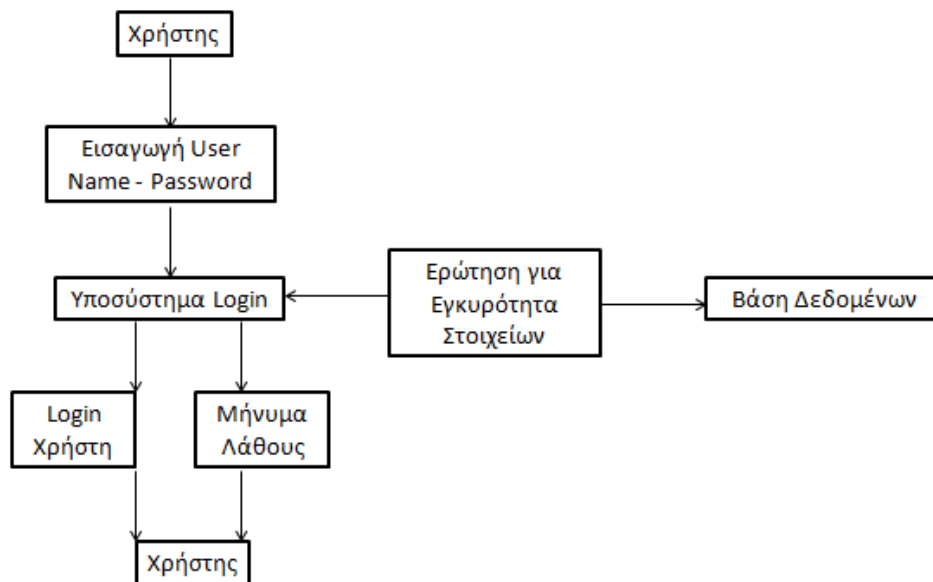
- Εφημερίδα Guardian
- Cinematch - Ενοικιάσεις Ταινιών
- Twitter – Αλγόριθμος (Follower of Follower)
- eBay
- Amazon
- YouTube

Στόχος της ανάλυσης είναι η τεκμηρίωση της άποψης ότι η επιτυχία των ιστότοπων σε θέματα σύστασης οφείλεται στην συνδυασμένη χρήση αλγόριθμων φιλτραρίσματος και του συστήματος διαχείρισης των βάσεων δεδομένων τους.

4.2 Εφημερίδα Guardian

Η ανάπτυξη ειδησεογραφικών ιστοσελίδων προέβαλε ως επιτακτική ανάγκη, στην εύρεση τρόπων που θα προσαρμόζονταν στις ανάγκες, τις ιδιαιτερότητες κάθε ατόμου, ώστε να του εξασφαλίσουν το επιθυμητό, δηλαδή την αδιάλειπτη και συνεχή εύκολη πρόσβαση στην πληροφορία. Μια από τις πρώτες σε επισκεψιμότητα σελίδες σε ευρωπαϊκό επίπεδο είναι η εφημερίδα Guardian.

Στην αρχική της σελίδα (www.guardian.co.uk) η εφημερίδα έχει πάνω δεξιά την επιλογή εγγραφής στο σύστημα. Η εγγραφή μπορεί να γίνει μέσα σε ελάχιστα δευτερόλεπτα μέσω του λογαριασμού του Facebook που διαθέτει ο χρήστης. Όταν ο χρήστης δώσει τα αναγνωριστικά του στοιχεία (username και password), εξετάζεται – μετά από επικοινωνία με τη βάση δεδομένων – αν αντιστοιχούν σε εγγεγραμμένο χρήστη. Αν ναι, εισάγεται στο σύστημα αρχίζει την πλοήγησή του στην πύλη. Επίσης προσφέρεται και η εναλλακτική, αν δεν είναι γραμμένος χρήστης, να συμπληρώσει τη φόρμα εγγραφής και να καταχωρηθεί στο σύστημα. Το διάγραμμα ροής δεδομένων της εφαρμογής φαίνεται στην επόμενη εικόνα



Εικόνα 17 Εφαρμογή εισόδου επισκέπτη

Μετά την εγγραφή του χρήστη στο σύστημα, κάθε άρθρο που διάβασει ο χρήστης καταγράφεται και αποθηκεύεται σε μια βάση δεδομένων με το ιστορικό των τελευταίων τριάντα ημερών. Ως αποτέλεσμα της διαδικασίας αυτής, κάθε φορά που ο

χρήστης διαβάζει ένα άρθρο στο τέλος της σελίδας του προτείνονται άρθρα που μοιότητες στο περιεχόμενο με αυτά που έχει διαβάσει στο παρελθόν.

The screenshot shows the top navigation bar of the Guardian website. The 'Sign in' link is circled in red. Below the navigation bar is a blue banner for a Xerox printer. The main content area features the Guardian logo, a search bar, and a list of navigation links including News, Sport, Comment, Culture, Business, Money, Life & style, Travel, Environment, Tech, TV, Video, Dating, Offers, and Jobs. A 'Breaking news' section highlights a story about Pinewood Studios. Below this are three article recommendations: 'A&E doctors urge overhaul of emergency care', 'The most awkward interview ever? An unsettling encounter with Ginger Baker', and 'guardianjobs'.

Εικόνα 18 Απόσπασμα αρχικής σελίδας

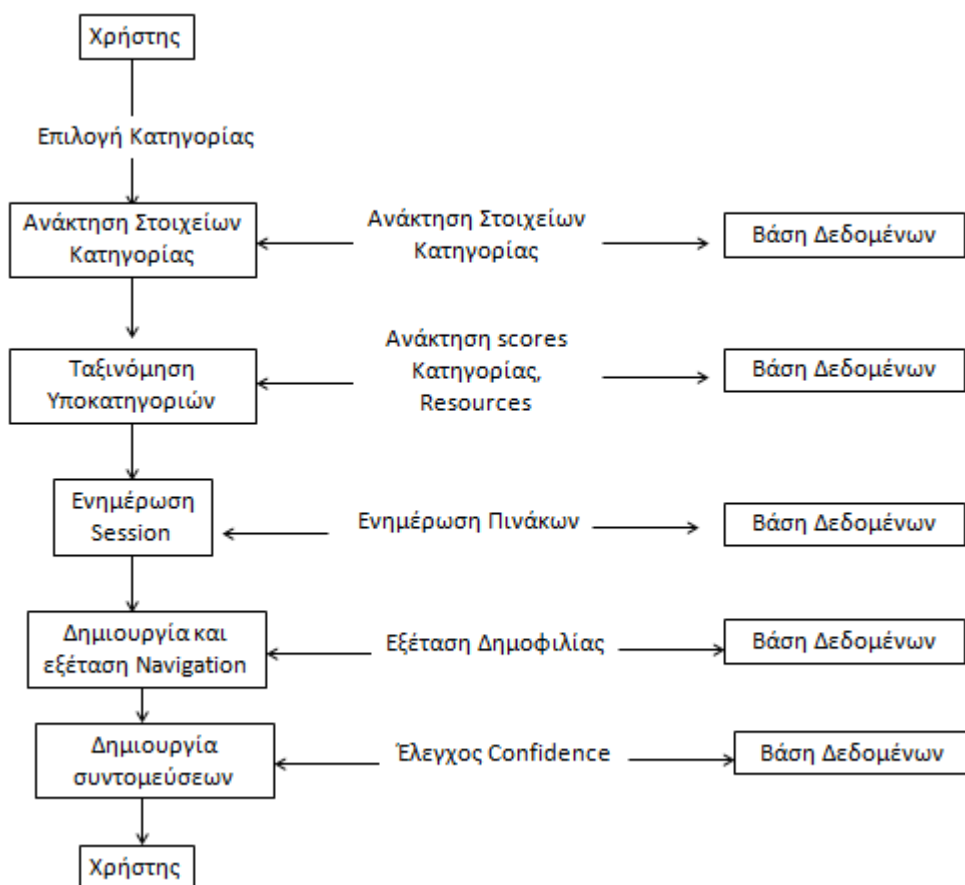
Ως στόχος του συστήματος είναι να καταφέρει να προτείνει άρθρα στον χρήστη τα οποία θα ήθελε πολύ αλλά θα ήταν δύσκολο να τα βρεί κάτω από άλλες συνθήκες.

Η ιστοσελίδα χρησιμοποιεί επίσης και μια μέθοδο μη προσωποποιημένης σύστασης που βασίζεται στα πιο δημοφιλή άρθρα που έχουν δημοσιευτεί. Η δημοτικότητα ενός άρθρου φαίνεται από το πόσες ενέργειες έχουν γίνει από τους χρήστες πάνω στο άρθρο. Ενέργεια μπορεί να θεωρηθεί η ανάγνωση ενός άρθρου, το ποστάρισμα του σε κάποιο blog ή το να γίνει αποστολή του άρθρου με e-mail. Η μέθοδος αυτή έχει στόχο να προσελκύσει νέους χρήστες που δεν έχουν κάνει εγγραφή ακόμα

The screenshot shows two columns of article recommendations. The left column is titled 'More from the Guardian' and lists articles such as 'Oxford child sex abuse ring: social services failed me, says victim', '10 gross ingredients you didn't know were in your food', 'The day I worked for the NHS 111 helpline', 'Photograph in Tia Sharp case 'appeared to have been taken after death'', and 'Young and poor hit hardest as UK cuts widen inequality, says OECD'. The right column is titled 'More from around the web' and lists articles such as 'Sanofi, Regeneron Announce Patient Enrollment in Two Phase 3 Sarilumab Trials for Rheumatoid Arthritis', 'Why You Should Be Drinking Lemon Water In The Morning', 'Burning Away High Blood Pressure', 'I question myself daily; will child abuse in Saudi Arabia ever change?', and 'Middle-Aged, Yet a Prince, Slouching but Haunted'.

