

ΤΕΙ ΠΑΤΡΑΣ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΥ ΣΧΕΔΙΑΣΜΟΥ & ΠΛΗΡΟΦΟΡΙΑΚΩΝ
ΣΥΣΤΗΜΑΤΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΤΕΧΝΟΛΟΓΙΕΣ ΓΙΑ ΕΞΑΤΟΜΙΚΕΥΜΕΝΕΣ ΥΠΗΡΕΣΙΕΣ ΣΤΟ
ΔΙΑΔΙΚΤΥΟ

ΟΝΟΜΑΤΕΠΩΝΥΜΟ ΣΠΟΥΔΑΣΤΡΙΩΝ

ΠΑΠΑΔΟΠΟΥΛΟΥ ΧΡΥΣΟΥΛΑ
ΠΑΛΛΑ ΑΘΑΝΑΣΙΑ
ΝΙΚΑ ΜΑΡΙΝΑ

ΕΠΟΠΤΕΥΩΝ ΚΑΘΗΓΗΤΗΣ

ΜΩΥΣΙΔΗΣ ΔΗΜΗΤΡΙΟΣ

ΠΑΤΡΑ - ΔΕΚΕΜΒΡΙΟΣ 2008

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΕΙΣΑΓΩΓΗ	1
ΚΕΦΑΛΑΙΟ 1	8
ΕΞΟΥΣΙΑ ΔΕΔΟΜΕΝΩΝ ΙΣΤΟΥ ΚΑΙ ΕΞΑΤΟΜΙΚΕΥΣΗ ΣΤΟ ΔΙΑΔΙΚΤΥΟ	
1.1 Κατηγορίες Εξόρυξης Δεδομένων στο Διαδίκτυο	8
1.1.1 Εξόρυξη Δομής Ιστού.....	10
1.1.2 Εξόρυξη Περιεχομένου Ιστού.....	11
1.1.3 Εξόρυξη Χρήσης Ιστού	11
1.2 Τι είναι Εξατομίκευση	12
1.3 Δεδομένα Εισόδου.....	13
1.4 Εξόρυξη Χρήσης Ιστού & Εξατομίκευση	14
1.4.1 Εξατομίκευση βασισμένη σε Εξόρυξη Χρήσης Ιστού	15
ΚΕΦΑΛΑΙΟ 2	20
Ο ΣΗΜΑΣΙΟΛΟΓΙΚΟΣ ΠΑΓΚΟΣΜΙΟΣ ΙΣΤΟΣ ΠΛΗΡΟΦΟΡΙΩΝ	
2.1 Γενικά για το Semantic Web	20
2.2 Η Αρχιτεκτονική του Σημασιολογικού Παγκόσμιου Ιστού	24
2.2.1 Ο ορισμός της Οντολογίας	29
2.2.2 Ταξινόμηση των Οντολογιών	30
2.3 Από το Web Mining στο Semantic Web Mining	33
2.3.1 Web Content Mining	34
2.3.2 Web Structure Mining	35
2.3.3 Web Usage Mining	36
2.4 Αξιολόγηση Συστημάτων Προσωπικοποίησης.....	39
ΚΕΦΑΛΑΙΟ 3	46
ΕΦΑΡΜΟΓΕΣ ΕΞΑΤΟΜΙΚΕΥΣΗΣ ΣΤΙΣ ΗΛΕΚΤΡΟΝΙΚΕΣ ΜΑΘΗΣΙΑΚΕΣ ΚΟΙΝΟΤΗΤΕΣ	
3.1 Συσχέτιση Προτύπων E-learning	46
3.2 Η Διαδικασία της Εξατομίκευσης σε ένα Τυπικό E-learning Σύστημα	49
3.3 Η E-learning Οντολογία.....	51

3.4 Διαδικασία Εξόρυξης Δεδομένων σε E- learning Περιβάλλοντα.....	56
3.4.1 Συλλογή των Δεδομένων	56
3.4.2 Καθαρισμός των Δεδομένων & Μοντελοποίηση του Χρήστη ...	62
3.4.3 Ανάλυση των Δεδομένων.....	63
<i>Κατηγοριοποίηση(Classification)</i>	63
<i>Συσταδοποίηση(Clustering)</i>	64
<i>Κανόνες Συσχέτισης (Association Rules)</i>	64
3.4.4 Ολοκλήρωση των Δεδομένων	65
ΚΕΦΑΛΑΙΟ 4.....	67
ΕΦΑΡΜΟΓΕΣ ΕΞΟΡΥΞΗΣ ΧΡΗΣΗΣ & ΕΞΑΤΟΜΙΚΕΥΣΗΣ	
4.1 Εξατομίκευση στο Περιβάλλον του Web.....	67
4.2 Παραγωγή Εξατομίκευσης	69
4.2.1 Ταξινόμηση μεθόδων Εξατομίκευσης	70
4.2.2 Προεπεξεργασία των Δεδομένων	71
4.2.3 Εξατομικευμένη Έξοδος	76
4.3 Παραδείγματα Εξατομίκευσης	77
4.3.1 Ηλεκτρονικές Βιβλιοθήκες	83
4.3.2 Τομείς Λειτουργίας Ηλεκτρονικών Βιβλιοθηκών.....	84
ΚΕΦΑΛΑΙΟ 5.....	87
ΕΡΓΑΛΕΙΑ ΕΞΑΤΟΜΙΚΕΥΜΕΝΩΝ ΥΠΗΡΕΣΙΩΝ ΣΤΟ ΔΙΑΔΙΚΤΥΟ	
5.1 Το Yahoo	90
5.1.1 Το προφίλ της Yahoo.....	96
5.2 Το Google	101
5.2.1 Το προφίλ του Google	106
5.3 Το MSN.....	111
5.4 Το AOL.....	115
5.5 Μεταμηχανές Αναζήτησης & Βάσεις Δεδομένων.....	116
5.5.1 Το Πρόγραμμα Κοπέρνικος	119
5.5.2 Λειτουργία Βάσεων Δεδομένων.....	121
5.6 Παραδείγματα Case Studies	123
5.6.1 Το Amazon & άλλα παραδείγματα Δικτυακής Αναζήτησης ...	125
5.6.2 Πρότυπη Πλατφόρμα Ηλεκτρονικού Καταστήματος.....	137

ΚΕΦΑΛΑΙΟ 6.....	141
ΣΥΜΠΕΡΑΣΜΑΤΑ & ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ	
ΚΕΦΑΛΑΙΟ 7	145
ΒΙΒΛΙΟΓΡΑΦΙΑ	
ΚΕΦΑΛΑΙΟ 8.....	151
ΠΑΡΑΡΤΗΜΑ	
<i>Λίστα Σχημάτων.....</i>	<i>151</i>
<i>Λίστα Πινάκων.....</i>	<i>152</i>
<i>Λίστα εικόνων.....</i>	<i>152</i>
<i>Γλωσσάρι Όρων.....</i>	<i>154</i>

ΕΙΣΑΓΩΓΗ

Η ευρεία χρήση και η ταχύτατη εξάπλωση του Παγκόσμιου Ιστού Πληροφοριών (World Wide Web ή Web) είχε σαν αποτέλεσμα τη μετατροπή του σε μια τεράστια αποθήκη πληροφοριών με δισεκατομμύρια σελίδες και περισσότερους από 800 εκατομμύρια χρήστες παγκοσμίως (Internet World Stats, 2005). Σήμερα θεωρείται ένα από τα πιο σημαντικά μέσα συλλογής, διαμοίρασης και διάδοσης πληροφοριών και υπηρεσιών.

Την ίδια όμως στιγμή ο όγκος των διαθέσιμων πληροφοριών προκαλεί αρκετά προβλήματα που σχετίζονται με τη συνεχώς αυξανόμενη δυσκολία αναζήτησης, εύρεσης, οργάνωσης, πρόσβασης και συντήρησης της αιτούμενης πληροφορίας από τους χρήστες. Αυτό το γεγονός έχει σημαντικές επιπτώσεις στον τρόπο που σχεδιάζουμε και υλοποιούμε εφαρμογές στο Web και τα συστήματα ηλεκτρονικής μάθησης (e-learning), ηλεκτρονικού εμπορίου (e-commerce) και επιχειρηματικής ευφυΐας (business intelligence) δεν θα μπορούσαν να αποτελέσουν εξαίρεση.

Μια λύση στο παραπάνω πρόβλημα έρχεται από την επιστημονική περιοχή του Data Knowledge Mining (Εξόρυξη Γνώσης από Δεδομένα) ή απλώς Data Mining (Εξόρυξη Δεδομένων) (Hand et al: 2001,pp.26-44).

Πράγματι η εξόρυξη γνώσης από μεγάλες αποθήκες δεδομένων έχει εξελιχθεί σε ένα από τα βασικότερα ερευνητικά ζητήματα. Μέχρι πρόσφατα η εξόρυξη γνώσης αφορούσε αποκλειστικά δομημένα (structured) δεδομένα δηλαδή δεδομένα που είναι αποθηκευμένα σε βάσεις δεδομένων (databases). Τα τελευταία χρόνια το ενδιαφέρον στράφηκε και σε μη δομημένα δεδομένα (semi-structured και unstructured) π.χ. κείμενα, εικόνες, ιστοσελίδες (web pages), κλπ. Αυτό οδήγησε στη δημιουργία νέων κλάδων όπως το Text Mining (Εξόρυξη Γνώσης από Κείμενα) και το Web Mining (Εξόρυξη Γνώσης από Δεδομένα του Web)(Hand et al: 2001,pp.26-44).

Με το συνδυασμό και την ολοκλήρωση τεχνικών και τεχνολογιών εξόρυξης γνώσης, οι e-εφαρμογές θα βελτιώσουν την ποιότητα των παρεχόμενων υπηρεσιών τους προς τους τελικούς χρήστες.

Συγκεκριμένα, για να το επιτύχουν αυτό μελετούν τη συμπεριφορά των χρηστών και προσπαθούν να εξάγουν χρήσιμα συμπεράσματα από αυτή. Η απαίτηση για πρόβλεψη των αναγκών των χρηστών με στόχο την ευκολία χρήσης αλλά και τη διατήρηση των χρηστών ενός δικτυακού τόπου μπορεί να ικανοποιηθεί με την εφαρμογή της εξατομίκευσης ή personalization.

Ορίζουμε σαν web personalization οποιαδήποτε ενέργεια που προσαρμόζει τις πληροφορίες ή τις υπηρεσίες που παρέχονται από ένα δικτυακό τόπο στις ανάγκες ενός συγκεκριμένου χρήστη ή ενός συνόλου χρηστών, χρησιμοποιώντας τη γνώση που πηγάζει από την ανάλυση της συμπεριφοράς πλοήγησης των χρηστών και από τα ενδιαφέροντά τους, σε συνδυασμό με τη δομή και το περιεχόμενο του δικτυακού τόπου.

Στόχος είναι η παροχή στους χρήστες των πληροφοριών που θέλουν ή χρειάζονται χωρίς να απαιτεί από αυτούς να τις ζητήσουν με ρητό τρόπο. Η διαδικασία της εξατομίκευσης περιλαμβάνει την κατηγοριοποίηση και την προεπεξεργασία των δεδομένων, την εξαγωγή συσχετίσεων μεταξύ των δεδομένων και τον καθορισμό των ενεργειών που θα προταθούν (Kevin R:2006,pp.24-28)

Ο όρος e-learning αφορά σε όλες τις μορφές εκπαίδευσης που βασίζονται στο Web και χρησιμοποιεί υπολογιστές και δίκτυα υπολογιστών για τη δημιουργία, την παράδοση, τη διαχείριση και την υποστήριξη online σειρών μαθημάτων (Germa B:2006,pp.187-194)

Τελευταία, χάρη στην προαναφερθείσα έκρηξη του Web αλλά και στην ευρεία χρήση των νέων Τεχνολογιών Πληροφορίας και Επικοινωνιών (ΤΠΕ), η έρευνα στο e-learning έχει κερδίσει όλο και περισσότερη προσοχή από τις επιστημονικές και όχι μόνο κοινότητες.

(Urdan T.A & Weggen C:2000).

Διάφοροι εκπαιδευτικοί και εμπορικοί οργανισμοί καταδεικνύουν ένα συνεχές ενδιαφέρον για την περιοχή, κίνηση που αποτελεί μια πολύ ισχυρή κατευθυντήρια δύναμη, εκφρασμένη μέσα από πολυάριθμη έρευνα και εμπορικές προσπάθειες (Wentling et al:2000). Η μεγάλη ποικιλία των διαθέσιμων συστημάτων και εφαρμογών e-learning αποτελεί άλλωστε μια ισχυρή ένδειξη της ωριμότητας στην περιοχή.

Τα πιο γνωστά εκπαιδευτικά περιβάλλοντα είναι: το webCT (<http://www.webct.com>), το Lotus Learning Space (<http://www.lotus.com>), το Librarian(<http://www.click2learn.com>), το Blackboard (<http://www.blackboard.net>), το TopClass(<http://www.wbtsystems.com>), το Intralearn (<http://www.intralearn.com>), το Ecollege (<http://www.ecollege.com>), το Embanet (<http://www.embanet.com>), το eduprise (<http://www.eduprise.com>), κλπ.

Η πραγματική δύναμη του e-learning δεν είναι το γεγονός ότι επιτρέπει στον καθένα, οπουδήποτε και οποτεδήποτε να αποκτήσει πρόσβαση στη γνώση αλλά η ικανότητά του να παρέχει την κατάλληλη πληροφορία, στους κατάλληλους ανθρώπους, την κατάλληλη χρονική στιγμή (Brusilovsky, P:1996,pp.87-129)

Στο μέλλον μάλιστα προβλέπεται ότι ο εκπαιδευόμενος θα μπορεί να ελέγχει και να προσαρμόζει σε πολύ μεγάλο βαθμό την προσωπική του εκπαιδευτική εμπειρία βελτιώνοντας παράλληλα και την απόδοσή του (Brusilovsky et al:1998,pp.87-129).

Εντούτοις, στην πλειοψηφία των e-learning συστημάτων του παρελθόντος, οι σειρές των μαθημάτων και το εκπαιδευτικό υλικό δεν ήταν αρκετά δυναμικά, παρέχοντας ένα μάλλον περιορισμένο σύνολο χαρακτηριστικών, με πολύπλοκη δομή και δυσκολίες στην πλοήγηση με αποτέλεσμα να μη μπορούν να ανταποκριθούν αποδοτικά στις ανάγκες και στις απαιτήσεις των χρηστών, οδηγώντας σε «φτωχές» εκπαιδευτικές εμπειρίες.

Γενικά, το εκπαιδευτικό υλικό με ενσωματωμένους υπερσυνδέσμους (hyperlinks) επιτρέπει στους εκπαιδευόμενους να ακολουθήσουν οποιοδήποτε μονοπάτι πλοήγησης επιθυμούν και να μη χρησιμοποιήσουν απαραίτητα τη δομή που έχουν καθορίσει οι σχεδιαστές των δικτυακών τόπων ή οι δημιουργοί του περιεχομένου (οι οποίοι έχουν ένα ορισμένο σχέδιο πλοήγησης στο μυαλό τους) (Blom, J:2000,pp.313-314)

Αυτή η ελευθερία μπορεί να αποδειχτεί αρνητικός παράγοντας, δεδομένου ότι σε πολλές περιπτώσεις οι εκπαιδευόμενοι δεν έχουν την απαραίτητη ωριμότητα και ικανότητα να ακολουθήσουν την κατάλληλη πορεία και συχνά οδηγούνται σε θεματικά που είναι είτε πάρα πολύ δύσκολα, είτε πάρα πολύ εύκολα, ή ακόμα και άσχετα με τις προσωπικές τους εκπαιδευτικές ανάγκες (Markellos et al:2004).

Μια λύση στο πρόβλημα αυτό αποτελεί η ενσωμάτωση προσαρμοστικών χαρακτηριστικών (adaptive features) που επιτρέπουν την παροχή εξατομικευμένης μάθησης (personalized learning) (Brusilovsky, 1996), (Brusilovsky, 1998).

Αυτά τα συστήματα είναι η απάντηση στην παραδοσιακή «one-size-fits-all» προσέγγιση που παρέχει το ίδιο στατικό υλικό εκπαίδευσης σε όλους τους χρήστες, χωρίς να λαμβάνει υπόψη τις ανάγκες και τις προτιμήσεις τους, τις εμπειρίες και τις γνώσεις τους, οι οποίες μπορεί να ποικίλουν εντυπωσιακά (De Bra et al:2000,pp.5-11).

Τα προσαρμοστικά εκπαιδευτικά συστήματα που βασίζονται στο Web (adaptive web-based educational systems) επιτρέπουν στις σειρές των μαθημάτων να δημιουργούνται, ικανοποιώντας τις ανάγκες των εκπαιδευόμενων στα διάφορα γνωστικά αντικείμενα χωρίς να σπαταλούν το χρόνο τους σε περιοχές που ήδη κατέχουν ή που δεν τους ενδιαφέρουν (Clarke, T:1994).

Για να επιτευχθεί αυτός ο στόχος είναι απαραίτητη η γνώση του προφίλ (profile) του κάθε χρήστη που περιλαμβάνει πληροφορίες για την ταυτότητά του, τις προτιμήσεις του, τα

ενδιαφέροντά του, τις δεξιότητές του, το εκπαιδευτικό υπόβαθρό του, κλπ. Συνήθως για τη συλλογή αυτών των πληροφοριών απαιτείται η συμπλήρωση εξαντλητικών ερωτηματολογίων ή φορμών καθώς και η καταγραφή της πλοηγητικής συμπεριφοράς των χρηστών κατά τη χρήση του e-learning συστήματος (Clarke T and Pitelis C:1995,pp.100-102).

Με αυτό τον τρόπο οι εφαρμογές e-learning παρέχουν υψηλής ποιότητας περιεχόμενο, αποδοτική δομή, εύκολη πλοήγηση καθώς επίσης και πλήρη υποστήριξη για τις διαφορετικές εργασίες όλων των προφίλ χρηστών που συμμετέχουν σε ένα τυπικό σενάριο εκπαίδευσης από απόσταση (Meyen, E.L:2000).

