



ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΔΙΠΛΩΜΑ ΕΙΔΙΚΕΥΣΗΣ (MSc)
στα ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**«Ανάλυση της χρήσης κοινωνικών ταξινομιών (folksonomies)
για την εξόρυξη μοντέλων χρηστών»**

Τερζάκης Φρίξος

M 3080008

ΑΘΗΝΑ, ΦΕΒΡΟΥΑΡΙΟΣ 2010

Η σελίδα αυτή είναι σκόπιμα λευκή.

ΜΕΤΑΠΤΥΧΙΑΚΟ ΔΙΠΛΩΜΑ ΕΙΔΙΚΕΥΣΗΣ (MSc)
στα ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**«Ανάλυση της χρήσης κοινωνικών ταξινομιών (folksonomies)
για την εξόρυξη μοντέλων χρηστών»**

Τερζάκης Φρίξος

M 3080008

Επιβλέπων Καθηγητής: Παπαθεοδώρου Χρήστος

Εξωτερικός Κριτής: Κωτίδης Ιωάννης

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΑΘΗΝΑ, ΦΕΒΡΟΥΑΡΙΟΣ 2010

(Υπογραφή)

.....

ΦΡΙΞΟΣ ΤΕΡΖΑΚΗΣ

Πτυχιούχος Επιστήμης Υπολογιστών, Πανεπιστημίου Κρήτης

© 2010 – All rights reserved

Περίληψη

Τα τελευταία χρόνια έχει παρατηρηθεί ότι οι επισκέψεις των διαδικτυακών τόπων αποκτούν προστιθέμενη αξία, η οποία δεν οφείλεται στον όγκο της πληροφορίας που μπορεί να βρει κανείς μέσα σε ένα ιστότοπο, αλλά στην ευκολότερη πρόσβαση σε απαιτούμενες πληροφορίες την κατάλληλη στιγμή και με την καταλληλότερη μορφή. Επομένως, η ανάλυση της συμπεριφοράς των χρηστών στο Διαδίκτυο αποτελεί πλέον ένα σημαντικό θέμα, καθώς η εστίαση σε συγκεκριμένες ατομικές ανάγκες και η παροχή προσαρμοσμένων προϊόντων και υπηρεσιών αποτελεί το κινητήριο μοχλό για τη δημιουργία ξεχωριστών υπηρεσιών με ολοένα αυξανόμενη προστιθέμενη αξία.

Λόγω της μεγάλης πληθώρας πληροφοριών στο Διαδίκτυο, τέθηκε η ανάγκη δημιουργίας συστημάτων που παρέχουν υπηρεσίες χαρακτηρισμού, αναζήτησης και ανάκτησης των πόρων που διαθέτουν, μέσω κοινωνικών επισημειώσεων. Για το σκοπό αυτό, δημιουργήθηκαν πλήρη σύνολα επισημειώσεων, βιβλιογραφικά καλούμενα ως κοινωνικές ταξονομίες, τα οποία χρησιμοποιούν οι χρήστες ενός συστήματος διαχείρισης διαμοιραζόμενου περιεχομένου σε ορισμένα κομμάτια του περιεχομένου, με σκοπό να τα ομαδοποιήσουν ή να τα ταξινομήσουν για μελλοντική ανάκτηση. Τα περισσότερα από αυτά τα συστήματα είναι προσβάσιμα από όλους και συνήθως δεν υπάρχει μηχανισμός καταγραφής της ταυτότητας του χρήστη, καθιστώντας την ανάλυση της συμπεριφοράς των χρηστών μια τεχνικά δύσκολη και επίπονη διαδικασία.

Η παρούσα μελέτη μελετά τη χρησιμότητα της κατασκευής κοινοτήτων χρηστών με πανομοιότυπα πρότυπα συμπεριφοράς χρήσης, μέσω της χρήσης αλγορίθμων εξόρυξης γνώσης. Συγκεκριμένα, για την εκπόνηση αυτής της εργασίας απαιτείται η επεξεργασία αρχείων καταγραφής (log files) ενός εξυπηρετητή του παγκόσμιου ιστού που παρέχει τη δυνατότητα χαρακτηρισμού των πόρων που διαθέτει, μέσω κοινωνικών επισημειώσεων. Απώτερος στόχος της επεξεργασίας είναι ο προσδιορισμός ομάδων χρηστών που χρησιμοποιούν κοινές επισημειώσεις για την αναζήτηση και ανάκτηση πόρων, δηλαδή τα μέλη τους παρουσιάζουν παρόμοια συμπεριφορά χρήσης και διαθέτουν κοινά ενδιαφέροντα. Σημαντική έμφαση δίνεται στην Web Usage Mining διαδικασία που αποτελεί τη θεμέλιο λίθο για το σχηματισμό κοινοτήτων χρηστών. Η ανάλυση και η παρουσίαση όλων των σταδίων της διαδικασίας αυτής αποτελεί τη μεθοδολογία που χρησιμοποιείται στις δυο μελέτες περίπτωσης που παρουσιάζονται στην παρούσα εργασία. Συνοπτικά, τα στάδια αυτά αφορούν στην εκκαθάριση των αρχείων καταγραφής (data cleaning), στον εντοπισμό ενεργειών των χρηστών (user sessions) που αναφέρονται στην αναζήτηση πληροφοριών με τη χρήση κοινωνικών επισημειώσεων, στην μετατροπή των επεξεργασμένων δεδομένων χρήσης σε χαρακτηριστικά, στην επιλογή και εκτέλεση αλγόριθμου εξόρυξης γνώσης για το σχηματισμό κοινοτήτων χρηστών και στην αξιολόγηση των κοινοτήτων που δημιουργήθηκαν. Ο αλγόριθμος εξόρυξης γνώσης που χρησιμοποιείται στη συγκεκριμένη μελέτη είναι ένας cluster mining αλγόριθμος. Σκοπός των δύο μελετών περίπτωσης, είναι η δημιουργία μοντέλων κοινοτήτων με κοινά χαρακτηριστικά. Συγκεκριμένα στην πρώτη μελέτη, σχηματίστηκαν κοινότητες από κοινωνικές επισημειώσεις χρηστών, ενώ στην δεύτερη μελέτη σχηματίστηκαν κοινότητες από ενέργειες χρηστών (user sessions) που αναφέρονται στην αναζήτηση πληροφοριών με τη χρήση κοινωνικών επισημειώσεων. Επομένως, και στις δυο περιπτώσεις τα μοντέλα κοινοτήτων αφορούν ομάδες χρηστών με κοινά ενδιαφέροντα και παρόμοια συμπεριφορά χρήσης κατά την αλληλεπίδραση τους με το σύστημα. Επίσης, παρουσιάζονται δύο μετρικές που αφορούν στην κάλυψη και μοναδικότητα (coverage, distinctiveness) των αποτελεσμάτων εξόρυξης, που αποτελούν αντικειμενικά κριτήρια αξιολόγησης των μοντέλων κοινοτήτων που σχηματίζονται, καθώς και βασικό κριτήριο επιλογής του καταλληλότερου μοντέλου για την κάθε μελέτη περίπτωσης ξεχωριστά.

Συνοψίζοντας, η αξιολόγηση των κοινοτήτων που σχηματίζονται κατά την εκτέλεση των πειραμάτων, αποτελεί βασικό ζήτημα της συγκεκριμένης μελέτης, καθώς η χρήση και η ανάλυση αντιπροσωπευτικών και «κατανοητών» κοινοτήτων χρηστών αντικατοπτρίζει σχεδόν πλήρως τις ανάγκες και τα ενδιαφέροντα κάθε τύπου χρήστη μέσα σε μια εφαρμογή.

Λέξεις Κλειδιά: κοινωνικό δίκτυο, κοινωνικές ταξινομίες, επισημειώσεις, Web Personalization, Web Usage Mining, μοντέλα κοινοτήτων

Η σελίδα αυτή είναι σκόπιμα λευκή.

Abstract

In the recent years, the added value for Web site visitors has not only been gained through larger quantity of information data existing on a site, but also through easier access to the required information in exact time and in the most suitable form. Hence, the release of personalized services has emerged, the analysis of user behavior on the Internet has increasingly become an important issue, while focusing specifically on individual needs and providing tailored products and services.

The large diversity of the information, provided on the Internet, introduced the need to provide services which support user identification, searching of terms and retrieval of available resources, through social tagging. For this purpose, a complete set of tags, literally called as social taxonomies, was created, in order that users of a shared content management system are able to group and classify similar type of content for future usage. The popularity of these systems is accessible to everyone and usually there is no method to record the identity of the user. Consequently, the analysis of user behavior becomes a technically difficult and timeconsuming process.

This research aims to study in deep data mining algorithms which compose user communities with similar patterns of services usage. Specifically, we process large log files, retrieved by a web server, which allows identification of the available resources on the web, through social tagging. The purpose of this approach is to identify groups of users who use common tags while searching and accessing resources, i.e. group members who have similar behaviour and interests. Significant emphasis is given on Web Usage Mining process which is the main task for the creation of user communities. This study focuses on the analysis and presentation of all stages proceeded while assessing two case studies. To address the problem, we process and apply data cleaning in log files, then we identify sessions of users who search information by using a social tag and we convert raw data to user features. In addition to all this, we use and evaluate a machine learning algorithm in order to form user communities and assess the whole procedure. In this study we use a cluster mining algorithm.

In both case studies, we model communities with common characteristics. Specifically, in the first case study, communities were formed by user's social tags, while in the second study communities were formed by sessions of users who search information by using a social tag. As a consequence, in both cases community models are equivalent to groups of users who have similar behaviour while interacting with the system. Additionally, we introduce coverage and distinctiveness as metrics in order to evaluate community models and main selection criteria for each model-based case study.

Summarizing, after forming user communities the results show that the appropriate use and analysis of the latter reflect the needs and interests of each user type in a web application. Moreover, the results should be taken into account in order to provide user centered and more personalized services.

Keywords: Social Networks, folksonomies, tags, Web Personalization, Web Usage Mining, Community Models

Η σελίδα αυτή είναι σκόπιμα λευκή.

Ευχαριστίες

Ολοκληρώνοντας τη μεταπτυχιακή μου εργασία, θα ήθελα να ευχαριστήσω όλους όσους βοήθησαν σε όλα τα στάδια ανάπτυξής της. Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή κ. Παπαθεοδώρου Χρήστο, ο οποίος μου έδωσε την ευκαιρία να ασχοληθώ με το θέμα της παρούσας εργασίας και με καθοδήγησε σε όλη τη διάρκειά της. Τον ευχαριστώ θερμά για την εξαιρετική συνεργασία, τις εύστοχες παρατηρήσεις του και για την ποιότητα πνεύματος και προβληματισμού. Ακόμα, θα ήθελα να ευχαριστήσω θερμά την οικογένειά μου για την αμέριστη αγάπη και αφοσίωση, την ηθική υποστήριξη καθώς και την εμπιστοσύνη που δείχνουν σε κάθε επιλογή μου.

Στην οικογένειά μου...

Πίνακας περιεχομένων

1	Εισαγωγή.....	5
1.1	Ανάλυση της χρήσης κοινωνικών ταξινομιών (folksonomies) για την εξόρυξη μοντέλων χρηστών.....	5
1.2	Αντικείμενο διπλωματικής.....	6
1.2.1	Συνεισφορά.....	7
1.3	Σχετικές δραστηριότητες και εργασίες.....	8
1.4	Οργάνωση κειμένου.....	8
2	Θεωρητικό υπόβαθρο	10
2.1	Web 2.0 και κοινωνικά δίκτυα.....	10
2.1.1	Από το Web 1.0 στο Web 2.0.....	10
2.1.2	Τα κοινωνικά δίκτυα και η έννοια της σχέσης.....	12
2.1.2.1	Βασικά χαρακτηριστικά.....	13
2.2	Ιστοτόποι κοινωνικής δικτύωσης.....	14
2.2.1	Εισαγωγή.....	14
2.2.2	Ορισμός.....	14
2.2.3	Ο ρόλος του προφίλ του χρήστη (user profile).....	16
2.2.4	Ο ρόλος των «Φίλων».....	17
2.2.5	Χαρακτηριστικά των ιστοτόπων κοινωνικής δικτύωσης.....	18
2.2.6	Δομή των ιστοτόπων κοινωνικής δικτύωσης.....	18
2.2.7	Ιδιωτικότητα (Privacy).....	19
3	Τεχνολογικό υπόβαθρο	20
3.1	Social Tagging συστήματα	20
3.1.1	Ο ρόλος των tags στα social tagging συστήματα.....	21
3.1.1.1	Ανάλυση επισημειώσεων από το Delicious	21
3.1.1.2	Ανάλυση επισημειώσεων από το Flickr	22
3.1.1.3	Ανάλυση επισημειώσεων από το YouTube.....	23
3.1.2	Μέτρα βελτίωσης των tagging συστημάτων	24
3.1.3	Συμπεράσματα για το προφίλ των κοινωνικών δικτύων.....	26

3.2	Folksonomies	27
3.2.1	<i>Η ανάγκη για κοινωνικές ταξονομίες</i>	27
3.2.2	<i>Ορισμός των κοινωνικών ταξονομιών</i>	28
3.2.3	<i>Η χρήση των επισημειώσεων στις κοινωνικές ταξονομίες.....</i>	29
3.2.4	<i>Το σύννεφο επισημειώσεων.....</i>	30
3.2.5	<i>Κατανομή των επισημειώσεων</i>	30
3.2.6	<i>Αναπαράσταση μοντέλων κοινωνικών ταξονομιών.....</i>	33
3.2.7	<i>Εφαρμογές.....</i>	34
3.2.8	<i>Πλεονεκτήματα.....</i>	36
3.2.9	<i>Μειονεκτήματα.....</i>	37
3.3	Web Personalization	38
3.3.1	<i>Εισαγωγή.....</i>	38
3.3.2	<i>Web Personalization Roadmap.....</i>	40
3.3.2.1	Λειτουργίες του Web Personalization	40
3.3.2.1.1	Απομνημόνευση (Memorization)	40
3.3.2.1.2	Καθοδήγηση	41
3.3.2.1.3	Παραμετροποίηση	41
3.3.2.1.4	Υποστήριξη εκτέλεσης διαδικασιών.....	42
3.3.2.2	Απαιτήσεις για το σχεδιασμό ενός Web Personalization συστήματος	43
3.3.2.3	Προσεγγίσεις στο Web Personalization	44
3.3.3	<i>Ο ρόλος του Web Usage Mining.....</i>	46
4	Μεθοδολογία	49
4.1	Δημιουργία και αξιολόγηση μοντέλων κοινοτήτων.....	49
4.1.1	<i>Εισαγωγή.....</i>	49
4.1.2	<i>Συλλογή και προ-επεξεργασία των δεδομένων</i>	49
4.1.3	<i>Εύρεση προτύπου.....</i>	51
4.1.3.1	Ο αλγόριθμος Cluster Mining	51
4.1.4	<i>Μετα-επεξεργασία προτύπου και αξιολόγηση</i>	53
4.1.4.1	Κατασκευή μοντέλων κοινοτήτων	53
4.1.4.2	Κριτήρια αξιολόγησης	54
4.2	Τεχνολογίες που χρησιμοποιήθηκαν.....	55
5	Υλοποίηση	57

5.1	Μελέτη περίπτωσης Α: Δημιουργία model-based γραφήματος για την εξαγωγή μοντέλων κοινοτήτων	57
5.1.1	Διαδικασία αρχικοποίησης	57
5.1.2	Αποτελέσματα.....	58
5.1.2.1	Κοινότητες με ελάχιστο κατώφλι 0.95	58
5.1.2.2	Κοινότητες με ελάχιστο κατώφλι 0.97	63
5.1.2.3	Κοινότητες με ελάχιστο κατώφλι 1.0	67
5.1.3	Αξιολόγηση αποτελεσμάτων	68
5.2	Μελέτη περίπτωσης Β: Δημιουργία item-based γραφήματος για την εξαγωγή μοντέλων κοινοτήτων	68
5.2.1	Διαδικασία αρχικοποίησης	68
5.2.2	Αποτελέσματα.....	69
5.2.2.1	Κοινότητες με ελάχιστο κατώφλι 0.95	69
5.2.2.2	Κοινότητες με ελάχιστο κατώφλι 0.97	72
5.2.2.3	Κοινότητες με ελάχιστο κατώφλι 1.0	75
5.2.3	Αξιολόγηση αποτελεσμάτων	75
6	Επίλογος	77
6.1	Σύνοψη και συμπεράσματα.....	77
6.2	Μελλοντικές επεκτάσεις	78
7	Βιβλιογραφία.....	79

1

Εισαγωγή

1.1 Ανάλυση της χρήσης κοινωνικών ταξινομιών (folksonomies) για την εξόρυξη μοντέλων χρηστών

Ο διαχωρισμός των δεδομένων και των πληροφοριών στο Διαδίκτυο, γίνεται ολοένα και πιο σημαντικός καθώς ο όγκος των διαθέσιμων πληροφοριών καθιστά αδύνατη τη διαχείρισή τους από μη έμπειρους χρήστες. Τα τελευταία χρόνια, έχουν εμφανιστεί στην αγορά αρκετοί πάροχοι, με σκοπό να βοηθήσουν τους χρήστες να διαχωρίσουν τις πληροφορίες που χρειάζονται μέσα από την πληθώρα των δεδομένων που υπάρχουν στο Διαδίκτυο. Ως αποτέλεσμα, νέες τεχνολογικά προσεγγίσεις, όπως *electronic customer relationship management (e-CRM)*, *Web usage analysis* και *Web mining* ήρθαν στο προσκήνιο και αναπτυχθήκαν ιδιαίτερα τα τελευταία χρόνια. Ο στόχος όλων αυτών των διαδικασιών είναι η κατανόηση των αναγκών, των ενδιαφερόντων και των γνώσεων των χρηστών κάθε εφαρμογής στο Διαδίκτυο. Το κίνητρο προς αυτή την κατεύθυνση, πηγάζει από το γεγονός ότι η προστιθέμενη αξία που έχουν οι επισκέψεις των διαδικτυακών τόπων, δεν έχει αποκτηθεί μέσω των μεγαλύτερων ποσοτήτων δεδομένων που μπορεί να βρει κανείς μέσα σε ένα ιστότοπο, αλλά μέσω της ευκολότερης πρόσβασης στις απαιτούμενες πληροφορίες την κατάλληλη στιγμή και με την πιο κατάλληλη μορφή.

Επιπλέον, η εμφάνιση νέων ηλεκτρονικών υπηρεσιών, όπως το ηλεκτρονικό εμπόριο, το e-learning και το e-banking, έχουν αλλάξει ριζικά τον τρόπο με τον οποίο χρησιμοποιείται το Διαδίκτυο, μετατρέποντας τις ιστοσελίδες σε ηλεκτρονικές επιχειρήσεις, αυξάνοντας παράλληλα τον ανταγωνισμό μεταξύ τους. Με τους ανταγωνιστές να είναι «ένα κλικ μακριά», η απαίτηση για ηλεκτρονικές υπηρεσίες προστιθέμενης αξίας έχει καταστεί αναγκαία συνθήκη μπροστά στο φαινόμενο δημιουργίας επισκεπτών-πελατών σε μια δικτυακή τοποθεσία. Η εστίαση σε συγκεκριμένες ατομικές ανάγκες και η παροχή προσαρμοσμένων προϊόντων και υπηρεσιών αποτελεί το κινητήριο μοχλό για τη δημιουργία ξεχωριστών συστημάτων με αυξημένη προστιθέμενη αξία.

Το μεγαλύτερο μέρος των προαναφερθέντων υπηρεσιών έχουν εγγεγραμένους χρήστες-πελάτες στο συστημά τους και προσαρμόζονται πιο εύκολα στις ανάγκες του, καθώς

κάθε αλληλεπίδραση του χρήστη με το σύστημα συνδέεται μονοσήμαντα με ένα συγκεκριμένο κωδικό (user code). Ωστόσο, οι περισσότερες υπηρεσίες στο Διαδίκτυο είναι προσβάσιμες από όλους και συνήθως δεν υπάρχει μηχανισμός καταγραφής της ταυτότητας του χρήστη. Ιδιαίτερα, με την ανάπτυξη του Web 2.0 και των κοινωνικών δικτύων τα τελευταία χρόνια, έχουν δημιουργηθεί πολλά συστήματα που χρησιμοποιούν τις τεχνολογίες αυτές και είναι προσβάσιμα από όλους. Τα περισσότερα από αυτά τα συστήματα, παρέχουν υπηρεσίες κοινωνικών επισημειώσεων (social tags) για την περιγραφή, αναζήτηση και δημιουργία πληροφοριών και πόρων στο Διαδίκτυο. Λόγω της πληθώρας των επισημειώσεων που υπάρχουν στο Διαδίκτυο, τέθηκε η ανάγκη δημιουργίας ενός πλήρους συνόλου επισημειώσεων (κοινωνικές ταξονομίες), όπου οι χρήστες ενός συστήματος διαχείρισης διαμοιραζόμενου περιεχομένου θα χρησιμοποιήσουν σε ορισμένα κομμάτια περιεχομένου, με σκοπό να τα ομαδοποιήσουν ή να τα ταξινομήσουν προς μελλοντική ανάγκη.

Με αυτό τον τρόπο, τα συστήματα κοινωνικών ταξονομιών επιτρέπουν στους χρήστες να επισημειώσουν δημόσια και να διαμοιραστούν το περιεχόμενο, έτσι ώστε όχι μόνο να το κατηγοριοποιούν για τον εαυτό τους, αλλά και να φυλλομετρούν ευκολότερα τις πληροφορίες που έχουν κατηγοριοποιηθεί από άλλους. Συμπερασματικά, οι επισημειώσεις αποτελούν τον ακρογωνιαίο λίθο τέτοιου είδους συστημάτων και η ανάλυση του σκοπού της χρήσης τους αποτελεί βασικό ζήτημα για τη δημιουργία στοχευμένων υπηρεσιών προς τους τελικούς χρήστες.

1.2 Αντικείμενο διπλωματικής

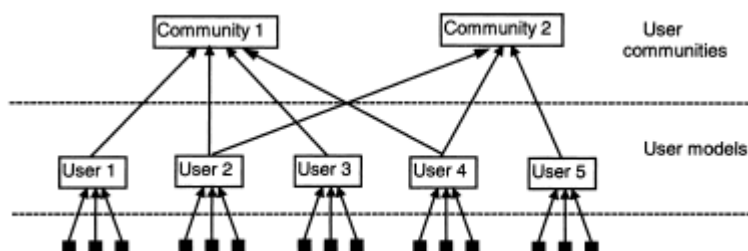
Η συγκεκριμένη διπλωματική εργασία αφορά στην «Ανάλυση της χρήσης κοινωνικών ταξονομιών (folksonomies) για την εξόρυξη μοντέλων χρηστών» και στόχος της είναι η υλοποίηση αλγορίθμων εξόρυξης γνώσης από αρχεία καταγραφής δικτυακών τόπων που παρέχουν υπηρεσίες κοινωνικών επισημειώσεων (social tags), για το σχηματισμό κοινοτήτων χρηστών.

Μια κοινότητα αντιστοιχεί σε μια ομάδα χρηστών που εμφανίζουν παρόμοια συμπεριφορά χρήσης κατά την αλληλεπίδραση τους με το σύστημα. Αυτό το είδος μοντέλου χρήστη έχει ως πρωταρχικό στόχο τη δημιουργία στοχευμένων διαδικτυακών υπηρεσιών, δίχως να χρειάζεται να αναλυθεί το περιεχόμενό τους. Σε αντίθεση με προηγούμενες προσεγγίσεις (memory-based collaborative filtering), η νέα αυτή διαδικασία, γνωστή και ως *model-based collaborative filtering*, οργανώνει τους χρήστες σε ομάδες, δίχως να καταγράφει πληροφορίες σχετικές με αυτούς μεμονωμένα.

Η εκπόνηση της παρούσας μελέτης απαιτεί την ανάλυση και επεξεργασία αρχείων καταγραφής (log files) ενός εξυπηρετητή του παγκόσμιου ιστού που παρέχει τη δυνατότητα χαρακτηρισμού των πόρων που διαθέτει, μέσω κοινωνικών επισημειώσεων. Στόχος της επεξεργασίας είναι να προσδιοριστούν ομάδες χρηστών που χρησιμοποιούν κοινές επισημειώσεις για την αναζήτηση και ανάκτηση πόρων, δηλαδή τα μέλη τους παρουσιάζουν παρόμοια συμπεριφορά χρήσης και διαθέτουν κοινά ενδιαφέροντα. Ένας δικτυακός τόπος που παρέχει υπηρεσίες κοινωνικής επισημείωσης είναι αυτός της Βιβλιοθήκης του Παντείου Πανεπιστημίου (<http://library.panteion.gr/oracial>). Ο δικτυακός τόπος αναπτύχθηκε από μέλη του Παντείου Πανεπιστημίου και της Μονάδας Ψηφιακής Επιμέλειας του Ινστιτούτου Πληροφοριακών Συστημάτων και Προσομοίωσης του Ερευνητικού Κέντρου «Αθήνα».

Η μεθοδολογία που χρησιμοποιείται για το σχηματισμό κοινοτήτων στην παρούσα μελέτη, είναι μια *Web Usage Mining* διαδικασία και αφορά πρωτίστως στην εκκαθάριση των αρχείων καταγραφής (data cleaning) του εξυπηρετητή Παγκόσμιου Ιστού (web server) της

Βιβλιοθήκης του Παντείου Πανεπιστημίου, που αναφέρονται αποκλειστικά στην εφαρμογή κοινωνικής επισημείωσης, και στον εντοπισμό ενεργειών των χρηστών (user sessions) που αναφέρονται στην αναζήτηση πληροφοριών με τη χρήση κοινωνικών επισημειώσεων. Στην συνέχεια, σχηματίζονται ατομικά μοντέλα χρήστη, τα οποία χρησιμοποιούνται στην υλοποίηση των αλγορίθμων εξόρυξης γνώσης για το σχηματισμό κοινοτήτων χρηστών. Αντικειμενικά κριτήρια για την αξιολόγηση των μεθόδων κατασκευής μοντέλων κοινοτήτων αποτελούν δυο μετρικές αξιολόγησης που παρουσιάζονται αναλυτικότερα στο τελευταίο στάδιο της μεθοδολογίας. Τέλος, η παρουσίαση των μοντέλων κοινοτήτων γίνεται με την μορφή πινάκων. Παρακάτω παρατίθεται ένα γενικό σχήμα απεικόνισης της διαδικασίας δημιουργίας κοινοτήτων χρηστών.



Εικόνα 1: Τα δύο επίπεδα κατασκευής μοντέλων χρηστών

Ολοκληρώνοντας, στα πλαίσια της διπλωματικής εργασίας υλοποιούνται δυο διαφορετικές μελέτες περιπτώσεων χρήσης, βάσει της μεθοδολογίας που αναλύθηκε παραπάνω, με σκοπό τη δημιουργία μοντέλων κοινοτήτων με κοινά χαρακτηριστικά. Συγκεκριμένα στην πρώτη μελέτη, σχηματίζονται κοινότητες από κοινωνικές επισημειώσεις χρηστών, ενώ στην δεύτερη μελέτη σχηματίζονται κοινότητες από ενέργειες χρηστών (user sessions) που αναφέρονται στην αναζήτηση πληροφοριών με τη χρήση κοινωνικών επισημειώσεων. Επομένως, και στις δυο περιπτώσεις τα μοντέλα κοινοτήτων που σχηματίζονται, αφορούν σε ομάδες χρηστών με κοινά ενδιαφέροντα και παρόμοια συμπεριφορά χρήσης κατά την αλληλεπίδραση τους με το σύστημα.

1.2.1 Συνεισφορά

Μέσα από την ανακάλυψη κοινοτήτων χρηστών, οποιαδήποτε υπηρεσία στο Διαδίκτυο, μπορεί να αντιμετωπιστεί ως ένα σύστημα βασισμένο σε κοινότητες (community-based system), στο οποίο όλοι οι χρήστες της υπηρεσίας αποτελούν μια μεγάλη κοινότητα, εντός της οποίας μπορούν να εντοπιστούν μικρότερες κοινότητες χρηστών με κοινά χαρακτηριστικά. Αυτά τα χαρακτηριστικά είναι συνήθως μετρήσιμοι παράμετροι της αλληλεπίδρασής τους με το σύστημα, όπως τα άρθρα ειδήσεων που βλέπουν ή τα βιβλία που αγοράζουν, αλλά μπορούν να αντιστοιχούν και σε πληροφορίες που λαμβάνονται ρητά από τους χρήστες, όπως η ηλικία τους ή το επίπεδο των γνώσεών τους. Συνοπτικά, η ανακάλυψη των κοινοτήτων χρηστών μπορεί να διευκολύνει την αλληλεπίδραση του ανθρώπου με ένα σύστημα, όπως είναι μια υπηρεσία στο Διαδίκτυο, με πολλούς τρόπους όπως οι εξής:

- **Βελτιστοποίηση Υπηρεσίας (service optimisation).** Παρέχει ενδείξεις στους παρόχους υπηρεσιών, βοηθώντας τους να αναδιοργανώσουν το σύστημα, ώστε να καταστεί πιο κατάλληλο για τις ανάγκες διαφορετικών τύπων χρηστών.
- **Εξατομίκευση Υπηρεσίας (service personalization).** Βοηθά τους χρήστες να αναγνωρίσουν τις πληροφορίες που τους ενδιαφέρουν.

- **Υποστήριξη Αλληλεπίδρασης (interaction support).** Με δεδομένη μια κατάλληλη υποδομή που προστατεύει την ιδιωτική ζωή των χρηστών, η υπηρεσία μπορεί να υποστηρίξει την αλληλεπίδραση μεταξύ των μελών της κοινότητας.

1.3 Σχετικές δραστηριότητες και εργασίες

Η τεχνολογία μοντελοποίησης χρηστών έχει ως στόχο να καταστήσει τα συστήματα πληροφοριών πιο φιλικά προς το χρήστη, προσαρμόζοντας τα στις ανάγκες του καθενός ξεχωριστά. Με την εκρηκτική ανάπτυξη του Διαδικτύου και τον όγκο των πληροφοριών που δημοσιεύονται σε αυτό, η ανάγκη για προσαρμοστικά συστήματα που θα βοηθούν τους πολίτες να εντοπίσουν πληροφορίες που τους ενδιαφέρουν, έχει καταστεί επιτακτική. Η τεχνολογία μοντελοποίησης χρηστών έχει χρησιμοποιηθεί με επιτυχία σε διάφορους τομείς που σχετίζονται με την πρόσβαση σε πληροφορίες, όπως η ανάκτηση πληροφοριών [27], το φιλτράρισμα [28], η εξόρυξη [29], οι προσαρμοστικές διεπαφές χρήστη [30] και οι προσαρμοστικές τοποθεσίες στο Διαδίκτυο [31].

Κατά κύριο λόγο, ένα μοντέλο χρήστη περιλαμβάνει στοιχεία που χαρακτηρίζουν την αλληλεπίδραση του με το σύστημα και άλλους χρήστες, εφόσον αυτή υποστηρίζεται. Για παράδειγμα, ένα μοντέλο χρήστη για μια υπηρεσία που παρέχει άρθρα ειδήσεων, περιέχει τις προτιμήσεις του χρήστη σχετικά με τα άρθρα αυτά. Επιπλέον, ένα μοντέλο χρήστη μπορεί να περιέχει προσωπικές πληροφορίες για αυτόν, όπως η ηλικία, το επάγγελμα του και άλλα.

Συνήθως, γενικότερα μοντέλα χρήστη, όπως είναι οι κοινότητες, δεν προϋποθέτουν τη ρητή παροχή προσωπικών πληροφοριών των χρηστών. Στην περίπτωση που υπάρχουν διαθέσιμες προσωπικές πληροφορίες χρηστών, τα μοντέλα κοινοτήτων αναφέρονται συχνά ως στερεότυπα. Το στερεότυπο είναι μια από τις πρώτες μορφές γενικών μοντέλων που εμφανίστηκαν σε βιβλιογραφίες [32] και σήμερα χρησιμοποιούνται συχνά σε ψηφιακές βιβλιοθήκες και μουσεία [33]. Τα στερεότυπα συνήθως βασίζονται σε εξωτερικές γνώσεις σχετικά με το χρήστη, όπως το επίπεδο της εμπειρογνωμοσύνης του ή άλλες προσωπικές πληροφορίες. Για παράδειγμα, ένα στερεότυπο μπορεί συνδέσει χρήστες μιας συγκεκριμένης ηλικίας με μια συγκεκριμένη κατηγορία πληροφοριών. Ωστόσο, η συλλογή προσωπικών πληροφοριών θεωρείται δύσκολη, ανακριβής και συχνά ανεπιθύμητη διαδικασία, καθώς αναγκάζει το χρήστη να παρέχει τις πληροφορίες που μπορούν να παραβιάσουν την ιδιωτική του ζωή.

Όπως αναφέρθηκε και παραπάνω, οι διαδικασίες μοντελοποίησης χρηστών χρησιμοποιούν μεθόδους μηχανικής μάθησης κυρίως για την απόκτηση μοντέλων μεμονωμένων χρηστών που αλληλεπιδρούν με ένα σύστημα πληροφοριών [34]. Συγκεκριμένα, ελέγχεται η χρήση του συστήματος από ένα άτομο και τα δεδομένα που συλλέγονται χρησιμοποιούνται για την κατασκευή του μοντέλου χρήστη. Οι τεχνικές αυτές έχουν χρησιμοποιηθεί σε πολλά agent-based και multi-agent συστήματα, όπως το FAB [28], το WebWatcher [35] και το Amalthea [36]. Τέλος, παρόμοιες τεχνικές για τη μοντελοποίηση χρηστών έχουν χρησιμοποιηθεί σε υπηρεσίες ψηφιακής βιβλιοθήκης, όπως το IDL [33] και σε νέα συστήματα φιλτραρίσματος, όπως το News Dude [37].

1.4 Οργάνωση κειμένου

Στο συγκεκριμένο κεφάλαιο, παρουσιάστηκε το αντικείμενο της παρούσας διπλωματικής εργασίας καθώς επίσης έρευνες και εργασίες που έχουν πραγματοποιηθεί στο

παρελθόν σχετικά με αυτό. Επίσης, δόθηκε έμφαση στο πεδίο εφαρμογής που εντάσσεται η παρούσα μελέτη και αναφέρθηκε ποια είναι η συνεισφορά της τόσο σε επίπεδο ευρύτερων εφαρμογών όσο και σε εξειδικευμένο επίπεδο όπως αποτέλεσε αυτό των υπηρεσιών κοινωνικών επισημειώσεων της Βιβλιοθήκης του Παντείου Πανεπιστημίου.

Στη συνέχεια, στο Κεφάλαιο 2 παρουσιάζεται το θεωρητικό υπόβαθρο στο οποίο στηρίχτηκε η παρούσα μελέτη, περιγράφοντας βασικούς όρους και έννοιες, όπως το Web 2.0, τα κοινωνικά δίκτυα και οι ιστότοποι κοινωνικής δικτύωσης. Στο Κεφάλαιο 3 παρουσιάζεται το τεχνολογικό υπόβαθρο της παρούσας μελέτης, στο οποίο γίνεται η περιγραφή των τεχνολογικών χαρακτηριστικών και των λειτουργιών συγκριμένων συστημάτων, όπως τα συστήματα κοινωνικών επισημειώσεων, τα συστήματα κοινωνικών ταξονομιών καθώς και τα Web Personalization συστήματα. Στο Κεφάλαιο 4 αναπτύσσεται η μεθοδολογία για τη δημιουργία και αξιολόγηση μοντέλων κοινοτήτων και παρουσιάζονται οι τεχνολογίες που χρησιμοποιήθηκαν. Στο Κεφάλαιο 5 παρατίθεται η υλοποίηση και τα αποτελέσματα των 2 μελετών περίπτωσης βάσει της μεθοδολογίας που αναλύεται στο Κεφάλαιο 4. Στο Κεφάλαιο 6 συνοψίζονται τα αποτελέσματα της παρούσας μελέτης και περιγράφονται τα συμπεράσματα που προέκυψαν καθώς και τυχούσες μελλοντικές επεκτάσεις. Τέλος, στο Κεφάλαιο 7 παρατίθεται η βιβλιογραφία που χρησιμοποιήθηκε για την περάτωση της παρούσας μελέτης.