Εικόνα 19 Οι σχετικές προτάσεις της εφημερίδας, βάσει του προηγούμενου άρθρου που αναγνώστηκε

Η ιστοσελίδα χρησιμοποιεί ως πηγή σύστασης τα κειμενικά χαρακτηριστικά καθώς και λέξεις κλειδιά όπως: θέμα και συγγραφέας. Με βάση ένα σύνολο χαρακτηριστικών γνωρισμάτων στοιχείων, το σύστημα προσπαθεί να δημιουργήσει ένα μοντέλο για κάθε χρήστη που του επιτρέπει να ταξινομήσει τα άγνωστα στοιχεία σε ενδιαφέροντα μη ενδιαφέροντα για αυτόν. Η σύσταση στηρίζεται στο φιλτράρισμα περιεχομένου και η εφαρμογή του αλγόριθμου είναι απλή καθώς δεν χρειάζονται σύνθετες συσχετίσεις, (π.χ. την γνώμη των άλλων αναγνωστών). Ο αλγόριθμος παρουσιάζει την ακόλουθη μορφή.



Εικόνα 20 Εφαρμογή πλοήγησης

4.3 Cinematch - Ενοικιάσεις Ταινιών

Η εταιρεία Netflix, που δραστηριοποιείται στον χώρο ενοικιάσεως ταινιών⁷ μέσω διαδικτύου θέλοντας να προσφέρει όσο το δυνατόν καλύτερες υπηρεσίες στους συνδρομητές, σχεδίασαν ένα σύστημα συστάσεων (αλγόριθμος) με την βοήθεια του οποίου θα προτεινόταν ταινίες στους συνδρομητές με βάση το προφίλ τους. Ο χρησιμοποιούμενος αλγόριθμος είναι ένα σύστημα συνεργατικού φιλτραρίσματος (Collaborative Filtering System) με την ονομασία. Το CineMatch είναι μια βάση δεδομένων που χρησιμοποιεί πληροφορίες από τρεις πηγές για να αποφασίσει ποιες ταινίες θα προτείνει στον χρήστη. (Μ. Κωνσταντίνου, 2012)

- Αρχικά οι ίδιες οι ταινίες, οι οποίες οργανώνονται σε ομάδες κοινών ταινιών, βάσει του περιεχομένου τους. Με τον όρο «περιεχόμενο ταινίας» καλείται ένα σύνολο στοιχείων που μπορούν να καθορίσουν λεκτικά τις παραμέτρους μιας ταινίας. Μια σημαντική παράμετρος είναι το είδος μιας ταινίας, ή καλύτερα τα είδη στα οποία ανήκει. Λαμβάνονται υπόψη ακόμα οι συντελεστές της ταινίας, δηλαδή ο σκηνοθέτης, ο σεναριογράφος και οι ηθοποιοί. Τέλος, εξετάζεται και η περίληψη της ταινίας. Προφανώς, από το κείμενο της περίληψης(μαζί με την προσθήκη του τίτλου) αφαιρούνται οι συνηθισμένες και χωρίς σημασία λέξεις (όπως σύνδεσμοι, αντωνυμίες κ.α.) και υπολογίζεται η συχνότητα των λέξεων για να καθοριστούν αυτές που είναι σημαντικές και χαρακτηρίζουν την ταινία.
- Δεύτερον οι βαθμολογίες που έχει δώσει ο χρήστης σε προηγούμενες ταινίες που έχει δει. Η βαθμολογία μπορεί να είναι από ένα έως πέντε αστέρια (πέντε είναι η καλλίτερη). Ταινίες άξιες σύστασης θεωρούνται αυτές που λαμβάνουν βαθμολογία 4-5, ενώ απορριπτέες αυτές που λαμβάνουν 1-3. Αυτό είναι απαραίτητο προκειμένου να γίνουν γνωστές οι προτιμήσεις του χρήστη και να κατασκευαστεί το προφίλ.
- Τέλος οι συνδυασμένες βαθμολογίες όλων των χρηστών της υπηρεσίας. (Χ. Χριστάκου, 2012)

⁷ Η εταιρεία παρέχει τη μεγαλύτερη συλλογή ταινιών, όμως αυτή είναι διαθέσιμη μόνο στην Αμερική και κυρίως γίνεται μέσω της ταχυδρομικής υπηρεσίας, ενώ μόνο ένας μικρός αριθμός τίτλων είναι διαθέσιμος για προβολή μέσω διαδικτύου.



Εικόνα 21 Αποτελέσματα Recommender System στο Netflix

Στην συνέχεια συνδυάζει αυτές τις πληροφορίες μεταξύ τους με σκοπό να βρεί παρόμοιες ταινίες με τις ταινίες που του άρεσαν στο παρελθόν. Το Cinematch, εβδομαδιαία, επεξεργάζεται τις παλιότερες αξιολογήσεις ταινιών από τους συνδρομητές με σκοπό να καταχωρίσει τις ταινίες σε λίστες ταινιών με ομοιότητες. Για να βρεθεί η ομοιότητα ανάμεσα στις ταινίες, χρησιμοποιείται ο συντελεστής γραμμικής συσχέτισης. (Φ. Μιχελινάκης, 2011)

Η συσχέτιση μετρά το βαθμό συνάφειας αλληλεπίδρασης ανάμεσα σε δύο ή περισσότερες ταινίες. Πρακτικά, σημαίνει ότι από την τιμή ενός δείκτη (συντελεστή συσχέτισης) κατανοούμε πόσο έντονη ή χαλαρή είναι η συσχέτιση δύο ταινιών. Η διαδικασία συσχέτισης παρουσιάζεται στις κατηγορικές μεταβλητές του Netflix με μια παραλλαγή του συντελεστή Pearson.

Η συσχέτιση μετρά το βαθμό συνάφειας-αλληλεπίδρασης ανάμεσα σε δύο ή περισσότερες ταινίες. Πρακτικά, σημαίνει ότι από την τιμή ενός δείκτη (συντελεστή συσχέτισης) κατανοούμε πόσο έντονη ή χαλαρή είναι η συσχέτιση δύο ταινιών. Η διαδικασία συσχέτισης παρουσιάζεται στις κατηγορικές μεταβλητές του Netflix με μια παραλλαγή του συντελεστή Pearson.

Το στατιστικό τεστ X^2 (Pearson Chi-Square test) είναι το πιο δημοφιλές μη παραμετρικό τεστ.

Καθώς έχουμε στη διάθεση μας ένα δείγμα ποιοτικών δεδομένων οργανωμένο σε ονομαστικές κατηγορίες, στοχεύουμε στην χρήση των δεδομένων αυτών ώστε να προσδιοριστεί η αναλογία του πληθυσμού που ανήκει στην κάθε κατηγορία. Για την επίτευξη αυτού του στόχου διατυπώνεται μια μηδενική υπόθεση, που είτε δηλώνει ότι

δεν υπάρχει κάποια συγκεκριμένη προτίμηση στις διαθέσιμες ονομαστικές κατηγορίες (*no-preference null-hypothesis*), είτε δηλώνει ότι τα ποσοστά που προτιμώνται από τα υποκείμενα δε διαφέρουν από τα ποσοστά άλλων πληθυσμών οι οποίοι αποτελούν το σημείο αναφοράς (*no difference from a comparison population*). Και στις δύο περιπτώσεις, αυτό που προσδιορίζει η μηδενική υπόθεση είναι ο αναμενόμενος αριθμός (expected frequency – fe) των υποκειμένων που ανήκει σε κάθε ονομαστική κατηγορία. Ο έλεγχος υποθέσεων που ακολουθεί αξιολογεί αυτή τη μηδενική υπόθεση, συγκρίνοντας τον αριθμό των υποκειμένων που αναμένεται σε κάθε ονομαστική κατηγορία με τον αριθμό των υποκειμένων που παρατηρείται ότι ανήκει σε κάθε ονομαστική κατηγορία (observed frequency - fo), με βάση τις μετρήσεις του δείγματος.

The screenshot shows a 'Your Queue' interface. At the top, there's a 'DVDs At Home' section with two items: 'A Scanner Darkly' and 'Neil Young: Heart of Gold', both with 3-star ratings and scheduled for 04/01/08. Below this is a promotional banner for a '3 at-a-time (Unlimited) plan'. The main section is 'DVD Queue (9)', which lists 9 movies. The first row, 'I Am Legend', is circled in red. It has a 3-star rating and is marked as 'Long Wait'. The other movies in the queue are 'Once', 'Beowulf: Director's Cut', 'A Prairie Home Companion', 'Meet the Robinsons', 'I, Robot', 'The Librarian: Quest for the Spear', and 'Enchanted', all with 3-star ratings.

List Order	Movie Title	Watch Instantly	Star Rating	Genre	Availability	Remove
1	I Am Legend		☆☆☆☆☆	Thrillers	Long Wait	<input type="checkbox"/>
2	Once		☆☆☆☆☆	Independent		<input type="checkbox"/>
3	Beowulf: Director's Cut		☆☆☆☆☆	Sci-Fi & Fantasy		<input type="checkbox"/>
4	A Prairie Home Companion		☆☆☆☆☆	Comedy		<input type="checkbox"/>
5	Meet the Robinsons		☆☆☆☆☆	Children & Family		<input type="checkbox"/>
6	I, Robot		☆☆☆☆☆	Sci-Fi & Fantasy		<input type="checkbox"/>
7	The Librarian: Quest for the Spear		☆☆☆☆☆	Action & Adventure		<input type="checkbox"/>
8	Enchanted		☆☆☆☆☆	Comedy		<input type="checkbox"/>

Εικόνα 22 Κατάταξη των ταινιών βάσει της λίστας αναμονής που υπάρχει για αυτές. Για κάθε μια ταινία δίνεται η συνολική αξιολόγηση που έχει λάβει από το κοινό

Το σύστημα έχει τρία ευδιάκριτα στρώματα, το γραφικό περιβάλλον διαπροσωπείας χρήστη (GUI), το κύριο στρώμα εφαρμογής που εφαρμόζει τις στρατηγικές σύστασης (π.χ., μηχανή ψηφοφορίας) και το στρώμα αποθήκευσης πληροφοριών που αποτελείται από τη βάση δεδομένων ταινιών και τη βάση δεδομένων των προτιμήσεων των χρηστών.