Συγκεκριμένα, ανάλογα με το γνωστικό υπόβαθρο του εκπαιδευομένου, τις γνώσεις και τις αδυναμίες του, το εκπαιδευτικό στυλ που προτιμά, την πρόοδο που έχει σημειώσει μέχρι τώρα, το σύστημα αποφασίζει τι εκπαιδευτικό υλικό θα του παρουσιάσει και με ποιον τρόπο (Brusilovsky:2001).

Πιθανές παράμετροι αποτελούν τα διαφορετικά εκπαιδευτικά μονοπάτια, οι διαφορετικοί τρόποι παρουσίασης του ίδιου περιεχομένου (π.χ. με ή χωρίς ήχο) ή η προσφορά ενός διαφορετικού συνόλου λειτουργιών που το περιβάλλον διεπαφής (user interface) του συστήματος εκπαίδευσης μπορεί να παρέχει για να μειώσει την πολυπλοκότητα (Gotschall, M:2000).

Τα τελευταία χρόνια, ένας σημαντικός αριθμός από ερευνητικά αλλά και εμπορικά e-learning περιβάλλοντα που ενσωματώνουν προσαρμοστικά χαρακτηριστικά έχουν σχεδιαστεί και αναπτυχθεί. Αυτά τα συστήματα συλλέγουν «γνώση» για κάθε χρήστη, την αναλύουν και έπειτα τη «μετατρέπουν» σε εξατομικευμένη εμπειρία μάθησης.

Παρέχουν δηλαδή προσωποποιημένες (personalized), αλληλεπιδραστικές (interactive), ανά πάσα στιγμή (just-in-time), άμεσες (current) και κεντρικοποιημένες προς το χρήστη

(user-centric) υπηρεσίες(Neumann & Zirvas, 1998,pp.688-693).

Τα συστήματα αυτά περιλαμβάνουν όλες τις λειτουργικότητες μιας σειράς μαθημάτων και παράλληλα μπορούν και διαφοροποιούνται κατά τη διάρκεια χρήσης τους ώστε να προσαρμόζονται στις απαιτήσεις των χρηστών και να γίνονται περισσότερο αποδοτικά(Brusilovsky et al:1998).

Στα πιο αντιπροσωπευτικά ανήκουν το InterBook το οποίο εστιάζει στην παροχή προσαρμοστικής πλοήγησης σε ένα e-learning σύστημα μέσω τεχνικών σήμανσης υπερσυνδέσμων, ενώ το AHA (Adaptive Hypermedia Architecture) χρησιμοποιεί την τεχνική απόκρυψης υπερσυνδέσμων (De Bra & Calvi:1998).

Το NetCoach προέκυψε από το ELM-ART, ένα από τα πρώτα web-based προσαρμοστικά e-learning συστήματα και επιτρέπει στους συγγραφείς να δημιουργούν τα εξατομικευμένα μαθήματα χωρίς να απαιτεί ιδιαίτερες προγραμματιστικές γνώσεις. Το WebPersonalizer (Mobasher et al., 2000) είναι ένα σύστημα γενικότερου σκοπού το οποίο παρέχει μια λίστα από προτεινόμενους υπερσυνδέσμους στο χρήστη καθώς αυτός πλοηγείται σε ένα web site(Weber G:1999, pp.371-378).

Τέλος, το INSPIRE (Intelligent System for Personalized Instruction in a Remote Environment) δημιουργεί μαθήματα με δυναμικό τρόπο τα οποία καλύπτουν σταδιακά το γνωστικό στόχο που έχει επιλέξει ο εκπαιδευόμενος ακολουθώντας το γνωστικό του επίπεδο, το στυλ μάθησής του αλλά και την εξέλιξή του (Papanikolaou K:2003,pp.213-267).

Ένα από τα ανοικτά προβλήματα της περιοχής σχετίζεται με το πώς θα συνδυαστούν οι κατάλληλες πηγές στοιχείων, θα βρεθούν οι σωστές και χρήσιμες πληροφορίες κάθε φορά και θα γίνει κατανοητή η συμπεριφορά του χρήστη, ώστε να υποστηριχθεί μια ικανοποιητική και εξατομικευμένη ηλεκτρονική μάθηση. Αυτό είχε σαν αποτέλεσμα πολλοί ερευνητές να στραφούν στην επιστημονική περιοχή του Data Knowledge Mining (Εξόρυξη

Γνώσης από Δεδομένα) ή απλώς Data Mining (Εξόρυξη Δεδομένων) (Hand et al:2001).

Τα τελευταία χρόνια το ενδιαφέρον στράφηκε στο Web Mining (Εξόρυξη Γνώσης από Δεδομένα του Web). Αυτός ο κλάδος αποτελεί ένα ελπιδοφόρο πεδίο ερευνητικής δραστηριότητας που θα μπορούσε να εκπληρώσει τους παραπάνω στόχους.

Επιπλέον η πρόσφατη εμφάνιση του Σημασιολογικού Παγκόσμιου Ιστού Πληροφοριών στοχεύει στην ενσωμάτωση ενός ακόμα επιπέδου ευφυΐας (intelligence) στο περιβάλλον των διαδικτυακών εφαρμογών. Σύμφωνα με τους Berners-Lee et al. (2001):

«Το Σημασιολογικό δίκτυο αποτελεί τη προέκταση του υπάρχοντος διαδικτύου στόχος του είναι η καλύτερη διάδοση της πληροφόρησης μέσα από την ευρύτερη συνεργασία των χρηστών και την εύρυθμη ανταλλαγή πληροφοριών μεταξύ τους.»

Ένας πιο επίσημος ορισμός από το W3C (2005) αναφέρει ότι: *« Το Σημασιολογικό δίκτυο είναι η αναπαράσταση και αναπαραγωγή των πληροφοριών στο διαδίκτυο. Αποτελεί ουσιαστικά μια διαδικασία η οποία κατευθύνεται από έναν αριθμό μελετητών αλλά και βιομηχανικών και καταναλωτικών χρηστών. Η μορφή αυτή περιλαμβάνει εκατοντάδες εφαρμογές αλλά και διαφορετικές γλώσσες προγραμματισμού βάση των οποίων αναπτύσσεται και λειτουργεί.»*

Η ικανότητα του Σημασιολογικού Παγκόσμιου Ιστού Πληροφοριών να προσθέτει νόημα στις πληροφορίες, αποθηκευμένο με τέτοιο τρόπο που να μπορεί να αναζητηθεί και να υποβληθεί σε επεξεργασία, παρέχει πολλές ευκαιρίες επέκτασης για τις web-based εφαρμογές (Kondo, H:2004,pp. 35-44). Ο συνδυασμός του με το Web Mining έχει δημιουργήσει μια νέα και γρήγορα-αναδυόμενη ερευνητική περιοχή που είναι γνωστή ως Semantic Web Mining. Η ιδέα πίσω από τη χρησιμοποίησή του στη δημιουργία εξατομικευμένων (personalized) web εμπειριών είναι να βελτιωθεί το Web Mining με την εκμετάλλευση των νέων σημασιολογικών δομών.

ΚΕΦΑΛΑΙΟ 1

ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΙΣΤΟΥ ΚΑΙ ΕΞΑΤΟΜΙΚΕΥΣΗ ΣΤΟ ΔΙΑΔΙΚΤΥΟ

1.1 ΚΑΤΗΓΟΡΙΕΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΣΤΟ ΔΙΑΔΙΚΤΥΟ

Ο παγκόσμιος ιστός (World Wide Web) στις μέρες μας αποτελεί ένα δημοφιλές και αλληλεπιδραστικό μέσο που είναι ιδανικό για τη δημοσίευση πληροφοριών. Έχει αναπτυχθεί τα τελευταία χρόνια από μια μικρή ερευνητική κοινότητα στο μεγαλύτερο και δημοφιλέστερο τρόπο επικοινωνίας και διάδοσης πληροφοριών. Η ανάπτυξη του ιστού κάθε μέρα είναι περίπου ένα εκατομμύριο ηλεκτρονικές σελίδες που προστίθενται σε εκατοντάδες εκατομμυρίων που είναι ήδη online. Η συνεχής αύξηση του μεγέθους και της χρήσης του ιστού επιβάλλει νέες μεθόδους για την επεξεργασία αυτής της τεράστιας ποσότητας των δεδομένων. Εξαιτίας αυτής της ταχύτητας και χαοτικής ανάπτυξης, το αποτέλεσμα είναι πως η πληροφορία στο δίκτυο είναι ανοργάνωτη και μη δομημένη.

Η έρευνα στην εξόρυξη ιστού είναι μια περιοχή που συγκλίνουν διάφορες ερευνητικές κοινότητες όπως είναι οι βάσεις δεδομένων, η ανάκτηση πληροφοριών και η μη γραμμική πρόβλεψη. Παρόλο που σχετίζεται με την εξόρυξη δεδομένων, η εξόρυξη ιστού δεν είναι ισοδύναμη μιας και επεξεργάζεται δεδομένα ιστού σε διαφορετικές κατηγορίες και μορφές κάτι το οποίο καθιστά πολύ βασικό το συνδυασμό πολλών τεχνικών που εφαρμόζονται στις ερευνητικές περιοχές που αναφέρουμε.

Η απόκτηση πληροφοριών από τον ιστό απαιτεί τεχνικές και μεθόδους που προκύπτουν από αυτές τις διάφορες ερευνητικές περιοχές. Η εξόρυξη ιστού χρησιμοποιεί τεχνικές εξόρυξης δεδομένων για την αυτόματη ανακάλυψη και εξαγωγή πληροφοριών από κείμενα και υπηρεσίες ιστού.

Υπάρχουν τρεις βασικές κατηγορίες εξόρυξης ιστού σύμφωνα με τα δεδομένα ιστού που χρησιμοποιούνται σαν είσοδος στην διεργασία της εξόρυξης δεδομένων και είναι η εξόρυξη δομής ιστού (web structure mining), η εξόρυξη περιεχομένου ιστού (web content mining) και η εξόρυξη ιστού χρήσης (web usage mining).

Ο στόχος της εξόρυξης δομής ιστού είναι η κατηγοριοποίηση των ιστοσελίδων και η εξαγωγή πληροφοριών όπως είναι οι ομοιότητες και οι σχέσεις μεταξύ τους, λαμβάνοντας υπόψη την τυπολογία των υπερσυνδέσεων.

Η εξόρυξη περιεχομένου ιστού σχετίζεται με την ανάκτηση πληροφοριών, που είναι διαθέσιμες στον ιστό, σε πιο δομημένες μορφές. Το περιεχόμενο μπορεί να είναι μη δομημένο (καθαρό κείμενο), ημιδομημένο (html κείμενο) ή δομημένο (δυναμικές ιστοσελίδες που εξήχθησαν από μια βάση δεδομένων). Η εξόρυξη κειμένου (text mining) είναι μια έννοια που σχετίζεται με την εξόρυξη περιεχομένου.

Η εξόρυξη χρήσης ιστού είναι η διεργασία του προσδιορισμού προτύπων πλοήγησης (browsing patterns) για την ανάλυση της συμπεριφοράς των χρηστών όταν πλοηγούνται στον ιστό. Σαν πηγές χρησιμοποιούνται τα δεδομένα χρήσης όπως είναι αυτά των καταλόγων ενός εξυπηρετητή ιστού (web server logs) που εγγράφουν τις επισκέψεις των χρηστών σε έναν δικτυακό τόπο.

Η εκτεταμένη έρευνα στην περιοχή αυτή οδήγησε στην εμφάνιση άλλων σχετικών ερευνητικών περιοχών όπως είναι η εξατομίκευση ιστού (web personalization), η οποία χρησιμοποιεί τα αποτελέσματα χρήσης προκειμένου να παρέχει κάποιες προτάσεις ή συμβουλές σε κάθε χρήστη.

Αν και υπάρχουν τρεις διαφορετικές περιοχές έρευνας στην εξόρυξη ιστού, οι διαφορές τους δεν είναι εύκολα διακριτές. Η εξόρυξη περιεχομένου μπορεί να χρησιμοποιήσει το κείμενο που περιέχεται σε υπερσυνδέσεις ή και το κείμενο που βρίσκεται γύρω από αυτές.

Από την άλλη πλευρά η εξόρυξη δομής μπορεί να έχει την ίδια πληροφορία για διαφορετικούς σκοπούς. Επίσης η εξόρυξη χρήσης δύναται να βελτιωθεί εάν ληφθούν υπόψη τα σημασιολογικά του περιεχομένου (content semantics).

Οι περισσότερες από τις ερευνητικές προσπάθειες σήμερα προτείνουν συστήματα που συνδυάζουν μεθόδους από τις τρεις ανωτέρω ερευνητικές περιοχές. Αυτό έχει σαν αποτέλεσμα να μετακινείται η εξόρυξη ιστού σε ένα πιο αφηρημένο επίπεδο όπου η αναπαράσταση των δεδομένων γίνεται με σημασιολογικά (semantics). Τα σημασιολογικά ορίζονται μέσω εργαλείων που προκύπτουν στο πλαίσιο του σημασιολογικού ιστού (semantic web) όπως είναι οι XML, RDF και οι οντολογίες (ontologies), και για τα οποία θα αναφερθούμε εκτενέστερα σε επόμενες ενότητες.

1.1.1 Εξόρυξη Δομής Ιστού

Αυτό που διαφοροποιεί τον παγκόσμιο ιστό από άλλες συλλογές κειμένων είναι το γεγονός ότι χαρακτηρίζεται επιπλέον από τις υπερσυνδέσεις ανάμεσα στα κείμενα που τα διασυνδέουν. Επομένως ο ιστός θεωρείται σαν ένας γράφος του οποίου οι κόμβοι είναι τα κείμενα και οι πλευρές του θεωρούνται οι υπερσυνδέσεις ανάμεσά τους.

Αρκετοί ερευνητές προσπάθησαν να αναλύσουν τις ιδιότητες ενός τέτοιου γράφου. Σε μια μελέτη υποστηρίζεται ότι η δομή του γράφου ιστού ομοιάζει με μια γιγαντιαία γραβάτα. Πολύ επιγραμματικά η γραβάτα αυτή περιέχει έναν δυνατά συνδεδεμένο πυρήνα (strongly connected core –SCC) από 56 εκατομμύρια σελίδες στο κέντρο και άλλα δυο μέρη με 44 εκατομμύρια σελίδες στα άκρα , που το ένα περιέχει σελίδες που δείχνουν στο SCC (σύνολο IN) και το άλλο που δείχνονται από το SCC (σύνολο OUT). Επιπλέον υπάρχουν κάποιοι “σωλήνες” (tubes) που συνδέονται κατευθείαν με τα σύνολα IN και OUT. Τέλος υπάρχουν και άλλα πολύ μικρότερα μέρη , οι ομάδες από σελίδες που δεν μπορούν να ενταχθούν σε αυτή τη δομή.

Είναι πασιφανές ότι ο ιστός είναι μια τεράστια δομή που αυξάνει τάχιστα. Αυτό το δίκτυο πληροφοριών στερείται οργάνωσης και δομής και κάθε εσωτερική συσχέτιση καθορίζεται από τις υπερσυνδέσεις.

1.1.2 Εξόρυξη Περιεχομένου Ιστού

Ο ιστός έχει γίνει τα τελευταία χρόνια ο μεγαλύτερος χώρος αποθήκευσης δεδομένων περιέχοντας εκατοντάδες εκατομμυρίων από κείμενα που βρίσκονται σε ημι-δομημένη ή δομημένη μορφή , εικόνες ή αρχεία πολυμέσων. Η εκθετική αύξηση του μεγέθους του ακολουθείται από εκείνη της χρήσης του καθώς ολοένα και περισσότεροι χρησιμοποιούν το δίκτυο σαν την κύρια πηγή πληροφοριών. Ο μεγάλος όγκος δεδομένων σε συνδυασμό με την ετερογενή φύση του ιστού καθιστά την ανάκτηση πληροφοριών μια πολύ δύσκολη υπόθεση για το μέσο χρήστη.

Η εξόρυξη περιεχομένου ιστού παρέχει μεθόδους για την αυτόματη ανακάλυψη, ανάκτηση, οργάνωση και διαχείριση μιας μεγάλης ποσότητας πληροφοριών και διαθέσιμων πόρων του ιστού. Στις μέρες μας η εξόρυξη περιεχομένου συνδέεται με την εξόρυξη δομής ιστού μιας και συνήθως και τα δυο χρησιμοποιούνται σε συνδυασμό για την εξαγωγή και οργάνωση πληροφορίας από τον ιστό.

1.1.3 Εξόρυξη Χρήσης Ιστού

Η εξόρυξη χρήσης ιστού , πέρα από την καταγραφή και επεξεργασία της δομής και του περιεχομένου των ιστοσελίδων που αποτελούν τη βάση της διαδικτυακής πληροφορίας προχωρά σε ζητήματα που σχετίζονται με τη χρήση της πληροφορίας αυτής από μέρους μιας μεγάλης ποικιλίας χρηστών που την προσπελαίνουν. Το πρόβλημα επομένως έγκειται στο γεγονός ότι άνθρωποι με διαφορετικές ανάγκες και προτιμήσεις πλοηγούνται μέσα σε πολύ μεγάλες δομές ιστοσελίδων χάνοντας πολλές φορές το στόχο αναζήτησής τους. Η εξόρυξη δεδομένων για την εξατομίκευση ιστού είναι μια πολλά υποσχόμενη περιοχή για την

αποφυγή του προβλήματος της υπερφόρτωσης πληροφοριών παρέχοντας στους χρήστες πληροφορίες προσαρμοσμένες σε αυτούς.

1.2 ΤΙ ΕΙΝΑΙ ΕΞΑΤΟΜΙΚΕΥΣΗ

Πρόκειται για μια περιοχή που έχει τη δυναμική να παίξει ουσιαστικό ρόλο στην πορεία εξέλιξης του Web. Οι ρίζες της τεχνολογίας εξατομίκευσης εντοπίζονται στον ορισμό των προσαρμοστικών υπερμέσων, τα οποία αποτέλεσαν την εναλλακτική λύση στην παραδοσιακή προσέγγιση one-size-fits-all, χτίζοντας ένα μοντέλο των γνώσεων, των προτιμήσεων ή/και των στόχων κάθε χρήστη.