2

Θεωρητικό υπόβαθρο

Στην ενότητα αυτή περιγράφονται εις βάθος οι βασικές έννοιες και ορολογίες, στις οποίες στηρίχθηκε η παρούσα μελέτη, με σκοπό την καλύτερη κατανόηση του αναγνώστη πριν από την παρουσίαση περεταίρω τεχνικών λεπτομερειών.

2.1 Web 2.0 και κοινωνικά δίκτυα

2.1.1 Από το Web 1.0 στο Web 2.0

Η έννοια του «Web 2.0» χρησιμοποιήθηκε για πρώτη φορά σε συνέδριο με αντικείμενο την ανταλλαγή ιδεών μεταξύ του εκδοτικού οίκου O'Reilly και του Medialive International. Μέσα από αυτή τη σύσκεψη έγινε φανερό ότι το Web είναι πιο σημαντικό από ποτέ, με εντυπωσιακές νέες εφαρμογές και ιστοτόπους οι οποίοι κάνουν την εμφάνισή τους ραγδαία. Μισό χρόνο μετά, ο όρος Web 2.0 λαμβάνει ξεκάθαρα ολόένα και περισσότερο χώρα, με περισσότερες από 9,5 εκατομμύρια αναφορές στη μηχανή αναζήτησης Google.

Πιο συγκεκριμένα, το νέο Web αλλάζει, επειδή αλλάζει η νοοτροπία των δημιουργών των ιστοτόπων, των προγραμματιστών αλλά και των απλών χρηστών. Το Web 2.0 είναι περισσότερο δημοκρατικό. Ο ρόλος των ισχυρών, παραδοσιακών δημιουργών και «εκδοτών» περιεχομένου αποδυναμώνεται. Η αλληλεπίδραση των χρηστών με το περιεχόμενο καθώς και άλλους χρήστες εντείνεται όλο και περισσότερο. Η νέα χρήση των ήδη υπάρχουσών τεχνολογιών και εργαλείων δίνει καινούριες διαστάσεις και προστιθέμενη αξία στο περιεχόμενο. Οι λέξεις «υλικό» και «λογισμικό» περνάνε σε δεύτερη μοίρα ενώ μια νέα, καθολική πλατφόρμα είναι αυτή που αναδεικνύεται.

Ο νέος Παγκόσμιος Ιστός που βιβλιογραφικά συναντάται ως Web 2.0 ενθαρρύνει τη συμμετοχή των χρηστών και την παραγωγή ενός πλουσιότερου, πιο σύγχρονου και δυναμικότερου περιεχομένου. Επίσης, προσφέρει σε όλους τους χρήστες τη δυνατότητα ανάληψης του ρόλου του δημιουργού και του εκδότη, αφού ταυτόχρονα με τους προγραμματιστές διαδικτύου, και οι απλοί χρήστες είναι σε θέση να δημιουργούν χρησιμοποιώντας τη θέληση και τη φαντασία τους. Παράλληλα με τη διαμόρφωση του

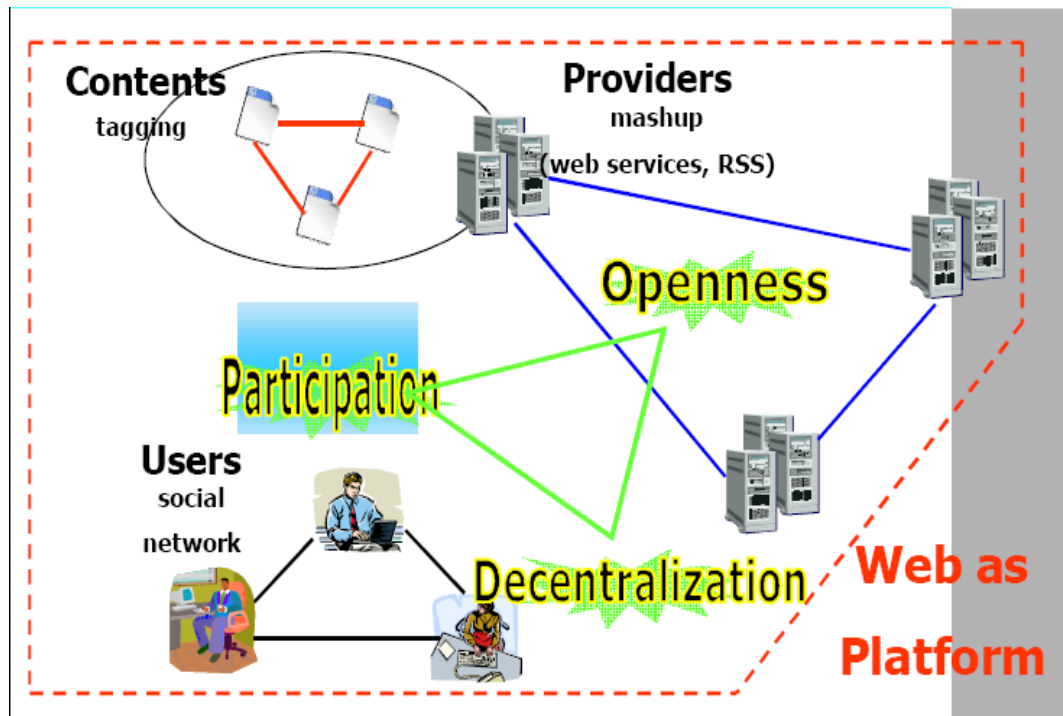
περιεχομένου σε διαφορετικές μορφές, όπως κείμενο, ήχος, εικόνα, βίντεο, επαγίεται στους χρήστες και η δυνατότητα κατηγοριοποίησης, αξιολόγησης και κατάταξης του περιεχομένου, όπως για παράδειγμα ο καθορισμός του ποια είδηση θεωρείται από αυτούς ως η περισσότερο σημαντική.

Το Web2.0 αναφέρεται σε ένα σύνολο νέων διαδικτυακών υπηρεσιών, οι οποίες επιτρέπουν στους χρήστες να συνεργάζονται και να ανταλλάζουν δεδομένα online, με πιο αποδοτικό τρόπο σε σχέση με αυτόν που προσέφεραν οι παλαιότερες υπηρεσίες. Η ειδοποιός διαφορά είναι ότι οι νέες υπηρεσίες παρέχουν στο χρήστη μια εμπειρία που πλησιάζει περισσότερο σε αυτή που έχει όταν εργάζεται στον προσωπικό του υπολογιστή. Με άλλα λόγια, οι εφαρμογές του Web 2.0 μοιάζουν με τις desktop εφαρμογές. Επιπλέον, οι νέοι δικτυακοί τόποι είναι κατά κανόνα «δυναμικοί» και περισσότερο αλληλεπιδραστικοί, διαφέροντας σε σημαντικό βαθμό από το «στατικό» Web1.0.

Το Web 2.0 είναι συμβατό με οποιοδήποτε λειτουργικό σύστημα χρησιμοποιεί ο εκάστοτε χρήστης. Ένας οποιοσδήποτε φυλλομετρητής ιστοσελίδων αρκεί για να συμμετέχει ένα χρήστης στο νέο, πιο ζωντανό και εκπληκτικό διαδίκτυο. Επιπρόσθετα, το Web 2.0 είναι εκτός από πλούσιο και «ελαφρύ», καθώς πολλές από τις εφαρμογές του έχουν σχεδιαστεί για να τρέχουν γρήγορα, χωρίς να βαραίνουν τους πόρους του συστήματος. Τέλος, το λογισμικό και το υλικό δεν απασχολούν πλέον τους προγραμματιστές στον ίδιο βαθμό με το παρελθόν, αφού το περιεχόμενο, η διαμόρφωση και η αξιοποίησή του είναι τα θέματα στα οποία επικεντρώνεται κυρίως το ενδιαφέρον.

Οι χρήστες του Web 2.0 μπορούν να λάβουν βοήθεια από την εγκυκλοπαίδεια Wikipedia, να ανεβάζουν φωτογραφίες στο Flickr, να εκφράζονται στα προσωπικά τους blogs, να σχολιάζουν απόψεις και σκέψεις άλλων χρηστών, να χρησιμοποιούν τα podcast, να ενημερώνονται από ειδήσεις που διαρκώς συγκεντρώνει ο προσωπικός τους RSS συνδρομητής.

Παρόλα αυτά υπάρχει και ένα μεγάλο ποσοστό που αντιμετωπίζει το Web 2.0 με επιφυλακτικότητα. Ενώ κάποιοι το θεωρούν μια καινοτομία, κάποιοι άλλοι το θεωρούν μια λέξη δίχως νόημα, που αποσκοπεί την επικράτηση στην αγορά και στην αύξηση των κερδών, των εφαρμογών και των υπηρεσιών που φέρονται ως αντιπρόσωποι του όρου [1].



Εικόνα 2: Η φιλοσοφία του Web 2.0

2.1.2 Τα κοινωνικά δίκτυα και η έννοια της σχέσης

Όπως ήδη προαναφέρθηκε, το Web 2.0 αποτελείται από μια δεύτερη γενιά υπηρεσιών βασισμένων στο διαδίκτυο, όπως είναι οι ιστότοποι κοινωνικής δικτύωσης (social networking sites), τα wikis, οι διάφορες εφαρμογές και τα εργαλεία επικοινωνίας, που δίνουν έμφαση στην ηλεκτρονική συνέργεια και ανταλλαγή πληροφορίας μεταξύ των χρηστών. Αυτές οι υπηρεσίες, με την παροχή συγκεκριμένων τεχνολογιών δίνουν τη δυνατότητα στους τελικούς χρήστες να έχουν ευκολότερη πρόσβαση σε πηγές καθώς και συμμετοχή σε κοινωνικά δίκτυα (social networks). Τα «κοινωνικά δίκτυα» δεν είναι νέος όρος, αλλά προέρχεται από τις κοινωνικές επιστήμες, που τα ορίζει ως τα νοητά δίκτυα που σχηματίζονται από κοινωνικές σχέσεις (οικογενειακές, πολιτικές, φιλικές) ανάμεσα σε άτομα που συνάπτονται, διατηρούνται και δυναμώνουν ή το αντίθετο, με βάση την ελεύθερη βούληση και επιλογή αυτών [2].

Οι αναλυτές των κοινωνικών δικτύων επιδιώκουν να περιγράψουν τέτοιου είδους δίκτυα σχέσεων όσο το δυνατόν πληρέστερα, να δώσουν έμφαση στα μοντέλα (patterns) που κυριαρχούν στα δίκτυα αυτά, να εντοπίσουν την ροή της πληροφορίας (καθώς και άλλων πόρων) μέσω αυτών και να ανακαλύψουν τι επιπτώσεις έχουν στους ανθρώπους και στους οργανισμούς. Μια τέτοιου είδους προσέγγιση αντικατοπτρίζει τη μετάβαση από τον «κοινό» ατομικισμό που επικρατεί στις κοινωνικές επιστήμες, προς την κατεύθυνση μιας πιο διαρθρωτικής ανάλυσης. Η μέθοδος αυτή προτείνει τον επαναπροσδιορισμό των θεμελιωδών μονάδων ανάλυσης και την ανάπτυξη νέων αναλυτικών μεθόδων. Κυρίαρχη μονάδα ανάλυσης θεωρείται πλέον η «σχέση» (relation), όπως οι σχέσεις συγγένειας μεταξύ των προσώπων, οι δεσμοί επικοινωνίας μεταξύ των μελών ενός οργανισμού, καθώς και το είδος της φιλίας που διαμορφώνεται μέσα από μια μικρή ομάδα. Το μοντέλο που χρησιμοποιείται για τον καθορισμό των σχέσεων αυτών είναι ανεξάρτητο από κοινωνικό-πολιτικά χαρακτηριστικά όπως είναι το φύλο, η ηλικία και το εισόδημα.

Οι σχέσεις χαρακτηρίζονται από το περιεχόμενο, την κατεύθυνση και τη δύναμη. Το περιεχόμενο μιας σχέσης αναφέρεται στον πόρο που ανταλλάσσεται μεταξύ των μελών ενός κοινωνικού δικτύου και μπορεί να διαχειρίζεται διαφορετικά είδη πληροφορίας, όπως ανακοινώσεις σχετικά με διοικητικά, προσωπικά ή ακόμα και κοινωνικά θέματα. Επίσης οι σχέσεις περιλαμβάνουν την αποστολή ενός αρχείου δεδομένων ή ενός προγράμματος ηλεκτρονικού υπολογιστή. Με την άνοδο του ηλεκτρονικού εμπορίου (όπως π.χ στις ηλεκτρονικές τραπεζικές συναλλαγές), οι πληροφορίες που ανταλλάσσονται μέσω τέτοιων δικτύων μπορούν να αντιστοιχούν σε ανταλλαγές χρημάτων, αγαθών ή υπηρεσιών στο «πραγματικό» κόσμο.

Μια σχέση μπορεί να είναι κατευθυνόμενη ή μη κατευθυνόμενη. Για παράδειγμα, ένα άτομο μπορεί να προσφέρει κοινωνική υποστήριξη σε ένα δεύτερο πρόσωπο δημιουργώντας με αυτόν τον τρόπο μια αμφίδρομη σχέση μεταξύ τους, παρέχοντας και λαμβάνοντας ταυτόχρονα κατά αυτόν τον τρόπο υποστήριξη. Εναλλακτικά, μπορεί να υπάρξει σχέση φιλίας χωρίς να υπάρχει συγκεκριμένη κατεύθυνση προς αυτήν. Παρόλα αυτά, η σχέση αυτή μπορεί να μην είναι ισορροπημένη και η επικοινωνία να μην είναι αμφίπλευρη με την ίδια συχνότητα.

Τέλος, οι σχέσεις διαφέρουν και σε δύναμη, η οποία μπορεί να εκφραστεί με διάφορους τρόπους. Όσον αφορά στην επικοινωνία, τα διάφορα μέλη του δικτύου μπορούν να επικοινωνούν κατά τη διάρκεια της ημέρας, μία φορά την ημέρα, εβδομαδιαία ή ετησίως. Μπορούν να ανταλλάσσουν μεγάλες ή μικρές ποσότητες πόρων όπως, χρήματα, αγαθά ή υπηρεσίες. Αυτά τα είδη σχέσεων αντιστοιχούν σε διαφορετικούς τύπους σχεσιακής δύναμης. Πολλοί από τους πιο σημαντικούς τύπους σχέσεων που χρησιμοποιούνται στα κοινωνικά δίκτυα, αφορούν την ανταλλαγή πολύπλοκων ή δύσκολων πληροφοριών, την παροχή επικοινωνίας για την παραγωγή νέων ιδεών (blogs) καθώς και τη δημιουργία ή την υποστήριξη εικονικών κοινοτήτων [3].

2.1.2.1 Βασικά χαρακτηριστικά

Όπως σε όλα τα είδη δικτύων, έτσι και στα κοινωνικά δίκτυα τα βασικά χαρακτηριστικά τους, τα οποία αποτελούν επίσης και βασικό κριτήριο διαφοροποίησής τους, είναι το μέγεθος και η ετερογένειά τους. Σε μεγαλύτερα κοινωνικά δίκτυα παρατηρείται αυξημένη ετερογένεια όσον αφορά στα κοινωνικά χαρακτηριστικά των μελών τους καθώς και μεγαλύτερη πολυπλοκότητα στη δομή τους. Αντιθέτως, τα μικρά ομοιογενή κοινωνικά δίκτυα αποτελούνται από μικρές ομάδες εργασίας (work groups) και κοινότητες οι οποίες είναι κατάλληλες για τη διατήρηση των υφιστάμενων πόρων των εν λόγω δικτύων. Αυτά τα δίκτυα, συχνά αποτελούν για τους ειδικούς τον κανόνα σύγκρισης ανάμεσα σε μικρά δίκτυα συνεργασίας, με την βοήθεια υπολογιστών, και στις εικονικές κοινότητες. Παρόλα αυτά, τα ετερογενή δίκτυα θεωρούνται καταλληλότερα για τη διάθεση νέων πόρων δικτύου.

Επίσης, ένα άλλο βασικό χαρακτηριστικό των κοινωνικών δικτύων αποτελεί το κριτήριο κεντρικής πρόσβασης σε αυτά. Πολλές φορές, θεωρείται απαραίτητο να εξεταστεί ποια οντότητα θεωρείται απομονωμένη ή κεντρική σε δίκτυα που συντηρούνται από διαφορετικά μέσα. Για παράδειγμα, ο διαχειριστής που δεν έχει λογαριασμό e-mail θεωρείται απομονωμένος από το δίκτυο ηλεκτρονικών μηνυμάτων, ενώ κατέχει κεντρικό ρόλο στο οργανωσιακό δίκτυο. Ως αποτέλεσμα, η πληροφορία και τα δεδομένα που θα μεταφερθούν μέσω ηλεκτρονικών μηνυμάτων, δεν θα είναι ποτέ προσβάσιμα στο διαχειριστή, ενώ η πληροφορία που θα ανταλλαχθεί σε οργανωσιακά δίκτυα δεν θα είναι προσβάσιμη σε χαμηλότερου επιπέδου οντότητες. Για το λόγο αυτό, έχει δημιουργηθεί ένας μεσάζων ρόλος που προσπαθεί να συνδέσει τα δυο αυτά δίκτυα, διαβιβάζοντας πληροφορίες από το ένα δίκτυο στο άλλο. Μέσω της ανάλυσης κοινωνικών δικτύων, έχουν αναπτυχθεί μέτρα

κεντρικής πρόσβασης τα οποία μπορούν να χρησιμοποιηθούν για τον εντοπισμό μελών του δικτύου που έχουν τις περισσότερες συνδέσεις με άτομα των οποίων η παραμονή τους στο δίκτυο κρίνεται απαραίτητη και κρίσιμη.

Οι ομοιότητες στη συμπεριφορά των μελών του δικτύου υποδηλώνουν την παρουσία ενός ρόλου δικτύου. Για παράδειγμα, όλοι οι καθηγητές έχουν τον ίδιο διαδικτυακό ρόλο όσον αφορά στους μαθητές τους, παρέχοντας συμβουλές, επιβλέποντας εργασίες και αναθέτοντας βαθμούς. Οι κανονικότητες που παρουσιάζονται στα πρότυπα των σχέσεων σε δίκτυα ή σε συμπεριφορές εντός του δικτύου, επιτρέπουν την εμπειρική αναγνώριση ρόλων μέσα σε ένα δίκτυο.

Τέλος η πυκνότητα του δικτύου θεωρείται ένα από τα πιο ευρέως χρησιμοποιούμενα μέτρα κοινωνικής δομής του δικτύου. Η πυκνότητα ορίζεται ως ο αριθμός των σχέσεων ή των δεσμών που υφίσταται πραγματικά ως ποσοστό των θεωρητικά δυνατών σχέσεων και δεσμών. Σε υψηλής πυκνότητας δίκτυα υπάρχει άμεση επικοινωνία μεταξύ όλων των μελών τους (work groups), ενώ σε χαμηλής πυκνότητας δίκτυα ελάχιστα μέλη επικοινωνούν άμεσα και συχνά. Αυτή η έλλειψη επικοινωνίας οφείλεται στο γεγονός ότι στα δίκτυα χαμηλής πυκνότητας το κάθε μέλος μπορεί να λειτουργεί αυτόνομα και να διαλέγει τι είδους σχέσεις θέλει να έχει, τις οποίες θα πρέπει να διατηρεί ξεχωριστά τη μια από την άλλη. Αντιθέτως στις ομάδες εργασίας όλα τα μέλη λειτουργούν σαν μια οντότητα καθιστώντας με αυτό τον τρόπο πιο ομαλές τις σχέσεις μεταξύ τους [3].

2.2 Ιστοτόποι κοινωνικής δικτύωσης

2.2.1 Εισαγωγή

Οι ιστοτόποι κοινωνικής δικτύωσης (Social Network Sites) όπως είναι οι *MySpace*, *Facebook*, *Cyworld* και *Bebo*, από την αρχή της εμφάνισής τους στο διαδίκτυο, έχουν προσελκύσει εκατομμύρια χρήστες, πολλοί από τους οποίους έχουν ενσωματώσει αυτές τις υπηρεσίες στις καθημερινές τους δραστηριότητες. Υπάρχουν εκατοντάδες ιστοτόποι κοινωνικής δικτύωσης, με διαφορετικές τεχνολογικές δυνατότητες που υποστηρίζουν ένα ευρύ φάσμα δραστηριοτήτων και πρακτικών. Παρόλο, που τα βασικά τεχνολογικά χαρακτηριστικά τους είναι σταθερά, οι κουλτούρες που προκύπτουν γύρω από αυτά ποικίλλουν σε μεγάλο βαθμό. Οι περισσότεροι ιστοτόποι υποστηρίζουν τη διατήρηση της δομής προϋπαρχόντων κοινωνικών δικτύων, αλλά υπάρχουν και άλλοι που βοηθούν άγνωστους χρήστες να συνδέονται σε αυτούς βάσει κοινών συμφερόντων, πολιτικών απόψεων ή δραστηριοτήτων. Επίσης, μερικοί ιστοτόποι προσελκύουν ανθρώπους βάσει κάποιων κοινωνικών χαρακτηριστικών τους όπως η γλώσσα, η φυλετική, σεξουαλική ή θρησκευτική ταυτότητα, καλύπτοντας με αυτό τον τρόπο ένα μεγάλο εύρος κοινωνικών στρωμάτων. Τέλος, οι ιστοτόποι ποικίλλουν ως προς το βαθμό στον οποίο ενσωματώνονται σε αυτούς νέα εργαλεία πληροφορίας και επικοινωνίας, όπως η κινητή συνδεσιμότητα (mobile connectivity), το *blogging* και η διαμοίραση φωτογραφιών και αρχείων video [4].

2.2.2 Ορισμός

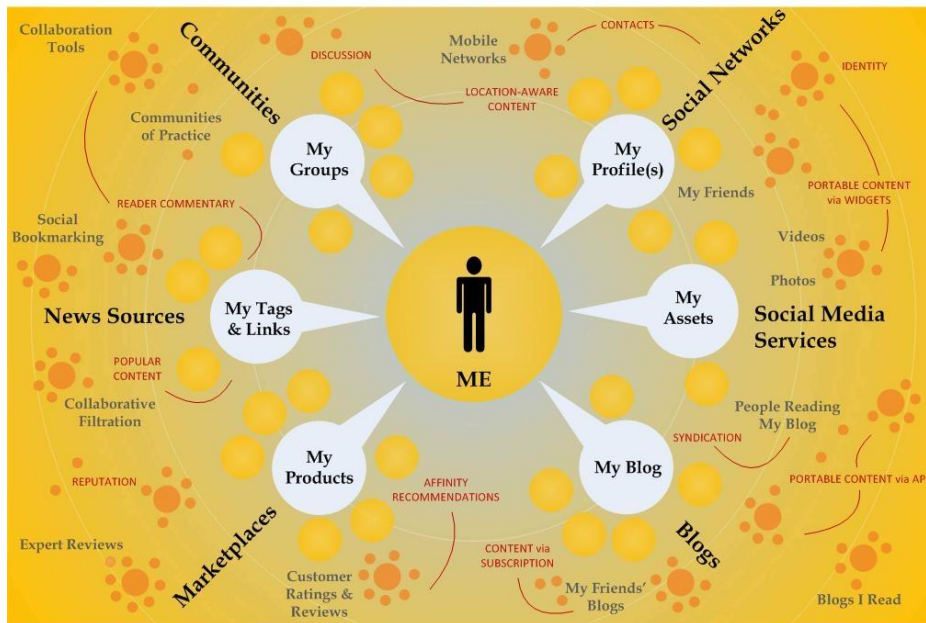
Ως ιστοτόποι κοινωνικής δικτύωσης ορίζονται οι υπηρεσίες διαδικτύου που επιτρέπουν στους χρήστες την κατασκευή ενός δημόσιου ή προσωπικού προφίλ, που οριοθετείται μέσα σε ένα σύστημα. Επίσης, επιτρέπουν τη δημιουργία μια λίστας με όλους τους χρήστες με τους οποίους συνδέονται και την πρόσβαση και εξερεύνηση της λίστας των δικών τους συνδέσεων στην υπηρεσία, καθώς και τις συνδέσεις που πραγματοποιήθηκαν από άλλους χρήστες του

συστήματος. Η φύση και η ονοματολογία αυτών των συνδέσεων μπορεί να ποικίλλουν από ιστότοπο σε ιστότοπο. Οι βασικές τους λειτουργίες περιλαμβάνουν:

- Αποστολή email μέσω του ιστοτόπου
- Δημιουργία ομάδων (groups - διαδικτυακές ομάδες στις οποίες άτομα τα οποία μοιράζονται κοινά ενδιαφέροντα και στις οποίες μπορούν να συμμετέχουν)
- Forums για οποιοδήποτε θέμα
- Κοινή χρήση πολυμεσικού υλικού (video and photo sharing)
- Chat rooms
- Blogs
- Δωρεάν διαφήμιση για οποιονδήποτε και οτιδήποτε (Classified / Marketplace section)
- Δημοσίευση πληροφοριών από μέλη του ιστότοπου και αποστολή σε φίλους τους (Bulletin boards)
- Προσωπικές σελίδες ή προφίλ οι οποίες μπορεί να θεωρηθούν ως portfolio, βιογραφικό ή ημερολόγιο

Σε αντίθεση με τους ιστοτόπους που έχουν σαν σκοπό την ανάπτυξη των ανθρώπινων σχέσεων (Facebook, MySpace) υπάρχουν και ιστότοποι με σκοπό την οργάνωση και διάχυση της πληροφορίας. Οι πιο δημοφιλείς από αυτούς είναι το **del.icio.us** (social bookmarking site), το **Flickr** (digital image storage) και το **Technorati** (αναζήτηση στο περιεχόμενο των ιστολογίων). Μια σειρά βέβαια από εφαρμογές όπως είναι η **Cannotea** ή η **CiteUlike** απευθύνονται στον ακαδημαϊκό χώρο, σε επιστήμονες και ερευνητές, καθώς συνδέονται με αξιόπιστες βάσεις δεδομένων όπως η **PubMed**, ελέγχοντας παράλληλα την χαμηλής ποιότητας πληροφορία επιτρέποντας τη επισήμανση σε συνδέσμους από έγκυρα μόνο περιοδικά.

Η μοναδικότητα των ιστοτόπων κοινωνικής δικτύωσης οφείλεται κυρίως στη δυνατότητα που δίνεται στους χρήστες να συναντούν άλλους άγνωστους χρήστες, να συγκεντρώνουν και να κάνουν ορατά τα δικά τους κοινωνικά δίκτυα σε άλλους. Αυτές οι δυνατότητες, μπορεί να οδηγήσουν πολλές φορές σε συνδέσεις μεταξύ ατόμων οι οποίες διαφορετικά δεν θα γίνονταν, παρόλο που δεν είναι συχνά αυτός ο κύριος στόχος. Στο μεγαλύτερο εύρος των ιστοτόπων κοινωνικής δικτύωσης, οι συμμετέχοντες δεν έχουν αναγκαστικά ως κύρια δραστηριότητα τη «δικτύωση» ή την αναζήτηση καινούργιων χρηστών, αλλά κυρίως την επικοινωνία με τους ανθρώπους που είναι ήδη μέρος του επεκταμένου κοινωνικού τους δικτύου. Οπότε, αυτό το αρθρωτό κοινωνικό δίκτυο θεωρείται το κύριο συστατικό των εν λόγω ιστότοπων [2].



Εικόνα 3: Αποτύπωση ενός ιστοτόπου κοινωνικής δικτύωσης

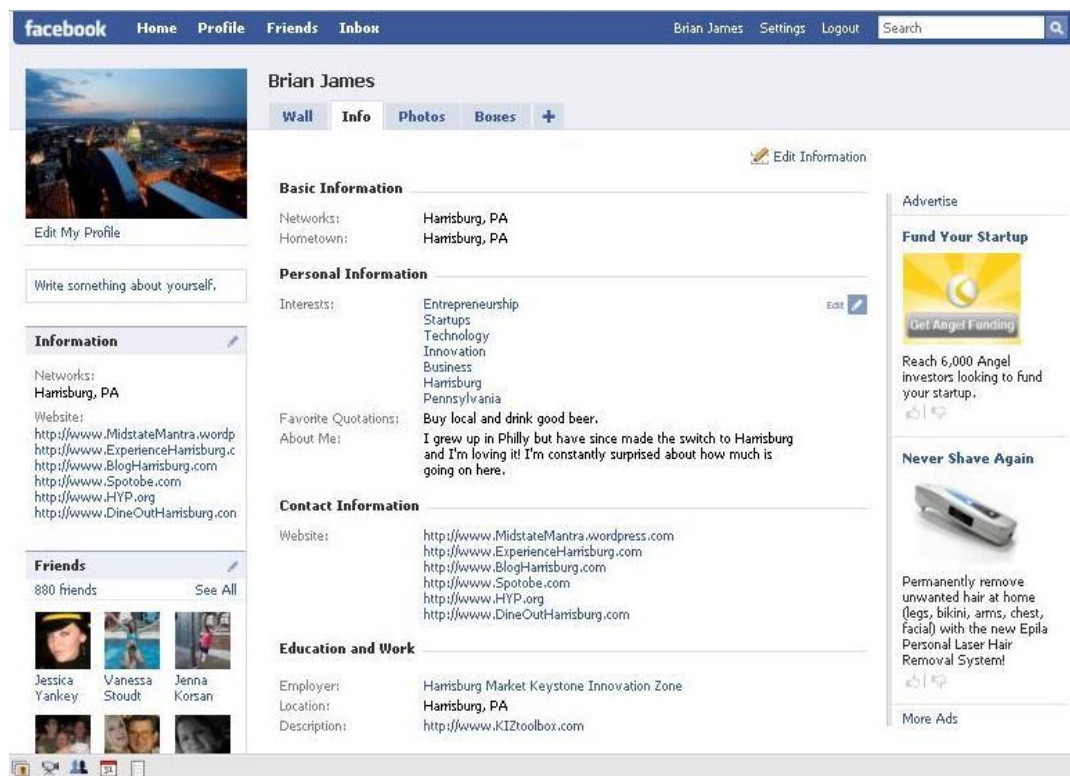
2.2.3 Ο ρόλος του προφίλ του χρήστη (user profile)

Παρόλο που οι ιστότοποι κοινωνικής δικτύωσης έχουν αναπτύξει μια ευρεία ποικιλία τεχνικών χαρακτηριστικών, η «υποδομή τους» (backbone) αποτελείται κυρίως από ορατά προφίλ, που εμφανίζουν ένα συγκεντρωτικό κατάλογο «φίλων» (Friends), οι οποίοι είναι και χρήστες του συστήματος. Τα προφίλ είναι μοναδικές σελίδες όπου μπορεί κανείς να «αποτυπώσει τον εαυτό του όπως ακριβώς είναι» (Sundén, 2003).

Η πρώτη διαδικασία που καλείται να ακολουθήσει ένα άτομο που εντάσσεται για πρώτη φορά σε ένα ιστότοπο κοινωνικής δικτύωσης, είναι η συμπλήρωση μιας φόρμας που περιέχει μια σειρά από ερωτήματα. Το προφίλ παράγεται βάσει των απαντήσεων που δίνονται σε αυτά τα ερωτήματα, τα οποία περιλαμβάνουν συνήθως γενικά χαρακτηριστικά, όπως ηλικία, γεωγραφική θέση, «τα ενδιαφέροντα μου», αλλά υπάρχει και ένα τμήμα που ο εκάστοτε χρήστης συμπληρώνει με προσωπικές του πληροφορίες (private section). Οι περισσότεροι ιστότοποι ενθαρρύνουν τους χρήστες να φορτώσουν (upload) μια φωτογραφία τους στο προφίλ. Επιπλέον, μερικοί ιστότοποι επιτρέπουν στους χρήστες να ενισχύσουν το προφίλ τους με την προσθήκη πολυμεσικού υλικού ή τροποποιώντας τον τρόπο εμφάνισης του προφίλ τους (look and feel). Τέλος, άλλοι ιστότοποι, όπως το Facebook, επιτρέπουν στους χρήστες να προσθέτουν «Εφαρμογές» που ενισχύουν το προφίλ τους.

Ο τρόπος προβολής του προφίλ του χρήστη ποικίλει από υπηρεσία σε υπηρεσία και έγγυται πάντα στη διακριτική του ευχέρεια. Για παράδειγμα, τα προφίλ στους ιστοτόπους *Friendster* και *Tribe.net* ανιχνεύονται εξ ορισμού από τις μηχανές αναζήτησης, καθιστώντας τα ορατά σε όλους, ανεξάρτητα από το αν ο χρήστης έχει λογαριασμό στις υπηρεσίες αυτές. Εναλλακτικά, το *LinkedIn* καθορίζει το τι μπορεί να δει ο κάθε χρήστης βάσει του είδους της συνδρομής που έχει με τη συγκεκριμένη υπηρεσία. Υπηρεσίες, όπως το *MySpace*, επιτρέπουν στους χρήστες να επιλέγουν αν θέλουν να είναι το προφίλ τους δημόσιο ή να το βλέπουν μόνο οι «φίλοι» τους. Το *Facebook* έχει μια διαφορετική προσέγγιση στο τρόπο προβολής του προφίλ καθώς οι χρήστες που είναι μέρος της ίδιου «υποδικτύου» μπορούν να δουν τα προφίλ των άλλων μελών, εκτός αν ο «ιδιοκτήτης» του προφίλ έχει απαγορεύσει την προβολή του προφίλ του στα άλλα μέλη του δικτύου του. Συμπερασματικά, οι διαρθρωτικές αποκλίσεις που υπάρχουν γύρω από την προβολή και την πρόσβαση ενός προφίλ χρήστη

αποτελεί έναν από τους βασικότερους τρόπους διαφοροποίησης των ιστότοπων κοινωνικής δικτύωσης [4].



Εικόνα 4: Το προφίλ του χρήστη στο Facebook

2.2.4 Ο ρόλος των «Φίλων»

Κατά τη διαδικασία ένταξης ενός χρήστη σε ένα ιστότοπο κοινωνικής δικτύωσης, του ζητείται από το σύστημα να προσδιορίσει τους χρήστες με τους οποίους έχει μια «σχέση». Μερικοί από τους πιο δημοφιλείς όρους για να περιγραφούν οι σχέσεις αυτές, είναι «Friends», «Επαφές» και «Fans». Οι περισσότερες υπηρεσίες απαιτούν αμοιβαία επιβεβαίωση για τη δημιουργία φιλίας μεταξύ δυο χρηστών. Ο όρος «Friends» μπορεί να είναι παραπλανητικός, διότι η σύνδεση αυτή δεν σημαίνει απαραίτητα τη φιλία όπως ορίζεται στην καθημερινή ζωή, καθώς οι λόγοι που συνδέονται οι άνθρωποι μεταξύ τους ποικίλλουν (Boyd, 2006).

Η δημόσια εμφάνιση των συνδέσεων αυτών αποτελεί βασικό συστατικό των εν λόγω υπηρεσιών. Η λίστα «Φίλων» περιέχει συνδέσεις με το προφίλ του κάθε φίλου, έτσι ώστε οι χρήστες να διασχίζουν το γράφημα του δικτύου κάνοντας κλικ μέσα από αυτές τις λίστες. Στις περισσότερες υπηρεσίες, η λίστα των «Φίλων» είναι ορατή σε όλους τους χρήστες που επιτρέπεται να δουν το προφίλ ενός χρήστη, αν και υπάρχουν εξαιρέσεις. Για παράδειγμα, ορισμένοι χρήστες του *MySpace* έχουν αποκρύψει με τέτοιο τρόπο το προφίλ τους ώστε να μην είναι ορατή η λίστα των φίλων τους.

Τέλος, οι περισσότεροι ιστότοποι κοινωνικής δικτύωσης παρέχουν στους χρήστες έναν μηχανισμό που τους επιτρέπει να αφήνουν μηνύματα στο προφίλ των φίλων τους. Το

χαρακτηριστικό αυτό ουσιαστικά δίνει τη δυνατότητα δημιουργίας «παρατηρήσεων» (Comments), το οποίο χρησιμοποιείται σαν όρος σε πολλές υπηρεσίες. Επιπλέον, οι εν λόγω ιστότοποι παρέχουν ένα επιπλέον χαρακτηριστικό το οποίο δίνει τη δυνατότητα αποστολής ιδιωτικών μηνυμάτων, χρησιμοποιώντας την λογική και την τεχνολογία του webmail.