4.4 Twitter – Follower of Follower

Το Twitter είναι μια από τις πιο αξιoσημείωτες micro-blogging υπηρεσίες, που απασχολεί ένα μοντέλο κοινωνικού δικτύου που ονομάζεται «following», στο οποίο κάθε χρήστης μπορεί να διαλέξει εκείνον που αυτός θέλει να «ακολουθεί» (follow), δηλαδή από τον οποίο μπορεί να λαμβάνει tweets χωρίς να απαιτείται ο τελευταίος να παρέχει άδεια πρώτα.

Το Micro-blogging είναι ένα είδος επικοινωνίας που εμφανίζεται ολοένα και περισσότερο στο προσκήνιο τα τελευταία χρόνια. Επιτρέπει στους χρήστες να δημοσιεύουν σύντομα μηνύματα ενημέρωσης, τα οποία μπορεί να υποβάλλονται σε πολλά διαφορετικά κανάλια, συμπεριλαμβανομένου του Web και της υπηρεσίας έκδοσης μηνυμάτων. Μία από τις περισσότερο διακεκριμένες υπηρεσίες του microblogging είναι το Twitter. Αναφερόμαστε στους χρήστες του twitter ως «Twitterers», και τα μικρά μηνύματα ενημέρωσης που δημοσιεύονται από τους χρήστες «tweets».

The image shows a screenshot of the Twitter homepage. On the left, there's a sidebar with the user's profile summary: 6 tweets, 25 followers, and 5 following. Below that is a 'Who to follow' section with suggestions like 'TV Χωρίς Σύνορα', 'Bloomberg News', and 'Greens in the EP'. A red box highlights this section. The main content area shows a list of tweets from various users and organizations, including silas serafim, Lizabeth Ronk, Huffington Post, stern.de, and pwned by.

Εικόνα 23 Απόσπασμα της αρχικής σελίδας του Twitter. Στην δεξιά στήλη υπάρχει το παράθυρο σύστασης ποιόν να ακολουθήσει ο χρήστης

Αυτό επιτρέπει στους twitterers να δημοσιεύουν tweets (με ένα όριο 140 χαρακτήρων). Το Twitter επίσης παρέχει τη λειτουργικότητα της κοινωνικής δικτύωσης. Αντίθετα προς τις υπηρεσίες άλλων κοινωνικών δικτύων που απαιτούν οι χρήστες να στέλνουν προσκλήσεις φιλίας προς τους άλλους χρήστες για να τους κάνουν φίλους, το Twitter περιλαμβάνει ένα μοντέλο κοινωνικής δικτύωσης καλούμενο «ακόλουθος» (following), στο οποίο κάθε twitterer επιτρέπεται να διαλέγει ποιόν θέλει να ακολουθεί χωρίς απαίτηση κάποιας άδειας. Αντιστρόφως, αυτός μπορεί επίσης να ακολουθείται από άλλους χωρίς τη χορήγηση άδειας πρώτα.

Το Twitter έγινε πολύ δημοφιλές από την πρώτη μέρα που εμφανίστηκε. Ενδεικτικά, το πρώτο εξάμηνο του 2010 ήταν καταγεγραμμένοι στο Twitter πάνω από 100 εκατομμύρια χρήστες, οι οποίοι συνέτασσαν πάνω από 65 εκατομμύρια tweets την ημέρα. Αυτό έχει τραβήξει το αυξανόμενο ενδιαφέρον της ερευνητικής κοινότητας. Έχουν γίνει εργασίες για τη μελέτη των τοπολογικών και γεωγραφικών ιδιοτήτων του κοινωνικού δικτύου που σχηματίστηκε από τους twitterers και από τους followers τους. Επιπροσθέτως, έχουν πραγματοποιηθεί έρευνες για τον προσδιορισμό της ταυτότητας των twitterers που επηρεάζουν («influential»).

Το Twitter χρησιμοποιεί ένα αλγόριθμο που βασίζεται στην υπόθεση ότι «εάν κάποιος που ακολουθεί ο χρήστης A, ακολουθεί με την σειρά του κάποιον χρήστη B πιθανώς ο χρήστης A να θέλει να ακολουθήσει τον B» Αυτή η μέθοδος είναι αρκετά αποτελεσματική καθώς, είναι πολύ πιθανόν, χρήστες που έχουν κοινούς ακόλουθους να θέλουν να ακολουθηθούν και μεταξύ τους.

η “following” σχέση να είναι ένας ισχυρός δείκτης της ομοιότητας μεταξύ των χρηστών. Με άλλα λόγια, ένας twitterer ακολουθεί ένα φίλο επειδή ενδιαφέρεται για τα θέματα που ο φίλος δημοσιεύει στα tweets και ο φίλος ακολουθεί πίσω επειδή βρίσκει ότι μοιράζονται θέμα όμοιου ενδιαφέροντος. Αυτό το φαινόμενο ονομάζεται «ομοφιλία», και έχει παρατηρηθεί σε πολλά κοινωνικά δίκτυα, πράγμα το οποίο είναι πολύ σημαντικό.

Στο περιβάλλον του Twitter η ομοφιλία υποδηλώνει ότι ένας twitterer ακολουθεί ένα φίλο επειδή ενδιαφέρεται για τα θέματα που ο φίλος δημοσιεύει, και ο φίλος ακολουθεί πάλι επειδή βρίσκει ότι μοιράζονται όμοια θεματικά ενδιαφέροντα.

Η παρουσία της ομοφιλίας υποδηλώνει ότι υπάρχουν twitterers οι οποίοι με σοβαρότητα διαλέγουν φίλους να ακολουθήσουν. Αυτό είναι βασικό, διότι η αναγνώριση των twitterers οι οποίοι επηρεάζουν βασιζόμενοι στις «following»

σχέσεις, δεν θα είχε κανένα νόημα εάν κανένας twitterer δεν είναι διαλέγει με σοβαρότητα ποιους να κάνει follow.

4.5 eBay

Το eBay είναι ένα κατάστημα ηλεκτρονικού εμπορίου που εκτελεί την μεσολάβηση πώλησης διάφορων προϊόντων. Η ιστοσελίδα χρησιμοποιεί ως πηγή σύστασης τα χαρακτηριστικά του αντικειμένου. Όλα τα προϊόντα που εμπορεύεται η σελίδα διαθέτουν λεπτομερή χαρακτηριστικά (λέξεις κλειδιά) όπως: ονομασία, κατασκευαστή, χρονολογία κατασκευής.

Στην εφαρμογή «ίσως να σου αρέσει αυτό» οι συστάσεις βασίζονται στην αγορά ενός προϊόντος από τον χρήστη. Από την στιγμή που ο χρήστης θα πραγματοποιήσει την αγορά, το σύστημα θα του κάνει προτάσεις που να σχετίζονται με τα προϊόντα που έχει αγοράσει στο παρελθόν. Η σύσταση στηρίζεται στην συσχέτιση με βάση το περιεχόμενο αντλώντας πληροφορίες από τα προϊόντα για τα οποία έχει ενδιαφερθεί ο καταναλωτής χωρίς να τον συσχετίζει με τους υπόλοιπους.

ation Intel Core i5 processor, 4
drive give you loads of
ide range of tasks.



[Click here to view full size.](#)

[Shipping](#) [Returns](#) [Contact Us](#)



Τα συστήματα διαχείρισης θεματικών καταλόγων στο διαδίκτυο, εφαρμόζουν διαδικασίες εξόρυξης γνώσης από τις πλοηγήσεις των χρηστών με σκοπό τη δημιουργία ομάδων χρηστών με κοινά ενδιαφέροντα και κατ'επέκταση την εξατομίκευση των καταλόγων σε επίπεδο ομάδων χρηστών.

Το περιεχόμενο του καταλόγου, η ιεραρχία του, καθώς και το σύνολο των πλοηγήσεων των χρηστών είναι αποθηκευμένο με τη μορφή μίας σχεσιακής βάσης δεδομένων.

Ένας θεματικός κατάλογος μπορεί να παρασταθεί με τη μορφή ενός κατευθυνόμενου γράφου, όπου οι κόμβοι αποτελούν τις κατηγορίες και οι ακμές τις σχέσεις κατηγορία-υποκατηγορία.

Στα συστήματα διαχείρισης θεματικών καταλόγων ενσωματώνεται η λειτουργία αναζήτησης σε κατηγορίες οι οποίες αποτελούν απογόνους της εκάστοτε κατηγορίας που διαλέγει ο χρήστης να πραγματοποιήσει την αναζήτησή του. Συνεπώς θα πρέπει να υπάρχει ένας αποδοτικός τρόπος να συγκεντρώνεται το σύνολο των υποκατηγοριών μίας κατηγορίας.

