



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

**ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ &
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

**Τομέας Επικοινωνιών, Ηλεκτρονικής &
Συστημάτων Πληροφορικής**

**ΑΛΓΟΡΙΘΜΙΚΗ ΠΡΟΣΕΓΓΙΣΗ ΓΙΑ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ
ΠΛΗΡΟΦΟΡΙΩΝ ΚΑΙ ΠΑΡΟΧΗ ΥΠΗΡΕΣΙΩΝ ΥΨΗΛΗΣ
ΠΡΟΣΤΙΘΕΜΕΝΗΣ ΑΞΙΑΣ ΜΕΣΩ ΔΙΑΔΙΚΤΥΟΥ**

**ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ
ΣΩΤΗΡΙΟΣ Δ. ΜΠΟΥΤΑΣ**

ΑΘΗΝΑ Ιούνιος 2012



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ &
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Τομέας Επικοινωνιών, Ηλεκτρονικής &
Συστημάτων Πληροφορικής

**Αλγοριθμική προσέγγιση για κατηγοριοποίηση πληροφοριών και
παροχή υπηρεσιών υψηλής προστιθέμενης αξίας μέσω διαδικτύου**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Σωτήριος Δ. Μπούτας

Συμβουλευτική Επιτροπή: Ελευθέριος Καγιάφας
Βασίλειος Λούμος
Ιωάννης Αβαριτσιώτης

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή

.....
Ε. Καγιάφας
Καθηγητής
Ε.Μ.Π.

.....
Β. Λούμος
Καθηγητής
Ε.Μ.Π.

.....
Ι. Αβαριτσιώτης
Καθηγητής
Ε.Μ.Π.

.....
Μ. Ιωαννίδου
Καθηγήτρια
Ε.Μ.Π.

.....
Κ. Παπαοδυσσεύς
Αναπληρωτής Καθηγητής
Ε.Μ.Π.

.....
Η. Κουκούτσης
Επίκουρος Καθηγητής
Ε.Μ.Π.

.....
Ι. Αναγνωστόπουλος
Επίκουρος Καθηγητής
Παν. Στερεάς Ελλάδος

Αθήνα, Ιούνιος 2012

Σωτήριος Δ. Μπούτας

Διδάκτωρ Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Σωτήριος Δ. Μπούτας, 2012

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ'ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Η διδακτορική διατριβή αφιερώνεται στους γονείς μου Δημήτρη και Άννα, στα αδέρφια μου Αγορίτσα, Ηλία και Παντελή, καθώς και στην αρραβωνιαστικιά μου Λήδα για την αμέριστη συμπαράστασή τους σε κάθε στάδιο της ζωής μου.

ΕΥΧΑΡΙΣΤΙΕΣ

Στο σημείο αυτό, θα ήθελα να αναφερθώ σε φίλους και συνεργάτες που μου στάθηκαν πραγματικά σε όλη τη διάρκεια του διδακτορικού μου.

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου κ. Ελευθέριο Καγιάφα και τον συν-επιβλέποντα καθηγητή μου κ. Βασίλειο Λούμο, των οποίων η καθοδήγηση, η άριστη συνεργασία και η αμέριστη βοήθεια υπήρξαν καταλυτικοί παράγοντες στο ερευνητικό μου έργο. Σας ευχαριστώ για την κατανόηση στις δύσκολες στιγμές και την εμπιστοσύνη που μου δείξατε.

Επίσης, ένα μεγάλο ευχαριστώ σε όλα τα παιδιά του εργαστηρίου που συνεργάστηκα, οι οποίοι αποδείχτηκαν πραγματικοί φίλοι και συνεργάτες. Ιδιαίτερα όμως θα ήθελα να ευχαριστήσω τον φίλο μου Ιωάννη Αναγνωστόπουλο για τη μεγάλη του συνεισφορά και συμπαράσταση στο διδακτορικό μου, ο οποίος ήταν πάντα δίπλα μου όταν τον χρειαζόμουνα, καθώς επίσης και τον φίλο μου Ιωάννη Ψωρούλα που μου στάθηκε στις δύσκολες και ευχάριστες στιγμές.

Τέλος, θα ήθελα να ευχαριστήσω τους Καθηγητές Ι. Αβαριτσιώτη, Μ. Ιωαννίδου, Κ. Παπαοδυσσέα και Η. Κουκούτση για την τιμή που μου προσέφεραν συμμετέχοντας στην επταμελή επιτροπή κρίσης της διατριβής μου.

ΠΕΡΙΛΗΨΗ

Η αλματώδης ανάπτυξη του Παγκόσμιου Ιστού τα τελευταία χρόνια τον έχει καταστήσει ως βασική πηγή πληροφόρησης για εκατομμύρια ανθρώπους. Ο τεράστιος όγκος δεδομένων που διακινείται μέσω αυτού, πολύ συχνά προκαλεί σύγχυση στο μέσο χρήστη, ο οποίος αδυνατεί να επιλέξει αυτό που πραγματικά τον ενδιαφέρει. Εδώ είναι που πρέπει διάχυτες τεχνικές πληροφορικής να παρέχουν τις σχετικές πληροφορίες σχετικά με προϊόντα και υπηρεσίες που σχετίζονται με τις ανάγκες των χρηστών. Η διατριβή αυτή παρουσιάζει ένα τέτοιο σύστημα, ο κύριος στόχος του οποίου είναι να επιλέξει, ανάμεσα από ένα σύνολο χρηστών, αυτούς που είναι πιο πιθανό να ενδιαφέρονται για συγκεκριμένα γεγονότα (events). Για την προσέγγιση, εφαρμόστηκε ένα εκτεταμένο σύστημα πληροφοριών μαζί με μία Web 2.0 γεωχωρική τεχνολογία API, το οποίο δοκιμάστηκε με προσομοιωμένα, αλλά και με αληθινά δεδομένα.

Λέξεις-κλειδιά: Φιλτράρισμα Πληροφορίας, Συστήματα Συστάσεων, Νευρωνικά Δίκτυα, Συστήματα Διαχείρισης Βάσεων Δεδομένων, Γεωκωδικοποίηση, Αλγόριθμοι Βαθμολόγησης

ABSTRACT

Due to its rapid development in the recent years, World Wide Web has become the basic source of information for millions of people around the world. However, the huge volume of data transferred through the Web often causes confusion to users, who are often unable to select the kind of information that is actually suitable to their needs. That is where ubiquitous computing techniques should provide relevant information regarding products and services related to the users' needs. This paper presents such a system, the main goal of which is to select among a set of users, those who are most likely to be interested in specific events. For our approach we employed an information system along with web 2.0 geospatial API technology and we tested it over simulated and real-life data.

Keywords: Information Filtering, Recommender Systems, Neural Networks, DBMS, Geocoding, Ranking Algorithms

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΕΥΧΑΡΙΣΤΙΕΣ	5
ΠΕΡΙΛΗΨΗ	6
ABSTRACT	7
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ	8
ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ - ΕΙΚΟΝΩΝ	12
ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ	15
1 ΕΙΣΑΓΩΓΗ	16
1.1 Αντικείμενο της διατριβής.....	16
1.2 Ανάλυση κεφαλαίων	17
2 ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ.....	19
2.1 Φιλτράρισμα Πληροφορίας και Συστήματα Προτάσεων	19
2.2 Κατηγοριοποιήσεις Συστημάτων Προτάσεων.....	20
2.2.1 Βάσει του Τρόπου Παραγωγής των Συστάσεων	20
Collaborative Filtering (Συνεργατικό Φιλτράρισμα).....	21
Content-Based Filtering (Φιλτράρισμα Βασισμένο στο Περιεχόμενο)	22
Demographic Filtering (Δημογραφικό Φιλτράρισμα)	22
Knowledge-Based Filtering (Φιλτράρισμα βασισμένο στη Γνώση).....	22
Υβριδικές (Hybrid) Μέθοδοι	23
2.2.2 Βάσει του Τρόπου Απόκτησης Γνώσης	24
2.2.3 Βάσει της Πρωτοβουλίας Δράσης (<i>Initiative of Operation</i>)	25
Ενεργά (Active) Συστήματα.....	25
Παθητικά (Passive) Συστήματα.....	25
2.3 Γεωγραφική Πληροφορία.....	26
2.4 Ταξινόμηση και Ταξινομητές.....	27
Περιγραφική Μοντελοποίηση	27
Προβλεπτικό Μοντέλο.....	28
2.4.1 Γενική Διαδικασία Επίλυσης Προβλήματος Ταξινόμησης	28
2.4.2 Δοκιμή και Επαλήθευση του Μοντέλου	30

2.4.3 Δένδρα Απόφασης.....	30
2.4.4 Ταξινομητές Πλησιέστερων Γειτόνων (Nearest Neighbor).....	32
2.4.5 Μπεϋσιανός Ταξινομητής (Bayes Classifier).....	35
Ταξινομητής Naïve Bayes.....	37
Λειτουργία ενός Ταξινομητή Naïve Bayes.....	37
2.5 Δίκτυα, Τεχνητά Νευρωνικά Δίκτυα και Recommender Τεχνικές Νευρωνικών Δικτύων.....	37
2.5.1 Δίκτυα.....	37
2.5.2 Τεχνητά Νευρωνικά Δίκτυα.....	38
2.5.3 Τεχνικές Προτάσεων Νευρωνικών Δικτύων.....	41
2.6 Συστήματα Διαχείρισης Βάσεων Δεδομένων.....	42
3 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΣΥΣΤΗΜΑΤΟΣ.....	44
3.1 Αρχιτεκτονική.....	44
3.2 Σχεσιακό Διάγραμμα Οντοτήτων.....	46
3.2.1 Οντότητες.....	46
3.2.2 Σχέσεις Οντοτήτων.....	48
4 ΑΝΑΛΥΣΗ ΣΥΣΤΗΜΑΤΟΣ.....	50
4.1 Event και User Keys.....	50
4.1.1 User Key.....	51
4.1.2 Event Key.....	52
4.2 Παράγοντες που Επηρεάζουν την Επιλογή Χρηστών.....	53
4.2.1 Φύλο.....	54
4.2.2 Ηλικία.....	54
4.2.3 Τοποθεσία.....	55
4.3 Αλγόριθμοι Συστήματος.....	58
4.3.1 Ταξινόμηση και Κατηγορίες.....	58
4.3.2 Αλγόριθμος Ομαδοποίησης.....	59
4.3.3 Προτεινόμενος Ταξινομητής Νευρωνικού Δικτύου.....	61
4.3.4 Θεωρητικό Υπόβαθρο και Αντιμετώπιση του Προβλήματος.....	61
4.3.5 Εκπαιδύοντας τον ταξινομητή.....	64
5 ΥΛΟΠΟΙΗΣΗ ΣΥΣΤΗΜΑΤΟΣ.....	67
5.1 Ανάπτυξη και Λειτουργία του Ιστότοπου.....	67

5.1.1 Home	67
5.1.2 UserProfile	68
5.1.3 UserProfileSubmitted	71
5.1.4 ManagerProfile	71
5.1.5 ManagerProfileSubmitted	72
5.1.6 Login	73
5.1.7 Usermenu	73
5.1.8 UserUpdate.....	75
5.1.9 UserDelete	75
5.1.10 UserDeleted.....	76
5.1.11 EventChoiceFeedback.....	76
5.1.12 UserFeedback	77
5.1.13 UserFeedbackSubmitted	78
5.1.14 UserLogout	78
5.1.15 ManagerMenu	79
5.1.16 ManagerUpdate.....	80
5.1.17 ManagerDelete.....	80
5.1.18 ManagerDeleted.....	81
5.1.19 Event.....	81
5.1.20 EventSubmitted	83
5.1.21 EventUpdate	84
5.1.22 EventChoiceDelete	84
5.1.23 EventConfirmDelete	85
5.1.24 EventDeleted	85
5.1.25 EventChoiceResults	86
5.1.26 EventResults	87
5.1.27 ManagerLogout	88
5.2 Service.....	88
5.2.1 Περιβάλλον Λειτουργίας και Controls	88
Menu Bar.....	89
Επιλογή Event	90
Πληροφορίες Event.....	90

Αποτελέσματα Event	91
Εκτύπωση Αποτελεσμάτων και Έξοδος	92
Status Bar	92
Προϋποθέσεις Επιλεγμένων Χρηστών	92
Εμφάνιση Χρηστών στο Χάρτη	92
Tab Control.....	93
5.2.2 Παρουσίαση των Αποτελεσμάτων του Ταξινομητή.....	94
5.2.3 Λειτουργία του Service.....	95
Αλλαγή στα Δεδομένα της Βάσης.....	95
Ενημέρωση Κατάστασης των Events	96
Αποστολή Ενημερωτικών Μηνυμάτων	96
6 ΑΞΙΟΛΟΓΗΣΗ ΣΥΣΤΗΜΑΤΟΣ	97
6.1 Απόδοση Συστήματος.....	97
6.2 Απόδοση Ταξινομητή	99
6.3 Αξιολόγηση της Απόδοσης του Συστήματος.....	102
6.4 Αξιολόγηση της Απόδοσης του Ταξινομητή.....	102
7 ΕΠΙΛΟΓΟΣ	107
7.1 Μελλοντική Ανάπτυξη του Συστήματος.....	107
7.2 Συζήτηση - Εφαρμογές του Συστήματος σε Διάφορους Τομείς	109
7.3 Δημοσιεύσεις.....	110
7.3.1 Περιοδικά	110
7.3.2 Πρακτικά Συνεδρίων.....	110
8 ΒΙΒΛΙΟΓΡΑΦΙΑ	112

ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ - ΕΙΚΟΝΩΝ

Σχήμα 2 - 1: Κατηγορίες των Recommender Systems	26
Σχήμα 2 - 2: Η ταξινόμηση ως διεργασία χαρτογράφησης ενός συνόλου x χαρακτηριστικών	27
Σχήμα 2 - 3: Επίλυση Προβλήματος Ταξινόμησης.....	29
Σχήμα 2 - 4: Παράδειγμα Δένδρου Απόφασης.....	31
Σχήμα 2 - 5: Ανάθεση Πινακίδας Κατηγορίας σε	32
Σχήμα 2 - 6: Εάν περπατά σαν πάπια, φωνάζει σαν πάπια και μοιάζει με πάπια τότε είναι πάπια.....	33
Σχήμα 2 - 7: Παραδείγματα Επίλυσης με k-nearest Neighbors, για k=1, 2, ή 3.....	34
Σχήμα 2 - 8: Αρνητικοί και Θετικοί Γείτονες γύρω από το Σημείο Δεδομένων.....	35
Σχήμα 2 - 9: Μη κατευθυνόμενο δίκτυο χωρίς βάρη με 6 κορυφές.....	38
Σχήμα 2 - 10: Τεχνητός Νευρώνας.....	39
Σχήμα 2 - 11: Πολυστρωματικό Τεχνητό Νευρωνικό Δίκτυο.....	40
Σχήμα 3 - 1: Χάρτης Ιστότοπου	45
Σχήμα 3 - 2: Αναπαράσταση Διασύνδεσης Συστήματος	46
Σχήμα 3 - 3: Σχεσιακό Μοντέλο της Βάσης Δεδομένων	49
Σχήμα 4 - 1: Αναπαράσταση User Key	52
Σχήμα 4 - 2: Αναπαράσταση Event Key	52
Σχήμα 4 - 3: Όρια Γεωγραφικού Πλέγματος.....	56
Σχήμα 4 - 4: Προσομοιωμένες Περιπτώσεις Event: (a) Θεωρητικό Μοντέλο Αναπαράστασης για όλες τις Κατηγορίες, (b) Gaussian Μοντέλο Ομαδοποίησης για τα Κεντροειδή Art - Recreation	60
Εικόνα 5 - 1: Home.aspx – Αρχική Σελίδα.....	68
Εικόνα 5 - 2: UserProfile.aspx – Καταχώρηση Προσωπικών Στοιχείων.....	69

Εικόνα 5 - 3: UserProfile.aspx – Καταχώρηση Προτιμήσεων / Ενδιαφερόντων	70
Εικόνα 5 - 4: UserProfileSubmitted.aspx – Επιτυχής Καταχώρηση Προφίλ Χρήστη	71
Εικόνα 5 - 5: ManagerProfile.aspx – Καταχώρηση Προφίλ Χρήστη	72
Εικόνα 5 - 6: ManagerProfileSubmitted.aspx – Επιτυχής Καταχώρηση Προφίλ Manager	72
Εικόνα 5 - 7: Login.aspx – Είσοδος στο Σύστημα	73
Εικόνα 5 - 8: UserMenu.aspx – Μενού Χρήστη	74
Εικόνα 5 - 9: UserUpdate.aspx – Ενημέρωση Προφίλ Χρήστη	75
Εικόνα 5 - 10: UserDelete.aspx – Επιβεβαίωση Διαγραφής Προφίλ Χρήστη.....	75
Εικόνα 5 - 11: UserDeleted.aspx – Διαγραφή Προφίλ Χρήστη.....	76
Εικόνα 5 - 12: EventChoiceFeedback.aspx – Επιλογή Event για Feedback.....	76
Εικόνα 5 - 13: UserFeedback.aspx – Feedback Χρήστη για Event	77
Εικόνα 5 - 14: UserFeedbackSubmitted.aspx – Καταχώρηση Feedback Χρήστη	78
Εικόνα 5 - 15: UserLogout.aspx – Αποσύνδεση Χρήστη	78
Εικόνα 5 - 16: ManagerMenu.aspx – Μενού Manager.....	79
Εικόνα 5 - 17: ManagerUpdate.aspx – Ενημέρωση Προφίλ Manager	80
Εικόνα 5 - 18: ManagerDelete.aspx – Επιβεβαίωση Διαγραφής Προφίλ Manager	80
Εικόνα 5 - 19: ManagerDeleted.aspx – Διαγραφή Προφίλ Manager	81
Εικόνα 5 - 20: Event.aspx – Καταχώρηση Event	82
Εικόνα 5 - 21: EventSubmitted.aspx – Επιτυχής Καταχώρηση Event	83
Εικόνα 5 - 22: EventUpdate.aspx – Επιτυχής Ενημέρωση Event	84
Εικόνα 5 - 23: EventChoiceDelete.aspx – Επιλογή Event για Διαγραφή	84
Εικόνα 5 - 24: EventConfirmDelete.aspx – Επιβεβαίωση Διαγραφής Event.....	85
Εικόνα 5 - 25: EventDeleted.aspx – Διαγραφή Event	85
Εικόνα 5 - 26: EventChoiceResults.aspx – Επιλογή Event για Αποτελέσματα.....	86
Εικόνα 5 - 27: EventResults.aspx – Αποτελέσματα Event.....	87
Εικόνα 5 - 28: ManagerLogout.aspx – Αποσύνδεση Manager	88
Εικόνα 5 - 29: Controls του Service	89
Εικόνα 5 - 30: Επιλογές του Συστήματος	90
Εικόνα 5 - 31: Επιλογή Event.....	90
Εικόνα 5 - 32: Πληροφορίες Event.....	91
Εικόνα 5 - 33: Αποτελέσματα Event.....	91

Εικόνα 5 - 34: Status Bar	92
Εικόνα 5 - 35: Προϋποθέσεις Επιλεγμένων	92
Εικόνα 5 - 36: Εμφάνιση Χρηστών στο Χάρτη.....	93
Εικόνα 5 - 37: Tab Control.....	93
Εικόνα 5 - 38: Παρουσίαση των Αποτελεσμάτων στο Χάρτη	94
Σχήμα 6 - 1: Διάγραμμα Χρόνων Εκτέλεσης	98
Σχήμα 6 - 2: Αποτελέσματα με πρώτο ισχυρό κριτήριο: a) τις προτιμήσεις των χρηστών, b) την τοποθεσία των χρηστών	105

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

Πίνακας 3 - 1: Περιγραφή Οντοτήτων	47
Πίνακας 3 - 2: Περιγραφή Σχέσεων Οντοτήτων.....	49
Πίνακας 4 - 1: Κατηγορίες και Υποκατηγορίες DMOZ	59
Πίνακας 4 - 2: Στάδιο Εκπαίδευσης του Confusion Matrix.....	66
Πίνακας 6 - 1: Χρόνοι Εκτέλεσης (sec)	98
Πίνακας 6 - 2: Χρόνοι Εκτέλεσης (sec) Βάσει Περιοχής Κάλυψης.....	99
Πίνακας 6 - 3: Χρόνοι Εκτέλεσης (sec) Βάσει Περιορισμών	99
Πίνακας 6 - 4: Κατάσταση Εκπαίδευσης του Πίνακα Σύγχυσης (Confusion Matrix) για τις κατηγορίες A, B, C, D και E και για όλες τις υποκατηγορίες τους	101

1 ΕΙΣΑΓΩΓΗ

1.1 Αντικείμενο της διατριβής

Τα τελευταία χρόνια ο Παγκόσμιος Ιστός έχει εξελιχθεί εντυπωσιακά και έχει μετατραπεί σε βασικό εργαλείο πληροφόρησης για πολύ μεγάλο αριθμό ανθρώπων. Οι πληροφορίες που παρέχει είναι πρακτικά ανεξάντλητες και πολύ συχνά ο μέσος χρήστης αντιμετωπίζει πρόβλημα στην επιλογή των στοιχείων που πραγματικά τον ενδιαφέρουν. Η σύγκριση αυτή γίνεται ακόμα μεγαλύτερη στην περίπτωση που ο χρήστης θέλει να επιλέξει μέσω του Web κάποιο προϊόν ή υπηρεσία. Τα δεδομένα αυτά γέννησαν την ανάγκη ύπαρξης μηχανισμών για φιλτράρισμα της πληροφορίας και πραγματοποίηση προτάσεων στους χρήστες. Η ανάγκη αυτή σε συνδυασμό με την αλματώδη ανάπτυξη των Συστημάτων Διαχείρισης Βάσεων Δεδομένων οδήγησε στη δημιουργία των αποκαλούμενων Συστημάτων Προτάσεων (Recommender Systems). Τα συστήματα αυτά χρησιμοποιούν αλγόριθμους συσχέτισης, οι οποίοι με βάση διαφόρων ειδών πληροφορίες για τους χρήστες, επιλέγουν και προτείνουν σε αυτούς συγκεκριμένα προϊόντα ή υπηρεσίες, ανάλογα με το είδος της αναζήτησης που πραγματοποιούν.

Η παρούσα διατριβή επικεντρώνεται στο σχεδιασμό και την υλοποίηση αλγορίθμων για Φιλτράρισμα Πληροφορίας (Information Filtering), για την κατηγοριοποίηση πληροφοριών, με σκοπό τη δημιουργία ενός ολοκληρωμένου Συστήματος Προτάσεων (Recommender System) για παροχή υπηρεσιών υψηλής προστιθέμενης αξίας. Συγκεκριμένα, γίνεται μία διεξοδική μελέτη τόσο στην περιοχή του Information Filtering όσο και των Recommender Systems, αναφέρονται οι διάφορες τεχνικές που υπάρχουν και χρησιμοποιούνται μέχρι σήμερα και προτείνεται η καλύτερη δυνατή λύση που ανταποκρίνεται στις απαιτήσεις για την υλοποίηση του συστήματος.

Ο κύριος στόχος του συστήματος, είναι να επιλέξει ανάμεσα από ένα σύνολο χρηστών, αυτούς που είναι πιο πιθανό να ενδιαφέρονται για συγκεκριμένα γεγονότα (events) (π.χ. εκθέσεις, αθλητικά, τέχνη, διαφημίσεις κλπ.). Οι χρήστες δημιουργούν λογαριασμούς στο σύστημα παρέχοντας πληροφορίες για τα δημογραφικά τους στοιχεία, τη φυσική τους τοποθεσία, καθώς και πληροφορίες για τις προτιμήσεις τους πάνω σε διάφορες κατηγορίες/θέματα. Ένας διαχειριστής γεγονότος (event manager), δημιουργεί επίσης λογαριασμό στο σύστημα και εισάγει όλες τις λεπτομέρειες του γεγονότος όπως το είδος της

κατηγορίας που ανήκει, τη φυσική του τοποθεσία, πληροφορίες για τους χρήστες που στοχεύει (π.χ. ηλικία, φύλο), καθώς επίσης και την επιθυμητή απόστασή τους από το γεγονός. Σύμφωνα με την προτεινόμενη μεθοδολογία, επιλέγεται ένα σύνολο χρηστών βάσει της ηλικίας τους, το φύλο τους και την τοποθεσία τους σε σχέση με το γεγονός. Μέσα από αυτό το σύνολο χρηστών, επιλέγονται αυτοί που είναι πιθανότερο να ενδιαφέρονται για το συγκεκριμένο γεγονός, βάση ενός μηχανισμού ομαδοποίησης που λαμβάνει υπόψη τις προτιμήσεις από τα προφίλ τους. Οι προτιμήσεις τους καταγράφονται και διαχειρίζονται μέσω ενός διανύσματος που ονομάζεται Preference Event Vector (PEV) (Anagnostopoulos I. et. al. 2004).

Το πληροφοριακό σύστημα που αναπτύχθηκε, είναι σχεδιασμένο έτσι ώστε να παρέχει χρήσιμες πληροφορίες στο χρήστη, σύμφωνα με τα ενδιαφέροντά του, ώστε να αποτελέσει ένα χρήσιμο και πολύτιμο εργαλείο για αυτόν και να μη κινδυνεύει να χαρακτηριστεί ως μία ενοχλητική *spamming* υπηρεσία.

Το σύστημα αποτελείται από ένα server-side web interface μαζί με μία γεωχωρική βάση δεδομένων, χρησιμοποιεί ευφυείς μεθόδους αναζήτησης πληροφοριών (Kawuu and Der-Jiunn, 2010; Sadaqat et. al. 2011; Shangguang et. al. 2011) και παρακολουθεί παράλληλα τις αλλαγές που συμβαίνουν στη βάση δεδομένων σε πραγματικό χρόνο (Kuai et. al. 2010).

1.2 Ανάλυση κεφαλαίων

Η διατριβή είναι δομημένη σε 8 κεφάλαια. Η περίληψη της διδακτορικής διατριβής παρουσιάζεται στην Ελληνική και στην Αγγλική γλώσσα. Στη συνέχεια αναφέρονται συνοπτικά τα περιεχόμενα των επιμέρους κεφαλαίων.

Στο πρώτο κεφάλαιο περιέχεται το εισαγωγικό μέρος της διατριβής. Στο κεφάλαιο αυτό περιγράφεται με συνοπτικό τρόπο το αντικείμενο της διατριβής και ακολουθείται από μία σύντομη ανάλυση των περιεχομένων κάθε κεφαλαίου.

Στο δεύτερο κεφάλαιο, παρουσιάζεται το θεωρητικό υπόβαθρο, στο οποίο βασίζεται το σύστημα. Αρχικά, γίνεται μια αρκετά εκτενής αναφορά στους όρους Φιλτράρισμα Πληροφορίας (Information Filtering) και Συστήματα Προτάσεων (Recommender Systems), ενώ αναλύονται και οι διάφορες κατηγορίες τους. Στη συνέχεια, γίνεται αναφορά στους τρόπους ανάκτησης γεωγραφικής πληροφορίας. Ακολουθεί μία περιγραφή της θεωρίας της ταξινόμησης και των ταξινομητών, καθώς και των νευρωνικών δικτύων και αναφέρονται τεχνικές συστάσεων βασισμένες σε αυτά. Τέλος, γίνεται αναφορά στα «Σύστημα Διαχείρισης

Βάσης Δεδομένων (DBMS)» και του ρόλου που διαδραμάτισαν στην ανάπτυξη του συστήματος.

Στο τρίτο κεφάλαιο, περιγράφεται η βασική αρχιτεκτονική του συστήματος και γίνεται η ανάλυση απαιτήσεων για τις λειτουργίες του. Έτσι, μέσω κειμένου και επεξηγηματικών διαγραμμάτων (Χάρτης Ιστότοπου, Διασύνδεση Συστήματος, Σχεσιακό Διάγραμμα Βάσης) παρουσιάζεται η δομή των επιμέρους υποσυστημάτων που απαρτίζουν το σύστημα.

Στο τέταρτο κεφάλαιο γίνεται η ανάλυση του συστήματος. Περιγράφεται, ο τρόπος με τον οποίο γίνεται η ανάκτηση της τοποθεσίας των χρηστών και των events, καθώς και τους αλγόριθμους που χρησιμοποιήθηκαν.

Στο πέμπτο κεφάλαιο αναλύεται η τελική μορφή των υλοποιημένων υποσυστημάτων, μέσω κειμένου, διαγραμμάτων και screenshots από την εκτέλεση των διάφορων λειτουργιών του συστήματος. Έτσι, παρουσιάζονται όλες οι ιστοσελίδες που απαρτίζουν το Web κομμάτι, αναφέρονται οι λειτουργίες που εκτελεί το σύστημα και περιγράφονται οι διάφορες μέθοδοι, τεχνικές και διαδικασίες που ακολουθούνται για να επιτευχθεί ο σκοπός του. Επίσης, γίνεται ξεχωριστή αναφορά στο κομμάτι των Google Maps και το πώς έγινε η ενσωμάτωσή τους στο σύστημα.

Στο έκτο κεφάλαιο γίνεται η αξιολόγηση του συστήματος. Μέσω κρίσιμων test cases και διαφοροποιημένων σεναρίων λειτουργίας, πραγματοποιούνται και αναλύονται διάφορες μετρήσεις τόσο των επιδόσεων του συστήματος, όσο και των επιδόσεων και της αξιοπιστίας των αποτελεσμάτων των αλγόριθμων που σχεδιάστηκαν και υλοποιήθηκαν στα πλαίσια της διατριβής.

Στο έβδομο κεφάλαιο παρουσιάζονται οι προοπτικές για περαιτέρω έρευνα σχετικά με τη θεματολογία που πραγματεύεται η παρούσα διατριβή, καθώς σεναρία μελλοντικών εφαρμογών του συστήματος σε διαφορετικό πλαίσιο.

Το όγδοο και τελευταίο κεφάλαιο περιλαμβάνει τη βιβλιογραφία και τις διαδικτυακές πηγές στις οποίες γίνεται αναφορά στα πλαίσια της διδακτορικής διατριβής.

2 ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

2.1 Φιλτράρισμα Πληροφορίας και Συστήματα Προτάσεων

Τα Συστήματα Προτάσεων (Recommender Systems), που εμφανίστηκαν για πρώτη φορά στα μέσα της δεκαετίας του 1990, αποτελούν μια συγκεκριμένη τεχνική Φιλτραρίσματος Πληροφορίας (Information Filtering). Στη γενική περίπτωση, σκοπός τους είναι να φέρουν ένα χρήστη σε επαφή με αντικείμενα (μουσική, ταινίες, ειδήσεις, βιβλία, ιστοσελίδες, κτλ) και υπηρεσίες που κρίνεται ότι τον ενδιαφέρουν, συνήθως βάσει ενός προσωποποιημένου προφίλ. Το προφίλ αυτό δημιουργείται είτε άμεσα, από το χρήστη του συστήματος, είτε έμμεσα, από το ίδιο το σύστημα, παρατηρώντας τη συμπεριφορά του εκάστοτε χρήστη. Σύμφωνα με τους Sobelcki (2006), Shardanand και Maes (1995) και Burke (2000), ως *Recommender System* ορίζεται οποιοδήποτε σύστημα παράγει σαν έξοδο εξατομικευμένες προτάσεις ή έχει σαν αποτέλεσμα το να οδηγήσει το χρήστη με προσωποποιημένο τρόπο σε ενδιαφέροντα ή χρήσιμα αντικείμενα μέσα από ένα μεγάλο εύρος πιθανών επιλογών. Στις περισσότερες περιπτώσεις, ο βαθμός στον οποίο ένα αντικείμενο κρίνεται ότι ταιριάζει στις ανάγκες και τις επιθυμίες κάποιου χρήστη αντικατοπτρίζεται μέσω μιας βαθμολογίας (*score*), εξατομικευμένης για κάθε ζεύγος χρήστη-αντικειμένου. Η κλίμακα και ο τρόπος υπολογισμού της βαθμολογίας αυτής είναι προφανές ότι είναι διαφορετικά σε κάθε σύστημα.

Παράλληλα με τη διευκολύνσεις που προσφέρουν στους απλούς χρήστες των διαφόρων sites, τα Recommender Systems (Adomavicius and Tuzhilin, 2005) αποτελούν βασικό εργαλείο στα χέρια των επιχειρήσεων που δραστηριοποιούνται στον τομέα του ηλεκτρονικού εμπορίου (E-Commerce). Τα τελευταία χρόνια, όλο και περισσότερες επιχειρήσεις στηρίζονται επιτυχημένα στο *Internet*, τόσο για την προώθηση των προσφερόμενων προϊόντων και υπηρεσιών, όσο και για την απευθείας διάθεση προϊόντων στους καταναλωτές. Οι περισσότερες εταιρίες που κατάφεραν να καθιερωθούν στο χώρο του ηλεκτρονικού εμπορίου, με χαρακτηριστικά παραδείγματα το amazon.com και το ebay, στηρίζουν την επιτυχία τους στην έννοια της εξατομίκευσης, στην οποία καθοριστικό ρόλο παίζουν τα RS. Πιο συγκεκριμένα, τα ηλεκτρονικά καταστήματα έχουν τη δυνατότητα να βελτιώσουν την απόδοσή τους, εντοπίζοντας τις ξεχωριστές ανάγκες και επιθυμίες των πελατών τους και τους προσφέρουν προϊόντα που έχουν στόχο στο να τις ικανοποιήσουν. Ο

καταναλωτής με τον τρόπο αυτό αντιμετωπίζεται ως μονάδα, γεγονός που ενισχύει τη σχέση επιχείρησης-πελάτη, χτίζοντας παράλληλα σχέσεις εμπιστοσύνης.

Πιο αναλυτικά, σύμφωνα με τους Schafer, Konstan και Riedl (2001) τα RS μπορούν να βελτιώσουν τις πωλήσεις των ιστότοπων ηλεκτρονικού εμπορίου (E-Commerce sites) με πολλούς τρόπους:

- **Μετατρέποντας τους επισκέπτες σε πελάτες.** Είναι σύνηθες φαινόμενο οι χρήστες να περιηγούνται σε κάποιο *site* ηλεκτρονικού εμπορίου χωρίς να κάνουν κάποια αγορά. Τα RS μπορούν να τους βοηθήσουν να βρουν προϊόντα που επιθυμούν να αγοράσουν.
- **Αυξάνοντας τις διασταυρούμενες (*cross*) πωλήσεις,** τις πωλήσεις δηλαδή επιπρόσθετων προϊόντων σε χρήστες που έχουν αποφασίσει να πραγματοποιήσουν μια αγορά. Εάν οι συστάσεις του RS έχουν αντίκτυπο στο χρήστη, το μέσο μέγεθος των παραγγελιών αυξάνεται. Παράδειγμα του παραπάνω αποτελεί η προσωποποιημένη πρόταση αγοράς επιπλέον προϊόντων κατά τη διαδικασία της εξόφλησης, σύμφωνα με όσα ο χρήστης έχει τοποθετήσει ήδη στο καλάθι αγορών του.
- **Χτίζοντας ηλεκτρονική εμπιστοσύνη.** Σε ένα κόσμο όπου οι ανταγωνιστές απέχουν μεταξύ τους όχι περισσότερο από ένα ή δύο κλικ του ποντικιού, το να κερδίσει κανείς την εμπιστοσύνη των πελατών είναι μία βασική επιχειρηματική στρατηγική. Τα *sites* των ηλεκτρονικών επιχειρήσεων επενδύουν στο να μαθαίνουν τους πελάτες τους και χρησιμοποιούν RS για να δημιουργήσουν τα προφίλ των χρηστών και να παρουσιάσουν εξατομικευμένα περιβάλλοντα που ταιριάζουν με τις ανάγκες των καταναλωτών. Οι τελευταίοι αναγνωρίζουν το παραπάνω και η ικανοποίηση τους από τις προσφερόμενες υπηρεσίες αυξάνεται, κάτι που μακροπρόθεσμα θα τους οδηγήσει στο να επιστρέψουν στα συγκεκριμένα *sites*.

2.2 Κατηγοριοποιήσεις Συστημάτων Προτάσεων

2.2.1 Βάσει του Τρόπου Παραγωγής των Συστάσεων

Όσον αφορά τον τρόπο παραγωγής των προτάσεων και συστάσεων ενός Recommender System, έχουν προταθεί και εφαρμοστεί αρκετές διαφορετικές προσεγγίσεις, κάθε μία από τις οποίες παρουσιάζει επιμέρους πλεονεκτήματα και μειονεκτήματα. Οι κυριότερες παρουσιάζονται παρακάτω, μαζί με παραδείγματα της εφαρμογής τους, ενώ γίνεται και μια αναφορά στα κυριότερα πλεονεκτήματα και μειονεκτηματά τους. Αξίζει να σημειωθεί ότι καμία από τις τεχνικές αυτές δεν μπορεί να θεωρηθεί ιδανική, κάτι που έχει ως

αποτέλεσμα στο να οδηγούμαστε στην εφαρμογή υβριδικών τεχνικών, που συνδυάζουν περισσότερες από μία προσεγγίσεις.

Collaborative Filtering (Συνεργατικό Φιλτράρισμα)

Σύμφωνα με τον Burke (2002) τα RS που ακολουθούν τη συγκεκριμένη προσέγγιση χρησιμοποιούν τις βαθμολογίες μιας κοινότητας χρηστών για να ανακαλύψουν συσχετίσεις ανάμεσα σε έναν συγκεκριμένο πρόσωπο και σε άλλους χρήστες του συστήματος. Στη συνέχεια, προτείνουν στο πρόσωπο αυτό αντικείμενα που βαθμολογήθηκαν με υψηλό βαθμό από τους υπόλοιπους χρήστες με παρόμοιες προτιμήσεις. Ουσιαστικά, η συγκεκριμένη προσέγγιση μοντελοποιεί την ανθρώπινη συμπεριφορά σύμφωνα με την οποία δεχόμαστε προτάσεις από φίλους.

Το Collaborative Filtering είναι ίσως η πιο διαδεδομένη τεχνική προτάσεων και χρησιμοποιείται κατά κόρον από το *amazon.com*, με χαρακτηριστικό παράδειγμα την παρουσίαση, στο χρήστη-μέλος του *site*, προϊόντων που αγοράστηκαν από άλλους χρήστες, οι οποίοι στο παρελθόν είχαν αγοράσει το ίδιο προϊόν με αυτόν.

Βασικό πλεονέκτημα της μεθόδου αυτής είναι το γεγονός ότι δεν απαιτείται εξαγωγή γνωρισμάτων των αντικειμένων ή κάποια πολύπλοκη αναπαράστασή τους, γεγονός που την κάνει απλούστερη στην υλοποίηση από άλλες μεθόδους. Επιπλέον, με την προσέγγιση αυτή, οι συστάσεις δεν περιορίζονται σε συγκεκριμένες κατηγορίες αντικειμένων. Από τη στιγμή που θα γίνει η συσχέτιση δύο ή παραπάνω χρηστών, μπορούν να εξαχθούν συμπεράσματα για κοινές προτιμήσεις τους σε διάφορους τομείς (*cross-domain recommendations*). Ακόμα, η συγκεκριμένη τακτική είναι χρήσιμη για τον εντοπισμό *niches*, στοχευόμενων δηλαδή κομματιών της αγοράς με συγκεκριμένες σταθερές και ξεκάθαρες προτιμήσεις

Ένα βασικό μειονέκτημα του Collaborative Filtering, είναι ότι για να προταθούν ικανοποιητικά ταιριάσματα χρηστών ως προς το γούστο και τις προτιμήσεις, πρέπει να έχει λάβει χώρα μεγάλος αριθμός βαθμολογήσεων στο σύστημα (*sparsity problem*). Παράλληλα, είναι εύκολα κατανοητό ότι προβλήματα μπορούν να εντοπιστούν κατά την είσοδο στο σύστημα ενός νέου αντικειμένου (*new-item problem*) ή ενός νέου χρήστη (*new-user problem*), αφού για κάποιο χρονικό διάστημα δεν υπάρχουν αρκετές βαθμολογήσεις που θα οδηγήσουν στις απαραίτητες, για την εφαρμογή της μεθόδου, συσχετίσεις και προτάσεις. Τέλος, ένα σύστημα που βασίζεται στο συνεργατικό φιλτράρισμα είναι ευάλωτο στις επιθέσεις κακόβουλων χρηστών, οι οποίοι μπορούν με τη δημιουργία πλαστών προφίλ να αναδείξουν ή να υποβαθμίσουν κάποιο αντικείμενο.

Content-Based Filtering (Φιλτράρισμα Βασισμένο στο Περιεχόμενο)

Σε αυτή την προσέγγιση, τα αντικείμενα κατηγοριοποιούνται βάσει συγκεκριμένων ιδιοτήτων τους και στη συνέχεια το σύστημα χρησιμοποιεί κάποιους αλγόριθμους εκμάθησης ώστε να συσχετίσει σταδιακά τις προτιμήσεις του χρήστη με άλλα αντικείμενα που πιθανώς τον ενδιαφέρουν. Αυτό συμβαίνει με την έμμεση δημιουργία προσωποποιημένων προφίλ για όλους τους χρήστες του συστήματος, τα οποία συνήθως χτίζονται σταδιακά μέσω της αλληλεπίδρασης με αυτό, σε συνδυασμό με μια σειρά προτιμήσεων που δηλώνεται από το χρήστη κατά την εγγραφή του στο σύστημα. Και σε αυτή τη μέθοδο βασικό ρόλο παίζει το feedback των χρηστών, μέσω της βαθμολόγησης αντικειμένων.

Εφαρμογές του Content-Based filtering μπορούν να εντοπιστούν σε διάφορες υπηρεσίες προτάσεων εστιατορίων, ειδήσεων κ.α. Επίσης, δείγματα της συγκεκριμένης μεθόδου μπορούν να βρεθούν σε διαδικτυακά κατάστημα όπως για παράδειγμα το amazon.com και συγκεκριμένα στην επιλογή προβολής των αγαπημένων κατηγοριών προϊόντων κάποιου χρήστη.