2.2.5 Χαρακτηριστικά των ιστοτόπων κοινωνικής δικτύωσης

Εκτός από τα προφίλ, τους φίλους και τα προσωπικά μηνύματα, οι ιστότοποι κοινωνικής δικτύωσης ποικίλλουν σε μεγάλο βαθμό και στα χαρακτηριστικά τους. Ορισμένοι ιστότοποι παρέχουν υπηρεσίες διαμοίρασης αρχείων video και φωτογραφιών. Επίσης, άλλοι έχουν ενσωματωμένο blogging και παρέχουν τεχνολογίες αποστολής άμεσων μηνυμάτων. Επιπρόσθετα, υπάρχουν κινητοί (mobile) ιστότοποι (π.χ., Dodgeball), αλλά και κάποιοι web-based ιστότοποι που υποστηρίζουν περιορισμένες κινητές αλληλεπιδράσεις (π.χ. Facebook, MySpace και Cyworld). Πολλοί ιστότοποι στοχεύουν σε ανθρώπους από συγκεκριμένες γεωγραφικές περιοχές ή γλωσσικές ομάδες, αν και αυτό δεν καθορίζει πάντοτε την ταυτότητα της υπηρεσίας. Για παράδειγμα το *Orkut*, ξεκίνησε στις Ηνωμένες Πολιτείες με μια μόνο αγγλική διεπαφή, αλλά οι πορτογαλόφωνοι Βραζιλιάνοι έγιναν γρήγορα η κυρίαρχη ομάδα χρηστών (Kopytoff, 2004). Επίσης, μερικοί ιστότοποι είναι σχεδιασμένοι βάσει συγκεκριμένων κατηγοριών όπως είναι η εθνότητα, η πολιτική είτε η σεξουαλική ταυτότητα. Υπάρχουν ακόμη ιστότοποι για σκύλους (Dogster) και γάτες (Catster), αν και οι ιδιοκτήτες τους θα πρέπει να διαχειρίζονται τα προφίλ τους.

Παρόλο που οι εν λόγω ιστότοποι σχεδιάζονται με τέτοιο έτσι ώστε να είναι ευρέως προσβάσιμοι, προσελκύοντας ομοιογενείς πληθυσμούς, αρχικά, δεν είναι ασυνήθιστο στη συνέχεια να βρεθούν ομάδες που χρησιμοποιούν τις υπηρεσίες αυτές για να διαχωριστούν με κριτήρια όπως η εθνικότητα, η ηλικία, το μορφωτικό επίπεδο, ή άλλους παράγοντες που τυπικά διαχωρίζουν μια κοινωνία πολιτών.

2.2.6 Δομή των ιστοτόπων κοινωνικής δικτύωσης

Οι ιστότοποι κοινωνικής δικτύωσης αποτελούν επίσης πλούσια πηγή δεδομένων φυσικής συμπεριφοράς. Τα δεδομένα που παράγονται από το προφίλ και τις διασυνδέσεις, μπορούν να συγκεντρωθούν είτε με τη χρήση τεχνικών αυτόματης συλλογής είτε μέσω συνόλων δεδομένων (datasets) που παρέχονται απευθείας από την εταιρεία, επιτρέποντας στους αναλυτές δικτύου να ερευνήσουν μεγάλης κλίμακας μοτίβα φίλων (friending) καθώς και χρήσης του δικτύου. Για παράδειγμα, το 2007 εξετάστηκε ένα ανώνυμο σύνολο δεδομένων που αποτελείται από 362 εκατομμύρια μηνύματα που ανταλλάσσονται μεταξύ τεσσάρων εκατομμυρίων χρηστών του Facebook για δραστηριότητες που αφορούν στις φιλίες και στα μηνύματα. Διερευνήθηκε επίσης η σχέση μεταξύ των στοιχείων ενός προφίλ και του αριθμού των φίλων στο Facebook και παρατηρήθηκε ότι τα πεδία του προφίλ που μειώνουν τις δραστηριότητες (transactions) τους μέσα στο δίκτυο είναι πιο πιθανό να συνδέονται με ένα μεγάλο αριθμό δεσμών φιλίας. Οι ερευνητές των εν λόγω ιστοτόπων εξέτασαν επίσης τη δομή του δικτύου «Φιλίας». Κατόπιν ανάλυσης των ρόλων που είχαν συγκεκριμένοι χρήστες για την ανάπτυξη του Flickr και Yahoo! υποστηρίζεται ότι υπάρχουν παθητικά μέλη, *inviters* και *linkers*, οι οποίοι συμμετέχουν πλήρως στην κοινωνική εξέλιξη του δικτύου. Βάσει συγκεκριμένων δεδομένων, προσδιορίστηκε μια τοπολογία χρηστών μέσω της συμμετοχής τους σε ορισμένες κοινότητες και το συμπέρασμα ήταν ότι οι συγκεκριμένοι ιστότοποι μπορούν να χρησιμοποιήσουν τη συγκεκριμένη τοπολογία για να προτείνουν άλλες κοινότητες παρόμοιου ενδιαφέροντος στους χρήστες. Τέλος, οι συνδέσεις μεταξύ φίλων δεν είναι η μόνη δομή του δικτύου που αξίζει να αναλυθεί. Για το σκοπό αυτό, εξετάστηκαν επίσης οι τρόποι με τους οποίους τα ενδιαφέροντα ενός χρήστη (αγαπημένη μουσική, βιβλία, ταινίες κλπ.), μπορούν να αποτελέσουν μια εναλλακτική δομή του δικτύου.

2.2.7 *Ιδιωτικότητα (Privacy)*

Πολλοί ερευνητές δικτύων, έχουν τονίσει τα δυνητικά προβλήματα και τις πιθανές απειλές πάνω σε θέματα ιδιωτικότητας που μπορούν να παρουσιαστούν στους ιστοτόπους κοινωνικής δικτύωσης, κυρίως όσον αφορά στην ασφάλεια των νεότερων χρηστών. Σε μία από τις πρώτες ακαδημαϊκές μελέτες της ιδιωτικής ζωής και των εν λόγω ιστοτόπων, αναλύθηκαν πάνω από 4.000 προφίλ του Facebook από το CarnegieMellon University και σκιαγραφήθηκαν οι πιθανές απειλές που μπορεί να προκύψουν σε θέματα ιδιωτικότητας, όσον αφορά στις προσωπικές πληροφορίες των φοιτητών που περιλαμβάνονται στον ιστότοπο (πόλη, ημερομηνία γέννησης). Τα στοιχεία αυτά, που βρίσκονται συχνά στο προφίλ των χρηστών, μπορούν να χρησιμοποιηθούν για την ανακατασκευή αριθμών κοινωνικής ασφάλισης των ιδίων.

Μια άλλη έρευνα υποστηρίζει ότι υπάρχει συχνά αναντιστοιχία μεταξύ της επιθυμίας των χρηστών για προστασία πάνω σε θέματα που αφορούν στην ιδιωτικότητα και στις συμπεριφορές τους μέσα στο διαδίκτυο, ειδικά όταν δεν γνωρίζουν τη δημόσια φύση του. Όσον αφορά στην εμπιστοσύνη που δημιουργείται στους ιστοτόπους κοινωνικής δικτύωσης, έρευνες υποστηρίζουν ότι η εμπιστοσύνη και η στοχευμένη χρήση υπηρεσιών, ενδέχεται να επηρεάσουν τι είδους πληροφορίες, οι χρήστες είναι πρόθυμοι να μοιραστούν μέσα στον ιστότοπο.

Μία επιπρόσθετη έρευνα πάνω σε θέματα ασφάλειας χρησιμοποίησε ελεύθερα προσβάσιμα στοιχεία από το προφίλ ενός χρήστη, με σκοπό να δημιουργήσει ένα «phishing» σύστημα που φαινομενικά ανήκε σε ένα φίλο του χρήστη μέσα στον ιστότοπο. Εν τέλει οι χρήστες-θύματα ήταν πολύ πιο πρόθυμοι να δώσουν πληροφορίες σε αυτό τον «φίλο» παρά σε έναν ξένο. Παρόλα αυτά, στοιχεία από παρόμοιες μελέτες παρέχουν μια πιο αισιόδοξη προοπτική για το συγκεκριμένο ζήτημα, γεγονός που υποδηλώνει ότι οι έφηβοι έχουν επίγνωση σχετικά με τις πιθανές απειλές πάνω σε θέματα ιδιωτικότητας στο διαδίκτυο και φαίνονται πρόθυμοι να συμμετάσχουν στη διαδικασία λήψης μέτρων για την ελαχιστοποίηση ορισμένων δυνητικών κινδύνων.

Η ιδιωτικότητα εμπλέκεται επίσης και με την ικανότητα των χρηστών να ελέγχουν και να διαχειρίζονται «κοινωνικά πλαίσια» (social contexts). Για παράδειγμα, η έννοια «News Feed» που εισήχθη για πρώτη φορά στο Facebook αποσυντόνισε τους φοιτητές, έστω και αν οι πληροφορίες που υπήρχαν μέσα στο feed ήταν από πριν προσβάσιμες. Μια άλλη έρευνα αναφέρει ότι οι επιλογές που προσφέρονται από τους ιστοτόπους κοινωνικής δικτύωσης πάνω σε θέματα ιδιωτικότητας, δεν παρέχουν στους χρήστες την ευελιξία που χρειάζονται για να διαχειριστούν τις επικείμενες συγκρούσεις που μπορεί να έχουν με φίλους που έχουν διαφορετικές αντιλήψεις πάνω σε τέτοια θέματα. Οπότε, προτείνεται να ακολουθηθεί ένα γενικό πλαίσιο για την προστασία της ιδιωτικής ζωής του κάθε χρήστη μέσα στον ιστότοπο με την προοπτική ότι θα βοηθήσει στην επίλυση αυτών των συγκρούσεων.

Ολοκληρώνοντας, η έννοια της ιδιωτικότητας στους ιστοτόπους κοινωνικής δικτύωσης και εν γένει στο διαδίκτυο έρχεται πολλές φορές σε σύγκρουση με νομικές αντιλήψεις περί προστασίας της ιδιωτικής ζωής. Ο τρόπος που αντιλαμβάνεται ο κάθε χρήστης την έννοια της ιδιωτικότητας τον καθιστά νομικά υπεύθυνο για τον τρόπο διάδοσης πληροφοριών και δεδομένων που αφορούν στο προσωπικό του προφίλ [4].

3

Τεχνολογικό υπόβαθρο

Στο κεφάλαιο αυτό ακολουθεί η περιγραφή των τεχνολογικών χαρακτηριστικών των εν λόγω συστημάτων καθώς και των λειτουργιών τους.

3.1 Social Tagging συστήματα

Οι υπηρεσίες που προσφέρουν τα κοινωνικά δίκτυα συνέβαλλαν στη δημιουργία online κοινοτήτων, επιτρέποντας στους χρήστες να συνεισφέρουν στην συλλογική γνώση τοποθετώντας tags (επισημειώσεις) σε online υλικό (resources).

Τα tags είναι λέξεις – κλειδιά, θεματικοί όροι, που επιλέγονται από τους χρήστες συχνά με μία σύντομη περιγραφή του πόρου στον οποίο αναφέρονται. Ουσιαστικά, πρόκειται για μία μορφή αδόμητης, αλλά ρητής δημιουργίας μεταδεδομένων της πληροφορίας από τους χρήστες. Η διαδικασία αυτή αποτελεί το λεγόμενο social tagging. Το social tagging, ή αλλιώς το collaborative tagging, είναι ένα σχετικά νέο φαινόμενο, μία μορφή οργάνωσης γνώσης και ένας νέος τρόπος ανακάλυψης πληροφοριακών πόρων, που επιτρέπει στον καθένα να δημιουργεί σελιδοδείκτες (bookmarks) ή tags για να οργανώσει και να αποθηκεύει, με ένα αυτόνομο σύστημα επισήμανσης, την πληροφορία. Το φαινόμενο αυτό αναφέρεται και με τους όρους κοινωνική ταξινόμηση (social classification), ή κοινωνική ευρετηρίαση (social indexing) [2].

Ανάμεσα στα κοινωνικά δίκτυα, το Delicious, το Flickr και το YouTube είναι τρία από τα πιο γνωστά και τα πιο δημοφιλή του είδους. Αυτά τα δίκτυα εγκαθίδρυσαν την κρατούσα θέση τους φιλοξενώντας ένα μεγάλο αριθμό αντικειμένων, δημιουργώντας μεγάλες κοινότητες ενεργών χρηστών και προκαλώντας μεγάλη δικτυακή κίνηση. Συγκεκριμένα, το 2007 το Flickr φιλοξενούσε περισσότερες από δύο δισεκατομμύρια εικόνες, το 2006 το Delicious είχε περισσότερους από ένα δισεκατομμύριο χρήστες και το 2005 είχαν πραγματοποιηθεί έως τότε 1.73 δισεκατομμύρια θεάσεις video. Ένα ερευνητικό ζήτημα που τέθηκε ήταν αν η ανάλυση της επισημειωμένης συμπεριφοράς των χρηστών θα μπορούσε να αποκαλύψει διακριτά χαρακτηριστικά των κοινωνικών δικτύων. Σε αυτά τα πλαίσια

πραγματοποιήθηκε μία προσπάθεια διερεύνησης της social tagging συμπεριφοράς για τα έτη 2005, 2006 και 2007 όπως παρουσιάζεται στον Πίνακα 1.

Προς αυτόν τον στόχο, αναπτύχθηκε ο UTO Tag Crawler για να ενοποιήσει διαφορετικά tagging δεδομένα από τα κοινωνικά δίκτυα. Το συγκεκριμένο εργαλείο ανασύρει τα tags από το Delicious, το Flickr και το YouTube και αποθηκεύει τα δεδομένα για τα tags και τη συμπεριφορά σε RDF τριάδες, σύμφωνα με την Upper Tag Ontology (UTO). Για την έρευνα χρησιμοποιήθηκαν 21 εκατομμύρια RDF τριάδες από το Delicious, 2.3 εκατομμύρια από το Flickr και 2.2 εκατομμύρια από το YouTube [5].

Social Network	Objects	Taggers	Tags	Tag/Object	Tag/Tagger	Objects/Tagger
Delicious	996,748	2,787,860	9,282,058	9.31	3.33	0.36
Flickr	295,837	153,778	1,351,201	4.57	8.79	1.92
YouTube	527,924	185,975	1,443,924	2.74	7.76	2.84
Total	1,820,509	3,127,613	12,077,183	5.54	6.63	1.71

Πίνακας 1: Σύνολο αντικειμένων και επισημειώσεων για την περίοδο 2005-2007

3.1.1 Ο ρόλος των tags στα social tagging συστήματα

3.1.1.1 Ανάλυση επισημειώσεων από το Delicious

Στην Εικόνα 5 παρουσιάζονται τα 20 πιο συχνά ανατεθειμένα tags στο Delicious για τα έτη 2005 – 2007. Αυτά τα tags εμφανίζονται σχετικά σταθερά τα τρία αυτά χρόνια. Τα tags xml, science, search, games, technology και security εμφανίζονται στα top 20 tags του 2005, αλλά δεν βρίσκονται στην λίστα των top 20 του 2006 και 2007. Τα tags imported, search και internet έπεσαν από την λίστα των top 20 του 2007. Τα tags development, howto, tutorial και Web 2.0 εμφανίζονται τόσο στο 2006 όσο και στο 2007. Το 2007 εισάγονται τα tags webdesign, free και opensource φανερώνοντας τα νέα ενδιαφέροντα των χρηστών. Συνολικά, το 85% των top 20 tags παραμένει σταθερό στα έτη 2007 και 2008, δείχνοντας ότι μπορεί να δημιουργείται ένα διαμοιραζόμενο κοινωνικό λεξιλόγιο στο Delicious.

Το προφίλ των χρηστών του Delicious επιχειρήθηκε να δημιουργηθεί μέσα από την ανάλυση των πιο δημοφιλών tags. Η κυριαρχία κάποιων tags όπως blog, programming, web και design φανερώνουν τα κομβικά ενδιαφέροντα των χρηστών, ενώ τα tags music, video και art και news δείχνουν ένα πιο γενικό ενδιαφέρον. Βάσει αυτών των στοιχείων μπορεί να πραγματοποιηθεί η υπόθεση ότι το Delicious είναι ένα κοινωνικό δίκτυο για άτομα που ενδιαφέρονται κατά κύριο λόγο για το web και τον προγραμματισμό. Επιπρόσθετα, τα tags που εμφανίστηκαν για πρώτη φορά το 2006 και το 2007 φανερώνουν ένα αυξανόμενο ενδιαφέρον για ελεύθερο λογισμικό, αλλά και για tutorials και how-to πηγές που υποστηρίζουν την μάθηση νέων προγραμματιστικών γλωσσών και δεξιοτήτων. Το tag music παρουσιάζει σημαντική πτώση, που μπορεί να αιτιολογηθεί από την άνοδο της δημοτικότητας του Last.fm ως κοινωνικό δίκτυο για μουσικές προτιμήσεις την ίδια χρονική περίοδο. Τέλος, το tag programming έπεσε από την 2η στην 4η θέση, γεγονός που σε καμία περίπτωση δεν φανερώνει αλλαγή στο προφίλ των χρηστών.

Rank	2005	2006	2007
1	blog/blogs	blog/blogs	blog/blogs
2	programming	programming	design
3	software	software	software
4	music	design	programming
5	design	reference	reference
6	web	music	tools
7	reference	web	Web2.0
8	java	tools	web
9	art	art	video
10	tools	java	music
11	linux	video	art
12	news	Web2.0	linus
13	xml	linux	webdesign
14	science	news	howto
15	search	tutorial	free
16	games	howto	tutorial
17	research	imported	news
18	technology	development	development
19	security	research	opensource
20	video	internet	Java

Εικόνα 5: Top 20 Delicious tags

3.1.1.2 Ανάλυση επισημειώσεων από το Flickr

Στην Εικόνα 6 παρουσιάζονται τα 20 πιο συχνά ανατεθειμένα tags στο Flickr για τα έτη 2005 – 2007. Στο Flickr παρατηρείται ότι οι χρήστες αρέσκονται στο να βάζουν tags σε φωτογραφίες με ημερομηνίες, τοποθεσίες, χρώματα και εποχές. Συγκεκριμένα, τοποθεσίες που γίνονται tag συχνά είναι: Hong Kong (2005), Germany (2005), USA (2006 και 2007), London (2005-2007), California (2006), και Japan (2007). Χρώματα που γίνονται tag συχνά είναι τα εξής: πορτοκαλί (2005), μπλε (2006 και 2007), κόκκινο (2006 και 2007), πράσινο (2006 και 2007) και μαυρόασπρο (2007).

Στο Flickr συχνά οι χρήστες χρησιμοποιούν ανεπίσημα tags στις φωτογραφίες (π.χ. me), το οποίο φανερώνει ότι ένας λόγος που κάνουν tags είναι για την αποθήκευση και την ανάκτηση από τους ίδιους, παρά από πρόθεση να τις μοιραστούν με άλλες. Όταν κάνουν tag φωτογραφίες οι χρήστες εστιάζονται στα στοιχεία που είναι έντονα στο μάτι, όπως το χρώμα,

το τοπίο (ουρανός, θάλασσα, παραλία) και τις συνθήκες λήψης σε σχέση με τον φωτισμό (π.χ. βραδινή λήψη).

Rank	2005	2006	2007
1	2005	usa	2007
2	d70	california	canon
3	tsimshatsui	2006	nature
4	hongkong	cameraphone	autumn
5	nightview	celltagged	art
6	germany	zonetag	nikon
7	newkie	sanfrancisco	water
8	ragbrai	blue	bw
9	art	light	red
10	wonder	sky	blue
11	night	urban	sky
12	buttersweet	red	japan
13	15fav	sea	fall
14	central	me	beach
15	light	water	portrait
16	marco	nature	london
17	london	marco	night
18	apargioides	london	green
19	orange	green	usa
20	ads1	music	november

Εικόνα 6: Top Flickr tags

Από την ανάλυση που έγινε, προέκυψε το συμπέρασμα ότι υπάρχουν δύο μεγάλες κοινότητες στο Flickr. Η πρώτη αποτελείται από μη επαγγελματίες που χρησιμοποιούν το Flickr ως πλατφόρμα για να μοιράζονται τις φωτογραφίες με τους φίλους και την οικογένειά τους, ενώ η δεύτερη ομάδα χρηστών αποτελείται από επαγγελματίες φωτογράφους που δεν κάνουν tags συχνά, αλλά παρέχουν με μεγάλη συχνότητα σχόλια στις φωτογραφίες άλλων επαγγελματιών φωτογράφων.

3.1.1.3 Ανάλυση επισημειώσεων από το YouTube

Στην Εικόνα 7 παρουσιάζονται τα 20 πιο συχνά ανατεθειμένα tags στο YouTube για τα έτη 2005 – 2007.

Rank	2005	2006	2007
1	music	the	the
2	funny	funny	music
3	video	music	funny
4	the	video	video
5	dance	live	girl
6	crazy	of	of
7	commercial	comedy	sexy
8	dancing	dance	live
9	live	rock	dj
10	AMV	cat	2007
11	fun	Halloween	dance
12	guitar	love	hot
13	hot	girl	comedy
14	girl	movie	rock
15	japan	dj	love
16	animee	in	and
17	Halloween	sexy	sex
18	cat	and	in
19	halo	fight	new
20	of	you	cat

Εικόνα 7: YouTube Top 20 tags

Στο YouTube αυξάνεται η tagging δραστηριότητα δραματικά από το 2005 μέχρι το 2007. Ο συνολικός αριθμός των tags αυξήθηκε από 4.735 το 2005 σε 1.073.042 το 2007. Το σύνολο με τα tags παραμένει σε μεγάλο βαθμό σταθερό, φανερώνοντας ότι οι προτιμήσεις της κοινότητας παραμένουν οι ίδιες. Το tag *year* είχε την μεγαλύτερη αύξηση το 2007, ακολουθούμενο από τα tag *new* και *sex/sexy*.

3.1.2 Μέτρα βελτίωσης των tagging συστημάτων

Προκειμένου να διευκολυνθεί η αναζήτηση σε tagging συστήματα, το κριτήριο σύγκλισης των tags μπορεί να βελτιωθεί χρησιμοποιώντας νέα μέτρα.

Μία σπουδαία παράμετρος είναι η συμπεριφορά των ίδιων των χρηστών, αφού άλλωστε αυτοί επιλέγουν τα tags που θα χρησιμοποιήσουν. Εκπαιδύοντας τους χρήστες να χρησιμοποιήσουν καλύτερα tags θα ήταν ένα σημαντικό βήμα προς τη σύγκλιση των tags σε μία κατεύθυνση. Σήμερα, οι χρήστες δεν δίνουν μεγάλη βαρύτητα όταν χρησιμοποιούν ένα tag και η χρήση πρόχειρων tags είναι μία πραγματικότητα σε αυτά τα συστήματα. Κύριες απώλειες είναι τα λάθη στην γραμματική, η κακή χρήση της κωδικοποίησης, tags που δεν ακολουθούν τις συμβάσεις ως προς τον ενικό / πληθυντικό και την χρήση μικρών / κεφαλαίων – αριθμών, προσωπικών tags που δεν έχουν νόημα για την κοινότητα, tags μίας χρήσης που εμφανίζονται μόνο μία φορά στην βάση δεδομένων.

Επίσης, υπάρχει η άποψη ότι η δημιουργία των tags οφείλει να είναι περισσότερο τεκμηριωμένη. Προς το παρόν, η κοινότητα των συστημάτων αυτών δεν έχει καθορισμένα πρότυπα για την βελτίωση των tags, αλλά έχουν προταθεί πολλές ιδέες. Πρώτα από όλα, η κοινότητα πρέπει να είναι έτοιμη να θέσει κανόνες και να συμφωνήσει σε ένα πρότυπο για τα tags. Κατά δεύτερο λόγο, οι χρήστες πρέπει να πληροφορηθούν αυτούς τους κανόνες και να συμφωνήσουν να τους ακολουθούν. Έχει προταθεί ένα σύνολο από καλύτερες πρακτικές (best practice), οι οποίες περιλαμβάνουν:

- Την χρήση πληθυντικού αντί ενικού
- Την χρήση μικρών γραμμμάτων
- Την ομαδοποίηση λέξεων με χρήση underscore
- Το να ακολουθεί ένας χρήστης μία σύμβαση για ένα tag, που έχει υιοθετήσει ο πρώτος που το χρησιμοποίησε
- Την προσθήκη συνωνύμων

Μία άλλη πρόταση βελτίωσης των εν λόγω συστημάτων, είναι ότι οι χρήστες που επισημαίνουν πρέπει να σκέφτονται ταυτόχρονα γενικά και συγκεκριμένα και να μην χρησιμοποιούν προσωπικά και γενικά tags που δεν χρησιμοποιούνται. Μία άλλη παραδοχή είναι ότι όσα περισσότερα tags, τόσο καλύτερα. Πολλά tagging συστήματα επιτρέπουν σε ένα χρήστη να διορθώσει ένα tag που έχει χρησιμοποιήσει. Ίσως μία καλή λύση να ήταν η δημιουργία μίας αρχής από τα πιο γνωστά tagging συστήματα, έτσι ώστε να καθορίσουν κάποιες κοινές καθοδηγητικές γραμμές. Το tagging θα μπορούσε να βελτιωθεί παρέχοντας στους χρήστες μεθόδους που προωθούν την καλή επιλογή tags. Ακόμα, η αντικατάσταση των tags που αποτελούνται από πολλαπλές λέξεις μπορεί να ευνοήσει την αναζήτηση.

Μαζί με την εκπαίδευση των χρηστών, πολλά μπορούν να κάνουν και οι δημιουργοί των tagging συστημάτων για να βελτιώσουν τα δεδομένα που δημιουργούνται. Υπάρχουν δύο κύριες πλευρές στις οποίες μπορούν να γίνουν βελτιώσεις. Το πρώτο σημείο που μπορούν να υπάρξουν βελτιστοποιήσεις είναι το σημείο όπου εισάγονται νέες πηγές στο σύστημα. Σε αυτό, θα μπορούσαν να γίνουν έλεγχοι για λάθη στα tags που εισάγονται. Επιπλέον, όπως ήδη εφαρμόζουν κάποιοι ιστότοποι, μπορούν να γίνονται προτάσεις για τη χρήση συγκεκριμένων tags. Επιπρόσθετα, τα συστήματα θα μπορούσαν να προτείνουν συνώνυμα, επεκτάσεις στα ακρώνυμα και άλλες προτάσεις σύμφωνα με το tag που εισάγει ο χρήστης. Δευτερευόντως, βελτιστοποιήσεις θα μπορούσαν να γίνουν και στον τρόπο όπου αναζητούνται τα ήδη υπάρχοντα στο σύστημα δεδομένα.

Μία περιοχή που δεν έχει διερευνηθεί ακόμα είναι η δημιουργία ενός εργαλείου που θα αξιοποιείται για τη συζήτηση μεταξύ των μελών ενός κοινωνικού δικτύου, για τον λόγο που τοποθετήθηκε ένα tag σε μία συγκεκριμένη πηγή. Αυτήν τη στιγμή, υπάρχει ελάχιστη συζήτηση στα tagging συστήματα για την καταλληλότητα των επισημειώσεων. Οι περισσότερες ιστοσελίδες δεν προσφέρουν τη δυνατότητα ανατροφοδότησης των επισημειώσεων που επιλέγονται, αν και μερικά επιτρέπουν την αλλαγή μεταδεδομένων άλλων χρηστών. Ακόμα, αρκετά συστήματα παρέχουν πολύ λίγη πληροφορία για τον χρήστη που επισημαίνει μία πηγή. Όπως είναι προφανές, πιο διεξοδικά προφίλ των χρηστών θα βελτιώναν την εμπειρία για τα άλλα μέλη του ίδιου κοινωνικού δικτύου. Για παράδειγμα, μία σημαντική πληροφορία είναι η προτιμώμενη γλώσσα για τις πηγές που τοποθετεί κάποιος ένα tag. Προς αυτήν την κατεύθυνση το project Collaborative Rank αξιολογεί τους χρήστες σύμφωνα με το κατά πόσο οι προτιμήσεις που έδωσαν ήταν χρήσιμες σε άλλα μέλη.

Εν τέλει, το ερώτημα που τίθεται είναι αν είναι προτιμότερο να υπάρχουν ορισμένα

δημοφιλή tags ή αν είναι καλύτερα να υπάρχει ένα μεγαλύτερο σύνολο, από άγνωστα tags, που πιθανόν να αναπαριστούν με μεγαλύτερη πληρότητα διαφορετικές πτυχές του ίδιου αντικειμένου [6].

3.1.3 Συμπεράσματα για το προφίλ των κοινωνικών δικτύων

Συγκρίνοντας, τα τρία κοινωνικά δίκτυα, το Delicious παρουσιάζει την πιο ισχυρή σύνδεση των tags ως μέσο παροχής εκτεταμένης πληροφόρησης για τους πόρους της ιστοσελίδας. Στο Delicious, κάθε χρήστης μπορεί να επισημειώσει ένα αντικείμενο με μία λέξη δικής του επιλογής και κάθε αντικείμενο μπορεί να επισημειωθεί πολλές φορές, ακόμα και από διαφορετικούς χρήστες, αποδεικνύοντας ότι ανήκει στο σύνολο της Delicious κοινότητας. Το Delicious είναι χαρακτηριστικό παράδειγμα κοινότητας που το μέλος της μπορεί να επισημάνει οποιονδήποτε online πόρο.

Αντίθετα στο Flickr, επισημειώσεις βάζει μόνο αυτός που ανεβάζει το περιεχόμενο και οι άλλοι χρήστες απλά σχολιάζουν την φωτογραφία. Επιπλέον, στο Flickr επιτρέπεται οι χρήστες που είναι φίλοι με το άτομο που ανέβασε το περιεχόμενο να κάνουν tag στην φωτογραφία. Όμως αυτό σε καμία περίπτωση δεν μπορεί να θεωρηθεί ως μία ανοιχτή κοινότητα, αλλά περισσότερο σαν ένα τρόπο οργάνωσης των φωτογραφιών του εκάστοτε χρήστη και δημιουργίας μικρών κοινοτήτων με τους φίλους του. Παρόμοιος είναι και ο τρόπος λειτουργίας του YouTube, όπου οι άλλοι χρήστες μπορούν να αξιολογήσουν το περιεχόμενο που ανέβασε κάποιος άλλος.

Αυτές οι διαφορές προκάλεσαν αποκλίσεις όχι μόνο στον ρόλο που τα tags παίζουν σε κάθε σύστημα, αλλά και στην φύση των tags που ανατίθενται. Βάση της ανάλυσης που προηγήθηκε, είναι φανερό ότι τα tags στο Delicious είναι προσανατολισμένα στο περιεχόμενο, ενώ στο Flickr έχουν να κάνουν με τα φυσικά χαρακτηριστικά των ίδιων των φωτογραφιών. Αντιθέτως, στο YouTube τα tags τείνουν να περιγράφουν το είδος του περιεχομένου και την γνώμη των θεατών για αυτό.

Ο ρόλος των tags στο Delicious επικεντρώνεται στην αναπαράσταση πηγών που προστέθηκαν στους σελιδοδείκτες ώστε να αξιοποιηθούν για μελλοντική ανάκτηση, αλλά και για να μπορεί να τους μοιραστεί ολόκληρη η κοινότητα. Τα tags διαδραματίζουν κυρίαρχο ρόλο στην λειτουργία της κοινότητας, αφού χωρίς την χρήση τους δεν θα γινόταν εφικτός ο διαμοιρασμός των σελιδοδεικτών και η ανάκτηση τους. Αντίθετα, στα άλλα δύο κοινωνικά δίκτυα ο ρόλος των tags δεν είναι ζωτικής σημασίας για την λειτουργία τους και πολλά μέλη τους δεν κατανοούν την αξία τους.

Η συμπεριφορά των social tagging κοινοτήτων σχετίζεται και με την κοινότητα των χρηστών σε κάθε δίκτυο. Το Delicious έχει μία κοινότητα που ενδιαφέρεται για θέματα σχετικά με την τεχνολογία της πληροφορίας. Τα μέλη της ενδιαφέρονται για το περιεχόμενο των πόρων που βρίσκουν, και τα tags παρέχουν ένα τρόπο προκειμένου να το παρουσιάσουν συνοπτικά. Συνεπώς, η χρήση των tags αποτελεί σημείο κλειδί για το σύστημα. Αντίθετα, το Flickr αποτελείται από τις 2 κοινότητες που αναλύθηκαν παραπάνω, ενώ το YouTube μπορεί να θεωρηθεί ως μία μικρογραφία του διαδικτύου, όπου οι χρήστες έχουν διαφορετικές απαιτήσεις και ανάγκες, κάνοντας την αξιολόγηση με τα αστεράκια να επισκιάζει τη σημασία του tagging.

Η κοινότητα του Delicious, σε αντίθεση με τις άλλες, μπορεί να χαρακτηριστεί ως δυναμική αφού το 10% – 20% του συνόλου των επισημειώσεων αλλάζει ως προς το χρόνο.

Συνεπώς, είναι πιο αντιπροσωπευτικό από το YouTube για τις τάσεις στα κοινωνικά λεξιλόγια επισημειώσεων και τη συμπεριφορά του γύρω από αυτά.

Μία ακόμα έρευνα ανέλυσε tags αποκλειστικά από το Delicious χρησιμοποιώντας δύο σύνολα δεδομένων που συγκεντρώθηκαν το 2005. Παρατηρήθηκε ότι οι χρήστες χρησιμοποιούν διαφορετική ποικιλία από tags. Άλλοι χρησιμοποιούν πολλά tags και άλλοι λίγα. Τα ίδια τα tags διαφέρουν στη συχνότητα της χρήσης τους, αλλά και στο τι περιγράφουν. Παρόλα αυτά, κάποια σταθερά πρότυπα αναδεικνύονται, αλλά και οι απόψεις της μειοψηφίας συνυπάρχουν με ιδέες της πλειονότητας. Φανερό ήταν ότι σε μεγάλο βαθμό οι χρήστες χρησιμοποιούσαν tags προς όφελος των ιδίων και όχι για να ευνοηθεί η ολόκληρη η κοινότητα.

Ακόμα και αν αυτός είναι ο λόγος που ωθεί τους περισσότερους χρήστες στο να επισημαίνουν ιστοσελίδες στο Delicious, η κοινότητα μπορεί να επωφεληθεί. Για παράδειγμα, αν πολλοί χρήστες χαρακτηρίσουν κάτι ως αστείο, είναι πολύ πιθανό την ίδια άποψη να έχουν και άλλοι. Επίσης για ένα θέμα που πολλοί χρησιμοποίησαν το tag «toread», είναι πολύ πιθανό να ενδιαφέρει και άλλους να το διαβάσουν. Έτσι το Delicious μπορεί να λειτουργήσει ως ένα σύστημα παροχής προτιμήσεων, αρκεί βέβαια τα μέλη της κοινότητας να κατανοούν με παρόμοιο τρόπο τις έννοιες που χρησιμοποιούν [7].

Από τα παραπάνω, η έρευνα που διεξήχθη κατέληξε στο συμπέρασμα ότι μπορεί να δημιουργηθεί το προφίλ ενός κοινωνικού δικτύου από τα tags και τη συμπεριφορά των μελών γύρω από αυτά. Ένα άλλο συμπέρασμα είναι ότι οι προτάσεις για τα tags από το σύστημα θα ήταν πολύ χρήσιμες, καθώς και η επιλογή ενικού / πληθυντικού, η επιλογή ακρωνυμίων / πλήρων ονομάτων, αλλά και η δημιουργία προτύπου για τη σύνταξη των tags θεωρείται ότι θα αύξανε δραματικά την αποτελεσματικότητα της ανάκτησης της επιθυμητής πληροφορίας σε κοινωνικά δίκτυα [5].