η λύση προσφέρει η χρήση ενός Interval Labeling Scheme. Σύμφωνα με αυτό, αρχικά εντοπίζεται το spanning tree του γράφου. Σε κάθε κόμβο του spanning tree (δηλαδή κάθε θεματική ενότητα) δίνεται ένα ζεύγος τιμών $[a, b]$, όπου το b είναι ο αριθμός που δίνεται στον κόμβο κατά την postorder διάσχιση του δέντρου και a το ελάχιστο b που εμφανίζεται στους απογόνους του κόμβου. Στη συνέχεια, η διαδικασία επεκτείνεται σε όλο το γράφο ως εξής: Για κάθε ζευγάρι κόμβων p, q , που αντιστοιχεί σε σχέση κατηγορία - υποκατηγορία, ο κόμβος p κληρονομεί το ζεύγος τιμών

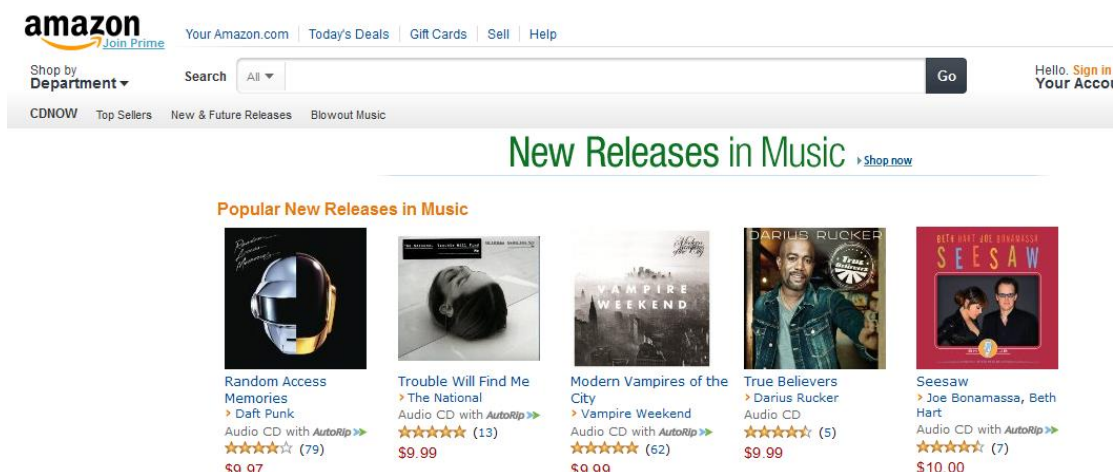
$[a, b]$ του q , εκτός αν το διάστημα $[a_p, b_p]$ περιλαμβάνει το $[a_q, b_q]$. Ικανή και αναγκαία συνθήκη για να είναι ένας κόμβος q απόγονος ενός κόμβου p , είναι κάποιο από τα διαστήματα $[a, b]$ του κόμβου p να περιλαμβάνει τον postorder αριθμό του q .

Η συνεισφορά της συγκεκριμένης τεχνικής στη λειτουργία της αναζήτησης είναι πολύ σημαντική, καθώς επιτρέπει τη γρήγορη και αποδοτική εύρεση του συνόλου των υποκατηγοριών μίας κατηγορίας (στην οποία εκτελείται αναζήτηση), με σκοπό να περιορίζεται σε αυτό η αναζήτηση κατηγοριών και resources.

4.6 Amazon

Στο ηλεκτρονικό εμπόριο χρησιμοποιείται κυρίως η ιστορική προσέγγιση, σύμφωνα με την οποία τα συστήματα διατηρούν έναν κατάλογο των αγορασμένων προϊόντων και των αντίστοιχων εκτιμήσεων των χρηστών, ως παραμέτρους χρήστη (προφίλ). Αυτό συμβαίνει στο δημοφιλέστερο σύστημα recommender στο ηλεκτρονικό εμπόριο, το Amazon.com Το Amazon είναι ένα κατάστημα ηλεκτρονικού εμπορίου που αναλαμβάνει την πώληση διάφορων προϊόντων, όλων των κατηγοριών εμπορίου. Η ιστοσελίδα χρησιμοποιεί διάφορες μεθόδους σύστασης που στηρίζονται σε τρεις πηγές για την τελική πρόταση προς τον χρήστη.

- Η πρώτη πηγή είναι η συμπεριφορά του χρήστη, δηλαδή ποια αντικείμενα έχει αγοράσει στο παρελθόν ή πρόσθεσε στο καλάθι αγορών ή πως αξιολόγησε το αντικείμενο μετά από την αγορά του (Rating-comment).
- Η δεύτερη πηγή είναι τα χαρακτηριστικά του αντικειμένου. Όλα τα προϊόντα που εμπορεύεται η σελίδα διαθέτουν λεπτομερή χαρακτηριστικά (λέξεις κλειδιά) όπως: ονομασία, κατασκευαστή, χρονολογία κατασκευής.
- Η τρίτη πηγή που λαμβάνεται υπόψη για την τελική σύσταση είναι οι ομοιότητες ανάμεσα στους χρήστες.



The screenshot shows the Amazon website interface. At the top, there is the Amazon logo with 'Join Prime' and navigation links: 'Your Amazon.com', 'Today's Deals', 'Gift Cards', 'Sell', and 'Help'. Below this is a search bar with 'Shop by Department' and 'Go' buttons. The main content area is titled 'New Releases in Music' with a 'Shop now' link. Underneath, there is a sub-section 'Popular New Releases in Music' displaying five music releases:

Album Cover	Title	Artist	Price	Rating
	Random Access Memories	Daft Punk	\$9.97	★★★★☆ (79)
	Trouble Will Find Me	The National	\$9.99	★★★★☆ (13)
	Modern Vampires of the City	Vampire Weekend	\$9.99	★★★★☆ (62)
	True Believers	Darius Rucker	\$9.99	★★★★☆ (5)
	Seesaw	Joe Bonamassa, Beth Hart	\$10.00	★★★★☆ (7)

Οι συστάσεις βασίζονται στην αγορά ενός προϊόντος από τον χρήστη. Από την στιγμή που ο χρήστης θα πραγματοποιήσει την αγορά, το σύστημα θα του κάνει προτάσεις που θα εμφανιστούν σε δύο διαφορετικές λίστες. Στην πρώτη λίστα το σύστημα

βρίσκει άλλους χρήστες που έχουν αγοράσει το ίδιο προϊόν με τον χρήστη. Στην συνέχεια προτείνει στον χρήστη προϊόντα που δεν έχει αγοράσει αλλά έχουν αγοράσει οι άλλοι χρήστες που έχουν αγοράσει το ίδιο προϊόν. Ουσιαστικά, αυτές οι προτάσεις στηρίζονται στο ότι χρήστες με κοινές αγορές θα έχουν και κοινά ενδιαφέροντα.

Customers Who Bought This Item Also Bought

		
<p>Charleston, SC 1966 Darius Rucker ★★★★★ (82) Audio CD \$9.00</p>	<p>Golden Lady Antebellum ★★★★☆ (21) Audio CD \$11.88</p>	<p>Learn To Live Darius Rucker ★★★★★ (139) Audio CD \$9.00</p>

Εικόνα 24 Το συνεργατικό φιλτράρισμα με βάση το προϊόν εφαρμόζεται ευρέως στο Amazon.com



amazon.com. [Help](#) | [Close window](#)

Recommended for you

Altered Carbon
 by Richard Morgan
 Our Price: **\$10.17**
 Used & new from **\$7.48**
 Add to Cart Add to Wish List

×|★★★★★
 I Own It
 Not interested

Because you rated...

Interface
 by Neal Stephenson, J. Frederick George
 Use to make recommendations

The Diamond Age
 by Neal Stephenson
 Use to make recommendations

The Diamond Age : Or, a Young Lady's Illustrated Primer (Bantam Spectra Book)
 by Neal Stephenson
 Use to make recommendations

Εικόνα 25 Παράδειγμα εξήγησης σύστασης που χρησιμοποιεί συνεργατικό φιλτράρισμα με βάση το προϊόν

Μια από τις πιο κοινές μεθόδους για να καθορίσει την ομοιότητα είναι ο υπολογισμός γωνίας συνημίτονου. Το σύστημα σύστασης Amazon.com χρησιμοποιεί το μέτρο συνημίτονου για να αποφασιστεί η ομοιότητα μεταξύ κάθε δύο στοιχείων που αγοράζονται από κάθε πελάτη και για να καθιερωθεί η μήτρα στοιχείων, η οποία περιέχει τις σχέσεις στοιχείο προς στοιχείο.

4.7 YouTube

Πρόσφατα, κάποια από τα δημοφιλέστερα με την μεγαλύτερη συμμετοχή Κοινωνικά Μέσα έχουν προσθέσει στις υπηρεσίες που προσφέρουν και μηχανές συστημάτων προτάσεων, όπως το YouTube που πλέον στην αρχική του σελίδα συμπεριλαμβάνει προτάσεις για video βάσει όσων έχει δει προηγουμένως ο χρήστης και όσων έχει δηλώσει ως αγαπημένα. Η εφαρμογή αυτή είχε ως αποτέλεσμα να αυξηθεί ο αριθμός των χρηστών που επισκέπτονται την αρχική σελίδα, η συχνότητα των επισκέψεων στο YouTube και ο αριθμός των πελατών που εγγράφονται στην συγκεκριμένη πλατφόρμα.



Εικόνα 26 Παράδειγμα εξήγησης σύστασης που χρησιμοποιεί συνεργατικό φιλτράρισμα με βάση το προϊόν

Ορισμένα από τα καλύτερα παραδείγματα συστημάτων προτάσεων όπως αυτό του YouTube εφαρμόζουν ευρέως συνεργατικό φιλτράρισμα με βάση την μνήμη για την εξαγωγή προβλέψεων. Ο βασικός λόγος που συμβαίνει αυτό είναι επειδή το Συνεργατικό Φιλτράρισμα στηρίζεται στους χρήστες, σε παρελθοντικές προτιμήσεις τους και σε κοινά ενδιαφέροντα με άλλους χρήστες και όχι στο περιεχόμενο των προϊόντων, το οποίο αλλάζει δυναμικά και πολύ συχνά δεν είναι αρκετά σαφές ώστε να οδηγήσει σε ακριβείς προβλέψεις. Επιπλέον, η επιλογή memory-based

συστημάτων οφείλεται μεταξύ άλλων στο ότι είναι εύκολα και απλά τόσο στην δημιουργία όσο και στην εφαρμογή τους, καθώς και στο ότι είναι σταθερά και δεν επηρεάζονται σε σημαντικό βαθμό από το συνεχώς μεταβαλλόμενο dataset (προσθήκη νέων χρηστών, προϊόντων ή βαθμολογιών).

Η εφαρμογή αυτή αποτελεί τμήμα της εφαρμογής πλοήγησης και αναλαμβάνει τη δημιουργία και εμφάνιση στο χρήστη συντομεύσεων, που προκύπτουν μετά από σύγκριση της τρέχουσας πλοήγησής του με δημοφιλή αποθηκευμένα τμήματα πλοηγήσεων.