Οι πρώτες προσπάθειες εφαρμογής εξατομίκευσης περιορίστηκαν σε παραμετροποιήσεις μέσω επιλογών check-box, όπου για παράδειγμα οι πληροφοριακές πύλες επέτρεπαν στους χρήστες να επιλέξουν τους συνδέσμους που θα εμφανίζονται στις προσωπικές τους σελίδες. Η εφαρμογή αυτής της τεχνικής απαιτεί από τους χρήστες να εισαγάγουν προσωπικές πληροφορίες σχετικά με τα ενδιαφέροντά τους, γεγονός που ενδέχεται να δημιουργήσει προβλήματα καθώς οι χρήστες του Παγκόσμιου Ιστού είναι γενικά διστακτικοί στο να αποκαλύψουν τέτοιου είδους στοιχεία.

Λόγω τέτοιων προβλημάτων, οι ερευνητές στράφηκαν προς την εξατομίκευση με βάση την παρατήρηση, η οποία στηρίζεται στη μελέτη της καταγεγραμμένης πλοηγητικής συμπεριφοράς των χρηστών με σκοπό να εντοπιστούν στοιχεία που στη συνέχεια θα καθορίσουν το πώς θα πρέπει να εξατομικευτούν οι πληροφορίες, οι υπηρεσίες ή τα προϊόντα που προσφέρει μια εφαρμογή.

Η εξατομίκευση, λοιπόν, ορίζεται ως οποιαδήποτε ενέργεια που προσαρμόζει τις πληροφορίες ή τις υπηρεσίες που παρέχονται από ένα web site στη γνώση που αποκτιέται

από την πλοηγική συμπεριφορά και τα ατομικά ενδιαφέροντα των χρηστών του, σε συνδυασμό με το περιεχόμενο και τη δομή του web site.

Η εξατομίκευση έχει περάσει από πολλά διαφορετικά στάδια και μπορεί να παίζει ρόλο-κλειδί σε διάφορες πτυχές του ηλεκτρονικού εμπορίου.

1.3 ΔΕΔΟΜΕΝΑ ΕΙΣΟΔΟΥ

Τα δεδομένα που συγκεντρώνονται και χρησιμοποιούνται από μια διαδικασία εξατομίκευσης διακρίνονται σε 4 κατηγορίες (Srivastava et al., 2000)

- Δεδομένα περιεχομένου (content data). Πρόκειται για τα δεδομένα που καταλήγουν να παρουσιαστούν στον τελικό χρήστη και έχουν την μορφή απλού κειμένου , εικόνων ή δομημένων δεδομένων όπως είναι οι πληροφορίες που λαμβάνονται από βάσεις δεδομένων.

- Δεδομένα δομής (structure data). Δείχνουν τον τρόπο που είναι οργανωμένο το περιεχόμενο του site. Μπορεί να είναι στοιχεία που χρησιμοποιούνται μέσα σε μια ιστοσελίδα (όπως ετικέτες tags HTML,ή XML) ή στοιχεία που διασυνδέουν ένα site (όπως οι υπερσύνδεσμοι).

- Δεδομένα χρήσης (usage data). Περιγράφουν τη χρήση μιας τοποθεσίας (δηλαδή ενός web site) όπως είναι η διευθύνσεις IP των επισκεπτών , ο χρόνος και η ημερομηνία προσπέλασης, τα λαμβανόμενα αρχεία , οι διευθύνσεις τους καθώς και άλλα χαρακτηριστικά που συναντώνται σε ένα κατάλογο προσπέλασης ιστού(web access log).

- Δεδομένα προφίλ χρήστη (user profile data). Παρέχουν πληροφορίες για τους χρήστες μίας τοποθεσίας (web site). Εντός του προφίλ περιέχονται δημογραφικές πληροφορίες για κάθε χρήστη καθώς και δεδομένα για τα ενδιαφέροντα και τις προτιμήσεις του. Οι πληροφορίες αυτές αποκτώνται μέσω ερωτηματολογίων κατά την εγγραφή του χρήστη στο σύστημα ή και έμμεσα μέσω των καταλόγων χρήσης .

1.4 ΕΞΟΡΥΞΗ ΧΡΗΣΗΣ ΙΣΤΟΥ ΚΑΙ ΕΞΑΤΟΜΙΚΕΥΣΗ

Οι ρίζες της εξόρυξης γνώσης και εξατομίκευσης , βρίσκονται στις προσαρμοσμένες εφαρμογές υπερμέσων (adaptive hypermedia applications), όπως δείχνει η μελέτη του Brusilovsky 2001. Τα προσαρμοσμένα υπερμέσα αποτέλεσαν την εναλλακτική λύση στην παραδοσιακή προσέγγιση όπου ένα μέγεθος ταιριάζει σε όλα, χτίζοντας ένα μοντέλο με τους στόχους, τις προτιμήσεις και τη γνώση κάθε ξεχωριστού χρήστη.

Όπως είναι ήδη γνωστό, οι αρχικές προσπάθειες εξατομίκευσης σχετίζονταν με την εξατομίκευση τσεκαρίσματος (check box personalization) όπου οι πύλες ιστού επέτρεπαν στους χρήστες να επιλέγουν τις συνδέσεις που ήθελαν στις προσωπικές τους σελίδες, πράγμα που είχε τελικά περιορισμένη χρήση μιας και οι χρήστες έπρεπε να ξέρουν εκ των προτέρων το περιεχόμενο των ενδιαφερόντων τους. Μια πιο έξυπνη προσέγγιση ήταν τα συνεργατικά φίλτρα (collaborative filtering) που βασίζονταν στην γνώση για το τι προτιμούσαν και τι όχι οι παλαιοί χρήστες που θεωρούνταν παρόμοιοι με το χρήστη που εξετάζεται. Η τεχνική αυτή χρειάζεται την παροχή πληροφοριών από τους χρήστες για τα ενδιαφέροντα , τις ανάγκες ή και τις προτιμήσεις τους. Πράγμα που είναι δύσκολο να γίνει μιας και η συνεργασία των χρηστών δεν είναι σύνηθες φαινόμενο στον ιστό για την παροχή τέτοιων δεδομένων.

Εξαιτίας αυτών των προβλημάτων , οι ερευνητές προσέφυγαν στην εξατομίκευση παρατήρησης (observation personalization) που βασίζεται στην υπόθεση ότι μπορούμε να βρούμε ενδείξεις για την εξατομίκευση πληροφοριών , υπηρεσιών και προϊόντων σε εγγραφές της προηγούμενης συμπεριφοράς πλοήγησης των χρηστών. Σε αυτό το σημείο έρχεται η έννοια της εξόρυξης του ιστού , που ορίζεται σαν η χρήση τεχνικών εξόρυξης δεδομένων για την ανακάλυψη και εξαγωγή πληροφοριών και υπηρεσιών , και που διαχωρίζεται σε εξόρυξη περιεχομένου , χρήσης και δομής.

Στην πλειοψηφία των περιπτώσεων , οι εφαρμογές ιστού βασίζουν την εξατομίκευση σε εξόρυξη χρήσης του, η οποία αναλαμβάνει την εργασία της συλλογής και εξαγωγής των δεδομένων που απαιτούνται για τη δόμηση των προφίλ των χρηστών βασισμένη στη συμπεριφορά των χρηστών όπως είναι εγγεγραμμένη στους καταλόγους των εξυπηρετητών (server logs). Η προσπέλαση μιας σελίδας από έναν χρήστη οδηγεί σε μια εγγραφή σε ένα κατάλογο προσπέλασης στον εξυπηρετητή που βρίσκεται αυτή η σελίδα .

1.4.1 Εξατομίκευση Βασισμένη σε Εξόρυξη Χρήσης Ιστού

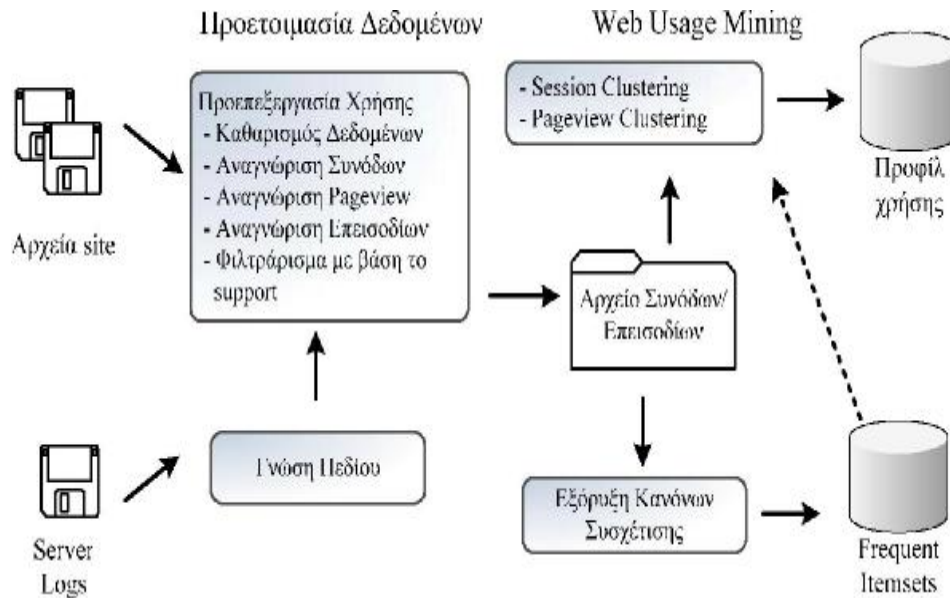
Η χρήση του web usage mining (Srivastava et al., 2000) ως τεχνολογία βάσης για ανάπτυξη εφαρμογών εξατομίκευσης προτάθηκε από τους Mobasher et al. (2000) και ήδη το πεδίο αυτό συγκεντρώνει το έντονο ενδιαφέρον της ερευνητικής κοινότητας (Eirinaki & Vazirgiannis, 2004; Pierrakos et al., 2003; Mobasher, 2004), ενώ έχουν κάνει την εμφάνισή τους και εμπορικές εφαρμογές που παρέχουν υπηρεσίες εξατομίκευσης χρησιμοποιώντας κάποια μορφή διαδικασίας web usage mining (Maurino & Fraternali, 2002). Ένα από τα πλεονεκτήματα αυτής της προσέγγισης είναι ότι τα δεδομένα που χρησιμοποιούνται σαν είσοδος, δεν αποτελούν υποκειμενικές περιγραφές από πλευράς χρηστών και έτσι δεν περιέχουν προκατάληψη. Παράλληλα, τα προφίλ εξάγονται (και ενημερώνονται) δυναμικά από τα πρότυπα που εντοπίζονται στα δεδομένα χρήσης, με αποτέλεσμα να μην μειώνεται η απόδοση του συστήματος με το χρόνο, λόγω μη ενημέρωσης των προφίλ. Επιπλέον, η κατασκευή των προφίλ αποκλειστικά με βάση ομοιότητες ως προς το περιεχόμενο (content similarities), έχει συχνά σαν αποτέλεσμα την απώλεια κάποιων σημασιολογικών συσχετίσεων μεταξύ των χρηστών ή των web αντικειμένων (σελίδων, προϊόντων, κλπ.).

Το web usage mining μειώνει την ανάγκη για συγκέντρωση υποκειμενικών βαθμολογήσεων από πλευράς χρηστών καθώς και την καταγραφή προσωπικών προτιμήσεων κατά την εγγραφή (registration) καθώς στηρίζεται στη συνεχή παρατήρηση (observation) της συμπεριφοράς των χρηστών και την ανάλυσή της, ώστε να εξάγει έμμεσα

προτιμήσεις, patterns, ομάδες ομοϊδεατών, κτλ. Η εξατομίκευση που βασίζεται στο web usage mining ονομάζεται *observational personalization* (παρατηρητική εξατομίκευση).

Οι θεμελιώδεις διεργασίες της εξατομίκευσης σε περιβάλλον web, περιλαμβάνουν τη μοντελοποίηση των web αντικειμένων (π.χ. προϊόντα ή σελίδες) και υποκειμένων (χρήστες), την κατηγοριοποίηση των αντικειμένων και των υποκειμένων και την αντιστοίχιση ανάμεσα σε αντικείμενα και/ή υποκείμενα ώστε να καθοριστούν οι ενέργειες που θα πρέπει να εκτελεστούν για να παραχθεί το εξατομικευμένο αποτέλεσμα.

Η συνολική διαδικασία της εξατομίκευσης με βάση δεδομένα χρήσης (usage-based) διασπάται τυπικά σε δύο φάσεις. Η off-line φάση αποτελείται από την προετοιμασία των δεδομένων και συγκεκριμένες εργασίες ανάκτησης. Οι εργασίες προετοιμασίας των δεδομένων έχουν σαν αποτέλεσμα ένα αρχείο συνόδου server, όπου κάθε σύνοδος είναι μια ακολουθία από pageviews που το καθένα αντιπροσωπεύεται από ένα μοναδικό URI. Στα όρια μιας συνόδου server, κάθε URI αντιστοιχείται σε ένα συγκεκριμένο χρήστη. Μόνο τα URIs που αντιπροσωπεύουν pageviews με κάποια σημασία ή σχετικότητα περιλαμβάνονται σε ένα αρχείο συνόδου server. Οι εργασίες usage mining μπορεί να αφορούν την ανακάλυψη κανόνων συσχέτισης (association rules) ή σειριακών προτύπων (sequential patterns), τον εντοπισμό ομάδων (clusters) από pageviews/χρήστες, ή οποιαδήποτε άλλη μέθοδο ανακάλυψης προτύπων (pattern discovery). Στο σχήμα που ακολουθεί παρουσιάζεται ενδεικτικά η off-line φάση της διαδικασίας εξατομίκευσης, όπως εφαρμόζεται από τους Mobasher et al. (2000).



Σχήμα 1. Η off-line φάση μιας διαδικασίας εξατομίκευσης βασισμένης σε δεδομένα χρήσης web που χρησιμοποιεί για ανακάλυψη προτύπων την τεχνική της εξαγωγής κανόνων συσχέτισης.

Τα πρότυπα που εντοπίζονται χρησιμοποιούνται στην *on-line φάση*, ώστε να παρέχουν στους χρήστες εξατομικευμένες διαμορφώσεις του website, με βάση την τρέχουσα συμπεριφορά τους κατά την πλοήγηση και τις υπόλοιπες τυχόν πληροφορίες που έχει αποθηκεύσει η εφαρμογή για το χρήστη (ατομικό προφίλ, ιστορικό παλαιότερων επισκέψεων, κλπ.). Στο σχήμα που ακολουθεί (Σχήμα 2), παρουσιάζεται η *on-line φάση* της διαδικασίας εξατομίκευσης που εφαρμόζουν οι Mobasher et al. (2000). Σύμφωνα με το σχήμα που προτείνουν, ενώ ο browser του χρήστη στέλνει HTTP αιτήσεις στο server, ο server κρατάει ιστορικό για την τρέχουσα σύνοδο. Η μηχανή παραγωγής συστάσεων εξετάζει την τρέχουσα σύνοδο server, σε συνδυασμό με τα πρότυπα που έχουν εντοπιστεί, ώστε να εξυπηρετεί τις αιτήσεις κάθε browser αποστέλλοντάς του το κατάλληλο περιεχόμενο. Οι εξατομικευμένες διαμορφώσεις μπορεί να πάρουν τη μορφή προτεινόμενων

συνδέσμων ή προϊόντων, στοχευμένων διαφημίσεων (targeted advertisements), ή κειμένων και γραφικών που να ικανοποιούν τις προτιμήσεις των χρηστών.

Η διαδικασία εξατομίκευσης που βασίζεται σε web usage mining είναι στην ουσία μια διαδικασία data mining και σαν τέτοια, μπορεί σύμφωνα με τους Pierrakos et al. (2003) να αναλυθεί στα παρακάτω στοιχειώδη στάδια data mining (από τις φάσεις αυτές μόνο η τελευταία εκτελείται σε πραγματικό χρόνο):

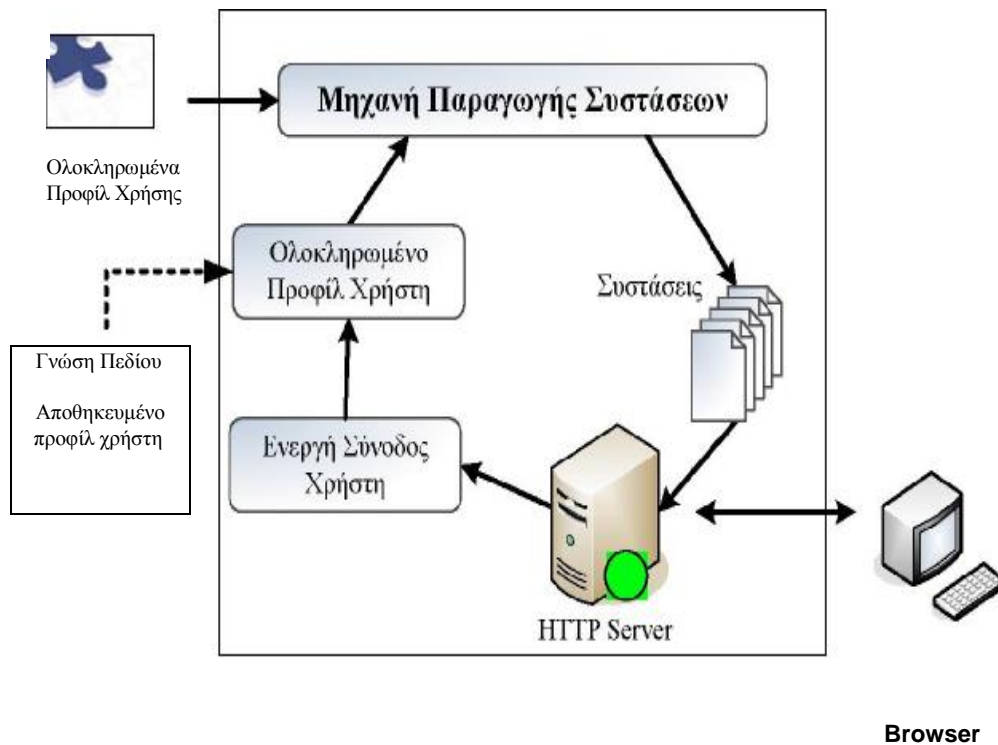
1. *Συλλογή Δεδομένων (Data Collection)*: συγκεντρώνονται δεδομένα χρήσης από διάφορες πηγές λαμβάνοντας υπόψη τη δομή και το περιεχόμενό τους (τυπικές πηγές δεδομένων για μια διαδικασία web usage mining είναι οι web servers, οι clients, και οι proxy servers).