Η Content-Based μέθοδος φιλτραρίσματος, μπορεί να είναι αρκετά αποτελεσματική σε περιπτώσεις καλά δομημένων αναπαραστάσεων των αντικειμένων, σε συστήματα με επαρκή αριθμό βαθμολογήσεων. Και εδώ παρουσιάζεται το μειονέκτημα της ανακρίβειας στις προτάσεις σε νέους χρήστες, για τους οποίους δεν έχουν συγκεντρωθεί δεδομένα και προτιμήσεις. Επιπλέον, η αναπαράσταση ενός αντικειμένου δεν είναι πάντα εύκολη υπόθεση, όπως για παράδειγμα στην περίπτωση που ως αντικείμενα θεωρούνται πολυμεσικά δεδομένα.

Demographic Filtering (Δημογραφικό Φιλτράρισμα)

Η τεχνική αυτή βασίζεται σε δημογραφικά δεδομένα όπως το φύλο, το επάγγελμα και η ηλικία για την εξαγωγή συμπερασμάτων και τον αποκλεισμό συγκεκριμένων δημογραφικών ομάδων από προτάσεις. Για παράδειγμα, είναι λογικό, προϊόντα καλλυντικών να είναι επιθυμητό να προταθούν αποκλειστικά σε γυναίκες.

Knowledge-Based Filtering (Φιλτράρισμα βασισμένο στη Γνώση)

Τα συστήματα που υιοθετούν αυτή την προσέγγιση χρησιμοποιούν τη γνώση χρηστών και αντικειμένων για τη δημιουργία προτάσεων που βασίζονται σε διαφορετικές και μάλλον ετερόκλητες μεθόδους, κανόνες και στατιστικές προσεγγίσεις. Η βασική διαφοροποίηση των Knowledge-Based συστημάτων από τις προσεγγίσεις που έχουν περιγραφεί μέχρι τώρα, είναι

το ότι δεν επιχειρούν σε καμία περίπτωση να εξάγουν μακροπρόθεσμες γενικεύσεις για τους χρήστες τους.

Ένα παράδειγμα πρότασης βασισμένο σε γνώση και πιο συγκεκριμένα βασισμένο σε κανόνα (rule-based), είναι ότι ένας χρήστης που αγόρασε ένα βιβλίο είναι πιθανό να ενδιαφέρεται να αγοράσει τη συνέχεια του βιβλίου.

Πλεονέκτημα της μεθόδου αυτής αποτελεί προφανώς το γεγονός ότι δε βασίζεται σε κάποιου είδους feedback και βαθμολογήσεων για να εξάγει συμπεράσματα, άρα μπορεί από την αρχή να παράγει ικανοποιητικές προτάσεις. Παράλληλα, από τη στιγμή που δεν εξάγονται μακροπρόθεσμα συμπεράσματα για τους χρήστες ενός τέτοιου συστήματος, οι προτάσεις του είναι ευαίσθητες στις αλλαγές των προτιμήσεων του χρήστη, χωρίς να χρειαστεί κάποιο επιπλέον διάστημα προσαρμογής. Από την άλλη, οι προτάσεις μιας τέτοιας μεθόδου είναι στατικές και δε βελτιώνονται με την πάροδο του χρόνου, ενώ δεν είναι πάντα εύκολη η σωστή διαχείριση της γνώσης για την εξαγωγή συμπερασμάτων. Τέλος, ένα άλλο πρακτικό πρόβλημα είναι η δυσκολία απόκτησης και διατήρησης ενημερωμένων δεδομένων για να τροφοδοτηθούν οι αλγόριθμοι που χρησιμοποιούνται.

Υβριδικές (Hybrid) Μέθοδοι

Όπως έγινε προφανές, οι διάφορες προσεγγίσεις που αναλύθηκαν παραπάνω παρουσιάζουν επιμέρους πλεονεκτήματα και αδυναμίες, ανάλογα με τα διαφορετικά σενάρια χρήσης. Για το λόγο αυτό συνήθως τακτική, η οποία τείνει να επικρατήσει στο συγκεκριμένο χώρο, είναι ο συνδυασμός δύο ή παραπάνω μεθόδων, με σκοπό τη βελτίωση της ποιότητας των προτάσεων. Γίνεται εύκολα αντιληπτό, ότι και σε αυτή την περίπτωση δεν μπορεί υπάρξει συνδυαστικός αλγόριθμος που θα αποδίδει άριστα υπό όλες τις συνθήκες και για όλες τις περιπτώσεις διαφορετικών συστημάτων και αντικειμένων. Η εύρεση νέων, καθώς και η βελτιστοποίηση των υπάρχοντων προτάσεων υβριδικών αλγορίθμων, απασχολεί σε τεράστιο βαθμό την έρευνα στον τομέα των *Recommender Systems* τα τελευταία χρόνια. Κατά καιρούς έχουν προταθεί και εφαρμοστεί αρκετές διαφορετικές προσεγγίσεις στο θέμα των υβριδικών τεχνικών, οι κυριότερες από τις οποίες παρουσιάζονται συνοπτικά στη συνέχεια.

- Στους **Σταθμισμένους (Weighted) Hybrid Recommenders** η βαθμολογία ενός υπό πρόταση προϊόντος, προκύπτει με συνδυασμό όλων των ενσωματωμένων στο σύστημα τεχνικών πρότασης. Όπως είναι προφανές από το όνομα της προσέγγισης, οι επιμέρους

τεχνικές συμμετέχουν στον τελικό υπολογισμό της βαθμολογίας με διαφορετικούς συντελεστές – βάρη.

- Τα **Εναλλάσσοντα (Switching)** υβριδικά μοντέλα επιλέγουν μια συγκεκριμένη προσέγγιση από τις προαναφερθείσες για να χρησιμοποιηθεί στη βαθμολόγηση των συσχετίσεων, ανάλογα με κάποια κριτήρια. Επίσης, το σύστημα μπορεί να δοκιμάσει κάποια προσέγγιση και αν ένας δείκτης εγκυρότητας (confidence), που έχει προηγουμένως οριστεί στα πλαίσια του συστήματος είναι ανεπαρκής, να επιλεγεί μια διαφορετική προσέγγιση.
- Μια **Μεικτή (Mixed)** προσέγγιση χρησιμοποιεί ταυτόχρονα παραπάνω από μία διαφορετικές τεχνικές και παρουσιάζει τα αποτελέσματα όλων τους στο χρήστη.
- Στα **Υβριδικά Συστήματα Συνδυασμού Γνωρισμάτων (Feature Combination)**, εφαρμόζεται σε πρώτη φάση συνεργατικό φιλτράρισμα, τα αποτελέσματα του οποίου αντιμετωπίζονται ως γνωρίσματα των αντικειμένων και τροφοδοτούν ως δεδομένα μια Content-Based μέθοδο.
- Η **Διαδοχική (Cascade)** υβριδική προσέγγιση χρησιμοποιεί κάποια από τις γνωστές μεθόδους, για να δημιουργήσει μια πρόχειρη προσέγγιση για τη βαθμολόγηση ή την κατάταξη των υπό εξέταση αντικειμένων, συνήθως αποκλείοντας δεδομένα που έχουν μικρό βαθμό (score) και κατ' επέκταση μικρή πιθανότητα να προταθούν. Στη συνέχεια, η κατάταξη βελτιώνεται και παίρνει την τελική της μορφή, με την εφαρμογή κάποιας άλλης γνωστής μεθόδου πάνω στο μικρότερο dataset που έχει προκύψει.
- Η μέθοδος **Επαυξημένων Γνωρισμάτων (Feature Augmentation)** μοιάζει αρκετά με τη διαδοχική μέθοδο. Στην περίπτωση αυτή όμως, η πρώτη βαθμολόγηση/κατάταξη χρησιμοποιείται από τη δεύτερη εφαρμοζόμενη μέθοδο ως πρόσθετο χαρακτηριστικό.
- Τέλος, η προσέγγιση **Μετα-Επιπέδου (Meta-Level)** συνδυάζει δύο τεχνικές Recommendation, χρησιμοποιώντας το μοντέλο που προέκυψε ως αποτέλεσμα της πρώτης, ως είσοδο για τη δεύτερη.

2.2.2 Βάσει του Τρόπου Απόκτησης Γνώσης

Μια άλλη κατηγοριοποίηση που μπορεί να γίνει στα Recommender Systems, αφορά στο πώς αυτά αποκτούν την απαραίτητη γνώση για το χρήστη ώστε να διαμορφώσουν το προφίλ του, πάνω στο οποίο θα βασιστούν οι προτάσεις τους. Διακρίνονται τρεις περιπτώσεις:

- **Άμεση Ανάδραση (*Explicit Feedback*):** Ο χρήστης ερωτάται ευθέως για τις προτιμήσεις του, καλούμενος να συμπληρώσει κάποια πεδία, να απαντήσει σε ερωτήσεις ή να βαθμολογήσει κάποια προϊόντα και προτάσεις. Σημειώνεται ότι ο συγκεκριμένος τρόπος συλλογής γνώσης για τους χρήστες του συστήματος παρουσιάζει αρκετές δυσκολίες ως προς το εύρος και την ακρίβεια των δεδομένων, αφού προϋποθέτει ότι ο χρήστης θα αφιερώσει προσωπικό χρόνο για να απαντήσει στις ερωτήσεις που τίθενται, κάτι που δεν είναι πάντα εφικτό.
- **Έμμεση Ανάδραση (*Implicit Feedback*):** Οι προτιμήσεις του χρήστη εντοπίζονται από το σύστημα, δημιουργώντας ένα εικονικό προφίλ για το πως το σύστημα βλέπει τον κάθε χρήστη. Τεχνικές που χρησιμοποιούνται για αυτόν το σκοπό, είναι η καταγραφή των αγορών που έχει πραγματοποιήσει ο χρήστης στο παρελθόν, η παρακολούθηση του ιστορικού αναζητήσεών του, ο χρόνος που αφιερώνει σε κάθε αντικείμενο, κ.α.
- **Συνδυασμός των παραπάνω (*Explicit & Implicit Feedback*)**

2.2.3 Βάσει της Πρωτοβουλίας Δράσης (*Initiative of Operation*)

Διακρίνονται δύο περιπτώσεις:

Ενεργά (Active) Συστήματα

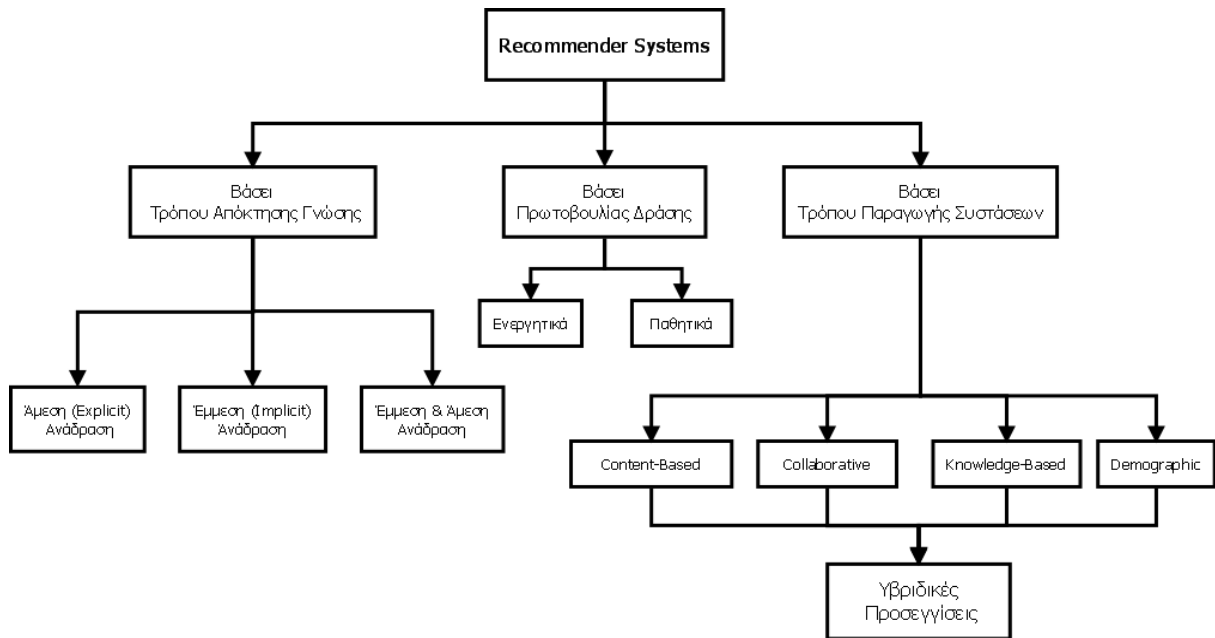
Στην περίπτωση αυτή, ο χρήστης τροφοδοτείται ενεργά με τα αντικείμενα που το σύστημα κρίνει ότι τον ενδιαφέρουν μέσω ειδοποιήσεων. Στη συγκεκριμένη προσέγγιση, που καλείται και προωθητική (push), η ακρίβεια των προτάσεων διαδραματίζει τεράστιο και καθοριστικό ρόλο στο πόσο επιτυχημένο κρίνεται ένα σύστημα. Σε μια εποχή που ο μέσος χρήστης κατακλύζεται από έναν τεράστιο αριθμό πληροφοριών, ένα από τα τελευταία πράγματα που θα αναζητήσει είναι το να γίνεται συνεχώς δέκτης επιπλέον πραγμάτων, τα οποία δεν παρουσιάζουν ενδιαφέρον για αυτόν. Κατά συνέπεια, το Recommender System θα έχει αποτύχει και κατά πάσα πιθανότητα ο χρήστης θα εγκαταλείψει τη χρήση του, καθώς θα το αντιμετωπίζει ως μια επιπρόσθετη ενόχληση.

Παθητικά (Passive) Συστήματα

Τα συγκεκριμένα συστήματα αφορούν κυρίως πεδία τα οποία περιλαμβάνουν ροές πληροφορίας, όπως είναι τα εισερχόμενα e-mails και τα Usenet Groups και στα οποία δεν απαιτείται επιπλέον προσπάθεια από το χρήστη για τη συλλογή των δεδομένων. Ρόλος τους είναι το να καθορίσουν τη σημασία συγκεκριμένων αντικειμένων στο χρήστη-αποδέκτη και είτε να τα κατατάξουν με σειρά προτεραιότητας, είτε να φιλτράρουν πληροφορίες που

κρίνουν άχρηστες. Έρευνες έχουν δείξει ότι η πλειονότητα των χρηστών θεωρεί προτιμότερη την πρώτη προσέγγιση, αφού στην περίπτωση του φιλτραρίσματος κινδυνεύουν να μη γίνουν ποτέ αποδέκτες κάποιας χρήσιμης για αυτούς πληροφορίας.

Όλες οι παραπάνω κατηγοριοποιήσεις συνοψίζονται περιληπτικά στο παρακάτω διάγραμμα:



Σχήμα 2-1: Κατηγορίες των Recommender Systems

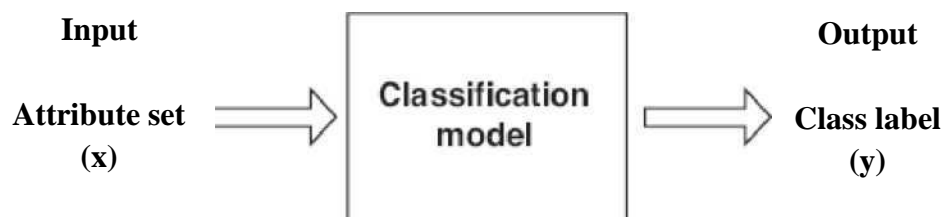
2.3 Γεωγραφική Πληροφορία

Οι περισσότερες πληροφορίες που απευθύνονται στους χρήστες συνδέονται συνήθως με κάποια γεωγραφικά όρια (π.χ. ένα event για ένα κονσέρτο στην Αθήνα). Υπάρχουν πολλοί τρόποι για τον εντοπισμό πληροφοριών με γεωγραφική σημασία για διάφορες εργασίες ανάκτησης, όπως η παροχή καλύτερης ταύτισης διαφημίσεων ώστε να παρέχονται ακριβείς πληροφορίες για τοπικά προϊόντα και υπηρεσίες κτλ. Πολλοί ερευνητές έχουν δείξει πως μπορεί να βελτιωθεί η απόδοση της ανάκτησης για τέτοιου είδους εργασίες, κυρίως με την ενσωμάτωση γεωγραφικής πληροφορίας όταν αυτή είναι γνωστή εκ των προτέρων, καθώς επίσης και με την ανάπτυξη τεχνικών για τον εντοπισμό γεωγραφικής πρόθεσης ακόμα και αν η πληροφορία απουσιάζει (Andrade and Silva, 2006; Yu and Cai, 2007; Yi, Reghavan and Leggetter, 2009; Kalantidis et al., 2011). Στην έρευνα για την ανάκτηση γεωγραφικής πληροφορίας, η σχέση μεταξύ των users και των events στον γεωγραφικό χώρο κρίνεται από

τις χωρικές σχέσεις (όπως περιέχεται, επικαλύπτεται, διασταυρώνεται, συνδέεται, κοντά κλπ.).

2.4 Ταξινόμηση και Ταξινομητές

Ταξινόμηση (Classification) (Tan P.N., Steinbach M. and Kumar V., 2006) είναι η διεργασία αντιστοίχισης αντικειμένων σε μία ή περισσότερες προκαθορισμένες κατηγορίες και είναι μια διεισδυτική διαδικασία που περιγράφει (αναλύει) και επιλύει πολλά διαφορετικά πολύπλοκα προβλήματα.



Σχήμα 2-2: Η ταξινόμηση ως διεργασία χαρτογράφησης ενός συνόλου x χαρακτηριστικών γνωρισμάτων – ιδιοτήτων ως προς την ετικέτα κατηγορίας (κλάσης) y

Τα δεδομένα εισόδου για μια διεργασία ταξινόμησης είναι ένα σύνολο εγγραφών. Κάθε εγγραφή ονομάζεται παράδειγμα ή δείγμα, χαρακτηρίζεται σαν ένα ζεύγος (x,y) , όπου x είναι το σύνολο των χαρακτηριστικών γνωρισμάτων και y είναι ένα καθορισμένο ιδιαίτερο χαρακτηριστικό γνώρισμα γνωστό ως ετικέτα κατηγορίας (class label - γνωστό και σαν κατηγορία χαρακτηριστικού γνωρίσματος) (Tom Mitchell and McGraw Hill, 1997).

Ταξινόμηση είναι η διεργασία εκμάθησης μιας συγκεκριμένης συνάρτησης f η οποία χαρτογραφεί κάθε σύνολο x χαρακτηριστικών γνωρισμάτων σε μία προκαθορισμένη ετικέτα κατηγορίας.

Η συγκεκριμένη συνάρτηση είναι επίσης γνωστή ως μοντέλο - πρότυπο ταξινόμησης (classification model). Ένα μοντέλο ταξινόμησης χρησιμεύει στις παρακάτω περιπτώσεις:

Περιγραφική Μοντελοποίηση

Ένα μοντέλο ταξινόμησης μπορεί να χρησιμοποιηθεί σαν επεξηγηματικό εργαλείο για να μπορέσουμε να διαχωρίσουμε τα αντικείμενα μεταξύ τους σε διαφορετικές κλάσεις.

Προβλεπτικό Μοντέλο

Ένα μοντέλο ταξινόμησης μπορεί επίσης να χρησιμοποιηθεί για να προβλέπει την ετικέτα κατηγορίας (κλάσης) αγνώστων εγγραφών. Μπορεί επίσης να θεωρηθεί σαν ένα μαύρο κουτί το οποίο αναθέτει αυτόματα μία ετικέτα κατηγορίας όταν εμφανιστεί με το σύνολο των χαρακτηριστικών γνωρισμάτων μιας άγνωστης εγγραφής.

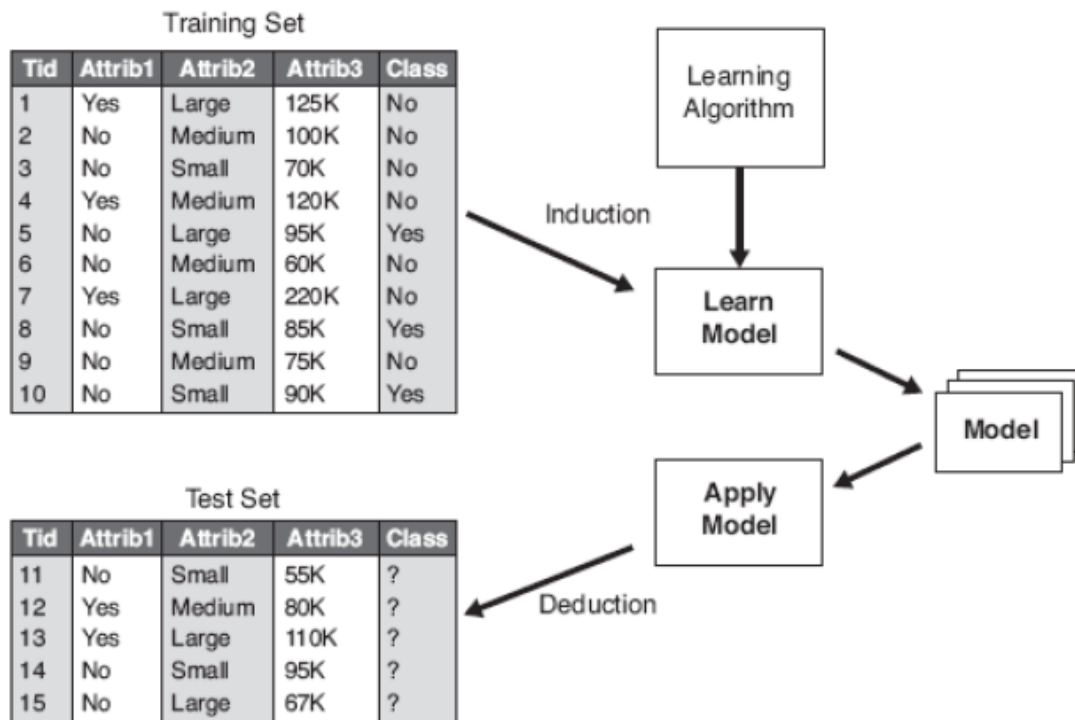
Οι τεχνικές ταξινόμησης είναι περισσότερο κατάλληλες στο να περιγράφουν ή να προβλέπουν σε σύνολα δεδομένων (data sets) με δυαδικές ή ονομαστικές κατηγορίες. Είναι λιγότερο αποδοτικές σε περιπτώσεις που υπάρχουν κανονικές - συνηθισμένες κατηγορίες γιατί δεν λαμβάνουν υπόψη τους την αναμενόμενη σειρά μεταξύ των κατηγοριών. Για παράδειγμα είναι δύσκολο να ταξινομήσουν ένα άτομο στο αν ανήκει στην κατηγορία των του υψηλού, μεσαίου ή χαμηλού εισοδήματος. Άλλες μορφές συσχετίσεων, όπως συσχετίσεις υποκλάσεων - υπερκλάσεων μεταξύ των κατηγοριών (π.χ. παράδειγμα οι άνθρωποι και οι πίθηκοι είναι ανθρωποειδή το οποίο συνεπάγεται ότι και οι δύο είναι θηλαστικά), αγνοούνται.

2.4.1 Γενική Διαδικασία Επίλυσης Προβλήματος Ταξινόμησης

Η τεχνική ταξινόμησης (ή ταξινομητής - Classifier) είναι μια συστηματική διαδικασία που γίνεται για να δημιουργηθούν μοντέλα ταξινόμησης (classification models) παίρνοντας ως είσοδο ένα σύνολο δεδομένων (data set). Παραδείγματα ταξινομητών είναι τα δένδρα απόφασης (decision trees), οι βασισμένοι σε κανόνες ταξινομητές (rule-based classifiers), τα νευρωνικά δίκτυα (neural networks), οι μηχανές διανυσμάτων υποστήριξης (Support Vector Machines - SVM) και οι ταξινομητές naïve Bayes. Κάθε τεχνική αναπτύσσει ένα αλγόριθμο εκμάθησης για να προσδιορίσει το μοντέλο που ταιριάζει καλύτερα στη σχέση μεταξύ του συνόλου των χαρακτηριστικών γνωρισμάτων και της ετικέτας κατηγορίας των εισαγόμενων δεδομένων. Το μοντέλο που δημιουργείται από έναν αλγόριθμο εκμάθησης θα πρέπει να προσαρμόζεται βάσει των εισαγόμενων δεδομένων και να προβλέπει σωστά την ετικέτα κατηγορίας εγγραφών που δεν έχει ξαναδεί. Επομένως μία βασική λειτουργία του αλγορίθμου εκμάθησης είναι να κατασκευάζει μοντέλα με υψηλή ικανότητα γενίκευσης. Για παράδειγμα μοντέλα τα οποία προβλέπουν σωστά και με απόλυτη ακρίβεια τις ετικέτες κατηγοριών αγνώστων εγγραφών.

Το *Σχήμα 2-3*, δείχνει μια γενική προσέγγιση επίλυσης ενός προβλήματος ταξινόμησης. Πρώτον, πρέπει να μας δίνεται ένα training set το οποίο περιέχει εγγραφές των οποίων οι ετικέτες κατηγορίας είναι γνωστές. Το training set χρησιμοποιείται για να φτιάξει το μοντέλο ταξινόμησης, το οποίο μετέπειτα εφαρμόζεται στο test set, το οποίο περιέχει

εγγραφές των οποίων οι ετικέτες κατηγορίας είναι άγνωστες.



Σχήμα 2-3: Επίλυση Προβλήματος Ταξινόμησης

Η Αξιολόγηση της απόδοσης ενός ταξινομητή (ή ενός μοντέλου ταξινόμησης) βασίζεται στον αριθμό των εγγραφών του test set που προβλέφθηκαν σωστά ή λάθος από τον ταξινομητή.

Για να μπορούμε να συγκρίνουμε τις αποδόσεις διαφορετικών μοντέλων χρησιμοποιούμε διάφορους δείκτες επίδοσης, όπως είναι η ακρίβεια (accuracy):

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

και αντίστροφα η αποτίμηση σφάλματος (error rate):

$$\text{Error rate} = \frac{\text{Number of wrong predictions}}{\text{Total number of predictions}}$$

Έτσι ο ταξινομητής με το μεγαλύτερο accuracy και αντίστοιχα με το μικρότερο error rate είναι ο αποτελεσματικότερος (κάνει τις πιο σωστές προβλέψεις).

2.4.2 Δοκιμή και Επαλήθευση του Μοντέλου

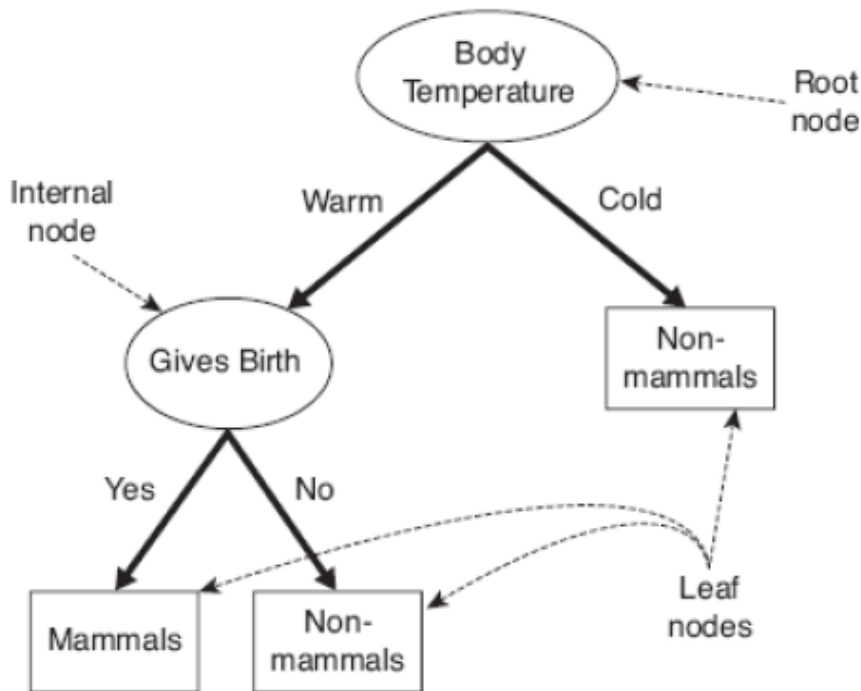
Για τον υπολογισμό των δεικτών επίδοσης του μοντέλου ώστε να οδηγηθούμε σε αντικειμενική αξιολόγησή του, δύο είναι οι ευρύτερα εφαρμοζόμενες πρακτικές:

- Τυχαία διαδικασία χωρισμού των δεδομένων σε διακριτά σύνολα εκπαίδευσης και δοκιμής (train and test sets).
- N-fold cross validation (10-fold, 5-fold, 20 fold κ.λ.π.). Το σύνολο των δεδομένων χωρίζεται με τυχαίο τρόπο σε N ισάριθμα υποσύνολα από τα οποία τα N-1 χρησιμοποιούνται για εκπαίδευση, ενώ αυτό που απομένει χρησιμοποιείται ως σύνολο δοκιμής. Η διαδικασία επαναλαμβάνεται N φορές και ως δείκτες επίδοσης λαμβάνονται οι αντίστοιχοι μέσοι όροι.

2.4.3 Δένδρα Απόφασης

Η λειτουργία ενός δένδρου απόφασης είναι σχετικά απλή. Μπορούμε να λύσουμε ένα πρόβλημα ταξινόμησης κάνοντας μια σειρά από σωστά δομημένες ερωτήσεις σχετικά με το κάθε ένα από τα χαρακτηριστικά γνωρίσματα κάθε εγγραφής του test set. Κάθε φορά που λαμβάνουμε μια απάντηση μία (καθορισμένη) επόμενη γίνεται και μόλις απαντηθεί και αυτή άλλη μία καθορισμένη επακόλουθη αυτής γίνεται και αυτό συνεχίζεται μέχρι να φτάσουμε σε ένα συμπέρασμα, δηλαδή τον καθορισμό της ετικέτας κατηγορίας για την συγκεκριμένη εγγραφή. Η σειρά των ερωτήσεων και οι πιθανές (δυνατές) απαντήσεις τους μπορούν να οργανωθούν υπό την μορφή ενός δένδρου απόφασης, το οποίο είναι μια ιεραρχική δομή που αποτελείται από κόμβους και κατευθυνόμενες ακμές.

Για παράδειγμα το *Σχήμα 2-4* δείχνει ένα δένδρο απόφασης για το υποθετικό πρόβλημα ταξινόμησης θηλαστικού - μη-θηλαστικού.



Σχήμα 2-4: Παράδειγμα Δένδρου Απόφασης

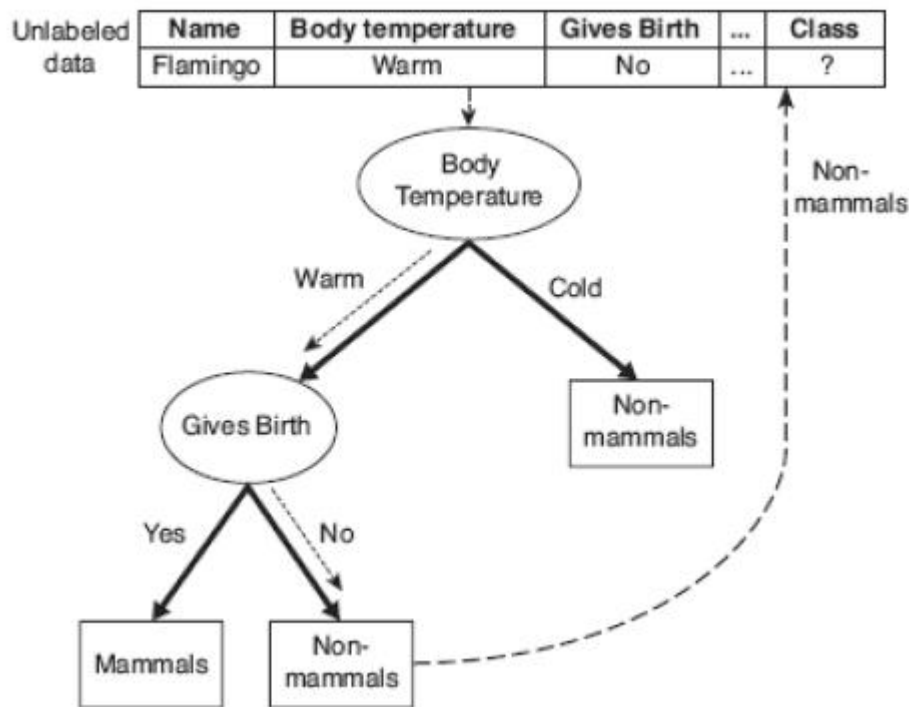
Όπως παρατηρούμε στο Σχήμα 2-4, το Δένδρο έχει τρεις τύπους κόμβων:

- **Μία ρίζα (Root Node):** Κόμβος ο οποίος δεν έχει εισερχόμενες ακμές και μηδέν ή περισσότερες εξερχόμενες ακμές.
- **Εσωτερικούς κόμβους (Internal Nodes):** Καθένας από αυτούς έχει μία εισερχόμενη ακμή και δύο ή περισσότερες εξερχόμενες.
- **Φύλλα ή τερματικούς κόμβους (Leaf or Terminal Nodes):** Κόμβος ο οποίος έχει μία εισερχόμενη ακμή και καμία εξερχόμενη. Συνήθως καταλήγει σε συμπέρασμα, δηλαδή στην ανάθεση ετικέτας κατηγορίας της συγκεκριμένης εγγραφής.

Σε ένα δένδρο απόφασης κάθε κόμβος φύλλο είναι προσδιορισμένο με μία ετικέτα κατηγορίας. Οι μη-τερματικοί κόμβοι, οι οποίοι είναι η ρίζα και οι άλλοι εσωτερικοί κόμβοι περιέχουν χαρακτηριστικά γνωρίσματα δοκιμαστικών καταστάσεων για να διαχωρίζουν τις εγγραφές που έχουν διαφορετικά χαρακτηριστικά γνωρίσματα.

Η ταξινόμηση μιας δοκιμαστικής εγγραφής είναι απλή μόλις κατασκευαστεί το δένδρο απόφασης. Ξεκινώντας από την ρίζα, εφαρμόζουμε μια δοκιμαστική κατάσταση στην εγγραφή και ακολουθούμε το κατάλληλο «κλαδί» (μονοπάτι) βασισμένοι στο αποτέλεσμα

του τεστ. Αυτό θα μας οδηγήσει είτε σε έναν άλλο εσωτερικό κόμβο, για τον οποίο μια νέα δοκιμαστική κατάσταση θα εφαρμοστεί, ή θα μας οδηγήσει σε ένα κόμβο φύλλο.

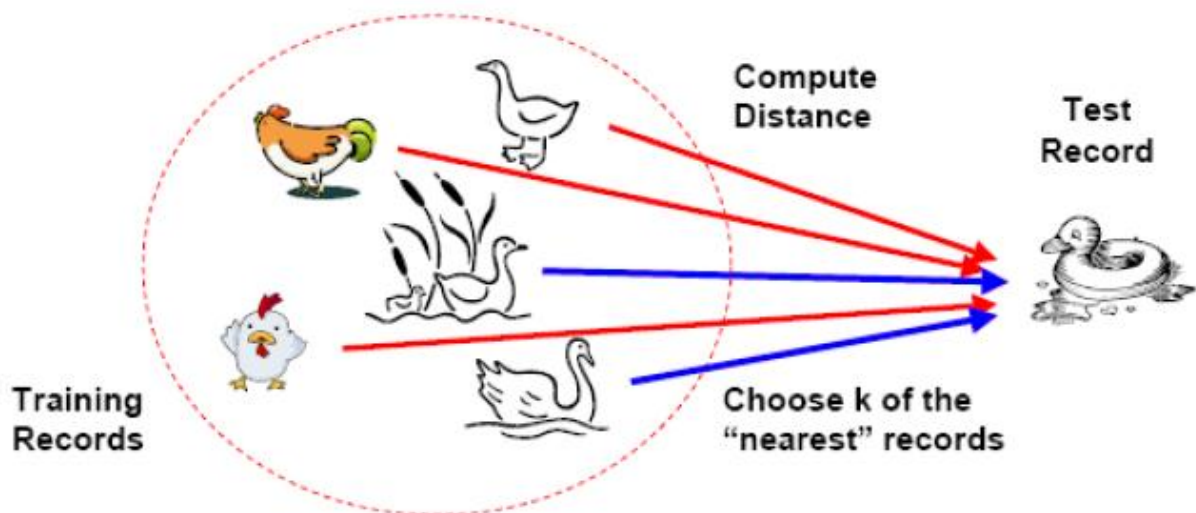


Σχήμα 2-5: Ανάθεση Πινακίδας Κατηγορίας σε μία Εγγραφή του Τεστ Σετ

2.4.4 Ταξινομητές Πλησιέστερων Γειτόνων (Nearest Neighbor)

Τα δένδρα απόφασης είναι ανυπόμονοι ταξινομητές εκμάθησης (eager learners) γιατί είναι σχεδιασμένα έτσι ώστε να μαθαίνουν ένα μοντέλο το οποίο χαρτογραφεί τα χαρακτηριστικά γνωρίσματα που εισάγονται και τα αναθέτει μία ετικέτα κατηγορίας μόλις το training set γίνει διαθέσιμο. Μια αντίθετη στρατηγική είναι να καθυστερήσουμε την διαδικασία της μοντελοποίησης – διαμόρφωσης του training set μέχρι αυτό να είναι απαραίτητο για την ταξινόμηση – classification του test set. Τεχνικές που αναπτύσσουν τέτοια στρατηγική ονομάζονται τεμπέλικοι ταξινομητές εκμάθησης (lazy learners). Ένα παράδειγμα τέτοιων ταξινομητών είναι ο ταξινομητής Rote, ο οποίος απομνημονεύει ολόκληρο το training set και πραγματοποιεί ταξινόμηση μόνο αν τα χαρακτηριστικά γνωρίσματα μίας δοκιμαστικής εγγραφής ταιριάζουν ακριβώς με αυτά που έχει μία εγγραφή του test set. Μία προφανής αδυναμία αυτής της διαδικασίας είναι ότι μερικές εγγραφές του test set μπορεί να μην μπορούν να ταξινομηθούν μιας και δεν ταιριάζουν με κάποιο παράδειγμα – εγγραφή (τα χαρακτηριστικά της γνωρίσματα) από το training set.

Ένας τρόπος για να βελτιωθεί αυτή η τεχνική και να γίνει πιο ευέλικτη είναι να βρούμε όλα εκείνα τα παραδείγματα του training set τα οποία έχουν σχετικά όμοια χαρακτηριστικά γνωρίσματα με την συγκεκριμένη εγγραφή του test set. Αυτά τα παραδείγματα, τα οποία είναι γνωστά σαν πλησιέστεροι γείτονες (nearest neighbors), μπορούν να χρησιμοποιηθούν για να προσδιορίσουμε την ετικέτα κατηγορίας της εγγραφής του test set. Η αιτιολόγηση του συγκεκριμένου παραδείγματος μπορεί να εξηγηθεί καλύτερα με την επόμενη φράση: «Εάν περπατά σαν πάπια, φωνάζει σαν πάπια και μοιάζει με πάπια τότε είναι πάπια» (Σχήμα 2-6).

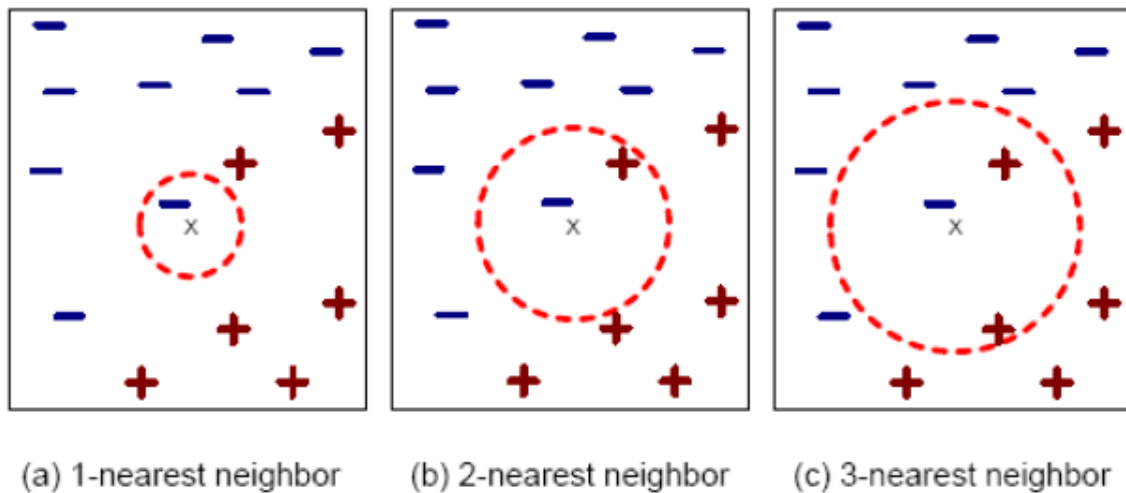


Σχήμα 2-6: Εάν περπατά σαν πάπια, φωνάζει σαν πάπια και μοιάζει με πάπια τότε είναι πάπια

Ένας ταξινομητής πλησιέστερων γειτόνων (nearest neighbor) αναπαριστά κάθε παράδειγμα σαν ένα σημείο δεδομένων σε ένα d -διάστατο χώρο, όπου d είναι ο αριθμός των χαρακτηριστικών γνωρισμάτων. Αν μας δοθεί ένα δοκιμαστικό παράδειγμα, τότε μπορούμε να υπολογίσουμε το κατά πόσο «κοντά» είναι στα υπόλοιπα σημεία δεδομένων (data points) του training set είναι. Οι k - nearest neighbors ενός δοσμένου παραδείγματος z αναφέρονται στα k σημεία τα οποία βρίσκονται πιο κοντά στο z .