Όσο ο χρήστης πλοηγείται στην πύλη, το σύστημα εξετάζει ένα παράθυρο ν-τελευταίων κατηγοριών που έχει επισκεφθεί (navigation window). Εάν οι κατηγορίες αυτές είναι δημοφιλείς - δηλαδή έχουν ανιχνευθεί δημοφιλή subpatterns μήκους 1 που ταυτίζονται με αυτές-, τότε ελέγχεται κατά πόσο αυτή η ακολουθία κατηγοριών εντοπίζεται σε δημοφιλή subpatterns μήκους n+1 που είναι αποθηκευμένα στη βάση. Εάν εντοπιστούν τέτοια subpatterns, τότε δημιουργούνται υποψήφια συντομεύσεις. Αν το confidence της μετάβασης προς την υποψήφια συντόμευση είναι μεγαλύτερο από ένα κάτω όριο, τότε η συντόμευση αυτή εμφανίζεται στο χρήστη. Όμοια και για την περίπτωση των δημοφιλών L-subpatterns, δηλαδή των ακολουθιών κατηγοριών στις οποίες οι χρήστες έχουν επιλέξει τουλάχιστον ένα resource. Το διάγραμμα ροής δεδομένων της εφαρμογής online personalization φαίνεται στην ακόλουθη εικόνα.

The image shows a screenshot of a YouTube interface. On the left is a sidebar for user 'Manos Vomvilas' with options like 'Watch Later', 'Watch History', 'Playlists', 'What to watch', 'My subscriptions', and 'Social'. The main area is titled 'Recommended for you' and displays four video thumbnails. The first is 'Moby "One of These Mornings"' (3:16), the second is 'Massive Attack - Mezzanine (full album)' (1:03:45), the third is a news clip (0:13), and the fourth is another news clip (2:41).

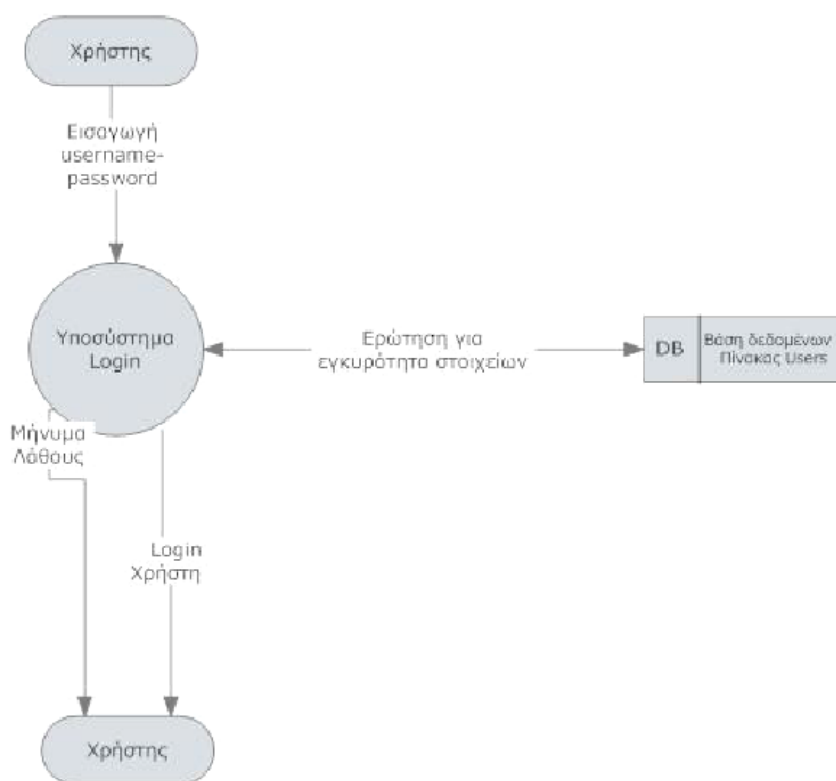
Επιχείρηση	Είδος Φιλτραρίσματος	Τομέας Ανάπτυξης	Σύστημα ERP	Χαρακτηρισμός
Εφημερίδα Guardian	Διήθηση με βάση το περιεχόμενο	Social Media	Αρχειοθέτηση άρθρων και αναγνωστών	Λειτουργεί έως σήμερα, χωρίς ιδιαίτερη καινοτομία
Cinematch	Συνδιαστική διήθηση με βάση το προϊόν και το πελάτη	e-commerce	Αρχειοθέτηση Ταινιών, Χρηστών, Συσχέτιση Χρηστών	Καινοτομεί στο χώρο, αλλά χρήζει βελτίωσης
Twitter	Συνδιαστική διήθηση με βάση τον πελάτη	Social Network	Συσχέτιση Χρηστών	Καινοτομεί και αναπτύσσεται
eBay	Διήθηση με βάση το περιεχόμενο	e-commerce	Αρχειοθέτηση προϊόντων και πελατών	Λειτουργεί έως σήμερα
Amazon	Συνδιαστική διήθηση με βάση το προϊόν	e-commerce	Αρχειοθέτηση προϊόντων και πελατών και Συσχέτιση Πελατών	Καινοτομεί και αναπτύσσεται
YouTube	Συνδιαστική διήθηση με βάση το προϊόν	Social Media	Αρχειοθέτηση Δεδομένων, Χρηστών και συσχέτιση δεδομένων	Καινοτομεί και αναπτύσσεται

4.8 Μελέτη Ηλεκτρονικού Καταστήματος

Στα πλαίσια της έρευνας, εξετάσαμε το σύστημα διαχείρισης πελατών και τα δεδομένα που προκύπτουν από την ανάλυσή τους. Για λόγους διασφάλισης της εταιρείας σε θέματα ανταγωνισμού, δεν δίνονται τα στοιχεία της εταιρείας.

4.8.1 Εφαρμογή εισόδου εγγεγραμμένου χρήστη στο σύστημα (login)

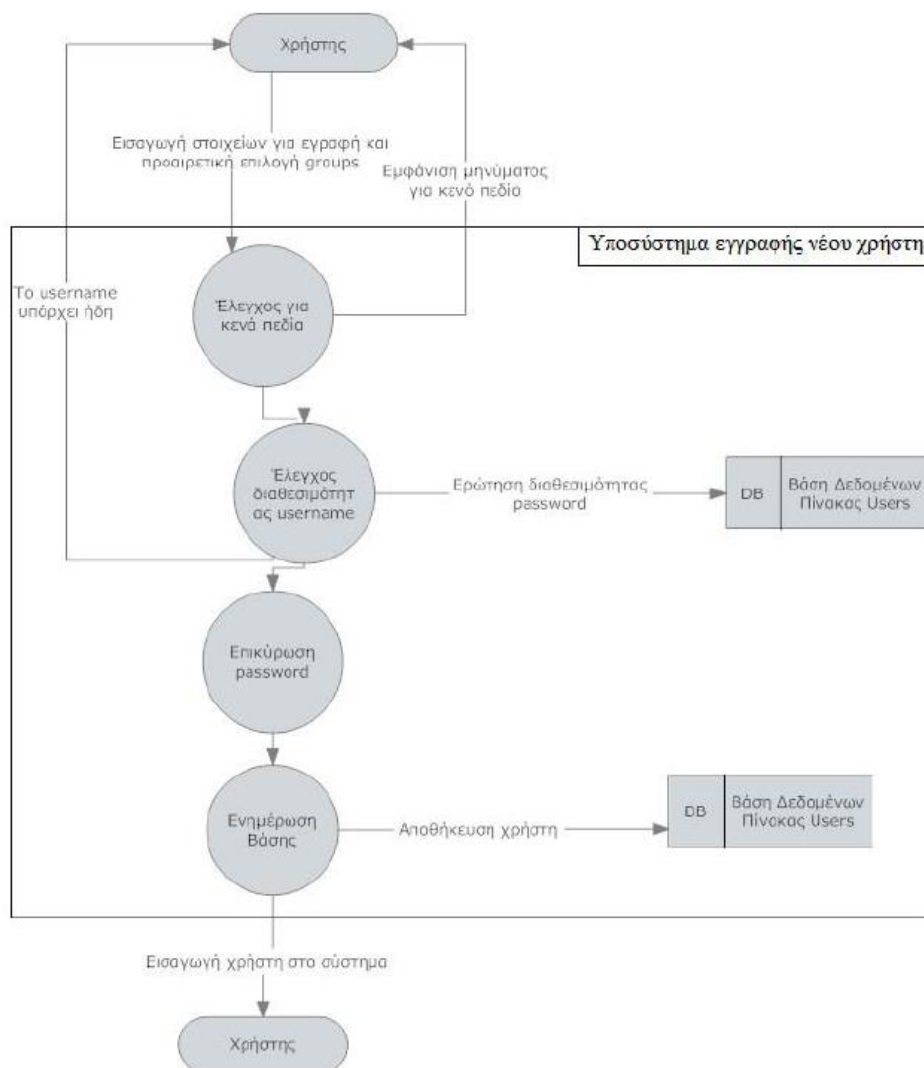
Η εφαρμογή αυτή είναι υπεύθυνη για την είσοδο ενός ήδη εγγεγραμμένου χρήστη του συστήματος σε αυτό (login). Όταν ο χρήστης δώσει τα αναγνωριστικά του στοιχεία (username και password), εξετάζεται – μετά από επικοινωνία με τη βάση δεδομένων – αν αντιστοιχούν σε εγγεγραμμένο χρήστη. Αν ναι, εισάγεται στο σύστημα αρχίζοντας την πλοήγησή του στην πύλη. Επίσης προσφέρεται και η εναλλακτική, αν δεν είναι εγγεγραμμένος χρήστης, να συμπληρώσει τη φόρμα εγγραφής και να καταχωρηθεί στο σύστημα. Το διάγραμμα ροής δεδομένων της εφαρμογής φαίνεται στην επόμενη εικόνα



Εικόνα 27 Εφαρμογή εισόδου επισκέπτη

4.8.2 Εφαρμογή εγγραφής νέου χρήστη στο σύστημα (signup)

Η εφαρμογή αυτή είναι υπεύθυνη για την εγγραφή ενός νέου χρήστη στο σύστημα. Ο χρήστης συμπληρώνει τα απαραίτητα αναγνωριστικά στοιχεία (username, password, όνομα, e-mail address), ενώ προαιρετικά μπορεί να επιλέξει τις ομάδες ενδιαφέροντος στις οποίες θέλει να γίνει μέλος και οι οποίες χαρακτηρίζονται από τις δημοφιλέστερες θεματικές κατηγορίες για κάθε μία από αυτές. Για την εμφάνιση των δημοφιλέστερων κατηγοριών για κάθε ομάδα ενδιαφέροντος, καλείται η αντίστοιχη εφαρμογή που αναζητά τα πιο δημοφιλή τμήματα πλοηγώσεων μήκους 1. Με την προϋπόθεση ότι έχει δώσει έγκυρα στοιχεία (το username δεν υπάρχει ήδη στη βάση δεδομένων), ο νέος χρήστης εγγράφεται στο σύστημα και ενημερώνεται η βάση δεδομένων. Το διάγραμμα ροής δεδομένων της εφαρμογής φαίνεται στην επόμενη εικόνα.