Συμπερασματικά, οι επισημειώσεις αποτελούν τον ακρογωνιαίο λίθο τέτοιου είδους συστημάτων και η ανάλυση του σκοπού της χρήσης τους αποτελεί βασικό ζήτημα για τη δημιουργία στοχευμένων υπηρεσιών προς τους τελικούς χρήστες. Ο κυριότερος λόγος που οι χρήστες επισημαίνουν ψηφιακά αντικείμενα, ιστοσελίδες και εγγραφές, είναι η μελλοντική ανάκτηση τους με βάση την έννοια που ο χρήστης το είχε χαρακτηρίσει και ανταποκρίνεται στις αναζητήσεις του. Πέρα από την ανάκτηση, υπάρχουν εξειδικευμένοι χρήστες που επισημαίνουν για να διορθώσουν ή να εμπλουτίσουν το ήδη υπάρχον σύστημα ανάκτησης πληροφορίας με σκοπό την βελτίωση του τρόπου ευρετηρίασης των όρων που χρησιμοποιούνται. Τέλος, τόσο οι χρήστες όσο και οι δημιουργοί περιεχομένου, επισημαίνοντας σελίδες ή αντικείμενα τους δίνουν τη δυνατότητα να συνδεθούν με άλλα συναφή αντικείμενα (Technorati), δημιουργώντας με αυτό τον τρόπο μια κοινωνική ταξινόμια. Η δημιουργία καταλόγων σχετικών άρθρων, μέσω της χρήσης των επισημειώσεων αποτελεί ένα άλλο είδος κοινωνικής ταξινόμιας που χρησιμοποιείται από διάφορα λογισμικά ιστολογίων (World Press).

3.2 Folksonomies

3.2.1 Η ανάγκη για κοινωνικές ταξονομίες

Με την έλευση των νέων τεχνολογιών πολλά μεγάλα αδόμητα σύνολα πληροφορίας προσφέρονται σε πολύ σύντομο χρονικό διάστημα. Καθώς αυτή η πληροφορία καθίσταται διαθέσιμη, γίνεται όλο και πιο δύσκολη η πλοήγηση μεταξύ αυτών χωρίς κάποιον οδηγό. Καθώς πλέον ο χρήστης μπορεί να έχει 10,000 τραγούδια στην τσέπη του, η επιλογή για το

ποιο τραγούδι θα διαλέξει να ακούσει γίνεται πιο δύσκολη. Επίσης, καθώς ένας τυπικός χρήστης μπορεί να έχει πρόσβαση σε 13 δισεκατομμύρια ιστότοπους, καθίσταται σχεδόν αδύνατο να ξέρει ποιοι από αυτούς είναι σχετικοί με το αντικείμενο που τον ενδιαφέρει.

Τα μεταδεδομένα ως τεχνολογική λύση ήρθαν να δώσουν μια γενική προσέγγιση και διευθέτηση σε αυτό το πρόβλημα. Συνήθως χαρακτηρίζονται ως τα «δεδομένα για τα δεδομένα». Αποτελούν πληροφορίες, συνήθως με ισχυρή δομή, σχετικά με έγγραφα, βιβλία, άρθρα, φωτογραφίες ή άλλα αντικείμενα και είναι σχεδιασμένα έτσι ώστε να υποστηρίζουν συγκεκριμένες λειτουργίες σχετικές με την πρόσβαση στην πληροφορία.

Παραδοσιακά, τα μεταδεδομένα δημιουργούνται από ειδικούς επαγγελματίες. Οι υπεύθυνοι δημιουργίας καταλόγων παράγουν τα μεταδεδομένα συνήθως σε εγγραφές της μορφής MARC (Machine-Readable Cataloging), κάτι που αποτελεί τη βάση πολλών καταλόγων OPAC (Online Public Access Catalogs) σε βιβλιοθήκες και άλλα ιδρύματα. Η διαδικασία αυτή συχνά απαιτεί ιδιαίτερη εκπαίδευση και εξάσκηση. Οι επιστήμες της βιβλιοθηκονομίας και των πληροφοριών έχουν αναπτύξει περίτεχνους κανόνες για την καταλογοποίηση, κατηγοριοποίηση και ταξινόμηση περιλαμβάνοντας σχήματα ταξινόμησης όπως το Dewey Decimal System και το Library of Congress Classification Scheme, όπως και τεράστια ελεγχόμενα λεξιλόγια όρων για την περιγραφή του θέματος των αντικειμένων, όπως το Library of Congress Subject Headings.

Όμως, παρόλο που τα μεταδεδομένα τα οποία παράγονται με τέτοιο επαγγελματικό τρόπο θεωρούνται υψηλής ποιότητας, έχουν μεγάλο κόστος σε χρόνο που απαιτείται για τη συγκέντρωσή τους και προσπάθεια για την παραγωγή τους. Το γεγονός αυτό κάνει πολύ δύσκολη την κλιμάκωση και τη διατήρηση της συνέπειας της πληροφορίας καθώς συνεχίζουν να παράγονται τεράστια ποσά περιεχομένου, ειδικά σε μέσα όπως ο Παγκόσμιος Ιστός. Μια εναλλακτική λύση είναι η δημιουργία μεταδεδομένων από τους αυθεντικούς δημιουργούς του εκάστοτε περιεχομένου. Η λύση αυτή θα αντιμετώπιζε το πρόβλημα της κλιμάκωσης σε σχέση με τα επαγγελματικά μεταδεδομένα, όμως και οι δύο προσεγγίσεις βρίσκονται απέναντι σε ένα κοινό πρόβλημα: οι πραγματικοί χρήστες των πληροφοριών μένουν αποκομμένοι από τη διαδικασία (Mathes). Επιπρόσθετα, κανένα από τα συστήματα που παράγουν μεταδεδομένα δεν προσαρμόζεται ικανοποιητικά στο γοργά μεταβαλλόμενο σύνολο πληροφοριών που βρίσκεται στο Διαδίκτυο.

Έτσι, το ρόλο των μεταδεδομένων έρχονται να καλύψουν οι επισημειώσεις που παράγονται από τους ίδιους τους χρήστες. Η δημιουργία επισημειώσεων αποτελεί μία από τις δεσπόζουσες λύσεις με τη συλλογή μεταδεδομένων από τους τελικούς χρήστες με τρόπο απλό καταλήγοντας στην ανάπτυξη κοινωνικών ταξονομιών [8], [9].

3.2.2 Ορισμός των κοινωνικών ταξονομιών

Ο όρος «κοινωνική ταξονομία» προέρχεται από τον αγγλικό όρο «folksonomy», μια επινόηση του Thomas Vander Wal [10], ο οποίος ένωσε τις λέξεις «folk» και «taxonomy» δημιουργώντας έτσι ένα νεολογισμό του οποίου το νόημα αποδίδεται από τον ίδιο ως «από κάτω προς τα πάνω κοινωνική ταξινόμηση» («bottom-up social classification»). Σύμφωνα με την περιγραφή του Clay Shirky [11], για τις κοινωνικές ταξονομίες, πρόκειται για «κοινωνικά παραγόμενους, τυπικά επίπεδους χώρους ονομάτων». Τα βασικά χαρακτηριστικά ορισμού των κοινωνικών ταξονομιών είναι επομένως η από κάτω προς τα πάνω κατασκευή τους, η έλλειψη ιεραρχικής δομής και η δημιουργία και χρήση τους εντός ενός κοινωνικού περιβάλλοντος. Σύμφωνα πάντως με τον Merholz [12], ο όρος folksonomy είναι ανακριβής εφόσον εμπεριέχει τη λέξη taxonomy, η οποία υποδηλώνει ιεραρχία και έλεγχο. Έτσι έχουν προταθεί και άλλοι όροι όπως «ethnoclassification» και «social classification».

Πρακτικά, μια κοινωνική ταξονομία είναι ένα πλήρες σύνολο επισημειώσεων που οι χρήστες ενός συστήματος διαχείρισης διαμοιραζόμενου περιεχομένου χρησιμοποιούν για ορισμένα κομμάτια περιεχομένου, με σκοπό να τα ομαδοποιήσουν ή να τα ταξινομήσουν για μελλοντική ανάκτηση. Οι χρήστες είναι σε θέση να προσθέσουν άμεσα όρους στην κοινωνική ταξονομία καθώς αυτοί γίνονται απαραίτητοι για τον προσδιορισμό των μονάδων περιεχομένου.

Η δημόσια φύση αυτών των όρων αποτελεί ένα ιδιαίτερο χαρακτηριστικό, καθώς επιτρέπει στους χρήστες να διαπιστώνουν άμεσα πως άλλοι χρήστες έχουν χρησιμοποιήσει ίδιους όρους για την κατηγοριοποίηση του δικού τους περιεχομένου και να δουν τους όρους που άλλοι έχουν προσθέσει. Μέσα από αυτόν τον κύκλο χρήσης και διερεύνησης, η κοινότητα καθίσταται ικανή να δημιουργήσει μια κοινωνική ταξονομία. Η δημιουργία επισημειώσεων πραγματοποιείται έτσι, σε ένα κοινωνικό περιβάλλον (συνήθως διαμοιραζόμενο και ανοιχτό προς τους άλλους) με την κοινωνική ταξονομία να παράγεται από τις επισημειώσεις των ατόμων που «καταναλώνουν» την πληροφορία [10].

Στη πράξη, η κοινωνική ταξονομία είναι μία πρακτική που αποδεικνύεται ιδιαίτερα χρήσιμη όταν δεν υπάρχει κάποιος στο ρόλο του βιβλιοθηκονόμου ή όταν υπάρχουν τεράστια ποσά πληροφοριών τα οποία θα έπρεπε κάποιος να ταξινομήσει. Και οι δύο αυτές περιπτώσεις συναντώνται στον Παγκόσμιο Ιστό, όπου έχει γίνει πια δημοφιλής η δημιουργία επισημειώσεων. Οι υπηρεσίες επισημειώσεων επιτρέπουν έτσι στους χρήστες να επισημειώσουν δημόσια και να διαμοιραστούν το περιεχόμενο, έτσι ώστε όχι μόνο να κατηγοριοποιούν το περιεχόμενο για τον εαυτό τους, αλλά και να φυλλομετρούν ευκολότερα τις πληροφορίες που έχουν κατηγοριοποιηθεί από άλλους [9].

3.2.3 Η χρήση των επισημειώσεων στις κοινωνικές ταξονομίες

Όπως έχει αναφερθεί, οι επισημειώσεις είναι λέξεις ή φράσεις που οι χρήστες επισυνάπτουν σε έναν ιστότοπο ή μια σελίδα. Έχουν το ρόλο της ετικέτας για τους διάφορους διαδικτυακούς πόρους και σκοπό έχουν να βοηθήσουν το χρήστη για μια μελλοντική ανάκτηση αυτών των πόρων. Μια πρόσθετη ιδιότητά τους είναι πως μπορούν να ομαδοποιήσουν τους σχετιζόμενους διαδικτυακούς πόρους. Δεν υπάρχει όμως κάποιο σταθερό σύνολο κατηγοριών ή επίσημα εγκεκριμένων επιλογών. Ένας χρήστης μπορεί να χρησιμοποιήσει λέξεις, ακρωνύμια, αριθμούς, οτιδήποτε έχει νόημα για εκείνον, χωρίς να τον ενδιαφέρουν οι ανάγκες, τα ενδιαφέροντα και οι απαιτήσεις των άλλων. Μέσω της δημιουργίας επισημειώσεων, καθένας είναι ελεύθερος να χρησιμοποιήσει όποιες λέξεις πιστεύει πως είναι κατάλληλες, χωρίς να χρειάζεται να συμφωνήσει με κανένα για το πώς θα πρέπει να επισημειωθεί κάποιο αντικείμενο.

Έτσι, η μεθοδολογία για την δημιουργία επισημειώσεων είναι αρκετά απλή: ανάθεση λέξεων-κλειδιών στο περιεχόμενο ως ετικέτες. Για παράδειγμα, κάποιος μπορεί να προσθέσει μια επισημείωση σε μια φωτογραφία διακοπών με όρους όπως «σκι», «φίλοι», κλπ. Η δημιουργία επισημειώσεων μπορεί να θεωρηθεί και ως ένας προσωπικός τρόπος να οργανώσει κάποιος τα δεδομένα του, ώστε κάποιος άλλος αργότερα γράφοντας απλά «σκι» να μπορέσει να δει φωτογραφίες κάποιου σχετικές με το «σκι». Παρόλο που η διαδικασία της δημιουργίας επισημειώσεων είναι πολύ απλή, οι τρόποι με τους οποίους μπορούν να συναθροιστούν τις επισημειώσεις μεταξύ των χρηστών και των περιεχομένων μπορούν να είναι αρκετά σύνθετοι. Ωστόσο, είναι απαραίτητο πως η ενέργεια της ταξινόμησης για τον τελικό χρήστη να γίνεται όσο το δυνατόν πιο απλά χωρίς να χάνεται η ευελιξία.

Τελικά, η αξία σε αυτή την εξωτερική επισημείωση προέρχεται από την χρήση του λεξιλογίου των ιδίων των ατόμων, με την προσθήκη ρητού νοήματος, το οποίο πηγάζει από την κατανόηση της πληροφορίας. Οι άνθρωποι δεν ενεργούν τόσο κατηγοριοποιώντας, όσο προσφέροντας ένα μέσο για να συνδέσουν μεταξύ τους αντικείμενα και να εκφράσουν τη σημασία τους σύμφωνα με τη δική τους αντίληψη [8].

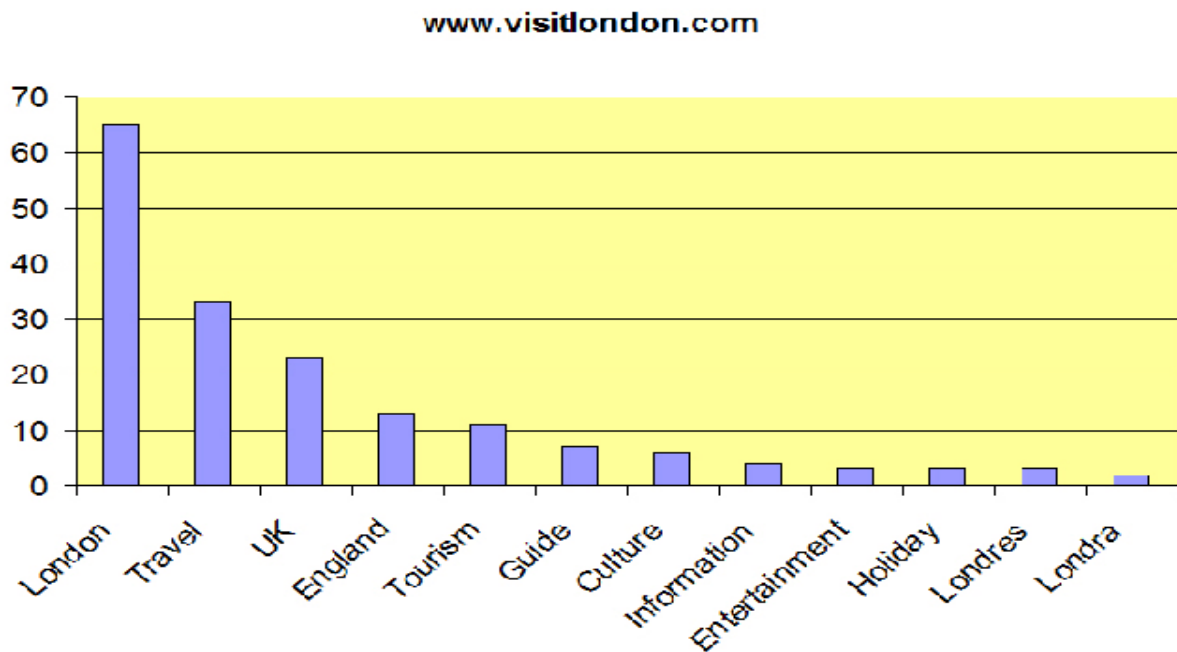
Το σύννεφο επισημειώσεων (tag cloud) χρησιμοποιείται σε πολλά συστήματα που βασίζονται σε κοινωνικές ταξονομίες. Πρόκειται για την οπτική απεικόνιση των επισημειώσεων που δημιουργούνται από τους χρήστες για να περιγράψουν το περιεχόμενο ορισμένων ιστοτόπων. Έτσι, σε αυτό εμφανίζονται οι πιο δημοφιλείς επισημειώσεις μιας σελίδας, με τις περισσότερες δημοφιλείς να είναι σε μεγαλύτερο μέγεθος γραμματοσειράς. Με αυτό τον τρόπο, οι χρήστες μπορούν με μια ματιά να αποκτήσουν μια άποψη για τις πιο δημοφιλείς επισημειώσεις μεταξύ των χρηστών [13].



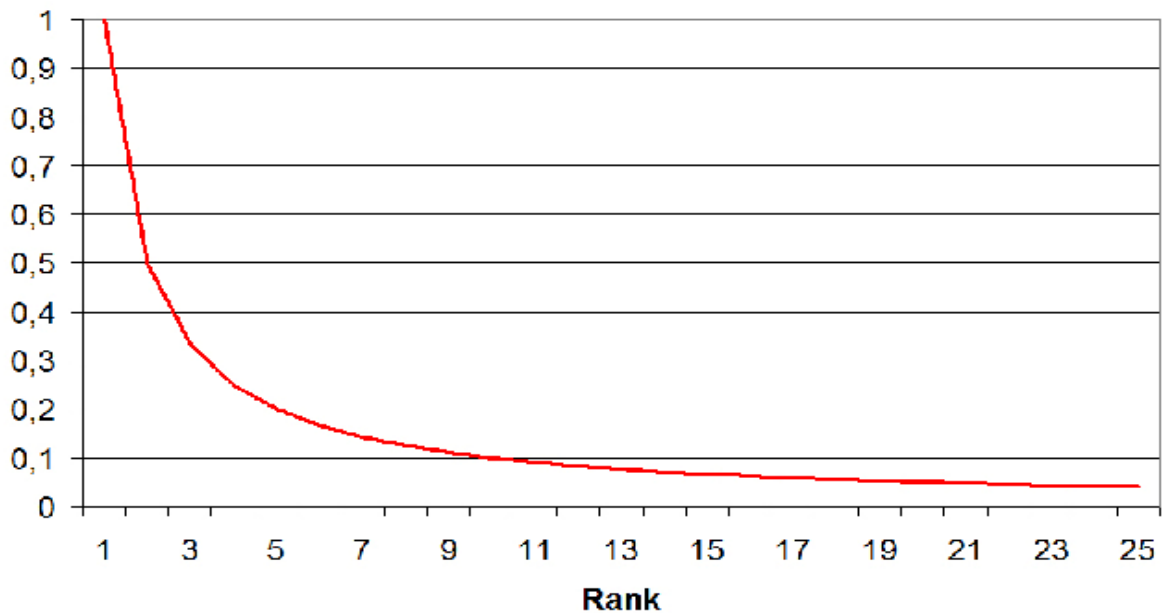
3.2.5 Κατανομή των επισημειώσεων

30

συγκεκριμένο ιστότοπο διαπιστώνεται πως οι επισημειώσεις αυτές ταιριάζουν σύμφωνα με το περιεχόμενό του και προσφέρουν μια ικανοποιητική περιγραφή του. Όπως φαίνεται στην Εικόνα 10, οι επισημειώσεις ακολουθούν ένα πρότυπο εκθετικής κατανομής (power law).

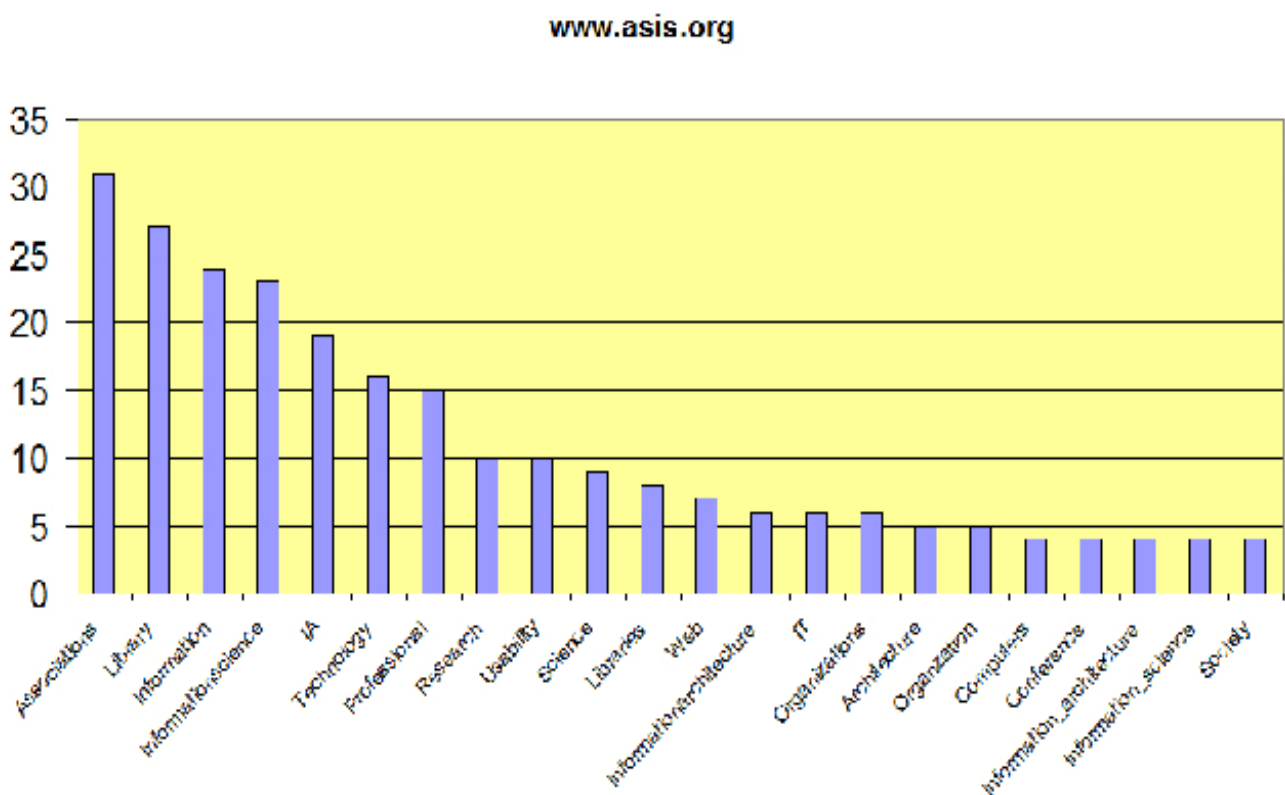


Εικόνα 9: Κατανομή επισημειώσεων για τον ιστότοπο www.visitlondon.com



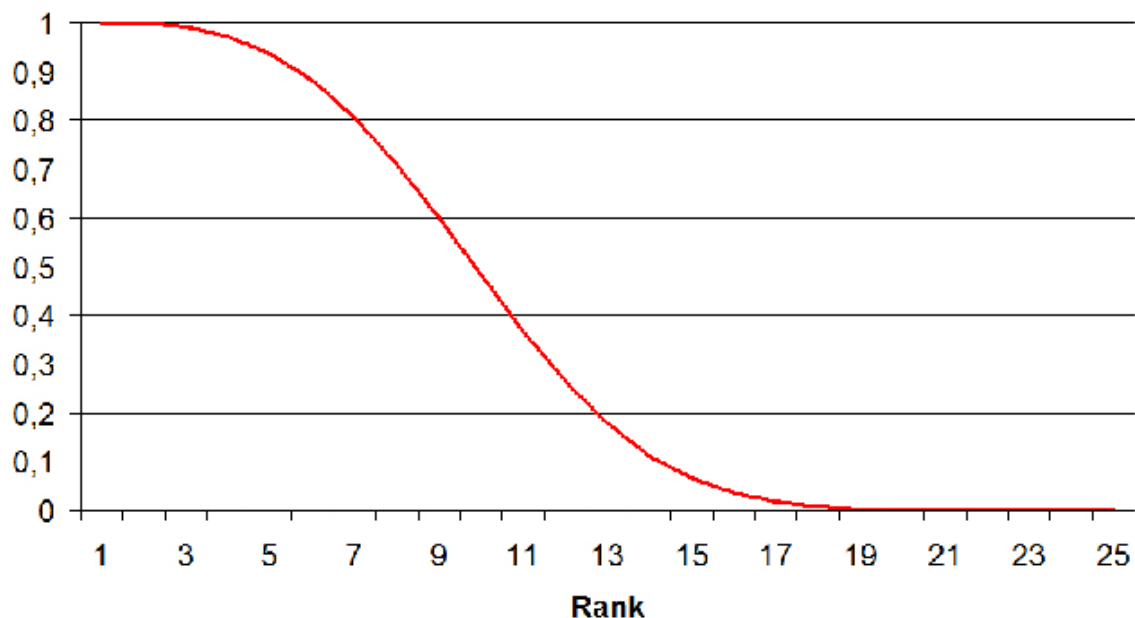
Εικόνα 10: Η εκθετική κατανομή

Το δεύτερο παράδειγμα είναι διαφορετικό καθώς υπάρχει μια λίστα κατάταξης επισημειώσεων που αναφέρονται στην ιστοσελίδα του ASIST (American Society for Information Science and Technology), όπως υπάρχει στο Del.icio.us. Ο ιστότοπος www.asis.org έχει σχολιαστεί περίπου 15 με 30 φορές από καθεμιά από τις επισημειώσεις «associations», «library», «information», «science», «IA», «technology», «professional». Όμως εδώ προφανώς όπως φαίνεται στην Εικόνα 11 δεν παρουσιάζεται εκθετική κατανομή. Υπάρχει μια «μακριά ουρά» («long tail») με επισημειώσεις όπως «research», «usability», «science», «libraries» στο δεξί τμήμα του σχήματος, όμως φαίνεται επίσης ότι υπάρχει μια δεύτερη μακριά ουρά στο αριστερό τμήμα. Οι επισημειώσεις που βρίσκονται πάνω σε αυτές τις ουρές περιγράφουν το περιεχόμενο του ιστοτόπου λιγότερο ή περισσότερο κατάλληλα. Σε αυτή την περίπτωση, η συλλεκτική ικανότητα του χρήστη παρουσιάζει δυσκολίες στο να δώσει μια αληθινή εικόνα για την ιστοσελίδα του ASIST [14].



Εικόνα 11: Κατανομή επισημειώσεων για τον ιστότοπο www.asis.org

Σύμφωνα με τον Stock [15], υπάρχουν δύο διαφορετικές κατανομές συμπεριφοράς: η γνωστή εκθετική κατανομή (ή αλλιώς «informetric») και ένα δεύτερο είδος κατανομής που ονομάζεται «αντίστροφη συμβολική λογική» («inverse logistic») που φαίνεται στην Εικόνα 12. Σε αυτή την κατανομή φαίνονται πολλά σχετικά tags τα οποία ο Stock ονομάζει το «μακρύ κορμό» («long trunk») και η γνωστή «μακριά ουρά». Ο μαθηματικός τύπος που ακολουθεί είναι ο $f(x) = e^{-C'(x-I)^b}$, όπου e είναι η σταθερά του Euler, το x είναι ο βαθμός κατάταξης της επισημείωσης, το C' είναι μια σταθερά και ο εκθέτης b είναι περίπου 3. Σε πολλές περιπτώσεις ο «μακρύς κορμός» είναι μικρότερος από ότι η «μακριά ουρά».



Εικόνα 12: Η κατανομή της «αντίστροφης συμβολικής λογικής»

3.2.6 Αναπαράσταση μοντέλων κοινωνικών ταξονομιών

Η αναπαράσταση των κοινωνικών ταξονομιών μοντελοποιείται σε δίκτυα που αναπαριστώνται ως τριμερείς γράφοι με υπερακμές (hyperedges). Το σύνολο των κορυφών διαχωρίζεται σε τρία σύνολα, το σύνολο των χρηστών, το σύνολο των επισημειώσεων - εννοιών και το σύνολο των αντικειμένων που επισημαίνονται. Έτσι, σε ένα σύστημα κοινωνικών ταξονομιών, οι χρήστες επισημαίνουν αντικείμενα δημιουργώντας τριαδικές συσχετίσεις μεταξύ του χρήστη, της επισημείωσης και του αντικειμένου. Επομένως, η κοινωνική ταξινόμια ορίζεται από ένα σύνολο T (υποσύνολο) $A \times C \times I$ (A : actor, C : concept, I : instance).

Ένα τέτοιο δίκτυο αναπαρίσταται πιο φυσικά ως ένας υπεργράφος (hypergraph) με τριαδικές ακμές, όπου κάθε ακμή αναπαριστά το γεγονός ότι ένας δεδομένος χρήστης συνέδεσε ένα συγκεκριμένο στιγμιότυπο αντικειμένου με μια συγκεκριμένη επισημείωση. Επειδή όμως οι τριμερείς γράφοι και οι υπερακμές είναι δύσκολες στη χρήση, μπορούμε να μετατρέψουμε αυτό τον υπεργράφο σε τρεις διμερείς γράφους με κανονικές ακμές. Μέσω αυτών των τριών γράφων μοντελοποιούνται οι συσχετίσεις μεταξύ των χρηστών και των επισημειώσεων (γράφος AC), των επισημειώσεων και των αντικειμένων (γράφος CO) και των χρηστών και των στιγμιότυπων (γράφος AI).

Έτσι ο διμερής γράφος AC συνδέει τα άτομα με τις επισημειώσεις που έχουν χρησιμοποιήσει για τουλάχιστον ένα αντικείμενο. Κάθε σύνδεση έχει ως ειδικό βάρος τον αριθμό των φορών που ένας χρήστης χρησιμοποίησε μια συγκεκριμένη επισημείωση. Αυτό

το είδος γράφου είναι γνωστό και ως δίκτυο δεσμού (affiliation network), συνδέοντας μεταξύ τους άτομα και υποδεικνύοντας τον βαθμό του δεσμού τους. Ένα δίκτυο δεσμού μπορεί να χρησιμοποιηθεί για να παράγει δυο απλούς γράφους με βάρη που δείχνουν τις ομοιότητες μεταξύ των χρηστών και των γεγονότων, αντίστοιχα.

Έτσι, ο γράφος AC, το δίκτυο δεσμού για χρήστες και επισημειώσεις, μπορεί να χωριστεί σε δύο γράφους: σε ένα κοινωνικό δίκτυο χρηστών που βασίζεται σε επικαλυπτόμενα σύνολα αντικειμένων και σε μια μικρή οντολογία επισημειώσεων που βασίζεται σε επικαλυπτόμενα σύνολα κοινοτήτων. Κοινώς, σε αυτό το απλό μοντέλο, τα κοινωνικά δίκτυα και η σημασιολογία είναι απλά οι δύο όψεις του ίδιου νομίσματος: ο αυθεντικός διμερής γράφος περιέχει όλη την πληροφορία που χρειάζεται για την παραγωγή αυτών των δικτύων, χωρίς όμως να είναι εφικτό να τον αναπαράγουμε αν τα έχουμε στη διάθεσή μας.

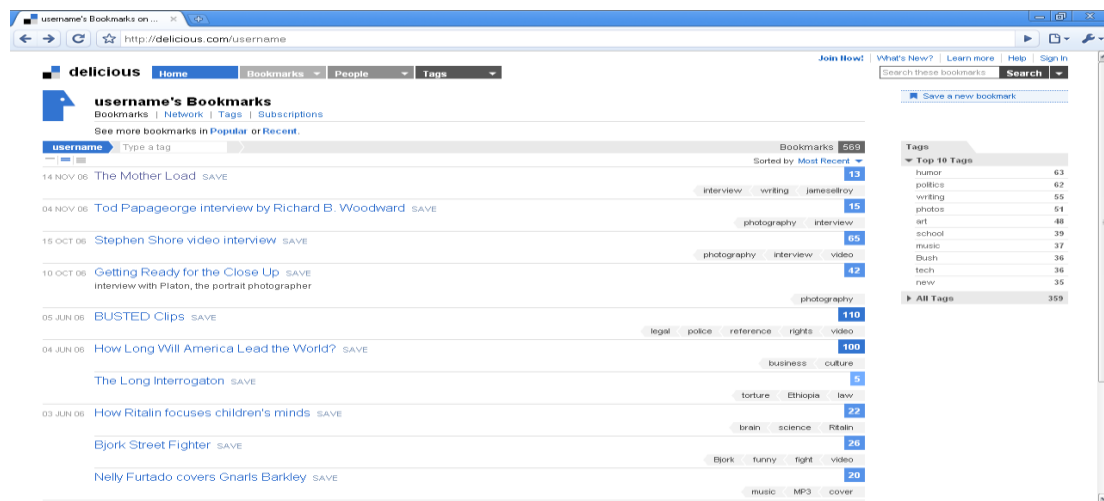
Ανάλογα ισχύουν και για τους δύο άλλους διμερείς γράφους που προέκυψαν από το αρχικό τριμερές μοντέλο. Έτσι, ο γράφος CI οδηγεί σε ένα άλλο σημασιολογικό δίκτυο, όπου οι συνδέσεις μεταξύ των όρων έχουν βάρος ανάλογο με τον αριθμό των αντικειμένων που επισημαίνονται και με τους δύο όρους. Αυτός ο τύπος σημασιολογικού δικτύου είναι πιο οικείος και μοιάζει με την βασική μέθοδο που εφαρμόζεται στην εξόρυξη κειμένου, όπου οι όροι συνήθως συσχετίζονται βάσει των συμφραζομένων τους. Ο γράφος AI οδηγεί σε ένα άλλο σημασιολογικό δίκτυο χρηστών, όπου το βάρος κάθε ζεύγους δίνεται από τον αριθμό των αντικειμένων που έχουν και τα δύο επισημειωθεί. Παίρνουμε επίσης ένα δίκτυο στιγμιότυπων με συσχετίσεις που δείχνουν τον αριθμό των χρηστών που έχουν επισημειώσει ένα δεδομένο ζευγάρι στιγμιότυπων [16].

3.2.7 Εφαρμογές

Στα συστήματα κοινωνικών ταξονομιών μόλις οι χρήστες δημιουργήσουν τους λογαριασμούς τους, μπορούν να αρχίσουν να κάνουν σελιδοποίηση (bookmark) σε διαδικτυακούς πόρους. Κάθε τέτοια εγγραφή καταγράφει τη διεύθυνση (URL), τον τίτλο καθώς και τον χρόνο που δημιουργήθηκε η εγγραφή αυτή. Με αυτόν τον τρόπο ο χρήστης αποθηκεύει έτσι τη διεύθυνση ενός διαδικτυακού πόρου που επιθυμεί να επισκεφθεί στο μέλλον. Για να δημιουργηθεί μια συλλογή μεμονωμένων σελιδοποιήσεων, ο χρήστης εγγράφεται σε ένα διαδικτυακό τόπο κοινωνικών ταξονομιών όπου έχει την δυνατότητα να κάνει σελιδοποίηση, να προσθέτει επισημειώσεις της επιλογής του και να υποδεικνύει τις μεμονωμένες σελιδοποιήσεις δημόσιες ή ιδιωτικές.

Στα συστήματα κοινωνικών ταξονομιών η ομαδοποίηση των πόρων του διαδικτύου στους οποίους έχει γίνει επισημείωση, γίνεται σύμφωνα με τη μέρα, κάθε σύνδεσμος αναγράφει το όνομα του χρήστη που δημιούργησε την επισημείωση, τον τίτλο του διαδικτυακού πόρου, το URL του και οποιαδήποτε επισημείωση λέξης κλειδιού, ή σχόλια που αναφέρονται σε αυτή την εγγραφή.

Οι χρήστες μπορούν να επισημαίνουν ένα σελιδοδείκτη με πολλαπλές επισημειώσεις ή λέξεις κλειδιά της επιλογής τους. Κάθε χρήστης έχει μια προσωπική σελίδα στην οποία οι παρουσιάζονται οι σελιδοδείκτες του, όπως στην Εικόνα 13. Όλοι οι σελιδοδείκτες που έχει γράψει ο χρήστης παρουσιάζονται σε αντίστροφη χρονολογική διάταξη μαζί με έναν κατάλογο όλων των επισημειώσεων που ο χρήστης έχει δώσει σε κάθε μια από αυτές.