Το Σχήμα 2.7, μας δείχνει τους 1-,2- και 3- nearest neighbors ενός σημείου δεδομένων, το οποίο βρίσκεται στο κέντρο του κάθε κύκλου. Η ταξινόμηση του σημείου δεδομένων βασίζεται στις ετικέτες κατηγορίας των γειτονικών του σημείων. Σε περίπτωση που οι γείτονες έχουν διαφορετικές ετικέτες κατηγορίας, τότε το σημείο δεδομένων ταξινομείται σύμφωνα με την πλειοψηφία των κοντινότερων γειτόνων. Στο Σχήμα 2.7 (a), ο 1-nearest neighbor του σημείου δεδομένων είναι ένα αρνητικό παράδειγμα. Επομένως το

σημείο δεδομένων προσδιορίζεται από αρνητική κλάση – κατηγορία. Εάν ο αριθμός των nearest neighbors είναι τρία (Σχήμα 2.7 (c)), τότε η γειτονιά περιέχει δύο θετικά και ένα αρνητικό παράδειγμα. Χρησιμοποιώντας την μέθοδο της πλειοψηφίας, στο σημείο δεδομένων θα ανατεθεί η στη θετική κλάση – κατηγορία. Σε περίπτωση που υπάρχει ισοπαλία μεταξύ δύο κλάσεων (Σχήμα 2-7 (b)), διαλέγουμε τυχαία μία από τις δύο για να ταξινομηθεί σύμφωνα με αυτήν το σημείο δεδομένων μας.

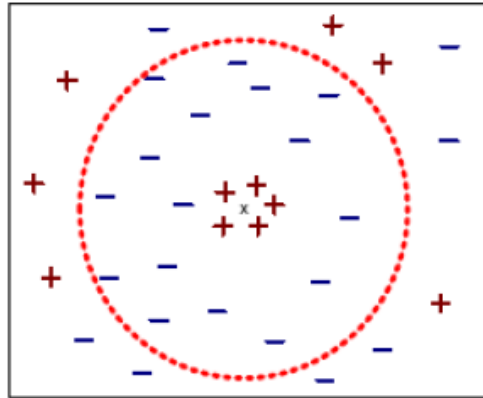


Σχήμα 2-7: Παραδείγματα Επίλυσης με k-nearest Neighbors, για k=1, 2, ή 3

Όλα τα παραπάνω δείχνουν το πόσο σημαντική είναι η επιλογή του κατάλληλου αριθμού για το k. Εάν το k είναι πολύ μικρό, τότε ο ταξινομητής nearest neighbor είναι ευάλωτος στο θόρυβο γιατί ο χώρος που λαμβάνεται υπόψη είναι ιδιαίτερα μικρός και οποιαδήποτε αλλαγή μπορεί να αλλοιώσει τα δεδομένα των λαμβανόμενων γειτόνων και άρα να γίνει λάθος η ταξινόμηση του σημείου δεδομένων. Από την άλλη, αν το k είναι πολύ μεγάλο, ο ταξινομητής nearest neighbor μπορεί να κάνει λάθος ανάθεση ετικέτας κατηγορίας του δοκιμαστικού παραδείγματος εξαιτίας του μεγέθους της λίστας των γειτόνων, η οποία θα είναι πολύ μεγάλη και θα παίρνει και τις ετικέτες κατηγορίας των πολύ μακρινών γειτόνων οι οποίοι μπορεί να έχουν διαφορετικές ετικέτες κατηγορίας και να μην έχουν τόσο όμοια χαρακτηριστικά γνωρίσματα αλλά βρίσκονται μέσα την περιοχή δειγμάτων. Για παράδειγμα στο Σχήμα 2-8, οι γείτονες γύρω από το σημείο δεδομένων είναι όλοι θετικοί, αλλά υπάρχουν περισσότεροι οι οποίοι είναι μακριά που είναι αρνητικοί. Αυτό έχει σαν αποτέλεσμα την λανθασμένη ανάθεση ετικέτας κατηγορίας στο σημείο δεδομένων.

Choosing the value of k:

- If k is too small, sensitive to noise points
- If k is too large, neighborhood may include points from other classes



Σχήμα 2-8: Αρνητικοί και Θετικοί Γείτονες γύρω από το Σημείο Δεδομένων

2.4.5 Μπεϋσιανός Ταξινομητής (Bayes Classifier)

Σε πολλές εφαρμογές η σχέση μεταξύ του συνόλου των χαρακτηριστικών γνωρισμάτων και της μεταβλητής της κατηγορίας είναι μη-ντετερμινιστική. Με άλλα λόγια, η ετικέτα κατηγορίας μιας εγγραφής του test set δεν μπορεί να προβλεφθεί με απόλυτη βεβαιότητα, ακόμα κι αν τα χαρακτηριστικά της γνωρίσματα είναι ίδια με μερικά μιας εγγραφής του training set. Η κατάσταση μπορεί να χειροτερέψει εξαιτίας δεδομένων που έχουν υποστεί κάποιο θόρυβο και μπορεί να έχουν παραμορφωθεί ή σε περίπτωση ύπαρξης συντελεστών που επηρεάζουν την ταξινόμηση και δεν έχουν ληφθεί υπόψη στην ανάλυση.

Οι Bayesian Classifiers, προσεγγίζουν ως τρόπο επίλυσης των προβλημάτων ταξινόμησης βασισμένο στις πιθανολογικές σχέσεις μεταξύ του συνόλου των χαρακτηριστικών γνωρισμάτων και της μεταβλητής της κλάσης – κατηγορίας. Χαρακτηριστικά γνωρίσματα των Bayesians μεθόδων εκμάθησης:

- Κάθε παράδειγμα παρατηρούμενο παράδειγμα εκμάθησης μπορεί να αυξήσει ή να μειώσει δραματικά την εκτιμώμενη πιθανότητα ότι μια υπόθεση – πρόβλεψη είναι σωστή.
- Η προγενέστερη γνώση μπορεί να συνδυαστεί με τα παρατηρηθέντα στοιχεία για να καθορίσει την τελική πιθανότητα μιας υπόθεσης.

- Οι Μπεϋσιανοί μέθοδοι μπορούν να προσαρμόσουν τις υποθέσεις που κάνουν τις πιθανολογικές προβλέψεις.
- Οι νέες περιπτώσεις μπορούν να ταξινομηθούν με το συνδυασμό των προβλέψεων των πολλαπλών υποθέσεων, που καθορίζονται από τις πιθανότητές τους.
- Μπορούν να παρέχουν πρότυπα για την λήψης της βέλτιστης απόφασης κατά την οποία μπορούν να μετρηθούν και άλλες πρακτικές μέθοδοι.

Βασικός Τύπος για Πιθανότητες

Κανόνας τομής: $P(A \cap B) = P(A|B) P(B)$

Κανόνας ένωσης: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Αθροιστική πιθανότητα: Αν A_1, A_2, \dots, A_n είναι ανεξάρτητα μεταξύ τους γεγονότα, τότε:

$$\sum_i P(A_i) = 1 \text{ ,οπότε } P(B) = \sum_i P(B | A_i) P(A_i)$$

Βασικοί Τύποι του Θεωρήματος Bayes

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (\text{class-conditional probability})$$

$P(A|B)$: Πιθανότητα να συμβεί το A δοθέντος του B.

$P(A)$: Πιθανότητα να συμβεί το A.

$P(B)$: Πιθανότητα να συμβεί το B.

$P(B|A)$: Πιθανότητα να συμβεί το B δοθέντος του A.

$$P(h|D) = P(D|h) P(h) / P(D)$$

$P(D)$: Προγενέστερη πιθανότητα των δεδομένων D, στοιχεία.

$P(h)$: Προγενέστερη πιθανότητα της υπόθεσης h, προγενέστερη.

$P(h|D)$: Μεταγενέστερη πιθανότητα την υπόθεση δεδομένων h των στοιχείων D, μεταγενέστερη.

$P(D|h)$: Πιθανότητα τα στοιχεία D δεδομένης της υπόθεσης h, πιθανότητα των στοιχείων.

Ταξινομητής Naïve Bayes

Ένας ταξινομητής Naïve Bayes υπολογίζει την class-conditional πιθανότητα θεωρώντας ότι τα χαρακτηριστικά γνωρίσματα είναι υπό όρους ανεξάρτητα μεταξύ τους, δεδομένης της ετικέτας κατηγορίας y . Η υπό όρους υποθετική ανεξαρτησία μπορεί να δηλωθεί τυπικά ως εξής:

$$P(X | Y = y) = \prod_{i=1}^d P(X_i | Y = y)$$

όπου κάθε σύνολο χαρακτηριστικών γνωρισμάτων $X = \{X_1, X_2, \dots, X_d\}$, έχει d χαρακτηριστικά γνωρίσματα για κάθε εγγραφή.

Λειτουργία ενός Ταξινομητή Naïve Bayes

Με την υπό όρους υποθετική ανεξαρτησία, αντί να υπολογίζουμε την class-conditional πιθανότητα για κάθε συνδυασμό του X , υπολογίζουμε μόνο την πιθανότητα του κάθε X_i , δοσμένου του Y . Η τελευταία προσέγγιση – μέθοδος είναι πιο πρακτική γιατί δεν απαιτεί ένα πολύ μεγάλο training set για να πετύχει μια καλή εκτίμηση της πιθανότητας.

Για να ταξινομήσουμε μία δοκιμαστική εγγραφή, ο Naïve Bayes ταξινομητής υπολογίζει τη μεταγενέστερη πιθανότητα για κάθε κατηγορία Y :

$$P(Y | X) = \frac{P(Y) \prod_{i=1}^d P(X_i | Y)}{P(X)}$$

Μιας και το $P(X)$ είναι προσαρμοσμένο για κάθε Y , το βέλτιστο είναι να επιλεγθεί η κατηγορία που μεγιστοποιεί τον αριθμητή $P(Y) \prod_{i=1}^d P(X_i | Y)$.

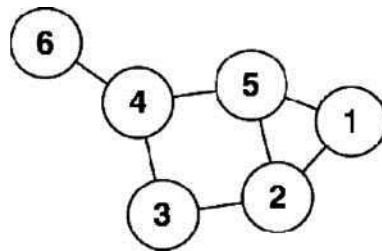
2.5 Δίκτυα, Τεχνητά Νευρωνικά Δίκτυα και Recommender Τεχνικές Νευρωνικών Δικτύων

2.5.1 Δίκτυα

Ο όρος δίκτυα (networks) χρησιμοποιείται για να περιγράψει τους γνωστούς, από την επιστήμη των μαθηματικών, γράφους που αναφέρονται στη βιβλιογραφία και με τον όρο γραφήματα. Ο όρος δίκτυα συναντάται κυρίως κατά την περιγραφή μεθοδολογιών βασισμένων σε γράφους στο χώρο της μηχανικής και γενικά τις εφαρμοσμένες επιστήμες.

Ένα δίκτυο αποτελείται από ένα σύνολο κορυφών ή κόμβων (vertices, nodes) και ένα σύνολο ακμών ή πλευρών (edges). Οι κορυφές είναι οντότητες του συστήματος που

μελετάται με τη χρήση του δικτύου, ενώ οι ακμές μεταξύ κορυφών χρησιμοποιούνται για να εκφράσουν τη σχέση μεταξύ των οντοτήτων του συστήματος. Η σχέση αυτή μπορεί να εκφράζεται με τη χρήση βαρών στις ακμές ή χωρίς βάρη στη περίπτωση μη βαθμωτής σχέσης. Στη δεύτερη περίπτωση η σχέση υποδηλώνεται με την ύπαρξη (ή όχι) ακμής μεταξύ των κορυφών. Επίσης, η ακμή σε ένα δίκτυο μπορεί να είναι προσανατολισμένη (κατευθυνόμενο δίκτυο) ή όχι (μη κατευθυνόμενο δίκτυο). Στο Σχήμα 2-9 παρουσιάζεται το παράδειγμα ενός μη κατευθυνόμενου δικτύου 6 κορυφών χωρίς βάρη.



Σχήμα 2-9: Μη κατευθυνόμενο δίκτυο χωρίς βάρη με 6 κορυφές

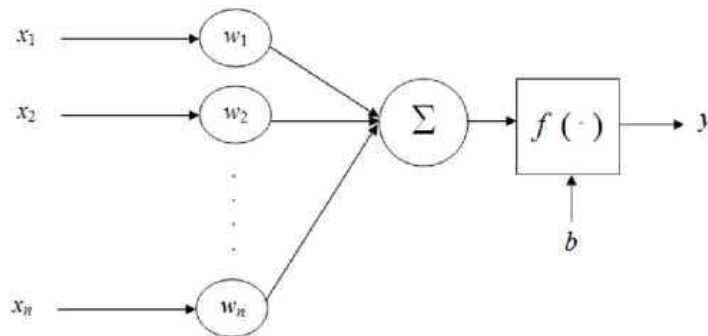
Τα δίκτυα αποτελούν σημαντικά εργαλεία των εφαρμοσμένων επιστημών και συντελούν στην επίλυση πρακτικών προβλημάτων. Έτσι μπορούν και αναπαριστούν πολύπλοκα συστήματα που απαντώνται στο φυσικό και τεχνολογικό κόσμο και εμφανίζουν δικτυακή δομή. Παραδείγματα αποτελούν τα βιολογικά και χημικά δίκτυα, η διατροφική αλυσίδα, τα δίκτυα υπολογιστών και το διαδίκτυο, αλλά και τα κοινωνιολογικά δίκτυα (Bar-Yam, 1997). Η χρήση των δικτύων στη μελέτη πολύπλοκων συστημάτων επιτρέπει την εφαρμογή αλγορίθμων της θεωρίας γράφων με σκοπό την επίλυση πρακτικών προβλημάτων, π.χ. την εφαρμογή αλγορίθμου για την εύρεση του πιο σύντομου μονοπατιού ανάμεσα σε δυο κορυφές. Παράλληλα, δίνεται η δυνατότητα εκτίμησης της συνδεσιμότητας που υπάρχει καθολικά στο μελετώμενο σύστημα και της επικινδυνότητας κατάρρευσης του συστήματος σε περίπτωση επιθέσεων σε αυτό, δηλαδή αφαίρεσης των κορυφών του αντίστοιχου δικτύου. Τελευταία, δε, έχουν αναπτυχθεί διάφορα θεωρητικά μοντέλα δικτύων, όπως τα δίκτυα μικρού κόσμου, τα τυχαία δίκτυα και τα ανεξάρτητα κλίμακας δίκτυα, που μοντελοποιούν τα συστήματα δικτυακής δομής που απαντώνται στον πραγματικό κόσμο (Albert and Barabasi, 2002).

2.5.2 Τεχνητά Νευρωνικά Δίκτυα

Τα τεχνητά νευρωνικά δίκτυα (artificial neural networks) (Haykin, 1999) ανήκουν στο

πεδίο της υπολογιστικής νοημοσύνης (computational intelligence) το οποίο περιλαμβάνει, επίσης, την ασαφή λογική (fuzzy logic) και τον εξελικτικό υπολογισμό (evolutionary computing). Αποτελούν συστήματα παράλληλης σύνδεσης ενός αριθμού απλών επεξεργαστών και μιμούνται τον ανθρώπινο εγκέφαλο με σκοπό την επίλυση πολύπλοκων προβλημάτων. Κάθε επεξεργαστής αντιστοιχεί σε ένα τεχνητό νευρώνα, η λειτουργία του οποίου βασίζεται στις αρχές του βιολογικού νευρώνα του νευρικού συστήματος. Τα τεχνητά νευρωνικά δίκτυα έχουν την ιδιότητα να εκπαιδεύονται μαθαίνοντας από την εμπειρία και να ταξινομούν τα εφαρμοζόμενα στην είσοδο τους δεδομένα ή να προβλέπουν την τιμή μιας συνεχής μεταβλητής εξόδου.

Ο τεχνητός νευρώνας (Σχήμα 2-10) αποτελεί τη θεμελιώδη μονάδα επεξεργασίας πληροφορίας στο τεχνητό νευρωνικό δίκτυο και αποτελείται από τρία στοιχεία: ένα σύνολο n κλάδων διασύνδεσης (συνάψεις), έναν κόμβο άθροισης και μια συνάρτηση ενεργοποίησης.



Σχήμα 2-10: Τεχνητός Νευρώνας

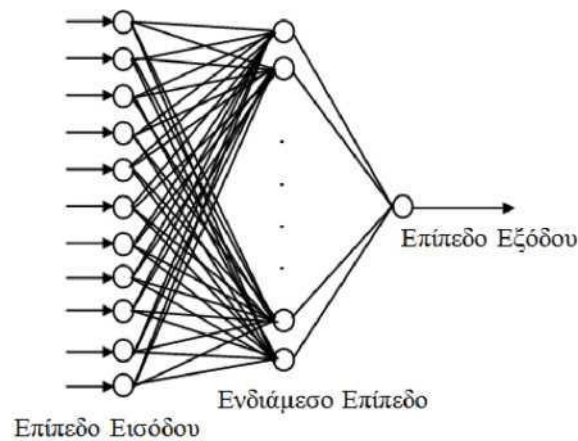
Κάθε κλάδος διασύνδεσης έχει ένα βάρος w_i ενώ ο κόμβος άθροισης αθροίζει τα σήματα εισόδου x_1, x_2, \dots, x_n πολλαπλασιαζόμενα με τα αντίστοιχα βάρη των συνάψεων. Στη συνάρτηση ενεργοποίησης εφαρμόζεται εξωτερικά μια πόλωση b που προστίθεται στην είσοδο της. Η συνάρτηση ενεργοποίησης περιορίζει το επιτρεπόμενο πλάτος του σήματος εξόδου του τεχνητού νευρώνα σε κάποιο διάστημα, π.χ. $[0,1]$ ή $[-1,1]$, και προσθέτει μη γραμμικότητα στην έξοδο του. Τελικά, η έξοδος του τεχνητού νευρώνα δίνεται από τη σχέση:

$$y = f\left(\sum_{i=1}^n w_i x_i + b\right)$$

Ανάλογα με την αρχιτεκτονική διασύνδεσης των τεχνητών νευρώνων διακρίνονται διάφορα είδη τεχνητών νευρωνικών δικτύων, όπως τα πολυστρωματικά τεχνητά νευρωνικά δίκτυα πρόσθιας τροφοδότησης (multilayer feed-forward neural networks) και τα αναδρομικά

τεχνητά νευρωνικά δίκτυα (recursive neural networks).

Τα πολυστρωματικά τεχνητά νευρωνικά δίκτυα πρόσθιας τροφοδότησης αποτελούν την πιο συνήθη και απλούστερη μορφή. Αποτελούνται από νευρώνες οργανωμένους σε επίπεδα: το επίπεδο εισόδου στο οποίο εφαρμόζεται το διάνυσμα εισόδου, ένα ή περισσότερα ενδιάμεσα (ή κρυμμένα) επίπεδα με νευρώνες που παρεμβαίνουν μεταξύ του επιπέδου εισόδου και του επιπέδου εξόδου, και το επίπεδο εξόδου που δίνει τη συνολική απόκριση του δικτύου στο διάνυσμα εισόδου. Η ροή του σήματος εισόδου γίνεται από το επίπεδο εισόδου προς το επίπεδο εξόδου. Παράδειγμα ενός τέτοιου δικτύου με ένα ενδιάμεσο επίπεδο παρουσιάζεται στο Σχήμα 2-11. Ας σημειωθεί εδώ ότι μπορεί κάποιος να φανταστεί απλοϊκά ένα πολυστρωματικό τεχνητό νευρωνικό δίκτυο πρόσθιας τροφοδότησης ως ένα κατευθυνόμενο δίκτυο, όπως αυτό που περιγράφηκε στη προηγούμενη παράγραφο.



Σχήμα 2-11: Πολυστρωματικό Τεχνητό Νευρωνικό Δίκτυο Πρόσθιας Τροφοδότησης με Ενδιάμεσα Επίπεδα

Όπως ήδη αναφέρθηκε, τα τεχνητά νευρωνικά δίκτυα εκπαιδεύονται ώστε να αποκτούν μια επιθυμητή συμπεριφορά στο υπό μελέτη πρόβλημα. Ανάλογα με το περιβάλλον εργασίας και τα διαθέσιμα δεδομένα διακρίνονται οι εξής τρόποι εκπαίδευσης: η επιβλεπόμενη, η ενισχυτική και η μη επιβλεπόμενη εκπαίδευση. Η επιβλεπόμενη εκπαίδευση είναι ο πιο συνήθης τρόπος εκπαίδευσης και χρησιμοποιεί δεδομένα για τα οποία είναι γνωστές οι εισοδοί αλλά και η επιθυμητή έξοδος. Κατά της διαδικασία εκπαίδευσης γίνεται χρήση των δεδομένων του λεγόμενου συνόλου εκπαίδευσης και κατάλληλου αλγόριθμου εκπαίδευσης (ή εκμάθησης), ώστε να επιτευχθεί η προσαρμογή των μεταβλητών παραμέτρων (π.χ. βάρη των συνάψεων και τιμές πόλωσης) ενός τεχνητού νευρωνικού δικτύου στο μελετώμενο πρόβλημα. Ως αποτέλεσμα, το τεχνητό νευρωνικό δίκτυο αποκτά την ικανότητα γενίκευσης, δηλαδή μπορεί και εκτιμά την τιμή της μεταβλητής εξόδου (συνεχής ή

κατηγορική) καθώς παρουσιάζονται νέα διανύσματα εισόδου, άγνωστα σε αυτό. Η ικανότητα γενίκευσης ελέγχεται σε ένα διαφορετικό σύνολο δεδομένων από το σύνολο εκπαίδευσης που συνήθως ονομάζεται σύνολο ελέγχου.

Για το συχνά απαντώμενο πολυστρωματικό τεχνητό νευρωνικό δίκτυο πρόσθιας τροφοδότησης διάσημος αλγόριθμος εκπαίδευσης είναι ο αλγόριθμος όπισθεν διάδοσης σφάλματος (error backpropagation) (Haykin, 1999). Ο συγκεκριμένος αλγόριθμος είναι αλγόριθμος επιβλεπόμενης εκπαίδευσης που προσαρμόζει τα βάρη των συνάψεων, έτσι ώστε να ελαχιστοποιηθεί το μέσο τετραγωνικό σφάλμα μεταξύ των επιθυμητών και πραγματικών αποκρίσεων του τεχνητού νευρωνικού δικτύου στο σύνολο εκπαίδευσης. Η προσαρμογή των βαρών γίνεται ανά επίπεδο νευρώνων με κατεύθυνση από το επίπεδο εξόδου προς το επίπεδο εισόδου. Η επαναληπτική διαδικασία εκπαίδευσης συνήθως τελειώνει με την ολοκλήρωση ενός προκαθορισμένου αριθμού «εποχών», κάθε μία από τις οποίες αντιστοιχεί στη χρήση για φορά του συνόλου εκπαίδευσης από τον αλγόριθμο και κατάλληλη προσαρμογή με βάση αυτό των βαρών των συνάψεων.

2.5.3 Τεχνικές Προτάσεων Νευρωνικών Δικτύων

Τεχνικές προτάσεων που βασίζονται στα τεχνητά νευρωνικά δίκτυα χρησιμοποιούνται στον τομέα του γεωγραφικού εντοπισμού θέσης για χωρικές σχέσεις (Lee and Jekeli, 2010; Rutecki and Kacprzak, 2008). Τέτοιου είδους συστήματα είναι μοντέλα που βασίζονται στη νευρωνική δομή του εγκεφάλου, ο οποίος μαθαίνει από την εμπειρία. Σε συναφή προβλήματα, οι υπολογιστικοί φορείς έχουν πρόβλημα στο να αναγνωρίσουν ακόμα και απλά σχέδια, ενώ το πρόβλημα γίνεται ακόμα πιο δύσκολο όταν πρόκειται για τη γενίκευση αυτών των προτύπων του παρελθόντος σε δράσεις του μέλλοντος. Έτσι, τέτοιου είδους συστημάτων προτάσεων είναι παρακινημένα βιολογικά και κυρίως βασισμένα στη στατιστική. Τέτοιου είδους υπολογιστικοί μηχανισμοί είναι γνωστοί για την ικανότητά τους να πραγματοποιούν ταχείς συσχετισμούς μνήμης αντί για υψηλής ακρίβειας υπολογιστική επεξεργασία. Επίσης παρέχουν τα μέσα για εργασίες που αφορούν μεγάλες ποσότητες δεδομένων και δυναμικά περιβάλλοντα. Ειδικότερα, ο αυτό-οργανωμένος χάρτης (Self-Organizing Map – SOM) είναι ένα γενικό εργαλείο για την παραγγελία μεγάλων διαστάσεων δεδομένων με τέτοιο τρόπο, ώστε παρόμοια πρότυπα εισόδου να ομαδοποιούνται χωρικά το ένα κοντά στο άλλο (Kohonen, 1995). Επίσης οι ταξινομητές στα ασαφή νευρωνικά δίκτυα (Fuzzy Neural Networks) εκτελούν ταξινόμηση ή ομαδοποίηση χρησιμοποιώντας στοιχεία που προέρχονται από θεωρίες για την ανθρώπινη αίσθηση της αντίληψης. Με βάση αυτές τις θεωρίες, η

εργασία της ομαδοποίησης / κατάταξης αναφέρεται ως πρόβλημα διάκρισης υφής που υλοποιήθηκε σαν προσεγγιζόμενη διαδικασία.

Στη σχετική βιβλιογραφία, το πρόβλημα της αναγνώρισης των προτιμήσεων του χρήστη όσον αφορά τα αντικείμενα / υπηρεσίες που θα μπορούσαν να τον ενδιαφέρουν, προσεγγίστηκε από διάφορες τεχνικές προτάσεων βασισμένες στα νευρωνικά δίκτυα, χρησιμοποιώντας μεταξύ άλλων την κοινωνική δικτύωση (social networking) και τις συστάσεις της σοφίας του πλήθους (crowd wisdom recommendations) (Kayaalp, Ozyer T. and Ozyer S., 2009).

Στην παρούσα διατριβή, το εργαλείο που χρησιμοποιήθηκε για το σύστημα προτάσεων είναι ένας ταξινομητής νευρωνικού δικτύου πιθανοτήτων (Probabilistic Neural Network – PNN), ο οποίος προσαρμόστηκε κατάλληλα ώστε να έχει ως στόχο τον εντοπισμό και την κατηγοριοποίηση event με γεωγραφική πληροφορία, σύμφωνα με τις προτιμήσεις του χρήστη, παρέχοντας παράλληλα πιο ακριβείς πληροφορίες στους χρήστες, σχετικά με τα γεγονότα που μπορεί να τους ενδιαφέρουν.

2.6 Συστήματα Διαχείρισης Βάσεων Δεδομένων

Με τον όρο βάση δεδομένων εννοείται μία συλλογή από συστηματικά οργανωμένα σχετιζόμενα δεδομένα, ηλεκτρονικά αποθηκευμένα. Πέρα από την εγγενή της ικανότητα να αποθηκεύει δεδομένα, η βάση δεδομένων παρέχει τη δυνατότητα γρήγορης άντλησης και ανανέωσης των δεδομένων. Οι ηλεκτρονικές βάσεις δεδομένων χρησιμοποιούν ιδιαίτερου τύπου λογισμικό προκειμένου να οργανώσουν την αποθήκευση των δεδομένων τους. Το διακριτό αυτό λογισμικό είναι γνωστό ως Σύστημα Διαχείρισης Βάσης Δεδομένων (DataBase Management System, συντομευμένα DBMS). Βασικός στόχος των DBMS είναι η παροχή μεθόδων αποθήκευσης και άντλησης των πληροφοριών των βάσεων δεδομένων, που να είναι βολικές, αποτελεσματικές και δομημένες.

Τα συστήματα διαχείρισης βάσεων δεδομένων είναι σχεδιασμένα με σκοπό το χειρισμό μεγάλου όγκου πληροφοριών. Ο σαφής ορισμός δομών απαραίτητων για την αποθήκευση των πληροφοριών και η παροχή μηχανισμών για το χειρισμό τους αποτελούν τα συστατικά στοιχεία της διαχείρισης των δεδομένων. Τα DBMS επιτρέπουν στους οργανισμούς να παραχωρούν τον έλεγχο της ανάπτυξης των βάσεων δεδομένων τους σε ειδικευμένους διαχειριστές βάσεων δεδομένων (Database Administrators, DBAs). Τα συστήματα αυτά μπορεί να χρησιμοποιούν οποιοδήποτε από μία ποικιλία μοντέλων βάσεων δεδομένων, όπως τα hierarchical model, network model και relational model. Το μοντέλο που

επικρατεί σήμερα είναι το σχεσιακό (relational model) (Abidi and Yu-N, 2000; Shasha, Bonnet and Bercich, 2004; Silberschatz, Korth and Sudarshan, 2002).

Επιπλέον ένα DBMS είναι απαραίτητο να εγγυάται την ασφάλεια των δεδομένων και τη συνέπεια των αποτελεσμάτων. Πιο συγκεκριμένα, ακόμα και σε περιπτώσεις κατάρρευσης του συστήματος ή προσπάθειας πρόσβασης από μη πιστοποιημένο χρήστη, οι πληροφορίες θα πρέπει να παραμένουν ακέραιες. Ακόμα, αν κάποια δεδομένα είναι κοινόχρηστα το σύστημα θα πρέπει να διασφαλίζει ότι δε θα υπάρξουν λανθασμένα αποτελέσματα.

Η αποτελεσματική οργάνωση και διαχείριση των πληροφοριών και η εύκολη και ασφαλής πρόσβαση και μετάδοσή τους αποτελεί σημαντική προτεραιότητα όλων των επιχειρήσεων και των οργανισμών, ιδιαίτερα καθώς το μέγεθός τους μεγαλώνει. Οι σημερινές τεχνικές διαχείρισης των δεδομένων παρέχουν αποτελεσματικούς τρόπους για την αποδοτική διεκπεραίωση αυτών των εργασιών. Επιπλέον οι δυνατότητες διατήρησης και επεξεργασίας δεδομένων από τις κινητές συσκευές έχουν αυξηθεί εντυπωσιακά. Όλο και μεγαλύτερος όγκος πληροφοριών γίνεται ευκολότερα προσβάσιμος από όλο και μικρότερες συσκευές. Ακόμα η αναζήτηση μέσω Internet βελτιώνεται συνεχώς και γίνεται πιο αποτελεσματική. Οι search engines του Web εκμεταλλεύονται την ανάπτυξη στον τομέα της διαχείρισης των δεδομένων και παρέχουν διαρκώς περισσότερες επιλογές για αποδοτικότερη και ευκολότερη εύρεση πληροφοριών. Ένας άλλος τομέας που έχει εκμεταλλευθεί τις εξελίξεις στη διαχείριση των δεδομένων είναι αυτός του ηλεκτρονικού εμπορίου. Οι online συναλλαγές έχουν γίνει πολύ πιο ασφαλείς και συνεπώς ιδιαίτερα δημοφιλείς.

Συμπερασματικά, οι εξελίξεις στην τεχνολογία της διαχείρισης των βάσεων δεδομένων έχουν επιφέρει αλλαγές σε μια πληθώρα διαφορετικών τομέων. Όμως, η ανάγκη περαιτέρω ανάπτυξης και καινοτομίας θα είναι πάντα υπαρκτή, όσο οι απαιτήσεις των χρηστών θα αυξάνονται και θα διαφοροποιούνται.

Αυτή τη στιγμή στην αγορά κυκλοφορεί μία πληθώρα διαφορετικών συστημάτων διαχείρισης βάσεων δεδομένων, καθένα από τα οποία έχει συγκεκριμένα ιδιαίτερα χαρακτηριστικά που το διαχωρίζουν από τα υπόλοιπα. Τα πλεονεκτήματα και τα μειονεκτήματα του καθενός είναι ποικίλα και η επιλογή ενός χρήστη ή μιας επιχείρησης βασίζεται στις συγκεκριμένες ανάγκες και απαιτήσεις που υπάρχουν από το σύστημα.

Για την υλοποίηση του συστήματος επιλέχθηκε η Oracle και συγκεκριμένα η έκδοση Oracle Database 11g Express Edition. Πρόκειται για μία έκδοση της βάσης, η οποία συνεργάζεται άψογα με εφαρμογές που υλοποιούνται με το Visual Studio .NET της Microsoft Visual Studio .Net και επιτρέπει τη δωρεάν χρήση, ανάπτυξη και διανομή εφαρμογών.

3 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΣΥΣΤΗΜΑΤΟΣ

Στην ενότητα αυτή ακολουθεί μια θεωρητική περιγραφή του πληροφοριακού συστήματος που δημιουργήθηκε, καθώς και μία ανάλυση των επιμέρους λειτουργιών του.

3.1 Αρχιτεκτονική

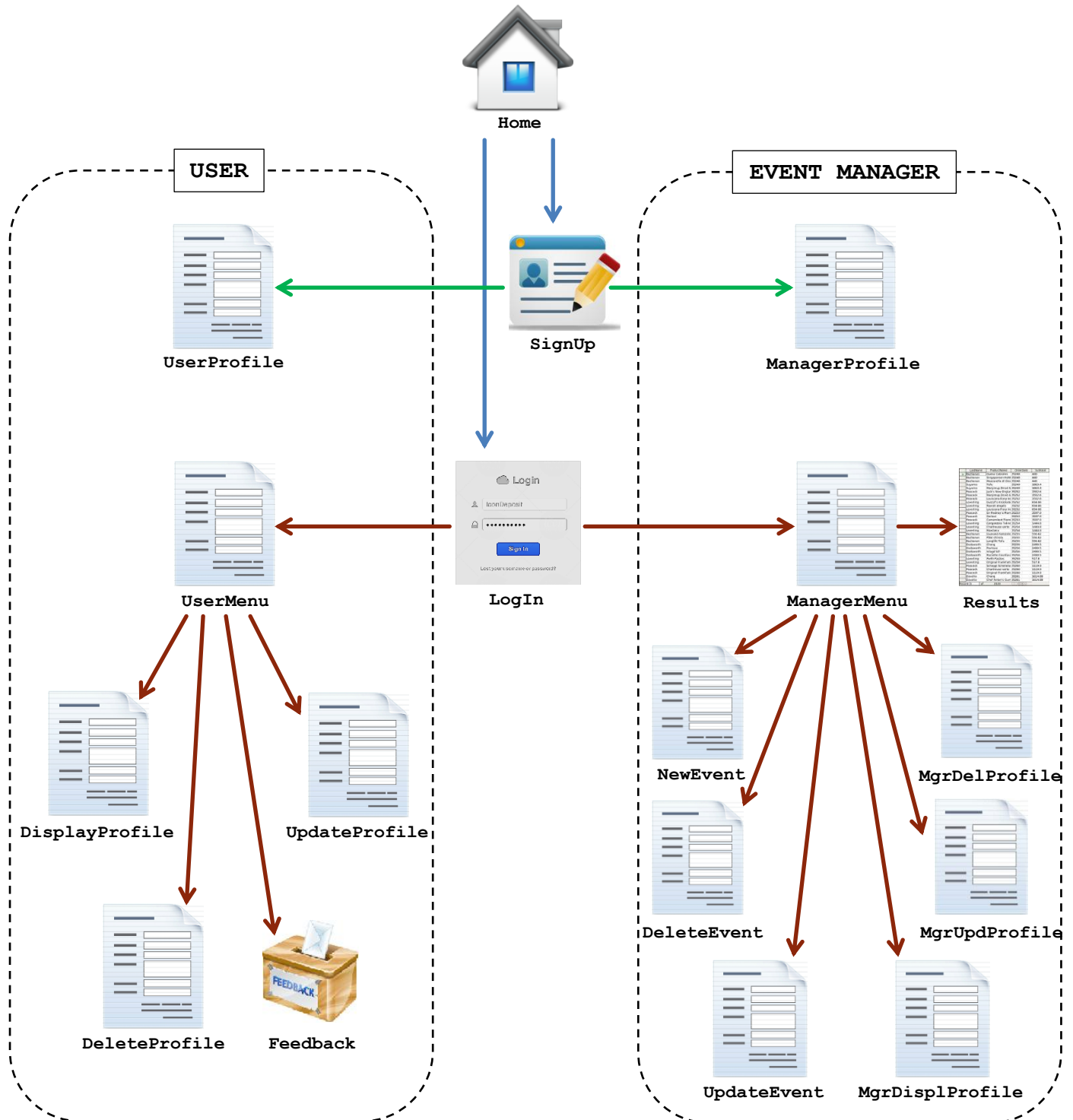
Σκοπός αυτής της διατριβής, είναι η δημιουργία ενός πληροφοριακού συστήματος παροχής υπηρεσιών υψηλής προστιθέμενης αξία μέσω διαδικτύου. Το σύστημα παρέχει πληροφορίες για διάφορα *events* που οργανώνονται στο λεκανοπέδιο της Αττικής σε εγγεγραμμένους χρήστες και απαρτίζεται από τρία υποσυστήματα.

Το πρώτο είναι ένας ιστότοπος, (Σχήμα 3-1) προσβάσιμος από το World Wide Web μέσω οποιουδήποτε web browser. Χρησιμοποιείται τόσο από τους χρήστες (users) του συστήματος που επιθυμούν να λαμβάνουν προσωποποιημένες προτάσεις για events, όσο και από τις εταιρίες-πελάτες (customers) που διοργανώνουν τα συγκεκριμένα events και έχουν σκοπό να τα προωθήσουν. Σε χρήστες και εταιρίες προσφέρεται η δυνατότητα εγγραφής στο σύστημα και η δημιουργία των εικονικών τους προφίλ, μέσω της συμπλήρωσης κατάλληλων forms. Στη συνέχεια, δίνεται η δυνατότητα και στις δύο κατηγορίες να προβάλουν και επεξεργαστούν τις πληροφορίες που καταχώρησαν κατά την εγγραφή. Από εκεί και πέρα, οι ενδιαφερόμενες εταιρίες μπορούν να προχωρήσουν στην καταχώρηση ενός νέου event, ενώ παράλληλα έχουν πρόσβαση στις λίστες με τα αποτελέσματα του αλγόριθμου προτάσεων, για κάθε event τους για το οποίο έχει ήδη τρέξει ο αλγόριθμος.

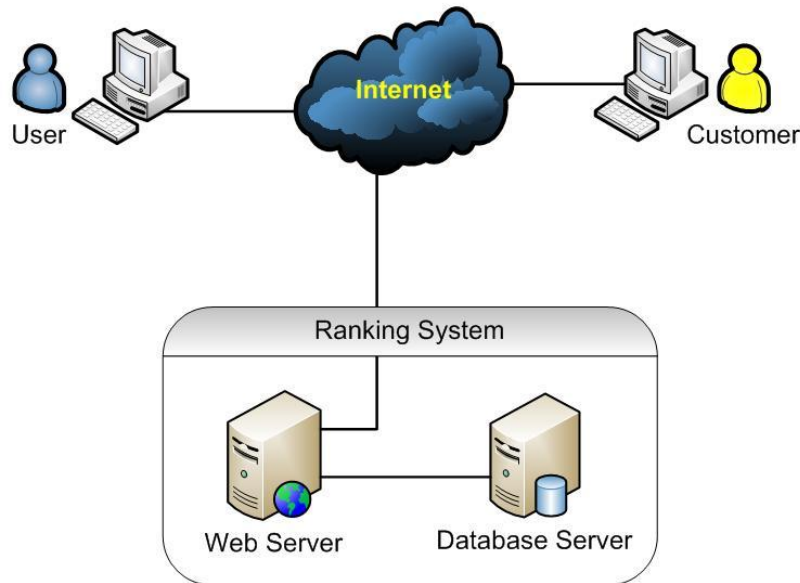
Το δεύτερο υποσύστημα είναι ένα Windows Service, το οποίο εκτελεί αυτόματα τον αλγόριθμο για οποιοδήποτε ενεργό καταχωρημένο event, αλλά και κάθε φορά που προστίθεται ένας καινούριος user ή event. Εντοπίζονται οι N καταλληλότεροι χρήστες για κάθε event, οι οποίοι αποθηκεύονται στη βάση και από εκεί, κάποια στοιχεία τους είναι προσβάσιμα στη διαφημιζόμενη εταιρία για να προβεί σε περαιτέρω ενέργειες, όπως ή αποστολή διαφημιστικών μηνυμάτων μέσω SMS ή e-mail.

Το τρίτο υποσύστημα είναι ο Database Server. Λειτουργίες του υποσυστήματος αυτού, αποτελούν οι procedures για την εισαγωγή δεδομένων στην database, καθώς και οι procedures για τη δημιουργία κλειδιών των χρηστών και events. Το Windows Service, τρέχει πάνω στον Database Server ώστε να ελαχιστοποιείται ο χρόνος μεταφοράς δεδομένων που

χρειάζονται για τους υπολογισμούς. Το Σχήμα 3-2 αναπαριστά, σε υψηλό επίπεδο τη διασύνδεση τους συστήματος.



Σχήμα 3-1: Χάρτης Ιστότοπου



Σχήμα 3-2: Αναπαράσταση Διασύνδεσης Συστήματος

3.2 Σχεσιακό Διάγραμμα Οντοτήτων

Ακολουθεί η περιγραφή των οντοτήτων (Πίνακας 3-1) του Σχεσιακού Μοντέλου της βάσης δεδομένων του συστήματος (Entity-Relationship Diagram) (Σχήμα 3-3), καθώς και η περιγραφή των μεταξύ τους σχέσεων (Πίνακας 3-2), σύμφωνα με τους Silberschatz, Korth και Sudarshan (2002).

3.2.1 Οντότητες

Οντότητα	Περιγραφή
Users	Αποθηκεύονται τα προφίλ των χρηστών του συστήματος και περιλαμβάνουν πληροφορίες όπως τα προσωπικά τους στοιχεία (όνομα, ηλικία, φύλο), η τοποθεσία τους, καθώς και στοιχεία επικοινωνίας (τηλέφωνο, email).
UserKeys	Αποθηκεύονται κωδικοποιημένες πληροφορίες που προκύπτουν κατά την δημιουργία των προφίλ των χρηστών στο σύστημα. Η πληροφορία ανανεώνεται κάθε φορά που οι χρήστες κάνουν αλλαγές στο προφίλ τους.