Εικόνα 28 Εφαρμογή εγγραφής νέου χρήστη

4.8.3 Εφαρμογή πλοήγησης

Η εφαρμογή αυτή αναλαμβάνει τη μετατροπή της πύλης από τη μορφή σχεσιακής βάσης σε κανονική ιστοσελίδα και την παρουσίαση στο χρήστη, ο οποίος μπορεί να περιηγηθεί στις θεματικές κατηγορίες.

Για κάθε θεματική κατηγορία που επισκέπτεται ο χρήστης, παρουσιάζονται από το σύστημα η περιγραφή της (δηλαδή σύντομη παρουσίαση του περιεχομένου της), οι υποκατηγορίες αυτής, οι related κατηγορίες και τα resources, με κλήση της αντίστοιχης εφαρμογής ανάκτησης στοιχείων της θεματικής κατηγορίας.

Σε περίπτωση που ο χρήστης είναι εγγεγραμμένος στο σύστημα και ανήκει σε κάποια ομάδα χρηστών, τα παραπάνω παρουσιάζονται ταξινομημένα σύμφωνα με κάποια από τις εξής μετρικές:

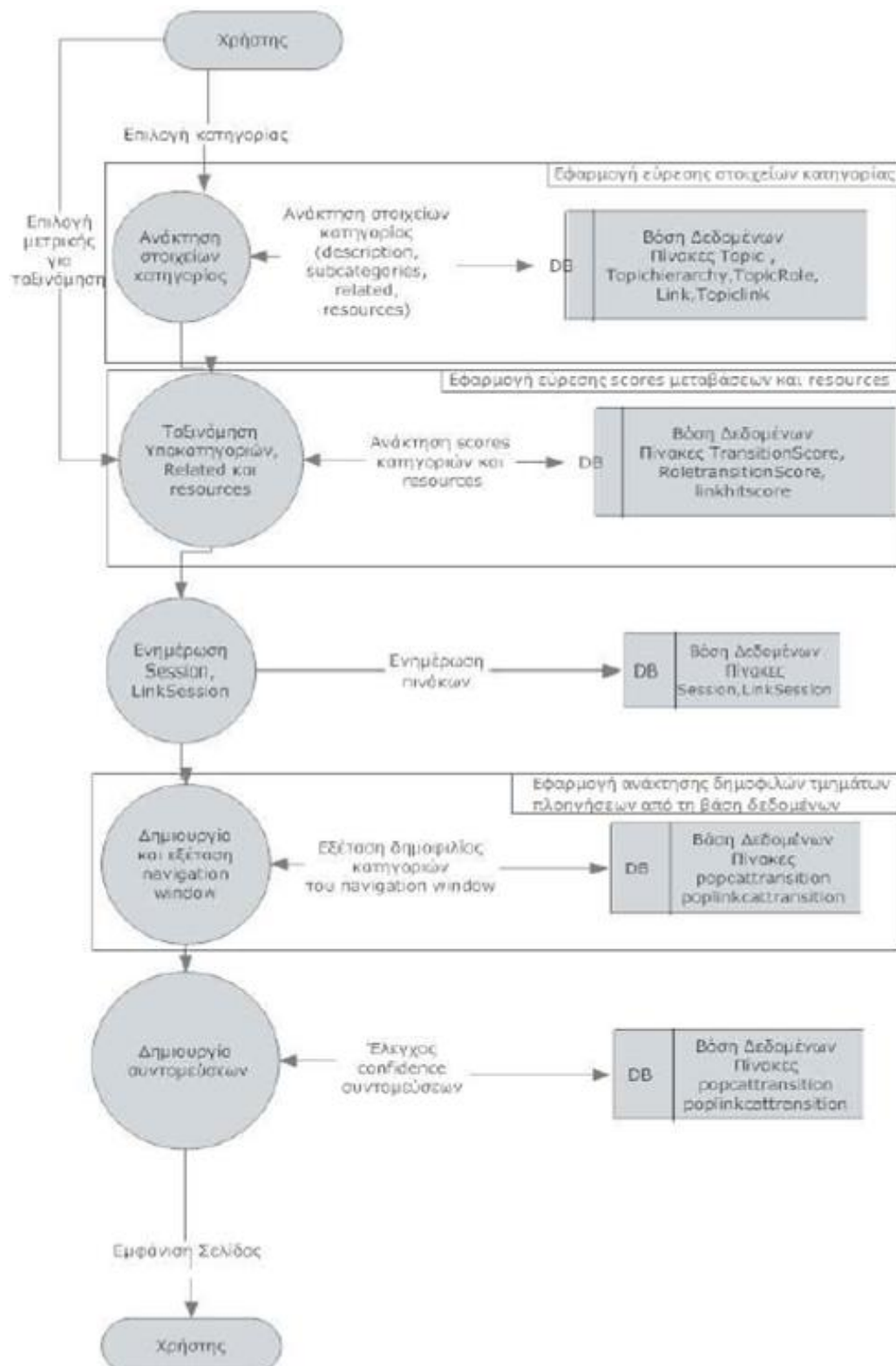
Transition Popularity μετάβασης A->B

Το πλήθος των άμεσων (χωρίς να μεσολαβεί περιήγηση σε άλλες κατηγορίες) μεταβάσεων από μια κατηγορία A σε μία κατηγορία B, είτε αυτές έχουν σχέση «category-subcategory» είτε «category-related category». Η Transition Popularity, δηλαδή το score της μετάβασης A->B, είναι διαφορετική για κάθε ομάδα χρηστών.

Transition L-Popularity μετάβασης A->B

Ο λόγος των επισκέψεων σε resources της κατηγορίας B, ύστερα από απλές μεταβάσεις A->B, προς το γινόμενο ($\#resources(B)$)*($\#μεταβάσεων A->B$). Η διαίρεση με το πλήθος το γινόμενο αυτό γίνεται για λόγους κανονικοποίησης, ώστε να μη δίνεται άδικα υψηλή κατάταξη σε κατηγορίες που συγκεντρώνουν μεγάλο αριθμό ετερογενών resources, άρα και μεγάλη επισκεψιμότητα. Και η Transition L-Popularity είναι διαφορετική για κάθε group χρηστών.⁸

⁸ Η μετρική που αφορά την κατάταξη των resources της κάθε κατηγορίας είναι η απλή δημοφιλία των resources, δηλαδή ο αριθμός των επισκέψεων σε κάθε ένα από αυτά για κάθε ομάδα χρηστών.



Εικόνα 29 Εφαρμογή πλοήγησης

Ο χρήστης μπορεί να επιλέξει τη μετρική με βάση την οποία επιθυμεί να γίνεται η ταξινόμηση των υποκατηγοριών και των related κατηγοριών, αφού του παρουσιάζονται οι επιλογές από το σύστημα. Ανάλογα με την επιλογή του, καλείται η κατάλληλη εφαρμογή που αναζητά στη βάση δεδομένων το score για κάθε κατηγορία. Η προεπιλογή είναι η Transition Popularity. Για την ταξινόμηση των links, η μετρική είναι σταθερή και ισούται με τις φορές που έχει επιλεγεί το κάθε link για τις ομάδες που ανήκει ο χρήστης.