Εικόνα 13: Προσωπική σελίδα στο delicious.us

Πολλές βιβλιοθήκες χρησιμοποιούν την τεχνολογία των κοινωνικών ταξονομιών και προβάλλουν διεξόδους, αναπτύσσοντας ή εφαρμόζοντας στα υπάρχοντα συστήματα τους (όπως OPAC) συστήματα κοινωνικών επισημειώσεων και δημιουργίας κοινωνικών ταξονομιών. Αναφορικά, η Βιβλιοθήκη του Πανεπιστημίου της Πενσυλβάνιας ανέπτυξε το **PennTags** (tags.library.upenn.edu/), ένα ιστότοπο όπου το προσωπικό και οι φοιτητές του πανεπιστημίου μπορούν να κάνουν σελιδοσήμανση σε ποιοτικές ιστοσελίδες και αρχεία του OPAC και να τις μοιραστούν μεταξύ τους. Στην Ελλάδα συγκεκριμένα, αναπτύχθηκε ο πρώτος opac 2.0 (OPACIAL), όπως φαίνεται στην Εικόνα 14 με ελεύθερο σύστημα κοινωνικής επισημείωσης (<http://library.panteion.gr/opacial>) με πολυεδρική (φασετική) αναζήτηση με βάση το έτος και το είδος του υλικού και των θεματικών επικεφαλίδων, αλλά και με παράλληλη ενοποίηση της ψηφιακής βιβλιοθήκης με βάση τις θεματικές επικεφαλίδες που προβάλλονται.

Επίσης, πολλά μουσεία αναπτύσσουν νέους καταλόγους εμπνευσμένους από τις τεχνολογίες του Web 2.0. Συγκεκριμένα, το *National Leadership Grant for Research από το U.S. Institute of Museum and Library Sciences* (<http://www.steve.museum/>) ανέπτυξε ένα σημαντικό πρόγραμμα, όπου με έρευνες απέδειξαν ότι η χρήση συστημάτων κοινωνικής επισημείωσης αποδείχτηκαν ιδιαίτερα χρήσιμα για τα μουσεία και τους επισκέπτες τους.

Επιπλέον, υπάρχουν εμπορικά λογισμικά όπως το *AquaBrowser* και το *PRIMO*, οι νέας γενιάς κατάλογοι όπως αναφέρονται, με ενσωματωμένες αρκετές λειτουργίες όπως κοινωνικών επισημειώσεις, κοινωνικός σχολιασμός καθώς και πολυεδρική φασετική αναζήτηση. Τέλος, μια πολύ διαδεδομένη διαδικτυακή εφαρμογή συλλογικής καταλογογράφησης βιβλίων από απομακρυσμένους χρήστες είναι το **LibraryThing** [2].

ανταπεξέλθουν στην πληθώρα πληροφοριών που διαχέεται μέσα από τα νέα μέσα (wikis, blogs). Επιπλέον, είναι σύγχρονες και προσφέρουν επίκαιρους και νέους όρους, το οποίο δεν είναι εύκολο σε μια ελεγχόμενη ιεραρχική ταξινόμια. Τέλος, οι κοινωνικές ταξονομίες προορίζονται και για την ανακάλυψη νέων πόρων σε αντίθεση με τις ιεραρχικές που σχεδιάζονται για την εύρεση συγκεκριμένων πόρων.

Υπάρχει μια θεμελιώδης διαφορά στις δραστηριότητες της περιήγησης για την εύρεση ενδιαφέροντος περιεχομένου, σε αντίθεση με την άμεση έρευνα για την εξεύρεση σχετικών εγγράφων μέσα από μια επερώτηση. Οι πληροφορίες που αναζητεί ο εκάστοτε χρήστης ποικίλλει ανάλογα με το περιεχόμενο που μπορεί να διαχειριστεί. Αν και θα μπορούσε κανείς να αξιολογήσει μια κοινωνική ταξονομία σε ένα σύστημα όπως το Delicious ή Flickr, χρησιμοποιώντας συγκεκριμένες επερωτήσεις από τους χρήστες και αξιολογώντας ποια έγγραφα, που έχουν επισημειωθεί με συγκεκριμένες λέξεις-κλειδιά, έχουν επιλεγεί από τους χρήστες και είναι σχετικά με το επερώτημα, θα αγνοήσει το ευρύτερο σύνολο των δραστηριοτήτων περιήγησης που το σύστημα φαίνεται να είναι «ισχυρότερο» μέσα σε αυτό.

Η αξιολόγηση της χρησιμότητας αυτής της πτυχής, θα απαιτούσε πιθανόν ποιοτική έρευνα, με τη μορφή συνεντεύξεων ή εθνογραφική μελέτης των χρηστών. Θα απαιτούσε επίσης συγκρίσεις όχι στην αναζήτηση βάσει συστημάτων ανάκτησης πληροφοριών, αλλά στις δραστηριότητες περιήγησης χρησιμοποιώντας διαφορετικές κατηγοριοποιήσεις και συστήματα ταξινόμησης [2].

3.2.9 Μειονεκτήματα

Ωστόσο, η ύπαρξη ενός μη ελεγχόμενου λεξιλογίου για την παροχή μιας γενικότερης ευελιξίας που φαίνεται ότι προσφέρουν οι κοινωνικές ταξονομίες δημιουργεί έναν αριθμό περιορισμών και κάποιες αδυναμίες στη χρήση τους, όπως αναλύεται στη συνέχεια.

Οι κοινωνικές ταξονομίες δεν πραγματοποιούν έλεγχο συνωνύμων. Η έλλειψη ελεγχόμενου λεξιλογίου συνεπάγεται την έλλειψη ενός επίσημου όρου για την περιγραφή μιας έννοιας ή μιας οντότητας. Έτσι, όταν διαφορετικοί χρήστες περιγράφουν τα αντικείμενα χρησιμοποιώντας διαφορετικούς όρους, πιθανώς οι όροι αυτοί να εννοούν το ίδιο πράγμα (να είναι δηλαδή συνώνυμα). Αποτέλεσμα αυτού είναι ότι δυσχεραίνεται η διαδικασία αναζήτησης. Όμως, ακόμα και για ίδιες λεκτικά επισημειώσεις, η διαφορετική μορφή επισημείωσής τους (π.χ. «New York City», «New_York_City», «New.York.City», «New York», «NYC», κλπ) από τους χρήστες θα οδηγούσε στο ίδιο πρόβλημα.

Σχετικό με το πρόβλημα συνωνύμων είναι επίσης και το πρόβλημα του πληθυντικού επισημειώσεων. Έτσι, για παράδειγμα οι επισημειώσεις «Cat» και «Cats» εφόσον θεωρούνται διακριτές, το σύστημα θα επιστρέψει αποτελέσματα μόνο για τον ένα από τους δύο όρους που θα ζητηθεί αν και είναι προφανές ότι θα ενδιέφεραν και των δύο.

Οι κοινωνικές ταξονομίες χαρακτηρίζονται από έλλειψη ακρίβειας. Όπως αναφέρθηκε και προηγουμένως οι κοινωνικές ταξονομίες αποτελούν συστήματα ανακάλυψης, χωρίς την υψηλή δυναμικότητα αναζήτησης των ιεραρχικών ταξονομιών και έτσι παρουσιάζουν χαμηλό βαθμό ακρίβειας. Επίσης δεν διαθέτουν ιεραρχία και αποτελούν επίπεδα συστήματα, χωρίς την ύπαρξη σχέσεων γονέα-παιδιού, κατηγοριών και υποκατηγοριών.

Επιπλέον, οι κοινωνικές ταξονομίες παρουσιάζουν το πρόβλημα «βασικού επιπέδου». Οι χρήστες έχουν διαφορετική ιδέα για το βάθος επισημείωσης συγκεκριμένων οντοτήτων, αν

δηλαδή θα χρησιμοποιήσουν ένα ευρύτερο ή ένα στενότερο όρο. Λόγω της διαφορετικότητάς τους, κάποιοι θεωρούν ορισμένους όρους πιο κατάλληλους για τα διάφορα επίπεδα λεπτομέρειας στην περιγραφή ενός αντικειμένου, κάτι που εξαρτάται από την προσωπική τους κρίση και εμπειρία ώστε να αντιλαμβάνονται το «βασικό επίπεδο» ενός αντικειμένου ως κάτι γενικότερο ή ειδικότερο.

Οι κοινωνικές ταξινομίες παρουσιάζουν έλλειψη ανάκλησης, καθώς δεν υπάρχει έλεγχος συνωνύμων, μια αναζήτηση κοινωνικής ταξινόμιας δε θα επιστρέφει μια πλήρη λίστα αποτελεσμάτων λόγω της χρήσης παρομοίων επισημειώσεων.

Τέλος, οι κοινωνικές ταξινομίες είναι ευάλωτες στο «gaming». Το «gaming» συνιστά μια ανάλογη διαδικασία με το spamming, όπου κάποιος κακόβουλος χρήστης διαδίδει εσκεμμένα λανθασμένες επισημειώσεις ή υπερβολικό αριθμό αυτών ώστε να παρενοχλεί την ορθή λειτουργία του συστήματος.

Συνοψίζοντας, με την ανάλυση και ορισμένων σημαντικών μειονεκτημάτων των κοινωνικών ταξινόμιών, κάποιος θα αναρωτηθεί αν αξίζει τελικά η χρήση τους συγκριτικά με τα παραδοσιακά συστήματα ταξινόμησης. Ξεκάθαρη απάντηση μπορεί να δοθεί ύστερα από αξιολόγηση διαφόρων παραγόντων ανά περίπτωση που μελετούνται. Συχνά όμως ένα σύστημα που προσπαθεί να συλλάβει την πλήρη αξία των αντικειμένων του, θα κοστίσει πολύ περισσότερο στον χρόνο του χρήστη και στην προσπάθειά του, οπότε και τελικά θα έχει μικρή αξία στην πράξη [9], [13], [17].

3.3 Web Personalization

3.3.1 Εισαγωγή

Τα τελευταία χρόνια το ενδιαφέρον για την ανάλυση της συμπεριφοράς των χρηστών στο Διαδίκτυο έχει αυξηθεί ραγδαία. Αυτό οφείλεται στο γεγονός ότι η προστιθέμενη αξία που έχουν οι επισκέψεις των διαδικτυακών τόπων, δεν έχει αποκτηθεί μέσω των μεγαλύτερων ποσοτήτων δεδομένων που μπορεί να βρει κανείς μέσα σε ένα ιστότοπο, αλλά μέσω της ευκολότερης πρόσβασης στις απαιτούμενες πληροφορίες την κατάλληλη στιγμή και με την πιο κατάλληλη μορφή.

Έρευνες σχετικές με την χρήση του διαδικτύου αναμένουν ότι ο αριθμός των χρηστών θα αυξηθεί πάνω από το 1.5 δισεκατομμύριο μέχρι το 2012. Η πλειοψηφία αυτών των χρηστών είναι μη ειδικοί πάνω σε θέματα πληροφορικής, καθιστώντας τους σχεδόν ανίκανους να συμβαδίσουν με την ταχεία ανάπτυξη των τεχνολογιών πληροφορικής, παρόλο που την ίδια στιγμή αναγνωρίζουν το Διαδίκτυο ως μία ανεκτίμητη πηγή πληροφορίας για την καθημερινή τους ζωή. Η αυξανόμενη χρήση του Διαδικτύου επιταχύνει επίσης το ρυθμό με τον οποίο οι πληροφορίες αυτές είναι διαθέσιμες online. Διάφορες έρευνες που έχουν γίνει πάνω στο Διαδίκτυο, εκτιμούν ότι περίπου ένα εκατομμύριο ιστοσελίδες προστίθενται κάθε μέρα, πάνω από 600 GB σελίδων αλλάζουν ανά μήνα και ένας δικτυακός εξυπηρετητής χρησιμοποιείται κάθε δύο ώρες. Σήμερα, πάνω από τρία δισεκατομμύρια ιστοσελίδες είναι διαθέσιμες στο Διαδίκτυο (σχεδόν μία σελίδα για κάθε δύο άτομα πάνω στη γη) [18].

Βάσει των παραπάνω στοιχείων, παρατηρείται μια ραγδαία και συνεχόμενη αύξηση των χρηστών που έχει σαν φυσικό επακόλουθο την συνεχόμενη αύξηση της πληροφορίας στο Διαδίκτυο, προσελκύνοντας με αυτό τον τρόπο νέους χρήστες. Αυτό το μοτίβο είναι υπεύθυνο για την «έκρηξη» του Διαδικτύου, η οποία έχει προκαλέσει το φαινόμενο της «υπερφόρτωσης

πληροφοριών» (information overload) που βρίσκει αντίκτυπο κυρίως στους διαδικτυακούς χρήστες.

Επιπλέον, η εμφάνιση νέων ηλεκτρονικών υπηρεσιών, όπως το ηλεκτρονικό εμπόριο, το e-learning και το e-banking, έχουν αλλάξει ριζικά τον τρόπο με τον οποίο χρησιμοποιείται το Διαδίκτυο, μετατρέποντας τις ιστοσελίδες σε ηλεκτρονικές επιχειρήσεις, αυξάνοντας παράλληλα τον ανταγωνισμό μεταξύ τους. Με τους ανταγωνιστές να είναι «ένα κλικ μακριά», η απαίτηση για προστιθέμενη αξία στις ηλεκτρονικές υπηρεσίες έχει καταστεί αναγκαία συνθήκη μπροστά στο φαινόμενο δημιουργίας επισκεπτών-πελατών σε μια δικτυακή τοποθεσία. Η εστίαση σε συγκεκριμένες ατομικές ανάγκες και η παροχή προσαρμοσμένων προϊόντων και υπηρεσιών αποτελεί το κινητήριο μοχλό για τη δημιουργία ξεχωριστών υπηρεσιών με αυξημένη προστιθέμενη αξία.

Η εξατομίκευση των υπηρεσιών που προσφέρονται από μια διαδικτυακή τοποθεσία είναι ένα σημαντικό βήμα προς την αποσυμφόρηση της υπερφόρτωσης των πληροφοριών, μετατρέποντας το διαδίκτυο σε ένα φιλικότερο περιβάλλον για τον κάθε χρήστη και δημιουργώντας αξιόπιστες σχέσεις μεταξύ του Παγκόσμιου Ιστού και του επισκέπτη-πελάτη. Με τον όρο **Web Personalization** ορίζεται η διαδικασία προσαρμογής των διαδικτυακών συστημάτων πληροφορίας στις ανάγκες, τις επιθυμίες και τα συμφέροντα του κάθε χρήστη ξεχωριστά [19].

Τυπικά, ένας εξατομικευμένος ιστότοπος αναγνωρίζει τους χρήστες του, συλλέγει πληροφορίες σχετικά με τις προτιμήσεις τους και προσαρμόζει τις υπηρεσίες του, έτσι ώστε να ταιριάζουν και να συμβαδίζουν με τις ανάγκες των χρηστών του. Το Web Personalization βελτιώνει τη δικτυακή εμπειρία του επισκέπτη παρουσιάζοντας τις πληροφορίες που θέλει να δει με τον καταλληλότερο τρόπο και στην κατάλληλη στιγμή. Για παράδειγμα στο ηλεκτρονικό επιχειρείν, το Web Personalization ορίζει επιπλέον μηχανισμούς για να αποκτήσει περισσότερες πληροφορίες σχετικά με τις ανάγκες των πελατών του, να προσδιορίσει τις μελλοντικές τάσεις και εν τέλει να αυξήσει την εμπιστοσύνη των πελατών προς την παρεχόμενη υπηρεσία.

Μέχρι σήμερα, το Web Personalization έχει συσχετιστεί κυρίως με «recommender» συστήματα, προσαρμοσμένες ιστοσελίδες για εφημερίδες και ηλεκτρονικά καταστήματα. Ωστόσο, πολλές έρευνες υποστηρίζουν ότι το Web Personalization περιλαμβάνει μια ποικιλία λειτουργιών που κυμαίνονται από την απλή αναγνώριση του χρήστη έως την πιο προηγμένη λειτουργικότητα, όπως η εκτέλεση ορισμένων καθηκόντων για λογαριασμό του χρήστη. Αυτή η λειτουργικότητα προσφέρεται από διαδικτυακά προσωποποιημένα συστήματα σύμφωνα με μια πολιτική εξατομίκευσης που καθορίζει τον τρόπο με τον οποίο η προσωποποιημένη υπηρεσία θα χρησιμοποιηθεί από τον τελικό χρήστη.

Στην πλειονότητα των υφιστάμενων εμπορικών προσωποποιημένων συστημάτων, η διαδικασία εξατομίκευσης απαιτεί σημαντικές χειροκίνητες εργασίες και τις περισσότερες φορές, σημαντική προσπάθεια από την πλευρά του χρήστη. Παρόλα αυτά, ο αριθμός των ιστοσελίδων που υποστηρίζουν εξατομικευμένες υπηρεσίες αυξάνεται συνεχώς. Συγκεκριμένα, αναφέρεται ότι οι παγκόσμιες επενδύσεις σε προσωποποιημένες τεχνολογίες έφτασαν τα 2,1 δισεκατομμύρια δολάρια το 2006, όπου το ήμισυ των επενδύσεων έγινε από εταιρείες χρηματοδοτικών υπηρεσιών και λιανικού εμπορίου. Επίσης οι συγκεκριμένες τεχνολογίες είναι δημοφιλείς στο τομέα των τηλεπικοινωνιών και της ψυχαγωγίας.

Ολοκληρώνοντας, τα τελευταία χρόνια παρατηρείται πλήρης αποδοχή των συγκεκριμένων υπηρεσιών και υιοθέτηση αυτής της νέας τεχνολογίας από τους χρήστες

καθώς τα συστήματα αυτά είναι πολύ ευέλικτα και προσαρμόζονται με ταχύ ρυθμό στις απαιτήσεις και στις ανάγκες του εκάστοτε χρήστη-πελάτη τους.

3.3.2 *Web Personalization Roadmap*

Όπως προαναφέρθηκε, ο όρος Web Personalization καλύπτει τις μεθόδους και τις τεχνικές που χρησιμοποιούνται, προκειμένου να βελτιωθεί η εμπειρία που αποκτούν οι επισκέπτες ενός ιστότοπου μέσω της περιήγησης τους σε αυτόν. Η βελτίωση αυτή μπορεί να αποκτηθεί μέσω μιας ποικιλίας λειτουργιών που μπορεί να προσφέρει ένα διαδικτυακό προσωποποιημένο σύστημα, το οποίο καθιστά ευκολότερη την αλληλεπίδραση με τον ιστότοπο, εξοικονομεί χρόνο στους χρήστες, και ως εκ τούτου ικανοποιείται ένας από τους κύριους στόχους των εφαρμογών, ο οποίος είναι η δημιουργία «πιστών επισκεπτών». Στις παρακάτω ενότητες, εξετάζονται οι προσωποποιημένες λειτουργίες που μπορούν να προσφερθούν από ένα Web Personalization σύστημα, μαζί με ένα σύνολο απαιτήσεων για το σχεδιασμό και την εφαρμογή ενός τέτοιου συστήματος.

3.3.2.1 *Λειτουργίες του Web Personalization*

Ένα διαδικτυακό προσωποποιημένο σύστημα μπορεί να προσφέρει μια ποικιλία λειτουργιών, αρχής γενομένης από το χαιρετισμό προς έναν απλό χρήστη, καθώς και πιο περίπλοκες λειτουργίες όπως η παράδοση προσωπικού περιεχομένου. Σε πολλές διαδικτυακές έρευνες προτείνεται η ταξινόμηση των Web Personalization λειτουργιών, η οποίες μπορούν να επεκταθούν σε ένα γενικότερο σχήμα ταξινόμησης. Το προτεινόμενο σχήμα λαμβάνει υπόψη αυτό που προσφέρεται σήμερα από τα εμπορικά συστήματα και τα ερευνητικά πρότυπα, καθώς και τι είναι δυνητικά εφικτό από τέτοιου είδους συστήματα. Έχουν διακριθεί τέσσερις βασικές κατηγορίες Web Personalization λειτουργιών, η απομνημόνευση, η καθοδήγηση, η παραμετροποίηση και η υποστήριξη εκτέλεσης διαδικασιών. Καθεμία από αυτές εξετάζεται λεπτομερέστερα παρακάτω [20], [23].

3.3.2.1.1 *Απομνημόνευση (Memorization)*

Η απομνημόνευση θεωρείται η απλούστερη μορφή μιας Web Personalization λειτουργίας, κατά την οποία το σύστημα καταγράφει και αποθηκεύει στη μνήμη πληροφορίες σχετικές με το χρήστη, όπως το όνομα του και το ιστορικό περιήγησης του. Συγκεκριμένα, όταν ο χρήστης επιστρέφει στην ιστοσελίδα, οι συγκεκριμένες πληροφορίες χρησιμοποιούνται ως υπενθύμιση για την μέχρι τώρα δραστηριότητα του, χωρίς περαιτέρω επεξεργασία. Συνήθως η απομνημόνευση δεν προσφέρεται ως αυτόνομη λειτουργία, αλλά ως μέρος μιας πιο ολοκληρωμένης προσωποποιημένης λύσης. Μερικά παραδείγματα της συγκεκριμένης κατηγορίας λειτουργιών παρατίθενται στη συνέχεια.

- **Χαιρετισμός Χρήστη:** Το διαδικτυακό προσωποποιημένο σύστημα αναγνωρίζει τον κάθε εγγεγραμμένο χρήστη και εμφανίζει το όνομά του μαζί με ένα μήνυμα καλωσορίσματος. Διάφορες εμπορικές εφαρμογές χρησιμοποιούν τους χαιρετισμούς για τους πελάτες τους ή τους εγγεγραμμένους χρήστες τους. Αν και αυτή θεωρείται μια απλή λειτουργία, είναι το πρώτο βήμα προς την αύξηση της εμπιστοσύνης των επισκεπτών, δεδομένου ότι οι χρήστες αισθάνονται πιο άνετα με τους ιστότοπους που τους αναγνωρίζουν ως άτομα, και όχι ως τακτικούς επισκέπτες.
- **Bookmarking:** Στην συγκεκριμένη λειτουργία το σύστημα αποθηκεύει τις ιστοσελίδες που ένας χρήστης έχει επισκεφθεί στο παρελθόν και τις παρουσιάζει μέσω ενός personalized bookmarking σχήματος που έχει η εκάστοτε εφαρμογή.

- **Εξατομικευμένα δικαιώματα πρόσβασης:** Μια ιστοσελίδα μπορεί να χρησιμοποιήσει εξατομικευμένα δικαιώματα πρόσβασης, προκειμένου να διαχωρίσει τους εξουσιοδοτημένους χρήστες από τους κοινούς. Διαφορετικά δικαιώματα πρόσβασης μπορεί να απαιτούνται για διαφορετικά είδη πληροφορίας, όπως σε αναφορές, στις τιμές των προϊόντων, καθώς και για την εκτέλεση συγκεκριμένων εφαρμογών στο Διαδίκτυο, όπως το ftp ή το e-mail.

3.3.2.1.2 Καθοδήγηση

Η συγκεκριμένη λειτουργία αναφέρεται στην προσπάθεια που κάνει ένα Web Personalization σύστημα να βοηθήσει τον χρήστη να πάρει γρήγορα τις πληροφορίες που αναζητά σε μια εφαρμογή και να παρέχει εναλλακτικές επιλογές περιήγησης. Αυτή η λειτουργία όχι μόνο αυξάνει την «πίστη» των χρηστών, αλλά βοηθάει σε μεγάλο βαθμό στο πρόβλημα υπερφόρτωσης πληροφοριών που οι χρήστες μιας μεγάλης εφαρμογής ενδέχεται να αντιμετωπίσουν. Μερικά παραδείγματα καθοδήγησης παρατίθενται παρακάτω.

- **Σύσταση υπερσυνδέσεων (recommendation of hyperlinks):** Η λειτουργία αυτή αναφέρεται στη σύσταση ενός συνόλου υπερσυνδέσεων που σχετίζονται με τα ενδιαφέροντα και τις προτιμήσεις του χρήστη. Η παρουσίαση των προτεινόμενων συνδέσεων γίνεται είτε σε ξεχωριστό πλαίσιο από την ιστοσελίδα, είτε σε ένα αναδυόμενο παράθυρο. Σε διάφορες έρευνες, η λειτουργία αυτή περιγράφεται ως μια προσαρμοστική σύσταση και μπορεί να λάβει τη μορφή της σύστασης συνδέσεων σε ορισμένα προϊόντα, θέματα πληροφοριών, ή μονοπάτια πλοήγησης που ένας χρήστης θα μπορούσε να ακολουθήσει. Επίσης, η σύσταση υπερσυνδέσεων είναι μια από τις πιο συχνά προσφερόμενες προσωποποιημένες λειτουργίες και υποστηρίζεται από μια σειρά συστημάτων, όπως το Web Personalizer [21].
- **Διδασκαλία Χρήστη:** Αυτή η λειτουργία «δανείζεται» τη βασική έννοια των προσαρμοστικών εκπαιδευτικών συστημάτων και την χρησιμοποιεί σε εφαρμογές στο Διαδίκτυο. Μια προσωπική ιστοσελίδα μπορεί να προσφέρει υπηρεσίες καθοδήγησης για ένα άτομο σε κάθε βήμα της αλληλεπίδρασης του με την ιστοσελίδα, βάσει των γνώσεων και των ενδιαφερόντων του. Αυτή η λειτουργία επιτυγχάνεται είτε με τη σύσταση άλλων ιστοσελίδων, είτε με την προσθήκη επεξηγηματικού περιεχομένου στις ίδιες τις εφαρμογές. Μια εφαρμογή αυτής της λειτουργίας μπορεί να βρεθεί σε *Webinars* (διαδικτυακά σεμινάρια), τα οποία γίνονται ζωντανά ή επαναλαμβάνουν πολυμεσικές παρουσιάσεις που πραγματοποιούνται σε μια διαδικτυακή τοποθεσία.

3.3.2.1.3 Παραμετροποίηση

Η παραμετροποίηση ως προσωποποιημένη λειτουργία αναφέρεται στη διαδικασία τροποποίησης της ιστοσελίδας όσον αφορά στο περιεχόμενο, τη δομή και τη διάταξη, προκειμένου να ληφθούν υπόψη οι γνώσεις, οι προτιμήσεις και τα ενδιαφέροντα του χρήστη. Ο κύριος στόχος της είναι η διαχείριση του φορτίου των πληροφοριών, μέσω της διευκόλυνσης της αλληλεπίδρασης του χρήστη με την εφαρμογή. Μερικά από τα πιο χαρακτηριστικά παραδείγματα αυτής της κατηγορίας λειτουργιών παρατίθενται παρακάτω.

- **Personalized layout:** Η συγκεκριμένη λειτουργία κληρονομήθηκε από τις «προσαρμοστικές διεπαφές χρήστη» (Adaptive User Interfaces), όπου μια συγκεκριμένη ιστοσελίδα αλλάζει τον τρόπο εμφάνισής της, το χρώμα και τις

εσωτερικές της πληροφορίες, βάσει του προφίλ του χρήστη. Επίσης η εν λόγω λειτουργία αξιοποιείται συνήθως από Διαδικτυακές Πύλες, όπως η Yahoo και η Altavista, οι οποίες προσφέρουν προσαρμοσμένες λειτουργίες προκειμένου να δημιουργούν personalized «My Portals» ιστότοπους.

- **Παραμετροποίηση περιεχομένου:** Το περιεχόμενο της ιστοσελίδας που παρουσιάζεται σε ένα χρήστη μπορεί να τροποποιηθεί προκειμένου να προσαρμοστεί στις γνώσεις, τα συμφέροντα και τις προτιμήσεις του. Για παράδειγμα, η ίδια σελίδα μπορεί να παρουσιαστεί σε διαφορετικούς χρήστες σε μια συνοπτική ή εκτεταμένη μορφή ανάλογα με το είδος του χρήστη. Μια τέτοια χαρακτηριστική παραμετροποιημένη ιστοσελίδα είναι η *UM2001*.
- **Παραμετροποίηση των υπερσυνδέσεων:** Η διαδικασία της παραμετροποίησης μπορεί να επιτευχθεί και για τις συνδέσεις εντός μιας σελίδας. Σε αυτή την περίπτωση, η ιστοσελίδα τροποποιείται με την προσθήκη ή την αφαίρεση συνδέσεων εντός μιας συγκεκριμένης σελίδας. Η συγκεκριμένη διαδικασία μπορεί να οδηγήσει στη βελτιστοποίηση ολόκληρης της δομής της ιστοσελίδας μέσω της αφαίρεσης των συνδέσμων που είναι αχρησιμοποίητοι και τροποποιώντας την τοπολογία της εφαρμογής με σκοπό να καταστεί πιο εύχρηστη.
- **Προσωποποιημένο σχήμα τιμολόγησης:** Μια ιστοσελίδα μπορεί να προσφέρει διαφορετικές τιμές και τρόπους πληρωμής για διαφορετικούς χρήστες, όπως είναι οι εκπώσεις και οι δόσεις σε χρήστες που έχουν αναγνωριστεί από την εφαρμογή ως «πιστοί πελάτες». Από την Amazon έγινε μια προσπάθεια παροχής παρόμοιας λειτουργικότητας, στην οποία χρεώνονταν διαφορετικοί πελάτες με διαφορετικές τιμές για το ίδιο προϊόν. Παρόλα αυτά, η συγκεκριμένη προσπάθεια αμφισβητήθηκε κυρίως από νομικής άποψης, λόγω της μη επικοινωνίας και μη δικαιολόγησης των λόγων που υπήρχαν αυτές οι διαφορές στις τιμές. Η συγκεκριμένη λειτουργία σε συνδυασμό με τη σύσταση υπερ-συνδέσεων, που αναλύθηκε παραπάνω, μπορούν να χρησιμοποιηθούν σε e-commerce ιστότοπους προκειμένου να προσελκύσουν επισκέπτες οι οποίοι δεν είναι τη συγκεκριμένη χρονική στιγμή και αγοραστές.
- **Προσωποποιημένη διαφοροποίηση των προϊόντων:** Σε όρους μάρκετινγκ, το personalization μπορεί να είναι και μια ισχυρή μέθοδος μετατροπής ενός προτύπου προϊόντος σε μια εξειδικευμένη λύση για ένα άτομο.

3.3.2.1.4 Υποστήριξη εκτέλεσης διαδικασιών

Η υποστήριξη εκτέλεσης διαδικασιών είναι μια λειτουργία που συνεπάγεται την εκτέλεση μιας συγκεκριμένης ενέργειας για λογαριασμό του χρήστη. Αυτή είναι η πιο προηγμένη personalized λειτουργία, η οποία κληρονομήθηκε από μια κατηγορία προσαρμοστικών συστημάτων που είναι γνωστή ως «προσωπικοί βοηθοί» (personalized assistants), η οποία μπορεί να θεωρηθεί ως client-side personalized σύστημα. Επίσης, η ίδια λειτουργικότητα μπορεί να χρησιμοποιηθεί και σε personalized συστήματα που απασχολούνται από ένα Web διακομιστή. Παραδείγματα αυτής της κατηγορίας των λειτουργιών παρατίθενται στη συνέχεια.

- **Προσωποποιημένη εξυπηρέτηση (personalized errands):** Το διαδικτυακό προσωποποιημένο σύστημα μπορεί να εκτελέσει μια σειρά από ενέργειες και να βοηθήσει το έργο του χρήστη, όπως κατά την αποστολή ενός e-mail. Ανάλογα με την πολυπλοκότητα του personalized συστήματος, αυτά τα «θελήματα» μπορεί να περιλαμβάνουν απλές ενέργειες-ρουτίνες μέχρι πιο περίπλοκες που λαμβάνουν υπόψη τις προσωπικές συνθήκες του χρήστη.
- **Προσωποποιημένη συμπλήρωση επερωτήματος:** Το σύστημα μπορεί να ολοκληρώσει ή ακόμη και να ενισχύσει μέσω εισαγωγής νέων όρων, τα επερωτήματα

του χρήστη που υποβάλλονται είτε σε μια μηχανή αναζήτησης είτε σε ένα σύστημα βάσεων δεδομένων στο Διαδίκτυο. Με τον τρόπο αυτό, το personalization μπορεί να βοηθήσει στη βελτίωση της απόδοσης ενός συστήματος ανάκτησης πληροφοριών.

- **Προσωποποιημένες διαπραγματεύσεις:** Το Web Personalization σύστημα μπορεί να λειτουργήσει ως διαπραγματευτής για λογαριασμό του χρήστη και να συμμετάσχει για παράδειγμα, σε δημοπρασίες στο Διαδίκτυο. Αυτή είναι μία από τις πιο προηγμένες λειτουργίες υποστήριξης εκτέλεσης διαδικασιών, που απαιτούν υψηλό βαθμό πολυπλοκότητας από το personalization σύστημα, με σκοπό να κερδίσει την εμπιστοσύνη των χρηστών του [22], [23].

3.3.2.2 Απαιτήσεις για το σχεδιασμό ενός Web Personalization συστήματος

Όπως αναλύθηκε και στο προηγούμενο κεφάλαιο, υπάρχει μια ευρεία ποικιλία λειτουργιών που μπορούν να χρησιμοποιηθούν σε ένα Web Personalization σύστημα. Αυτές οι λειτουργίες επιβάλλουν ορισμένες απαιτήσεις σχετικά με το σχεδιασμό ενός τέτοιου συστήματος, οι οποίες στοχεύουν στην ανάπτυξη ενός ισχυρού και ευέλικτου συστήματος.

Αρχικά, η λειτουργικότητα που προσφέρει ένα Web Personalization σύστημα είναι *domain sensitive*. Το ίδιο σύστημα λειτουργώντας σε διαφορετικούς τομείς όπως το ηλεκτρονικό εμπόριο, η ψηφιακή βιβλιοθήκη και τα portals μπορεί να προσφέρει διαφορετικές προσωποποιημένες λειτουργίες. Έτσι, το domain κάτω από το οποίο θα λειτουργήσει ένα διαδικτυακό προσωποποιημένο σύστημα θα πρέπει να προσδιοριστεί και να περιγραφεί διεξοδικά.

Επίσης, η ταυτοποίηση του χρήστη που έχει πρόσβαση σε μια τοποθεσία στο διαδίκτυο είναι σημαντική, προκειμένου να γίνει διαχωρισμός μεταξύ εγγεγραμμένων χρηστών και χρηστών που χρησιμοποιούν την εφαρμογή για πρώτη φορά. Ωστόσο, υπάρχουν προβληματισμοί πάνω σε θέματα ιδιωτικότητας και ανωνυμίας στο Διαδίκτυο που δεν το καθιστούν πάντοτε αυτό δυνατό. Αυτό όμως που μπορεί να είναι δυνατό, είναι ο προσδιορισμός του στόχου και των κινήτρων του χρήστη που έχει πρόσβαση στην εφαρμογή. Το συγκεκριμένο θεωρείται ένα κρίσιμο ζήτημα, δεδομένου ότι οι ακριβείς εκτιμήσεις των στόχων του χρήστη συμβάλλουν στην ορθή λειτουργία της προσωποποίησης, δίνοντας με αυτό τον τρόπο στον χρήστη μια βελτιωμένη εμπειρία περιήγησης.

Ένα personalized σύστημα θα πρέπει να είναι σε θέση να συλλέγει όλα τα σχετικά δεδομένα από το χρήστη που χρειάζονται για την ορθή λειτουργία του personalization. Το είδος και η ποσότητα των δεδομένων που απαιτούνται, εξαρτάται από τις personalization λειτουργίες που έχουν επιλεγεί να πραγματοποιηθούν. Η συλλογή των δεδομένων του χρήστη είναι μια συνεχής διαδικασία και λόγω της φύσης του Διαδικτύου, το σύστημα θα πρέπει να είναι σε θέση να χειριστεί μεγάλο όγκο δεδομένων, καθώς και να αυξάνει την μεταβλητότητα των δεδομένων αυτών.

Επιπλέον, τα δεδομένα που συλλέγονται θα πρέπει να υποβάλλονται σε επεξεργασία, προκειμένου να διαχωριστούν από άλλα σχετικά δεδομένα, να συσχετιστούν, να αξιολογηθούν και εν τέλει να μορφοποιηθούν, έτσι ώστε να είναι έτοιμα για personalization. Η επεξεργασία δεδομένων είναι μια domain-dependent διαδικασία και συνεπώς απαιτείται μεγάλη ευελιξία από το personalization σύστημα έτσι ώστε να είναι σε θέση να προσαρμοστεί σε διαφορετικά domains και στις αντίστοιχες προσωποποιημένες λειτουργίες.