2. *Προεπεξεργασία Δεδομένων (Data Pre-processing)*: καθαρίζονται τα δεδομένα από θόρυβο, επιλύονται ασυμβατότητες, ολοκληρώνονται και ενοποιούνται, ώστε να χρησιμοποιηθούν σαν είσοδος στο στάδιο της ανακάλυψης προτύπων.

3. *Ανακάλυψη Προτύπων (Pattern Discovery)*: ανακαλύπτεται νέα γνώση με την εφαρμογή τεχνικών από το χώρο του machine learning και της στατιστικής, όπως το clustering, το classification, η εξόρυξη κανόνων συσχέτισης (association rule mining) και η ανακάλυψη σειριακών προτύπων (sequential pattern discovery). Τα πρότυπα που χρησιμοποιούνται από τις τεχνικές εξατομίκευσης αντιστοιχούν στη συμπεριφορά και τα ενδιαφέροντα των χρηστών. Στο στάδιο αυτό, εφαρμόζονται μέθοδοι μάθησης με σκοπό να αυτοματοποιηθεί η κατασκευή των μοντέλων χρηστών.

4. *Μετα-επεξεργασία Γνώσης (Knowledge Post-processing)*: αξιολογείται η γνώση που έχει αποκτηθεί και συνήθως παρουσιάζεται σε κάποια μορφή που είναι εύκολα κατανοητή στο χρήστη (αναφορές, διαγράμματα και άλλες μορφές οπτικοποίησης). Ειδικά για τις ανάγκες της εξατομίκευσης, η νεοαποκτηθείσα γνώση διοχετεύεται στο υποσύστημα εξατομίκευσης για να χρησιμοποιηθεί στον υπολογισμό και εφαρμογή των κατάλληλων

προσαρμογών, ώστε να παραχθούν οι αντίστοιχες συνολικές εξατομικευμένες διαμορφώσεις της web εφαρμογής.



Σχήμα 2. Η on-line φάση μιας διαδικασίας εξατομίκευσης βασισμένης σε δεδομένα χρήσης web, που στοχεύει στον υπολογισμό και την αποστολή εξατομικευμένων συστάσεων.

Θα πρέπει να σημειωθεί ότι το web mining είναι μια περιοχή που, σε σύγκριση με το data mining είναι πολύ λιγότερο ώριμη. Αποτέλεσμα αυτού του γεγονότος είναι ότι πολλά από τα θέματα που έχουν μελετηθεί από τον ερευνητικό χώρο του data mining (όπως ο ορισμός του προβλήματος και η αξιολόγηση της γνώσης που εξορύσσεται και που αντιμετωπίζονται σαν διακριτά στάδια της διαδικασίας data mining), να αποτελούν ανεξερεύνητη περιοχή για το web mining (Pierrakos et al., 2003).

ΚΕΦΑΛΑΙΟ 2

Ο ΣΗΜΑΣΙΟΛΟΓΙΚΟΣ ΠΑΓΚΟΣΜΙΟΣ ΙΣΤΟΣ ΠΛΗΡΟΦΟΡΙΩΝ

2.1 ΓΕΝΙΚΑ ΓΙΑ ΤΟ SEMANTIC WEB

Ο Σημασιολογικός Παγκόσμιος Ιστός Πληροφοριών αποτελεί σήμερα μια από τις πιο σημαντικές προσπάθειες για την αυτόματη ενοποίηση των συστημάτων και των εφαρμογών ώστε να συνεργάζονται διαλειτουργικά σε παγκόσμιο επίπεδο. Ο Tim Berners-Lee που δημιούργησε το Web το 1989 οραματιζόταν ένα ιστό δεδομένων αυτόματα επεξεργάσιμων από τις εφαρμογές βάσει του νοήματος και όχι της μορφής της πληροφορίας. Αυτό θα έχει σαν αποτέλεσμα από το απλό κείμενο που το καταλαβαίνει πλήρως μόνο ο άνθρωπος να κινούμαστε προς τις ημιδομημένες (semi-structured) και δομημένες (structured) πληροφορίες που μπορούν να γίνουν κατανοητές από τις διαδικτυακές εφαρμογές όπως για παράδειγμα συμβαίνει στα συστήματα ηλεκτρονικού εμπορίου (e-commerce), ηλεκτρονικού επιχειρείν (ebusiness) και ηλεκτρονικής μάθησης (e-learning).

Ο Σημασιολογικός Παγκόσμιος Ιστός Πληροφοριών αποτελεί πρωτοβουλία της Κοινοπραξίας του Παγκοσμίου Ιστού ή World Wide Web Consortium ή W3C (<http://www.w3.org>). Η ιδέα στην οποία βασίζεται αφορά στην οργάνωση και στη διασύνδεση της διαθέσιμης πληροφορίας στο Διαδίκτυο ώστε να χρησιμοποιηθεί πιο αποτελεσματικά στην ανακάλυψη, στην αυτοματοποίηση, στην ομαδοποίηση και στην επαχρησιμοποίησή της από διαφορετικές μεταξύ τους διαδικτυακές εφαρμογές.

Ο Σημασιολογικός Παγκόσμιος Ιστός Πληροφοριών στοχεύει δηλαδή στη μετεξέλιξη του σημερινού Web ώστε να μετατραπεί σε ένα περιβάλλον όπου οι πληροφορίες που υπάρχουν και διακινούνται να είναι επεξεργάσιμες από τους ηλεκτρονικούς υπολογιστές. Κινείται προς την κατεύθυνση της επέκτασης και της βελτίωσης της δόμησης της πληροφορίας για να είναι προσπελάσιμη από προγράμματα υπολογιστών με τελικό στόχο

την αυτοματοποίηση πολλών λειτουργιών στο Web. Σε αυτό το πλαίσιο, τα σημερινά κείμενα στις σελίδες του Web θα αντικατασταθούν με δομημένα κείμενα και δεδομένα σε μορφή XML και RDF. Από αυτή την άποψη ο Σημασιολογικός Παγκόσμιος Ιστός Πληροφοριών είναι ένα πρόβλημα αναπαράστασης γνώσης (knowledge representation) από και για τους υπολογιστές και απαιτεί ένα μηχανισμό που θα επιτρέπει την επεξεργασία αυτής της γνώσης.

Αυτός ο μηχανισμός θα πρέπει να υποστηρίζει τη δυνατότητα λογικής επεξεργασίας των πληροφοριών με σκοπό την εξαγωγή συμπερασμάτων, τη δημιουργία νέας γνώσης, την υποστήριξη στη λήψη αποφάσεων και την αυτόματη εκτέλεση ενεργειών. Η υλοποίηση του παραπάνω οράματος είναι πολύ δύσκολη προς το παρόν. Προς αυτή την κατεύθυνση η χρήση των οντολογιών μπορεί να αποδειχθεί ιδιαίτερα κρίσιμη παράμετρος. Μια οντολογία (ontology) είναι μια αυστηρή (formal) περιγραφή των αντικειμένων-πόρων και των σχέσεων μεταξύ τους (Dogac A:2004).

Οι οντολογίες δίνουν τη δυνατότητα σε μια κοινότητα χρηστών να έχουν κοινή ονοματολογία και κοινή αντίληψη για τα αντικείμενα-πόρους τα οποία δηλώνουν ή χρησιμοποιούν. Επίσης, τους επιτρέπουν να περιγράψουν τον «κόσμο» που χρησιμοποιούν (που τους αφορά) με ένα τέτοιο τρόπο, ώστε οι υπολογιστές να μπορούν να επεξεργαστούν και να «συνδέσουν» τα δεδομένα από διαφορετικούς κόσμους. Η δυσκολία στις οντολογίες έγκειται στο γεγονός ότι οι κοινότητες των χρηστών θα πρέπει να συμφωνήσουν στην οντολογική περιγραφή του πεδίου ενδιαφέροντός τους (Berners-Lee:2000).

Η λέξη "Σημασιολογία" έχει ρίζα τις Ελληνικές λέξεις "σημάδι", "σημαίνω" και "σημαντικός" και σήμερα αναφέρεται στο νόημα συχνά σε επίπεδο γλώσσας. Μπορούμε να πούμε ότι ο Σημασιολογικός Ιστός αποτελεί το μεγαλύτερο σε παγκόσμιο επίπεδο έργο έξυπνης ενσωμάτωσης συστημάτων ώστε να συνεργάζονται δια-λειτουργικά.

Ο Tim Berners-Lee, που επινόησε τον Παγκόσμιο Ιστό το 1989, είχε το όραμα, που

τώρα συμμερίζονται πολλοί - ενός ιστού δεδομένων που μπορούν να επεξεργαστούν από μηχανές.

"Ο Σημασιολογικός Ιστός είναι μια επέκταση του σημερινού ιστού όπου η πληροφορία έχει καλά καθορισμένο νόημα, καθιστώντας τη συνεργασία μεταξύ ανθρώπων και υπολογιστών πιο αποτελεσματική".

Το κέντρο βάρους του περιεχομένου του Ιστού μετατοπίζεται συνεχώς από τον άνθρωπο προς τα δεδομένα. Για να φτάσει ο Ιστός το μέγιστο των δυνατοτήτων του, πρέπει να εξελιχθεί σε ένα Σημασιολογικό Ιστό, ο οποίος παρέχει μια διεθνώς προσβάσιμη πλατφόρμα που επιτρέπει σε αυτοματοποιημένα εργαλεία αλλά και σε ανθρώπους να μοιράζονται και να επεξεργάζονται δεδομένα. Ο Σημασιολογικός Ιστός αποτελεί πρωτοβουλία της Κοινοπραξίας του Παγκοσμίου Ιστού (W3C) και η σχετική Δραστηριότητα (W3C Semantic Web Activity) έχει δημιουργηθεί για να εξυπηρετήσει έναν ηγετικό ρόλο, τόσο στο σχεδιασμό προδιαγραφών, όσο και στην ανοικτή ανάπτυξη της τεχνολογίας μέσω της συνεργασίας.

Ο Berners-Lee, Lassila, Hendler στο άρθρο του Scientific American αναφέρει ότι

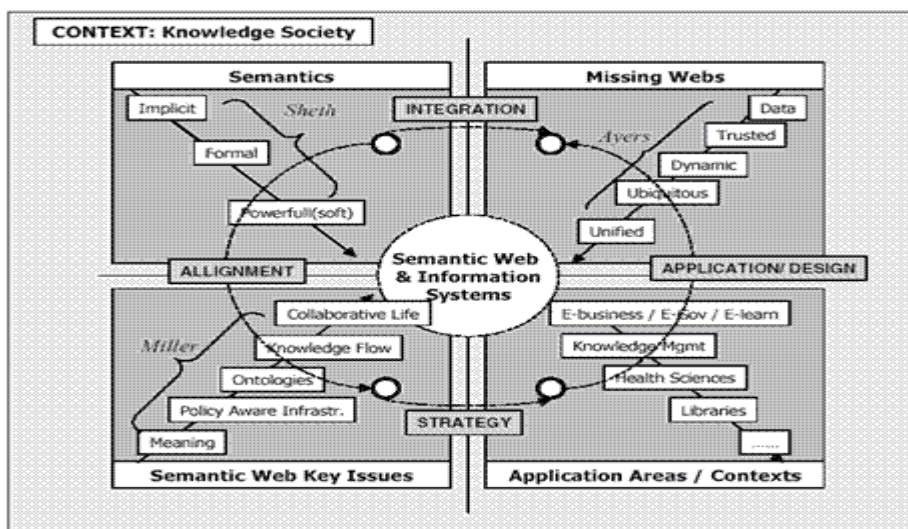
«Ο σημασιολογικός ιστός δεν είναι ένας ξεχωριστός ιστός αλλά η επέκταση του συντακτικού ιστού, στον οποίο η πληροφορία είναι καλά καθορισμένη κάνοντας καλύτερη τη συνεργασία ανθρώπων και υπολογιστών»

Ο σημασιολογικό ιστός είναι ένα σύνολο από πηγές και συνδέσεις. Οι υπολογιστές μπορούν να παράγουν συμπεράσματα έχοντας πρόσβαση σε κατανοητή σε αυτούς γνώση. Για τους χρήστες αυτό είναι κάτι πάρα πολύ πιο συναρπαστικό.

Για τη καλύτερη κατανόηση του σημασιολογικού ιστού ο συγγραφέας αναφέρει το ακόλουθο παράδειγμα : «Φανταστείτε να πηγαίνατε σε μία μηχανή αναζήτησης και να θέλετε να μάθετε « πόσες γραμμές τραίνου υπάρχουν στην Ιαπωνία?». Με βάση το σημερινό δίκτυο σύμφωνα με το συγγραφέα ο χρήστης θα πάρει πολλές σελίδες για

απάντηση αλλά μπορεί σε καμιά να μην είναι η σωστή απάντηση.

Στο σημασιολογικό ιστό αναφέρεται ότι οι γραμμές τραίνου είναι πάνω από 5000, είτε αναφέρεται ότι υπάρχει μία βάση που παρέχει τέτοιες πληροφορίες αλλά χρειάζεται κωδικό πρόσβασης. Ο ίδιος ιστός μπορεί να δώσει ως απάντηση ότι υπάρχει μια υπηρεσία διαδικτύου που μπορεί να υπολογίσει αλλά κοστίζει 500 yen. Τέλος μπορεί να απαντήσει ότι μπορεί να βρει την απάντηση αλλά θα πάρει λίγες ώρες να το υπολογίσει. Ουσιαστικά ο σημασιολογικός ιστός δίνει συγκεκριμένη απάντηση, κάτι που δε κάνει το σημερινό Ίντερνετ, το οποίο θα δώσει εκατομμύρια πληροφορίες και θα αναγκάσει το χρήστη να ψάχνει για ώρες, πολλές φορές και χωρίς αποτέλεσμα. Παρακάτω παρουσιάζουμε παραδείγματα του σημασιολογικού ιστού. Συγκεκριμένα:



Σχήμα 3. Σημασιολογικός παγκόσμιος ιστός.

Πηγή: Lytras M. (2005) “Semantic Web and Information Systems: An Agenda Based on Discourse with Community Leaders”, in International Journal on Semantic Web and Information Systems, Inaugural Issue, 1(1), pp: i-xii

Η σημασιολογική πληροφορία που είναι συσχετισμένη με πηγές στο διαδίκτυο (ως μεταδεδομένα) μπορεί να ερμηνευτεί και να συνδυαστεί από τις μηχανές με στόχο την

παραγωγή λογικής. Μία πλατφόρμα για μηχανές εύρεσης και έξυπνους πράκτορες Εξέλιξη του σημερινού διαδικτύου έτσι ώστε η πληροφορία να μπορεί να επεξεργαστεί από ανθρώπους και μηχανές. Οι σημασιολογικές πληροφορίες θα μπορούσαν επίσης να χρησιμοποιηθούν για την ενοποίηση του ορατού (πληροφορίες που βρίσκονται στις μηχανές αναζήτησης) και του άορατου ιστού (πληροφορίες που δεν βρίσκονται στις μηχανές αναζήτησης).

Ο σημασιολογικός ιστός δεν αντικαθιστά το σημερινό παγκόσμιο ιστό. Ο σημασιολογικός ιστός κάθεται δίπλα στο σημερινό παγκόσμιο ιστό. Οι παρούσες προσπάθειες αναφέρονται στα ακόλουθα (Βασίλης Τ.2007):

1. Καθορίζουν τις γλώσσες που θα πλαισιώσουν το σημασιολογικό ιστό-NA
ΚΑΘΟΡΙΣΤΕΙ

2. Καθορίζουν και αναπτύσσουν την αρχιτεκτονική και τα εργαλεία που θα χρησιμοποιηθούν στο σημασιολογικό ιστό-NA ΦΤΙΑΧΤΕΙ

3. Χρησιμοποιούν τα εργαλεία και τις γλώσσες για να περιγράψουν το υλικό σε εφαρμογές-NA ΕΦΑΡΜΟΣΤΕΙ

Οι υπηρεσίες που παρέχει ο σημασιολογικός ιστός αναφέρονται στην περιγραφή των αρχείων, στην ενοποίηση της πληροφορίας από διαφορετικές πηγές, στην ελάττωση του χρόνου περιγραφής, στην ανάκτηση και ενοποίηση ανομοιογενούς πληροφορίας.

2.2 Η ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΤΟΥ ΣΗΜΑΣΙΟΛΟΓΙΚΟΥ ΠΑΓΚΟΣΜΙΟΥ ΙΣΤΟΥ ΠΛΗΡΟΦΟΡΙΩΝ

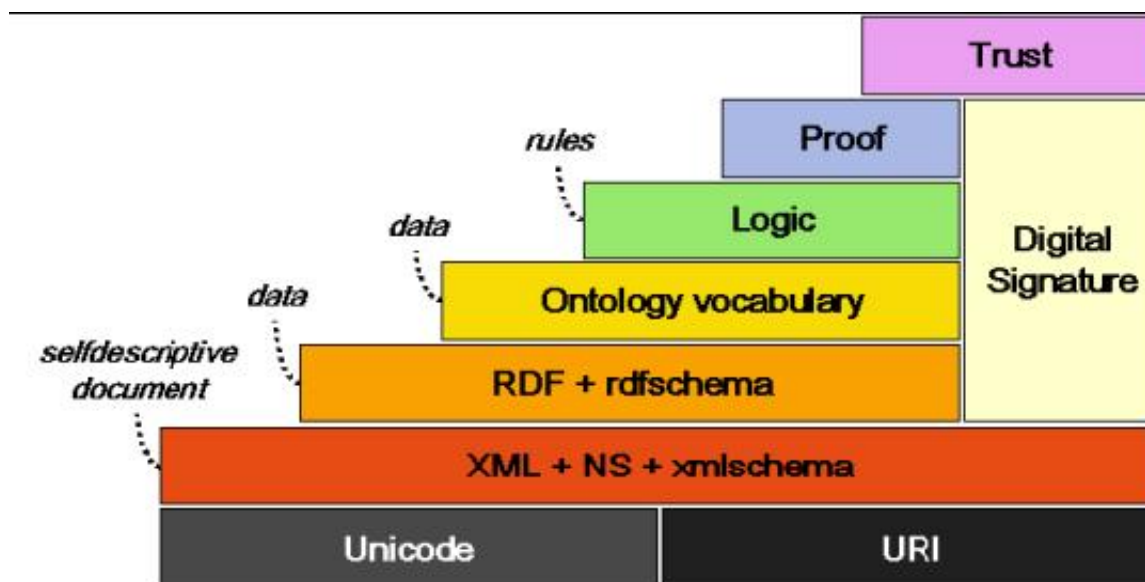
Η βασική ιδέα του Σημασιολογικού Παγκόσμιου Ιστού Πληροφοριών είναι να συλλαμβάνει και να εκφράζει έννοιες και σημασίες. Για το σκοπό αυτό στηρίζεται σε μια αρχιτεκτονική διαφορετικών επιπέδων, τα σημαντικότερα από τα οποία απεικονίζονται στο

Σχήμα 1 (Nilson M:2001):

- *Unicode και URI επίπεδα*: διασφαλίζουν ότι διεθνή σύνολα χαρακτήρων και μέσα χρησιμοποιούνται για την αναγνώριση των αντικειμένων του Σημασιολογικού Παγκόσμιου Ιστού Πληροφοριών.