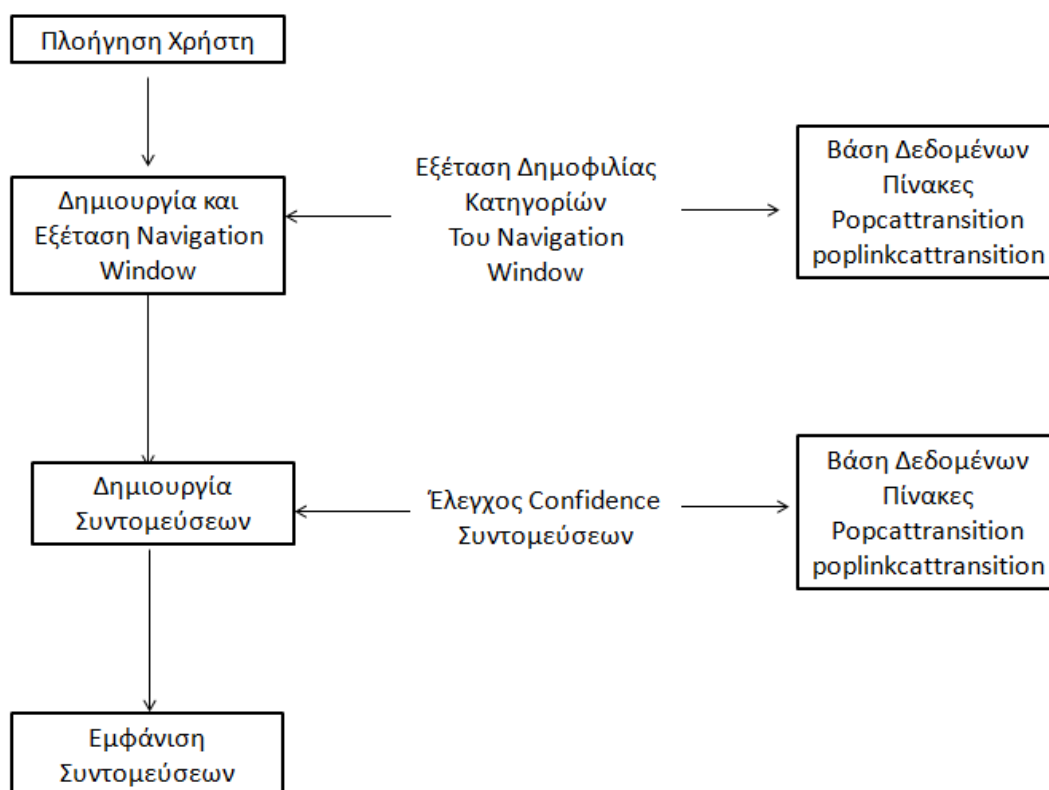
Η εφαρμογή πλοήγησης αναλαμβάνει να δημιουργήσει ένα νέο session όταν ξεκινά η πλοήγηση του χρήστη, να το αποθηκεύσει στη βάση δεδομένων και να το ενημερώνει για κάθε νέα θεματική κατηγορία που αυτός επισκέπτεται. Η ενημέρωση του session αφορά την αλλαγή του χρόνου λήξης της συνεδρίας, καθώς και την προσθήκη της θεματικής κατηγορίας στο navigation pattern που σχηματίζει η αλυσίδα των κατηγοριών που έχει ήδη επισκεφθεί ο χρήστης. Σε περίπτωση που ο χρήστης πατήσει κάποιο link, το υποσύστημα αυτό ενημερώνει τον κατάλληλο πίνακα της βάσης, για να υπάρχει η πληροφορία ότι στην τρέχουσα συνεδρία επελέγη το συγκεκριμένο link. Κατά τη διάρκεια της πλοήγησης του χρήστη, ενημερώνονται και οι πίνακες της βάσης που συγκεντρώνουν τα scores, τα οποία αντιπροσωπεύουν τις μετρικές ταξινόμησης των κατηγοριών και των resources.

Το διάγραμμα ροής δεδομένων της εφαρμογής πλοήγησης φαίνεται στην προηγούμενη εικόνα. Στο διάγραμμα παρατηρούμε ότι η εφαρμογή πλοήγησης χρησιμοποιεί πολλές εφαρμογές ανάκτησης πληροφοριών από τη βάση δεδομένων, οι οποίες ανήκουν στο υποσύστημα εξόρυξης δεδομένων και περιγράφονται παρακάτω.

4.8.4 Εφαρμογή online personalization

Η εφαρμογή αυτή αποτελεί τμήμα της εφαρμογής πλοήγησης και αναλαμβάνει τη δημιουργία και εμφάνιση στο χρήστη συντομεύσεων, που προκύπτουν μετά από σύγκριση της τρέχουσας πλοήγησής του με δημοφιλή αποθηκευμένα τμήματα πλοηγήσεων.

Όσο ο χρήστης πλοηγείται στην πύλη, το σύστημα εξετάζει ένα παράθυρο ν-τελευταίων κατηγοριών που έχει επισκεφθεί (navigation window). Εάν οι κατηγορίες αυτές είναι δημοφιλείς - δηλαδή έχουν ανιχνευθεί δημοφιλή subpatterns μήκους 1 που ταυτίζονται με αυτές-, τότε ελέγχεται κατά πόσο αυτή η ακολουθία κατηγοριών εντοπίζεται σε δημοφιλή subpatterns μήκους n+1 που είναι αποθηκευμένα στη βάση.



Εικόνα 30 Εφαρμογή on line personalization

Εάν εντοπιστούν τέτοια subpatterns, τότε δημιουργούνται υποψήφιες συντομεύσεις. Αν το confidence της μετάβασης προς την υποψήφια συντόμευση είναι μεγαλύτερο από ένα κάτω όριο, τότε η συντόμευση αυτή εμφανίζεται στο χρήστη. Όμοια και για

την περίπτωση των δημοφιλών L-subpatterns, δηλαδή των ακολουθιών κατηγοριών στις οποίες οι χρήστες έχουν επιλέξει τουλάχιστον ένα resource. Το διάγραμμα ροής δεδομένων της εφαρμογής online personalization φαίνεται στην ακόλουθη εικόνα.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Η ραγδαία εξέλιξη του διαδικτύου, έχει επιφέρει το φαινόμενο της υπερπληροφόρησης. Οι χρήστες του διαδικτύου, αδυνατούν πολλές φορές να παρακολουθήσουν τα δρώμενα και τις εξελίξεις, για θέματα που τους ενδιαφέρουν, όπως επίσης και να διαχειριστούν τον τεράστιο όγκο δεδομένων. Τα συστήματα συστάσεων, σχεδιάστηκαν με σκοπό να διηθήσουν και να οργανώσουν την πληθώρα αυτής της πληροφορίας, όπως επίσης και για να καταστήσουν ευκολότερη την εύρεση χρήσιμης πληροφορίας, για τους χρήστες. Πλέον, οι αλγόριθμοι συστάσεων, υποστηρίζονται από πολλές γνωστές εφαρμογές, όπως είναι το youtube.com ή η amazon.com και ο ρόλος τους είναι να προτείνουν στους χρήστες «προϊόντα», αντίστοιχα με τα ενδιαφέροντά τους. Υπάρχουν διάφοροι αλγόριθμοι συστάσεων. Οι κύριες κατηγορίες στις οποίες χωρίζονται, είναι οι αλγόριθμοι βάσει περιεχομένου (content based) και οι αλγόριθμοι βάσει συνεργατικής διήθησης (user/item based). Καθεμία από αυτές, έχει τα πλεονεκτήματα και τα μειονεκτήματά της και είναι κατάλληλη για συγκεκριμένες περιστάσεις.

Τα Συστήματα Προτάσεων ευνοούν τόσο τους υπεύθυνους παροχής ηλεκτρονικών υπηρεσιών όσο και τους χρήστες των υπηρεσιών αυτών. Τα άτομα που παρέχουν υπηρεσίες και προϊόντα ηλεκτρονικά μέσω του διαδικτύου έχουν πολλά πλεονεκτήματα από την αξιοποίηση ενός τέτοιου συστήματος, με βασικότερο την αύξηση των πωλήσεων των προϊόντων τους. Αυτό οφείλεται στο γεγονός ότι οι προτάσεις του συστήματος ταιριάζουν με τις ανάγκες και τα ενδιαφέροντα του χρήστη. Το χαρακτηριστικό αυτό των συστημάτων προτάσεων συμβάλει στην ικανοποίηση του χρήστη, καθώς οι ενδιαφέρουσες συστάσεις βελτιώνουν την εμπειρία του, και στην αύξηση της εμπιστοσύνης του στο σύστημα. Επίσης, μία λειτουργία που καθιστά πετυχημένο ένα Recommender System είναι η δυνατότητά του να προτείνει στον χρήστη προϊόντα που θα δυσκολευόταν να βρει από μόνος του, γιατί για παράδειγμα δεν ανήκουν στην λίστα με τα πιο δημοφιλή προϊόντα.

Σε περιπτώσεις στις οποίες ο αριθμός των χρηστών ξεπερνάει κατά πολύ των αριθμό των προϊόντων (οι περιπτώσεις αυτές είναι και οι συχνότερες στις διαδικτυακές εφαρμογές συστημάτων προτάσεων), τότε τα συστήματα προτάσεων με βάση τα προϊόντα δείχνουν να έχουν καλύτερες επιδόσεις. Σε αυτή τη διαπίστωση μπορούμε να στηριχτούμε για να προβλέψουμε ότι το Amazon μελλοντικά θα έχει πολύ καλλίτερα αποτελέσματα συστάσεων σε σχέση με το ebay καθώς το πρώτο

χρησιμοποιεί αλγόριθμο συνεργατικής διήθησης με βάση το προϊόν ενώ το δεύτερο χρησιμοποιεί απλή συσχέτιση με βάση το περιεχόμενο.

Στην περίπτωση οπου, το φιλτράρισμα γίνεται με βάση τον χρήστη οι προβλέψεις έχουν μεγαλύτερη ακρίβεια όταν τα αποθηκευμένα προϊόντα είναι κατά πολύ περισσότερα από τους χρήστες.

Μια μελλοντική τάση στα Συστήματα Προτάσεων είναι η εξαγωγή συστάσεων στα Κοινωνικά Μέσα. Πιστεύουμε ότι η χρήση δεδομένων που προέρχονται αυτόματα από τα κοινωνικά μέσα αλλά και η εξαγωγή προτάσεων στηριγμένα σε κοινωνικά χαρακτηριστικά αποτελούν το μέλλον στο συγκεκριμένο επιστημονικό πεδίο, και συγκεντρώνουν ήδη ιδιαίτερης προσοχής.