AttendanceFeedback	Αποθηκεύονται οι απαντήσεις των χρηστών για το αν θα παρακολουθήσουν ή όχι το event για το οποίο έχουν λάβει μήνυμα, αν τους ενδιαφέρει γενικά σαν αντικείμενο, καθώς και κάποια σχόλια που μπορούν να αφήσουν οι χρήστες.
RatingFeedback	Αποθηκεύεται η βαθμολογία των χρηστών για το event που παρακολούθησαν.
EventManagers	Αποθηκεύονται τα προφίλ των διοργανωτών events και περιλαμβάνουν πληροφορίες όπως το όνομα της εταιρίας, καθώς και στοιχεία επικοινωνίας (τηλέφωνο, email).
Events	Αποθηκεύονται τα event του συστήματος και περιλαμβάνουν πληροφορίες όπως ο τίτλος τους, η τοποθεσία τους, η διάρκεια τους, η κατάστασή τους (αν είναι ενεργά ή έχουν λήξει), καθώς και πληροφορίες για τους χρήστες που απευθύνονται (ηλικία, φύλο).
EventKeys	Όμοια με το UserKeys. Αποθηκεύονται κωδικοποιημένες πληροφορίες που προκύπτουν κατά την δημιουργία των events στο σύστημα. Η πληροφορία ανανεώνεται κάθε φορά που οι διοργανωτές κάνουν αλλαγές στα στοιχεία των events.
Ranking	Αποθηκεύεται η βαθμολογία των χρηστών ως προς τα event, η οποία προκύπτει από τον αλγόριθμο του συστήματος και εκφράζει το πόσο μπορεί να ενδιαφέρονται για να τα παρακολουθήσουν. Επίσης αποθηκεύεται η κατάσταση της βαθμολογίας (αν είναι ενεργή ή όχι), καθώς και η ημερομηνία με την τελευταία φορά που ανανεώθηκε.
PrefCategories	Αποθηκεύονται όλες οι προτιμήσεις των χρηστών, καθώς και ο χαρακτηρισμός των events για τις υποκατηγορίες των 5 κύριων κατηγοριών (Τέχνη, Αθλητικά, Επιστήμη, Ψυχαγωγία και Υπολογιστές) .
EventNotifications	Αποθηκεύονται όλα τα μηνύματα που αποστέλλονται για να ειδοποιήσουν τους χρήστες για τα events που μπορεί να τους ενδιαφέρουν.
Credentials	Αποθηκεύονται τα διαπιστευτήρια με τα οποία εισέρχονται στο σύστημα οι χρήστες και οι διοργανωτές των events.

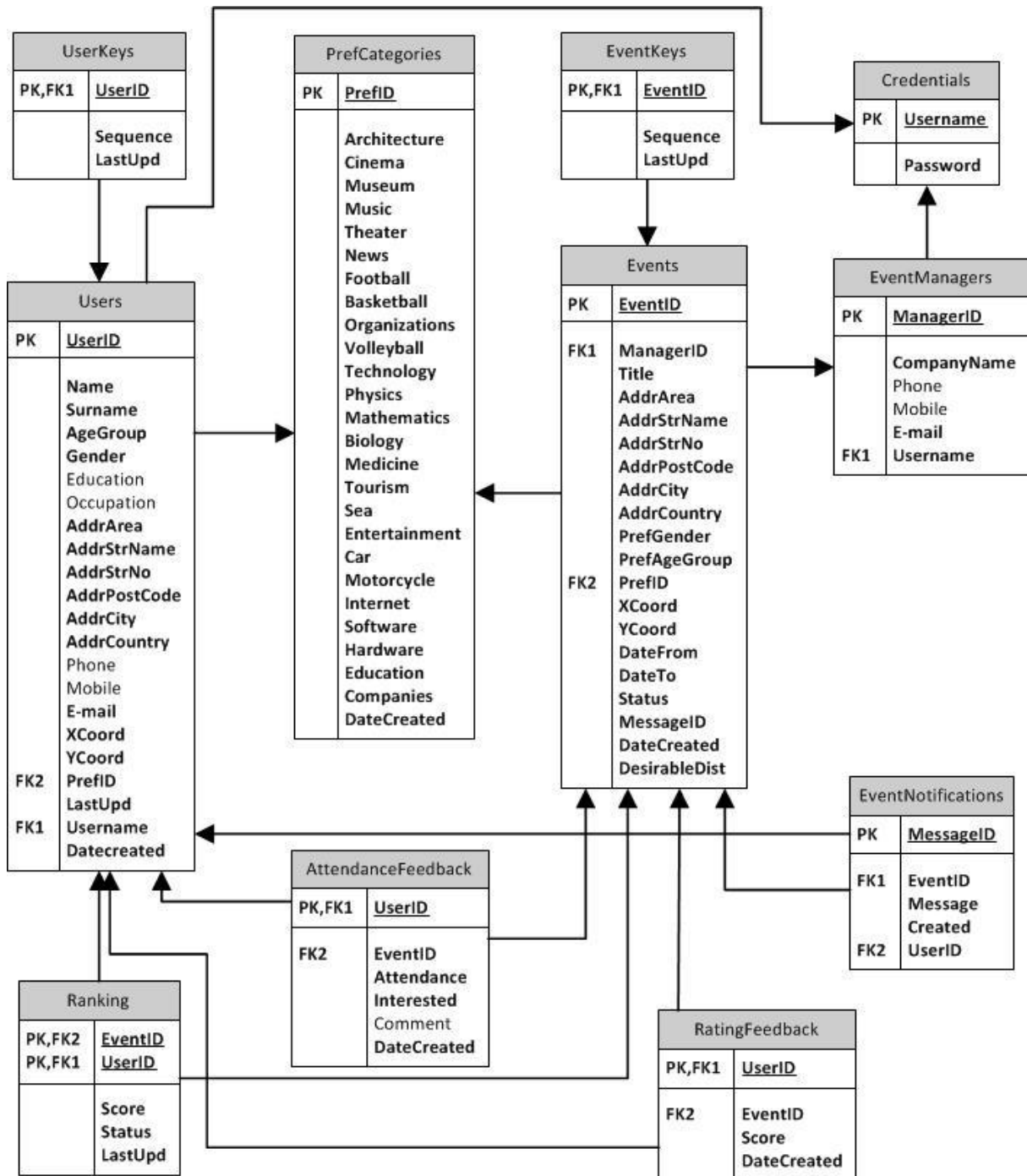
Πίνακας 3-1: Περιγραφή Οντοτήτων

3.2.2 Σχέσεις Οντοτήτων

Σχέση Οντοτήτων	Περιγραφή
Credentials - Users	Η σχέση τους είναι 1 προς 1 αφού ένας χρήστης μπορεί να έχει ένα ζευγάρι διαπιστευτηρίων για την είσοδο του στο σύστημα
Credentials - EventManagers	Η σχέση τους είναι 1 προς 1 αφού ένας διοργανωτής event μπορεί να έχει ένα ζευγάρι διαπιστευτηρίων για την είσοδο του στο σύστημα
Users – UserKeys	Η σχέση τους είναι 1 προς 1 αφού για κάθε χρήστη μπορεί να δημιουργηθεί ένα μόνο κλειδί. Ακόμα και όταν ο χρήστης αλλάζει τα στοιχεία του, το κλειδί ανανεώνεται.
Events - EventKeys	Η σχέση τους είναι 1 προς 1 αφού για κάθε event μπορεί να δημιουργηθεί ένα μόνο κλειδί. Ακόμα και όταν ο διοργανωτής αλλάζει τα στοιχεία του event, το κλειδί ανανεώνεται.
EventManagers - Events	Η σχέση τους είναι 1 προς N μιας και κάθε διοργανωτής μπορεί να διοργανώνει πολλά events.
Ranking – Users - Events	Συνδέει τις οντότητες Users και Events και η σχέση τους είναι N προς M, αφού κάθε event μπορεί να συνδέεται με πολλούς χρήστες που έχουν βαθμό ως προς αυτό και αντίστροφα ένας χρήστης μπορεί να συνδέεται με πολλά events που έχει βαθμό ως προς αυτά.
PrefCategories – Users - Events	Συνδέει τις οντότητες Users και Events και η σχέση τους είναι 1 προς 1, αφού κάθε χρήστης και κάθε event μπορεί να έχει από ένα σετ προτιμήσεων και χαρακτηρισμού αντίστοιχα.
AttendanceFeedback – Users - Events	Συνδέει τις οντότητες Users και Events και η σχέση τους είναι N προς M, αφού κάθε χρήστης μπορεί να δηλώσει ενδιαφέρον παρακολούθησης για πολλά events και για κάθε event μπορεί να δηλωθεί ενδιαφέρον από πολλούς χρήστες.
RatingFeedback – Users - Events	Συνδέει τις οντότητες Users και Events και η σχέση τους είναι N προς M, αφού κάθε χρήστης μπορεί να βαθμολογήσει πολλά events που παρακολούθησε και κάθε event να βαθμολογηθεί από πολλούς χρήστες.
EventNotifications –	Συνδέει τις οντότητες Users και Events και η σχέση τους είναι N

Users - Events	προς Μ, αφού κάθε χρήστης μπορεί να λάβει πρόσκληση από πολλά events και κάθε event να αποστείλει μηνύματα σε πολλούς χρήστες.
-----------------------	--

Πίνακας 3-2: Περιγραφή Σχέσεων Οντοτήτων



Σχήμα 3-3: Σχεσιακό Μοντέλο της Βάσης Δεδομένων

4 ΑΝΑΛΥΣΗ ΣΥΣΤΗΜΑΤΟΣ

Όπως αναφέρθηκε προηγουμένως, βασικός σκοπός του συστήματος είναι να βρει χρήστες που μπορεί να ενδιαφέρονται για ένα συγκεκριμένο event. Προς αυτή την κατεύθυνση ακολουθήθηκε μία content-based προσέγγιση για τη δημιουργία μιας λίστας βαθμολόγησης χρηστών, ξεκινώντας από αυτούς που είναι πιθανότερο να παρακολουθήσουν ένα συγκεκριμένο event και καταλήγοντας σε αυτούς που έχουν λιγότερες πιθανότητες να το παρακολουθήσουν.

Η απαραίτητη πληροφορία εξάγεται από τις φόρμες του προφίλ των χρηστών, οι οποίες συμπληρώνονται κατά τη διάρκεια της εγγραφής τους στο σύστημα. Παράλληλα οι διοργανωτές που ενδιαφέρονται να προωθήσουν ένα event συμπληρώνουν φόρμες περιγράφοντας τα χαρακτηριστικά του event, αλλά και των χρηστών που θα ήθελαν να προσελκύσουν. Οι παράγοντες που λαμβάνονται υπόψη για τη βαθμολόγηση των χρηστών που πιθανώς να ενδιαφέρονται για ένα συγκεκριμένο event είναι οι προτιμήσεις τους, η ηλικιακή ομάδα που ανήκουν, το φύλο τους, αλλά και η τοποθεσία διαμονής τους. Η επιρροή που έχουν αυτοί οι παράγοντες στα αποτελέσματα που επιστρέφονται είναι διπλή. Πρώτα ελαχιστοποιούν τον αριθμό των χρηστών που θα επιλεγθούν (ηλικία, φύλο και τοποθεσία) και δεύτερον επηρεάζουν την θέση κατάταξής τους στον πίνακα βαθμολόγησης βάση των προτιμήσεων τους. Ένα τέτοιο παράδειγμα είναι η διαφήμιση ενός γυναικείου προϊόντος. Στην περίπτωση αυτή θα επιλεγθούν μόνο χρήστες γυναικείου φύλου.

Όλοι οι παραπάνω υπολογισμοί παράγουν ένα non-normalized βαθμό, ο οποίος είναι ακέραιος και περιγράφει την σχέση κάθε χρήστη με το event. Οι χρήστες με τη μεγαλύτερη βαθμολογία έχουν και τη μεγαλύτερη πιθανότητα να ενδιαφέρονται και να παρακολουθήσουν το συγκεκριμένο event.

4.1 Event και User Keys

Ένα από τα βασικά δομικά στοιχεία του συστήματος, πάνω στο οποίο στηρίζουν τη λειτουργία τους οι αλγόριθμοι, είναι τα Event και User Keys. Στην ουσία, πρόκειται για δομές που ορίστηκαν στα πλαίσια του συστήματος με βάση συγκεκριμένους κανόνες και περιέχουν συμπυκνωμένες, κωδικοποιημένες πληροφορίες για διάφορες ιδιότητες των εκδηλώσεων (event keys) ή των χρηστών (user keys) της υπηρεσίας. Οι πληροφορίες που κωδικοποιούνται

αφορούν δημογραφικά δεδομένα (φύλο, ηλικία, κτλ.), τις προτιμήσεις τους σε διάφορες κατηγορίες και γεωγραφικά δεδομένα (τόπος διαμονής τους).

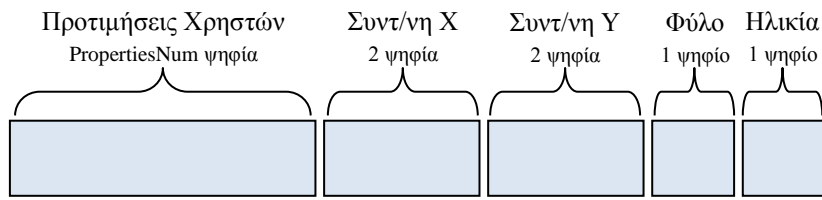
Τα keys δημιουργούνται κατά την καταχώριση ενός χρήστη ή event στο σύστημα και αποθηκεύονται σε ξεχωριστούς πίνακες της βάσης δεδομένων, μαζί με τα αντίστοιχα IDs, ώστε να γίνεται η αντιστοίχιση με την εγγραφή στην οποία αναφέρονται. Παράλληλα, ανανεώνονται σε περίπτωση τροποποίησης κάποιων από τα στοιχεία που κωδικοποιούν. Για παράδειγμα, αν ένας χρήστης μετακομίσει και ενημερώσει το προφίλ του, το user key που του αντιστοιχεί θα ενημερωθεί ανάλογα.

Όσον αφορά τη μορφή, τα κλειδιά αποτελούνται από μια ακολουθία αριθμητικών χαρακτήρων. Οι λόγοι που επιλέχθηκε η χρήση των κλειδιών είναι οι ακόλουθοι:

- Μείωση του αριθμού συναλλαγών με τη βάση, του αριθμού των queries που εκτελούνται και του fetching δεδομένων. Με τον τρόπο αυτό μειώνεται ο φόρτος εργασίας της βάσης δεδομένων.
- Αποθήκευση σύνθετων meta-πληροφοριών για τα δεδομένα. Η συγκεκριμένη τεχνική μπορεί να παίξει καθοριστικό ρόλο σε νέες, βελτιωμένες εκδόσεις των αλγόριθμων κατάταξης. Έτσι, σε μεταγενέστερες εκδόσεις, δίνεται η δυνατότητα εφαρμογής εξειδικευμένων αλγόριθμων πάνω στα δεδομένα χρηστών και events, με σκοπό την εξαγωγή επιπρόσθετων χρήσιμων συμπερασμάτων για τη βαθμολόγηση.
- Αποφυγή δαπανηρών Join άνω του ενός table της βάσης δεδομένων, αφού όλες οι απαραίτητες, για την εφαρμογή των αλγόριθμων, πληροφορίες βρίσκονται συγκεντρωμένες και κωδικοποιημένες σε ένα μόνο table για κάθε είδος κλειδιού.
- Προστασία προσωπικών δεδομένων. Στους διαχειριστές και τους αλγόριθμους του συστήματος αρκεί μόνο να έχουν δικαίωμα πρόσβασης στα keys, που όπως αναφέρθηκε αποθηκεύονται σε ξεχωριστά tables της βάσης, χωρίς να είναι απαραίτητο να τους δίνεται η δυνατότητα να προσπελάσουν πίνακες που περιέχουν ευαίσθητα προσωπικά δεδομένα των χρηστών.

4.1.1 User Key

Τα κλειδιά που αναπαριστούν τις πληροφορίες για τους χρήστες του συστήματος, έχουν την παρακάτω μορφή για κάθε χρήστη:



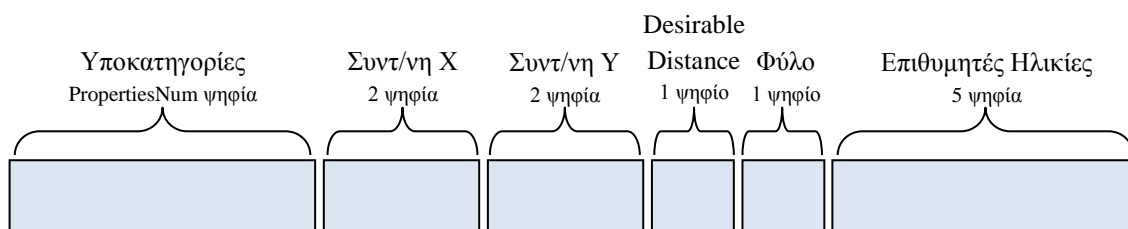
Σχήμα 4-1: Αναπαράσταση User Key

Όπως φαίνεται στο παραπάνω σχήμα, το User Key αποτελείται από τα ακόλουθα:

- Το πεδίο που αναπαριστά τις προτιμήσεις του χρήστη, αποτελείται από τόσα ψηφία-χαρακτήρες όσες και οι υποκατηγορίες των βασικών κατηγοριών του συστήματος (Τέχνη, Αθλητισμός, Επιστήμη, Ψυχαγωγία και Υπολογιστές). Κάθε χαρακτήρας αντιστοιχεί σε μία ιδιότητα και παίρνει τιμές από το 1 (καθόλου ενδιαφέρον) μέχρι το 5 (πάρα πολύ ενδιαφέρον), δηλώνοντας το πόσο ενδιαφέρει τον εν λόγω χρήστη η συγκεκριμένη κατηγορία.
- Οι γεωγραφικές συντεταγμένες, αναπαρίστανται με δύο πεδία των δύο χαρακτήρων που αναφέρονται στο γεωγραφικό μήκος και πλάτος της κατοικίας του χρήστη. Σε αυτά, απεικονίζονται με τη μορφή διψήφιων ακέραιων οι συντεταγμένες όπου τοποθετείται ο χρήστης στο πλέγμα του χάρτη που καλύπτει το σύστημα.
- Το φύλο του χρήστη, κωδικοποιείται με ένα χαρακτήρα, ο οποίος παίρνει την τιμή 1 για τους άντρες και 2 για τις γυναίκες.
- Η ηλικία, παριστάνεται με ένα ψηφίο που παίρνει τιμές από το 1 μέχρι το 5 και αντιστοιχεί σε ένα από τα πέντε ηλικιακά γκρουπ (18-24, 25-34, 35-49, 50-64, 65+) που έχουν οριστεί για να χαρακτηρίζουν τους χρήστες.

4.1.2 Event Key

Τα κλειδιά που αναπαριστούν τις πληροφορίες για τα *events* του συστήματος, έχουν την παρακάτω μορφή για κάθε *event*:



Σχήμα 4-2: Αναπαράσταση Event Key

Όπως φαίνεται στο παραπάνω σχήμα, το Event Key αποτελείται από τα ακόλουθα:

- Τόσα ψηφία αριθμού όσες και οι υποκατηγορίες των βασικών κατηγοριών του συστήματος. Ο συνολικός αριθμός και η διάταξη των ιδιοτήτων, ταυτίζεται με αυτά του User Key.
- Δύο πεδία των δύο χαρακτήρων που αναφέρονται στο γεωγραφικό μήκος και πλάτος του τόπου διεξαγωγής του *event*.
- Ένα χαρακτήρα που λαμβάνει τιμές από το 0 έως 2 και αντιπροσωπεύει την επιθυμητή απόσταση των χρηστών από το event (0 – Ανεξαρτήτως Περιοχής, 1 – Ίδια Περιοχή και 2 – Κοντινές Περιοχές).
- Ένα ψηφίο που λαμβάνει τις τιμές 0, 1 και 2 ανάλογα με το ποιο είναι το φύλο του κοινού στο οποίο απευθύνεται η εκδήλωση (0 - όλους, 1 - άντρες και 2 - γυναίκες).
- Ένα πεδίο των 5 χαρακτήρων, για την αναπαράσταση των ηλικιακών γκρουπ (18-24, 25-34, 35-49, 50-64, 65+) στα οποία στοχεύει η προώθηση του event. Κάθε χαρακτήρας, αντιπροσωπεύει ένα από τα πέντε ηλικιακά γκρουπ και αν η τιμή του είναι ίση με 1 το συγκεκριμένο age group είναι επιθυμητό, ενώ αν ισούται με 0 δεν είναι.

4.2 Παράγοντες που Επηρεάζουν την Επιλογή Χρηστών

Βασικός σκοπός του συστήματος είναι η εύρεση και η εμφάνιση των καταλληλότερων χρηστών για ένα event. Συγκεκριμένα, το σύστημα πρέπει να αποφασίσει ποιοι από τους χρήστες είναι πιο πιθανό να ενδιαφερθούν για ένα συγκεκριμένο event και να το παρακολουθήσουν. Ο αλγόριθμος που δημιουργήθηκε είναι ένας ταξινομητής που ακολουθεί μια content-based προσέγγιση για την κατηγοριοποίηση των χρηστών, ανάλογα με το πόσο πιθανό κρίνεται να παρακολουθήσουν το event.

Τα δεδομένα που εισάγονται στο σύστημα λαμβάνονται από τις φόρμες δημιουργίας προφίλ που συμπληρώνουν οι χρήστες κατά την εγγραφή τους, καθώς επίσης και από τους διοργανωτές που επιθυμούν να προωθήσουν ένα event, οι οποίοι συμπληρώνουν μια αντίστοιχη φόρμα εισάγοντας τα χαρακτηριστικά του, καθώς και τις προϋποθέσεις που πρέπει να πληρούν οι χρήστες στους οποίους απευθύνεται.

Αρχικά, οι χρήστες πρέπει να φιλτραριστούν προτού εισαχθούν στον ταξινομητή, βάσει των προϋποθέσεων του εκάστοτε event. Οι προϋποθέσεις αυτές αφορούν το φύλο, την ηλικία και την τοποθεσία των χρηστών και βοηθούν στο να περιοριστεί το σύνολο τους.

4.2.1 Φύλο

Στη φόρμα εισαγωγής των events, δίνεται η δυνατότητα στους διοργανωτές να καθορίσουν το επιθυμητό φύλο των χρηστών στους οποίους απευθύνεται το event. Η πληροφορία αυτή αποθηκεύεται στο event key με τη μορφή ενός χαρακτήρα, ο οποίος λαμβάνει την τιμή: α) 0 αν το event απευθύνεται σε όλους τους χρήστες ανεξαρτήτως φύλου, β) 1 αν απευθύνεται μόνο σε άντρες και γ) 2 αν απευθύνεται μόνο σε γυναίκες. Κατά το πρώτο στάδιο φιλτραρίσματος, ελέγχεται αυτό το ψηφίο του event key και έτσι επιλέγονται οι ανάλογοι χρήστες.

4.2.2 Ηλικία

Το δεύτερο στάδιο φιλτραρίσματος αφορά την ηλικία του χρήστη. Κάθε χρήστης εντάσσεται σε μία από τις 5 ηλικιακές ομάδες. Η ηλικιακή ομάδα υπολογίζεται κατά τη δημιουργία του αντίστοιχου user key και αντιστοιχεί σε ένα ψηφίο του, που λαμβάνει τιμές από το 1 μέχρι το 5. Τα ηλικιακά εύρη που απαρτίζουν κάθε ομάδα, ορίστηκαν βάσει των κανόνων του σύγχρονου μάρκετινγκ, όπως αποτυπώνονται στις περισσότερες έρευνες αγοράς. Ο διαχωρισμός σε πέντε ηλικιακές ομάδες έγινε με βάση την βαθμολόγηση MOS (Mean Opinion Score) που είναι ευρέως γνωστή στον τομέα της μηχανικής και της επιστήμης των ηλεκτρονικών υπολογιστών. Η κωδικοποίηση των προαναφερθέντων ηλικιακών ομάδων στα πλαίσια του συστήματος, είναι: α) 1 για τις ηλικίες 18-24, β) 2 για τις ηλικίες 25-34, γ) 3 για τις ηλικίες 35-49, δ) 4 για τις ηλικίες 50-64 και ε) 5 για τις ηλικίες από 65 έτη και πάνω.

Όσον αφορά τα event keys, ο τρόπος αναπαράστασης των ηλικιακών γκρουπ στα οποία απευθύνονται, έχει εντελώς διαφορετική λογική. Είναι προφανές ότι δεν μπορεί να ακολουθηθεί η προηγούμενη προσέγγιση, από τη στιγμή που ένα event ενδέχεται να απευθύνεται σε περισσότερες από μία ηλικιακές ομάδες. Για το λόγο αυτό, χρησιμοποιούνται 5 ψηφία στο event key. Κάθε ένα από τα 5 ψηφία αντιστοιχεί σε μία από τις 5 ηλικιακές ομάδες, με το πρώτο κατά σειρά ψηφίο να αντιστοιχεί στις ηλικίες 18-24 και το πέμπτο στις ηλικίες άνω των 65 ετών. Ο διοργανωτής, επιλέγει κατά τη συμπλήρωση της φόρμας εισαγωγής event τις ηλικιακές ομάδες στις οποίες απευθύνεται. Κατά τη δημιουργία του event key κάθε ένα από τα 5 ψηφία που χρησιμοποιούνται για τις ηλικιακές ομάδες παίρνουν την τιμή 1, αν η αντίστοιχη ηλικιακή ομάδα είναι επιθυμητή ή την τιμή 0 αν δεν είναι.

Σε αυτό το στάδιο του φιλτραρίσματος ελέγχονται τα 5 ψηφία event key που αφορούν την ηλικιακές ομάδες και ανάλογα με το ποιες είναι επιθυμητές επιλέγονται οι κατάλληλοι χρήστες μέσα από το σύνολο αυτών που προέκυψαν από το πρώτο στάδιο φιλτραρίσματος.

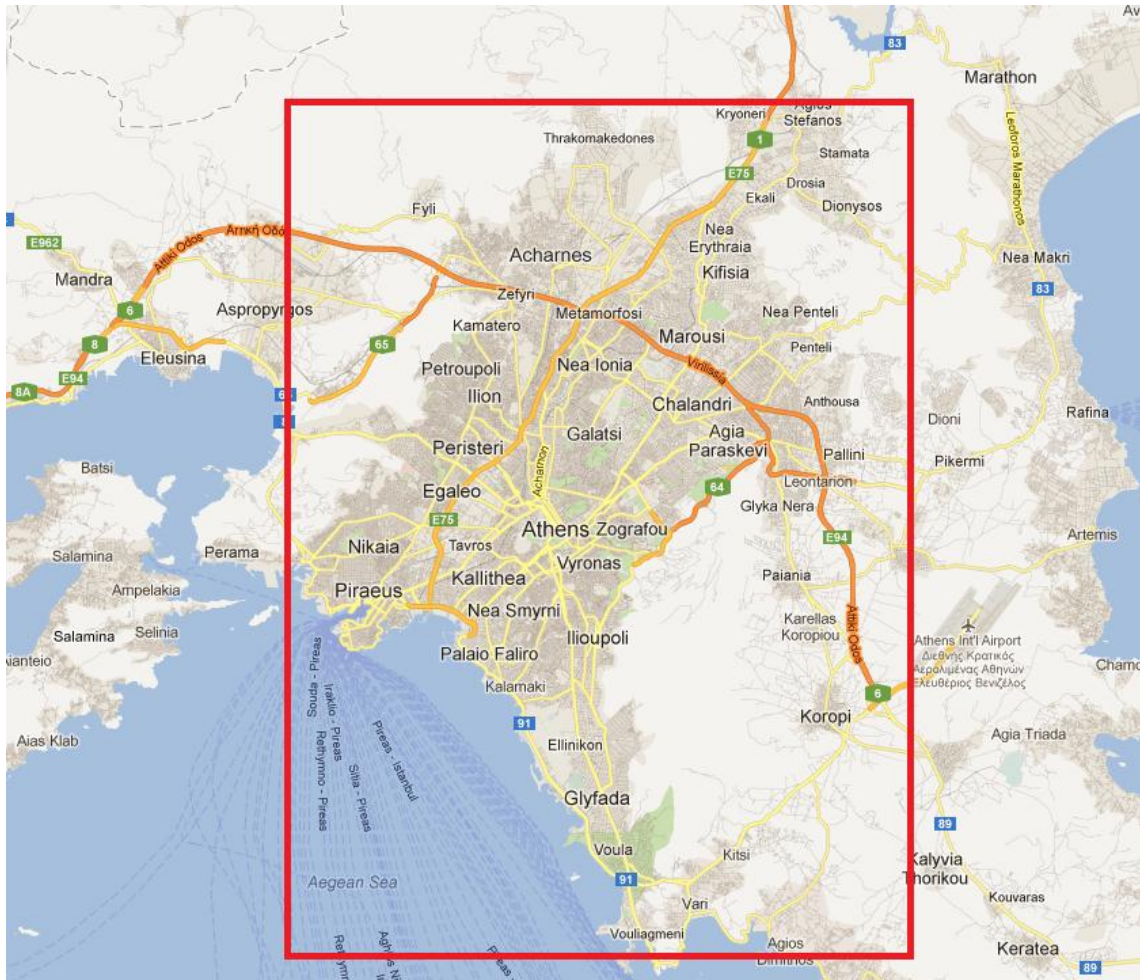
4.2.3 Τοποθεσία

Το τρίτο στάδιο φιλτραρίσματος αφορά την τοποθεσία των χρηστών σε σχέση με το event. Ο διοργανωτής του event, καθορίζει για το αν αυτό απευθύνεται προς όλους τους χρήστες ανεξάρτητα από την τοποθεσία τους ή μόνο σε χρήστες που βρίσκονται κοντά στην τοποθεσία του event. Για παράδειγμα, αν ένα event είναι ένας ποδοσφαιρικός αγώνας, τότε θα απευθύνεται προς όλους τους χρήστες ανεξαρτήτως τοποθεσίας. Ενώ, αν αφορά μία εορταστική εκδήλωση κάποιου δήμου, τότε θα απευθύνεται μόνο στους χρήστες που διαμένουν σε αυτή την περιοχή. Προκειμένου να υπολογιστεί αυτή η απόσταση, πρέπει να εξαχθούν οι γεωγραφικές συντεταγμένες της διεύθυνσης κάθε χρήστη, καθώς και της τοποθεσίας του εκάστοτε event. Για το σκοπό αυτό, χρησιμοποιήθηκε η τεχνολογία των Google Maps, μέσω του Google Maps API. Το γεωγραφικό μήκος και πλάτος, αποθηκεύονται με τη μορφή δεκαδικών αριθμών που εκφράζουν τις μοίρες.

Προτού η πληροφορία για τη γεωγραφική πληροφορία των users και των events αποθηκευτεί στα αντίστοιχα κλειδιά πρέπει να επιλυθούν τα εξής:

- Οι γεωγραφικές συντεταγμένες αποθηκεύονται με τη μορφή δεκαδικού αριθμού, ενώ οι δομές UserKey και EventKey πρέπει να αποτελούνται από μια ακολουθία ψηφίων που θα περιέχουν κωδικοποιημένες πληροφορίες για τους χρήστες και τα events αντίστοιχα.
- Η αντιστοιχία μίας μοίρας γεωγραφικού μήκους σε χιλιομετρική απόσταση, δεν είναι ίδια με αυτήν της μίας μοίρας γεωγραφικού πλάτους. Επομένως, για τον υπολογισμό της απόστασης μεταξύ δύο σημείων, χρειάζεται μετατροπή της γωνιακής απόκλισής τους σε μονάδα μέτρησης μήκους. Επιπλέον, η μετατροπή αυτή για το γεωγραφικό μήκος διαφέρει καθώς απομακρυνόμαστε από τον Ισημερινό, αφού η διάμετρος των παράλληλων μικραίνει.

Για την αντιμετώπιση αυτών των προβλημάτων θεωρείται ένα παραλληλόγραμμο πλέγμα από ίσα τετράγωνα πάνω από την περιοχή που καλύπτει το σύστημα, που στην τρέχουσα έκδοση είναι το λεκανοπέδιο Αττικής και κάποια επιπλέον προάστια. Το πλέγμα αυτό φαίνεται παρακάτω:



Σχήμα 4-3: Όρια Γεωγραφικού Πλέγματος

Τα όρια αυτού του πλέγματος προκαθορίστηκαν ως εξής:

- Γεωγραφικό μήκος: 23,60 έως 23,90 μοίρες δυτικά
- Γεωγραφικό πλάτος: 37,80 έως 38,10 μοίρες βόρεια

Κάθε ένα από τα τετράγωνα του πλέγματος χαρακτηρίζεται από δύο αριθμούς (συντεταγμένες). Αν θεωρήσουμε σύστημα αξόνων με αρχή την κάτω αριστερά γωνία του πλέγματος, οι αριθμοί αυτοί δείχνουν την απόσταση του τετραγώνου στους άξονες x και y. Κάθε διεύθυνση ανήκει στο εσωτερικό ενός από τα τετράγωνα του πλέγματος.

Για την πραγματοποίηση των υπολογισμών, έπρεπε στη συνέχεια να υπολογιστεί το μήκος σε χιλιόμετρα μεταξύ των ορίων του πλέγματος. Για το σκοπό αυτό χρησιμοποιήθηκε το εργαλείο μετατροπής μοιρών σε μήκος, που βρίσκεται στην ιστοσελίδα του National Geospatial Intelligence Agency. Το εργαλείο αυτό, επιτρέπει τον υπολογισμό της χιλιομετρικής απόστασης μίας μοίρας γεωγραφικού μήκους και μίας μοίρας γεωγραφικού πλάτους, σε δεδομένο γεωγραφικό πλάτος πάνω στον πλανήτη, αφού όπως έχει ήδη αναφερθεί οι αποστάσεις αυτές μεταβάλλονται καθώς απομακρυνόμαστε από τον Ισημερινό.

Υπολογίστηκε πως στο γεωγραφικό πλάτος της Αθήνας η απόκλιση 0,3 μοιρών γεωγραφικού πλάτους αντιστοιχεί σε 33298 μέτρα, ενώ η απόκλιση 0,3 μοιρών γεωγραφικού μήκους αντιστοιχεί σε 26367 μέτρα. Επομένως αυτό είναι και το μήκος των πλευρών του πλέγματος. Το μήκος πλευράς που επιλέχτηκε για τα τετράγωνα του πλέγματος είναι 500 μέτρα και ομοίως υπολογίστηκε ότι αντιστοιχεί σε 0,0045 μοίρες γεωγραφικού πλάτους και 0,0057 μοίρες γεωγραφικού μήκους. Αυτό το νούμερο θεωρήθηκε επαρκές ώστε να ληφθεί ως δεδομένο ότι η απόσταση των χρηστών, που κατοικούν εντός ενός τετραγώνου, από κάποιο *event* επηρεάζει ουσιαστικά με τον ίδιο τρόπο την απόφασή τους για το αν θα παρευρεθούν σε αυτό ή όχι, επομένως και ο τρόπος που η απόσταση επηρεάζει τη βαθμολόγησή τους στο σύστημα θα πρέπει να είναι κοινός. Η κωδικοποίηση των διευθύνσεων στα κλειδιά γίνεται με την ακόλουθη διαδικασία:

- Αφαιρείται από το γεωγραφικό πλάτος κάθε διεύθυνσης η τιμή 37,80 μοίρες και από το γεωγραφικό μήκος η τιμή 23,60 μοίρες.
- Η διαφορά στο γεωγραφικό πλάτος διαιρείται με 0,0045 ώστε να υπολογιστεί η θέση του τετραγώνου στο οποίο ανήκει η διεύθυνση στον άξονα των y . Το αποτέλεσμα στρογγυλοποιείται προς τα πάνω.
- Αντίστοιχα η διαφορά στο γεωγραφικό μήκος διαιρείται με 0,0057 για να υπολογιστεί η θέση του τετραγώνου στον άξονα των x . Το αποτέλεσμα και πάλι στρογγυλοποιείται προς τα πάνω.
- Οι δύο αριθμοί που προκύπτουν αποθηκεύονται στις αντίστοιχες θέσεις στην ακολουθία ψηφίων που αποτελεί το UserKey ή το EventKey.

Κατά τον υπολογισμό της βαθμολογίας ενός χρήστη σε σχέση με κάποιο *event*, τα κλειδιά αποκωδικοποιούνται και οι αριθμοί αυτοί χρησιμοποιούνται για την εύρεση της απόστασης μεταξύ *user* και *event*, σύμφωνα με την Εξίσωση 1.

$$\text{Γεωγραφική Απόσταση (E,U)} = (x_e - x_u)^2 + (y_e - y_u)^2 \quad \mathbf{1}$$

Όπου x_e, y_e οι συντεταγμένες του τετραγώνου του *event* και x_u, y_u οι συντεταγμένες του τετραγώνου του *user*.

Σε αυτό το τελευταίο στάδιο του φιλτραρίσματος επιλέγονται οι κατάλληλοι χρήστες μέσα από το σύνολο αυτών που προέκυψαν κατά το φιλτράρισμα του δεύτερου σταδίου και έτσι με το σύνολο χρηστών που απομένει τροφοδοτείται ο ταξινομητής.

4.3 Αλγόριθμοι Συστήματος

Όπως αναφέρθηκε και στην εισαγωγή, η τεχνική προτάσεων που χρησιμοποιήθηκε είναι ένας ταξινομητής, ο οποίος στοχεύει στην αναγνώριση και στην κατηγοριοποίηση διαφόρων *events* που περιέχουν γεωγραφική πληροφορία βάση των προτιμήσεων των χρηστών, παρέχοντάς τους παράλληλα αξιόπιστη πληροφορία για διάφορα *events* που μπορεί να τους ενδιαφέρουν.

4.3.1 Ταξινόμηση και Κατηγορίες

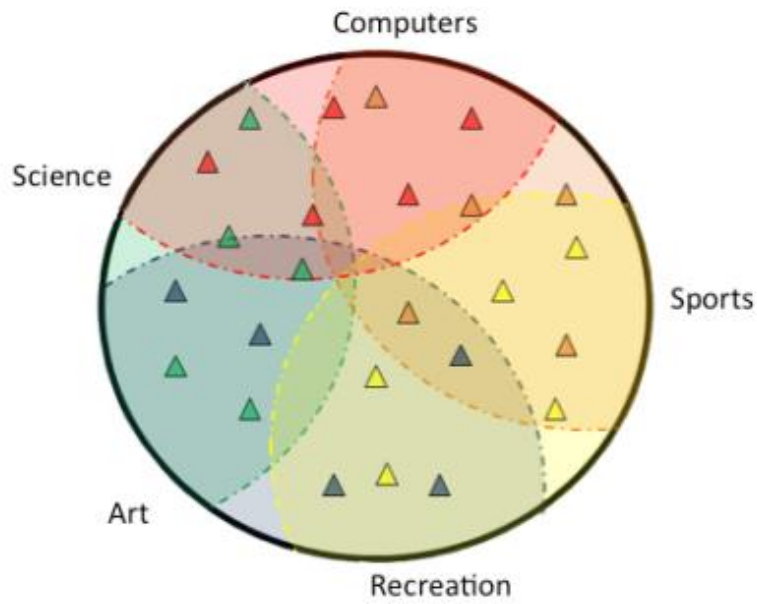
Ο ταξινομητής εκπαιδεύτηκε βάση μιας ταξινόμησης που προέρχεται από το Open Directory Project (DMOZ, <http://www.dmoz.org/>). Η διαδικασία της ταξινόμησης, εκτελείται εκτιμώντας την πιθανότητα ενός διανύσματος εισόδου σύμφωνα με τις πιθανότητες του Bayes. Το διάνυσμα αποτελείται από 25 βάρη που ανταποκρίνονται στις τιμές των προτιμήσεων του χρήστη βάση των 5 κατηγοριών (A, B, C, D και E), με κάθε μία από αυτές να έχει 5 υποκατηγορίες όπως φαίνονται στον Πίνακα 4-1. Από εδώ και πέρα αυτό το διάνυσμα θα ονομάζεται Preference Event Vector (PEV, Διάνυσμα Προτιμήσεων Event). Αυτό το διάνυσμα προσδιορίζει ένα μοναδικό προφίλ σε κάθε προσομοιωμένη περίπτωση *event*. Για τη δημιουργία αυτού του διανύσματος, δημιουργήθηκε ένας μεγάλος αριθμός προσομοιωμένων περιπτώσεων προτιμήσεων χρηστών που ήταν προσανατολισμένες προς τις κατηγορίες που επιλέχθηκαν από το DMOZ. Στη συνέχεια, ομαδοποιήθηκαν όλες οι προσομοιωμένες περιπτώσεις χρησιμοποιώντας ένα Gaussian μοντέλο ομαδοποίησης, προκειμένου να δημιουργηθεί ένα ισχυρό σετ εκπαίδευσης για τον ταξινομητή. Η μέθοδος ομαδοποίησης που χρησιμοποιήθηκε, υλοποιήθηκε με το γνωστό εργαλείο WEKA (www.cs.waikato.ac.nz/ml/weka/).