- *XML επίπεδο*: αναπαριστά τη δομή των δεδομένων. Το επίπεδο αυτό μαζί με τους ορισμούς namespace (NS) και xmlschemas διασφαλίζουν ότι είναι δυνατή η ενσωμάτωση σημασιολογικών ορισμών με τα άλλα XML πρότυπα.

- *RDF επίπεδο*: αναπαριστά τη σημασιολογία των δεδομένων. Με RDF και rdfschema είναι δυνατό να γίνουν οι δηλώσεις για τα αντικείμενα με URIs και να καθοριστούν τα λεξιλόγια που μπορούν να αναφερθούν από URIs. Αυτό είναι το επίπεδο όπου μπορούν να δοθούν οι τύποι στους πόρους και τις συνδέσεις.



Σχήμα 4. Τα επίπεδα της αρχιτεκτονικής του Σημασιολογικού Παγκόσμιου Ιστού Πληροφοριών.

Πηγή: Nilson M.,(2001), The Semantic Web: How RDF will change learning technology *standards*, Center for User-Oriented IT-design, Royal Institute of Technology, Stockholm, online at: <http://cid.nada.kth.se/pdf/CID-157.pdf>

- *Ontology επίπεδο*: αναπαριστά την επίσημη κοινή «συμφωνία» για το νόημα των δεδομένων. Το επίπεδο αυτό υποστηρίζει την εξέλιξη των λεξιλογίων δεδομένου ότι μπορεί να καθορίσει τις σχέσεις μεταξύ των διαφορετικών εννοιών.

- *Digital signature επίπεδο*: το επίπεδο αυτό στοχεύει στην ανίχνευση των αλλαγών των εγγράφων και μαζί με το ontology επίπεδο αποτελούν τα επίπεδα που τυποποιούνται αυτή την περίοδο από τις ομάδες εργασίας του W3C.

- *Logic, proof και trust επίπεδα*: αποτελούν τα ανώτερα επίπεδα της αρχιτεκτονικής τα οποία ερευνώνται αυτή την περίοδο, ενώ παράλληλα υλοποιούνται και απλές εφαρμογές επίδειξής τους. Το επίπεδο λογικής επιτρέπει το γράψιμο των κανόνων, ενώ το επίπεδο απόδειξης εκτελεί τους κανόνες και αξιολογεί μαζί με το μηχανισμό του επιπέδου εμπιστοσύνης για τις εφαρμογές αν θα εμπιστευθεί τη δεδομένη απόδειξη ή όχι.

Αξίζει να σημειωθεί ότι η πραγματική δύναμη του Σημασιολογικού Παγκόσμιου Ιστού Πληροφοριών θα υλοποιηθεί όταν οι άνθρωποι δημιουργήσουν συστήματα τα οποία θα συλλέγουν τα περιεχόμενα του Web από διαφορετικές πηγές, θα επεξεργάζονται τις πληροφορίες και θα ανταλλάσσουν τα αποτελέσματα με άλλους ανθρώπους ή μηχανές. Με αυτό τον τρόπο, η αποτελεσματικότητα του Σημασιολογικού Παγκόσμιου Ιστού Πληροφοριών θα βελτιωθεί δραστικά αφού μεγαλύτερο ποσοστό περιεχομένου Web θα είναι «αναγνώσιμο» από τους υπολογιστές και περισσότερες αυτοματοποιημένες υπηρεσίες θα είναι διαθέσιμες. Δύο σημαντικές τεχνολογίες για την ανάπτυξη του Σημασιολογικού Παγκόσμιου Ιστού Πληροφοριών είναι ήδη σε ισχύ: η XML (extensible Markup Language) και το RDF (Recourse Description Framework).

Η XML (<http://www.w3.org/XML>) επιτρέπει στον καθένα να δημιουργήσει ετικέτες (tags) που σχολιάζουν τις ιστοσελίδες ή τμήματα κειμένου σε μια ιστοσελίδα. Τα προγράμματα που διαχειρίζονται XML μπορούν να αξιοποιήσουν την πληροφορία αυτών των ετικετών με πολύπλοκους τρόπους, αρκεί να ξέρουν τι αυτές σημαίνουν (πως δηλαδή ο συγγραφέας της ιστοσελίδας χρησιμοποιεί κάθε ετικέτα). Ουσιαστικά, η XML επιτρέπει μια στοιχειώδη και αυθαίρετη δόμηση των δεδομένων, αλλά όχι την προσθήκη σημασιολογίας σε αυτά σε ικανοποιητικό βαθμό (Erdmann & Studer, 2000).

Η έννοια των XML εγγράφων είναι «διαισθητικά» σαφής λόγω της σήμανσης και των ετικετών. Εντούτοις, οι υπολογιστές δεν διαθέτουν αυτή τη «διαίσθηση». Επίσης, τόσο τα DTDs (Data Definition Types) όσο και το XMLSchema δεν λύνουν το πρόβλημα. Τα πρώτα αποτελούν μια δυνατότητα να «κτιστεί» το περιεχόμενο των εγγράφων, όμως δεν είναι κατάλληλα για να περιγράψουν τη σημασιολογία ενός XML εγγράφου. Το ίδιο ισχύει και για το XMLSchema (<http://www.w3.org/XML/Schema>) αφού μπορεί μόνο να καθορίσει τη δομή μέσω μιας πλουσιότερης γλώσσας.

Άρα, η XML από μόνη της δεν προσφέρει πλήρως τη δυνατότητα διάδοσης σημασιολογίας μέσω του Σημασιολογικού Παγκόσμιου Ιστού Πληροφοριών αλλά μπορεί να παίξει το ρόλο του ενδιάμεσου σαν ένα σχήμα δεδομένων εύκολα επεξεργάσιμου από ηλεκτρονικούς υπολογιστές.

Το RDF (<http://www.xml.com/xml/pub/98/06/rdf.html>) παρέχει πραγματική σημασιολογία στα δεδομένα, τα οποία περιγράφει μέσω δηλώσεων (statements). Οι δηλώσεις αυτές είναι κατανοήσιμες από τον υπολογιστή και οι μηχανές αναζήτησης ή οι πράκτορες μπορούν να χρησιμοποιήσουν αυτή την πληροφορία. Το RDF είναι ανεξάρτητο υλοποίησης και μπορεί η σύνταξή του να καθοριστεί σε XML. Η διαδικασία με την οποία σημασιολογικές πληροφορίες προστίθενται σε Web έγγραφα είναι γνωστή σαν σημασιολογικός σχολιασμός (semanticannotation) (Barreto C.M:2002,pp.221-240).

Το RDF σε συνδυασμό με το RDFS ή RDFSchema (<http://www.w3.org/TR/PRrdf-schema>) προσφέρει διαμορφώσεις οι οποίες μπορούν να επεκταθούν ανάλογα με τις ανάγκες. Οι βασικές ιεραρχίες των κλάσεων και οι σχέσεις μεταξύ αυτών και των αντικειμένων μπορούν να εκφραστούν σε RDFS. Γενικά, θα μπορούσαμε να πούμε ότι το μοντέλο RDF(S) δεν είναι «διαισθητικό», γιατί η σημασιολογία δεν δίνεται με τυπικό τρόπο και συχνά είναι δύσκολη η διερμηνεία της με αποτέλεσμα να αποτελεί μια επιρρεπή σε λάθη διαδικασία(Barreto C.M:2002,pp.221-240).

Μια λύση σε αυτό το πρόβλημα παρέχεται από το τρίτο βασικό συστατικό του Σημασιολογικού Παγκόσμιου Ιστού Πληροφοριών, δηλαδή τις οντολογίες. Η χρήση του όρου είναι δανεισμένη από τη Φιλοσοφία όπου ασχολείται με την ύπαρξη και τις θεωρίες για τη φύση και τα διάφορα είδη της.

Στην Τεχνητή Νοημοσύνη (Artificial Intelligence) και γενικότερα στην έρευνα στο Web ο όρος χρησιμοποιείται για να ορίσει μια επίσημη και κοινή «σύλληψη» μιας συγκεκριμένης περιοχής ενδιαφέροντος. Είναι δε κατάλληλη για να περιγράψει ετερογενείς, κατακεκομμένες και ήμι-δομημένες πηγές πληροφοριών που μπορούν να βρεθούν στο Web(Αλεξιάκης Σ:2002) .

Με τον καθορισμό των κοινών περιοχών, οι οντολογίες βοηθούν τόσο τους ανθρώπους όσο και τους υπολογιστές να επικοινωνήσουν μεταξύ τους, ενισχύοντας την ανταλλαγή της σημασιολογίας και όχι μόνο της σύνταξης. Οι οντολογίες τυπικά αποτελούνται από τους ορισμούς των εννοιών που είναι σχετικές με την περιοχή, τις σχέσεις τους και τα αξιώματα για αυτές τις έννοιες και τις σχέσεις. Διάφορες γλώσσες και συστήματα αναπαράστασης υπάρχουν π.χ. η OIL (Ontology Interchange Language) (<http://www.ontoknowledge.org/oil>) (Fensel et al., 2001), η DAML+OIL (DARPA Agent Markup Language + Ontology Interchange Language) (<http://www.daml.org/2001/03/reference.html>) (Horrocks, 2002), η OWL (Web Ontology Language) (<http://www.w3.org/TR/owl-ref>) (McGuinness & van

Harmelen, 2004), κλπ. Ενώ μεταξύ των πιο γνωστών εργαλείων είναι: το KAON (<http://kaon.semanticweb.org>), το OilEd (<http://oiled.man.ac.uk>), το Ontolingua (<http://www.wksl.stanford.edu>), το OntoSaurus (<http://www.isi.edu>), το Protégé2000 (<http://protege.stanford.edu>), το OntoEdit (<http://ontoserver.aifb.unikarlsruhe.de/ontoedit>), το WebOntoform (<http://kmi.open.ac.uk/projects/webonto>), το WebODE (<http://webode.dia.fi.upm.es/webODE>), κλπ.

2.2.1 Ο ορισμός της Οντολογίας

Οι οντολογίες αποτελούν τον πυρήνα για την υλοποίηση του Σημασιολογικού Παγκόσμιου Ιστού Πληροφοριών καθώς είναι το μέσο με το οποίο τα σημασιολογικά δεδομένα γίνονται αναγνώσιμα από τον υπολογιστή). Κατά καιρούς έχουν δοθεί διάφοροι ορισμοί για το τι είναι οντολογία. Ένας επίσημος κοινά αποδεκτός ορισμός είναι ο ακόλουθος (Davies et al:2004):

«Η οντολογία ορίζεται με βάση και τη φιλοσοφία των αρχαίων ελλήνων ως η συστηματική ανάλυση και παρουσίαση του όντος και της ύπαρξης.»

Ο όρος conceptualization αφορά σε ένα αφηρημένο μοντέλο που αποδίδει κάποιο φαινόμενο του κόσμου καθορίζοντας τις σχετικές έννοιες αυτού του φαινομένου. Ο όρος explicit σημαίνει ότι οι τύποι των εννοιών που χρησιμοποιούνται και οι περιορισμοί στη χρήση τους είναι επακριβώς καθορισμένοι.

Σε αυτή την περίπτωση ο όρος formal σημαίνει ότι η οντολογία είναι αναγνώσιμη από τον υπολογιστή, shared ότι συλλαμβάνει προφορική γνώση η οποία δεν είναι ιδιωτική αλλά κοινή, ενώ το domain of interest δηλώνει ότι κάποιος δεν ενδιαφέρεται να αναπαραστήσει όλο τον κόσμο αλλά κάποιο συγκεκριμένο τμήμα του που παρουσιάζει ενδιαφέρον (Fensel

D:2004).

Παράλληλα, ο Fensel D το 2004 εισάγει σαν μέρος του «Frame Ontology» τα ακόλουθα χαρακτηριστικά που χρησιμοποιούνται για να μοντελοποιήσουμε μια οντολογία:

- classes ή κλάσεις (συχνά αναφέρονται και σαν concepts). Οι classes μπορούν να οργανωθούν σε ταξινομίες μέσω των οποίων μηχανισμοί κληρονομικότητας μπορούν να εφαρμοστούν. Παραδείγματα κλάσεων από την περιοχή του e-learning μπορεί να είναι ο Learner, το Course, το Lesson, το Topic, κλπ.

- relations ή σχέσεις (συχνά, κυρίως στην RDF κοινότητα αναφέρονται σαν properties). Τα relations αναπαριστούν ένα τύπο αλληλεπίδρασης μεταξύ των concepts του domain. Ένα παράδειγμα σχέσης μπορεί να είναι το SubclassOf που ορίζει ότι μια κλάση αποτελεί υποκλάση μιας άλλης κλάσης π.χ. το Lesson είναι υποκλάση του Course, το Topic είναι υποκλάση του Lesson, κλπ.

- axioms ή αξιώματα. Τα axioms χρησιμοποιούνται για να μοντελοποιήσουν προτάσεις οι οποίες είναι πάντα αληθείς. Επίσης, χρησιμοποιούνται για να αναπαραστήσουν γνώση που δεν μπορεί να καθοριστεί από τα άλλα στοιχεία. Παράδειγμα αξιώματος μπορεί να είναι ότι ένας Learner για να απαντήσει ένα ερώτημα πρέπει να έχει διαβάσει την εκφώνησή του.

- instances. Τέλος, τα instances αποτελούν στιγμιότυπα των concepts και των relations. Ένα παράδειγμα της κλάσης Learner είναι ο εκπαιδευόμενος «Πέννυ Μαρκέλλου» που παρακολουθεί το Topic «Πως το Web Usage Mining μπορεί να βοηθήσει στην παροχή εξατομικευμένων εκπαιδευτικών εμπειριών» που ανήκει στο Lesson «Web Mining».

2.2.2 Ταξινόμηση των Οντολογιών

Οι οντολογίες μπορούν να ταξινομηθούν με διάφορους τρόπους και σε διάφορες

κατηγορίες. Αρχικά, οι Mizoguchi et al., (1995) πρότειναν τις ακόλουθες τέσσερις κατηγορίες οντολογιών (Mizoguchi et al.:1995)

- Οντολογίες περιεχομένου για επαναχρησιμοποίηση γνώσης (content ontologies for reusing knowledge). Οι οντολογίες αυτής της κατηγορίας είναι δυνατόν να διακριθούν περαιτέρω στις υποκατηγορίες: οντολογίες στόχου (task ontologies), οντολογίες περιοχών (domain ontologies) και γενικές ή κοινές οντολογίες (general ή common ontologies).

- Οντολογίες επικοινωνίας (λέω & ρωτάω) για διαμοιρασμό της γνώσης (communication (tell & ask) ontologies for sharing knowledge).

- Οντολογίες ευρετηρίου για την ανάκτηση περιπτώσεων ετα-οντολογίες (meta-ontologies), οι οποίες για τους συγγραφείς είναι ισοδύναμες με τις οντολογίες αναπαράστασης γνώσης knowledge representation ontologies).

Οι Van Heijst et al., (2003) ταξινόμησαν τις οντολογίες σύμφωνα με: α) το βαθμό και τον υπό δόμησης του μοντέλου που αποδίδει το φαινόμενο και β) του θέματός του. Με βάση το α) οι οντολογίες διακρίνονται σε τρεις κατηγορίες: τις ονοματολογικές ή ορολογικές οντολογίες (terminological ontologies) π.χ. λεξικά, τις οντολογίες πληροφοριών information ontologies) π.χ. σχήματα βάσεων δεδομένων και τις οντολογίες μοντελοποίησης της γνώσης (knowledge modeling ontologies) που αποσαφηνίζουν τις αντιλήψεις της γνώσης. Με βάση το (β) προσδιορίζονται οι ακόλουθες τέσσερις κατηγορίες: οντολογίες αναπαράστασης (representation ontologies), γενικές οντολογίες (generic ontologies), οντολογίες περιοχών (domain ontologies) και οντολογίες εφαρμογών (applications ontologies).

Ο Guarino (1998) ταξινόμησε τους τύπους των οντολογιών σύμφωνα με το επίπεδο εξάρτησής τους με κάποιο συγκεκριμένο στόχο ή άποψη. Σύμφωνα με αυτή την ταξινόμηση (Σχήμα 2) έχουμε:

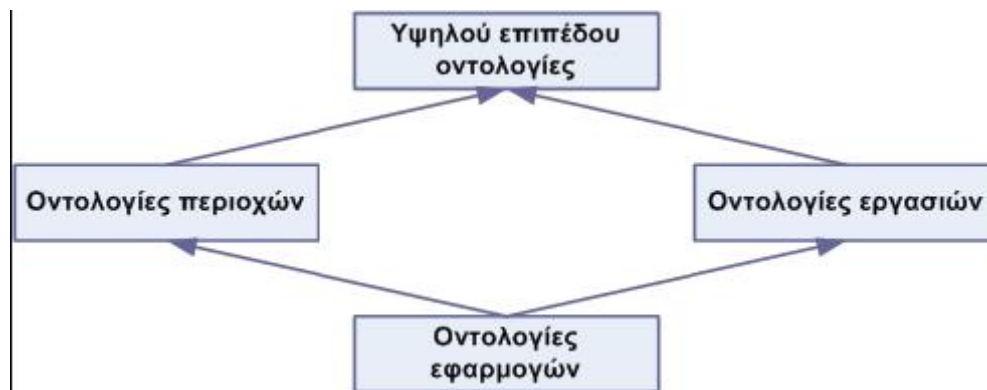
- Υψηλού επιπέδου οντολογίες (top-level ontologies) που περιγράφουν πολύ γενικές

έννοιες (όπως ο χώρος, ο χρόνος, το γεγονός) και οι οποίες είναι ανεξάρτητες από ένα συγκεκριμένο πρόβλημα ή περιοχή γνώσης. Είναι λογική η ύπαρξη ενοποιημένης οντολογίας υψηλού επιπέδου για μεγάλες κοινότητες χρηστών.