Το Web Personalization σύστημα θα πρέπει να είναι σε θέση να δημιουργεί και να διατηρεί αποτελεσματικά και με ακρίβεια τα μοντέλα χρήστη, όπως είναι οι πληροφορίες που κατέχει το σύστημα για τα ενδιαφέροντα, τη γνώση, τους στόχους και τις προτιμήσεις των χρηστών. Η κατασκευή μοντέλων χρήστη μπορεί να γίνει είτε χειροκίνητα είτε μέσω μιας μηχανής μάθησης. Το εγχειρίδιο της διαδικασίας της κατασκευής τέτοιων μοντέλων περιλαμβάνει τον αυτο-προσδιορισμό ενός μοντέλου από ένα χρήστη και / ή την ανάλυση δεδομένων χρήστη από έναν ειδικό για την κατασκευή κανόνων ταξινόμησης των χρηστών σε διαφορετικούς τύπους με συγκεκριμένα χαρακτηριστικά. Η αυτοματοποιημένη κατασκευή των μοντέλων χρήστη αξιοποιεί τεχνικές μηχανικής μάθησης, προκειμένου να δημιουργήσει και να περιγράψει την επιλογή των μοντέλων. Η επιλογή μεταξύ της χειροκίνητης δημιουργίας ή της χρήσης μιας αυτοματοποιημένης μεθόδου μάθησης, εξαρτάται κατά ένα μεγάλο βαθμό από το πεδίο εφαρμογής και τις λειτουργίες που προσφέρονται σε αυτό.

Ολοκληρώνοντας, οι προσωποποιημένες λειτουργίες πρέπει να είναι συναφείς με τους στόχους και τα κίνητρα του χρήστη και να τηρούν διάφορους σημαντικούς πρακτικούς περιορισμούς, όπως είναι ο χρόνος απόκρισης. Επιπλέον, τα προσωπικά στοιχεία του χρήστη θα πρέπει να προστατεύονται σε κάθε περίπτωση και θα πρέπει ο εκάστοτε χρήστης να γνωρίζει τον τρόπο με τον οποίο συλλέγονται και χρησιμοποιούνται τέτοιου είδους πληροφορίες [23].

3.3.2.3 Προσεγγίσεις στο Web Personalization

Κατά τη διάρκεια εξέλιξης του Διαδικτύου, το personalization έχει αναγνωριστεί ως η λύση για το πρόβλημα υπερφόρτωσης πληροφοριών και ως μέσο για την αύξηση της εμπιστοσύνης των επισκεπτών σε μια τοποθεσία στο Διαδίκτυο. Λόγω της σημασίας της προσωποποίησης στις διαδικτυακές υπηρεσίες, τα τελευταία χρόνια έχουν προταθεί πολλές Web Personalization τεχνικές. Στη συνέχεια παρατίθενται τρεις γενικές προσεγγίσεις στις οποίες κατατάσσονται οι Web Personalization τεχνικές.

- **Manual decision rule systems:** Σύμφωνα με αυτή την προσέγγιση, μια Web-based υπηρεσία «προσωποποιείται» μέσω της χειροκίνητης παρέμβασης του σχεδιαστή της και συνήθως μαζί με την συνεργασία του χρήστη. Τυπικά, τα στατικά μοντέλα χρηστών δημιουργούνται μέσω μιας διαδικασίας εγγραφής του χρήστη και μια σειρά από κανόνες που αφορούν στο περιεχόμενο που παρέχεται στους χρήστες του διαδικτύου οι οποίοι ανήκουν σε διαφορετικά μοντέλα και καθορίζονται χειροκίνητα. Δυο χαρακτηριστικά παραδείγματα που υιοθετούν αυτή την προσέγγιση είναι το *Yahoo! Personalization engine* και το *Websphere Personalization* (IBM) [24].
- **Content-based filtering systems:** Αυτό το είδος τεχνικών χρησιμοποιεί μεθόδους μηχανικής μάθησης για το περιεχόμενο στον Παγκόσμιο Ιστό, που είναι κυρίως κείμενο, προκειμένου να ανακαλύψει τις προσωπικές προτιμήσεις του χρήστη. Ένα εργαλείο που υιοθετεί την προσέγγιση αυτή είναι το *NewsWeeder*, το οποίο είναι σε θέση να κατασκευάζει προσαρμοσμένα μοντέλα χρήστη, με βάση την ομοιότητα που μπορεί να υπάρξει μεταξύ σχετικών δικτυακών εγγράφων που περιέχουν νέες πληροφορίες. Αυτά τα μοντέλα μπορεί να χρησιμοποιηθούν για το φιλτράρισμα των νέων αντικειμένων σύμφωνα με τις απαιτήσεις του κάθε χρήστη.
- **Social or collaborative filtering systems:** Ο σκοπός της προσέγγισης αυτής είναι η προσωποποίηση μιας υπηρεσίας, δίχως να απαιτείται η ανάλυση του περιεχομένου της. Εδώ, το personalization επιτυγχάνεται μέσω της αναζήτησης κοινών χαρακτηριστικών στις προτιμήσεις διαφορετικών χρηστών, οι οποίες εκφράζονται ρητά από αυτούς, με τη μορφή αξιολογήσεων (item ratings) και καταγράφονται από το σύστημα. Το *Recommendation Engine* (Net Perceptions) και το *Websphere*

Personalization (IBM) αποτελούν χαρακτηριστικά παραδείγματα υπηρεσιών που χρησιμοποιούν τη μέθοδο αυτή.

Τα manual decision rule systems πάσχουν από τα ίδια προβλήματα που έχουν και άλλα manual constructed πολύπλοκα συστήματα, καθώς απαιτείται σημαντική προσπάθεια στην κατασκευή και τη συντήρησή τους. Επιπλέον απαιτείται συνήθως και η συμμετοχή του χρήστη, η οποία αποτελεί σημαντικό αντικίνητρο για τη χρήση ενός τέτοιου συστήματος.

Οι δύο προσεγγίσεις αυτόματου φιλτραρίσματος προσπαθούν να αντιμετωπίσουν αυτά τα προβλήματα μέσω της χρήσης τεχνικών μηχανικής μάθησης, οι οποίες βοηθούν στην ανάλυση των δεδομένων και στην κατασκευή των απαιτούμενων μοντέλων χρήστη. Η βασική τους διαφορά βρίσκεται στο τι στοιχεία δίνει έμφαση η καθεμιά από αυτές για να κάνει το φιλτράρισμα. Τα content-based filtering συστήματα χρησιμοποιούν τεχνικές εκμάθησης πάνω στο περιεχόμενο των ιστοσελίδων, στις οποίες δίνεται έμφαση στο τι ενδιαφέρει το χρήστη. Αντιθέτως, τα collaborative filtering συστήματα βασίζονται στις ομοιότητες μεταξύ των χρηστών, δίνοντας έμφαση στο ποιος άλλος ενδιαφέρεται για τα ίδια πράγματα με τον εκάστοτε χρήστη.

Το κύριο πρόβλημα στα content-based filtering συστήματα είναι η δυσκολία που υπάρχει στην ανάλυση του περιεχομένου των ιστοσελίδων και στην μετατροπή του σε σημασιολογικές οντότητες. Ακόμη και αν κάποιος αγνοεί το πολυμεσικό περιεχόμενο, η ίδια η φυσική γλώσσα είναι μια πλούσια και αδόμητη πηγή δεδομένων. Παρόλη τη σημαντική εξέλιξη που επιτεύχθηκε σε τομείς της έρευνας που ασχολούνται με την ανάλυση δεδομένων κειμένου (textual data), απέχει ακόμη πολύ μακριά η κατασκευή μιας μηχανής που θα καταλαβαίνει τη φυσική γλώσσα με τον ίδιο τρόπο που το κάνουν οι άνθρωποι. Επίσης, η συγκεκριμένη προσέγγιση υιοθετεί μια ποικιλία στατιστικών μεθόδων για την εξαγωγή χρήσιμων πληροφοριών από τα δεδομένα που είναι υπό μορφή κειμένου (*TF-IDF vector representation* και *Latent Semantic Indexing*). Ωστόσο, το πρόβλημα της ανάλυσης του περιεχομένου στο Διαδίκτυο παραμένει και γίνεται ακόμη πιο κρίσιμο όταν υπάρχει περιορισμένο περιεχόμενο. Με τη μείωση της έμφασης πάνω στο περιεχόμενο, τα collaborative filtering συστήματα, προσπαθούν να επιλύσουν αυτό το σημαντικό πρόβλημα. Επιπλέον, αυτά τα συστήματα διευκολύνουν την εκμετάλλευση των πρότυπων χρήσης (usage patterns) που δεν περιορίζονται εντός αυστηρών σημασιολογικών ορίων.

Ωστόσο, τα collaborative filtering συστήματα παρουσιάζουν και αυτά αρκετά προβλήματα. Χρήστες που αξιολογούν πρώτοι νέα στοιχεία δεν μπορούν να τους δοθούν επιπλέον συστάσεις. Επίσης, η ποιότητα της σύστασης εξαρτάται από τον αριθμό των αξιολογήσεων που ένας συγκεκριμένος χρήστης έχει κάνει, οδηγώντας με αυτό τον τρόπο σε χαμηλής ποιότητας συστάσεις για τους χρήστες που έχουν αξιολογήσει ένα μικρό αριθμό αντικειμένων.

Τέλος, οι collaborative filtering μέθοδοι που χρησιμοποιούν αποκλειστικά memory-based προσεγγίσεις εκμάθησης, πάσχουν από δύο επιπλέον προβλήματα. Δεν κλιμακώνονται σε μεγάλο αριθμό χρηστών και δεν παρέχουν καμία πληροφορία σχετικά με τα μοντέλα χρήσης που εξάγονται από τα δεδομένα. Πρόσφατα, αυτά τα προβλήματα έχουν αρχίσει να αντιμετωπίζονται, με την ανάπτυξη model-based collaborative μεθόδων [23].

3.3.3 Ο ρόλος του *Web Usage Mining*

Κατά τη διάρκεια των τελευταίων ετών, οι ερευνητές έχουν προτείνει μία νέα κοινή ονομασία για όλες τις μεθόδους εξόρυξης δεδομένων στο Διαδίκτυο, που ονομάζεται *Web Mining*. Τα *Web Mining* εργαλεία έχουν σα στόχο την εξαγωγή γνώσης από το Διαδίκτυο και όχι την ανάκτηση πληροφοριών. Αυτά τα εργαλεία κατατάσσονται στις ακόλουθες τρεις κατηγορίες.

- *Web Content Mining*
- *Web Usage Mining*
- *Web Structure Mining*

Το *Web Content Mining* ασχολείται με την εξαγωγή χρήσιμων γνώσεων από το περιεχόμενο των ιστοσελίδων, χρησιμοποιώντας μηχανισμούς εξόρυξης δεδομένων. Το *Web Usage Mining*, στοχεύει στην ανακάλυψη σημαντικών προτύπων χρήσης, αναλύοντας τα δικτυακά δεδομένα χρήσης. Τέλος, το *Web Structure Mining* είναι ένας νέος τομέας, που ασχολείται με την εφαρμογή μεθόδων εξόρυξης δεδομένων πάνω στη δομή του γράφου του Διαδικτύου [23], [25], [26].

Το *Web Mining* θεωρείται μια ολοκληρωμένη διαδικασία παρά ένας απλός αλγόριθμος. Στην περίπτωση του *Web Usage Mining*, ολόκληρη η διαδικασία αυτή στηρίζεται στην ανακάλυψη της γνώσης που αφορά στη συμπεριφορά των χρηστών. Αρχικά, ο στόχος του *Web Usage Mining* ήταν να υποστηρίξει την διαδικασία λήψης αποφάσεων από τον άνθρωπο και ως εκ τούτου, το αποτέλεσμα της διαδικασίας αυτής είναι συνήθως μια σειρά μοντέλων δεδομένων που αποκαλύπτουν «απόλυτη γνώση» σχετικά με αντικείμενα δεδομένων, όπως ιστοσελίδες ή προϊόντα που διατίθενται σε μια συγκεκριμένη τοποθεσία στο Διαδίκτυο. Αυτά τα μοντέλα αξιολογούνται και αξιοποιούνται από εμπειρογνώμονες, όπως ο αναλυτής της αγοράς ο οποίος αναζητά επιχειρηματικές πληροφορίες (*business intelligence*), ή ο διαχειριστής που θέλει να βελτιστοποιήσει τη δομή της εφαρμογής του και να ενισχύσει την εμπειρία περιήγησης των επισκεπτών του.

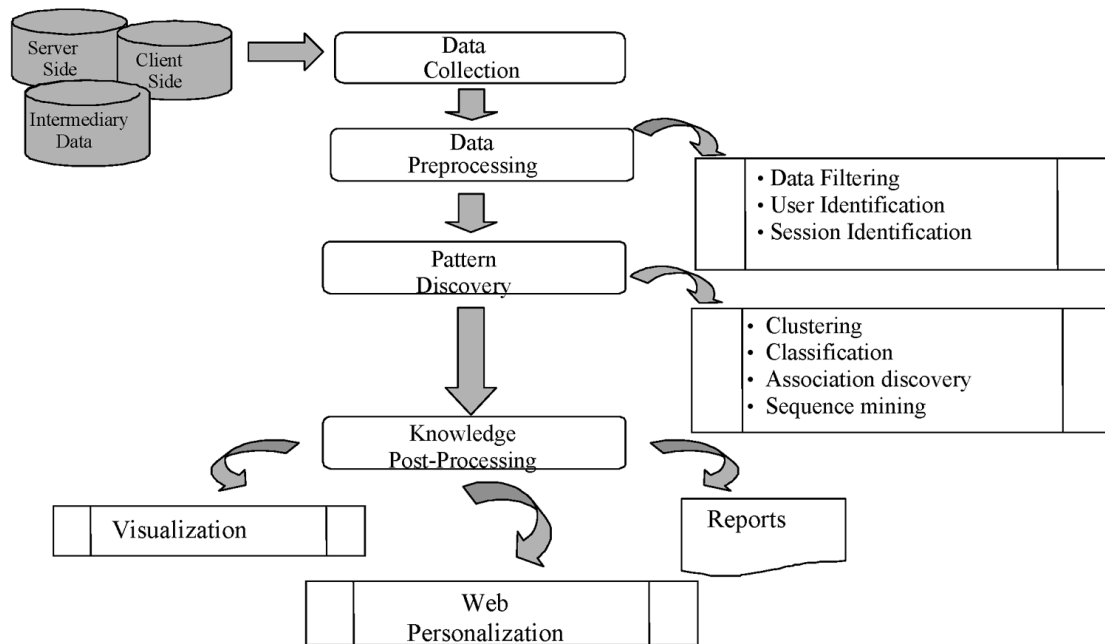
Παρά το γεγονός ότι το μεγαλύτερο μέρος των εργασιών που έχουν γίνει στον τομέα του *Web Usage Mining* δεν ασχολείται με το *personalization*, η σχέση του με αυτοματοποιημένα εργαλεία προσωποποίησης είναι ξεκάθαρη και σαφής. Οι αναλύσεις και οι έρευνες που γίνονται πάνω στο *Web Usage Mining* μπορεί να αποτελέσουν πηγή ιδεών και λύσεων για την πραγματοποίηση και την περάτωση του *Web Personalization*. Δεδομένα χρήσης, όπως για παράδειγμα αυτά που μπορούν να συλλεχθούν όταν ο χρήστης περιηγείται σε μια συγκεκριμένη τοποθεσία στο Διαδίκτυο, αντιπροσωπεύουν την αλληλεπίδραση μεταξύ του χρήστη και της συγκεκριμένης τοποθεσίας. Το *Web Usage Mining* παρέχει μια προσέγγιση για τη συλλογή και προ-επεξεργασία των δεδομένων αυτών και κατασκευάζει τα μοντέλα που αντιπροσωπεύουν τη συμπεριφορά και τα ενδιαφέροντα των χρηστών.

Τα μοντέλα αυτά μπορούν να χρησιμοποιηθούν από ένα προσωποποιημένο σύστημα αυτόματα, δίχως την παρέμβαση κάποιου ανθρώπινου παράγοντα, όπως ενός εμπειρογνώμονα, για την πραγματοποίηση των απαιτούμενων προσωποποιημένων λειτουργιών. Αυτό το είδος της γνώσης, όπως είναι τα μοντέλα χρήστη, αποτελεί «λειτουργική γνώση» για το *Web Personalization*. Ως εκ τούτου, ένα *Web Personalization* σύστημα μπορεί να χρησιμοποιήσει *Web Usage Mining* μεθόδους, προκειμένου να επιτευχθεί η απαιτούμενη ευρωστία και ευελιξία που χρειάζεται σε ένα τέτοιο σύστημα.

Η στενή σχέση μεταξύ των *Web Usage Mining* μεθόδων και του *Web Personalization* αποτελεί κύρια μελέτη πολλών ερευνών. Έχοντας υπόψη τη χρήση του στο *Web Personalization*, και όντας ουσιαστικά μια διαδικασία εξόρυξης δεδομένων, το *Web Usage Mining* αποτελείται από τα εξής βασικά στάδια.

- **Συλλογή δεδομένων:** Σε αυτό το στάδιο, συγκεντρώνονται τα δεδομένα χρήσης από διάφορες πηγές και αναγνωρίζεται το περιεχόμενο και η δομή τους. Τα στοιχεία αυτά συλλέγονται από *Web* διακομιστές, από τους πελάτες που συνδέονται σε ένα διακομιστή, ή από ενδιάμεσες πηγές όπως διακομιστές μεσολάβησης (*proxy servers*) και *packet sniffers*. Ορισμένες τεχνικές που έχουν εφαρμοστεί σε αυτό το στάδιο, μπορούν να χρησιμοποιηθούν για την αποτελεσματική συλλογή δεδομένων χρήστη για τους σκοπούς της προσωποποίησης.
- **Προ-επεξεργασία δεδομένων:** Στο στάδιο αυτό, τα δεδομένα έχουν καθαριστεί, έχουν επιλυθεί τυχόν ασυνέπειες σε αυτά και είναι πλέον ολοκληρωμένα και ενοποιημένα, έτσι ώστε να χρησιμοποιηθούν στην επόμενη φάση της εύρεσης προτύπων. Στο *Web Usage Mining*, αυτή η διαδικασία συνιστά κυρίως φιλτράρισμα δεδομένων, ταυτοποίηση του χρήστη καθώς και αναγνώριση της συνεδρίας του εκάστοτε χρήστη (*user session identification*). Οι τεχνικές που χρησιμοποιούνται σε αυτό το στάδιο μπορούν να παράσχουν αποτελεσματικά δεδομένα για επεξεργασία (*data elaboration*).
- **Εύρεση Προτύπων:** Σε αυτό το στάδιο, γίνεται εξόρυξη γνώσης μέσω μηχανικής μάθησης και στατιστικών τεχνικών, όπως είναι η ομαδοποίηση, η ταξινόμηση, η «εύρεση συνδέσεων» (*association discovery*) καθώς και η διαδοχική εύρεση προτύπων για τα δεδομένα. Τα πρότυπα που απαιτούνται για την ορθή λειτουργία του *Web Personalization*, αντιστοιχούν στην συμπεριφορά και στα ενδιαφέροντα των χρηστών. Αυτό είναι το στάδιο όπου εφαρμόζονται μέθοδοι εκμάθησης για την αυτοματοποιημένη κατασκευή μοντέλων χρηστών.
- **Μετά-επεξεργασία της γνώσης (*knowledge post-processing*):** Σε αυτό το τελευταίο στάδιο, η εξαγόμενη γνώση αξιολογείται και παρουσιάζεται συνήθως σε μορφή που να είναι κατανοητή από τον άνθρωπο, χρησιμοποιώντας εκθέσεις (*reports*) ή τεχνικές απεικόνισης. Για το *Web Personalization*, η γνώση που εξάγεται, ενσωματώνεται σε ένα *Personalization module* προκειμένου να διευκολύνει τις προσωποποιημένες λειτουργίες.

Ολοκληρώνοντας, στην Εικόνα 15 συνοψίζονται σχηματικά τα προαναφερθέντα στάδια. Καθένα από αυτά παρουσιάζει διαφορετικές δυσκολίες και έχουν αναλυθεί σε διάφορες μελέτες. Θα πρέπει να σημειωθεί ότι το *Web Usage Mining* είναι λιγότερο «ώριμο» από άλλες μεθόδους εξόρυξης δεδομένων. Ως αποτέλεσμα, μερικά από αυτά τα θέματα έχουν μελετηθεί σε έρευνες εξόρυξης δεδομένων και θεωρούνται ως ξεχωριστά στάδια της διαδικασίας αυτής. Παρόλα αυτά, προβλήματα όπως η αξιολόγηση των γνώσεων που εξάγονται, χρειάζονται περαιτέρω μελλοντική ανάλυση στον τομέα του *Web Usage Mining* [23].



Εικόνα 15: Web Usage Mining διαδικασία

4

Μεθοδολογία

4.1 Δημιουργία και αξιολόγηση μοντέλων κοινοτήτων

4.1.1 Εισαγωγή

Η κατασκευή μοντέλων κοινοτήτων στο Διαδίκτυο με τη χρήση μεθόδων μηχανικής μάθησης είναι κατά πρωτίστως μια **Web Usage Mining** διαδικασία. Όπως έχει προαναφερθεί, οι κοινότητες κατασκευάζονται με βάση τα δεδομένα που συλλέγονται από τους χρήστες κατά την αλληλεπίδρασή τους με το σύστημα. Για την ορθή περάτωση της συγκεκριμένης διαδικασίας, κρίνεται απαραίτητος ο προσδιορισμός ενδιαφερόντων προτύπων συμπεριφοράς μέσα από τα δεδομένα χρήσης που έχουν συλλεχθεί καθώς και η δημιουργία κοινοτήτων βάσει αυτών των προτύπων. Στο επίπεδο ανακάλυψης ενδιαφερόντων προτύπων, οι Web Usage Mining τεχνικές παρέχουν περισσότερες λεπτομερείς και χρήσιμες πληροφορίες σχετικά με τη χρήση της υπηρεσίας από ότι κάποια ευρέως χρησιμοποιούμενα στατιστικά στοιχεία [38], [39].

Επομένως, κρίνεται απαραίτητη η χρησιμοποίηση Web Usage Mining τεχνικών για την δημιουργία μοντέλων κοινοτήτων που θα αντικατοπτρίζουν τις ανάγκες και τα ενδιαφέροντα του εκάστοτε χρήστη σε μια εφαρμογή στο Διαδίκτυο. Παρακάτω αναλύονται τα προβλήματα και η μεθοδολογία που εν τέλει χρησιμοποιήθηκε στα 4 βασικά στάδια του Web Usage Mining.

4.1.2 Συλλογή και προ-επεξεργασία των δεδομένων

Οι αλγόριθμοι μάθησης συνήθως απαιτούν τα δεδομένα που έχουν συλλεχθεί, να μετατραπούν σε σύνολα αντικειμένων, όπου κάθε αντικείμενο να περιγράφεται από ένα σύνολο χαρακτηριστικών. Τα δεδομένα χρήσης στο Διαδίκτυο εμφανίζονται σε διάφορες μορφές, ανάλογα με τον τύπο πρόσβασης που η εκάστοτε υπηρεσία πληροφοριών παρέχει. Λόγω αυτής της ποικιλίας, η μετατροπή των δεδομένων σε ένα σύνολο αντικειμένων,

διατηρώντας τις πληροφορίες που είναι χρήσιμες για τις διαδικασίες μοντελοποίησης, είναι μια σημαντική μηχανική διαδικασία.

Ένα σημαντικό θέμα που προκύπτει στο στάδιο της προ-επεξεργασίας των δεδομένων είναι η δυνατότητα προσδιορισμού μεμονωμένων χρηστών. Σε υπηρεσίες που διαθέτουν εγγεγραμμένους χρήστες, αυτό επιτυγχάνεται εύκολα, καθώς κάθε αλληλεπίδραση του χρήστη με το σύστημα συνδέεται μονοσήμαντα με ένα συγκεκριμένο κωδικό (user code). Ωστόσο, οι περισσότερες υπηρεσίες στο Διαδίκτυο είναι προσβάσιμες από όλους και συνήθως δεν υπάρχει μηχανισμός καταγραφής της ταυτότητας του χρήστη. Μια από αυτές είναι και το ελεύθερο σύστημα κοινωνικής επισημείωσης (<http://library.panteion.gr/opacial>) της ηλεκτρονικής βιβλιοθήκης του Πάντειου Πανεπιστημίου, του οποίου τα δεδομένα χρήσης θα χρησιμοποιηθούν στις περιπτώσεις μελέτης που θα αναλυθούν παρακάτω. Σε αυτές τις περιπτώσεις, δεν υπάρχουν πρότυπα για τους μεμονωμένους χρήστες και τα μοντέλα κοινοτήτων κατασκευάζονται άμεσα από τα χαμηλά επίπεδα δεδομένων χρήσης. Δεδομένου ότι τα δεδομένα δεν μπορούν να κατηγοριοποιηθούν βάσει των κωδικών χρήστη, άλλες «μονάδες αλληλεπίδρασης» πρέπει να οριστούν, όπως οι συνεδρίες πρόσβασης (access sessions), οι οποίες θα αποτελέσουν τη βάση για την ανακάλυψη ομοιοτήτων στις συμπεριφορές των χρηστών.

Επίσης, ένα άλλο σημαντικό θέμα που προκύπτει σε αυτό το στάδιο είναι η επιλογή του συνόλου των χαρακτηριστικών που θα περιγράψουν τα αντικείμενα. Ουσιαστικά, κάθε αντικείμενο είναι ένας πίνακας χαρακτηριστικών με συγκεκριμένο μέγεθος. Λόγω αυτής της ιδιαιτερότητας του πίνακα, πρέπει να επιλεγεί ένα ορισμένο σύνολο χαρακτηριστικών ανάλογα με την μελέτη περίπτωσης που εξετάζεται. Συγκεκριμένα, στις μελέτες περίπτωσης που θα εξεταστούν παρακάτω τα αντικείμενα που δημιουργούνται αντιστοιχούν σε συνεδρίες πρόσβασης, όπου κάθε χαρακτηριστικό του αντικειμένου αντιστοιχεί σε μια συγκεκριμένη σελίδα της εφαρμογής μέσα στην συνεδρία. Αυτή η προσέγγιση είναι γνωστή και ως *bag-of-pages approach* [40].

Όπως ήδη αναφέρθηκε, οι δυο μελέτες περίπτωσης που εξετάζονται, αφορούν στην υλοποίηση αλγορίθμων εξόρυξης γνώσης από αρχεία καταγραφής δικτυακών τόπων που παρέχουν υπηρεσίες κοινωνικών επισημειώσεων, για το σχηματισμό κοινοτήτων χρηστών. Για τα συγκεκριμένα πειράματα χρησιμοποιήθηκαν log αρχεία πρόσβασης από το ελεύθερο σύστημα κοινωνικής επισημείωσης (<http://library.panteion.gr/opacial>) της ηλεκτρονικής βιβλιοθήκης του Πάντειου Πανεπιστημίου. Αρχικά, τα αρχεία αυτά περιείχαν πάνω από 4 εκατομμύρια εγγραφές και κάλυπταν μια χρονική περίοδο 7 μηνών (Ιούλιος 2009 – Ιανουάριος 2010). Κάθε εγγραφή περιέχει την ημερομηνία και την ώρα πρόσβασης του χρήστη στην εφαρμογή, την IP διεύθυνση του, το όνομα του domain και το url της σελίδας που χρησιμοποίησε. Όπως ήταν φυσικό, υπήρχε μεγάλος όγκος δεδομένων που ήταν άχρηστος για τα πειράματα αυτά και έπρεπε να εφαρμοστούν κάποιοι μέθοδοι φιλτραρίσματος για την μείωση της πλεονάζουσας και κατά συνέπεια μη αξιοποιήσιμης πληροφορίας. Στη συνέχεια παρατίθενται συνοπτικά οι μέθοδοι προσέγγισης.

- Αφαίρεση πολυμεσικού υλικού, όπως εικόνες (.jpg αρχεία) και βίντεο (.avi αρχεία).
- Αφαίρεση script αρχείων όπως Javascript αρχεία και κλάσεις Java.
- Ομαδοποίηση και αφαίρεση συγκεκριμένων *HTTP error* κωδικών που προκύπτουν από εσφαλμένες κλήσεις από μεριάς του χρήστη ή από μη εξουσιοδοτημένη πρόσβαση σε σελίδα της εφαρμογής.
- Συλλογή και αφαίρεση κακόβουλων IP που δεν αντιστοιχούν σε χρήστες αλλά σε *bots* με σκοπό την πλήρη ευρετηρίαση της εφαρμογής.
- Συλλογή των εγγραφών που έχουν σαν κριτήριο αναζήτησης τις κοινωνικές επισημειώσεις τους συστήματος.

Ως αποτέλεσμα, το νέο log αρχείο περιείχε μόνο 4.130 εγγραφές (από ένα αρχικό σύνολο ενός 1.000.000 εγγραφών περίπου) και ήταν έτοιμο για το επόμενο στάδιο της Web Usage Mining διαδικασίας, όπου χρησιμοποιήθηκαν δυο post-processing τεχνικές, η εξαγωγή συνεδριών και η μετατροπή των μονοπατιών που καταγράφηκαν σε πίνακες χαρακτηριστικών (feature vectors).

Η διαδικασία εξαγωγής συνεδριών θεωρείται αρκετά πολύπλοκη και πολλές φορές η αβεβαιότητα παίζει σημαντικό ρόλο. Στη συνέχεια ακολουθούν τα στάδια της συγκεκριμένης διαδικασίας.

- Ομαδοποίηση των κλήσεων ανά ημέρα και IP διεύθυνση
- Επιλογή χρονικού πλαισίου μέσα στην οποία 2 κλήσεις από την ίδια IP διεύθυνση μπορεί να θεωρηθούν ότι ανήκουν στην ίδια συνεδρία
- Ομαδοποίηση των σελίδων που χρησιμοποιούνται από την ίδια IP διεύθυνση μέσα στο χρονικό πλαίσιο που έχει επιλεγεί για την δημιουργία συνεδρίας.

Συνήθως, η επιλογή του χρονικού πλαισίου γίνεται μέσω απεικόνισης της κατανομής της συχνότητας μετάβασης σελίδων (page transitions) σε λεπτά. Βάσει αυτής της κατανομής, παρατηρήθηκε ότι οι μεταβάσεις από μια σελίδα σε μια άλλη, μέσα στο χρονικό περιθώριο των 6 λεπτών, είχε σχεδόν μηδενική συχνότητα. Οπότε στα πλαίσια αυτής της διπλωματικής εργασίας, μια συνεδρία πρόσβασης (access session) ορίζεται ως η ακολουθία μεταβάσεων σελίδων από την ίδια IP διεύθυνση, όπου κάθε μετάβαση γίνεται μέσα σε ένα χρονικό πλαίσιο κάτω των 5 λεπτών. Βάσει αυτού του ορισμού, δημιουργήθηκαν 519 συνεδρίες που περιείχαν 3000 επισημειώσεις.

Τέλος, για την μετατροπή των συνεδριών που δημιουργήθηκαν σε πίνακες χαρακτηριστικών, χρησιμοποιήθηκαν δυο διαφορετικές προσεγγίσεις. Οι συνεδρίες αυτές αποτελούνται από επισημειώσεις που επιλέχθηκαν από τους χρήστες σαν κριτήριο αναζήτησης πόρων (βιβλία, έγγραφα, δοκίμια) από την ηλεκτρονική βιβλιοθήκη του Πάντειου Πανεπιστημίου. Στη πρώτη μελέτη περίπτωσης κάθε χαρακτηριστικό αναπαριστά την κάθε μοναδική επισημείωση που υπάρχει σε τουλάχιστον μια συνεδρία, δημιουργώντας με αυτό τον τρόπο 425 μοναδιαία χαρακτηριστικά. Αντιθέτως, στη δεύτερη μελέτη περίπτωσης κάθε χαρακτηριστικό αναπαριστά τις συνεδρίες στις οποίες περιλαμβάνεται κάθε μοναδιαία επισημείωση, δημιουργώντας με αυτό τον τρόπο 519 μοναδιαία χαρακτηριστικά.

4.1.3 Εύρεση προτύπου

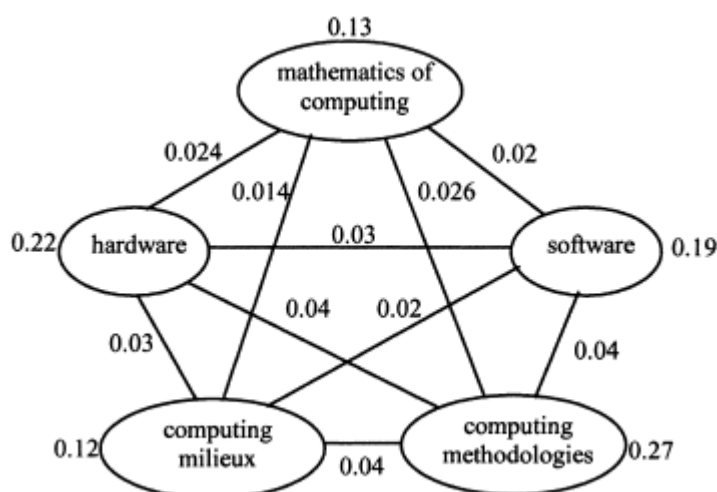
Σε αυτό το στάδιο αξιολογείται η cluster mining διαδικασία, η οποία επεξεργάζεται τα δεδομένα που προέκυψαν από το προηγούμενο στάδιο και προσδιορίζει ενδιαφέροντα πρότυπα που αντιστοιχούν σε διαφορετικές ομάδες χρηστών. Παρακάτω παρουσιάζεται αναλυτικά η εν λόγω διαδικασία.

4.1.3.1 Ο αλγόριθμος Cluster Mining

Η Cluster Mining μέθοδος ανακαλύπτει πρότυπα κοινής συμπεριφοράς, εξετάζοντας τις κλίκες σε ένα γράφημα που αντιστοιχεί στα χαρακτηριστικά των χρηστών. Ως κλίκα ορίζεται ένας πλήρης συνδεδεμένος υπογράφος, δηλαδή ένα σύνολο κορυφών, που

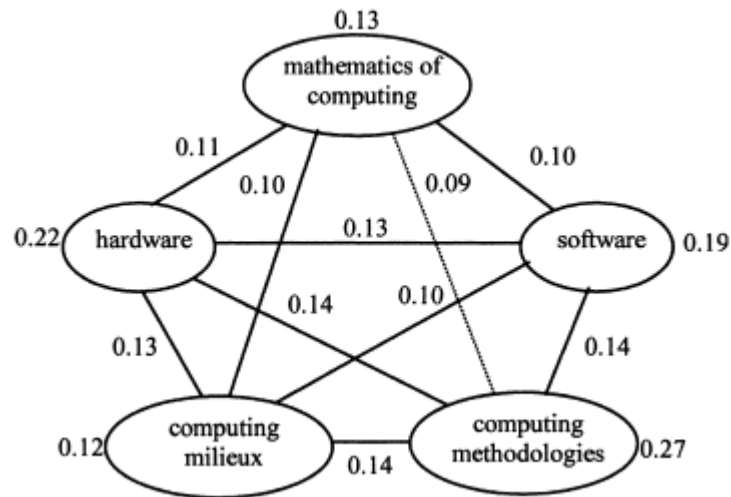
συνδέονται όλες μεταξύ τους με ακμές. Ο αλγόριθμος ξεκινά με την κατασκευή ενός γραφήματος του οποίου οι κορυφές αντιστοιχούν στα χαρακτηριστικά των χρηστών, ενώ οι ακμές του αντιστοιχούν σε συνδυασμούς χαρακτηριστικών όπως αυτά παρατηρούνται κατά την αλληλεπίδραση των χρηστών με το σύστημα [41].

Για παράδειγμα, εάν η υπηρεσία που χρησιμοποιείται είναι μια ηλεκτρονική βιβλιοθήκη άρθρων πάνω στο τομέα της πληροφορικής και οι χρήστες χρησιμοποιούν τις επισημειώσεις «hardware» και «computing methodologies» για να κάνουν επερωτήσεις στο σύστημα, δημιουργείται μια ακμή μεταξύ των δυο αυτών σχετικών κορυφών, όπως φαίνεται και στην Εικόνα 16. Οι κορυφές και οι ακμές του γραφήματος συσχετίζονται με βάρη, τα οποία υπολογίζονται βάσει των συχνοτήτων των επιλογών των χρηστών και των συνδυασμών τους, αντίστοιχα. Όπως φαίνεται και στην Εικόνα 16, η επισημείωση «computing methodologies» φαίνεται να είναι η πιο δημοφιλής μεταξύ των προτιμήσεων των χρηστών της υπηρεσίας και συνδυάζεται συχνά με επισημειώσεις λογισμικού όπως «software», «hardware», «computing milieu» και λιγότερο με την επισημείωση «mathematics of computing».