Κατηγορίες / Υποκατηγορίες DMOZ				
ART (A)	SPORTS (B)	SCIENCE (C)	RECREATION (D)	COMPUTERS (E)
Architecture (a1)	News (b1)	Technology (c1)	Tourism (d1)	Internet (e1)
Cinema (a2)	Football (b2)	Physics (c2)	Sea (d2)	Software (e2)
Museum (a3)	Basketball (b3)	Mathematics (c3)	Entertainment (d3)	Hardware (e3)
Music (a4)	Organizations (b4)	Biology (c4)	Car (d4)	Education (e4)
Theater (a5)	Volleyball (b5)	Medicine (c5)	Motorcycle (d5)	Companies (e5)

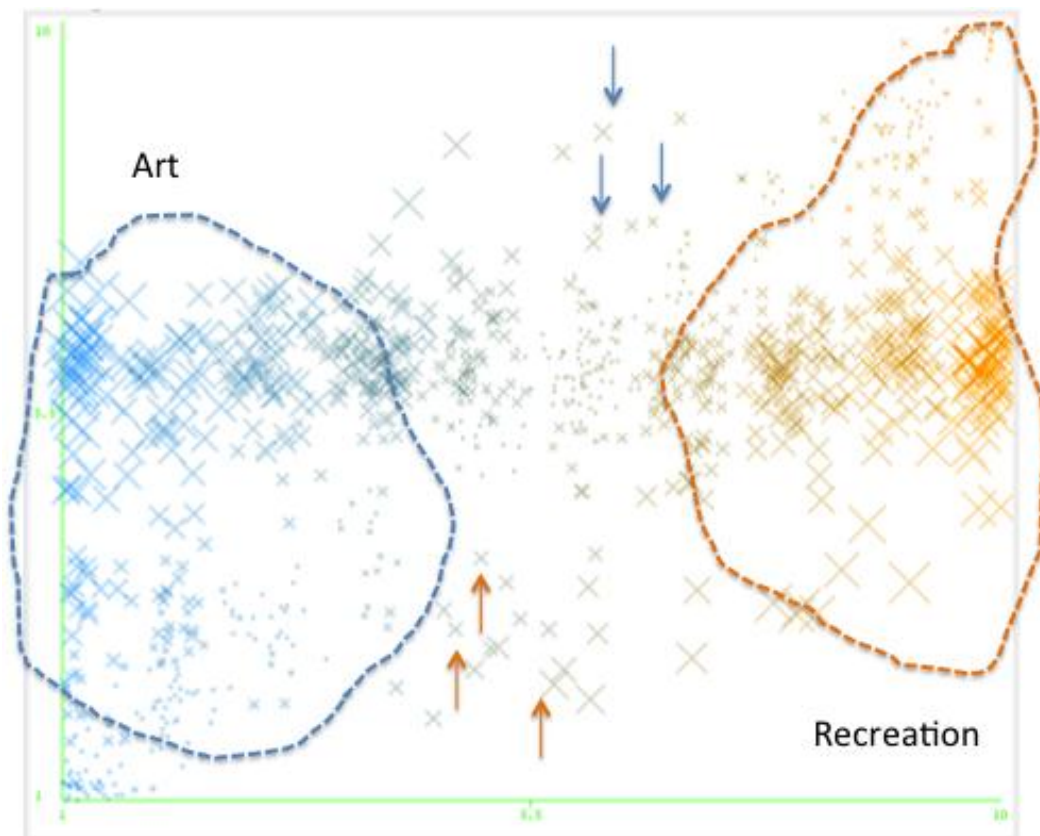
Πίνακας 4-1: Κατηγορίες και Υποκατηγορίες DMOZ

4.3.2 Αλγόριθμος Ομαδοποίησης

Στο μοντέλο ομαδοποίησης Gaussian, τα δεδομένα θεωρούνται ως ξεχωριστές Gaussian κατανομές, όπου κάθε διαφορετική ομάδα, αντιπροσωπεύει κάθε κατανομή. Το *Σχήμα 4-4* απεικονίζει το θεωρητικό μοντέλο αναπαράστασης και για τις πέντε κατηγορίες, καθώς και τις ομάδες για τις κατηγορίες Art και Recreation όπως αυτές προκύπτουν από το Gaussian μοντέλο που εφαρμόστηκε. Επιπλέον, οι τονισμένες περιοχές στο *Σχήμα 4-4α* δείχνουν τις πέντε επιλεγμένες κατηγορίες μαζί με τα πέντε κεντροειδή τους, που αντιστοιχούν στις πέντε υποκατηγορίες για κάθε κατηγορία. Είναι προφανές πως κάποια κεντροειδή μπορεί να ανήκουν σε περισσότερες από μία κατηγορίες, καθώς η θεματική κατηγοριοποίηση των εν λόγω πέντε κατηγοριών αποτελείται από σημασιολογικές επικαλύψεις και σίγουρα δεν θεωρείται ρητός και σαφής ο διαχωρισμός μεταξύ των events διαφορετικών κατηγοριών. Για παράδειγμα, ένα event που ανήκει στην υποκατηγορία Music (a4), μπορεί σίγουρα να επισημανθεί ως ένα event Recreation/Entertainment (υποκατηγορία d3). Τα βέλη στο *Σχήμα 4-4b* επισημαίνουν κάποια κεντροειδή από την περιοχή της ομάδας Recreation (D) που βρίσκονται κοντά στην περιοχή της ομάδας Art (A) και το αντίστροφο.



(a)



(b)

Σχήμα 4-4: Προσομοιωμένες Περιπτώσεις Event: (a) Θεωρητικό Μοντέλο Αναπαράστασης για όλες τις Κατηγορίες, (b) Gaussian Μοντέλο Ομαδοποίησης για τα Κεντροειδή Art - Recreation

4.3.3 Προτεινόμενος Ταξινομητής Νευρωνικού Δικτύου

Επιλέχθηκε να χρησιμοποιηθεί ένα Νευρωνικό Δίκτυο Πιθανοτήτων (PNN, Probabilistic Neural Network) ως ταξινομητής μιας και αυτό το είδος των τεχνητών νευρωνικών δικτύων συνδυάζουν μερικά από τα καλύτερα χαρακτηριστικά της στατιστικής αναγνώρισης και των feed-forward νευρωνικών δικτύων. Τα PNNs είναι η υλοποίηση νευρωνικού δικτύου της διακρίνουσας ανάλυσης πυρήνα και εισήχθησαν στη λογοτεχνία των νευρωνικών δικτύων από τον Specht (1990). Τα PNNs χαρακτηρίζονται για του πολύ γρήγορους χρόνους εκπαίδευσης και παράγουν εξόδους με πιθανότητες Bayes (Bishop, 1997). Ένα PNN χρησιμοποιεί ένα εποπτευόμενο σετ εκπαίδευσης για να αναπτύξει λειτουργίες διανομής μέσα σε ένα μοτίβο. Στη συνέχεια, αυτές οι λειτουργίες χρησιμοποιούνται για την εκτίμηση της πιθανότητας, ένα διάνυσμα εισόδου να είναι μέρος μιας κατηγορίας. Αυτά τα μοτίβα μάθησης, μπορούν επίσης να συνδυαστούν ή να σταθμιστούν με μία priori πιθανότητα που ονομάζεται σχετική συχνότητα κάθε κατηγορίας προκειμένου να προσδιοριστεί η πιο πιθανή κατηγορία για ένα δεδομένο διάνυσμα εισόδου. Εάν η σχετική συχνότητα των κατηγοριών είναι άγνωστη, τότε όλες οι κατηγορίες μπορεί να θεωρηθούν ότι είναι εξίσου πιθανές και ο προσδιορισμός της κατηγορίας βασίζεται αποκλειστικά και μόνο από το πόσο κοντά είναι το διάνυσμα εισόδου στην συνάρτηση κατανομής μιας κατηγορίας. Τα PNNs περιέχουν ένα στρώμα εισόδου με τόσα στοιχεία, όσες και οι ξεχωριστές παράμετροι που απαιτούνται για να περιγραφούν τα αντικείμενα της κατηγοριοποίησης (25 όσα και οι υποκατηγορίες), καθώς και ένα μεσαίο στρώμα το οποίο οργανώνει το σετ εκπαίδευσης έτσι ώστε ένα μεμονωμένο στοιχείο επεξεργασίας να αντιπροσωπεύει κάθε διάνυσμα εισόδου. Τέλος, έχουν ένα στρώμα εξόδου που ονομάζεται επίσης και στρώμα άθροισης, το οποίο έχει τόσα στοιχεία επεξεργασίας όσες και οι κατηγορίες που πρέπει να αναγνωριστούν (5 σε αυτή την περίπτωση).

4.3.4 Θεωρητικό Υπόβαθρο και Αντιμετώπιση του Προβλήματος

Στο πρόβλημα που προκύπτει, η ταξινόμηση μπορεί να δηλωθεί σαν δειγματοληψία ενός πολυμεταβλητού τυχαίου διανύσματος s -συνιστωσών $x = [x_1, x_2, \dots, x_s]$, όπου τα δείγματα είναι τοποθετημένα κατά u , $u = 1, \dots, U$ (Masters, 1995). Γνωρίζοντας τις συναρτήσεις πυκνότητας πιθανότητας για όλους τους πληθυσμούς διανυσμάτων, οι αποφάσεις ταξινόμησης $d_m(X)$ γίνονται κατά συνέπεια σύμφωνα την *Εξίσωση 2*, η οποία καθορίζει τον κανόνα της βέλτιστης απόφασης του Bayes, όπου το h_m συμβολίζει την πιθανότητα ότι ένα δείγμα πρέπει να εξαχθεί από τον πληθυσμό m , το c_m που συμβολίζει το

κόστος της λάθος ταξινόμησης ενός δείγματος και το $f_m(X)$ συμβολίζει τη λειτουργία απόφασης που χρειάζεται να εξεταστεί.

$$d_m(X) = h_m c_m f_m(X) \quad 2$$

αν $h_m c_m f_m(X) > h_n c_n f_n(X)$ για όλους τους πληθυσμούς όπου $m \neq n$.

Η τοπολογία του προτεινόμενου PNN είναι 25-2500-5. Το επίπεδο εισόδου αποτελείται από 25 κόμβους, που αντιστοιχούν στον αριθμό των προσομοιωμένων προτιμήσεων χρήστη. Το δεύτερο επίπεδο είναι το μεσαίο επίπεδο, το οποίο οργανώνει το σετ εκπαίδευσης με τέτοιο τρόπο ώστε ένα ανεξάρτητο στοιχείο επεξεργασίας να αντιπροσωπεύει κάθε normalized διάνυσμα εισόδου. Ως εκ τούτου, αποτελείται από 2500 κόμβους που αντιστοιχούν στο συνολικό ποσό των μοτίβων εκπαίδευσης που χρησιμοποιούνται. Τέλος, το δίκτυο έχει ένα επίπεδο εξόδου που αποτελείται από 5 κόμβους (A, B, C, D και E), που αντιπροσωπεύουν τις 5 κατηγορίες που πρέπει να αναγνωριστούν.

Ένας συνειδητός πλήρης ανταγωνιστικός μηχανισμός μάθησης ανάμεσα στα βάρη του στρώματος εισόδου και του μεσαίου στρώματος παρακολουθεί πόσο συχνά οι έξοδοι κερδίζουν τον ανταγωνισμό με μία εικόνα εξισορρόπησης των κερδών, δημιουργώντας ένα επιπλέον στρώμα ανταγωνισμού μεταξύ των στοιχείων για να καθορίσει πιο στοιχείο επεξεργασίας θα ενημερωθεί. Έτσι, αν υποθέσουμε ότι το O είναι ο αριθμός των εξόδων, η λειτουργία ενημέρωσης του βάρους για τη νικήτρια έξοδο καθορίζεται από την *Εξίσωση 3*, όπου y_i είναι το i^{th} διάνυσμα εξόδου που μετράει την απόσταση μεταξύ των διανυσμάτων βάρους των νευρώνων εισόδου και εξόδου, x_j είναι το j^{th} διάνυσμα εισόδου, iw_{ij} είναι το βάρος σύνδεσης που συνδέει το στοιχείο επεξεργασίας j με το στοιχείο επεξεργασίας i , το f_i αντιστοιχεί στη συχνότητα κέρδους της εξόδου όπου $0 \leq f_i \leq \frac{1}{O}$ και το b_i καθορίζει το αντίστοιχο διάνυσμα προκατάληψης (bias) που δημιουργείται από το μηχανισμό συνείδησης.

$$U(y) = \max_i (y_i + b_i) = \max_i \left(\sqrt{\sum_{j=1}^{25} (x_j - iw_{ij})^2} + b_i \right) \quad 3$$

$i=1, 2, \dots, 2500$

όπου $b_i = \gamma \cdot [O \cdot (\beta \cdot (1 - f_i))]$, i : νικήτης

$b_i = \gamma \cdot [O \cdot (\beta \cdot f_i)]$, i : διαφορετικά

$$\beta = 0.015, \gamma = 0.25, O = 2500$$

Λαμβάνοντας υπόψη ότι μεταξύ των στρωμάτων η ενεργοποίηση των συνάψεων δίνεται από την Ευκλείδεια μετρική απόσταση και η λειτουργία, η οποία μιμείται τις συναπτικές

διασυνδέσεις του νευρώνα καθορίζεται από $m(x_j(t), iw_{ij}) = \sqrt{\sum_j (x_j(t) - iw_{ij})^2}$, το κόστος

της συνάρτησης δίνεται από $J(t) = \frac{1}{2} \sum_j (d_j(t) - m(x_j(t), iw_{ij}))^2$, όπου d_j είναι κάθε

επιθυμητή απάντηση κατά την διάρκεια εκπαίδευσης. Για την βελτιστοποίηση της

συνάρτησης κόστους, $\frac{\partial J(t)}{\partial m(x_j(t), iw_{ij})} = 0$ πρέπει να ικανοποιείται.

Ο συνολικός αριθμός των συνάψεων ανάμεσα στο στρώμα εισόδου και το μεσαίο στρώμα είναι 25 x 2500, ενώ ανάμεσα στο μεσαίο στρώμα και στο στρώμα εξόδου είναι 2500 x 5.

Παράλληλα, το προτεινόμενο PNN χρησιμοποιεί ένα επιβλεπόμενο σετ εκπαίδευσης για να αναπτύξει λειτουργίες διανομής εντός του μεσαίου στρώματος. Αυτές οι λειτουργίες χρησιμοποιούνται για να εκτιμήσουν την πιθανότητα το PEV εισόδου να είναι μέρος μιας περίπτωσης εκμάθησης event. Το μεσαίο στρώμα αντιπροσωπεύει μία νευρωνική εφαρμογή ενός ταξινομητή Bayes, όπου οι συναρτήσεις πυκνότητας πιθανότητας εξαρτώμενες από την κατηγορία προσεγγίζονται, χρησιμοποιώντας τον εκτιμητή Parzen, ο οποίος εκφράζεται

γενικά από $\frac{1}{n\sigma} \sum_{i=0}^{n-1} W\left(\frac{x - x_i}{\sigma}\right)$, όπου n είναι το μέγεθος του δείγματος, x και x_i είναι η

είσοδος και τα σημεία δειγματοληψίας, σ είναι η παράμετρος κλίμακας που ελέγχει το πλάτος της περιοχής εξετάζοντας την επιρροή της αντίστοιχης απόστασης και W είναι η συνάρτηση βάρους (Parzen, 1962). Η προσέγγιση αυτή παρέχει ένα βέλτιστο ταξινομητή μοτίβου όσον αφορά την ελαχιστοποίηση του αναμενόμενου κινδύνου της λάθος ταξινόμησης ενός αντικειμένου. Με τον εκτιμητή, η προσέγγιση πλησιάζει κοντά στις πραγματικές συναρτήσεις πυκνότητας κατηγορίας, καθώς ο αριθμός των δειγμάτων εκπαίδευσης αυξάνει, με την προϋπόθεση ότι το σετ εκπαίδευσης εκπροσωπεί επαρκώς τις διακρίσεις των κατηγοριών. Η πιθανότητα ενός αγνώστου PEV να ανήκει σε κάποια από τις κατηγορίες υπολογίζεται σύμφωνα με την *Εξίσωση 4*.

$$g_{i(WPV)} = \frac{1}{(2\pi)^{\frac{p}{2}} \sigma^p N_i} \sum_{j=0}^{(N_i-1)} e^{\frac{-(PEV - \bar{x}_{ij})^T (PEV - \bar{x}_{ij})}{2\sigma^2}} \quad 4$$

Στην παραπάνω εξίσωση, το i εκφράζει τον αριθμό της κατηγορίας, το j είναι μονάδα του μεσαίου στρώματος, το \bar{x}_{ij} αντιστοιχεί στο j^{th} διάνυσμα εκπαίδευσης της κατηγορίας i , το PEV είναι το διάνυσμα που δοκιμάζεται, το N_i αντιπροσωπεύει τα αντίστοιχα εκπαιδευτικά διανύσματα για την i^{th} κατηγορία, το p ισούται με τη διάσταση του PEV ($p=25$), το σ είναι η τυπική απόκλιση και το $(2\sigma)^{-2}$ σκιαγραφεί τον συντελεστή (β). Με άλλα λόγια η *Εξίσωση 4* ορίζει το άθροισμα των σφαιρικών Gaussian συναρτήσεων πολλαπλών μεταβλητών που επικεντρώνονται σε κάθε ένα από τα διανύσματα εκπαίδευσης \bar{x}_{ij} για την εκτίμηση της συνάρτησης πυκνότητας πιθανότητας της i^{th} κατηγορίας.

Επιπλέον, στο μεσαίο στρώμα, υπάρχει ένα στοιχείο επεξεργασίας για κάθε διάνυσμα εισόδου από το σετ εκπαίδευσης και ίσα ποσά στοιχείων επεξεργασίας για κάθε κατηγορία εξόδου, προκειμένου να αποφευχθεί να λοξεύσουν λανθασμένα μία ή περισσότερες κατηγορίες. Κάθε στοιχείο επεξεργασίας σε αυτό το επίπεδο έχει εκπαιδευτεί να δημιουργεί μία υψηλή τιμή εξόδου όταν ένα PEV εισόδου ταιριάζει με το διάνυσμα εκπαίδευσης. Ωστόσο τα διανύσματα εκπαίδευσης δεν πρέπει να έχουν κάποια ειδική διάταξη στο σετ εκπαίδευσης, δεδομένου ότι η κατηγορία ενός συγκεκριμένου διανύσματος καθορίζεται από την επιθυμητή έξοδο. Η λειτουργία εκμάθησης επιλέγει απλά το πρώτο ανεκπαιδευτο στοιχείο επεξεργασίας στη σωστή κατηγορία εξόδου και τροποποιεί τα βάρη του ώστε να ταιριάζουν με το διάνυσμα εκπαίδευσης. Το μεσαίο στρώμα λειτουργεί ανταγωνιστικά, έτσι ώστε να επικρατεί μόνο το υψηλότερο ταίριασμα σε ένα PEV εισόδου και να παράγει μία έξοδο.

4.3.5 Εκπαιδώντας τον ταξινομητή

Σε γενικές γραμμές, η διαδικασία εκπαίδευσης προσάπτει ένα ερέθισμα στο νευρωνικό δίκτυο και τελικά παράγει το επιθυμητό αποτέλεσμα. Η εκπαίδευσης είναι μία συνεχής διαδικασία ταξινόμησης των ερεθισμάτων εισόδου. Όταν ένα ερέθισμα εμφανίζεται στο δίκτυο, τότε το δίκτυο είτε το αναγνωρίζει είτε δημιουργεί μία νέα ταξινόμηση. Όταν η πραγματική απάντηση εξόδου είναι η ίδια με την επιθυμητή, τότε το δίκτυο έχει ολοκληρώσει

τη διαδικασία εκμάθησης (Simpson, 1990). Το πιο σημαντικό στάδιο της προετοιμασίας του εκπαιδευτικού υλικού από δεδομένα κειμένου είναι ο προσδιορισμός των αριθμητικών τιμών – διανυσμάτων για τα αντικείμενα δεδομένων. Παρά τα μειονεκτήματα, η διανυσματική προσέγγιση επιλέχθηκε για τη δημιουργία αριθμητικών δεδομένων από το αρκετά μεγάλο σετ όρων.

Με βάση τα παραπάνω, το σετ εκπαίδευσης του προτεινόμενου PNN αποτελείται από ένα μεγάλο αριθμό προσομοιωμένων περιπτώσεων event για κάθε μία κατηγορία που δοκιμάστηκε. Κάθε περίπτωση event μετατρέπεται στο αντίστοιχο PEV. Η ποικιλία αυτών των δειγμάτων εξασφαλίζει την επιτυχή υλοποίηση του PNN μιας και η απόδοσή του είναι άμεση και δεν εξαρτάται από χρονοβόρα εκπαίδευση. Ο ταξινομητής υλοποιήθηκε με την C++. Για τον έλεγχο της απόδοσης του ταξινομητή χρησιμοποιήθηκε η τεχνική επαλήθευσης 10-fold cross.

Ο μέσος χρόνος που χρειάζεται για την ολοκλήρωση της περιόδου εκπαίδευσης είναι 183 δευτερόλεπτα. Οι *Εξισώσεις 5 και 6*, περιγράφουν το κριτήριο πληροφορίας του Akaike (AIC, Akaike's Information Criterion), καθώς και το ελάχιστο μήκος περιγραφής του Rissanen (MDL, Minimum Description Length) αντίστοιχα. Οι τιμές $|d_{ij} - y_{ij}|$ αντιστοιχούν στις αποστάσεις μεταξύ της επιθυμητής και της πραγματικής εξόδου του δικτύου για το υπόδειγμα i^{th} που ανήκει στο στοιχείο επεξεργασίας j^{th} .

$$AIC(k) = N * \ln(MSE) + 2 * K \quad 5$$

$$MDL(k) = N * \ln(MSE) + 0.5 * K * \ln(N) \quad 6$$

$$\text{Mean Square Error: } MSE = \frac{\sum_{j=1}^P \sum_{i=1}^N (d_{ij} - y_{ij})^2}{N \cdot P}$$

Στα παραπάνω κριτήρια, το P ισούται με τον αριθμό των στοιχείων επεξεργασίας εξόδου, ενώ το N και K καθορίζουν το ποσό των υποδειγμάτων στο σετ εκπαίδευσης και τον αριθμό των βαρών του δικτύου αντίστοιχα. Το AIC μετράει την εξισορρόπηση (trade-off) μεταξύ της απόδοσης της εκπαίδευσης και του μεγέθους του δικτύου, ενώ το MDL συνδυάζει το σφάλμα του μοντέλου με τον αριθμό των βαθμών ελευθερίας για τον καθορισμό του επιπέδου γενίκευσης. Οι προαναφερθέν δείκτες που απεικονίζονται στον *Πίνακα 4-2*, λήφθηκαν υπόψη προκειμένου να ρυθμιστούν σωστά οι μέσες διαφορές και οι διακυμάνσεις των αντίστοιχων τοπικών προσεγγιστών (local approximators) και παράγουν ένα πίνακα σύγχυσης (confusion matrix) με τις καλύτερες δυνατές τιμές στα διαγώνια κελιά.

		Classifier Decision					Active Training Parameters	
		A	B	C	D	E	P	5
Actual Class	A	0,959	0,009	0,000	0,000	0,031	N	2500
	B	0,000	0,980	0,000	0,019	0,000	K	2345202
	C	0,015	0,000	0,975	0,000	0,010	MSE	0,00842
	D	0,000	0,011	0,000	0,988	0,000	AIC	4261436,818
	E	0,018	0,000	0,006	0,000	0,976	MDL	10183055,121

Πίνακας 4-2: Στάδιο Εκπαίδευσης του Confusion Matrix

Κατά την αξιολόγηση παρατηρήθηκε ότι το νευρωνικό δίκτυο δεν ήταν κατάλληλα εκπαιδευμένο όταν $0,1 < \beta < 1$, λόγω των μεγάλων τιμών του MSE. Αντιθέτως, το MSE μειώθηκε σημαντικά όταν το β ορίστηκε ως 1 και το confusion matrix της εκπαίδευσης έδωσε αποδεκτές ποσοστιαίες τιμές στα διαγώνια κελιά. Ωστόσο, περαιτέρω έρευνα πάνω στην επίδραση του β στη δυνατότητα εκμάθησης του ταξινομητή, έδειξε ότι το MSE και οι τιμές των AIC/MDL αυξήθηκαν ελάχιστα για τις τιμές του β που κυμαίνονται μεταξύ 1 και 10 και στη συνέχεια μειώθηκαν περαιτέρω όταν $10 < \beta < 100$. Τέλος, στην περίπτωση του $\beta \geq 100$, μετρήθηκε ότι τα κριτήρια AIC και MDL βελτιστοποιήθηκαν.

Ο Πίνακας 4-2 παρουσιάζει το confusion matrix, καθώς και τις αντίστοιχες παραμέτρους που εμπλέκονται για την βέλτιστη περίοδο εκπαίδευσης του ταξινομητή. Αυτός ο πίνακας ορίζεται, επισημαίνοντας την επιθυμητή κατηγοριοποίηση στις σειρές και την προβλεπόμενη κατηγοριοποίηση στις στήλες. Δεδομένου ότι η προβλεπόμενη κατηγοριοποίηση πρέπει να είναι η ίδια με την επιθυμητή, η ιδανική κατάσταση είναι να έχουμε όλα τα υποδείγματα να καταλήγουν στα διαγώνια κελιά του πίνακα. Αυτό σημαίνει πως στα διαγώνια κελιά το ιδανικό ποσοστό πρέπει να είναι 100. Ο ταξινομητής αναγνώρισε επιτυχώς το 98.8% των στοιχείων εκπαίδευσης που αφορούν την κατηγορία D (Recreation) κατά την περίοδο εκμάθησης, ενώ μόλις το 1,2% θεωρήθηκε εσφαλμένα ως κατηγορία B (Sports). Η κατηγορία D ήταν αυτή που παρουσίασε την καλύτερη ακρίβεια εκπαίδευσης. Όπως είναι προφανές, υψηλότερες τιμές στα μη διαγώνια κελιά σημαίνουν μεγαλύτερη αβεβαιότητα, η οποία θα οδηγήσει τελικά σε μεγαλύτερες πιθανότητες εσφαλμένης ταξινόμησης μεταξύ των κατηγοριών. Βάση της αναπαράστασης του Πίνακα 4-2, η μεγαλύτερη αβεβαιότητα προέκυψε για την κατηγορία A (Art), δεδομένου ότι περισσότερο από το 3% των προσομοιωμένων περιπτώσεων χαρακτηρίστηκαν εσφαλμένα ως κατηγορία E (Computers) και μόλις το 1% ως κατηγορία B (Sports).

5 ΥΛΟΠΟΙΗΣΗ ΣΥΣΤΗΜΑΤΟΣ

Σε αυτή την ενότητα περιγράφεται η υλοποίηση και η λειτουργία των υποσυστημάτων που απαρτίζουν το σύστημα. Αρχικά, περιγράφεται η λειτουργία του ιστότοπου μέσω του οποίου καταχωρούνται τα δεδομένα στην βάση δεδομένων του συστήματος και στη συνέχεια παρουσιάζεται το Service που υλοποιήθηκε, το οποίο διαχειρίζεται τα δεδομένα αυτά μέσω των αλγορίθμων που αναπτύχθηκαν στα πλαίσια της παρούσας διατριβής.

5.1 Ανάπτυξη και Λειτουργία του Ιστότοπου

Όπως αναφέρθηκε και στο τρίτο κεφάλαιο, ένα από τα υποσυστήματα του συστήματος που αναπτύχθηκε είναι ο Ιστότοπος, μέσω του οποίου εισάγονται τα δεδομένα στην βάση δεδομένων του συστήματος. Αν και για τις δοκιμές που έγιναν για την αξιολόγηση του συστήματος, που παρουσιάζονται στο επόμενο κεφάλαιο, χρησιμοποιήθηκε ένα script για τη γρήγορη δημιουργία τεχνητών δεδομένων (χρηστών, διοργανωτών και events), ο ιστότοπος αποτελεί το βασικό εργαλείο για την εισαγωγή πραγματικών δεδομένων στο σύστημα. Ο ιστότοπος αναπτύχθηκε με το Visual Studio 2010 .Net της Microsoft σε γλώσσα προγραμματισμού Visual Basic. Στην παρούσα φάση, το σύστημα δεν προορίζεται για εμπορική χρήση και έτσι δεν έχουν δημιουργηθεί οι διάφοροι έλεγχοι που θα περίμενε κανείς κατά την εισαγωγή δεδομένων. Για παράδειγμα δεν γίνεται ορθογραφικός έλεγχος, έλεγχος εισαγωγής έγκυρης ημερομηνίας, αν κάποιο πεδίο δέχεται μόνο αριθμούς αντί για γράμματα κ.λ.π., παρά μόνο ο έλεγχος για το αν έχουν εισαχθεί δεδομένα σε όποια πεδία απαιτείται. Σε περίπτωση μελλοντικής εμπορικής χρήσης της εφαρμογής, πρέπει να επανασχεδιαστεί τόσο εμφανισιακά, όσο και λειτουργικά, ώστε να καλύψει τις σύγχρονες απαιτήσεις και ανάγκες των χρηστών. Στη συνέχεια, ακολουθεί η περιγραφή των λειτουργιών κάθε ιστοσελίδας του ιστότοπου.

5.1.1 Home

Είναι η αρχική σελίδα της εφαρμογής (*Εικόνα 5-1*) και παρέχει τρεις επιλογές:

- Είσοδος στο σύστημα. Οδηγεί σε επόμενη ιστοσελίδα όπου ο χρήστης ή ο διοργανωτής event (manager) εισάγει τα στοιχεία του για να εισέλθει στο σύστημα.

- Εγγραφή Χρήστη. Οδηγεί σε επόμενη ιστοσελίδα όπου ο χρήστης δημιουργεί για πρώτη φορά το προφίλ του και εγγράφεται στο σύστημα.
- Εγγραφή Event Manager. Ομοίως με το χρήστη, ο manager δημιουργεί για πρώτη φορά το προφίλ του και εγγράφεται στο σύστημα.



Εικόνα 5-1: Home.aspx – Αρχική Σελίδα

5.1.2 UserProfile

Είναι η φόρμα όπου ο χρήστης καταχωρεί για πρώτη φορά τα στοιχεία του και εγγράφεται στο σύστημα. Βάσει αυτών των στοιχείων δημιουργείται και καταχωρείται το UserKey του χρήστη στη βάση δεδομένων. Η φόρμα περιλαμβάνει δύο Tabs. Στο πρώτο Tab (Εικόνα 5-2) ο χρήστης καταχωρεί τα προσωπικά του στοιχεία όπως:

- Όνομα, Επώνυμο.
- Έτος Γέννησης, Φύλο.
- Μόρφωση, Μέσω της drop down list ο χρήστης επιλέγει το επίπεδο μόρφωσής του (Λύκειο, ΙΕΚ, ΤΕΙ, ΑΕΙ, Μεταπτυχιακό, Διδακτορικό).
- Επάγγελμα. Μέσω της drop down list επιλέγει μέσα από μία λίστα επαγγελμάτων αυτό που του αντιστοιχεί.
- Διεύθυνση. Ο χρήστης εισάγει τα στοιχεία της διεύθυνσης του και το στίγμα του εμφανίζεται στον χάρτη με τη χρήση του Google Maps API.

- Σταθερό τηλέφωνο, το οποίο δεν είναι υποχρεωτικό.
- Κινητό Τηλέφωνο, το οποίο δεν είναι υποχρεωτικό, αλλά επιθυμητό σε περίπτωση που θέλει να λαμβάνει ειδοποιήσεις από το σύστημα μέσω SMS.
- E-mail, το οποίο είναι υποχρεωτικό για να λαμβάνει ειδοποιήσεις από το σύστημα, καθώς επίσης αποτελεί και το username του για την εισαγωγή στο σύστημα.

Firefox

Untitled Document

http://localhost:36785/EventLocator/UserProfile.aspx

Google

Καταχώρηση Προφίλ Χρήστη

Προσωπικά Στοιχεία Προτιμήσεις / Ενδιαφέροντα

Όνομα: Σωτήριος - Επώνυμο: Μπούτας

Έτος Γέννησης: 02/09/1978 - Φύλο: Άνδρας Γυναίκα

Μόρφωση: Μεταπτυχιακό - Επάγγελμα: Μηχανικός

Στοιχεία Διεύθυνσης και Επικοινωνίας:

Περιοχή: Νέα Σμύρνη - Οδός: Σακτούρη - Αριθμός: 10

Πόλη: Αθήνα - Ταχ. Κωδ.: 17124 - Χώρα: Ελλάδα

Σταθερό Τηλ.: 2109730967 - Κινητό: 6944322785

E-mail: sboutas@medialab.ntua.gr

Καταχώρηση Ακύρωση

Εικόνα 5-2: UserProfile.aspx – Καταχώρηση Προσωπικών Στοιχείων

Στο δεύτερο Tab (Εικόνα 5-3) ο χρήστης δηλώνει το ενδιαφέρον του για τις υποκατηγορίες των πέντε βασικών κατηγοριών που υποστηρίζονται από το σύστημα όπως είδαμε στο τέταρτο κεφάλαιο. Μέσω μιας drop down list που υπάρχει για κάθε μία μπορεί να επιλέξει μεταξύ των τιμών: καθόλου, ελάχιστα, αρκετά, πολύ και πάρα πολύ. Είναι υποχρεωτικό να δηλώσει ενδιαφέρον για τουλάχιστον μία υποκατηγορία.

Καταχώρηση Προφίλ Χρήστη

Προσωπικά Στοιχεία Προτιμήσεις / Ενδιαφέροντα

Παρακαλώ βαθμολογήστε τις παρακάτω κατηγορίες ανάλογα με το πόσο σας ενδιαφέρουν:
(Είναι υποχρεωτικό να βαθμολογήσετε τουλάχιστον μία κατηγορία.)

Art
Architecture: Ελάχιστα - Cinema: Πάρα Πολύ - Museum: Αρκετά - Music: Πάρα Πολύ - Theater: Πολύ

Sports
News: Καθόλου - Football: Ελάχιστα - Basketball: Ελάχιστα - Organizations: Καθόλου - Volleyball: Ελάχιστα

Science
Technology: Πάρα Πολύ - Physics: Πολύ - Mathematics: Ελάχιστα - Biology: Πολύ - Medicine: Πολύ

Recreation
Tourism: Πολύ - Sea: Πάρα Πολύ - Entertainment: Πολύ - Car: Πολύ - Motorcycle: Ελάχιστα

Computers
Internet: Πάρα Πολύ - Software: Πολύ - Hardware: Καθόλου
Καθόλου
Ελάχιστα
Αρκετά
Πολύ
Πάρα Πολύ - Education: Πολύ - Companies: Αρκετά

Καταχώρηση Ακύρωση

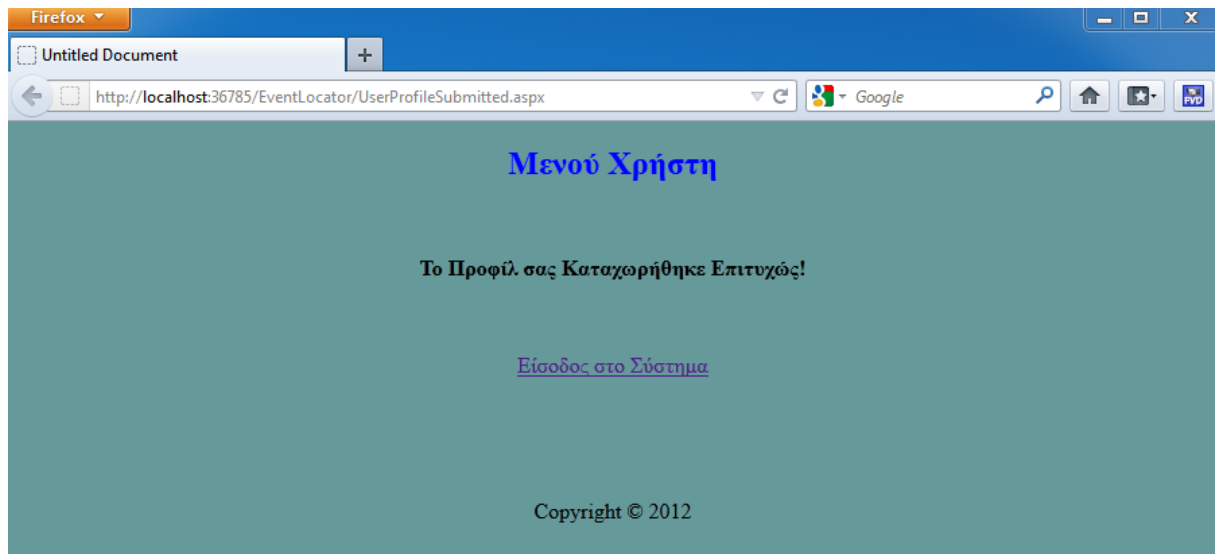
Εικόνα 5-3: UserProfile.aspx – Καταχώρηση Προτιμήσεων / Ενδιαφερόντων

Έχοντας συμπληρώσει τα απαιτούμενα στοιχεία και στα δύο Tabs, πατώντας το κουμπί «Καταχώρηση» τα δεδομένα εισάγονται στη βάση δεδομένων και δημιουργούνται ταυτόχρονα το UserKey και ο λογαριασμός του. Ο κωδικός του για την εισαγωγή του στο σύστημα είναι τα 4 πρώτα ψηφία του email του. Αντίθετα, μπορεί να πατήσει το κουμπί

«Ακύρωση» και να ακυρώσει τη διαδικασία της εγγραφής, επιστρέφοντας στην αρχική σελίδα.

5.1.3 UserProfileSubmitted

Η σελίδα αυτή (Εικόνα 5-4) επιβεβαιώνει πως το προφίλ του χρήστη έχει καταχωρηθεί επιτυχώς και η μόνη επιλογή είναι το link για την είσοδο στο σύστημα.



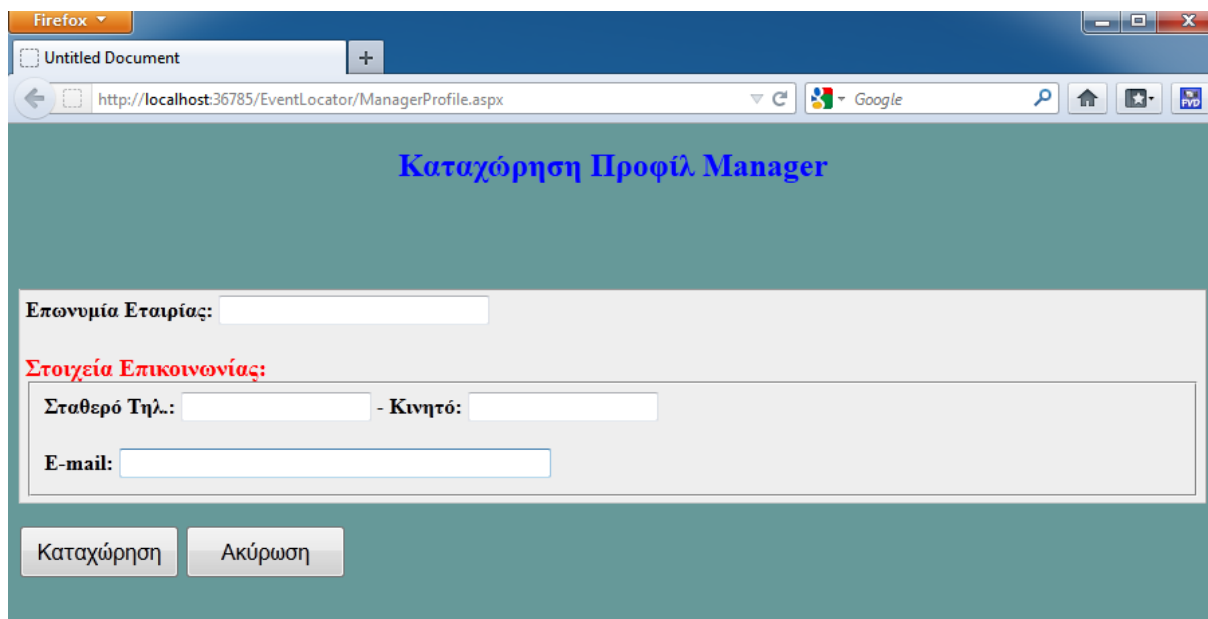
Εικόνα 5-4: UserProfileSubmitted.aspx – Επιτυχής Καταχώρηση Προφίλ Χρήστη

5.1.4 ManagerProfile

Είναι η φόρμα (Εικόνα 5-5) όπου ο διοργανωτής event (manager) καταχωρεί για πρώτη φορά τα στοιχεία του και εγγράφεται στο σύστημα. Τα στοιχεία που καταχωρεί είναι:

- Επωνυμία Εταιρίας.
- Σταθερό και κινητό τηλέφωνο.
- E-mail.

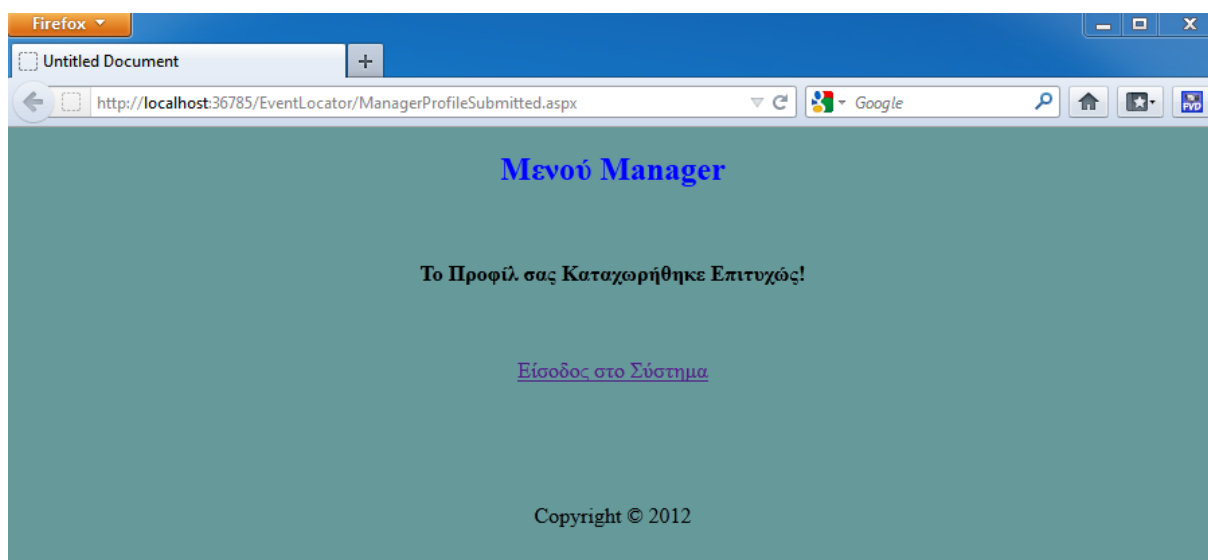
Σε αντίθεση με το χρήστη, ο manager δεν χρειάζεται να αποθηκεύσει περαιτέρω στοιχεία, αλλά ούτε και κάποιες προτιμήσεις.



Εικόνα 5-5: ManagerProfile.aspx – Καταχώρηση Προφίλ Χρήστη

5.1.5 ManagerProfileSubmitted

Ομοίως με του χρήστη, η σελίδα αυτή (Εικόνα 5-6) επιβεβαιώνει πως το προφίλ του manager έχει καταχωρηθεί επιτυχώς και η μόνη επιλογή είναι το link για την είσοδο στο σύστημα.