- Οντολογίες περιοχών (domain ontologies) που περιγράφουν ένα λεξιλόγιο σχετικό με ένα γενικό πεδίο γνώσης εξειδικεύοντας τις έννοιες που εισήχθησαν στην οντολογία υψηλού επιπέδου.

- Οντολογίες εργασιών (tasks ontologies) που περιγράφουν ένα λεξιλόγιο σχετικό με μια γενική δραστηριότητα ή εργασία εξειδικεύοντας την οντολογία υψηλού επιπέδου.

- Οντολογίες εφαρμογών (application ontologies) οι οποίες είναι οι πιο συγκεκριμένες οντολογίες. Οι έννοιες σε οντολογίες εφαρμογών συχνά ανταποκρίνονται σε ρόλους οντοτήτων της περιοχής γνώσης καθώς εκτελούν συγκεκριμένη δραστηριότητα.



Σχήμα 5. Ταξινόμηση οντολογιών σύμφωνα με τον Guarino (1998).

Guarino G.,(1998), An Environment for Merging and Testing Large Ontologies. In Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR-2000), Breckenridge, Colorado,. San Francisco: Morgan Kaufman, pp.14-18

Τέλος, οι Lassila & McGuinness (2001) ταξινόμησαν τις οντολογίες σύμφωνα με τις πληροφορίες που αυτές χρειάζονται για να εκφράσουν τον πλούτο της εσωτερικής δομής τους. Συγκεκριμένα, όρισαν τις ακόλουθες κατηγορίες: ελεγχόμενα λεξιλόγια (controlled vocabularies), γλωσσάρια (glossaries), θησαυρούς (thesauri), ανεπίσημες is-a ιεραρχίες (informal is-a hierarchies), επίσημες is-a ιεραρχίες (formal is-a hierarchies), επίσημα στιγμιότυπα (formal instances), πλαίσια (frames), περιορισμένων τιμών (value restriction) και γενικών λογικών περιορισμών (general logical constraints).

2.3 ΑΠΟ ΤΟ WEB MINING ΣΤΟ SEMANTIC WEB MINING

Ο όρος Web Mining πρωτοεμφανίστηκε στην εργασία και είναι η εφαρμογή τεχνικών Data Mining στο περιεχόμενο, τη δομή και τη χρήση δεδομένων Web με στόχο την ανακάλυψη «κρυμμένης» γνώσης (patterns ή προτύπων). Το Web Mining μπορεί να ωφεληθεί από τη συγκεκριμένη δομή των δεδομένων (π.χ. πίνακες των βάσεων δεδομένων) αλλά μπορεί επίσης να εφαρμοστεί και σε ημι-δομημένα ή αδόμητα δεδομένα π.χ. κείμενα emails(Kosala & Blockeel:2000).

Μια ευρέως αποδεκτή ταξινόμια του Web Mining το διακρίνει στις τρεις κατηγορίες που φαίνονται στο Σχήμα 3: το Web Content Mining, το Web Structure Mining και το Web Usage Mining (Kosala & Blockeel:2000).



Σχήμα 6. Οι κατηγορίες του semantic web mining.

Πηγή: Kosala G & Blockeel P.,(2000), Intelligent agents for web-based tasks: An advice-taking approach, Scientific American.

2.3.1 Web Content Mining

Το Web Content Mining αφορά στην εξόρυξη γνώσης από δεδομένα περιεχομένου Web. Στην ουσία αποτελεί μια μορφή Text Mining. Εστιάζει στο κείμενο των ιστοσελίδων που περιέχεται στον HTML, script, κλπ. κώδικα και μπορεί να εκμεταλλευτεί την ημι-δομημένη φύση αυτών των τύπων κειμένων. Τα HTML tags καθώς και τα XML markup των web pages ενσωματώνουν πληροφορίες που αφορούν όχι μόνο στη διαμόρφωση (layout) αλλά και στη λογική δομή. Το Web Content Mining μπορεί να χρησιμοποιηθεί για να ανιχνεύσει εμφανίσεις όρων στα κείμενα.

Για παράδειγμα, μπορεί να βρεθεί ότι η λέξη «champion» συναντιέται πολύ συχνά με τη λέξη «Olympiacos» στην Ελλάδα και με τη λέξη «Real» στην Ισπανία. Επίσης, μπορούν να ανακαλυφθούν τάσεις στη διάρκεια του χρόνου υποδεικνύοντας μια άνοδο ή μια πτώση ενδιαφέροντος για ορισμένα θέματα π.χ. η γλώσσα προγραμματισμού Java. Ένας άλλος τομέας εφαρμογής αποτελεί η ανίχνευση γεγονότων δηλαδή ο προσδιορισμός ιστοριών στα

συνεχιζόμενα δελτία ειδήσεων που αντιστοιχούν σε νέα ή προηγούμενα μη αναγνωρισμένα γεγονότα (Missikoff, M. et al:2003).

2.3.2 Web Structure Mining

Το Web Structure Mining αφορά στην εξόρυξη γνώσης από δεδομένα δόμησης Web. Η αρχική πηγή των δεδομένων από την οποία εξάγεται η γνώση είναι ένα σύνολο σελίδων και μπορεί να αφορά ένα απλό web site ή και ολόκληρο το Web. Στην ουσία εκμεταλλεύεται τις πρόσθετες πληροφορίες που περιλαμβάνονται (συχνά όχι με ρητό αλλά με υπονοούμενο τρόπο) στη δομή του υπερκειμένου (hypertext). Ένας σημαντικός τομέας εφαρμογής είναι ο προσδιορισμός της σχετικότητας των διαφορετικών σελίδων που εμφανίζονται εξίσου κατάλληλες όταν αναλύονται σε σχέση με το περιεχόμενό τους ή όταν αντιμετωπίζονται ξεχωριστά.

Η δομή των συνδέσμων υπερκειμένου χρησιμοποιείται σε διάφορες προσεγγίσεις π.χ. ο Kleinberg (1999) αναλύει την τοπολογία των υπερσυνδέσμων και ορίζει τα hubs και τα authorities, ενώ η μηχανή αναζήτησης Google οφείλει την επιτυχία της στον αλγόριθμο PageRank που καθορίζει τη σχετικότητα μιας σελίδας με βάση άλλες σελίδες που «αναφέρεται» (η σχετικότητα μιας σελίδας αυξάνεται με τον αριθμό των συνδέσμων προς αυτή από άλλες σελίδες και ειδικότερα από άλλες σχετικές σελίδες) (Βασίλης T:2007),.

Απλές σελίδες μπορούν επίσης να αναλυθούν όσον αφορά στη δομή τους, η οποία δίνει πληροφορίες για τη λειτουργία τους. Ο Pearce, M(1997), διακρίνει πέντε τύπους web pages α) «head pages» αποτελούν τα σημεία εισόδου στο site, β) «navigation pages» περιέχουν συνδέσμους και λίγη πληροφορία, γ) «content pages» περιέχουν μικρό αριθμό συνδέσεων και κυρίως οι χρήστες τις επισκέπτονται για το περιεχόμενό τους, δ) «look-up pages» έχουν αρκετές συνδέσεις εισόδου, λίγες εξερχόμενες συνδέσεις και κανένα σημαντικό

περιεχόμενο, όπως οι σελίδες που χρησιμοποιούνται για να παρέχουν έναν ορισμό ή κάποιο ακρωνύμιο και (ε) «personal pages» έχουν πολύ διαφορετικά χαρακτηριστικά και όχι σημαντική κυκλοφορία.

Από μια άλλη σκοπιά, οι σύνδεσμοι αποτελούν μέρος του περιεχομένου μιας ιστοσελίδας. Αυτό ισχύει ιδιαίτερα για τις σελίδες στις οποίες οι σύνδεσμοι όπως και άλλα στοιχεία τους, είναι σημασιολογικά χαρακτηρισμένα. Για αυτό το λόγο σε πολλές περιπτώσεις, οι κατηγορίες Web Content Mining και Web Structure Mining αντιμετωπίζονται σαν μία. Συχνά χρησιμοποιούνται για να εξάγουν πληροφορίες από ιστοσελίδες, να καθορίσουν λέξεις κλειδιά περιγράφοντας το περιεχόμενο, να αναθέσουν ιστοσελίδες σε ένα μοντέλο, να ανιχνεύσουν τα γεγονότα ή τις διαδρομές σε πόρους του Web, κλπ.

2.3.3 Web Usage Mining

Το Web Usage Mining αφορά στην εξόρυξη γνώσης από δεδομένα χρήσης Web. Αυτό σημαίνει ότι κυρίως χρησιμοποιούνται τα αρχεία logs τα οποία καταγράφουν τα αιτήματα των επισκεπτών ενός web site . Το περιεχόμενο και η δομή των web pages και ιδιαίτερα ενός web site απεικονίζουν τις προθέσεις των σχεδιαστών των σελίδων καθώς και την αρχιτεκτονική των πληροφοριών. Η πραγματική συμπεριφορά των χρηστών αυτών των σελίδων μπορεί να αποκαλύψει πρόσθετη γνώση αποκαλύπτοντας επιπλέον συσχετίσεις.

Αρχικά οι σχέσεις μπορούν να εξαχθούν από τα δεδομένα χρήσης χωρίς να έχει σχεδιαστεί κάποια ιδιαίτερη δομή ή οργάνωση. Για παράδειγμα, έστω ο ηλεκτρονικός κατάλογος ενός online καταστήματος ο οποίος δεν διαθέτει κάποια ιεραρχική οργάνωση με βάση την κατηγορία των προϊόντων, τον κατασκευαστή, κλπ. Η εφαρμογή εξόρυξης στα δεδομένα των επισκέψεων σε αυτό το site μπορεί να δείξει ότι οι περισσότεροι χρήστες π.χ.

σε ποσοστό 80% που ενδιαφέρθηκαν για το προϊόν A ενδιαφέρθηκαν επίσης για το προϊόν B.

Το «ενδιαφέρον» μπορεί να καθορίζεται από τα αιτήματα για τις σελίδες περιγραφής των προϊόντων, ή την τοποθέτηση του συγκεκριμένου προϊόντος στο καλάθι αγορών, κλπ. Αυτοί οι κανόνες αποτελούν βασικές στρατηγικές cross-selling και up-selling ενός site ηλεκτρονικού εμπορίου π.χ. όταν ένας νέος χρήστης ενδιαφέρεται για το προϊόν A τότε θα λάβει μια σύσταση για το προϊόν B (Mobasher et al., 2000), (Lin et al., 2002). Κατά δεύτερον, σχέσεις μπορούν να εξαχθούν από δεδομένα χρήσης σε περιπτώσεις που διαφορετικές σχέσεις έχουν προδιαγραφεί. Για παράδειγμα, η εξόρυξη ακολουθιών μπορεί να δείξει ότι οι περισσότεροι από εκείνους τους χρήστες που επισκέφτηκαν τη σελίδα C πήγαν αργότερα στη σελίδα (Roberts, L:1997).

Αυτό μπορεί να ερμηνευθεί ότι οι επισκέπτες επιθυμούν να φθάσουν στη D από τη C, αλλά ότι αυτό δεν είχε προβλεφθεί στην αρχιτεκτονική και ως εκ τούτου δεν υπάρχει απευθείας σύνδεσμος από τη C στο D. Αυτή η γνώση μπορεί να χρησιμοποιηθεί για τη στατική βελτίωση του site για όλους τους χρήστες (προσθέτοντας μια σύνδεση από τη C στη D) ή για την παραγωγή δυναμικών εξατομικευμένων συστάσεων για το υποσύνολο των χρηστών που πηγαίνουν στη C π.χ. «μπορεί να σας ενδιαφέρει να δείτε επίσης τη D».

Είναι επίσης χρήσιμο να συνδυαστεί το Web Usage Mining με το Web Content Mining και το Web Structure Mining προκειμένου να «κατανοηθούν» τα πιο συχνά παρατηρημένα μονοπάτια και οι σελίδες που τα απαρτίζουν. Αυτό μπορεί να γίνει με διάφορες μεθόδους όπως: ταξινόμηση των σελίδων από την u940 άποψη μιας προκαθορισμένης οντολογίας, εξαγωγή των λέξεων κλειδιών που περιέχονται σε αυτές τις σελίδες και εν συνεχεία ονομασία των ομάδων των λέξεων κλειδιών που αντιπροσωπεύουν τα συχνότερα μονοπάτια. Η οντολογία μπορεί να δημιουργηθεί με το χέρι ή να εξαχθεί (ήμι)αυτόματα, ενώ και η ταξινόμηση των σελίδων από την άποψη της οντολογίας μπορεί να γίνει με

διάφορους τρόπους.

Στην πλειοψηφία των περιπτώσεων, οι εφαρμογές e-learning βασίζουν την εξατομίκευση στη τεχνική Web Usage Mining, η οποία αναλαμβάνει το ρόλο συλλογής και εξαγωγής όλων των δεδομένων που απαιτούνται για την κατασκευή και τη συντήρηση των προφίλ των εκπαιδευόμενων βασισμένη στη συμπεριφορά κάθε χρήστη όπως αυτή καταγράφεται στα server logs .

Από την άλλη πλευρά, ο συνδυασμός του Σημασιολογικού Παγκόσμιου Ιστού Πληροφοριών και του Web Mining έχει δημιουργήσει μια νέα και γρήγορα αναδυόμενη ερευνητική περιοχή γνωστή ως Semantic Web Mining. Η ιδέα πίσω από τη χρησιμοποίηση του Σημασιολογικού Παγκόσμιου Ιστού Πληροφοριών για τη δημιουργία προσωποποιημένων Web εμπειριών είναι να βελτιωθεί η διαδικασία του Web Mining με την εκμετάλλευση των νέων σημασιολογικών δομών (Μάρκελλος Κ:2006).

Συνδέοντας νοηματικά ένα σύνολο από πόρους (κείμενα, ιστοσελίδες, εικόνες, ανθρώπους, έννοιες, κλπ.) είναι δυνατή η μετάβαση σε ένα πιο εκφραστικό και σημασιολογικά πλούσιο Web, όπου μπορούμε σταδιακά να ορίζουμε ένα σύνολο από νέες σχέσεις μεταξύ των δεδομένων που μέχρι τώρα μπορούσαμε μόνο έμμεσα να υπονοήσουμε.

Αυτό θα επιτρέψει την εφαρμογή εκλεπτυσμένων τεχνικών mining που απαιτούν δομημένη είσοδο για αποδοτική ολοκλήρωση των δεδομένων, καλύτερη διαχείρισή τους και θα οδηγήσουν σε πιο ευφείς e-learning εφαρμογές που θα «αυτοπροσαρμόζονται» ώστε να ταιριάζουν καλύτερα σε κάθε ξεχωριστό χρήστη.

2.4 ΑΞΙΟΛΟΓΗΣΗ ΣΥΣΤΗΜΑΤΩΝ ΠΡΟΣΩΠΙΚΟΠΟΙΗΣΗΣ

Η προσωπικοποίηση αποτελεί σήμερα αντικείμενο πολλών εκφάνσεων της κοινωνίας της πληροφορίας. Τέτοιες εκφάνσεις είναι το ηλεκτρονικό εμπόριο αλλά και η υποστηριζόμενη και αναπτυσσόμενη από τους υπολογιστές εκπαίδευση. Συγκεκριμένα αναφέρεται σύμφωνα με τους B. Mobasher, R. Cooley, J. Srivastava το 2000 στο άρθρο τους «*Automatic Personalization Based on Web Usage Mining. Communications of the ACM*» ότι: «η προσωπικοποίηση ως ζητούμενο πολλών εκφράσεων στη σημερινή κοινωνία της πληροφορίας, όπως είναι το ηλεκτρονικό εμπόριο (*electronic commerce*) και η βοηθούμενη από υπολογιστές εκπαίδευση, καλύπτει ένα ευρύ φάσμα: από την απλή εμφάνιση του ονόματος του χρήστη σε ένα δικτυακό τόπο μέχρι την προσαρμογή της «συμπεριφοράς» μιας εφαρμογής με βάση καλά οργανωμένα μοντέλα προτιμήσεων και αναγκών του τελικού χρήστη»(Mobasher, R. Cooley, J. Srivastava: 2000)

Η προσωπικοποίηση σύμφωνα με τους U. Manber, A. Patel, J. Robison αποτελεί τη μόνη διαδικασία, η οποία δίνει προσωπικότητα στο Ίντερνετ και ειδικά στις εμπορικές συναλλαγές, ενώ το καθιστά και ένα εργαλείο με το οποίο ο χρήστης νοιώθει εξοικειωμένος, φιλικός ενώ του έχει μεγαλύτερη εμπιστοσύνη, λόγω του τρόπου με τον οποίο αντιμετωπίζει. Οι συγγραφείς συνεχίζουν ότι λέγοντας ότι μέσα από αυτή τη σχέση και κάτω από τις όποιες υπάρχουσες συνθήκες, οι επιθυμίες του «πελάτη» επιχειρείται να κατανοηθούν και να ικανοποιηθούν(U. Manber, A. Patel, J. Robison: 2000).

Η έννοια της προσωπικοποίησης περιλαμβάνει την μοντελοποίηση διαδικτυακών αντικειμένων (προϊόντα ή ιστοσελίδες) και υποκειμένων (χρήστες), την κατηγοριοποίηση τους, την εύρεση ομοιοτήτων ανάμεσα τους και τον καθορισμό των απαραίτητων ενεργειών που τελικά θα προταθούν, υλοποιώντας την προσωπικοποίηση.