Εικόνα 16: Μη κανονικοποιημένος γράφος

Μία τροποποίηση που έχει γίνει στην βασική μέθοδο δημιουργίας του γράφου, είναι η κανονικοποίηση των συχνοτήτων των άκρων μέσω της διαίρεσης τους με το μέγιστο των συχνοτήτων των δύο κορυφών του συνδέονται. Αυτό έχει σαν αποτέλεσμα, την αφαίρεση των χαρακτηριστικών που εμφανίζονται πολύ συχνά σε πολλούς χρήστες. Στην Εικόνα 17 παρουσιάζεται το κανονικοποιημένο γράφημα της Εικόνα 16.



Εικόνα 17: Κανονικοποιημένος γράφος

Όπως παρατηρείται, η συνδεσιμότητα του γραφήματος που προκύπτει είναι συνήθως πολύ υψηλή. Για το λόγο αυτό, απαιτείται η χρήση ενός ορίου (threshold) με σκοπό να μειωθεί ο αριθμός των άκρων. Στο συγκεκριμένο παράδειγμα, αν το κατώτατο όριο ισούται με 0,1 τότε η ακμή μεταξύ «mathematics of computing» και «computing methodologies» αφαιρείται. Η κατάλληλη τιμή του ορίου διαφέρει από εφαρμογή σε εφαρμογή και δεν μπορεί να θεωρηθεί ως μια σταθερή αξία. Στις μελέτες περίπτωσης που θα ακολουθήσουν, θα χρησιμοποιηθούν 3 διαφορετικές τιμές ορίου.

Τέλος, ένα άλλο σημαντικό στοιχείο που προκύπτει από τον κανονικοποιημένο γράφο είναι ότι δεν περιορίζεται το μέγεθος των κλικών, αλλά δημιουργούνται όλες οι μέγιστες κλίκες του γραφήματος (κλίκες που δεν είναι υπογράφοι άλλων κλικών). Παρά τη μεγάλη θεωρητική πολυπλοκότητα που δημιουργεί το πρόβλημα εύρεσης κλικών, στην πράξη, ο αλγόριθμος που έχει εφαρμοστεί είναι αρκετά γρήγορος και η απόδοση του επέτρεψε την περαιτέρω έρευνα πάνω σε θέματα σχετικά με την επίδραση που έχει το όριο διασύνδεσης (connectivity threshold) στον αλγόριθμο [42].

4.1.4 Μετα-επεξεργασία προτύπου και αξιολόγηση

4.1.4.1 Κατασκευή μοντέλων κοινοτήτων

Το τελικό στάδιο της διαδικασίας αυτής αφορά στη δημιουργία ενός συνόλου μοντέλων, το οποίο αντιστοιχεί σε κοινά πρότυπα συμπεριφοράς για διαφορετικούς τύπους χρηστών. Η ομαδοποίηση των χρηστών σε κοινότητες με κοινά χαρακτηριστικά συμπεριφοράς είναι το πρώτο βήμα προς αυτή την κατεύθυνση, αλλά δεν παρέχει τα επιθυμητά μοντέλα. Οπότε, ο προσδιορισμός των περιγραφικών χαρακτηριστικών της κάθε κοινότητας αποτελεί βασική προϋπόθεση για την εξαγωγή των μοντέλων αυτών. Ο τρόπος με τον οποίο αυτό επιτυγχάνεται διαφέρει ανάλογα με τη μέθοδο ομαδοποίησης που χρησιμοποιείται, αλλά οι βασικές ιδέες είναι κοινές και παρατίθενται παρακάτω.

Ένα μοντέλο κοινότητας μπορεί να εκφραστεί με όρους των ίδιων χαρακτηριστικών που υπάρχουν και στα δεδομένα χρήσης. Για παράδειγμα, εάν τα

δεδομένα χρήσης καταγράφουν απλώς τις σελίδες ενός ιστότοπου που επισκέπτεται ο χρήστης, τότε οι κοινότητες θα πρέπει επίσης να περιγράφονται με βάση τις σελίδες του ιστότοπου αυτού. Με έναν πιο επιστημονικό ορισμό, δεδομένου ότι υπάρχει ένα σύνολο χαρακτηριστικών A που περιγράφει τα αντικείμενα μέσα σε ένα σύνολο δεδομένων, το μοντέλο μιας κοινότητας C_j αποτελείται από ένα υποσύνολο του A , το A_j , που χαρακτηρίζει τα μέλη της κοινότητας και βρίσκεται στα μοντέλα των μελών αυτών.

Η επιλογή των περιγραφικών χαρακτηριστικών γίνεται με τη βοήθεια απλών μετρικών, που βασίζονται στην ιδέα ότι ένα χαρακτηριστικό περιγράφει μια κοινότητα, αν η συχνότητα του εντός της κοινότητας είναι σημαντικά υψηλότερη από τη συχνότητα του σε όλο το σύνολο δεδομένων. Η φυσική επιλογή των μετρικών διαφέρει ανάλογα με την cluster μέθοδο που θα χρησιμοποιηθεί. Συγκεκριμένα, οι ομάδες (clusters) που δημιουργούνται από την cluster mining μέθοδο συσχετίζουν τα χαρακτηριστικά των χρηστών άμεσα. Οπότε, δεν υπάρχει ανάγκη για επιπρόσθετο στάδιο χαρακτηρισμού κοινότητας, καθώς κάθε κλίκα που προσδιορίζεται από τον cluster mining αλγόριθμο είναι ένα πρότυπο συμπεριφοράς. Τέλος, ο χρήστης μπορεί να συσχετιστεί με κλίκες που ταιριάζουν καλύτερα στο ατομικό του μοντέλο. Αυτή η διαδικασία δεν επιχειρείται στα συγκεκριμένα πειράματα, καθώς το επίκεντρο της μελέτης είναι η ανακάλυψη των προτύπων συμπεριφοράς και όχι των ομάδων χρηστών αυτών καθαυτών.

4.1.4.2 Κριτήρια αξιολόγησης

Έχοντας περιγράψει την μέθοδο κατασκευής μοντέλων κοινοτήτων, θα πρέπει να αποφασιστούν ποιες θα είναι οι επιθυμητές ιδιότητες των μοντέλων αυτών. Πρωταρχικός στόχος της διαδικασίας αυτής, είναι η παροχή χρήσιμων μοντέλων κοινοτήτων. Προκειμένου τα μοντέλα αυτά να είναι χρήσιμα για τον άνθρωπο (π.χ. χρήστες ή πάροχοι υπηρεσιών), θα πρέπει να είναι σχετικά λίγα σε αριθμό και μικρά σε μέγεθος. Ως αποτέλεσμα, ο αριθμός των μοντέλων και το μέσο μέγεθος τους, αποτελούν δύο σημαντικά μετρήσιμα κριτήρια για την επιτυχία αυτής της μεθόδου. Η εύρεση ακριβών στοιχείων για τα κριτήρια αυτά εξαρτάται από τη φύση του προβλήματος, όπως είναι το μέγεθος του συνόλου των χαρακτηριστικών.

Ωστόσο, ένα «εύπεπτο» σύνολο προτύπων δεν είναι κατ' ανάγκη ενδιαφέρον. Όταν υπάρχουν μόνο μικρές διαφορές μεταξύ των μοντέλων, αντιπροσωπεύοντας εκδοχές της ίδιας κοινότητας, η κατάτμηση των χρηστών σε κοινότητες δεν θεωρείται ενδιαφέρουσα. Οπότε, αναζητούνται μοντέλα κοινοτήτων που είναι όσο το δυνατόν διαφορετικά μεταξύ τους, δηλαδή να μοιράζονται όσον το δυνατόν λιγότερα χαρακτηριστικά.

Το πρώτο μετρήσιμο κριτήριο αξιολόγησης αφορά την ιδιαιτερότητα ενός συνόλου κοινοτήτων και χρησιμοποιείται ευρέως με τον όρο *distinctiveness*. Η ιδιαιτερότητα ενός συνόλου κοινοτήτων M υπολογίζεται από την αναλογία μεταξύ του αριθμού των ξεχωριστών χαρακτηριστικών που καλύπτονται μέσα από αυτά τα μοντέλα και του μεγέθους του συνόλου των κοινοτήτων M . Με άλλα λόγια, υπολογίζεται ο αριθμός των ξεχωριστών χαρακτηριστικών που εμφανίζονται σε ένα τουλάχιστον μοντέλο κοινότητας και διαιρείται με το άθροισμα των μεγεθών όλων των μοντέλων συνολικά, όπου το μέγεθος του κάθε μοντέλου αντιστοιχεί στον αριθμό των χαρακτηριστικών που περιέχει. Με πιο επιστημονικούς όρους, αν υπάρχουν J κοινότητες μέσα σε ένα σύνολο M και A_j τα χαρακτηριστικά που χρησιμοποιούνται σε κάθε j μοντέλο τότε ο όρος *distinctiveness* υπολογίζεται από την παρακάτω εξίσωση [40], [43].

$$\text{Distinctiveness}(M) = \frac{\left| \bigcup_j A_j \right|}{\sum_j |A_j|}$$

Στην καλύτερη περίπτωση, όπου τα μοντέλα κοινοτήτων δεν έχουν κάποιο κοινό χαρακτηριστικό, ο όρος *distinctiveness* παίρνει την τιμή 1, καθώς ο αριθμητής γίνεται ίδιος με το παρονομαστή. Αντιθέτως, στη χειρότερη περίπτωση όπου όλα τα μοντέλα είναι πανομοιότυπα, ο όρος *distinctiveness* γίνεται $1/|J|$, όπου J είναι το σύνολο των μοντέλων κοινότητας που έχουν ανακαλυφθεί.

Σε ένα πιο ρεαλιστικό παράδειγμα, υπάρχει δυνατότητα να δημιουργηθεί ένα μικρό σύνολο διακριτών μοντέλων, τα οποία αντιπροσωπεύουν μόνο ένα μικρό μέρος της χρήσης του συστήματος. Προκειμένου να αποφευχθεί αυτό το πρόβλημα, πολλές έρευνες εισήγαγαν ένα επιπλέον κριτήριο αξιολόγησης, το οποίο αντισταθμίζει το *distinctiveness*. Το νέο κριτήριο αφορά τη συνολική κάλυψη των μοντέλων κοινοτήτων, το ποσοστό δηλαδή των χαρακτηριστικών που καλύπτονται από τα μοντέλα. Χρησιμοποιώντας τις ίδιες μεταβλητές, η κάλυψη του συνόλου των μοντέλων M υπολογίζεται από την παρακάτω εξίσωση [40], [43].

$$\text{Coverage}(M) = \frac{\left| \bigcup_j A_j \right|}{|A|}$$

Όπως παρατηρείται, η διαφορά των δυο αυτών εξισώσεων βρίσκεται στον παρονομαστή, ο οποίος στην δεύτερη εξίσωση υπολογίζει τον συνολικό αριθμό των χαρακτηριστικών που χρησιμοποιούνται για την περιγραφή των αντικειμένων. Όταν αριθμητής γίνεται ίσος με αυτόν τον αριθμό τότε ο όρος *coverage* παίρνει τη μέγιστη τιμή του 1, ενώ η ελάχιστη τιμή 0 αντιστοιχεί στην περίπτωση που δεν ανακαλύπτονται μοντέλα κοινοτήτων.

Η ταυτόχρονη βελτιστοποίηση των 2 αυτών κριτηρίων από ένα σύνολο μοντέλων κοινοτήτων δείχνει την παρουσία χρήσιμων γνώσεων στο σύνολο αυτό. Θα πρέπει να τονιστεί, ότι αυτές οι μετρικές αξιολόγησης είναι ανεξάρτητες από τις μεθόδους κατασκευής κοινοτήτων και αποτελούν αντικειμενικά κριτήρια για την αξιολόγηση των μεθόδων αυτών.

4.2 Τεχνολογίες που χρησιμοποιήθηκαν

Στο στάδιο της συλλογής και προ-επεξεργασίας δεδομένων χρησιμοποιήθηκε η γλώσσα **PHP** για όλες τις διαδικασίες φιλτραρίσματος και εκκαθάρισης των αρχείων καταγραφής (log files). Επίσης δημιουργήθηκαν **PHP** συναρτήσεις για το διαχωρισμό συνεδριών πρόσβασης, δίνοντας σαν όρισμα το επιθυμητό χρονικό πλαίσιο.

Στο στάδιο εύρεσης προτύπου, για την εκτέλεση του cluster mining αλγορίθμου χρησιμοποιήθηκε ένα open-source λογισμικό που αναπτύχθηκε από το *Ινστιτούτο Πληροφορικής και Τηλεπικοινωνιών του Εθνικού Κέντρου Έρευνας Φυσικών Επιστημών*

«Δημόκριτος». Το σύστημα αυτό υλοποιήθηκε σε γλώσσα **Java** και παρέχει ένα σχήμα επικοινωνίας πελάτη – εξυπηρετητή το οποίο συνδέεται με μια **SQL** βάση δεδομένων. Το συγκεκριμένο λογισμικό εγκαταστάθηκε πάνω στην πλατφόρμα του εργαλείου **Net Beans**. Οι λειτουργίες που υποστηρίζει, κατηγοριοποιούνται σε 3 διαφορετικές μεθόδους (Personal Mode, Community Mode και Stereotype Mode). Στην την παρούσα μελέτη χρησιμοποιούνται οι δυο πρώτες μέθοδοι.

Στο *Personal Mode*, οι λειτουργίες του εξυπηρετητή δημιουργούν και αποθηκεύουν το προφίλ των χρηστών. Το προφίλ του εκάστοτε χρήστη είναι ένα σύνολο από εγγραφές της μορφής (χαρακτηριστικό, τιμή). Αρχικά εισάγονται όλα τα χαρακτηριστικά της εφαρμογής δίνοντας τους προκαθορισμένες τιμές και στη συνέχεια ενημερώνονται ανάλογα με το είδος συσχέτισης του κάθε χρήστη με τα χαρακτηριστικά αυτά.

Στο *Community Mode*, δημιουργείται ο κανονικοποιημένος γράφος και εκτελείται ο cluster mining αλγόριθμος, με συγκεκριμένο κάθε φορά ελάχιστο κατώφλι, για την δημιουργία μοντέλων κοινοτήτων. Οι κοινότητες που δημιουργούνται, περιέχουν τα χαρακτηριστικά των χρηστών.

Αξίζει να σημειωθεί, ότι οι χρήστες και τα χαρακτηριστικά τους αντιμετωπίζονται ως οντότητες στο σύστημα αυτό και προσαρμόζονται στις ανάγκες του κάθε πειράματος. Συγκεκριμένα για τη μελέτη της πρώτης περίπτωσης, ως χρήστες ορίζονται οι συνεδρίες πρόσβασης και παράγονται κοινότητες από επισημειώσεις, που είναι τα χαρακτηριστικά των χρηστών. Αντιθέτως, στην δεύτερη μελέτη περίπτωσης, ως χρήστες ορίζονται οι επισημειώσεις που υπάρχουν μέσα στις συνεδρίες με αποτέλεσμα να δημιουργούνται κοινότητες από συνεδρίες.

Ολοκληρώνοντας, η εμφάνιση των μοντέλων κοινοτήτων γίνεται με την μορφή πινάκων σε **HTML** αρχεία, μέσω της γλώσσας **PHP**.

5

Υλοποίηση

5.1 Μελέτη περίπτωσης A: Δημιουργία model-based γραφήματος για την εξαγωγή μοντέλων κοινοτήτων

5.1.1 Διαδικασία αρχικοποίησης

Όπως αναφέρθηκε και στο κεφάλαιο που προηγήθηκε, για τη συγκεκριμένη μελέτη αναλύθηκαν αρχεία καταγραφής (log files) ενός εξυπηρετητή του παγκόσμιου ιστού που παρέχει τη δυνατότητα χαρακτηρισμού των πόρων που διαθέτει, μέσω κοινωνικών επισημειώσεων. Στόχος της επεξεργασίας είναι να προσδιοριστούν ομάδες χρηστών που χρησιμοποιούν κοινές επισημειώσεις για την αναζήτηση και ανάκτηση πόρων, δηλαδή τα μέλη τους παρουσιάζουν παρόμοια συμπεριφορά χρήσης και διαθέτουν κοινά ενδιαφέροντα.

Σύμφωνα με την μεθοδολογία που παρουσιάστηκε στα στάδια συλλογής και προ-επεξεργασίας δεδομένων, δημιουργήθηκαν 519 συνεδρίες πρόσβασης χρηστών που περιείχαν τουλάχιστον δυο επισημειώσεις η καθεμία και χρησιμοποιήθηκαν συνολικά 3.000 επισημειώσεις. Επίσης, οι συνεδρίες αυτές μετατράπηκαν σε χαρακτηριστικά, όπου κάθε χαρακτηριστικό αναπαριστούσε κάθε μοναδική επισημείωση που υπήρχε σε τουλάχιστον μια συνεδρία, δημιουργώντας με αυτό τον τρόπο 424 μοναδιαία χαρακτηριστικά.

Οπότε, ο κανονικοποιημένος γράφος που δημιουργήθηκε στο τρίτο στάδιο επεξεργασίας, περιέχει σαν κόμβους τα χαρακτηριστικά των χρηστών και σαν ακμές την συνύπαρξη δυο χαρακτηριστικών μέσα σε μια συνεδρία πρόσβασης. Κοινώς, δυο κόμβοι-επισημειώσεις συνδέονται μεταξύ τους στην περίπτωση που συνυπάρχουν σε μια συνεδρία πρόσβασης. Επίσης, το βάρος του κάθε κόμβου υπολογίζεται βάσει της συχνότητας εμφάνισης της κάθε επισημείωσης ξεχωριστά μέσα σε όλες τις συνεδρίες συνολικά, ενώ το βάρος της κάθε ακμής υπολογίζεται βάσει της συχνότητας συνύπαρξης δυο επισημειώσεων μέσα σε μια συνεδρία.

Το γράφημα αυτό αποτελείται από 424 κόμβους, όσες και οι μοναδιαίες κοινωνικές επισημειώσεις που ανακτήθηκαν, και θεωρείται ένα model – based γράφημα. Κατόπιν, έχοντας υπολογίσει τα βάρη των ακμών και των κόμβων του, εκτελείται ο αλγόριθμος cluster mining, όπως αναλύθηκε παραπάνω, για την εξαγωγή μοντέλων κοινοτήτων. Στη συγκεκριμένη μελέτη χρησιμοποιήθηκαν τρία διαφορετικά ελάχιστα κατώφλια (thresholds) για την εκτέλεση του συγκεκριμένου αλγορίθμου με τιμές 0.95, 0.97 και 1.0. Στην παρακάτω ενότητα παρατίθενται τα αποτελέσματα.

5.1.2 Αποτελέσματα

5.1.2.1 Κοινότητες με ελάχιστο κατώφλι 0.95

community_1	cultural heritage, museums,	session_263, session_324
community_2	bible, xslt,	session_364, session_374
community_3	fear of crime, victimisation,	session_323, session_365
community_4	διανοούμενοι, ρομαντισμός,	session_259, session_260
community_5	urban history,	session_412, session_503
community_6	κοινωνιολογία,	session_415, session_420, session_482, session_68
community_7	αξιολογικός μηδενισμός,	session_304, session_414, session_444, session_84
community_8	ιστορία του πολέμου,	session_173, session_268
community_9	apa - οδηγός συγγραφής,	session_161, session_393, session_421, session_65, session_97

community_10	φιλοσοφία,	session_255, session_416
community_11	αλλοτρίωση,	session_176, session_178, session_223, session_225, session_42, session_57, session_66
community_12	διεθνείς σχέσεις,	session_15, session_242, session_293, session_313, session_40, session_46, session_59,
community_13	μάης 68,	session_162, session_183
community_14	μεσαίωνα,	session_191, session_45
community_15	διεθνής πολιτική, διεθνείς σχέσεις, στρατηγικές σπουδές,	session_168, session_172
community_16	verstehen,	session_1, session_167, session_422, session_457, session_460
community_17	άμεση προσέγγιση, έμμεση προσέγγιση, αθήνα, αλλαγή, διεθνές σύστημα, διεθνής ασφάλεια, διεθνής πολιτική, διεθνείς σχέσεις, επιθετικός ρεαλισμός, θεωρία πολέμου, θουκυδίδης, ισχύς, κριτική, μεθοδολογία, πολιτικός ρεαλισμός, πυρηνική στρατηγική, στρατηγικές σπουδές, στρατηγική της εκμηδένισης, στρατηγική της εξουθένωσης, συστημικός ρεαλισμός, σχολή του chicago, φιλελευθερισμός, χερσαία ισχύς, ψυχρός πόλεμος, dellburck, offensive realism, status quo,	session_166, session_35
community_18	νεομαρξισμός, φιλοσοφία,	session_49, session_53, session_64
community_19	στρατηγικές σπουδές,	session_103, session_131, session_152,

		session_187, session_314, session_459, session_48, session_504, session_79, session_85
community_20	behavior analysis,	session_23, session_47
community_21	απεγκληματοποίηση, αποποινικοποίηση, θετικισμός, ποινικοποίηση,	session_216, session_275
community_22	princeton,	session_372, session_379
community_23	apa, apa - οδηγός συγγραφής,	session_270, session_337
community_24	avatar,	session_401, session_452,
community_25	θεωρία πολέμου, φιλοσοφία πολέμου,	session_157, session_189, session_195, session_203, session_371, session_387, session_51,
community_26	γερμανοί φιλόσοφοι, max weber,	session_155, session_277,
community_27	δειγματοληπτικές έρευνες, κοινωνικές έρευνες,	session_204, session_244, session_481,
community_28	γερμανοί φιλόσοφοι, σχολή της φρανκφούρτης, verstehen,	session_116, session_121, session_129, session_174, session_184, session_214, session_221, session_383, session_388, session_441,

community_29	αντεγκληματική πολιτική,	session_316, session_462, session_50,
community_30	μουσειολογία, μουσειολογικά,	session_177, session_179,
community_31	έμμεση προσέγγιση, στρατηγικές σπουδές,	session_235, session_489,
community_32	κοινωνικές έρευνες,	session_149, session_280, session_430, session_56,
community_33	προφορικότητα,	session_12, session_245,
community_34	γερμανοί φιλόσοφοι, max weber,	session_120, session_190, session_236, session_256,
community_35	semantic web, xml,	session_104, session_192, session_193,
community_36	Έρευνα οικογενειακών προϋπολογισμών,	session_101, session_343,
community_37	οκτωβριανή επανάσταση,	session_100, session_133,
community_38	θεωρία κράτους, παγκοσμιοποίηση,	session_303, session_390,
community_39	behavioral psychology,	session_305, session_312,
community_40	ονε,	session_25, session_28,
community_41	opac, usability,	session_200, session_22,
community_42	opac, usability,	session_115, session_126, session_22, session_70,

community_43	γερμανοί φιλόσοφοι, φιλοσοφία,	session_10, session_188,
community_44	πολιτικός ρεαλισμός,	session_110, session_472, session_63,
community_45	μμε,	session_394, session_428, session_94,
community_46	china, science,	session_330, session_395,
community_47	ρεπορτάζ, τεχνικές έρευνας,	session_114, session_136,
community_48	γαλλική επανάσταση,	session_113, session_231,
community_49	πολιτική κοινωνία, πόλεμος θέσεων,	session_33, session_60,
community_50	έρευνες μετρήσεων εγκληματικότητας, έρευνες παραγόντων εγκληματογένεσης, criminological research,	session_339, session_340,
community_51	κριτική θεωρία, σχολή της φραγκφούρτης,	session_135, session_137, session_14,
community_52	xhtml, xml,	session_338, session_344,
community_53	marx,	session_150, session_479,
community_54	ευρωπαϊκό δικαστήριο, φιλελευθερισμός,	session_207, session_463,

Όπως φαίνεται και στο παραπάνω πίνακα, δημιουργήθηκαν 54 διαφορετικές κοινότητες επισημειώσεων που η καθεμία αναπαριστά μια διαφορετική ομάδα χρηστών με παρόμοια πρότυπα συμπεριφοράς. Η τελευταία στήλη παρουσιάζει τις συνεδρίες που ανήκουν οι επισημειώσεις της κάθε κοινότητας ξεχωριστά. Το μεγαλύτερο μέρος των κοινοτήτων αποτελείται από ένα μικρό εύρος επισημειώσεων και συνεδριών καθώς το κατώφλι που επιλέχτηκε θεωρείται σχετικά μικρό. Αξίζει να σημειωθεί ότι υπάρχουν κοινότητες με μεγάλο εύρος επισημειώσεων, όπως η κοινότητα 17, που ανήκουν σε αυτή μόνο δυο συνεδρίες και κοινότητες με μικρό εύρος επισημειώσεων στις οποίες όμως ανήκουν

πάρα πολλές συνεδρίες, όπως η κοινότητα 28. Και στις δυο περιπτώσεις, ο υπογράφος που δημιουργείται έχει ακμές με σχετικά μεγάλο βάρος καθώς οι επισημειώσεις των κοινοτήτων αυτών συνυπάρχουν είτε σε μικρό είτε σε μεγάλο εύρος και μέγεθος διαφορετικών συνεδριών. Για παράδειγμα, στην κοινότητα 28 οι επισημειώσεις «γερμανοί φιλόσοφοι», «σχολή της φρανκφούρτης» και «verstehen» συνυπάρχουν σε δεκα διαφορετικές συνεδρίες.

Συνοψίζοντας, στο συγκεκριμένο κατώφλι παρουσιάζεται μια αναλογία όσον αφορά στο εύρος των επισημειώσεων και των συνεδριών που ανήκουν στην εκάστοτε κοινότητα και παρουσιάζεται μια σχετική ομοιομορφία στο μέγεθος των κοινοτήτων.

5.1.2.2 Κοινότητες με ελάχιστο κατώφλι 0.97

community_1	cultural heritage, museums,	session_263, session_324,
community_2	bible, xslt,	session_364, session_374
community_3	fear of crime, victimisation,	session_323, session_365
community_4	διανοούμενοι, ρομαντισμός,	session_259, session_260
community_5	urban history,	session_412, session_503
community_6	κοινωνιολογία,	session_415, session_420, session_482, session_68,
community_7	αξιολογικός μηδενισμός,	session_304, session_414, session_444, session_84,
community_8	ιστορία του πολέμου,	session_173, session_268,
community_9	apa - οδηγός συγγραφής,	session_161, session_393, session_421, session_65, session_97,
community_10	φιλοσοφία,	session_255, session_416,
community_11	αλλοτρίωση,	session_176, session_178, session_223, session_225, session_42, session_57, session_66,

community_12	διεθνείς σχέσεις,	session_15, session_242, session_293, session_313, session_40, session_46, session_59,
community_13	μάης 68,	session_162, session_183,
community_14	μεσαίωνα,	session_191, session_45,
community_15	διεθνής πολιτική, διεθνείς σχέσεις, στρατηγικές σπουδές,	session_168, session_172,
community_16	verstehen,	session_1, session_167, session_422, session_457, session_460,
community_17	άμεση προσέγγιση, έμμεση προσέγγιση, αθήνα, αλλαγή, διεθνές σύστημα, διεθνής ασφάλεια, διεθνής πολιτική, διεθνείς σχέσεις, επιθετικός ρεαλισμός, θεωρία πολέμου, θουκυδίδης, ισχύς, κριτική, μεθοδολογία, πολιτικός ρεαλισμός, πυρηνική στρατηγική, στρατηγικές σπουδές, στρατηγική της εκμηδένισης, στρατηγική της εξουθένωσης, συστημικός ρεαλισμός, σχολή του chicago, φιλελευθερισμός, χειρσαία ισχύς, ψυχρός πόλεμος, dellburck, offensive realism, status quo,	session_166, session_35,
community_18	νεομαρξισμός, φιλοσοφία,	session_49, session_53, session_64,
community_19	στρατηγικές σπουδές,	session_103, session_131, session_152, session_187, session_314, session_459, session_48, session_504, session_79, session_85,
community_20	behavior analysis,	session_23, session_47,
community_21	princeton,	session_372, session_379,
community_22	apa, apa - οδηγός συγγραφής,	session_270, session_337,

community_23	avatar,	session_401, session_452,
community_24	θεωρία πολέμου, φιλοσοφία πολέμου,	session_157, session_189, session_195, session_203, session_371, session_387, session_51,
community_25	γερμανοί φιλόσοφοι, max weber,	session_155, session_277,
community_26	δειγματοληπτικές έρευνες, κοινωνικές έρευνες,	session_204, session_244, session_481,
community_27	γερμανοί φιλόσοφοι,	session_116, session_121, session_129, session_174, session_214, session_221, session_383, session_388, session_441,
community_28	αντεγκληματική πολιτική,	session_316, session_462, session_50,
community_29	μουσειολογία, μουσειολογικά,	session_177, session_179,
community_30	έμμεση προσέγγιση, στρατηγικές σπουδές,	session_235, session_489,
community_31	κοινωνικές έρευνες,	session_149, session_280, session_430, session_56,
community_32	προφορικότητα,	session_12, session_245,
community_33	γερμανοί φιλόσοφοι, max weber,	session_120, session_190, session_236, session_256,
community_34	semantic web, xml,	session_104, session_192, session_193,
community_35	Έρευνα οικογενειακών προϋπολογισμών,	session_101, session_343,
community_36	οκτωβριανή επανάσταση,	session_100, session_133,
community_37	θεωρία κράτους, παγκοσμιοποίηση,	session_303, session_390,
community_38	behavioral psychology,	session_305, session_312,

community_39	ove,	session_25, session_28,
community_40	opac, usability,	session_115, session_126, session_22, session_70,
community_41	γερμανοί φιλόσοφοι, φιλοσοφία,	session_10, session_188,
community_42	πολιτικός ρεαλισμός,	session_110, session_472, session_63,
community_43	μμε,	session_394, session_428, session_94,
community_44	china, science,	session_330, session_395,
community_45	ρεπορτάζ, τεχνικές έρευνας,	session_114, session_136,
community_46	γαλλική επανάσταση,	session_113, session_231,
community_47	πολιτική κοινωνία, πόλεμος θέσεων,	session_33, session_60,
community_48	κριτική θεωρία, σχολή της φραγκφούρτης,	session_135, session_137, session_14,
community_49	xhtml, xml,	session_338, session_344,
community_50	marx,	session_150, session_479,
community_51	ευρωπαϊκό δικαστήριο, φιλελευθερισμός,	session_207, session_463,

Όπως παρατηρείται και στο παραπάνω πίνακα, δεν υπάρχουν σημαντικές διαφορές στα μοντέλα κοινοτήτων που δημιουργήθηκαν με κατώφλι 0.97 σε σχέση με τα μοντέλα που δημιουργήθηκαν με κατώφλι 0.95. Συγκεκριμένα, δημιουργήθηκαν 51 κοινότητες επισημειώσεων, οι οποίες έχουν παρόμοια χαρακτηριστικά με τις προηγούμενες κοινότητες διατηρώντας τη σχετική ομοιομορφία που υπήρχε και πριν, όσον αφορά στο μέγεθος των κοινοτήτων συνολικά. Οπότε και σε αυτή την περίπτωση, παρουσιάζεται μια σχετική αναλογία όσον αφορά στο εύρος των επισημειώσεων και των συνεδριών που ανήκουν στην εκάστοτε κοινότητα.

5.1.2.3 Κοινότητες με ελάχιστο κατώφλι 1.0

community_1	cultural heritage, museums,	session_263, session_324,
community_2	bible, xslt,	session_364, session_374,
community_3	fear of crime, victimisation,	session_323, session_365,
community_4	semantic web, xml,	session_104,session_192, session_193,
community_5	δειγματοληπτικές έρευνες, κοινωνικές έρευνες,	session_204,session_244, session_481
community_6	νεομαρξισμός, φιλοσοφία,	session_49, session_53, session_64
community_7	πολιτική κοινωνία, πόλεμος θέσεων,	session_33, session_60
community_8	θεωρία κράτους, παγκοσμιοποίηση,	session_303, session_390,
community_9	γερμανοί φιλόσοφοι, φιλοσοφία,	session_10, session_188
community_10	xhtml, xml,	session_338, session_344,
community_11	apa, apa - οδηγός συγγραφής,	session_270, session_337
community_12	ρεπορτάζ, τεχνικές έρευνας,	session_114, session_136
community_13	κριτική θεωρία, σχολή της φραγκφούρτης,	session_135,session_137, session_14

Όπως φαίνεται και στο παραπάνω πίνακα, παρατηρείται μια σημαντική μείωση στον αριθμό των μοντέλων κοινοτήτων που εξάγονται από τον αλγόριθμο cluster mining. Από τις 51 κοινότητες που υπήρχαν με κατώφλι 0.97, πλέον υπάρχουν μόνο 13, γεγονός που υποδηλώνει ότι το συγκεκριμένο κατώφλι είναι πολύ ισχυρό. Εδώ παρατηρείται μια πλήρης ομοιομορφία στο μέγεθος των κοινοτήτων που έχουν δημιουργηθεί καθώς όλες οι κοινότητες αποτελούνται από δυο επισημειώσεις οι οποίες έχουν και σημασιολογική συνάφεια μεταξύ τους. Για παράδειγμα, η κοινότητα 4 αποτελείται από τις επισημειώσεις «semantic web» και «xml», η οποία αναπαριστά μια ομάδα χρηστών που έχουν σαν κριτήριο αναζήτησης έννοιες του σημασιολογικού ιστού. Αντιθέτως, υπάρχουν πολλές κοινότητες που χρησιμοποιούν επισημειώσεις σχετικές με τον τομέα των πολιτικών και κοινωνικών επιστημών, λόγω και της φύσης των δεδομένων χρήσης τα οποία προέρχονται από το Πάντειο Πανεπιστήμιο.

Επιπλέον, παρατηρείται ότι μόνο δύο κοινότητες έχουν μια κοινή επισημείωση, τη «φιλοσοφία», πράγμα που δείχνει ότι η καθεμιά τους είναι μοναδική και αντιπροσωπευτική της κάθε ομάδας χρηστών που αναπαριστά. Τέλος, οι κοινότητες αυτές έχουν προκύψει από τα υπογραφήματα με το μεγαλύτερο βάρος στις ακμές τους. Στην παρακάτω ενότητα θα εξεταστούν τα 2 κριτήρια αξιολόγησης, που αναλύθηκαν παραπάνω, για να επιλεγεί το καταλληλότερο και πιο αντιπροσωπευτικό σύνολο μοντέλων κοινοτήτων για τη συγκεκριμένη μελέτη.

5.1.3 Αξιολόγηση αποτελεσμάτων

Στο προηγούμενο κεφάλαιο παρουσιάστηκαν και αναλύθηκαν τα δυο κριτήρια αξιολόγησης των μοντέλων κοινοτήτων, βάσει των οποίων θα επιλεγεί ποιο σύνολο μοντέλων είναι το καταλληλότερο και πιο αντιπροσωπευτικό για το πείραμα αυτό. Παρακάτω παρατίθενται οι μετρικές αυτές για τα τρία διαφορετικά σύνολα που δημιουργήθηκαν.

Threshold	<i>Distinctiveness</i>	<i>Coverage</i>
0.95	90/112 (0.803)	90/424 (0.212)
0.97	82/101 (0.811)	82/424 (0.193)
1.0	25/26 (0.960)	25/424 (0.058)

Παρατηρείται ότι όσο μεγαλώνει το ελάχιστο κατώφλι τόσο αυξάνεται η μετρική *Distinctiveness*, ενώ αντίστοιχα η μετρική *Coverage* μειώνεται συνεχώς. Αυτό οφείλεται στην σημαντική μείωση του μεγέθους συνολικά όλων των κοινοτήτων που προκαλεί η αύξηση του ελάχιστου κατωφλίου. Λόγω αυτής της ιδιαιτερότητας που παρουσιάζουν τα μοντέλα αυτά, θα πρέπει να συνυπολογιστούν και τα 2 κριτήρια επί ίσοις όροις χωρίς να δοθεί σε κανένα κάποια προτεραιότητα.