Εικόνα 5-6: ManagerProfileSubmitted.aspx – Επιτυχής Καταχώρηση Προφίλ Manager

5.1.6 Login

Στη σελίδα αυτή (Εικόνα 5-7), ο χρήστης ή ο manager εισάγουν τα διαπιστευτήρια του λογαριασμού τους για να εισέλθουν στο σύστημα που θα τους οδηγήσει στα αντίστοιχα μενού. Όπως αναφέρθηκε προηγουμένως, το username είναι το E-mail του χρήστη / manager και το password τα 4 πρώτα του ψηφία. Σε περίπτωση εισαγωγής λανθασμένων διαπιστευτηρίων, το σύστημα εμφανίζει μήνυμα λάθους και ζητάει πάλι από το χρήστη / manager να τα καταχωρήσει.



Εικόνα 5-7: Login.aspx – Είσοδος στο Σύστημα

5.1.7 Usermenu

Η σελίδα αυτή (Εικόνα 5-8) είναι το κεντρικό μενού του χρήστη. Οι επιλογές που έχει ο χρήστης είναι:

- Προβολή Προφίλ. Να δει τα στοιχεία του προφίλ του (Εικόνες 5-2 και 5-3).
- Ενημέρωση Προφίλ. Εισάγεται και πάλι στην φόρμα καταχώρησης προφίλ όπου μπορεί είτε να κάνει αλλαγές στα προσωπικά του στοιχεία (Εικόνα 5-2) εκτός από τα τέσσερα πρώτα (όνομα, επώνυμο, έτος γέννησης και φύλο), είτε να αλλάξει τη βαθμολογία του στις κατηγορίες ενδιαφέροντος (Εικόνα 5-3).

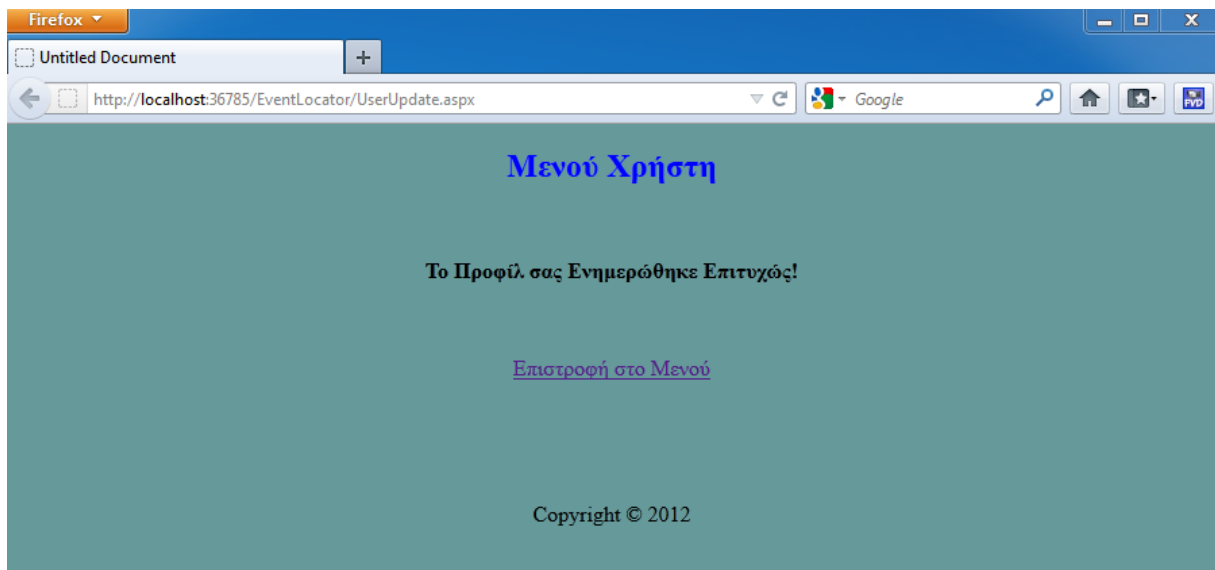
- Διαγραφή Προφίλ. Διαγράφει το λογαριασμό του από το σύστημα. Μαζί διαγράφονται το UserKey, τα RankingScores, καθώς και τα Feedbacks για τα events που συμμετείχε.
- Feedback. Ο χρήστης μπορεί να αφήσει Feedback για τα Events που έχει επιλεγεί.
- Αποσύνδεση. Αποσυνδέεται από το σύστημα και επιστρέφει στην αρχική σελίδα.



Εικόνα 5-8: UserMenu.aspx – Μενού Χρήστη

5.1.8 UserUpdate

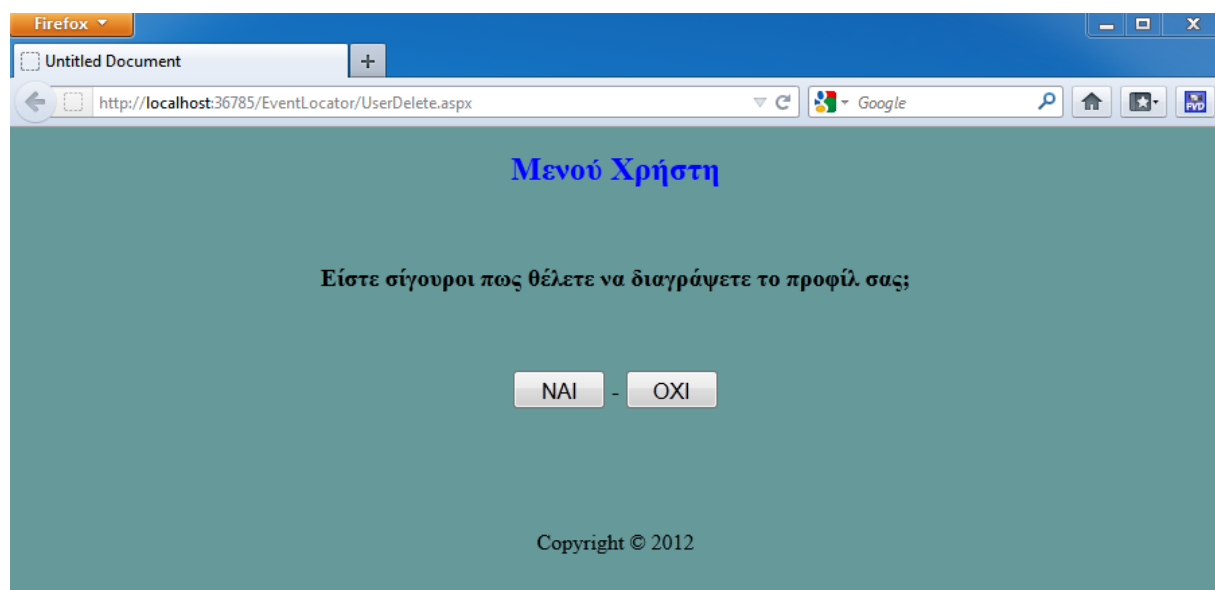
Η σελίδα αυτή (Εικόνα 5-9) επιβεβαιώνει πως το προφίλ του χρήστη έχει ενημερωθεί επιτυχώς και η μόνη επιλογή είναι το link για την επιστροφή στο μενού χρήστη.



Εικόνα 5-9: UserUpdate.aspx – Ενημέρωση Προφίλ Χρήστη

5.1.9 UserDelete

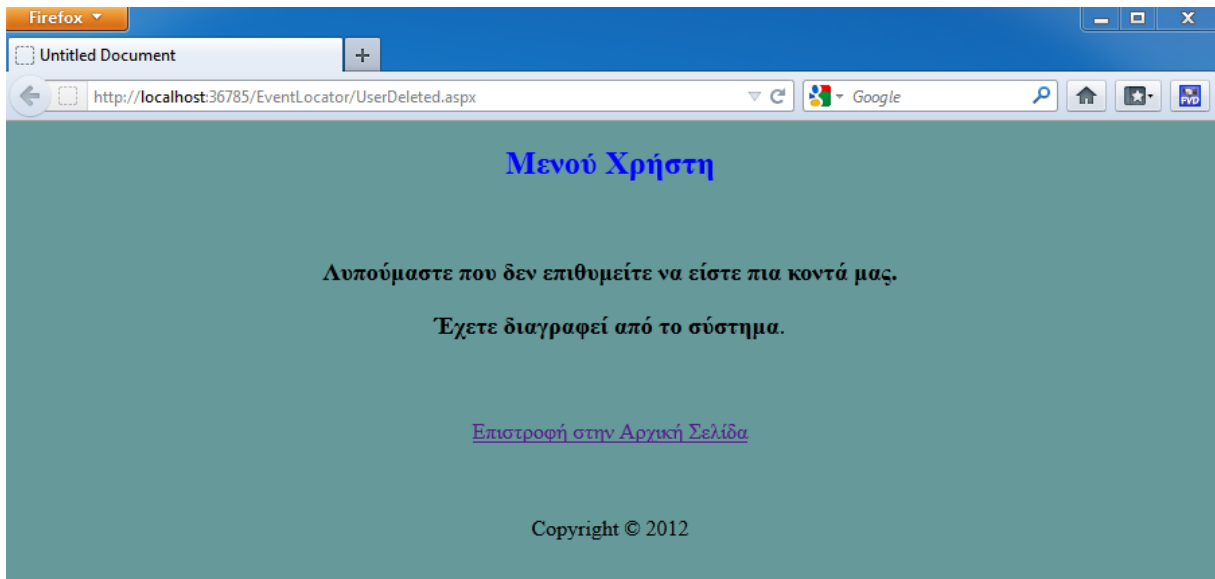
Η σελίδα αυτή (Εικόνα 5-10) ζητάει από το χρήστη να επιβεβαιώσει αν επιθυμεί την οριστική διαγραφή του προφίλ του. Πατώντας στο κουμπί «Ναι», το προφίλ διαγράφεται οριστικά μαζί με το UserKey, τα RankingScores, καθώς και τα Feedbacks για τα events που συμμετείχε.



Εικόνα 5-10: UserDelete.aspx – Επιβεβαίωση Διαγραφής Προφίλ Χρήστη

5.1.10 UserDeleted

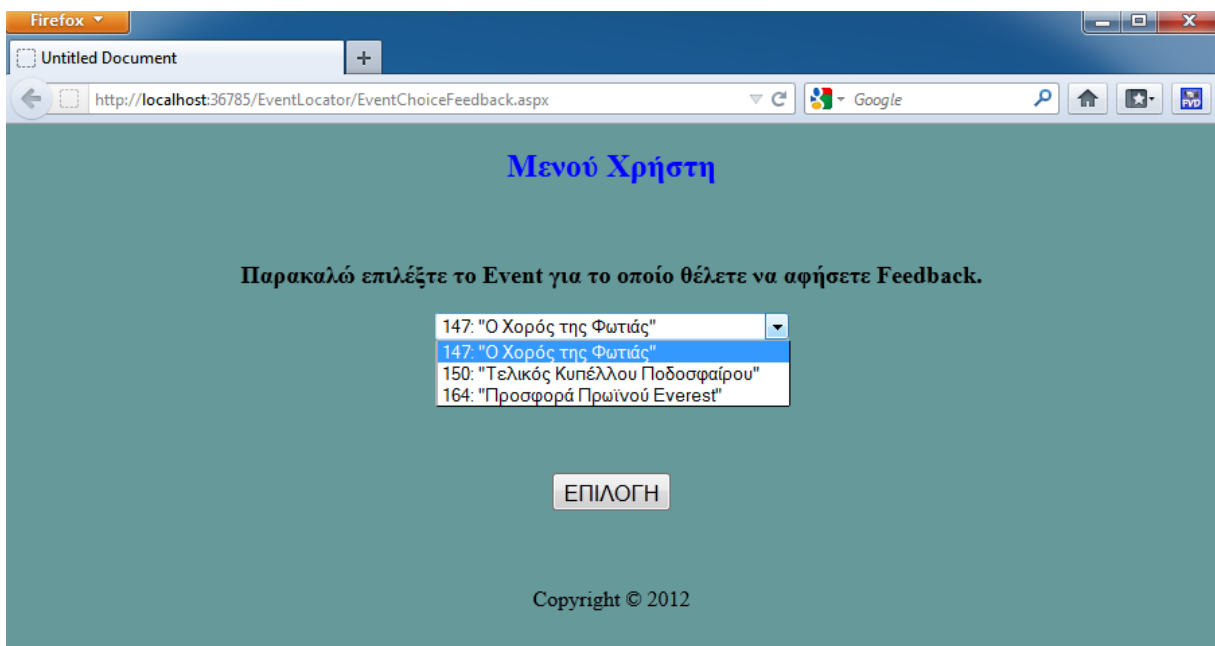
Η σελίδα αυτή (Εικόνα 5-11) επιβεβαιώνει πως το προφίλ του χρήστη έχει διαγραφεί επιτυχώς και η μόνη επιλογή είναι το link για την επιστροφή στην αρχική σελίδα.



Εικόνα 5-11: UserDeleted.aspx – Διαγραφή Προφίλ Χρήστη

5.1.11 EventChoiceFeedback

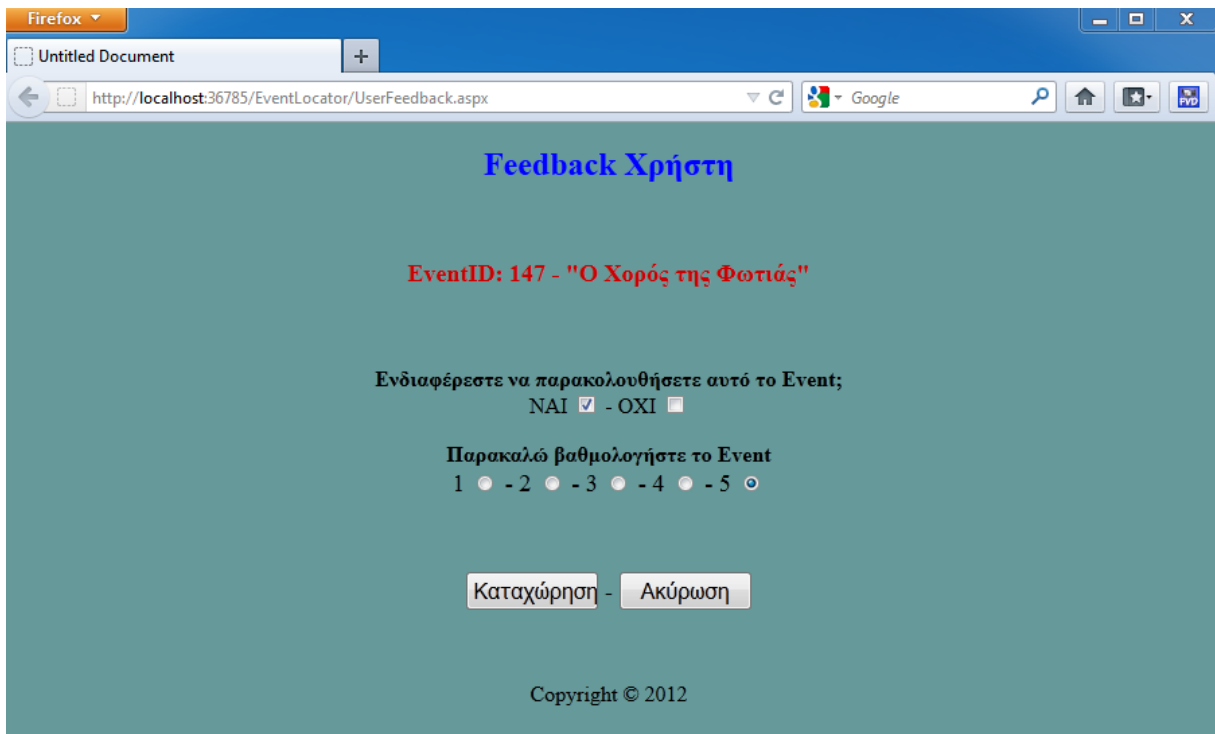
Στη σελίδα αυτή (Εικόνα 5-12) ο χρήστης επιλέγει από μία drop down list το event για το οποίο θέλει να αφήσει Feedback.



Εικόνα 5-12: EventChoiceFeedback.aspx – Επιλογή Event για Feedback

5.1.12 UserFeedback

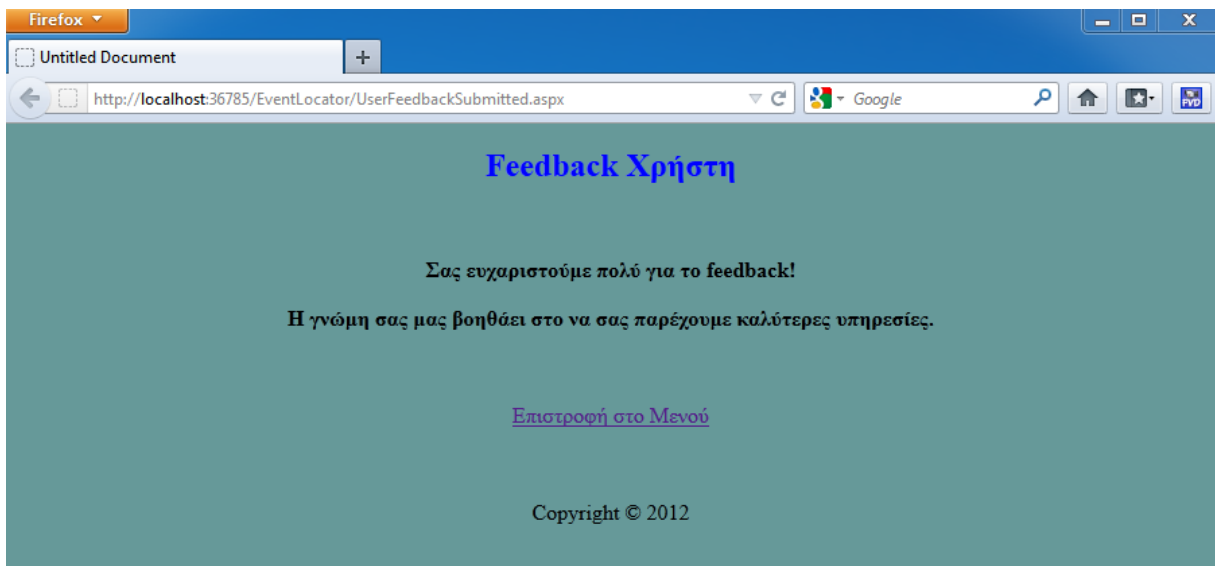
Στη σελίδα αυτή (Εικόνα 5-13) ο χρήστης επιλέγει για το αν ενδιαφέρεται να παρακολουθήσει το event, καθώς επίσης και να το βαθμολογήσει, ανάλογα με το πόσο τον ενδιαφέρει (1 - Καθόλου έως 5 - Πάρα Πολύ). Πατώντας «Καταχώρηση» το Feedback του αποθηκεύεται στο σύστημα, Αντίθετα, πατώντας «Ακύρωση» επιστρέφει στο μενού χρήστη.



Εικόνα 5-13: UserFeedback.aspx – Feedback Χρήστη για Event

5.1.13 UserFeedbackSubmitted

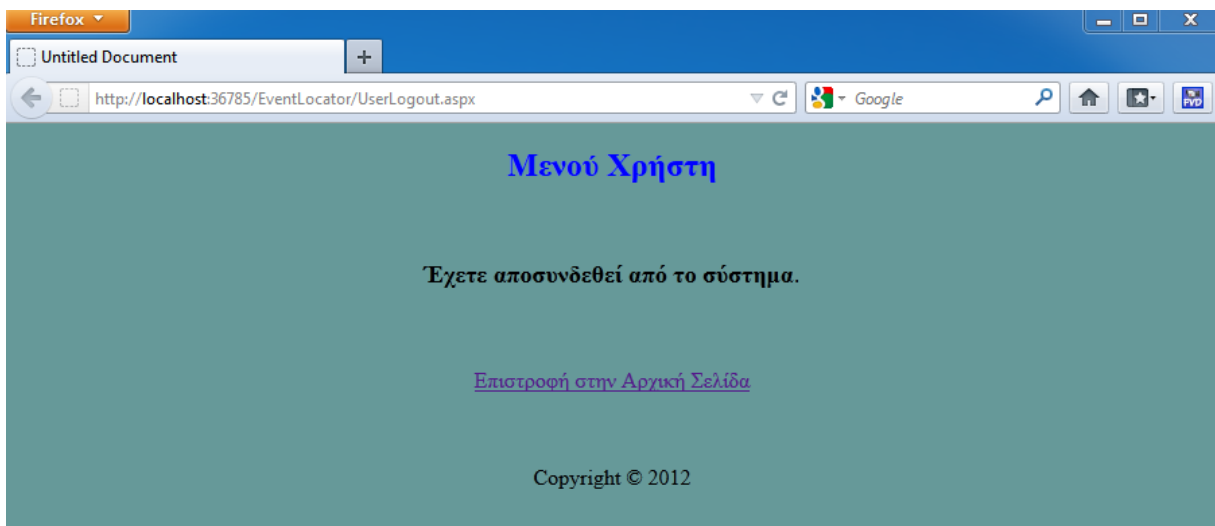
Η σελίδα αυτή (Εικόνα 5-14) επιβεβαιώνει πως το Feedback του χρήστη έχει καταχωρηθεί επιτυχώς και η μόνη επιλογή είναι το link για την επιστροφή στο μενού χρήστη.



Εικόνα 5-14: UserFeedbackSubmitted.aspx – Καταχώρηση Feedback Χρήστη

5.1.14 UserLogout

Η σελίδα αυτή (Εικόνα 5-15) επιβεβαιώνει πως ο χρήστης έχει αποσυνδεθεί επιτυχώς από το σύστημα και η μόνη επιλογή είναι το link για την επιστροφή στην αρχική σελίδα.



Εικόνα 5-15: UserLogout.aspx – Αποσύνδεση Χρήστη

5.1.15 ManagerMenu

Η σελίδα αυτή (Εικόνα 5-16) είναι το κεντρικό μενού του manager. Οι επιλογές που έχει ο manager είναι:

- Προβολή Προφίλ. Να δει τα στοιχεία του προφίλ του (Εικόνα 5-5).
- Ενημέρωση Προφίλ. Εισάγεται και πάλι στην φόρμα καταχώρησης προφίλ όπου μπορεί να κάνει αλλαγές στα προσωπικά του στοιχεία (Εικόνα 5-5) εκτός από την επωνυμία της εταιρίας.
- Διαγραφή Προφίλ. Διαγράφει το λογαριασμό του από το σύστημα. Μαζί διαγράφονται και όλα τα events μαζί με τα αντίστοιχα δεδομένα τους που συνδέονται με αυτόν.
- Καταχώρηση Event. Καταχωρεί ένα νέο event στο σύστημα.
- Ενημέρωση Event. Ενημερώνει τα στοιχεία ενός ήδη καταχωρημένου event.
- Διαγραφή Event. Διαγράφει ένα event από το σύστημα μαζί με όλα τα δεδομένα που σχετίζονται με αυτό.
- Αποτελέσματα. Εμφανίζει τα αποτελέσματα των Events, όπως για παράδειγμα πόσοι χρήστες έχουν επιλεγθεί, πόσοι από αυτούς άφησαν Feedback ότι θα το παρακολουθήσουν, καθώς και τη βαθμολογία για το πόσο ενδιαφέρον το βρίσκουν.
- Αποσύνδεση. Αποσυνδέεται από το σύστημα και επιστρέφει στην αρχική σελίδα.



Εικόνα 5-16: ManagerMenu.aspx – Μενού Manager

5.1.16 ManagerUpdate

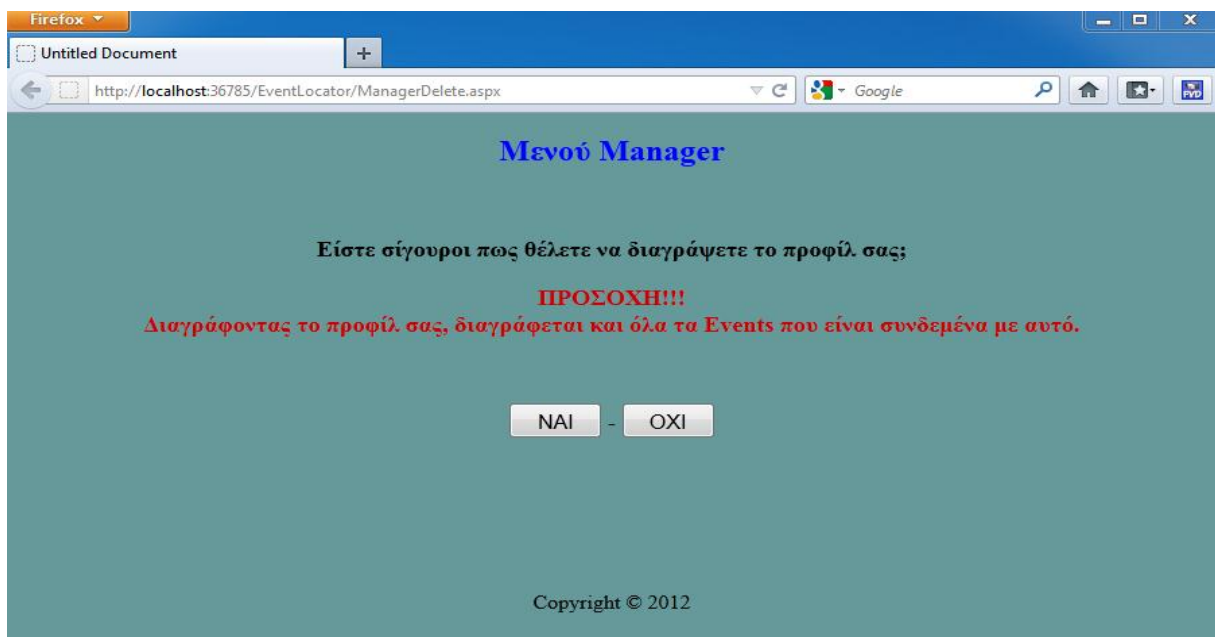
Ομοίως με του χρήστη, η σελίδα αυτή (Εικόνα 5-17) επιβεβαιώνει πως το προφίλ του manager έχει ενημερωθεί επιτυχώς και η μόνη επιλογή είναι το link για την επιστροφή στο μενού manager.



Εικόνα 5-17: ManagerUpdate.aspx – Ενημέρωση Προφίλ Manager

5.1.17 ManagerDelete

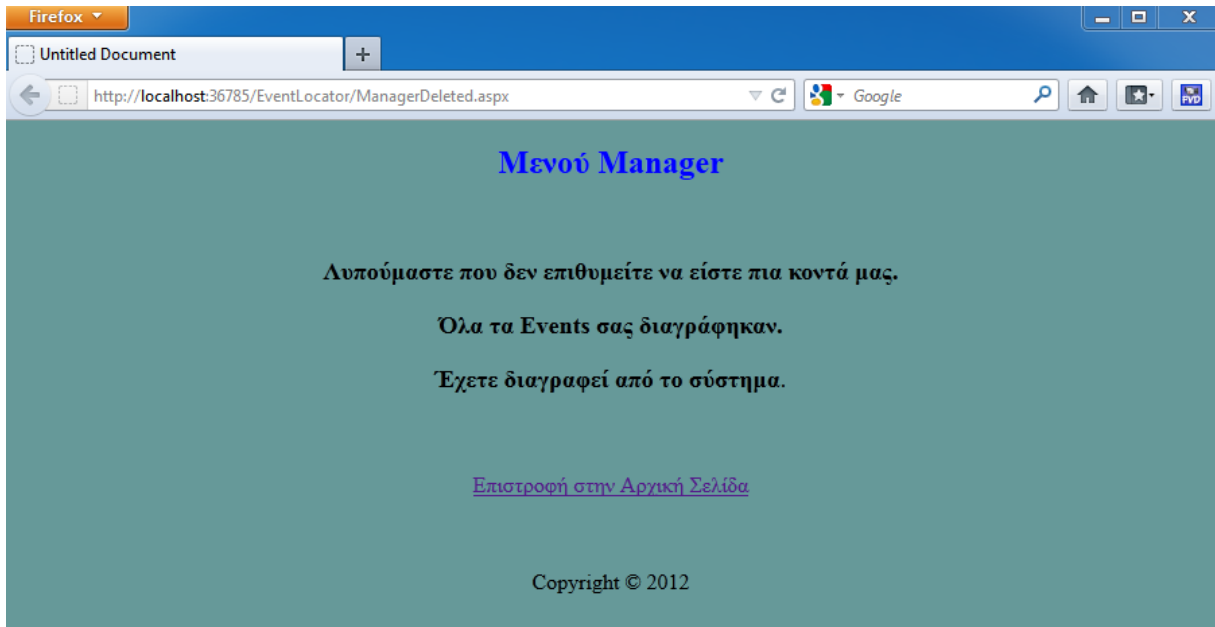
Ομοίως με του χρήστη, η σελίδα αυτή (Εικόνα 5-18) ζητάει από τον manager να επιβεβαιώσει αν επιθυμεί την οριστική διαγραφή του προφίλ του. Πατώντας στο κουμπί «Ναι», το προφίλ διαγράφεται οριστικά μαζί με όλα τα events και τα αντίστοιχα δεδομένα τους που συνδέονται με αυτόν.



Εικόνα 5-18: ManagerDelete.aspx – Επιβεβαίωση Διαγραφής Προφίλ Manager

5.1.18 ManagerDeleted

Ομοίως με του χρήστη, η σελίδα αυτή (Εικόνα 5-19) επιβεβαιώνει πως το προφίλ του manager έχει διαγραφεί επιτυχώς και η μόνη επιλογή είναι το link για την επιστροφή στην αρχική σελίδα.



Εικόνα 5-19: ManagerDeleted.aspx – Διαγραφή Προφίλ Manager

5.1.19 Event

Είναι η φόρμα (Εικόνα 5-20) όπου ο manager καταχωρεί ένα νέο event στο σύστημα. Βάσει αυτών των στοιχείων δημιουργείται και καταχωρείται το EventKey του event στη βάση δεδομένων. Τα στοιχεία που καταχωρούνται είναι:

- Τίτλος. Ο τίτλος του event.
- Ημερομηνία Από και Ημερομηνία Έως. Ορίζουν τη διάρκεια του event.
- Κατηγορία. Η κατηγορία στην οποία ανήκει το συγκεκριμένο event.
- Επιθυμητό Φύλο. Ο manager επιλέγει από μία drop down list το φύλο των χρηστών στους οποίους απευθύνεται το συγκεκριμένο event (Όλοι, Άντρας ή Γυναίκα).
- Επιθυμητή Απόσταση. Ο manager επιλέγει από μία drop down list την επιθυμητή απόσταση των χρηστών από το event (Ανεξαρτήτως Περιοχής, Ίδια Περιοχή ή Κοντινές Περιοχές).

- Επιθυμητές Ηλικιακές Ομάδες. Ο manager επιλέγει τα checkbox των ηλικιακών ομάδων που επιθυμεί να προσελκύσει στο event.
- Διεύθυνση. Ο manager εισάγει τα στοιχεία της τοποθεσίας του event και το στίγμα του εμφανίζεται στον χάρτη με τη χρήση του Google Maps API.
- Μήνυμα προς Αποστολή. Ο manager εισάγει το μήνυμα, το οποίο θα αποσταλεί για να ενημερωθούν οι χρήστες που θα επιλεγούν για το συγκεκριμένο event.

Καταχώρηση Event

Τίτλος: "Ο Χορός της Φωτιάς" - Βαγγέλης Παπαθανασίου

Ημ/νία Από: 16/06/2012 - Ημ/νία Έως: 17/06/2012 - Κατηγορία: Music

Επιθυμητό Φύλο: Όλοι - Επιθυμητή Απόσταση: Ανεξαρτήτως Περιοχής

Επιθυμητές Ηλικιακές Ομάδες:
 18-24 - 25-34 - 35-49 - 50-64 - 65+

Τοποθεσία Event:

Περιοχή: Αθήνα - Οδός: Πρόσβαση Λυκαβηττού - Αριθμός:

Πόλη: Αθήνα - Ταχ. Κωδ.: 10677 - Χώρα: Ελλάδα

Μήνυμα προς Αποστολή:
 Μουσικό Event στον Λυκαβηττό!

Καταχώρηση Ακύρωση

Εικόνα 5-20: Event.aspx – Καταχώρηση Event

Έχοντας συμπληρώσει τα απαιτούμενα στοιχεία και πατώντας το κουμπί «Καταχώρηση» τα δεδομένα εισάγονται στη βάση δεδομένων δημιουργείται ταυτόχρονα το EventKey. Αντίθετα, πατώντας το κουμπί «Ακύρωση» η διαδικασία καταχώρησης event ακυρώνεται και επιστρέφει στο μενού manager.

5.1.20 EventSubmitted

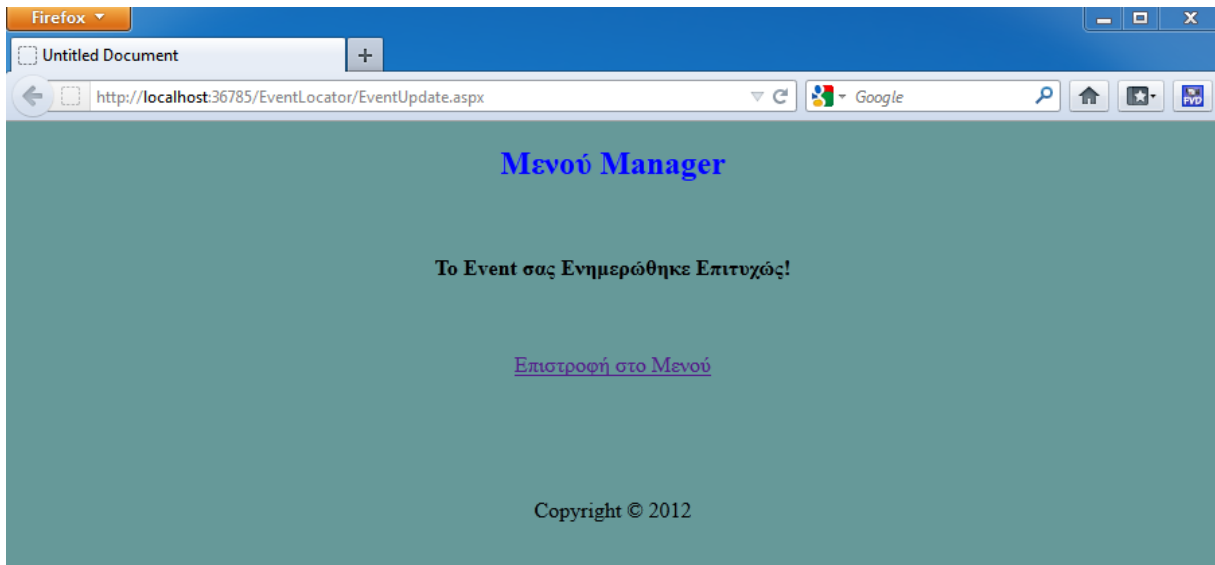
Η σελίδα αυτή (Εικόνα 5-21) επιβεβαιώνει πως το event έχει καταχωρηθεί επιτυχώς και η μόνη επιλογή είναι το link για την επιστροφή στο μενού manager.



Εικόνα 5-21: EventSubmitted.aspx – Επιτυχής Καταχώρηση Event

5.1.21 EventUpdate

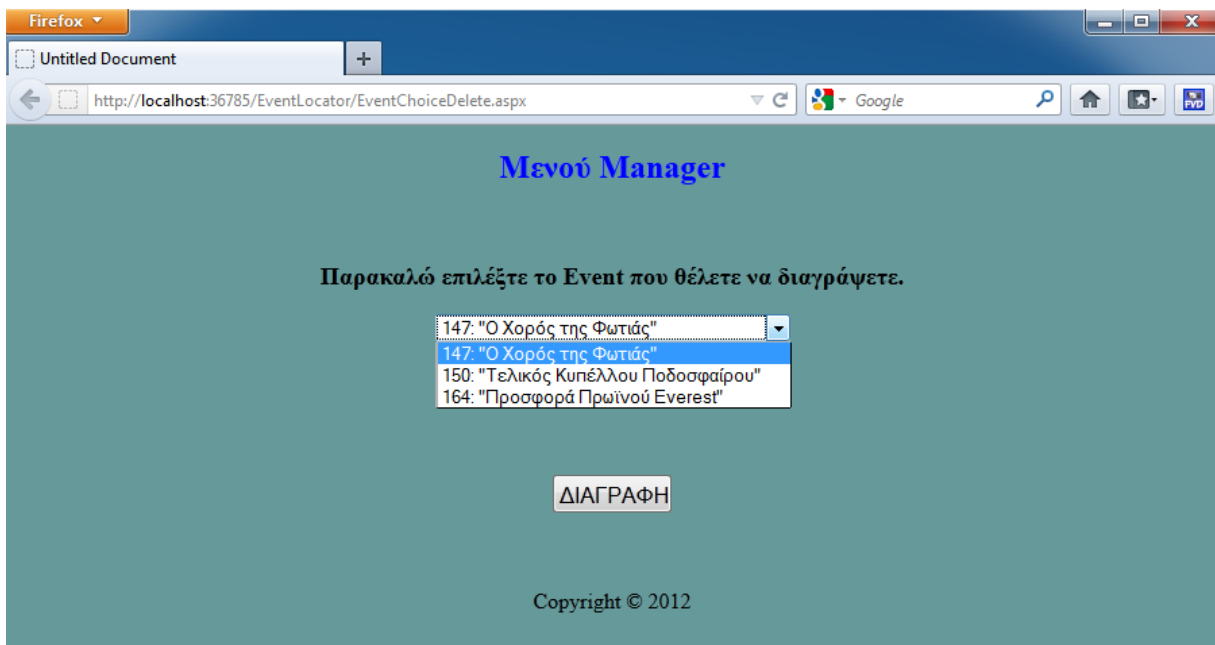
Η σελίδα αυτή (Εικόνα 5-22) επιβεβαιώνει πως τα στοιχεία του event έχουν ενημερωθεί επιτυχώς και η μόνη επιλογή είναι το link για την επιστροφή στο μενού manager.



Εικόνα 5-22: EventUpdate.aspx – Επιτυχής Ενημέρωση Event

5.1.22 EventChoiceDelete

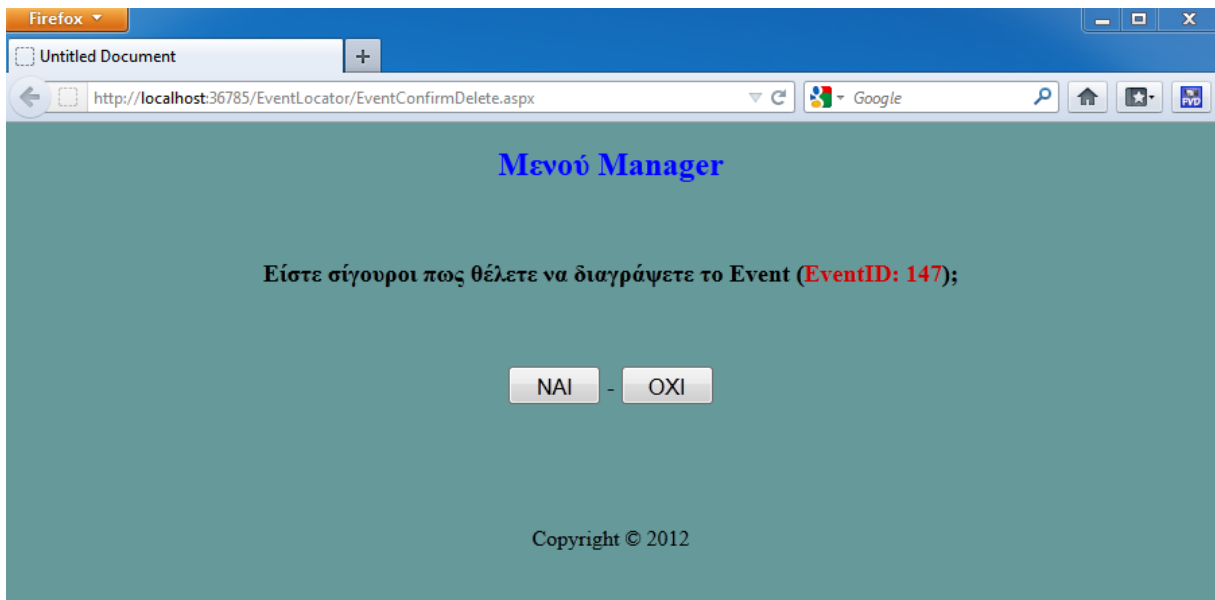
Στη σελίδα αυτή (Εικόνα 5-23) ο manager επιλέγει από μία drop down list το event που θέλει να διαγράψει.



Εικόνα 5-23: EventChoiceDelete.aspx – Επιλογή Event για Διαγραφή

5.1.23 EventConfirmDelete

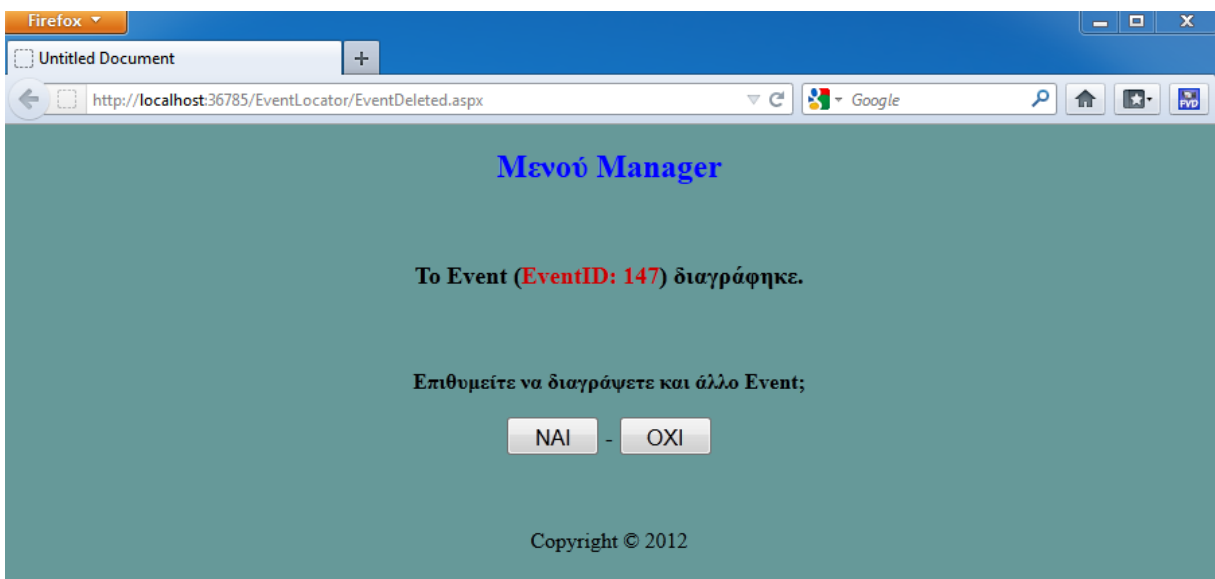
Η σελίδα αυτή (Εικόνα 5-24) ζητάει από τον manager να επιβεβαιώσει αν επιθυμεί την οριστική διαγραφή του event που επέλεξε. Πατώντας στο κουμπί «Ναι», το event διαγράφεται οριστικά μαζί με όλα τα δεδομένα που σχετίζονται με αυτό. Αντίθετα, πατώντας «Όχι» ακυρώνεται η διαγραφή και επιστρέφει στο μενού manager.