Στη κλασσική πλοήγηση η προσωπικοποίηση εκφράζεται μέσα από αλλαγές στο περιεχόμενο, την εμφάνιση και τη δομή ενός δικτυακού τόπου. Στο ηλεκτρονικό εμπόριο σύμφωνα με το μοντέλο των Paul B. Kantor et al το 2000 η αξιολόγηση και μελέτη των

μέσων προσωποποίησης γίνεται μέσα από τα ακόλουθα δεδομένα:

1. Αναγνώριση των προτιμήσεων του υποψήφιου αγοραστή.
2. Εντοπισμός πιθανόν του ελλείψεων και προτιμήσεων.

Από την άλλη στο τομέα της εκπαίδευσης με βάση και πάλι τους ίδιους συγγραφείς η αξιολόγηση γίνεται με βάση:

1. Τη δυνατότητα εύρεσης των δυνατών και των αδύνατων σημείων του κάθε μαθητή ώστε να προσαρμοστεί κατάλληλα τόσο η ύλη που θα του διδαχτεί όσο και ο τρόπος διδασκαλίας
2. Τη ξεκάθαρη παράθεση της ύλης και των δεδομένων αυτής.

Στην επόμενη ενότητα θα προσπαθήσουμε να αξιολογήσουμε με βάση τα άνωθεν κριτήρια συστήματα προσωποποίησης.

Στη συνέχεια θα αξιολογήσουμε τα ακόλουθα συστήματα έχοντας ως γνώμονα τα κριτήρια που παραθέσαμε παραπάνω. Η αξιολόγηση θα γίνει αρχικά θεωρητικά και στη συνέχεια πρακτικά:

1. Manual Decision Rule Systems: Αναφέρονται σε συστήματα, όπως αυτό της Broad vision, όπου οι κατασκευαστές τους χρησιμοποιούν κανόνες (rules) βασισμένους σε δημογραφικά στοιχεία (user demographics) και στατικά προφίλ (static profiles) χρηστών, που συνήθως δημιουργούνται κατά την εγγραφή των χρηστών (user registration) στο συγκεκριμένο σύστημα. Ακόμα οι κανόνες μπορούν να προκύψουν και από αρχεία αλληλεπίδρασης του χρήστη με το σύστημα (session histories). Οι κανόνες που τελικά θα δημιουργηθούν, χρησιμοποιούνται για να αλλάξουν το περιεχόμενο, τη δομή ή την εμφάνιση των στοιχείων που θα παρουσιαστούν εν συνεχεία στο χρήστη. Σε αυτή την κατηγορία ανήκει το σύστημα προσωποποίησης του Yahoo! Που θα αξιολογηθεί παρακάτω με βάση τα κριτήρια που παραθέσαμε στη προηγούμενη ενότητα

2. Content-Based Systems: Είναι συστήματα όπως το Web Watcher, τα οποία αναφέρονται στο εντοπισμό ομοιοτήτων μεταξύ των αντικειμένων ενδιαφέροντος του χρήστη και στα λεγόμενα προφίλ του. Τα προφίλ προκύπτουν από πληροφορίες που έχουν δώσει ο χρήστες, είτε από τη πληροφόρηση που συλλέγει ο χρήστης και τη μετουσιώνει στο προφίλ του χρήστη. Τα προφίλ των χρηστών μπορούν να έχουν προκύψει είτε από πληροφορίες που έχει δώσει ο ίδιος ο χρήστης (explicit data) είτε από πληροφορίες που έχουν συλλέγει έμμεσα και τον αφορούν (implicit data)(Terpsidis I et al:2000)

Οι τεχνικές content-based filtering έχουν τις ρίζες τους στην ανάκτηση πληροφοριών (information retrieval) και την έρευνα που γίνεται σχετικά με case-based reasoning. Η επιτυχία τους βασίζεται στην δυνατότητα να παρασταθούν με μεγάλη ακρίβεια αντικείμενα που πρόκειται να προταθούν, υπό τη μορφή των κατάλληλων γνωρισμάτων περιεχομένου (content features). Παράλληλα είναι αναγκαία και η αναπαράσταση των πληροφοριών από το προφίλ του χρήστη χρησιμοποιώντας το ίδιο σύνολο γνωρισμάτων περιεχομένου. Ο συσχετισμός ανάμεσα σε ένα αντικείμενο και σε ένα χρήστη-«στόχο» είναι ανάλογος με την ομοιότητα αυτού του αντικειμένου με το προφίλ του συγκεκριμένου χρήστη(Terpsidis I et al:2000).

Το σημαντικότερο πρόβλημα με τις συγκεκριμένες μεθόδους είναι η ανάγκη για μια καλή απεικόνιση των γνωρισμάτων περιεχομένου. Αναλυτικά τα προβλήματα που απορρέουν στο συγκεκριμένο μηχανισμό είναι τα ακόλουθα:

1. Η εύρεση της απεικόνισης μπορεί να αποδειχτεί προβληματική και χρονοβόρα. Σε μερικές περιπτώσεις δεν είναι δυνατό να βρεθεί μια τέτοια απεικόνιση, σχετική με το χρήστη.

2. Δυσκολία στην επιλογή των αντικειμένων τα οποία προτείνουν στο χρήστη. Αυτό γίνεται διότι οι προτάσεις που κάνει ο μηχανισμός βασίζονται σε αντικείμενα παρόμοια με αντικείμενα που ο χρήστης επέλεξε στο παρελθόν, το προφίλ του χρήστη και τα αντικείμενα που αυτό περιλαμβάνει, περιορίζει αυτόματα το πλήθος των προτάσεων που

μπορούν να του γίνουν στο μέλλον. Το πρόβλημα αυτό είναι εντονότερο σε περιπτώσεις νέων χρηστών των οποίων τα προφίλ περιέχουν περιορισμένο αριθμό αντικειμένων. Ο μηχανισμός δυσκολεύεται να ανανεωθεί και να βρει νέα αντικείμενα.

3. Collaborative filtering Systems: Αναφέρονται σε συστήματα, όπως το Firefly και το Net Perceptions τα οποία δέχονται πληροφορίες με τη ρητή μορφή προτιμήσεων και βαθμολογιών (ratings) ενός χρήστη, και στη συνέχεια με τη βοήθεια μιας μηχανής συσχετίσεων (correlation engine) επιστρέφουν αποτελέσματα που θεωρητικά ταιριάζουν σε μεγάλο βαθμό με τις προτιμήσεις του τελικού χρήστη.

Η βασική ιδέα πίσω από τα συστήματα collaborative filtering είναι η άντληση πληροφοριών από τις εμπειρίες μιας ολόκληρης κοινωνίας χρηστών και όχι μόνο από έναν μεμονωμένο άτομο. Τυπικά κάθε χρήστης που αποτελεί «στόχο» του συστήματος (target user) σχετίζεται με ένα σύνολο από άλλους χρήστες, των οποίων τα προφίλ παρουσιάζουν τις μεγαλύτερες ομοιότητες (nearest-neighbor users) με τις πληροφορίες που υπάρχουν για το χρήστη-«στόχο». Τα πλεονεκτήματα του συγκεκριμένου συστήματος είναι τα ακόλουθα(Βαζόλης Μ:2000,5-10)

1. Οι χρήστες που επιλέγονται είναι εκείνοι που εμφανίζουν το μεγαλύτερο συσχετισμό με το χρήστη-«στόχο». Οι χρήστες αυτοί στη συνέχεια αποτελούν “recommendation partners” για το χρήστη-«στόχο» και αντικείμενα που εμφανίζονται στα προφίλ τους (και όχι στο προφίλ του χρήστη-«στόχου») μπορούν να προταθούν σε αυτόν.

2. Οι τελικές προτάσεις που θα γίνουν στο χρήστη-«στόχο» προκύπτουν με βάση την ομοιότητα ανάμεσα στους χρήστες (user similarity) και όχι την ομοιότητα ανάμεσα στα αντικείμενα (item similarity) - όπως γίνεται στα content based συστήματα.

Στη προσπάθεια μας να συγκρίνουμε τα συστήματα collaborative filtering και content based μεθόδων βλέπουμε ότι τα συστήματα collaborative filtering παρουσιάζουν μια σειρά προτερημάτων σε σχέση με τα content based συστήματα . Αυτά είναι:

1. Τα συστήματα collaborative filtering δεν εστιάζουν στην εύρεση μιας αναπαράστασης των γνωρισμάτων περιεχομένου (content features representation) κάτι που έχει ως αποτέλεσμα η διαδικασία εύρεσης γνώσης (knowledge engineering) να είναι πολύ πιο απλή.

2. Η ποιότητα των συστημάτων collaborative filtering βελτιώνεται όσο το μέγεθος του πληθυσμού των χρηστών αυξάνεται, επειδή οι προτάσεις που γίνονται ωφελούνται από την ποικιλία που παρουσιάζουν τα πολλά και διαφορετικά προφίλ χρηστών. Αυτό δεν συμβαίνει στα content based συστήματα, όπου ανεξάρτητα από τον αριθμό των χρηστών, ένας νέος χρήστης δεν θα μπορεί να πάρει ικανοποιητικές προτάσεις, εξαιτίας του περιορισμένου του προφίλ.

Τα μειονεκτήματα των collaborative filtering είναι τα ακόλουθα(Βαζόλης Μ:2000,5-10):

1. Δεν είναι κατάλληλα όταν πρόκειται να κρίνουν και να προτείνουν στο χρήστη καινούργια αντικείμενα γιατί οι τεχνικές αυτές μπορούν να προτείνουν μόνο αντικείμενα που έχουν ήδη κριθεί από άλλους χρήστες. Αυτό σημαίνει μεγάλες καθυστερήσεις οι οποίες σχετίζονται με τη διαδικασία εύρεσης άλλων χρηστών με παρόμοια ενδιαφέροντα. Οι Mobasher B, Cooley P και Srivastava J το 2000 αναφέρουν το πρόβλημα αυτό ως “the latency problem”.

2. Τα collaborative filtering συστήματα δεν παρουσιάζουν ικανοποιητική απόδοση στην περίπτωση του «ασυνήθιστου χρήστη» (unusual user). Συγκεκριμένα, δεν υπάρχουν εγγυήσεις ότι θα βρεθούν οι κατάλληλοι recommendation partners για ένα χρήστη, ειδικά αν δεν υπάρχει ικανοποιητική κάλυψη ανάμεσα στο προφίλ του και στα προφίλ των υπολοίπων χρηστών.

4. Web Usage Mining Systems: Αναφέρονται σε συστήματα που ενσωματώνουν τεχνικές για την εξεύρεση μοτίβων (pattern discovery) απο δεδομένα χρήσης του Διαδικτύου (Web usage data).

Τα συστήματα αυτά τρέχουν μια σειρά από αλγορίθμους εξόρυξης δεδομένων (data mining) πάνω σε δεδομένα που προέρχονται από δικτυακούς τόπους που επισκέφτηκε ένας χρήστης, ώστε να ανακαλύψουν το αληθινό προφίλ του. Το θετικό με αυτά τα συστήματα είναι ότι δεν βασίζονται σε υποκειμενικές περιγραφές των χρηστών για τους εαυτούς τους. Απεναντίας, τα προφίλ χρηστών προκύπτουν με δυναμικό τρόπο από τα μοτίβα που εμφανίζει η συμπεριφορά τους με αποτέλεσμα η απόδοση του συστήματος να μην μειώνεται με το χρόνο, όσο τα προφίλ παλαιώνουν. Τα στάδια που τα καθορίζουν σύμφωνα με τους Mobasher B, Cooley R, Srivastava J. Το 2000 είναι τα ακόλουθα (Mobasher B, Cooley R, Srivastava J:2000)

1. Το πρώτο στάδιο αφορά την προετοιμασία των δεδομένων (data preparation). Το βήμα αυτό είναι προπαρασκευαστικό και έχει ως στόχο να ξεχωρίσει από ένα πλήθος δεδομένων χρήσης Διαδικτύου που αφορούν το χρήστη, εκείνα τα δεδομένα που θα βοηθήσουν περισσότερο το σύστημα στη συνέχεια.

2. Το δεύτερο στάδιο σχετίζεται με την «ανακάλυψη» των προφίλ χρήσης (discovery of usage profiles). Συγκεκριμένα, τα αρχεία δεδομένων που αποκτήθηκαν στο προηγούμενο στάδιο μπορούν πλέον να χρησιμοποιηθούν από μια σειρά αλγορίθμων εξόρυξης όπως αυτοί που αφορούν την ανακάλυψη κανόνων συσχέτισης (association rules) ή σειριακών μοτίβων (sequential patterns), την ομαδοποίηση (clustering) και την ταξινόμηση (classification). Κρίσιμο βήμα για το στάδιο αυτό είναι η χρησιμοποίηση των όποιων μοτίβων ανακαλύφθηκαν από τους αλγορίθμους εξόρυξης δεδομένων για την εξαγωγή ποιοτικών και χρήσιμων «αθροιστικών προφίλ» (aggregate profiles).

3. Το τρίτο και τελευταίο στάδιο αφορά την κρίσιμη μετάβαση από τα προφίλ στις προτάσεις (from profiles to recommendations). Το βήμα αυτό πρέπει να εκτελεστεί online σε αντίθεση με τα 2 προηγούμενα που μπορούν να εκτελεστούν offline. Για την υλοποίηση του χρησιμοποιείται η μηχανή προτάσεων (recommendation engine) του συστήματος, η οποία πρέπει να παράγει ένα σύνολο προτάσεων (recommendation set) με βάση τα ενεργά

δεδομένα χρήσης του χρήστη. Το σύνολο προτάσεων περιλαμβάνει συνήθως αντικείμενα - συνδέσμους, κείμενα, προϊόντα, διαφημίσεις- τα οποία ταιριάζουν στο μεγαλύτερο δυνατό βαθμό με το «αθροιστικό προφίλ» που έχει δημιουργηθεί για το συγκεκριμένο χρήστη. Η παραγωγή των προτάσεων πρέπει να είναι μια online διαδικασία, που να δίνει άμεσα αποτελέσματα στο χρήστη.

ΚΕΦΑΛΑΙΟ 3

ΕΦΑΡΜΟΓΕΣ ΕΞΑΤΟΜΙΚΕΥΣΗΣ ΣΤΙΣ ΗΛΕΚΤΡΟΝΙΚΕΣ ΜΑΘΗΣΙΑΚΕΣ ΚΟΙΝΟΤΗΤΕΣ

3.1 ΣΥΣΧΕΤΙΣΗ ΠΡΟΤΥΠΩΝ E-LEARNING

Αν θεωρήσουμε ότι «Η ηλεκτρονική μάθηση είναι η άμεση χρονικά εκπαίδευση που ενσωματώνεται σε αλυσίδες αξιών υψηλής ταχύτητας. Είναι η παράδοση του εξατομικευμένου, περιεκτικού, δυναμικού περιεχομένου εκμάθησης σε πραγματικό - χρόνο που βοηθά την ανάπτυξη των κοινοτήτων της γνώσης, που συνδέει τους αρχαίους και τους επαγγελματίες με τους εμπειρογνώμονες» (Fensel, D:2002) φαίνεται ότι το Semantic Web Mining ικανοποιεί τις βασικές απαιτήσεις του. Καταρχήν, το εκπαιδευτικό υλικό μπορεί να σχολιαστεί σημασιολογικά και να συνδεθεί με την οντολογία.

Έτσι κάθε φορά που θα προκύπτει μια νέα ανάγκη μάθησης να είναι δυνατός και εύκολος ο συνδυασμός του υλικού αυτού σε μια νέα σειρά μαθημάτων. Θα δημιουργηθεί λοιπόν ένα εξατομικευμένο περιβάλλον για το χρήστη που θα σχετίζεται με τα ενδιαφέροντά του.

Η δυνατότητα αυτή εξυπηρετεί την απαίτηση για διανομή πληροφορίας ατομικά στον εκπαιδευόμενο, με ύλη που εκείνος ορίζει και που αυτή ανταποκρίνεται στις εκάστοτε ανάγκες του.

Επιπλέον, οι διάφοροι web agents θα δίνουν τη δυνατότητα για επερωτήσεις που θα αφορούν σε ετερόκλητα τμήματα της πληροφορίας. Αυτό σημαίνει ότι το περιεχόμενο δεν παρέχεται με «γραμμικό» τρόπο σε συνέχειες, αλλά η πρόσβαση γίνεται σε διάφορα τμήματά του. Η δυνατότητα αυτή σε ένα περιβάλλον e-learning προσφέρει το πλεονέκτημα της γρήγορης και επικεντρωμένης στην ουσία μάθησης και δεδομένου ότι αυτή θα

ανανεώνεται συνεχώς, ο εκπαιδευόμενος επιτυγχάνει συνεχή πρόοδο. Τα παραπάνω αναδεικνύουν την αλληλεπίδραση του χρήστη με το Web, ως βασική προϋπόθεση του e-learning, γεγονός που σημαίνει ότι ο τρόπος λειτουργίας δεν είναι κεντριοποιημένος αλλά κατακευματισμένος (Fensel, D:2002).

Για να επιτευχθεί η διαλειτουργικότητα, η προσβασιμότητα και η επαναχρησιμοποίηση του εκπαιδευτικού περιεχομένου θα πρέπει να ακολουθηθούν κάποιοι κανόνες. Τα πρότυπα e-learning μεταδεδομένων (metadata) αποτελούν τυπικές περιγραφές των όρων που χρησιμοποιούνται για το σημασιολογικό σχολιασμό του εκπαιδευτικού υλικού. Ένα e-learning πρότυπο μπορεί να αναφέρεται στη δομή ενός αντικειμένου μάθησης (με τη χρήση μεταδεδομένων) ή στο διαμοιρασμό του περιεχομένου μέσω της χρήσης μοντέλων δομής περιεχομένου (content structure models), ενώ υπάρχουν πρότυπα που αναφέρονται και στα δύο.

Η διαρκώς αυξανόμενη ανάγκη για τον καθορισμό προτύπων οδήγησε πολλούς γνωστούς Οργανισμούς και Ιδρύματα να ασχοληθούν με το αντικείμενο και να καταθέσουν τις προτάσεις τους:

- ADL/SCORM – Advanced Distributed Learning/Sharable Content Object Reference Model (<http://www.adlnet.org>).
- AICC – Aviation Industry Computer Based Training Committee (<http://www.aicc.org>).
- ARIADNE – Alliance of Remote Instructional Authoring & Distribution Network for Europe (<http://www.ariadne-eu.org>).