Συγκεκριμένα, τα μοντέλα κοινοτήτων με κατώφλι 1.0 παρουσιάζουν πάρα πολύ υψηλό *Distinctiveness*, καθώς οι 25 από τις 26 επισημειώσεις που ανήκουν σε κοινότητες είναι διαφορετικές μεταξύ τους, αλλά εξαιρετικά χαμηλό *Coverage* (μόνο 25 από τις 424 επισημειώσεις χρησιμοποιήθηκαν). Οπότε, τα συγκεκριμένα μοντέλα δεν μπορούν να θεωρηθούν αντιπροσωπευτικά του τρόπου χρήσης των επισημειώσεων από τους χρήστες του συστήματος, παρόλο που οι κοινότητες που δημιουργούν είναι ως επί των πλείστων τελείως διαφορετικές μεταξύ τους.

Αντιθέτως, τα μοντέλα που παράγονται από τα κατώφλια 0.95 και 0.97 έχουν πολλά κοινά χαρακτηριστικά και παράγουν αρκετές όμοιες κοινότητες επισημειώσεων. Ως αποτέλεσμα, και οι μετρικές των δυο αυτών συνόλων έχουν σχεδόν την ίδια τιμή. Λόγω αυτής της σύγκλισης τιμών, επιλέχθηκε να δοθεί μια μικρή προτεραιότητα στο *Distinctiveness* έτσι ώστε να επιλεγθούν τα λιγότερα επικαλυπτόμενα μοντέλα που ούτως ή άλλως έχουν σχεδόν το ίδιο ποσοστό κάλυψης των συνολικών επισημειώσεων του πειράματος.

Οπότε, για την συγκεκριμένη μελέτη και βάσει των αποτελεσμάτων και των αναλύσεων που έγιναν, επιλέχθηκε **το σύνολο των μοντέλων κοινοτήτων που δημιουργήθηκαν με ελάχιστο κατώφλι 0.97**, ως το καταλληλότερο για το πείραμα αυτό.

5.2 Μελέτη περίπτωσης B: Δημιουργία item-based

γραφήματος για την εξαγωγή μοντέλων κοινοτήτων

5.2.1 Διαδικασία αρχικοποίησης

Για τη συγκεκριμένη μελέτη χρησιμοποιήθηκαν τα ίδια δεδομένα χρήσης με σκοπό τον προσδιορισμό ομάδων χρηστών που έχουν παρόμοια συμπεριφορά χρήσης και διαθέτουν κοινά ενδιαφέροντα. Σύμφωνα με την μεθοδολογία που παρουσιάστηκε στα στάδια συλλογής

και προ-επεξεργασίας δεδομένων, δημιουργήθηκαν 519 συνεδρίες πρόσβασης χρηστών που περιείχαν τουλάχιστον δυο επισημειώσεις η καθεμία και χρησιμοποιήθηκαν συνολικά 3.000 επισημειώσεις.

Η βασική διαφορά της συγκεκριμένης μελέτης με την προηγούμενη, έγκυται στη μετατροπή των αντικειμένων, που δημιουργήθηκαν στα πρώτα δυο στάδια, σε χαρακτηριστικά. Συγκεκριμένα, από τις 3000 επισημειώσεις που χρησιμοποιήθηκαν σαν κριτήριο αναζήτησης πόρων, υπήρχαν 424 μοναδικές επισημειώσεις οι οποίες εμφανίζονταν σε τουλάχιστον μια συνεδρία πρόσβασης. Στην μελέτη αυτή, κάθε χαρακτηριστικό αναπαριστά την κάθε συνεδρία στην οποία υπάρχει τουλάχιστον μια μοναδική επισημείωση, δημιουργώντας με αυτό τον τρόπο 519 μοναδικά χαρακτηριστικά.

Επομένως, ο κανονικοποιημένος γράφος που δημιουργήθηκε στο τρίτο στάδιο επεξεργασίας, περιέχει σαν κόμβους τα χαρακτηριστικά των χρηστών και σαν ακμές την ύπαρξη τουλάχιστον μιας κοινής επισημείωσης σε 2 χαρακτηριστικά. Κοινώς, δυο κόμβοι-συνεδρίες πρόσβασης συνδέονται μεταξύ τους στην περίπτωση που υπάρχει τουλάχιστον μια κοινή επισημείωση στις δυο αυτές συνεδρίες. Επίσης, το βάρος του κάθε κόμβου υπολογίζεται βάσει της συχνότητας εμφάνισης των μοναδικών κοινωνικών επισημειώσεων που υπάρχουν σε κάθε συνεδρία ξεχωριστά, ενώ το βάρος της κάθε ακμής υπολογίζεται βάσει της συχνότητας εμφάνισης κοινών επισημειώσεων μέσα σε δυο συνεδρίες πρόσβασης.

Το γράφημα αυτό αποτελείται από 519 κόμβους, όσες και οι συνεδρίες πρόσβασης, και θεωρείται ένα item – based γράφημα. Έχοντας υπολογίσει τα βάρη των ακμών και των κόμβων του, στη συνέχεια εκτελείται ο αλγόριθμος cluster mining, όπως αναλύθηκε παραπάνω, για την εξαγωγή μοντέλων κοινότητων. Στη συγκεκριμένη μελέτη χρησιμοποιήθηκαν τρία διαφορετικά ελάχιστα κατώφλια (thresholds) για την εκτέλεση του συγκεκριμένου αλγορίθμου με τιμές 0.95, 0.97 και 1.0. Στην παρακάτω ενότητα παρατίθενται τα αποτελέσματα.

5.2.2 Αποτελέσματα

5.2.2.1 Κοινότητες με ελάχιστο κατώφλι 0.95

community_1	session_76, session_95,	κοινωνικές στάσεις, κοσμοθεωρία, μηδενική ανοχή, ποινικό σύστημα, υποκειμενική τιμωρητικότητα,
community_2	session_455,	μαρξιστική θεωρία, marxist theory, social theory,
community_3	session_505,	ancient, classics, college, social history,
community_4	session_514,	cultural theory, english literature,
community_5	session_509,	french, textbooks,

community_6	session_140, session_141, session_142, session_143, session_325, session_346, session_373,	ecce homo, ηθική στην διεθνή πολιτική, ιταλοί φιλόσοφοι, περικλής, θαλάσσια ισχύς,
community_7	session_512,	management, socialism,
community_8	session_497,	anthology, christianity, colonialism, feminism, film studies, linguistics, us history, world history,
community_9	session_499,	academia, greek,
community_10	session_5,	ασφαλιστικοί γοσ, γενική ασφάλεια προϊόντων, καταναλωτική πώληση, συμβάσεις από απόσταση, συμβάσεις εκτός εμπορικού καταστήματος, συστημικός ρεαλισμός,
community_11	session_501,	religion, religious history,
community_12	session_471,	death visual arts, postmodernism, postmodernism history, tech, θανατος, visual culture, visual rhetoric, visual thinking,
community_13	session_456,	class, class structure, class struggle,
community_14	session_491,	economic and social issues, european neighbourhood policy: political,
community_15	session_240,	αυθόρμητο, ολότητα,
community_16	session_432,	graphics, illuminated manuscripts, intellectual history, ir, marxism, marxism-leninism, mesopotamia, political science, political theory, social liberalism, wishlist,
community_17	session_451,	essays, history of ideas, russia, unionism, united states, us,
community_18	session_140, session_143,	collodion, κοινωνία πολιτών, organizational knowledge, saladin, studies in politics, virtual environments,
community_19	session_140, session_143, session_166, session_89,	dellburck, πυρηνική στρατηγική,

community_20	session_140, session_143, session_261,	anti, qweqwe,
community_21	session_436,	επιστήμη θετικής σκέψης, ηδονή του πόνου,
community_22	session_140, session_143, session_166,	dellburck, ελληνοτουρκικός πόλεμος, ευρωπαϊκή ολοκλήρωση, περιγραφική θεωρία,
community_23	session_286,	ακροπολη, λένα, μάης, οικονομια, χαραλαμποπούλου,
community_24	session_439,	nicola acocicella, principes of economy nicola apicella,
community_25	session_110, session_13, session_140, session_141, session_142, session_143, session_159, session_166, session_169, session_306, session_35, session_472, session_5, session_63, session_72, session_80, session_82, session_9, session_92,	διεθνές σύστημα, πολιτικός ρεαλισμός,
community_26	session_480,	art psychology, mellon lectures, vision,
community_27	session_510,	greek literature, own, statistics,
community_28	session_488,	φλωρινα, πολιτικη φλωρινα, πολιτιστικη πολιτικη,
community_29	session_450,	disability, international relations,
community_30	session_516,	europe, sexuality,
community_31	session_515,	america, research,
community_32	session_110, session_13, session_140, session_141, session_142, session_143, session_159, session_166, session_169, session_306, session_35, session_472, session_5, session_63, session_72, session_80, session_82, session_9, session_92,	κριτική, πολιτικός ρεαλισμός, θουκυδίδης, χερσαία ισχύς,
community_33	session_215,	δημογραφία - μέθοδοι - ανάλυση, προβολές,
community_34	session_267,	εγκλημα, μοντέλα κοινωνικού τύπου,

community_35	session_294,	διεθνές συνέδριο παντείου, ίδρυμα τύπος α.ε.,
community_36	session_140, session_142, session_143,	αποτροπή, structural funds,
community_37	session_248,	διαρθρωτικά ταμεία, κεππα,
community_38	session_493,	africa, autobiography,
community_39	session_507,	imperialism, literature,
community_40	session_490,	asian history, japanese,
community_41	session_508,	history of science, reference,

Όπως φαίνεται και στο παραπάνω πίνακα, δημιουργήθηκαν 41 διαφορετικές κοινότητες συνεδριών πρόσβασης που η καθεμία αναπαριστά μια διαφορετική ομάδα χρηστών με παρόμοια πρότυπα συμπεριφοράς. Η τελευταία στήλη παρουσιάζει τις επισημειώσεις που συνυπάρχουν μέσα στις συνεδρίες της κάθε κοινότητας ξεχωριστά. Το μεγαλύτερο μέρος των κοινοτήτων αποτελείται από ένα μικρό εύρος συνεδριών και επισημειώσεων καθώς το κατώφλι που επιλέχτηκε θεωρείται σχετικά μικρό. Αξίζει να σημειωθεί ότι υπάρχουν κοινότητες με μεγάλο αριθμό συνεδριών πρόσβασης όπως η κοινότητα 25, που συνυπάρχουν σε αυτή μόνο δυο επισημειώσεις καθώς και κοινότητες με μικρό αριθμό συνεδριών πρόσβασης στις οποίες όμως συνυπάρχουν πάρα πολλές επισημειώσεις, όπως η κοινότητα 16. Και στις δυο αυτές περιπτώσεις, ο υπογράφος που δημιουργείται έχει ακμές με πολύ μεγάλο βάρος καθώς στις συνεδρίες των κοινοτήτων αυτών συνυπάρχει είτε μικρό είτε μεγάλο εύρος διαφορετικών επισημειώσεων. Αξίζει να σημειωθεί, ότι υπάρχουν και πολλές κοινότητες με μια μόνο συνεδρία πρόσβασης, όπως η κοινότητα 16, στις οποίες ο κόμβος που αποτελεί την εκάστοτε κοινότητα έχει πολύ μεγάλο βάρος.

Συνοψίζοντας, για το συγκεκριμένο κατώφλι παρουσιάζεται μια δυσαναλογία όσον αφορά το εύρος των επισημειώσεων και των συνεδριών που ανήκουν στην εκάστοτε κοινότητα, αλλά παρουσιάζεται μια σχετική ομοιομορφία στο μέγεθος των κοινοτήτων.

5.2.2.2 Κοινότητες με ελάχιστο κατώφλι 0.97

community_1	session_76, session_95,	κοινωνικές στάσεις, κοσμοθεωρία, μηδενική ανοχή, ποινικό σύστημα, υποκειμενική τιμωρητικότητα,
community_2	session_455,	μαρξιστική θεωρία, marxist theory, social theory,
community_3	session_505,	ancient, classics, college, social history,

community_4	session_514,	cultural theory, english literature,
community_5	session_509,	french, textbooks,
community_6	session_140, session_143, session_373,	session_141, session_325, session_142, session_346, ecce homo, ηθική στην διεθνή πολιτική, ιταλοί φιλόσοφοι, περικλής, θαλάσσια ισχύς,
community_7	session_512,	management, socialism,
community_8	session_497,	anthology, christianity, colonialism, feminism, film studies, linguistics, us history, world history,
community_9	session_499,	academia, greek,
community_10	session_5,	ασφαλιστικοί γοσ, γενική ασφάλεια προϊόντων, καταναλωτική πώληση, συμβάσεις από απόσταση, συμβάσεις εκτός εμπορικού καταστήματος, συστημικός ρεαλισμός,
community_11	session_471,	death visual arts, postmodernism, postmodernism history, tech, θανατος, visual culture, visual rhetoric, visual thinking,
community_12	session_501,	religion, religious history,
community_13	session_456,	class, class structure, class struggle,
community_14	session_491,	economic and social issues, european neighbourhood policy: political,
community_15	session_240,	αυθόρμητο, ολότητα,
community_16	session_432,	graphics, illuminated manuscripts, intellectual history, ir, marxism, marxism- leninism, mesopotamia, political science, political theory, social liberalism, wishlist,
community_17	session_451,	essays, history of ideas, russia, unionism, united states, us,
community_18	session_140, session_143,	collodion, κοινωνία πολιτών, organizational knowledge, saladin, studies in politics, virtual environments,

community_19	session_436,	επιστήμη θετικής σκέψης, ηδονή του πόνου,
community_20	session_286,	ακροπολη, λένα, μάης, οικονομια, χαλαραμποπούλου,
community_21	session_439,	nicola acocicella, principes of economy nicola apicella,
community_22	session_110, session_13, session_140, session_141, session_142, session_143, session_159, session_166, session_169, session_306, session_35, session_472, session_5, session_63, session_72, session_80, session_82, session_9, session_92,	διεθνές σύστημα, πολιτικός ρεαλισμός,
community_23	session_480,	art psychology, mellon lectures, vision,
community_24	session_510,	greek literature, own, statistics,
community_25	session_488,	φλωρινα, πολιτικη φλωρινα, πολιτιστικη πολιτικη,
community_26	session_450,	disability, international relations,
community_27	session_516,	europa, sexuality,
community_28	session_515,	america, research,
community_29	session_110, session_13, session_140, session_141, session_142, session_143, session_159, session_166, session_169, session_306, session_35, session_472, session_5, session_63, session_72, session_80, session_82, session_9, session_92,	κριτική, πολιτικός ρεαλισμός, θουκυδίδης, χερσαία ισχύς,
community_30	session_215,	δημογραφία - μέθοδοι - ανάλυση, προβολές,
community_31	session_267,	εγκλημα, μοντέλα κοινωνικού τύπου,
community_32	session_294,	διεθνές συνέδριο παντείου, ίδρυμα τύπος α.ε.,
community_33	session_140, session_142, session_143,	αποτροπή, structural funds,
community_34	session_248,	διαρθρωτικά ταμεία, κεππα,

community_35	session_493,	africa, autobiography,
community_36	session_507,	imperialism, literature,
community_37	session_490,	asian history, japanese,
community_38	session_508,	history of science, reference,
community_39	session_140, session_143,	ελληνοτουρκικός πόλεμος, ευρωπαϊκή ολοκλήρωση, περιγραφική θεωρία,

Όπως παρατηρείται και στο παραπάνω πίνακα, δεν υπάρχουν σημαντικές διαφορές στα μοντέλα κοινοτήτων που δημιουργήθηκαν με κατώφλι 0.97 σε σχέση με τα μοντέλα που δημιουργήθηκαν με κατώφλι 0.95. Συγκεκριμένα, δημιουργήθηκαν 39 κοινότητες συνεδριών πρόσβασης, οι οποίες έχουν παρόμοια χαρακτηριστικά με τις προηγούμενες κοινότητες διατηρώντας την σχετική ομοιομορφία που υπήρχε και πριν, όσον αφορά στο μέγεθος των κοινοτήτων συνολικά. Επίσης και σε αυτή την περίπτωση, παρουσιάζεται μια σχετική δυσαναλογία όσον αφορά το εύρος των επισημειώσεων και των συνεδριών που ανήκουν στην εκάστοτε κοινότητα.

5.2.2.3 Κοινότητες με ελάχιστο κατώφλι 1.0

community_1	session_140, session_143,	collodion, κοινωνία πολιτών, organizational knowledge, saladin, studies in politics, virtual environments,
community_2	session_143, session_166,	κριτική, θουκυδίδης, χερσαία ισχύς,

Όπως φαίνεται και στο παραπάνω πίνακα, παρατηρείται μια κατακόρυφη μείωση στον αριθμό των μοντέλων κοινοτήτων που εξάγονται από τον αλγόριθμο cluster mining. Από τις 39 κοινότητες που υπήρχαν με κατώφλι 0.97, πλέον υπάρχουν μόνο 2, γεγονός που υποδηλώνει ότι το συγκεκριμένο κατώφλι είναι πολύ ισχυρό. Οι κοινότητες αυτές έχουν προκύψει από τα υπογραφημάτα με το μεγαλύτερο βάρος στους κόμβους και στις ακμές τους.

Στο παρακάτω υποκεφάλαιο εξετάζονται τα 2 κριτήρια αξιολόγησης, που αναλύθηκαν παραπάνω, για να επιλεγεί το καταλληλότερο και πιο αντιπροσωπευτικό σύνολο μοντέλων κοινοτήτων για την συγκεκριμένη μελέτη.

5.2.3 Αξιολόγηση αποτελεσμάτων

Στην προηγούμενη ενότητα παρουσιάστηκαν και αναλύθηκαν τα δυο κριτήρια αξιολόγησης των μοντέλων κοινοτήτων, βάσει των οποίων θα επιλεγεί ποιο σύνολο μοντέλων είναι το καταλληλότερο και πιο αντιπροσωπευτικό για το πείραμα αυτό. Παρακάτω παρατίθενται οι μετρικές αυτές για τα τρία διαφορετικά σύνολα που δημιουργήθηκαν.

Threshold	<i>Distictiveness</i>	<i>Coverage</i>
0.95	57/94 (0.60)	57/519 (0.107)
0.97	55/86 (0.63)	55/519 (0.105)
1.0	3/4 (0.75)	3/519 (0.005)

Όπως και στην προηγούμενη μελέτη, παρατηρείται ότι όσο μεγαλώνει το ελάχιστο κατώφλι τόσο αυξάνεται η μετρική ***Distictiveness***, ενώ αντίστοιχα η μετρική ***Coverage*** μειώνεται συνεχώς. Αυτό οφείλεται στη σημαντική μείωση του μεγέθους όλων των κοινοτήτων συνολικά που προκαλεί η αύξηση του ελαχίστου κατωφλίου. Λόγω αυτής της ιδιαιτερότητας που παρουσιάζουν τα μοντέλα αυτά, θα πρέπει να συνυπολογιστούν και τα 2 κριτήρια επί ίσοις όροις χωρίς να δοθεί σε κανένα κάποια προτεραιότητα.

Συγκεκριμένα, τα μοντέλα κοινοτήτων με κατώφλι 1.0 παρουσιάζουν πάρα πολύ υψηλό ***Distictiveness***, καθώς υπάρχουν μόνο 2 κοινότητες με 2 συνεδρίες πρόσβασης η καθεμία, αλλά απελπιστικά χαμηλό ***Coverage*** (μόνο 3 από τις 519 συνεδρίες πρόσβασης χρησιμοποιήθηκαν). Όπως είναι κατανοητό, το εύρος και το μέγεθος των συγκεκριμένων μοντέλων δεν τα καθιστούν αντιπροσωπευτικά του τρόπου χρήσης των επισημειώσεων από τους χρήστες του συστήματος.

Αντιθέτως, τα μοντέλα που παράγονται από τα κατώφλια 0.95 και 0.97 έχουν πολλά κοινά χαρακτηριστικά και παράγουν αρκετές όμοιες κοινότητες επισημειώσεων. Ως αποτέλεσμα, και οι μετρικές των δυο αυτών συνόλων έχουν σχεδόν την ίδια τιμή. Λόγω αυτής της σύγκλισης τιμών, επιλέχθηκε να δοθεί μια μικρή προτεραιότητα στο ***Distictiveness*** έτσι ώστε να επιλεγθούν τα λιγότερα επικαλυπτόμενα μοντέλα που ούτως ή άλλως έχουν σχεδόν το ίδιο ποσοστό κάλυψης των συνολικών επισημειώσεων του πειράματος.

Οπότε, για την συγκεκριμένη μελέτη και βάσει των αποτελεσμάτων και των αναλύσεων που έγιναν, επιλέχθηκε **το σύνολο των μοντέλων κοινοτήτων που δημιουργήθηκαν με ελάχιστο κατώφλι 0.97**, ως το καταλληλότερο για το πείραμα αυτό.

6

Επίλογος

Στο κεφάλαιο αυτό ακολουθεί μια γενική σύνοψη της παρούσας μελέτης και περιγράφονται τα συμπεράσματα που προέκυψαν καθώς και μελλοντικές επεκτάσεις.

6.1 Σύνοψη και συμπεράσματα

Στα πλαίσια της συγκεκριμένης διπλωματικής εργασίας, υλοποιήθηκαν αλγόριθμοι εξόρυξης γνώσης από αρχεία καταγραφής δικτυακών τόπων που παρέχουν υπηρεσίες κοινωνικών επισημειώσεων (social tags), για το σχηματισμό κοινοτήτων χρηστών. Η εκπόνηση της συγκεκριμένης μελέτης απαιτούσε την επεξεργασία αρχείων καταγραφής (log files) ενός εξυπηρετητή του παγκόσμιου ιστού που παρέχει τη δυνατότητα χαρακτηρισμού των πόρων που διαθέτει, μέσω κοινωνικών επισημειώσεων. Στόχος της επεξεργασίας ήταν ο προσδιορισμός ομάδων χρηστών που χρησιμοποιούν κοινές επισημειώσεις για την αναζήτηση και ανάκτηση πόρων, δηλαδή τα μέλη τους παρουσιάζουν παρόμοια συμπεριφορά χρήσης και διαθέτουν κοινά ενδιαφέροντα. Παρέχοντας το κατάλληλο θεωρητικό και τεχνολογικό υπόβαθρο, παρουσιάστηκε και αναλύθηκε η μεθοδολογία που χρησιμοποιήθηκε στις δύο μελέτες περιπτώσεων που παρουσιάστηκαν σε αυτήν την εργασία. Ο κοινός στόχος και στις δυο αυτές μελέτες ήταν η ανάλυση της συμπεριφοράς των χρηστών μιας εφαρμογής με σκοπό την παροχή χρήσιμων πληροφοριών, προκειμένου να βελτιωθεί η εκάστοτε εφαρμογή μέσω της δημιουργίας στοχευμένων υπηρεσιών για τους χρήστες. Το τελικό αποτέλεσμα ήταν η δημιουργία αντιπροσωπευτικών μοντέλων κοινοτήτων με κοινά χαρακτηριστικά για την κάθε μελέτη αντίστοιχα, με την χρήση του cluster mining αλγόριθμου, που αναπαριστούσαν τα κοινά ενδιαφέροντα και προτιμήσεις διαφορετικών τύπων χρηστών. Συγκεκριμένα, δημιουργήθηκαν τρεις διαφορετικές ομάδες μοντέλων κοινοτήτων, καθώς ο cluster mining αλγόριθμος εκτελέστηκε με τρία διαφορετικά ελάχιστα κατώφλια (0.95, 0.97, 1.0) και αξιολογήθηκαν με βάση τις μετρικές Coverage και Distinctiveness, που αναλύθηκαν διεξοδικά στο κεφάλαιο 4. Η επιλογή του καταλληλότερου μοντέλου για την καθεμιά μελέτη ξεχωριστά έγινε βάσει αυτών των μετρικών και τα τελικά αποτελέσματα παρουσιάστηκαν με την μορφή πινάκων.

Συμπερασματικά, η παρούσα μελέτη καθιστά το σχηματισμό μοντέλων κοινοτήτων, με την χρήση αλγορίθμων εξόρυξης γνώσης, ως ένα πολύτιμο εργαλείο για την ανάλυση της συμπεριφοράς των χρηστών σε εφαρμογές στο Διαδίκτυο. Η ανάλυση αυτή αποτελεί το κινητήριο μοχλό για τη δημιουργία ξεχωριστών υπηρεσιών με αυξημένη προστιθέμενη αξία, δημιουργώντας ένα φιλικότερο και πιο «έμπιστο» περιβάλλον προς τον χρήστη. Ολοκληρώνοντας, καθιστάται πλέον σαφές, ότι η προστιθέμενη αξία που έχουν οι επισκέψεις των διαδικτυακών τόπων, δεν έχει αποκτηθεί μέσω των μεγαλύτερων ποσοτήτων δεδομένων που μπορεί να βρει κανείς μέσα σε ένα ιστότοπο, αλλά μέσω της ευκολότερης πρόσβασης στις απαιτούμενες πληροφορίες την κατάλληλη στιγμή και με την πιο κατάλληλη μορφή.

6.2 Μελλοντικές επεκτάσεις

Ο *cluster mining* αλγόριθμος, που παρουσιάστηκε και αναλύθηκε στη συγκεκριμένη μελέτη θεωρείται ένας ευέλικτος και γρήγορος αλγόριθμος που θα μπορούσε να χρησιμοποιηθεί για το σχηματισμό και άλλων τύπων μοντέλων κοινοτήτων. Συγκεκριμένα, θα μπορούσε να γίνει μια μελέτη σχετικά με την χρήση αλγορίθμων εξόρυξης γνώσης για την κατασκευή στερεοτύπων, τα οποία θεωρούνται μοντέλα κοινοτήτων εμπλουτισμένα με προσωπικές πληροφορίες για τον εκάστοτε χρήστη. Η εννοιολογική σχέση μεταξύ των στερεοτύπων και των κοινοτήτων καθιστά αναγκαία τη χρήση διαδικασιών και μεθόδων που χρησιμοποιήθηκαν στην παρούσα μελέτη. Ωστόσο, η απόκτηση των προσωπικών πληροφοριών από υπηρεσίες του Διαδικτύου θεωρείται μια αρκετά δύσκολη και «λεπτή» διαδικασία.

Μια πιο ουσιαστική διαφοροποίηση της παρούσας μελέτης, θα ήταν να επεκταθεί το πεδίο εφαρμογής της μεθολογίας που χρησιμοποιήθηκε και σε άλλους τύπους διαδικτυακών υπηρεσιών. Ένα ενδιαφέρον παράδειγμα θα ήταν η μελέτη των χρηστών ενός πάροχου υπηρεσιών στο Διαδίκτυο (Internet Service Provider), οι οποίοι θα αποτελούν μία λιγότερο ομοιογενή ομάδα από τους χρήστες μιας συγκεκριμένης τοποθεσίας στο Διαδίκτυο. Επιπλέον, θα ήταν αρκετά ενδιαφέρουσα η μοντελοποίηση άλλων παραμέτρων της συμπεριφοράς των χρηστών, πέρα από τα «ενδιαφέροντα» και τις «προτιμήσεις». Για παράδειγμα, σε πολλές e-learning εφαρμογές θα μπορούσαν τα μοντέλα κοινοτήτων να βασιστούν στο επίπεδο γνώσεων των χρηστών. Το γεγονός ότι οι μέθοδοι που παρουσιάζονται στη συγκεκριμένη μελέτη δεν εξαρτώνται από τη σημασιολογία του προβλήματος, τις καθιστούν κατάλληλες για την λύση παρόμοιων προβλημάτων, μέσω της κατάλληλης μετατροπής των δεδομένων χρήσης σε ένα σύνολο χαρακτηριστικών.

Ολοκληρώνοντας, στην παρούσα μελέτη δημιουργήθηκαν μοντέλα κοινοτήτων χρηστών που περιείχαν κοινωνικές επισημειώσεις που χρησιμοποιήθηκαν για την αναζήτηση και ανάκτηση πόρων. Το μέγεθος και το εύρος των κοινοτήτων που προέκυψαν ήταν σχετικά μικρό και μια ενδιαφέρουσα μελέτη θα ήταν η σημασιολογική ανάλυση των επισημειώσεων αυτών, έτσι ώστε να μπορούν να προταθούν στο μέλλον σημασιολογικά συναφείς επισημειώσεις για το σχηματισμό πιο εμπλουτισμένων και αντιπροσωπευτικών μοντέλων κοινοτήτων.

7

Βιβλιογραφία

- [1] Γαβαλάς Δαμιανός (2008), «Δικτυακά Πολυμέσα II: Διάλεξη 11 Web2.0», Πανεπιστήμιο Αιγαίου Τμήμα Πολιτισμικής Τεχνολογίας και Επικοινωνίας.
- [2] Κωνσταντία Κακάλη, Κοινωνικές Ταξονομίες στο Web 2.0, Πανεπιστημιακές διαλέξεις, Οικονομικό Πανεπιστήμιο Αθηνών, Αθήνα 2009.
- [3] Garton, L., Haythornthwaite, C. & Wellman, B. (1997), «Studying online social networks», Journal of Computer-Mediated Communication.
- [4] Boyd, D. M.; Ellison, N.B (2008) «Social network sites: Definition, history, and scholarship», Journal of Computer-Mediated Communication, 1 - 13, Article 11.
- [5] Ying Ding, Elin K. Jacob, James Caverlee, Michael Fried, Zhixiong Zhang, «Profiling Social Networks: A Social Tagging Perspective», D-lib Magazine, Volume 15 Number 3/4, March/April, 2009.
- [6] Marieke Guy, Emma Tonkin, «Folksonomies: Tidying up tags», D-lib Magazine, Volume 12 Number 1, January, 2006.
- [7] Golder, S.A., Huberman, B.A., «Usage patterns of collaborative tagging systems», Journal of Information Science, 32(2):198-208, 2006.
- [8] Shaw B. (2005), «Building a Better Folksonomy: Web-based Aggregation of Metadata», Columbia university. <http://www.metablake.com/webfolk/web-paper.pdf>
- [9] Mathes A. (2004), «Folksonomies – cooperative classification and communication through shared metadata», Computer Mediated Communication. <http://adammathes.com/academic/computer-mediatedcommunication/folksonomies.pdf>
- [10] Vander Wal T. (2004), «Folksonomy». <http://vanderwal.net/folksonomy.html>
- [11] Shirky(2004),«Folksonomy»,Corante. <http://many.corante.com/archives/2004/08/25/folksonomy.php>
- [12] Merholz P. (2004), «Ethnoclassification and vernacular vocabularies». <http://www.peterme.com/archives/000387.htm>
- [13] Kroski, E. (2005), «The Hive Mind: Folksonomies and User-Based Tagging», InfoTangle [blog].
- [14] Stock W. G., Peters, I. (2007), «Folksonomy and Information Retrieval», Joining Research and Practice: Social Computing and Information Science. American

Society for Information Science and Technology.

- [15] Stock W. (2007), «Folksonomies and science communication», Information Services & Use, Department of Information Science, University Dusseldorf.
- [16] Mikaz P. (2005), «Ontologies are us: A Unified Model of Social Networks and Semantics», 4th International Semantic Web Conference (ISWC 2005), 522-536.
- [17] Noruzi A. (2006), «Folksonomies: (Un) Controlled Vocabulary?», University of Tehran, Department of Library and Information Science, Tehran, Iran.
- [18] Chakrabarti, S. (2000), «Data mining for hypertext: A tutorial survey» ACM SIGKDD Explorations, 1(2), 1 - 11.
- [19] Mobasher, B., Cooley, R. and Srivastava, J.: 1999a, Automatic personalization based on Web usage mining. TR-99010, Department of Computer Science. DePaul University.
- [20] Kobsa, A., Koenemann, J. and Pohl, W.: 2001, Personalized Hypermedia Presentation Techniques for Improving Online Customer Relationships, The Knowledge Engineering Review 16(2), 111-155.
- [21] Mobasher, B., Cooley, R. and Srivastava, J.: 2000b, Automatic personalization based on Web usage mining, Communications of the ACM, 43(8), 142 - 151.
- [22] Bouganis C., Koukopoulos, D. and Kalles, D.: 1999, A Real Time Auction System over the WWW, Conference on Communication Networks and Distributed Systems Modeling and Simulation, San Francisco, CA, USA, 1999.
- [23] Pierrakos, D., Paliouras, G., Papatheodorou, C., Spyropoulos, C. (2003). Web Usage Mining as a Tool for Personalization: A Survey. J. User Modeling and User Adapted Interaction, 13, 311-372, Kluwer Academic Publishers: Netherlands.
- [24] Manber, U., Patel, A. and Robison, J.: 2000, Experience with Personalization on Yahoo, Communications of the ACM, 43(8), 35 - 39.
- [25] Cooley, R., Srivastava, J. and Mobasher, B.: 1997b, Web Mining: Information and Pattern Discovery on the World Wide Web, In : Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), p p. 558 - 567.
- [26] Kosala, R. and Blockeel, H.: 2000, WebMining Research: A Survey, SIGKDD Explorations, 2(1), 1-15.
- [27] Brajnik, G., Tasso, C., 1994. A shell for developing non-monotonic user modelling systems. International Journal of Human-Computer Studies 40, 31-62.
- [28] Balabanovic, M., Shoham, Y., 1997. Content-based, collaborative recommendation. Communications of the ACM 4 (3), 66-72. Basu, C., Hirsh, H., Cohen, W., 1998. Recommendation.
- [29] Benaki, E., Karkaletsis, V., Spyropoulos, C.D., 1997. Integrating user modelling into information extraction: the umie prototype. Proceedings Sixth International Conference on User Modeling (UM'97), 55-57.
- [30] Langley, P., 1999. User modelling in adaptive interfaces. Proceedings Seventh International Conference on User Modelling (UM'99), Banff, Canada, 357-370.
- [31] Perkowitz, M., Etzioni, O., 1998. Adaptive web sites: automatically synthesizing web pages. Proceedings Fifteen National Conference in Artificial Intelligence (AAAI 98).
- [32] Rich, E., 1983. Users are individuals: individualizing user models. International Journal of Man-Machine Studies 18, 199-214.
- [33] Esposito, F., Malerba, D., Semeraro, G., Fanizzi, N., Ferilli, S., 1998. Adding machine learning and knowledge intensive techniques to a digital library service. International Journal on Digital Libraries 2, 3-19.
- [34] Bloedorn, E., Mani, I., MacMillan, T.R., 1996. Machine learning of user profiles:

- representational issues. Proceedings Thirteen National Conference on Artificial Intelligence (AAAI'96), 433–438.
- [35] Joachims, T., Freitag, D., Mitchell, T., 1997. WebWatcher: a tour guide for the World Wide Web. Proceedings Fifteenth International Joint Conference in Artificial Intelligence (IJCAI'97).
 - [36] Moukas, A., 1997. Amalthaea: information discovery and filtering using a multiagent evolving ecosystem. Applied Artificial Intelligence: An International Journal 11 (5), 437–457.
 - [37] Billsus, D., Pazzani, M., 1999. A hybrid user model for news story classification. Proceedings Seventh International Conference on User Modelling (UM'99), 99–108
 - [38] Cooley, R., 2000. Web usage mining: discovery and applications of interesting patterns from web data. PhD Thesis, University of Minnesota, Department of Computer Science and Engineering.
 - [39] Paliouras, G., Papatheodorou, C., Karkaletsis, V., Spyropoulos, C.D., Tzitziras, P., 1999. From web usage statistics to web usage analysis. Proceedings IEEE International Conference on Systems Man and Cybernetics, II-159–II-164.
 - [40] Paliouras, G., Papatheodorou, C., Karkaletsis, V., Spyropoulos, 2002, C.D.: Discovering User Communities on the Internet using Unsupervised Machine Learning Techniques., Interacting with Computers Journal, 14,6, 761-791.
 - [41] Perkowitz, M., Etzioni, O., 1998. Adaptive web sites: automatically synthesizing web pages. Proceedings Fifteen National Conference in Artificial Intelligence (AAAI 98).
 - [42] Bron, C., Kerbosch, J., 1973. Finding all cliques of an undirected graph. Communications of the ACM 16 (9), 575–577.
 - [43] Paliouras, G., C. Papatheodorou, V. Karkaletsis, C.D. Spyropoulos. 2000. Clustering the users of large web sites into communities. *Proc. of the Internat. Conf. on Machine Learning*, Stanford, CA, 719–726.