Εικόνα 5-24: EventConfirmDelete.aspx – Επιβεβαίωση Διαγραφής Event

5.1.24 EventDeleted

Η σελίδα αυτή (Εικόνα 5-25) επιβεβαιώνει πως το event έχει διαγραφεί επιτυχώς και δίνεται η επιλογή για διαγραφή άλλου event.

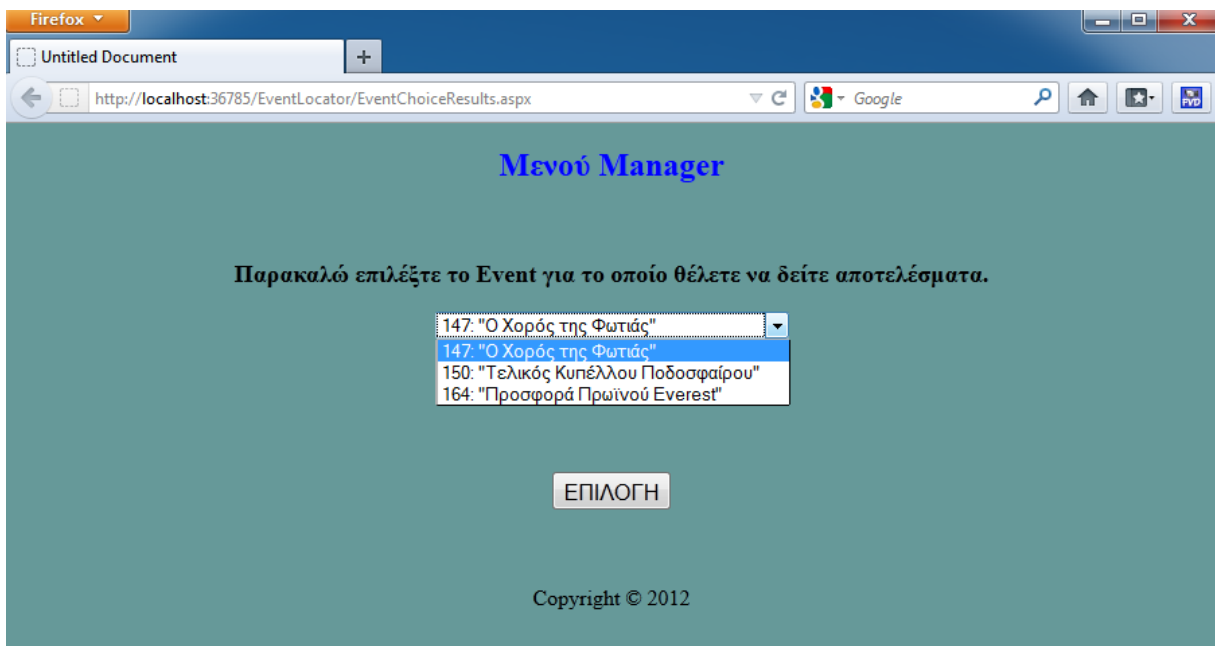


Εικόνα 5-25: EventDeleted.aspx – Διαγραφή Event

Πατώντας στο κουμπί «Ναι», επιστρέφει στη σελίδα διαγραφής event για την επιλογή άλλου event προς διαγραφή (Εικόνα 5-23), ενώ πατώντας «Όχι» επιστρέφει στο μενού manager.

5.1.25 EventChoiceResults

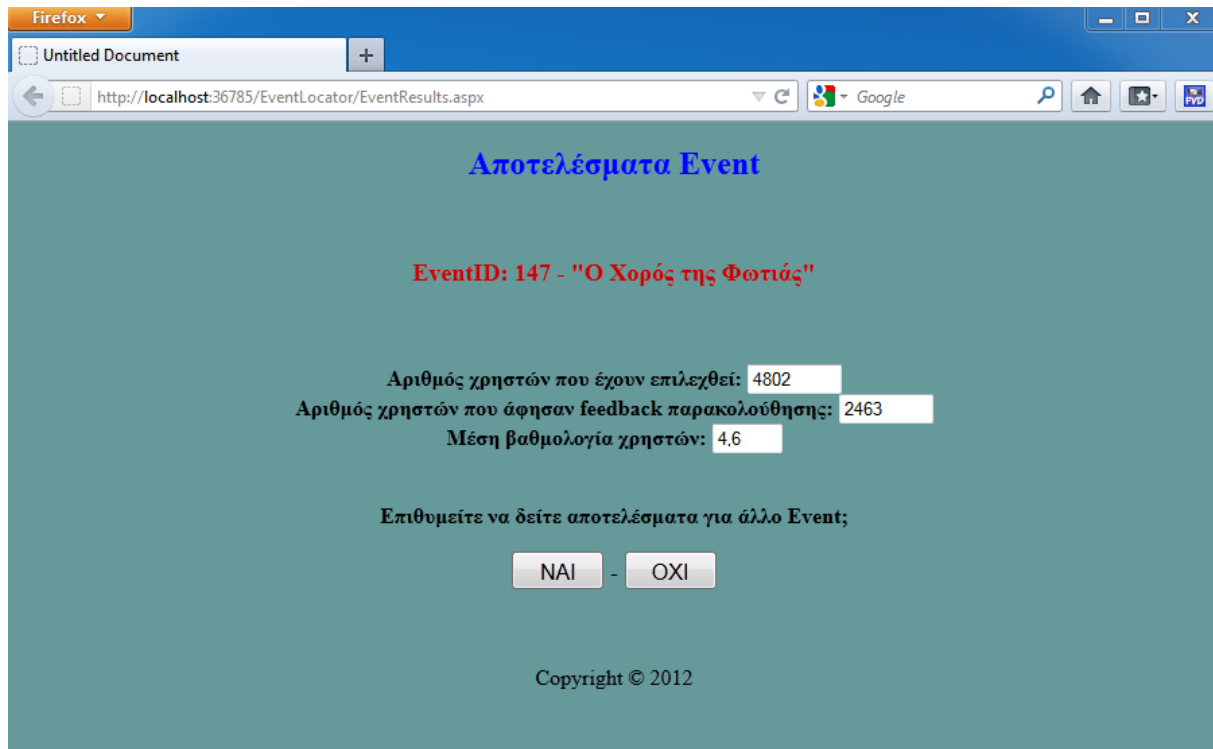
Στη σελίδα αυτή (Εικόνα 5-26) ο manager επιλέγει από μία drop down list το event για το οποίο θέλει να δει τα αποτελέσματα.



Εικόνα 5-26: EventChoiceResults.aspx – Επιλογή Event για Αποτελέσματα

5.1.26 EventResults

Στη σελίδα αυτή (Εικόνα 5-27) ο manager βλέπει τα αποτελέσματα για το event που επέλεξε. Τα αποτελέσματα δείχνουν πόσοι χρήστες έχουν επιλεγθεί για αυτό το event, πόσοι από αυτούς άφησαν Feedback ότι θα το παρακολουθήσουν, καθώς και τη βαθμολογία για το πόσο ενδιαφέρον το βρίσκουν. Δίνεται επίσης η δυνατότητα επιλογής άλλου event για παρουσίαση των αποτελεσμάτων.

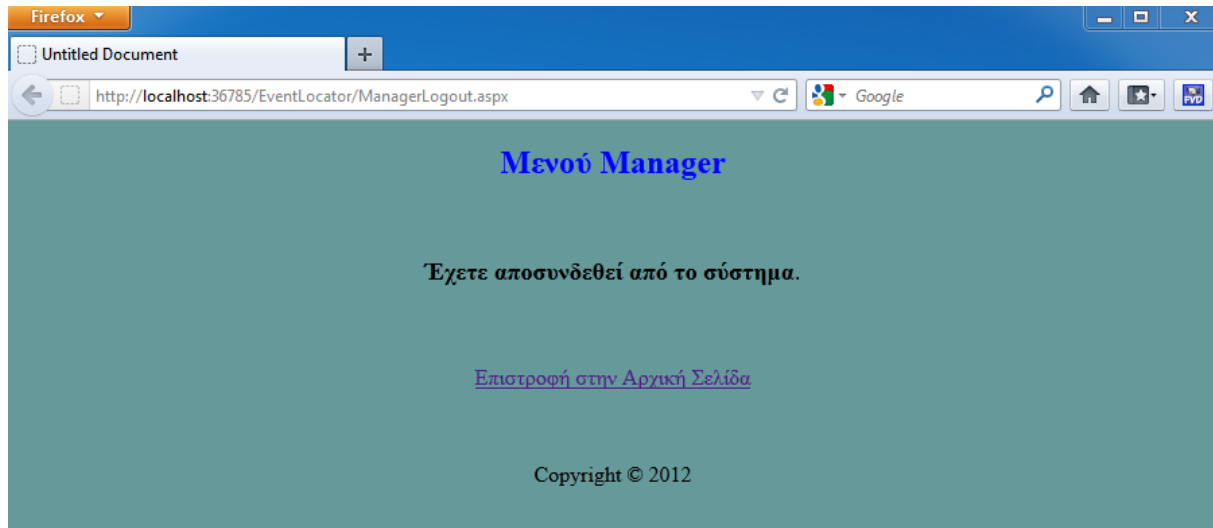


Εικόνα 5-27: EventResults.aspx – Αποτελέσματα Event

Πατώντας στο κουμπί «Ναι», επιστρέφει στη σελίδα επιλογής νέου event για προβολή αποτελεσμάτων (Εικόνα 5-26), ενώ πατώντας «Όχι» επιστρέφει στο μενού manager.

5.1.27 ManagerLogout

Ομοίως με του χρήστη, η σελίδα αυτή (Εικόνα 5-28) επιβεβαιώνει πως ο manager έχει αποσυνδεθεί επιτυχώς από το σύστημα και η μόνη επιλογή είναι το link για την επιστροφή στην αρχική σελίδα.



Εικόνα 5-28: ManagerLogout.aspx – Αποσύνδεση Manager

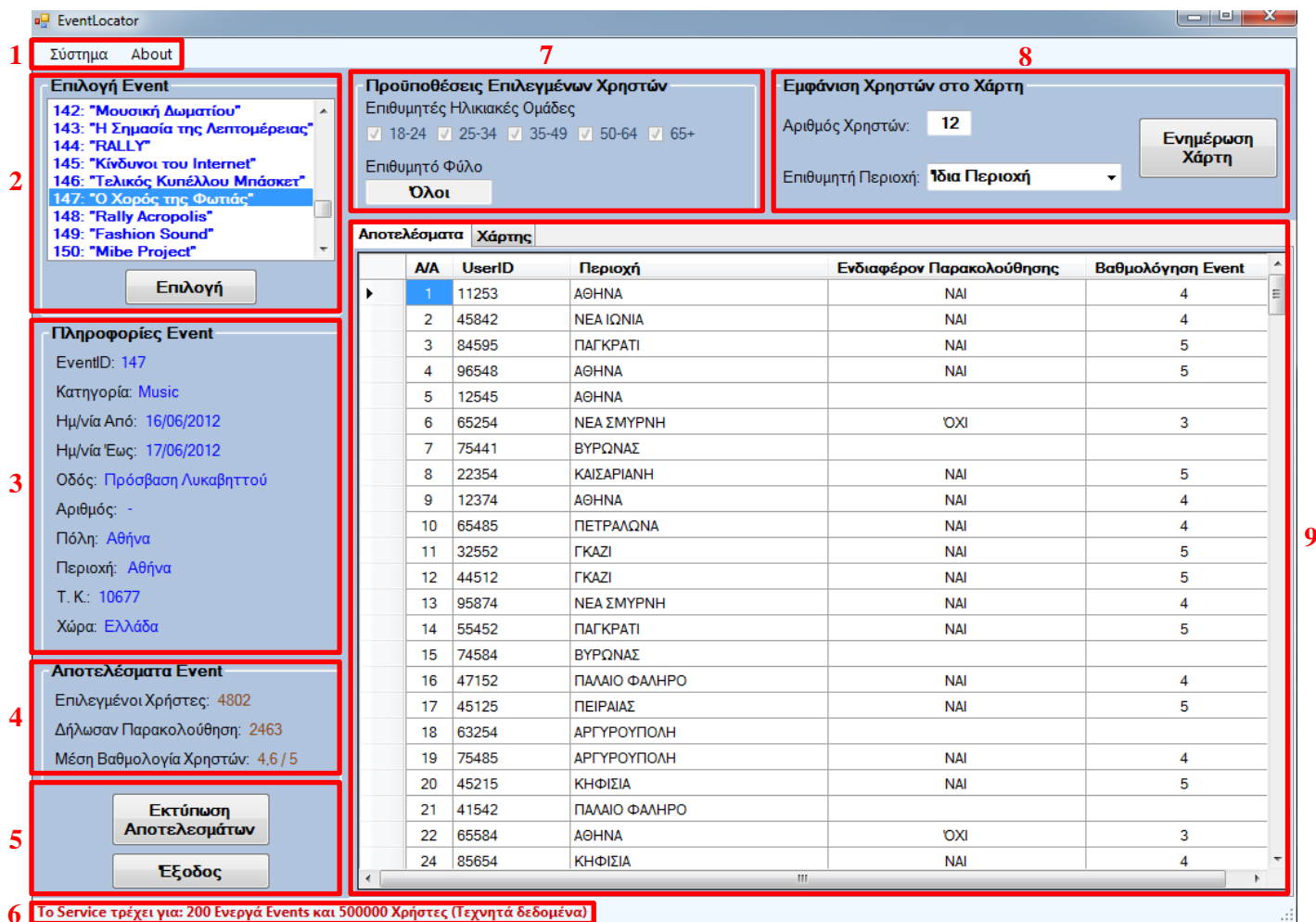
5.2 Service

Στο σημείο αυτό, παρουσιάζεται το Service που υλοποιήθηκε και αποτελεί τον πυρήνα του συστήματος μιας και είναι υπεύθυνο για την εκτέλεση των αλγορίθμων, καθώς και για ένα πλήθος διαφόρων λειτουργιών. Όπως και ο ιστότοπος, έτσι και το Service αναπτύχθηκε με το Visual Studio 2010 .Net της Microsoft σε γλώσσα προγραμματισμού Visual Basic. Μόνη εξαίρεση αποτελεί ο ταξινομητής (PNN), ο οποίος, όπως αναφέρθηκε και στο τέταρτο κεφάλαιο (4.3.5), αναπτύχθηκε σε C++ και η εισαγωγή του στο Service έγινε με τη μορφή ενός DLL αρχείου.

Η λειτουργία του Service είναι πλήρως αυτοματοποιημένη αν και δίνεται η δυνατότητα στον διαχειριστή του συστήματος να σταματήσει ή να ξεκινήσει και πάλι τη λειτουργία του. Στη συνέχεια παρουσιάζεται το περιβάλλον λειτουργίας του Service όπου εξηγείται η λειτουργία του και τα διάφορα controls.

5.2.1 Περιβάλλον Λειτουργίας και Controls

Στην Εικόνα που ακολουθεί έχουν μαρκαριστεί με αριθμημένα πλαίσια τα διάφορα controls και εξηγείται η λειτουργία τους.

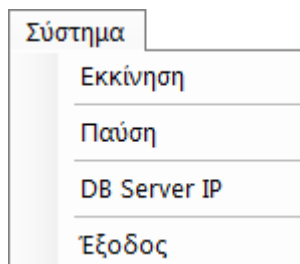


Εικόνα 5-29: Controls του Service

Menu Bar

Το 1^ο πλαίσιο δείχνει την μπάρα του μενού, η οποία έχει δύο επιλογές: Σύστημα και About. Κάνοντας κλικ στην επιλογή σύστημα, εμφανίζεται το μενού που βλέπουμε στην Εικόνα 5-30 και έχει 4 επιλογές. Η επιλογή «Εκκίνηση» ξεκινάει χειροκίνητα το Service σε περίπτωση που έχει σταματήσει, ενώ η «Παύση» το σταματάει. Με την «DB Server IP» δίνεται η δυνατότητα να εισάγουμε την IP διεύθυνση της βάσης δεδομένων σε περίπτωση που αυτή είναι εγκατεστημένη σε διαφορετικό μηχάνημα, από ότι το Service. Τέλος με την επιλογή «Εξοδος» βγαίνουμε από το παράθυρο του προγράμματος δίχως όμως να σταματήσουμε το Service, το οποίο συνεχίζει να εκτελείται.

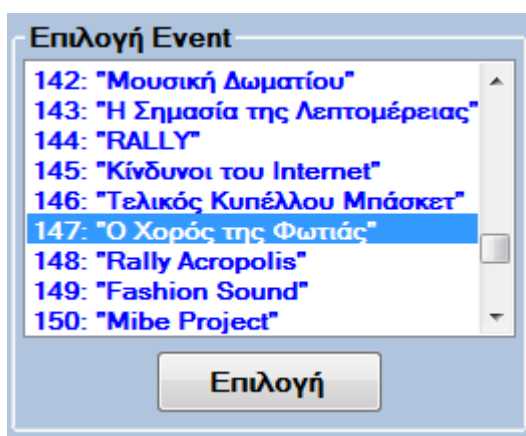
Επιλέγοντας το About, εμφανίζεται ένα πλαίσιο με πληροφορίες για το σύστημα όπως το πότε δημιουργήθηκε και σε τι περιβάλλον, καθώς και πληροφορίες για τον κατασκευαστή και τρόπου επικοινωνίας μαζί του.



Εικόνα 5-30: Επιλογές του Συστήματος

Επιλογή Event

Το 2^ο πλαίσιο δείχνει μία λίστα με όλα τα events (Εικόνα 5-31) που υπάρχουν στο σύστημα. Μέσα από αυτή τη λίστα μαρκάρουμε το event που θέλουμε και πατάμε το κουμπί «Επιλογή» ώστε να φορτωθούν τα στοιχεία του στην οθόνη του προγράμματος.



Εικόνα 5-31: Επιλογή Event

Πληροφορίες Event

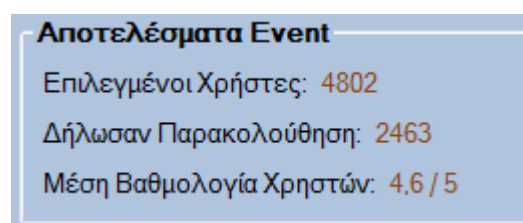
Το 3^ο πλαίσιο δείχνει πληροφορίες για το event που επιλέχθηκε (Εικόνα 5-32). Αυτές περιλαμβάνουν το EventID, την κατηγορία που ανήκει, το πότε ξεκινάει και πότε τελειώνει, καθώς και την τοποθεσία του.



Εικόνα 5-32: Πληροφορίες Event

Αποτελέσματα Event

Το 4^ο πλαίσιο δείχνει τα αποτελέσματα του επιλεγμένου event (Εικόνα 5-33) που προκύπτουν από τους αλγόριθμους του συστήματος, αλλά και από το feedback των χρηστών. Ο αριθμός 4802 των επιλεγμένων χρηστών προκύπτει από τον αλγόριθμο ταξινόμησης και είναι οι χρήστες που πιθανόν να ενδιαφέρονται να παρακολουθήσουν το συγκεκριμένο event. Ο αριθμός 2463 στο δήλωσαν παρακολούθηση, δείχνει τους χρήστες που άφησαν feedback ότι θα παρακολουθήσουν το συγκεκριμένο event. Αυτό δε σημαίνει όμως ότι δεν θα το παρακολουθήσουν περισσότεροι, απλά μπορεί να μην έτυχε να αφήσουν feedback. Ο αριθμός 4,6 / 5 στην βαθμολογία των χρηστών δείχνει τη μέση βαθμολογία που δήλωσαν οι χρήστες στο feedback για το συγκεκριμένο event βάσει του ενδιαφέρον τους για αυτό. Η βαθμολογία αυτή προκύπτει μόνο από τους χρήστες που έχουν αφήσει feedback.



Εικόνα 5-33: Αποτελέσματα Event

Εκτύπωση Αποτελεσμάτων και Έξοδος

Το 5^ο πλαίσιο περιλαμβάνει δύο κουμπιά. Το κουμπί «Εκτύπωση Αποτελεσμάτων» εκτυπώνει μία λίστα με τους χρήστες που φαίνεται στην *Εικόνα 5-29 Πλαίσιο 9* περιλαμβάνοντας στο τέλος της τα αποτελέσματα του 4^{ου} Πλαισίου και ομαδοποιεί τους χρήστες ανά περιοχή.

Το κουμπί «Έξοδος» έχει την ίδια λειτουργία με την επιλογή «Έξοδος» του Συστήματος στην μπάρα του μενού.

Status Bar

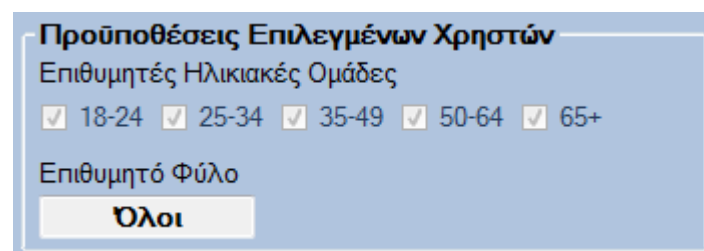
Το 6^ο πλαίσιο δείχνει την μπάρα κατάστασης του συστήματος (*Εικόνα 5-34*), η οποία δίνει πληροφορίες για το νούμερο των αποθηκευμένων event και χρηστών στο σύστημα.

To Service τρέχει για: 200 Ενεργά Events και 500000 Χρήστες (Τεχνητά δεδομένα)

Εικόνα 5-34: Status Bar

Προϋποθέσεις Επιλεγμένων Χρηστών

Το 7^ο πλαίσιο (*Εικόνα 5-35*) εμφανίζει πληροφορίες σχετικά με που απευθύνεται το συγκεκριμένο event και με αυτές τις προϋποθέσεις έγινε η επιλογή των χρηστών στους οποίους εφαρμόστηκε η ταξινόμηση. Εδώ βλέπουμε πως για το συγκεκριμένο event επιλέχθηκαν όλες οι ηλικιακές ομάδες, καθώς και όλοι οι χρήστες ανεξαρτήτως φύλου.



Εικόνα 5-35: Προϋποθέσεις Επιλεγμένων

Εμφάνιση Χρηστών στο Χάρτη

Το 8^ο πλαίσιο (*Εικόνα 5-36*) περιλαμβάνει δύο controls. Ένα που ορίζει τον αριθμό των χρηστών που θα εμφανιστούν στο χάρτη και πληρούν τις καλύτερες προϋποθέσεις για να παρακολουθήσουν το event και ένα που ορίζει από ποια περιοχή του χάρτη θα επιλεγούν

αυτοί οι χρήστες. Πατώντας το κουμπί «Ενημέρωση Χάρτη», ο χάρτης ενημερώνεται και προβάλλει τα στοιχεία που ζητήθηκαν.

Εικόνα 5-36: Εμφάνιση Χρηστών στο Χάρτη

Tab Control

Τέλος, το 9^ο πλαίσιο αφορά το Tab control (Εικόνα 5-37), το οποίο περιλαμβάνει δύο Tabs. Μία για τα αποτελέσματα με τους χρήστες που επιλέχθηκαν από το σύστημα και μία που δείχνει την τοποθεσία του event και των χρηστών στο χάρτη. Τα αποτελέσματα των χρηστών περιλαμβάνουν το UserID, την περιοχή του χρήστη, αν δήλωσε ενδιαφέρον παρακολούθησης και το πόσο βαθμολόγησε το συγκεκριμένο event. Ο χάρτης ενημερώνεται βάσει των τιμών που εισάγουμε στα προαναφερθέν controls (Εικόνα 5-36).

Αποτελέσματα		Χάρτης			
Α/Α	UserID	Περιοχή	Ενδιαφέρον Παρακολούθησης	Βαθμολόγηση Event	
1	11253	ΑΘΗΝΑ	ΝΑΙ	4	
2	45842	ΝΕΑ ΙΩΝΙΑ	ΝΑΙ	4	
3	84595	ΠΑΓΚΡΑΤΙ	ΝΑΙ	5	
4	96548	ΑΘΗΝΑ	ΝΑΙ	5	
5	12545	ΑΘΗΝΑ			
6	65254	ΝΕΑ ΣΜΥΡΝΗ	ΌΧΙ	3	
7	75441	ΒΥΡΩΝΑΣ			
8	22354	ΚΑΙΣΑΡΙΑΝΗ	ΝΑΙ	5	
9	12374	ΑΘΗΝΑ	ΝΑΙ	4	
10	65485	ΠΕΤΡΑΛΩΝΑ	ΝΑΙ	4	
11	32552	ΓΚΑΖΙ	ΝΑΙ	5	
12	44512	ΓΚΑΖΙ	ΝΑΙ	5	
13	95874	ΝΕΑ ΣΜΥΡΝΗ	ΝΑΙ	4	
14	55452	ΠΑΓΚΡΑΤΙ	ΝΑΙ	5	
15	74584	ΒΥΡΩΝΑΣ			
16	47152	ΠΑΛΑΙΟ ΦΑΛΗΡΟ	ΝΑΙ	4	
17	45125	ΠΕΙΡΑΙΑΣ	ΝΑΙ	5	
18	63254	ΑΡΓΥΡΟΥΠΟΛΗ			
19	75485	ΑΡΓΥΡΟΥΠΟΛΗ	ΝΑΙ	4	
20	45215	ΚΗΦΙΣΙΑ	ΝΑΙ	5	
21	41542	ΠΑΛΑΙΟ ΦΑΛΗΡΟ			
22	65584	ΑΘΗΝΑ	ΌΧΙ	3	
24	85654	ΚΗΦΙΣΙΑ	ΝΑΙ	4	

Εικόνα 5-37: Tab Control

5.2.2 Παρουσίαση των Αποτελεσμάτων του Ταξινομητή

Τα αποτελέσματα του ταξινομητή για κάθε event παρουσιάζονται είτε σε λίστα όπως είδαμε στην *Εικόνα 5-37* είτε στο χάρτη (*Εικόνα 5-38*).

The screenshot shows the EventLocator application interface. On the left, there is a sidebar with the following sections:

- Επιλογή Event:** A list of 15 events, with event 147, "Ο Χορός της Φωτιάς", selected.
- Πληροφορίες Event:** EventID: 147, Category: Music, Start Date: 16/06/2012, End Date: 17/06/2012, Route: Πρόσβαση Λυκαβηττού, City: Αθήνα, Area: Αθήνα, Phone: 10677, Country: Ελλάδα.
- Αποτελέσματα Event:** Selected Users: 4802, Total Reviews: 2463, Average Rating: 4.6/5.

At the top right, there are filters for "Προϋποθέσεις Επιλεγμένων Χρηστών" (Age groups: 18-24, 25-34, 35-49, 50-64, 65+; Gender: Όλοι) and "Εμφάνιση Χρηστών στο Χάρτη" (Number of Users: 12, Area: Ίδια Περιοχή, Refresh Map button).

The main area displays a map of Athens with several red location pins. The map is titled "Αποτελέσματα Χάρτης".

At the bottom of the application window, a status bar reads: "Το Service τρέχει για: 200 Ενεργά Events και 500000 Χρήστες (Τεχνητά δεδομένα)".

Εικόνα 5-38: Παρουσίαση των Αποτελεσμάτων στο Χάρτη

Μέσα από αυτές τις δύο εικόνες (5-37 και 5-38) βλέπουμε το πόσο καλά λειτουργεί ο ταξινομητής, κάτι που θα συζητηθεί αναλυτικότερα στο επόμενο κεφάλαιο. Στην Εικόνα 5-37 βλέπουμε πως οι χρήστες που έχουν δηλώσει πως δε θα παρακολουθήσουν το event το έχουν βαθμολογήσει με 3. Από αυτό συμπεραίνουμε πως ο ταξινομητής σωστά τους επέλεξε για το event αυτής της κατηγορίας, αλλά οι συγκεκριμένοι χρήστες δεν ενδιαφέρονται τόσο για αυτό το είδος μουσικής. Όσοι βαθμολόγησαν με 4 και 5 το συγκεκριμένο event έχουν δηλώσει επίσης ότι θα το παρακολουθήσουν.

Όπως αναφέρθηκε προηγουμένως, τα αποτελέσματα του χάρτη της Εικόνας 5-38 προκύπτουν βάσει των επιλογών του διαχειριστή στο πλαίσιο «Εμφάνιση Χρηστών στο Χάρτη». Στη συγκεκριμένη περίπτωση έχει επιλεχθεί ο αριθμός των χρηστών να είναι 12 και η επιθυμητή περιοχή αφορά την ίδια περιοχή που είναι και η τοποθεσία του event, δηλαδή το

κέντρο της Αθήνας. Έτσι, εμφανίζονται στον χάρτη οι 12 καλύτεροι χρήστες από αυτή την περιοχή που έχουν τις καλύτερες πιθανότητες για να παρακολουθήσουν αυτό το event. Οι χρήστες αυτοί προέρχονται από την λίστα της *Εικόνας 5-37*. Οι άλλες επιλογές της επιθυμητής περιοχής είναι: Κοντινές Περιοχές και Ανεξαρτήτως Περιοχές. Σε περίπτωση που επιλέγαμε μία από αυτές, τα αποτελέσματα μπορεί να ήταν τελείως διαφορετικά και οι 12 καλύτεροι χρήστες να προερχόντουσαν από διαφορετικές περιοχές.

5.2.3 Λειτουργία του Service

Εγκαθιστώντας το πρόγραμμα στο σύστημα, ο χρήστης ερωτείται αν θέλει το Service να τρέχει κατά την εκκίνηση του λειτουργικού. Επιλέγοντας «Ναι», κάθε φορά που ξεκινάει τα Windows, σηκώνεται και το Service και αρχίζει να εκτελεί ένα πλήθος από λειτουργίες σε πραγματικό χρόνο. Το Service πρέπει να αντιλαμβάνεται οποιαδήποτε αλλαγή συμβαίνει στο σύστημα όπως για το αν υπάρχουν οποιεσδήποτε αλλαγές στα δεδομένα της βάσης (εισαγωγή νέων, ενημέρωση και διαγραφή προφίλ χρηστών, διαγραφή manager, καθώς και προσθήκη νέων, ενημέρωση ή διαγραφή event), να ενημερώνει την κατάσταση των αποθηκευμένων events και να τα κατατάσσει σε ενεργά ή ανενεργά και να στέλνει τα ενημερωτικά μηνύματα στους χρήστες που έχουν επιλεγθεί.

Αλλαγή στα Δεδομένα της Βάσης

Για οποιαδήποτε αλλαγή στα δεδομένα της βάσης, το service εκτελεί την κατάλληλη λειτουργία:

- **Εισαγωγή νέου ή ανανέωση προφίλ χρήστη.** Εκτελείται τη διαδικασία ταξινόμησης για το συγκεκριμένο χρήστη για όλα τα ενεργά events που υπάρχουν στο σύστημα.
- **Διαγραφή χρήστη.** Διαγράφονται από το σύστημα όλα τα δεδομένα που έχουν να κάνουν με αυτόν (Ranking, Feedback κ.λ.π.).
- **Εισαγωγή νέου ή ανανέωση event.** Εκτελείται τη διαδικασία ταξινόμησης για το συγκεκριμένο event για όλους τους χρήστες του συστήματος.
- **Διαγραφή Event.** Διαγράφονται από το σύστημα όλα τα δεδομένα που έχουν να κάνουν με αυτό (Ranking, Feedback κ.λ.π.).
- **Διαγραφή Manager.** Διαγράφονται από το σύστημα όλα τα events που σχετίζονται με αυτόν.

Ενημέρωση Κατάστασης των Events

Καθημερινά ελέγχει τις διάρκειες των events και τα κατατάσσει σε ενεργά και ανενεργά. Η ταξινόμηση συνεχίζει να εφαρμόζεται μόνο στα ενεργά event του συστήματος. Τα ανενεργά παραμένουν στο σύστημα για χρήση από τους managers, κυρίως για στατιστικούς λόγους.

Αποστολή Ενημερωτικών Μηνυμάτων

Κάθε φορά που εφαρμόζεται ο αλγόριθμος ταξινόμησης για κάθε χρήστη ή event, αποστέλλει μήνυμα ειδοποίησης στο χρήστη για τα events που επιλέχθηκε ή στους χρήστες που επιλέχθηκαν για το event αντίστοιχα. Για την ώρα έχει υλοποιηθεί η αποστολή μηνυμάτων μέσω email από τον mail server που έχει εγκατασταθεί στο ίδιο μηχάνημα με τον web server. Πολύ σημαντική θα ήταν η υλοποίηση αποστολής μηνυμάτων μέσω SMS, για την παροχή καλύτερων υπηρεσιών από το σύστημα, όπως αναφέρεται στο έβδομο κεφάλαιο.

6 ΑΞΙΟΛΟΓΗΣΗ ΣΥΣΤΗΜΑΤΟΣ

Στην ενότητα αυτή παρουσιάζονται τα τεστ που πραγματοποιήθηκαν για τη μέτρηση των χρόνων εκτέλεσης των αλγορίθμων, καθώς επίσης και δοκιμαστικά σενάρια (test cases) για να προσδιοριστεί η ακρίβεια των αποτελεσμάτων του ταξινομητή.

6.1 Απόδοση Συστήματος

Τα ακόλουθα τεστ εξετάζουν όλες τις παραμέτρους που μπορεί να επηρεάσουν τους χρόνους εκτέλεσης των αλγορίθμων, όπως το μέγεθος των δεδομένων που είναι αποθηκευμένο στη βάση δεδομένων του συστήματος, καθώς επίσης και η επιλογή των προτιμήσεων που ποικίλει τόσο για τους χρήστες και για τα events. Το σύστημα που χρησιμοποιήθηκε έχει τα ακόλουθα χαρακτηριστικά:

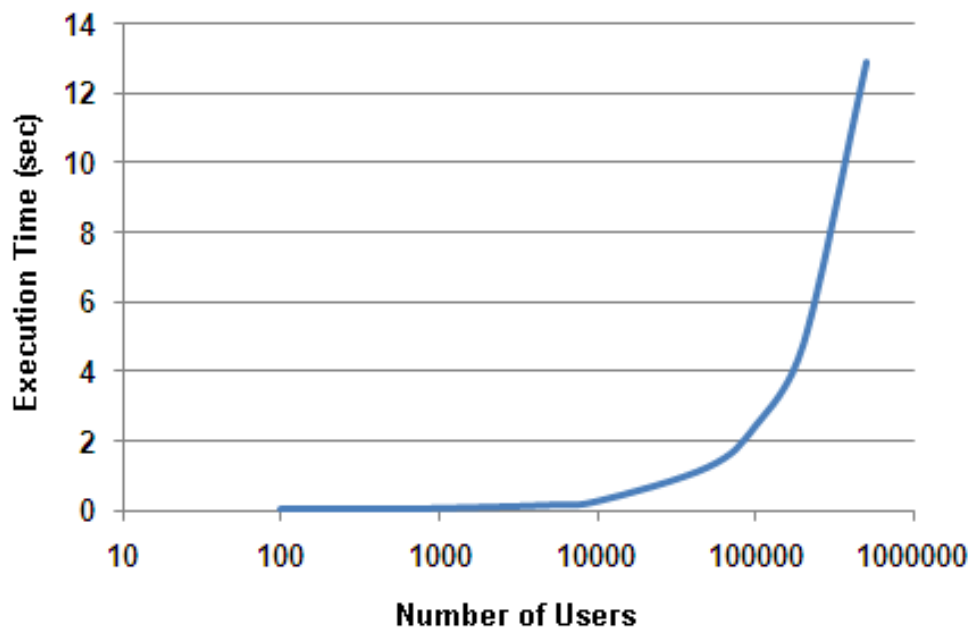
- CPU: Pentium 4 @ 3Ghz
- RAM: 4GB DDR2 @ 800Mhz
- HDD: WD Velociraptor 150GB, 10000rpm, 16MB Cache
- Operating System: Windows XP 32bit Enterprise Edition με SP3
- Oracle 11g Express Edition

Στα τεστ που πραγματοποιήθηκαν εξετάζεται αν οι χρόνοι εκτέλεσης τους συστήματος επηρεάζονται: α) από την αύξηση του αριθμού των χρηστών που εγγράφονται στο σύστημα, β) όταν επιλέγονται χρήστες από όλο το χάρτη κάλυψης ή από μία συγκεκριμένη περιοχή του και γ) όταν υπάρχουν περιορισμοί που αφορούν το φύλο και την ηλικία. Για αυτές τις δοκιμές και προκειμένου να αξιολογηθεί η απόδοση του συστήματος, χρησιμοποιήθηκαν τεχνητά αλλά και πραγματικά δεδομένα τόσο για τους χρήστες όσο και για τα events.

Ο Πίνακας 6-1 και το Σχήμα 6-1 παρουσιάζουν τους μέσους χρόνους εκτέλεσης όταν ζητούνται χρήστες μέσα από ένα εύρος των 100 και 500000 χρηστών που πληρούν τις προϋποθέσεις για ένα συνολικό σετ από 200 events.

Αριθμός Χρηστών	Μέσος Χρόνος Εκτέλεσης (sec)
100	0,01
1000	0,03
10000	0,2
100000	2,2
500000	12,1

Πίνακας 6-1: Χρόνοι Εκτέλεσης (sec)



Σχήμα 6-1: Διάγραμμα Χρόνων Εκτέλεσης

Ο Πίνακας 6-2 παρουσιάζει τα αποτελέσματα των μέσων χρόνων εκτέλεσης για την επιλογή χρηστών μέσα από 500000 που είναι αποθηκευμένοι στο σύστημα, τόσο για ολόκληρο τον χάρτη κάλυψης όσο και για μία συγκεκριμένη περιοχή του, που πληρούν τις προϋποθέσεις για 200 events.

Περιοχή Κάλυψης	Μέσος Χρόνος Εκτέλεσης (sec)
Ολόκληρος Χάρτης	13,8
Συγκεκριμένη Περιοχή	12,1

Πίνακας 6-2: Χρόνοι Εκτέλεσης (sec) Βάσει Περιοχής Κάλυψης

Ο Πίνακας 6-3 παρουσιάζει τα αποτελέσματα των μέσων χρόνων εκτέλεσης για την επιλογή χρηστών μέσα από 500000 που είναι αποθηκευμένοι στο σύστημα, για 200 events, εφαρμόζοντας ένα σύνολο από διάφορους περιορισμούς (π.χ. φύλο και ηλικία). Συγκρίνοντας τα αποτελέσματα με και χωρίς τους περιορισμούς, το ποσοστό μείωσης είναι 62%.

Περιορισμοί	Μέσος Χρόνος Εκτέλεσης (sec)
Ναι	4,6
Όχι	12,1

Πίνακας 6-3: Χρόνοι Εκτέλεσης (sec) Βάσει Περιορισμών

6.2 Απόδοση Ταξινομητή

Το προτεινόμενο σύστημα πληροφοριών δοκιμάστηκε με πραγματικά δεδομένα 218 χρηστών με γνωστές προτιμήσεις event για τις κατηγορίες Art, Sports, Science, Recreation και Computers. Επιπλέον, το σύνολο δοκιμών αποτελείτο από 46, 42, 39, 47 και 44 χρήστες που δήλωσαν, μέσω ενός ερωτηματολογίου, ισχυρή προτίμηση για τις κατηγορίες A, B, C, D και E αντίστοιχα. Ο Πίνακας 6-4 σχηματίζει το πίνακα σύγχυσης (confusion matrix) των δοκιμών μεταξύ των πέντε κατηγοριών. Τα διαγώνια κελιά αντιστοιχούν στους σωστά ταξινομημένους χρήστες για κάθε κατηγορία αντίστοιχα (απόλυτες τιμές), ενώ τα άλλα κελιά δείχνουν τους χρήστες που ταξινομήθηκαν εσφαλμένα. Η ακρίβεια του συστήματος από άποψη σωστής αναγνώρισης αναγράφεται στην τελευταία στήλη του Πίνακα 6-4.

Για παράδειγμα, έχοντας ένα συνολικό αριθμό 46 χρηστών που έχουν δηλώσει την προτίμησή τους για την κατηγορία A (Art), το σύστημα ταξινομήσε σωστά 39 από αυτούς ως περιπτώσεις τις κατηγορίας A, ενώ οι υπόλοιποι 7 χρήστες ταξινομήθηκαν εσφαλμένα στις κατηγορίες B (Sports), C (Science), D (Recreation) και E (Computers). Αυτό σημαίνει ότι από τη μία πλευρά το σύστημα ταξινομήσε εσφαλμένα τις προτιμήσεις των 7 χρηστών στις κατηγορίες αυτές, ενώ από την άλλη, σχεδόν το 85% των υπό δοκιμή χρηστών αναγνωρίστηκαν σωστά από το σύστημα (Ακρίβεια = $39/46 = 0,848$). Ως αποτέλεσμα, η συνολική απόδοση του συστήματος περιγράφει την ικανότητα να εντοπίζει σωστά τους χρήστες σε σχέση με τις προτιμήσεις τους για τις πέντε κατηγορίες. Η συνολική ακρίβεια για όλες τις κατηγορίες, είναι το συνολικό ποσό των σωστά ταξινομημένων περιπτώσεων που βρίσκονται στα διαγώνια κελιά προς όλους τους χρήστες του συνόλου των δοκιμών (Συνολική Απόδοση = 84,4%).