Βιβλιογραφία

1. **Markus Jessenitschnig and Markus Zanker** ISeller A Flexible Personalization Infrastructure for e-Commerce Applications, Lecture Notes in Computer Science, 2013.
2. **Adam F. and O'Doherty P.** , Lessons from enterprise resource-planning implementations in Ireland: towards smaller and shorter ERP project [Βιβλίο]. - [s.l.] : Journal of Information Technology V.15, 2000.
3. **Addicted A** Κατασκευή Mobile Site για Smart phones [Ηλεκτρονικό] // <http://www.addicted.gr/mobile-sites/>. - 2012.
4. **Aha D. Bankert R.**, A Comparative Evaluation of Sequential Feature Selection Algorithms [Βιβλίο]. - [s.l.] : in proceedings of AI & Statistics Workshop, 1995.
5. **Altec Software A** Τεχνικά Χαρακτηριστικά Altec Software Atlantis ERP [Βιβλίο]. - Αθήνα : [s.n.], 2008.
6. **Auriol E. Manago M. , Althoff S. Wess S., Dittrich S.** , Intergrating Induction and Case Based Reasoning [Βιβλίο]. - France : Springer, 1994.
7. **Burke R. B** Hybrid Recommender Systems. Survey and Experiments. User Modeling and User-Adapted Interaction [Βιβλίο]. - 2002.
8. **Deloitte Digital D** The Dawn of Mobile Influence [Βιβλίο]. - USA : Delloite Digital, 2012.
9. **Dien D. D** E-business development for competitive advantagesQ a case study [Βιβλίο]. - [s.l.] : Information and Management, 2002.
10. **Ellsworth J. & Ellsworth W. E** Επιχειρηματικές Εφαρμογές με το Internet [Βιβλίο]. - Αθήνα : Εκδόσεις Γκιούρδας, 1997.

11. **Forrester F** Mobile Commerce Forecast 2011 to 2012 Forrester Report [Βιβλίο]. - USA : [s.n.], 2012.
12. **H. Marmanis D. Babenko** , Algorithms of the Intelligent Web, in Artificial Intelligence [Βιβλίο]. - 2008.
13. **Hausman A. H** A multi-method investigation og consumer motivations in impulse buying behavior [Βιβλίο]. - Journal of Consumer Marketing, 2000. - Τόμ. 17.
14. **info.magento.com** Magento
http://info.magento.com/rs/magentocommerce/images/Magento_Mobile_Data_sheet.pdf, 2011.
15. **Manber U. Patel A. , Robinson J. ,** Experience with Personalization on Yahoo! [Βιβλίο]. - [s.l.] : Communications of the ACM, 2000.
16. **Mobasher B. Cooley R., Srivastava J.** Automatic Personalization Based on Web Usage MINING [Βιβλίο]. - [s.l.] : Communications of the ACM, 2000.
17. **Montainer M. Beatriz Lopez, Josef De la Rosa** , A taxonomy of Recommender agents on the internet [Βιβλίο]. - [s.l.] : Artificial Intelligence Review, 2003.
18. **Sven R. Beck M. Freitag B. ,** Generating Recommendation Dialogues from Product Models [Βιβλίο]. - Passau : University of Passau - Germany, 2003.
19. **Synergic Software S** Κατασκευή ιστοσελίδων για smart phones και tablets [Ηλεκτρονικό] // <http://www.synergic.gr/blog>. - 2012.
20. **T. Wailgum T** EPR Definition and Solutions [Ηλεκτρονικό] // www.cio.com. - 18 3 2013.

21. **Turban E., J. Lee, D. King, and H.M. Chung** , Electronic commerce: a managerial perspective [Βιβλίο]. - NJ : Prentice Hall Upper Saddle River, 2000.
22. **www.espa.gr** [Ηλεκτρονικό]. - 2013.
23. **www.magentocommerce.com** [Ηλεκτρονικό]. - 2013.
24. **Z. I. Magabe Z** Open Access Technology [Βιβλίο]. - Stockholm : Royal Institute of Technology, 2006.
25. **Α.Σ.Αδριανοπούλου Β. Ασίκη, Ε. Βασιλειάδη, Ι. Μίνη, Γ. Παναγιωτοπούλου, Ι. Παπακυριακοπούλου** , Τα Πληροφοριακά Συστήματα Enterprise Resource Planning (ERP) στην Ελληνική Επιχείρηση [Βιβλίο]. - Αθήνα : [s.n.], 2000.
26. **Αθανασάκης Εμ. Α** Διείδυση και Ανάπτυξη του Ηλεκτρονικού Εμπορίου στις Ελληνικές Επιχειρήσεις [Βιβλίο]. - Κρήτη : Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης, 2012.
27. **Β. Γκιντσιούδης Β** Σχεδίαση και Ανάπτυξη Εξατομικευμένης Εφαρμογής Κινητού Ηλεκτρονικού Εμπορίου [Βιβλίο]. - Θεσσαλονίκη : Πανεπιστήμιο Μακεδονίας - Μεταπτυχιακή Εργασία, 2008.
28. **Δ. Παρούτσας Δ** Αναζήτηση Πληροφοριών στον Παγκόσμιο Ιστό [Ηλεκτρονικό] // Η Εκπαίδευση στο Δημοτικό Σχολείο. - <http://paroutsas.jmc.gr/search.htm>, 2013.
29. **Διάλεξη 10η : Κινητό Εμπόριο Δ** Ψηφιακό Περιεχόμενο και Ηλεκτρονικό Εμπόριο (Δ' Εξάμηνο) [Βιβλίο]. - Πανεπιστήμιο Αιγαίου : Σχολή Κοινωνικών Επιστημών , 2011.
30. **Καρακατσούλης Δ. Κ** Υλοποίηση Ηλεκτρονικού Καταστήματος YouBooks [Βιβλίο]. - Πάτρα : Πανεπιστήμιο Πατρών Πολυτεχνική Σχολή - Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής, 2011.

31. **Κατσίκας Σ. & Μήτρου Λ. Κ** Ασφάλεια Πληροφοριακών & Επικοινωνιακών Συστημάτων στο Χώρο του Ηλεκτρονικού Επιχειρείν [Βιβλίο]. - Αθήνα : Ομάδα Εργασίας Β1 του e-businessforum, 2002. - Τόμ. διαθέσιμο στον διαδικτυακό τόπο "www.ebusinessforum.gr".
32. **Κατσουλάκος Γ. Κ** Νέα Οικονομία, Διαδίκτυο και Ηλεκτρονικό Εμπόριο [Βιβλίο]. - Αθήνα : Εκδόσεις Κέρκυρα, 2001.
33. **Λ. Στάμκου Α** Πληροφοριακά Συστήματα στο Λιανικό Εμπόριο [Βιβλίο]. - Θεσσαλονίκη : Μεταπτυχιακό Τμήμα Εφαρμοσμένης Επιχειρηματικής Πληροφορίας, 2011.
34. **Μ. Κωνσταντίνου Μ** Συστήματα Συστάσεων σε Ηλεκτρονικά καταστήματα Λιανικής [Βιβλίο]. - Λάρισα : Τμήμα Τεχνολογίας Πληροφορικής και Τηλεπικοινωνιών, 2012.
35. **Μ. Ρήγκου Μ** Personalization, Τεχνολογίες & Υπηρεσίες [Βιβλίο]. - Πάτρα : Πανεπιστήμιο Πατρών, 2012.
36. **Οικονόμου & Αργυρόπουλος Ο** [Βιβλίο]. - 1995.
37. **Π. Παναγιωτόπουλος Π** Εφαρμογή Πολυκριτήριας Μεθοδολογίας ΑHP για την Επιλογή ERP [Βιβλίο]. - Αθήνα : Εθνικό Μετσόβιο Πολυτεχνείο, 2007.
38. **Πασχόπουλος Α. Π** Ηλεκτρονικό Εμπόριο [Βιβλίο]. - Αθήνα : Εκδόσεις Κλειδάριθμος, 2007.
39. **Πασχόπουλος Α. Σκάλτσας Π. ,** Ηλεκτρονικό Εμπόριο [Βιβλίο]. - Αθήνα : Κλειδάριθμος 2η Έκδοση, 2001.
40. **Σιωμίκος Γ.** Συμπεριφορά Καταναλωτή και Στρατηγικό Μάρκετινγκ [Βιβλίο]. - Αθήνα : Εκδόσεις Σταμούλη, 2002.

41. **Σωχωράκη Ε. Ραφαηλία Β.** , Ηλεκτρονικό Εμπόριο μέσω Κινητών Τηλεφώνων, Μελέτη για την Ασφάλεια των Συναλλαγών [Βιβλίο]. - Ηράκλειο : Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης, 2009.
42. **Τζιάστα Α.** Ηλεκτρονικό Εμπόριο και Ηλεκτρονικό Μάρκετινγκ [Βιβλίο]. - Αθήνα : Χαροκόπειο Πανεπιστήμιο - Τμήμα Οικιακής Οικονομίας και Οικολογίας, 2011.
43. **Τσόπογλου Σ.** Συγκριτική Ανάλυση και Μελέτη ERP Συστημάτων - Μακεδονία : Μεταπτυχιακό Πρόγραμμα Εφαρμοσμένης Πληροφορικής - Πανεπιστήμιο Μακεδονίας, 2013.
44. **Φ. Μιχελινάκης** Μετάδοση σε Πραγματικό Χρόνο Ροών Πολυμέσων Πάνω απο Δίκτυα Ομοτίμων Κόμβων - Αθήνα : Εθνικό Μετσόβειο Πολυτεχνείο, 2011.
45. **Φούκη Ι Φ** Έρευνα Αγοράς για Κινητές Συσκευές [Βιβλίο]. - Λάρισα : Σχολή Τεχνολογικών Εφαρμογών - Τμήμα Τεχνολογίας Πληροφορικής και Τηλεπικοινωνιών, 2013.
46. **Χ. Χριστάκου Χ** Εισηγητικά Συστήματα Βασισμένα σε Μοντελοποίηση Προτιμήσεων Χρήστη και Μεθόδους Διήθησης της Πληροφορίας [Βιβλίο]. - Αθήνα : Εθνικό Μετσόβιο Πολυτεχνείο, 2012.
47. **Χρυσοχού Χ. Χ** Ανάλυση και Σχεδιασμός Συστημάτων ERP. Υλοποίηση και Ανάπτυξη σε Εταιρεία Επεξεργασίας Ύδατος - Αθήνα : Πανεπιστήμιο Πειραιά - Μεταπτυχιακό Πρόγραμμα Σπουδών Οργάνωση και Διοίκηση Βιομηχανικών Συστημάτων, 2008.

ΤΕΛΟΣ