Ομοίως με όλα τα παραπάνω, ο Πίνακας 6-4 σχηματίζει τον πίνακα σύγχυσης (confusion matrix) δοκιμών μεταξύ των πέντε υποκατηγοριών της κατηγορίας A (Architecture, Cinema, Museum, Music και Theater). Τα διαγώνια κελιά αντιστοιχούν στους σωστά ταξινομημένους χρήστες για κάθε υποκατηγορία αντίστοιχα, ενώ η ακρίβεια του συστήματος όσον αφορά τη σωστή ταυτοποίηση πάνω στο γνωστό σύνολο δοκιμών μετρήθηκε στο 63%. Αυτό σημαίνει πως αν αγνοήσουμε την πρώτη ταξινόμηση του συστήματος για τις πέντε κατηγορίες (A, B, C, D και E) και ελέγξουμε τις περιπτώσεις που γνωρίζουμε ότι ανήκουν στην κατηγορία A, τότε η ακρίβεια του συστήματος είναι 63%. Κατά μία έννοια, η μέτρηση αυτή καθορίζει την ευαισθησία του συστήματος όσον αφορά την ακρίβεια στις πέντε ξεχωριστές υποκατηγορίες της κατηγορίας A, η οποία είναι χωρίς αμφιβολία πιο δύσκολο έργο, δεδομένου ότι οι σημασιολογική επικάλυψη είναι μεγαλύτερη.

		Classifier Decision					Total	Precision
		A	B	C	D	E		
Actual Class	A	39	2	1	3	1	46	0,848
	B	1	36	1	3	1	42	0,857
	C	1	0	34	0	4	39	0,872
	D	4	3	0	38	2	47	0,809
	E	0	2	4	1	37	44	0,841
Total		45	43	40	45	45	218	0,844
Pi		0,867	0,837	0,850	0,844	0,822	Pi	
		Classifier Decision					Total	Precision
		a1	a2	a3	a4	a5		
Actual Class	a1	4	0	2	0	1	7	0,571
	a2	1	6	0	1	3	11	0,545
	a3	1	1	3	0	1	6	0,500
	a4	0	1	0	10	2	13	0,769
	a5	0	2	0	1	6	9	0,667
Total		6	10	5	12	13	46	0,630
Pi		0,667	0,600	0,600	0,833	0,462	Pi	
		Classifier Decision					Total	Precision
		b1	b2	b3	b4	b5		
Actual Class	b1	5	0	2	0	1	8	0,625
	b2	1	5	0	1	3	10	0,500
	b3	1	1	2	0	1	5	0,400
	b4	1	1	0	7	3	12	0,583
	b5	0	2	0	1	4	7	0,571
Total		8	9	4	9	12	42	0,500
Pi		0,625	0,556	0,500	0,778	0,333	Pi	
		Classifier Decision					Total	Precision
		c1	c2	c3	c4	c5		
Actual Class	c1	5	1	2	0	1	9	0,556
	c2	1	6	0	0	2	9	0,667
	c3	0	1	2	0	1	4	0,500
	c4	1	0	0	6	3	10	0,600
	c5	0	1	0	1	5	7	0,714
Total		7	9	4	7	12	39	0,522
Pi		0,714	0,667	0,500	0,857	0,417	Pi	
		Classifier Decision					Total	Precision
		d1	d2	d3	d4	d5		
Actual Class	d1	6	1	1	0	1	9	0,667
	d2	1	7	0	1	2	11	0,636
	d3	1	1	4	0	1	7	0,571
	d4	1	0	0	8	3	12	0,667
	d5	0	1	0	0	7	8	0,875
Total		9	10	5	9	14	47	0,696
Pi		0,667	0,700	0,800	0,889	0,500	Pi	
		Classifier Decision					Total	Precision
		e1	e2	e3	e4	e5		
Actual Class	e1	5	1	1	0	0	7	0,714
	e2	1	7	1	1	1	11	0,636
	e3	1	1	5	0	1	8	0,625
	e4	0	0	0	6	2	8	0,750
	e5	1	1	1	0	7	10	0,700
Total		8	10	8	7	11	44	0,652
Pi		0,625	0,700	0,625	0,857	0,636	Pi	

Πίνακας 6-4: Κατάσταση Εκπαίδευσης του Πίνακα Σύγχυσης (Confusion Matrix) για τις κατηγορίες A, B, C, D και E και για όλες τις υποκατηγορίες τους

6.3 Αξιολόγηση της Απόδοσης του Συστήματος

Στον Πίνακα 6-1 και Σχήμα 6-1 παρατηρούμε ότι η αύξηση του χρόνου εκτέλεσης είναι σχεδόν γραμμική με την αύξηση του αριθμού των χρηστών στην βάση δεδομένων που κυμαίνονται μεταξύ 10000 και 100000. Μετά από αυτό το σημείο τον προσομοιωμένων εγγραφών (που κυμαίνεται μεταξύ 100.000 και 500.000), παρατηρείται μία εκθετική αύξηση στο χρόνο εκτέλεσης του συστήματος που οφείλεται στο αποτέλεσμα της διαχείρισης των SQL statements.

Τα αποτελέσματα του Πίνακα 6-2 παρουσιάζουν το χρονικό διάστημα που χρειάζεται το PNN για να επεξεργαστεί τις περιπτώσεις από το επίπεδο εισόδου μέχρι και την τελική απόφαση. Μία μικρή διαφορά των 1,7 sec εμφανίζεται στους μέσους χρόνους εκτέλεσης, η οποία είναι αναμενόμενη, αφού για όλη την περιοχή (όλοι οι χρήστες) του πλέγματος τοποθεσίας το PNN έχει να επεξεργαστεί περισσότερες περιπτώσεις στο σημείο εισόδου σε αντίθεση με μία συγκεκριμένη περιοχή (λιγότεροι χρήστες). Αυτό δείχνει πως το PNN μπορεί να επεξεργαστεί ένα συγκεκριμένο αριθμό περιπτώσεων κάθε φορά και τροφοδοτώντας το με περισσότερες έχουμε συμφόρηση (bottleneck effect). Από την άλλη πλευρά, αυτό αποδεικνύει ότι ακόμα και αν έχουμε περισσότερες περιπτώσεις στο σημείο εισόδου (500000 αποθηκευμένοι χρήστες), δεν αυξάνεται σημαντικά το φορτίο του συστήματος και ως εκ τούτου δεν επηρεάζεται σημαντικά η απόδοσή του.

Τέλος, Σύμφωνα με τα αποτελέσματα του Πίνακα 6-3, εξαιρώντας τους χρήστες που δεν πληρούν τις προδιαγραφές του event (βάσει των προτιμήσεών τους) έχει σημαντικό αντίκτυπο στο χρόνο εκτέλεσης του αλγορίθμου, δεδομένου ότι ο χρόνος εκτέλεσης μειώθηκε σημαντικά (62%).

6.4 Αξιολόγηση της Απόδοσης του Ταξινομητή

Όσον αφορά την απόδοση του ταξινομητή, η τιμή με την υψηλότερη ακρίβεια παρουσιάστηκε για την κατηγορία C (87%), ενώ παρόμοια επίπεδα ακρίβειας εμφανίστηκαν για τις κατηγορίες A, B και E (84% - 85%). Κατώτερη απόδοση διάκρισης μετρήθηκε για την κατηγορία D (σχεδόν 81%). Σε σχέση με τα ποσοστά εσφαλμένης ταξινόμησης προκύπτουν τα ακόλουθα συμπεράσματα. Για την κατηγορία C το σύστημα παρουσίασε ένα μεγάλο ποσοστό από εσφαλμένες περιπτώσεις ταξινόμησης στην κατηγορία E. Ομοίως, σύμφωνα με τον Πίνακα 6-4, πολλές περιπτώσεις που ανήκουν στις κατηγορίες A και B χαρακτηρίστηκαν εσφαλμένα ως περιπτώσεις της κατηγορίας D. Τέλος, το ίδιο ισχύει και για την κατηγορία E, δεδομένου ότι τέσσερις περιπτώσεις ταξινομήθηκαν εσφαλμένα στην κατηγορία C, ενώ για

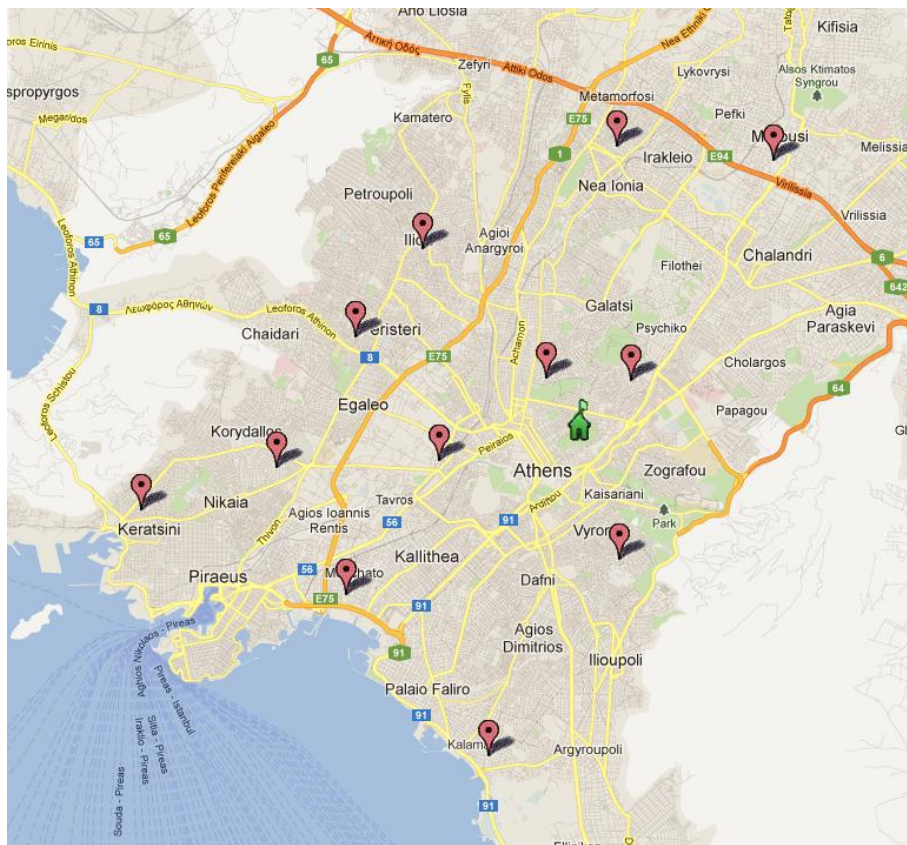
την κατηγορία D, η οποία παρουσιάζει το χαμηλότερο επίπεδο ακρίβειας, οι επτά περιπτώσεις διασκορπίστηκαν στις κατηγορίες A και B. Ωστόσο, αυτό ήταν αναμενόμενο, μιας και όπως διερευνήθηκε κατά την φάση εκπαίδευσης κάποιες περιπτώσεις ήταν πολύ κοντά σε κεντροειδή άλλων κατηγοριών, πράγμα που οφείλεται σε σημασιολογικές επικαλύψεις μεταξύ των διαφορετικών κατηγοριών (όπως απεικονίζεται συγκεκριμένα για τις κατηγορίες A και D στο Σχήμα 4-4b).

Παρά το γεγονός ότι ο πίνακας σύγκρισης της φάσης εκπαίδευσης παρέχει κάποιες μετρήσεις επιδόσεων των δεξιοτήτων ταξινόμησης του προτεινόμενου ταξινομητή, πρέπει όμως να λαμβάνεται υπόψη ότι αυτές οι μετρήσεις προέρχονται από σύνολο χρηστών με προκαθορισμένες προτιμήσεις. Αυτό έχει ως αποτέλεσμα οι πιθανότητες για τη σωστή ταξινόμηση ενός συνόλου άγνωστων χρηστών διαφέρουν. Έτσι, υπό την προϋπόθεση ότι οι προτιμήσεις των χρηστών είναι κατανομημένες ισότιμα μεταξύ των πέντε κατηγοριών και ορίζοντας παράλληλα ότι το $P(x_i / \bar{x}_j)$ περιγράφει την πιθανότητα μιας περίπτωσης προτιμήσεων ενός χρήστη να ταξινομηθεί στην κατηγορία i , ενώ στην πραγματικότητα ανήκει στην j , η Εξίσωση 7 εκτιμά την ακρίβεια της ταξινόμησης σε ένα άγνωστο δείγμα συνόλου περιπτώσεων προτιμήσεων. Ως εκ τούτου, αν το σύστημα προσδιορίσει μία άγνωστη περίπτωση ότι ανήκει στην κατηγορία A, η απόφαση αυτή αναμένεται να είναι κατά 86,7% ακριβής σύμφωνα με την Εξίσωση 7, όπου $P(\bar{x}_i / x_i)$ είναι η πιθανότητα που έχει μία περίπτωση προτιμήσεων να ταξινομηθεί σωστά στην κατηγορία x_i πάνω από ένα γνωστό «από πριν» σύνολο περιπτώσεων που ανήκουν στην κατηγορία x_i , $N(x_i)$ είναι ο πληθυσμός των εξεταζόμενων περιπτώσεων στην κατηγορία x_i σε σχέση με τον πληθυσμό του συνολικού δείγματος, και το $N(\bar{x}_i)$ είναι ο πληθυσμός των εξεταζόμενων περιπτώσεων στην κατηγορία \bar{x}_i σε σχέση με τον πληθυσμό του συνολικού δείγματος. Τέλος, η τελευταία σειρά του Πίνακα 6-4 κατέχει όλες τις αντίστοιχες τιμές πιθανοτήτων του P_i για $i = 1, \dots, 5$.

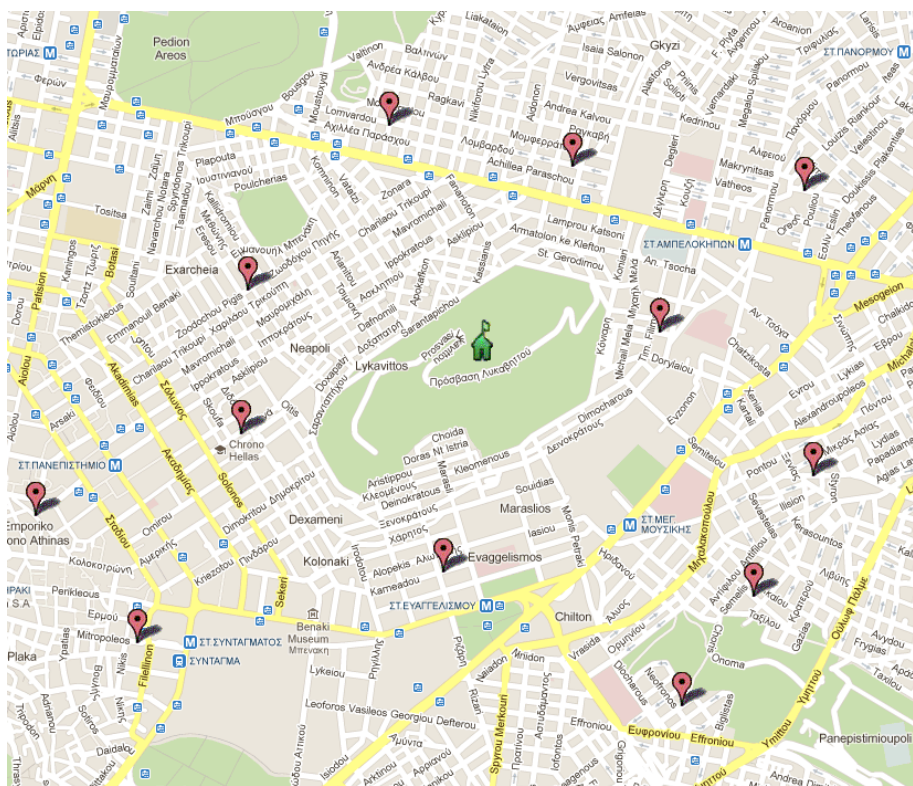
$$P_i = P(x_i / \bar{x}_i) = \frac{P(\bar{x}_i / x_i) \cdot N(\bar{x}_i)}{N(x_i)} \quad 7$$

Σύμφωνα με την ίδια λογική που περιγράφεται παραπάνω, το P_i υπολογίζει την ακρίβεια πρόβλεψης πάνω από ένα άγνωστο δείγμα περιπτώσεων προτιμήσεων για τις υποκατηγορίες a1 έως a5. Όλες οι υπολογιζόμενες (ακρίβεια / πρόβλεψη) τιμές που αφορούν

τις υποκατηγορίες είναι σε σημαντικά χαμηλότερα επίπεδα από τις τιμές των γενικών κατηγοριών A, B, C, D και E (βλ. *Πίνακας 6-4*), αλλά αυτό ήταν αναμενόμενο δεδομένου ότι η διάκριση του κεντροειδή κάθε ομάδας σε κάθε υποκατηγορία είναι πολύ πιο δύσκολο έργο (οι υποκατηγορίες παρουσιάζουν μεγαλύτερη πιθανότητα επικάλυψης). Συγκρίσιμες μετρήσεις εμφανίζονται για τις υποκατηγορίες των A, B, C, D, και E όπως φαίνεται στον *Πίνακα 6-4*. Σύμφωνα με τα παραγόμενα αποτελέσματα, η καλύτερη ακρίβεια στην κατηγορία A (Art) μετρήθηκε για την υποκατηγορία a4 (music), ενώ η χειρότερη για την υποκατηγορία a3 (museum). Ομοίως, τα καλύτερα/χειρότερα επίπεδα ακρίβειας για τις κατηγορίες B (Sports), C (Science), D (Recreation) και E (Computers) μετρήθηκαν για τις υποκατηγορίες b1 (news) / b3 (basketball), c5 (medicine) / c3 (mathematics), d5 (motorcycle) / d3 (entertainment) και e4 (education) / e3 (hardware) αντίστοιχα. Επιπλέον, οι καλύτερες/χειρότερες μετρήσεις ζευγαριών τιμών P_i για όλες τις κατηγορίες ήταν a4 (music) / a5 (theatre), b4 (organizations) / b5 (volleyball), c4 (biology) / c5 (medicine), d4 (car) / d5 (motorcycle), και e4 (education) / e1 (internet) – e3 (hardware) αντίστοιχα. Ορίζοντας ότι το ζευγάρι {ακρίβεια, P_i } ορίζει τις σχέσεις μεταξύ των κατηγοριών και υποκατηγοριών, καταλήγουμε στο αποτέλεσμα ότι από τη μία οι κατηγορίες C (Science) και A (Art) είναι πιο εύκολο να διαχωριστούν (χαλαρά αλληλένδετα), ενώ από την άλλη οι κατηγορίες D (Recreation) και E (Computers) είναι στενά αλληλένδετες. Ομοίως, ζευγάρια χαλαρών/στενών σχέσεων για τις υποκατηγορίες εμφανίζονται για a4 (music) / a3 - a5 (museum - theatre), b1 - b4 (news - organizations) / b3 - b5 (basketball - volleyball), c4 - c5 (biology - medicine) / c3 - c5 (mathematics - medicine), d4 - d5 (car - motorcycle) / d3 - d5 (entertainment - motorcycle) και e4 (education) / e1 - e3 (internet - hardware). Σε όλες τις παραπάνω περιπτώσεις οι καλύτερες/χειρότερες τιμές κελιών επισημαίνονται στον *Πίνακα 6-4* με πράσινο και κόκκινο χρώμα αντίστοιχα.



(a)



(b)

Σχήμα 6-2: Αποτελέσματα με πρώτο ισχυρό κριτήριο: α) τις προτιμήσεις των χρηστών, β) την τοποθεσία των χρηστών

Το Σχήμα 6-2a, δείχνει τους δώδεκα ενδιαφερόμενους χρήστες, σύμφωνα με τον Πίνακα 4-1, σχετικά με ένα event με τα ακόλουθα χαρακτηριστικά:

- Είδος event: Συναυλία
- Κατηγορία: Art (A)
- Υποκατηγορία: Music (a4)
- Τοποθεσία: Πρόσβαση Λυκαβηττού, Αθήνα, 10677

Τα αποτελέσματα δείχνουν μία γεωγραφική διασπορά που οφείλεται στο γεγονός ότι πρόκειται για μία αναζήτηση χρηστών σε ολόκληρο το χάρτη, αλλά η ανάκληση του συστήματος είναι σε υψηλά επίπεδα (10 σωστά ταξινομημένοι χρήστες στην υποκατηγορία a4, καθώς και 2 εσφαλμένα ταξινομημένοι στις υποκατηγορίες a2 και a5. Ωστόσο, η τοποθεσία είναι συνήθως ένα επιθυμητό μέτρο ποιότητας, προκειμένου να εντοπιστούν αποτελεσματικά οι χρήστες σε μία συγκεκριμένη περιοχή αναζήτησης. Έτσι, το Σχήμα 6-2b δείχνει έναν ίσο αριθμό από ενδιαφερόμενους χρήστες που εξαρτώνται σε μεγάλο βαθμό από την απόστασή τους από το event. Αυτό έχει ως αποτέλεσμα, η ανάκληση του συστήματος να είναι σε χαμηλότερα επίπεδα (7 από τους 12 χρήστες που ενδιαφέρονται για το συγκεκριμένο event). Η μικρότερη γεωγραφική διασπορά που επιτυγχάνεται με τον αλγόριθμο, αποδεικνύει ότι στην περίπτωση που είναι επιθυμητός συγκεκριμένος εντοπισμός, το σύστημα μπορεί να συσχετίζει σε αρκετά ικανοποιητικά επίπεδα την τοποθεσία των χρηστών σε σχέση με τις προτιμήσεις τους.

7 ΕΠΙΛΟΓΟΣ

Η παρούσα διατριβή επικεντρώθηκε στην έρευνα και ανάπτυξη αλγοριθμικών τεχνικών ομαδοποίησης και ταξινόμησης, καθώς και εντοπισμού θέσης, με απώτερο σκοπό τη δημιουργία ενός ολοκληρωμένου συστήματος προτάσεων. Το σύστημα που αναπτύχθηκε, εφαρμόστηκε αρχικά για παροχή υπηρεσιών υψηλής προστιθέμενης αξίας μέσω διαδικτύου. Συγκεκριμένα, επιλέγει ανάμεσα από ένα σύνολο χρηστών αυτούς που είναι πιθανό να ενδιαφέρονται για συγκεκριμένα γεγονότα (events) (π.χ. εκθέσεις, αθλητικά, τέχνη, διαφημίσεις κλπ.) από διάφορους διοργανωτές. Στη συνέχεια, παρουσιάζονται μερικές περαιτέρω βελτιώσεις που θα μπορούσαν να γίνουν στο υπάρχον σύστημα, καθώς και οι διάφοροι τομείς που θα μπορούσε να έχει εφαρμογή.

7.1 Μελλοντική Ανάπτυξη του Συστήματος

Το ζητούμενο σε αυτή τη διατριβή ήταν η ανάπτυξη αλγορίθμων φιλτραρίσματος για τη δημιουργία ενός συστήματος προτάσεων, το οποίο κάνει χρήση του εντοπισμού τοποθεσίας για καλύτερα αποτελέσματα. Το πρωτότυπο έδειξε πως παράγει πάρα πολύ καλά αποτελέσματα για το λεκανοπέδιο της Αττικής, οπότε μία σημαντική βελτίωση που θα μπορούσε να γίνει είναι η επέκταση της γεωγραφικής κάλυψης συμπεριλαμβάνοντας περισσότερα πλέγματα (grids) για ολόκληρη την Ελλάδα. Έτσι θα μπορούν να προστεθούν πολύ περισσότεροι χρήστες και events και να δούμε πως θα ανταποκριθεί το σύστημα όταν για ένα event θα πρέπει να επιλέξει ανάμεσα σε χρήστες από όλο το χάρτη (π.χ. μία συναυλία ενός ξένου γνωστού Rock συγκροτήματος στην Αθήνα).

Επίσης θα ήταν πολύ ενδιαφέρον να εξεταστούν και άλλες υπηρεσίες, σχετικές με αυτή που αναπτύχθηκε, που χρησιμοποιούν τα ίδια σύνολα δεδομένων, καθώς και να γίνει περαιτέρω έρευνα πάνω σε άλλες μεθόδους ομαδοποίησης και ταξινόμησης, με αποτέλεσμα να επανασχεδιαστούν οι αλγόριθμοι του συστήματος με διαφορετικό τρόπο, ώστε να συγκριθούν και να σχολιαστούν τα αποτελέσματα.

Παρόλο που το σύστημα αποθηκεύει το feedback των χρηστών τόσο για το αν θα παρακολουθήσουν κάποιο event, όσο και για το πόσο το βαθμολόγησαν, δεν το χρησιμοποιεί για τίποτα άλλο πέρα από να δείχνει στους διοργανωτές των events το ποσοστό παρακολούθησης και βαθμολογίας των events τους. Έτσι, μία επόμενη σημαντική βελτίωση

θα ήταν η υλοποίηση κάποιου αλγόριθμου που θα εκμεταλλεύεται το feedback των χρηστών, ώστε να αναπροσαρμόζονται ανάλογα τα προφίλ τους και να βελτιώνει περισσότερο την απόδοση του συστήματος.

Ακόμα μία πάρα πολύ σημαντική βελτίωση του συστήματος, θα ήταν η επιλογή να προσφέρει συνεχή ανανέωση της τοποθεσίας των χρηστών. Σήμερα, όλοι μας έχουμε από κοντά κάποιο κινητό τηλέφωνο ή tablet PC που έχει ενσωματωμένο GPS tracker. Με την υλοποίηση μιας μικρής εφαρμογής για τις συσκευές αυτές, θα μπορούσε ο χρήστης να ενημερώνει συνεχώς το σύστημα για την τρέχουσα τοποθεσία του. Έτσι, θα μπορούσε να επωφεληθεί από διάφορες προσφορές διαφημιστικών events, ανάλογα με το που βρίσκεται. Για παράδειγμα, ένας χρήστης φεύγει από το σπίτι του και πάει στη δουλειά του. Κοντά στη δουλειά του είναι κάποια καφετέρια, η οποία έχει καταχωρήσει ένα event στο σύστημα, το οποίο διαφημίζει μία προσφορά για πρωινό (καφές και τοστ), αλλά ο διοργανωτής έχει επιλέξει η επιλογή των χρηστών να γίνεται μόνο από τη συγκεκριμένη περιοχή. Φθάνοντας ο χρήστης στην συγκεκριμένη περιοχή, η τοποθεσία του ανανεώνεται στο σύστημα, οπότε συμπεριλαμβάνεται και αυτός πια στη λίστα για την αποστολή μηνύματος για τη συγκεκριμένη προσφορά.

Ακολουθώντας τη μόδα της εποχής, το σύστημα θα μπορούσε να βελτιωθεί κάνοντας χρήση των κοινωνικών δικτύων (social networks) όπως για παράδειγμα το Facebook. Κατά την εγγραφή των χρηστών στο σύστημα, θα μπορούσε να τους ζητηθεί να συνδέσουν το προφίλ τους με το λογαριασμό που μπορεί να έχουν σε κάποιο κοινωνικό δίκτυο. Δεδομένου ότι στα κοινωνικά δίκτυα οι χρήστες κάνουν συχνά επιλογές για διάφορα θέματα, το προφίλ των προτιμήσεων τους δημιουργείται ad-hoc (Huachun et. al. 2010; Booyesen, Zeadally και Rooyen 2011). Μέσω του λογαριασμού αυτού, θα μπορούσαν να αντληθούν αυτόματα επιπλέον πληροφορίες για τις προτιμήσεις τους, όπως περιγράφεται από τους Bonhard και Sasse (2006), Woemdl και Groh (2007), καθώς και Bonhard et. al. (2006), για την απόκτηση πιο αποτελεσματικών εξατομικευμένων συστάσεων.

Τέλος, μία ακόμα επιθυμητή βελτίωση θα ήταν η δημιουργία μιας καλύτερης ιστοσελίδας. Όπως αναφέρθηκε και στο πέμπτο κεφάλαιο, η ιστοσελίδα δημιουργήθηκε παρέχοντας μόνο τα βασικά για τη λειτουργία του συστήματος. Δεν περιλαμβάνει ελέγχους στα πεδία καταχώρησης δεδομένων (π.χ. ημερομηνίες, αριθμούς αντί για γράμματα, λατινικούς χαρακτήρες κλπ.) παρά μόνο αν έχουν εισαχθεί δεδομένα στα υποχρεωτικά πεδία. Θα μπορούσε λοιπόν να σχεδιαστεί μία σύγχρονη ιστοσελίδα που ακολουθεί τους κανόνες του HCI (Human Computer Interaction), ώστε να είναι πιο φιλική προς τους χρήστες

παρέχοντας καλύτερα οργανωμένα αποτελέσματα και περισσότερες επιλογές. Αυτό αποτελεί προϋπόθεση αν η εφαρμογή χρησιμοποιηθεί για εμπορικούς σκοπούς.

7.2 Συζήτηση - Εφαρμογές του Συστήματος σε Διάφορους Τομείς

Το σύστημα που δημιουργήθηκε μπορεί να έχει πάρα πολλές εφαρμογές σε διάφορους τομείς. Αρχικά σχεδιάστηκε σαν ένα σύστημα ενημέρωσης χρηστών σχετικά με διάφορα events, που λαμβάνουν χώρα στην πόλη τους και μπορεί να τους ενδιαφέρει να παρακολουθήσουν. Με μικρές αλλαγές στον τρόπο λειτουργίας του, μπορεί να αποδειχθεί ένα πολύ χρήσιμο εργαλείο στην καθημερινή μας ζωή.

Μία χρήση του θα ήταν στα μέσα μαζικής μεταφοράς (π.χ. Μετρό, Ηλεκτρικός, κ.λ.π.) εισάγοντας στο σύστημα διάφορες στάσεις και σταθμούς, καθώς και το χρονοδιάγραμμα των δρομολογίων τους. Οι χρήστες που έχουν δηλώσει στην υπηρεσία ότι ενδιαφέρονται να λαμβάνουν αυτού του είδους πληροφορία θα ενημερώνονται αυτόματα μέσω SMS για τα δρομολόγια των μέσων μεταφοράς ανάλογα με την περιοχή που βρίσκονται. Έτσι, ένας χρήστης που βρίσκεται στην περιοχή της Αγίας Παρασκευής θα λάβει πληροφορίες σχετικά με την άφιξη και αναχώρηση των επόμενων συρμών στη συγκεκριμένη στάση και θα προγραμματίσει το χρόνο του ανάλογα. Πέρα από αυτό, η χρήση του θα μπορούσε να γίνει και από τα ίδια μέσα μεταφοράς για διαφημιστικούς σκοπούς. Για παράδειγμα, κάνοντας χρήση των οθονών που είναι ήδη εγκατεστημένες στις στάσεις του μετρό, ανάλογα με την περιοχή θα μπορούσαν να διαφημίζονται διάφορες προσφορές. Έστω ότι ένας χρήστης αποβιβάζεται στη στάση Πανόρμου, στις οθόνες μπορούν να διαφημίζονται διάφορες προσφορές με προϊόντα από τα καταστήματα της περιοχής.

Άλλη μία εφαρμογή θα ήταν η χρήση του για την ενημέρωση της κυκλοφορίας. Ανανεώνοντας συχνά το σύστημα με πληροφορίες σχετικά με την κίνηση σε κάθε περιοχή, ο χρήστης θα ενημερώνεται νωρίτερα για την κατάσταση της κυκλοφορίας ανάλογα με τη θέση του και έτσι αν υπάρχει κίνηση στη διαδρομή που ακολουθεί, θα ακολουθήσει εναλλακτική πορεία.

Σημαντική θα ήταν και η εφαρμογή του ως τουριστικός οδηγός. Για παράδειγμα ένας χρήστης βρίσκεται στο ιστορικό κέντρο της Αθήνας και η τοποθεσία του ενημερώνεται ανά τακτά χρονικά διαστήματα μέσω του GPS Tracker του τηλεφώνου του. Το σύστημα εντοπίζει μόλις περάσει κοντά από κάποιο αξιοθέατο και του στέλνει τις σχετικές πληροφορίες με αυτό.

Ακόμα, θα μπορούσε να έχει χρήση και σαν υπηρεσία καιρού. Ενημερώνοντας συχνά τη βάση δεδομένων του συστήματος με μετεωρολογικά δεδομένα, ο χρήστης μπορεί να ενημερώνεται ανάλογα με την τοποθεσία του για τις καιρικές συνθήκες που επικρατούν και έτσι να προγραμματίζει τις δραστηριότητές του ανάλογα (π.χ. αθλητισμός, ψάρεμα κ.λ.π.)

Τέλος η πιο κερδοφόρα εφαρμογή του θα ήταν η εμπορική χρήση του ως διαφημιστική υπηρεσία. Για παράδειγμα θα μπορούσε να χρησιμοποιηθεί από κάποιον πάροχο κινητής τηλεφωνίας για την αποστολή διάφορων διαφημιστικών μηνυμάτων προς τους συνδρομητές του. Οι χρήστες θα μπορούν να εγγράφονται στην υπηρεσία δίχως κόστος και να δημιουργούν τα προφίλ τους. Οι διοργανωτές διαφόρων events και οι διαφημιστές προϊόντων θα καταχωρούν τα events και τις προσφορές τους με κάποιο κόστος. Ο πάροχος γνωρίζει μέσω των κυψελών του τη θέση κάθε συνδρομητή και έτσι ανάλογα με τη θέση τους και τις προτιμήσεις τους θα τους αποστέλλονται και οι ανάλογες πληροφορίες. Η εφαρμογή του αυτή θα ήταν άμεση μιας και δεν έχει μεγάλη απόκλιση από την αρχική υλοποίηση του συστήματος.

7.3 Δημοσιεύσεις

7.3.1 Περιοδικά

- “An Intelligent Web Recommendation System for Ubiquitous Geolocation Awareness”, Boutas, S., Anagnostopoulos, I., Loumos, V., Kayafas, E., Accepted for publication by the Inderscience Publishers at the International Journal of Ad Hoc and Ubiquitous Computing in 2012.
- “Template-guided Inspection of Arbitrarily Oriented Targets”, Alexandropoulos, T., Boutas, S., Loumos, V., Kayafas, E., (2008) IET Computer Vision, 2 (3), pp. 150-163, doi: 10.1049/iet-cvi: 20070025

7.3.2 Πρακτικά Συνεδρίων

- “Real-time Change Detection for Surveillance in Public Transportation”, Alexandropoulos, T., Boutas, S., Loumos, V., Kayafas, E., (2005) IEEE International Conference on Advanced Video and Signal Based Surveillance - Proceedings of AVSS 2005, art. no. 1577243, pp. 58-63

- “A Template-guided Approach to Vehicle Surveillance and Access Control”, Anagnostopoulos, C., Alexandropoulos, T., Boutas, S., Loumos, V., Kayafas, E., (2005) IEEE International Conference on Advanced Video and Signal Based Surveillance - Proceedings of AVSS 2005, art. no. 1577325, pp. 534-539
- “An Intelligent Inspection System for Template-based Change Detection”, Alexandropoulos, T., Boutas, S., Loumos, V., Kayafas, E., (2005) Proceedings of the 14th International Symposium on New Technologies in Measurement and Instrumentation and 10th Workshop on ADC Modeling and Testing, 2, pp. 504-509

8 ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Abidi, S. S. R., & Yu-N, C. (2000). A Convergence of Knowledge Management and Data Mining: Towards " Knowledge-Driven" Strategic Services.
2. Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 734-749.
3. Albert R. and Barabasi A.L., "Statistical Mechanics of Complex Networks", *Reviews of Modern Physics*, 2002, 74, 47.
4. Anagnostopoulos I., Anagnostopoulos C., Kouzas G. and Vergados D., "A Generalised Regression algorithm for web page categorisation", *Neural Computing & Applications journal*, Springer-Verlag, Vol. 13, no. 3, pp. 229 – 236, 2004.
5. Anagnostopoulos I. and Anagnostopoulos C.N., "Cache adaptation of Internet services and user modeling for improving ranking precision in third-party results", *Second International Workshop on Semantic Media Adaptation and Personalization (SMAP 2007)*, pp. 122-127, London, UK, 2007.
6. Andrade L. and Silva M. J., 'Relevance ranking for geographic ir', in *ACM GIR*, 2006.
7. Bar-Yam Y., "Dynamics of Complex Systems", Addison-Wesley, Reading, 1997, MA.
8. Bishop C.M., 'Neural Networks for pattern recognition', Oxford University Press, New York, 1997.
9. Bonhard P., and Sasse M. A., "'Knowing me, Knowing you' – Using Profiles and Social Networking to Improve Recommender Systems", *BT Technology Journal*, Vol. 24, No. 3, 2006a, pp. 84–98
10. Bonhard P., Harries C., McCarthy J. and Sasse M. A., "Accounting for taste: using profile similarity to improve recommender systems", *CHI 2006b Proceedings*, pp 1057-1066

11. Booyesen M. J., Zeadally S., and G.-J. van Rooyen, "Survey of media access control protocols for vehicular ad hoc networks", IET Communications, 2011 - Vol. 5, pp. 1619-1631.
12. Burke, R. (2000). Knowledge-based recommender systems. *Encyclopedia of Library and Information Systems*, 69(Supplement 32), 175-186.
13. Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331-370.
14. Candillier, L., Jack, K., Fessant, F., Meyer, F.: "State-of-the-Art Recommender Systems". In: Chevalier, M., Julien, C., Soule-Dupuy, C. (eds.): *Collaborative and Social Information Retrieval and Access - Techniques for Improved User Modeling*. Idea Group Publishing, 2009, 1-22.
15. Haykin S., "Neural Networks: A Comprehensive Foundation", 1999, Prentice-Hall, New Jersey, USA.
16. Huachun Zhou, Hongbin Luo, Hongke Zhang, Chi-Hsiang Lo, Han-Chieh Chao, "A network-based global mobility management architecture", *Int. J. of Ad Hoc and Ubiquitous Computing* 2010 - Vol. 5, No.1 pp. 1-6.
17. Jong Ki Lee and Christopher Jekeli, Neural Network Aided Adaptive Filtering and Smoothing for an Integrated INS/GPS Unexploded Ordnance Geolocation System, *Journal of Navigation* (2010), 63: 251-267.
18. Kalantidis Y., Tolia G., Avrithis Y., Phinikettos M., Spyrou E., Mylonas Ph. and Kollias S., 'VIRaL: Visual Image Retrieval and Localization', *Multimedia Tools and Applications*, Springer, Volume 51, Number 2, pp. 555-592, February 2011
19. Kawuu W. Lin, Der-Jiunn Deng, "A novel parallel algorithm for frequent pattern mining with privacy preserved in cloud computing environments", *Int. J. of Ad Hoc and Ubiquitous Computing* 2010 - Vol. 6, No.4 pp. 205-215.
20. Kohonen T., *Self-organizing Maps*, Springer, Berlin, 1995
21. Kuai Xu, Feng Wang, Supratik Bhattacharyya, Zhi-Li Zhang, "Real-time behaviour profiling for network monitoring", *Int. J. of Internet Protocol Technology* 2010 - Vol. 5, No.1/2 pp. 65-80.

22. Maciej Rutecki, Tomasz Kacprzak, Use of Neural Networks for Design of Geolocation System in Indoor Applications, International Symposium on Wireless Pervasive Computing 2008 (2008), pp. 451-454.
23. Masters T., 'Advanced Algorithms for Neural Networks', John Wiley: New York, 1995.
24. Mehmet Kayaalp, Tansel Özyer, Sibel Tariyan Özyer, "A Collaborative and Content Based Event Recommendation System Integrated with Data Collection Scrapers and Services at a Social Networking Site," *ASONAM*, pp.113-118, 2009 International Conference on Advances in Social Network Analysis and Mining, 2009.
25. Mitchell T. and Hill M., "Machine Learning", 1997.
26. Parzen E., 'On estimation of a probability density function and mode', of *Mathematical Statistics*, vol.33, pp. 1065-1076, 1962.
27. Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems.
28. Ren-Chiun Wang, Wen-Shenq Juang, Chin-Laung Lei, "A privacy and delegation-enhanced user authentication protocol for Portable Communication Systems", *Int. J. of Ad Hoc and Ubiquitous Computing* 2010 - Vol. 6, No.3 pp. 183-190.
29. Sadaqat Jan, Maozhen Li, Ghaidaa Al-Sultany, Hamed Al-Raweshidy and Ibrar Ali Shah, "Semantic file annotation and retrieval on mobile devices", *Mobile Information Systems*, April 2011 - Vol. 7 Issue 2, pp. 107-122.
30. Schafer, J. B., Konstan, J. A., & Riedl, J. (2001). E-commerce recommendation applications. *Data mining and knowledge discovery*, 5(1), 115-153.
31. Shangguang Wang, Qibo Sun, Hua Zou, Fangchun Yang, "Web Service Selection Based on Adaptive Decomposition of Global QoS Constraints in Ubiquitous Environment", *Journal of Internet Technology*, Vol. 12 No. 5, P.757-768 (2011/9).
32. Shardanand, U., & Maes, P. (1995). Social information filtering: algorithms for automating "word of mouth".
33. Shasha, D., Bonnet, P., & Bercich, N. H. (2004). Database tuning principles, experiments, and troubleshooting techniques. *ACM SIGMOD Record*, 33(2), 116.

34. Silberschatz, A., Korth, H. F., & Sudarshan, S. (2002). Database system concepts: McGraw-Hill New York.
35. Simpson P., “Artificial Neural Systems: Foundation, Paradigms, Applications and Implementations”, Pergamon Press, Elmsford, NY, 1990.
36. Sobecki, J. (2006). Implementations of Web-based Recommender Systems Using Hybrid Methods. International Journal of Computer Science & Applications, 3(3), 52-64.
37. Specht D.F., ‘Probabilistic Neural Networks’, Neural Networks, vol.3, no.1, pp.109-118, 1990.
38. Tan P.N., Steinbach M. and Kumar V., “Introduction to Data Mining”, 2006, Addison-Wesley.
39. Woerndl W. and Groh G., “Utilizing Physical and Social Context to Improve Recommender Systems”, WI-IATW '07 Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology – Workshops, 2007, pp 123 – 128
40. Xing Yi, Hema Raghavan and Chris Leggetter, ‘Discovering user’s specific Geo Intention in Web Search’, WWW Madrid 2009.
41. Yu B. and Cai G., ‘A query-aware document ranking method for geographic information retrieval’, in ACM GIR, pages 49-54, 2007.