Τα πρότυπα αυτά δεν είναι ανεξάρτητα μεταξύ τους αλλά σχετίζονται στενά όπως φαίνεται και στο Σχήμα 7. Η πλειονότητα των προτύπων που υπάρχουν αυτή τη στιγμή, χρησιμοποιούν ως γλώσσα κωδικοποίησης των περιγραφών (specifications) την XML

3.2 Η ΔΙΑΔΙΚΑΣΙΑ ΤΗΣ ΕΞΑΤΟΜΙΚΕΥΣΗΣ ΣΕ ΕΝΑ ΤΥΠΙΚΟ E-LEARNING ΣΥΣΤΗΜΑ

Η ανάγκη για εξατομίκευση στο Web είναι αποτέλεσμα της τεράστιας αύξησης του αριθμού και της πολυπλοκότητας των πληροφοριών και των υπηρεσιών. Η δυνατότητα ενός web site και ειδικότερα ενός e-learning site να χειριστεί τους χρήστες του με διαφορετικό τρόπο και να τους καθοδηγεί επιτυχώς στην πλοήγησή τους αποτελεί κρίσιμο παράγοντα της επιτυχίας του και έχει μετατρέψει την εξατομίκευση σε ένα ιδιαίτερα χρήσιμο εργαλείο τόσο για τα ίδια τα sites όσο και για τους χρήστες τους.

Σαν web personalization μπορούμε να θεωρήσουμε οποιαδήποτε ενέργεια που προσαρμόζει τις πληροφορίες ή τις υπηρεσίες που παρέχονται από ένα δικτυακό τόπο στις ανάγκες ενός συγκεκριμένου χρήστη ή ενός συνόλου χρηστών, χρησιμοποιώντας τη γνώση που πηγάζει από την ανάλυση της συμπεριφοράς πλοήγησης των χρηστών και από τα ενδιαφέροντά τους, σε συνδυασμό με τη δομή και το περιεχόμενο του δικτυακού τόπου. Στόχος είναι η παροχή στους χρήστες των πληροφοριών που θέλουν ή χρειάζονται χωρίς να απαιτεί από αυτούς να το ζητήσουν με ρητό τρόπο. Η διαδικασία της εξατομίκευσης περιλαμβάνει τη μοντελοποίηση των web αντικειμένων (π.χ. σελίδες, μαθήματα, προϊόντα, κλπ.) και των υποκειμένων (π.χ. χρήστες, εκπαιδευόμενοι, πελάτες, κλπ.), την κατηγοριοποίηση των αντικειμένων/υποκειμένων, το ταίριασμα μεταξύ των αντικειμένων ή/και των υποκειμένων και τον προσδιορισμό ενός συνόλου ενεργειών που θα προταθούν για εξατομίκευση.

Τα τελευταία χρόνια, το Web Usage Mining, (Mobasher et al., 2000) χρησιμοποιείται σαν η βασική τεχνική για την επίτευξη εξατομίκευσης στο πεδίο του e-learning. Στόχος

είναι(Μάρκελλος et al., 2004):

- η συλλογή πληροφοριών για το χρήστη (στόχοι, γνωστικό επίπεδο, προτιμήσεις, ενδιαφέροντα) καθώς και άλλων πρόσθετων δεδομένων που αφορούν στο περιβάλλον χρήσης (υλικό, λογισμικό, σύνδεση, εκδόσεις), τα οποία θα χρησιμοποιηθούν για τη δημιουργία και ενημέρωση του προφίλ του,

- η ανάλυση της συμπεριφοράς του χρήστη με βάση τεχνικές όπως κανόνες συσχέτισης (association rules), ομαδοποίηση (clustering), κατηγοριοποίηση (classification), ανακάλυψη προτύπων (patterns discovery),

- η εξαγωγή συμπερασμάτων,

- και τέλος η παραγωγή εξατομικευμένων διεπαφών χρήσης.

Παραδείγματα τέτοιων εξατομικευμένων αλληλεπιδράσεων των εκπαιδευόμενων με μια εφαρμογή ηλεκτρονικής μάθησης θα μπορούσαν να είναι τα ακόλουθα: ένας χρήστης επιθυμεί περισσότερο οπτικό υλικό από ότι κείμενο, ενώ κάποιος άλλος το αντίστροφο ή ένας χρήστης θέλει πρώτα να ολοκληρώσει μια ενότητα και μετά να προχωρήσει σε άλλες, ή προτείνεται σε ένα χρήστη να λύσει μια άσκηση μετά τη μελέτη μιας συγκεκριμένης ενότητας την οποία συνήθως λύνουν οι χρήστες που έχουν παρακολουθήσει την ίδια ενότητα (Brusilovsky: 2001), (Markellou et al., 2004a). Επίσης, θα μπορούσαν να αφορούν στο πλάνο μελέτης, σε προτάσεις για περαιτέρω μελέτη, σε σχετικούς υπερσυνδέσμους, στην εμφάνιση ή στην απόκρυψη υπερσυνδέσμων, κλπ. Για να επιτευχθούν τα ανωτέρω, τη βάση αποτελεί ένα καλό μοντέλο χρηστών (user modeling), στο οποίο φυλάσσονται πάντα ενημερωμένες πληροφορίες για αυτούς δεδομένου ότι οι χρήστες δεν είναι στατικοί και αλλάζουν χαρακτηριστικά και συμπεριφορά με την πάροδο του χρόνου (Barrett, N:1998).

3.3 Η E-LEARNING ONTOΛΟΓΙΑ

Ας υποθέσουμε ένα απλό e-learning σενάριο σύμφωνα με το οποίο ένας φοιτητής ενός Τμήματος «Πληροφορικής» επιθυμεί να διαβάσει εκπαιδευτικό υλικό σχετικά με τις τεχνικές «Web Mining» για να μπορέσει να ετοιμάσει μια εργασία για το μάθημα «Web Data Mining for Business Intelligence». Για το σκοπό αυτό αποφασίζει να εγγραφεί στο πιλοτικό e-learning site που έχουμε κατασκευάσει ώστε να παρακολουθήσει το μάθημα και να πλοηγηθεί στο εκπαιδευτικό περιεχόμενό του. Ένας άλλος φοιτητής, ο οποίος ήδη γνωρίζει τα βασικά θέματα του «Web Mining» θέλει να μάθει για το συγκεκριμένο topic «Web Usage Mining».

Είναι σαφές ότι ο πρώτος φοιτητής θα ακολουθήσει αρχικά το πλάνο του μαθήματος όπως έχει καθοριστεί από το σχεδιασμό του. Στη συνέχεια, ανάλογα με την πορεία του, το πλάνο αυτό μπορεί να αλλάξει για να προσαρμοστεί καλύτερα στην πλοήγησή του. Ο δεύτερος φοιτητής δεν θα παρακολουθήσει από την αρχή το μάθημα αλλά θα στείλει ένα ερώτημα στο site με το οποίο θα ζητάει συγκεκριμένες σελίδες που περιέχουν εκπαιδευτικό υλικό για το «Web Usage Mining». Οι σελίδες που θα του επιστρέψει το σύστημα θα έχουν δείκτες που θα προσδιορίζουν τη σχετικότητα της σελίδας και θα βασίζονται αρχικά στα στοιχεία του προφίλ του π.χ. εκπαιδευτικό επίπεδο, προτιμήσεις, κλπ. Για την υλοποίηση του παραπάνω σεναρίου απαιτείται πληροφορία τόσο για το εκπαιδευτικό υλικό όσο και για τον εκπαιδευόμενο.

Για το σκοπό αυτό επιλέχθηκε η υιοθέτηση μιας οντολογίας η οποία χρησιμοποιείται για να δημιουργήσει μια αναπαράσταση της e-learning περιοχής με τον καθορισμό όλων των εννοιών, των πιθανών σχέσεων μεταξύ τους και άλλων ιδιοτήτων, όρων ή κανονισμών της περιοχής. Η ανάπτυξη της οντολογίας είναι συγγενής με τον καθορισμό ενός συνόλου στοιχείων και της δομής τους. Κατά αυτό τον τρόπο, η οντολογία μπορεί να θεωρηθεί σαν μια βάση γνώσεων που χρησιμοποιείται περαιτέρω για την εξαγωγή χρήσιμης «κρυμμένης»

γνώσης και την παραγωγή εξατομικευμένων μορφών του e-learning συστήματος. Τα βήματα που ακολουθήθηκαν για την κατασκευή της οντολογίας του πιλοτικού e-learning συστήματος βασίζονται σε μια middle-out προσέγγιση η οποία ξεκινά από τον καθορισμό των πιο σημαντικών εννοιών και σταδιακά οδηγείται τόσο προς τη γενίκευση όσο και προς τη συγκεκριμενοποίηση. Τα βήματα αυτά παρουσιάζονται στη συνέχεια:

- Αρχικά καθορίστηκε το πεδίο και ο σκοπός της οντολογίας. Στην περίπτωση μας, το πεδίο εφαρμογής είναι μια σειρά online μαθημάτων για το θεματικό «Web Mining» και ο στόχος είναι η αναπαράσταση όλων των εννοιών που το απαρτίζουν. Πιθανές ερωτήσεις που θα πρέπει να απαντηθούν είναι της μορφής «μετά το θεματικό αντικείμενο Web Usage Mining τι πρέπει να μελετήσω;» ή «τι είδους σύνδεση υπάρχει μεταξύ δυο παραδειγμάτων;», κλπ.

- Στη συνέχεια καθορίστηκαν οι σημαντικοί όροι όπως: Course, Lesson, Teacher, Learner, LearningResource, Example, κλπ.

- Ακολούθησε ο ορισμός των κλάσεων, των υποκλάσεων και της ιεραρχίας τους.

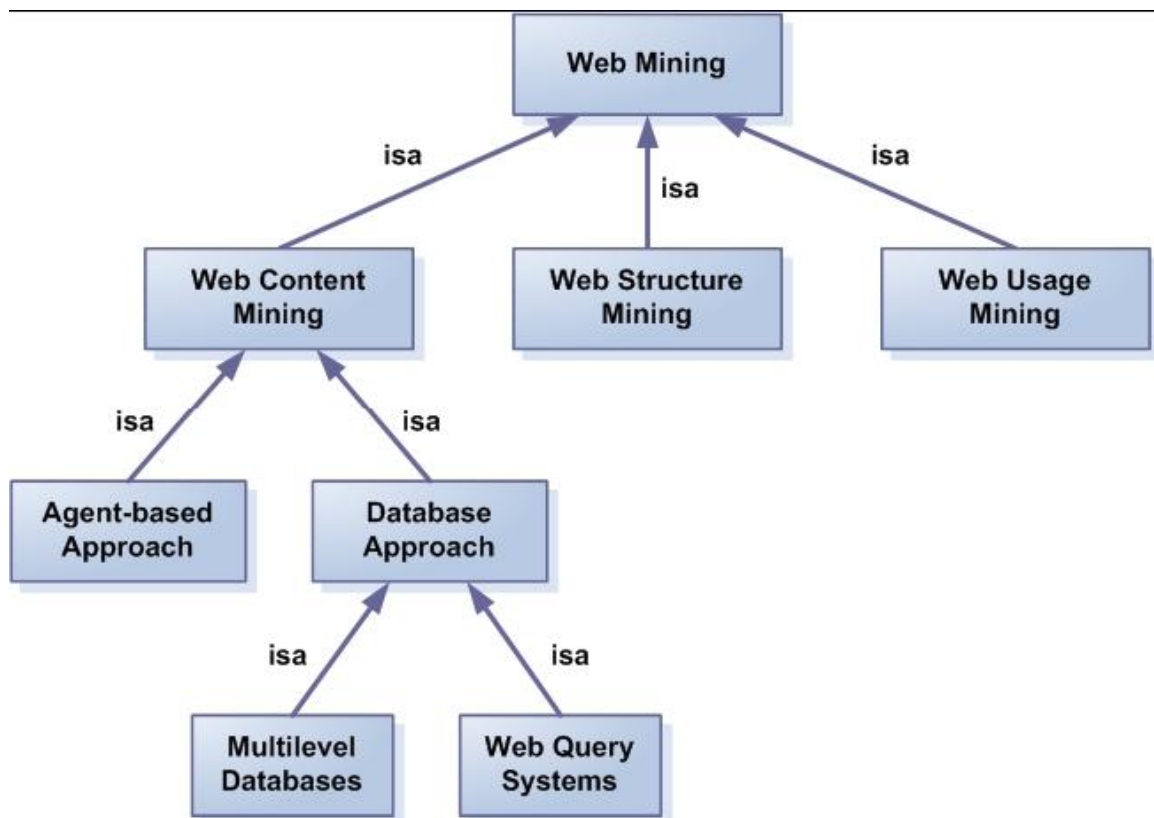
- Το επόμενο βήμα ήταν ο καθορισμός των συσχετίσεων. Στο βήμα αυτό ορίστηκαν οι σχέσεις μεταξύ των κλάσεων της οντολογίας (επίπεδο πολλαπλότητας μιας συσχέτισης και το είδος της τιμής της).

Για λόγους απλότητας της υλοποίησης επιλέχθηκε η χρήση standards όπου αυτό ήταν εφικτό. Αυτά καθορίζουν για παράδειγμα τις ιδιότητες κάθε αντικειμένου και συνήθως τα ταξινομούν σε συγκεκριμένες κατηγορίες.

Για το μάθημα «Web Mining» υιοθετήθηκε η ταξινόμια που προτάθηκε από τον Cooley et al., (1997). Ένα τμήμα της απεικονίζεται στο παρακάτω Σχήμα. Τα συγκεκριμένα θεματικά αντικείμενα περιγράφονται από τα concepts τους και τις αμοιβαίες σχέσεις τους στην περιοχή. Τα concepts της περιοχής μπορεί να έχουν πολύπλοκες δομές. Στο παράδειγμα απεικονίζεται ένα τμήμα της γνώσης της περιοχής που καλύπτει το θεματικό

του «Web Mining» και περιλαμβάνει τη σχέση isa (subConceptOf) μεταξύ των concepts.

Συνεπώς, το concept «Web Mining» έχει sub concepts το «Web Content Mining», το «Web Structure Mining» και το «Web Usage Mining». Τα «Agent-based Approach» και «Database Approach» αποτελούν sub concepts του «Web Content Mining», κ.ο.κ.



Σχήμα 8. Web Mining (a)

Πηγή: Stratakis et al(2003),Ontology Infrastructure, IEEE INTELLIGENT SYSTEMS

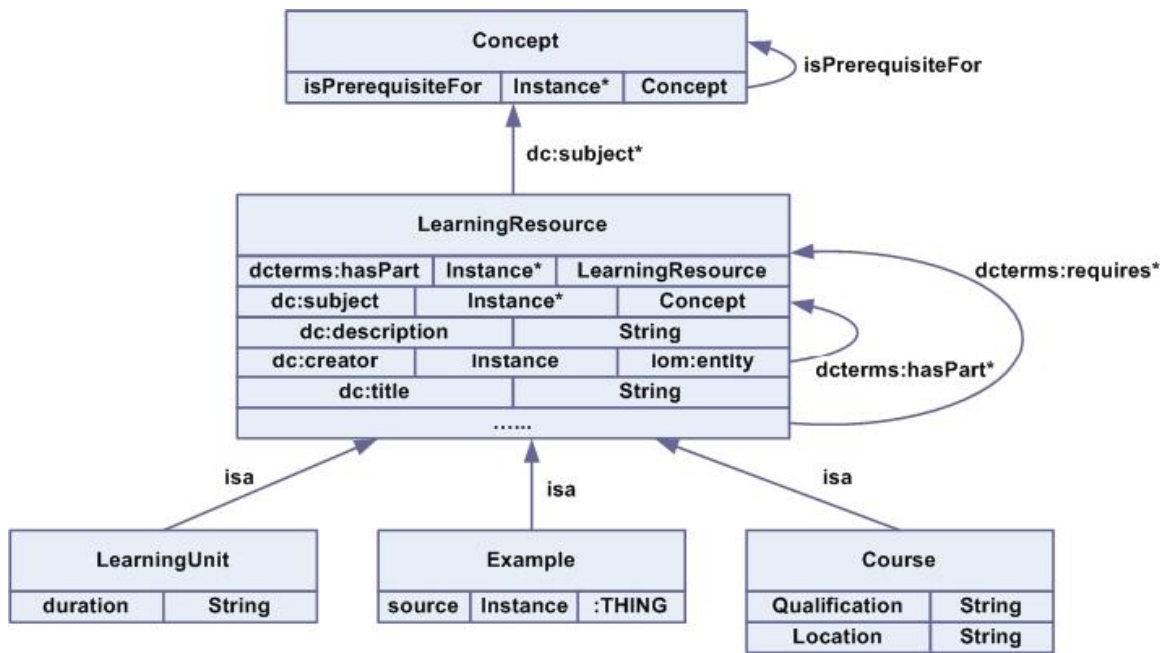
Επίσης, ένα τμήμα της οντολογίας που παριστάνει ένα εκπαιδευτικό πόρο φαίνεται στο Σχήμα 6. Συγκεκριμένα, η κλάση LearningResource καθορίζει τα χαρακτηριστικά που χρησιμοποιούνται για να περιγράψουν τους πόρους όπως αυτά έχουν υιοθετηθεί από

συνδυασμό των Duplin Core ή DC (<http://duplincore.org>) και Learning Objects Metadata Standard ή LOM (<http://ltsc.ieee.org/wg12/index.html>). Οι υποκλάσεις που παρουσιάζονται στο Σχήμα 8 αφορούν σε ειδικούς τύπους εκπαιδευτικού υλικού(Dolog et al., 2004):

- Course: είναι ένα LearningResource που έχει ένα Location και οδηγεί σε συγκεκριμένα Qualifications.
- LearningUnit: είναι ένα LearningResource που έχει ένα συγκεκριμένο Duration.
- Examples: εξηγούν το περιβάλλον μιας έννοιας ή ενός αντικειμένου που έχει διδαχθεί και αναπαρίσταται από Source.

Τα LearningResources και οι υποκατηγορίες τους δημιουργούν δομές (dcterms: hasPart σχέση) ή μπορεί να συνδεθούν με σχέση αναγκαίας προϋπόθεσης (dcterms: requires).

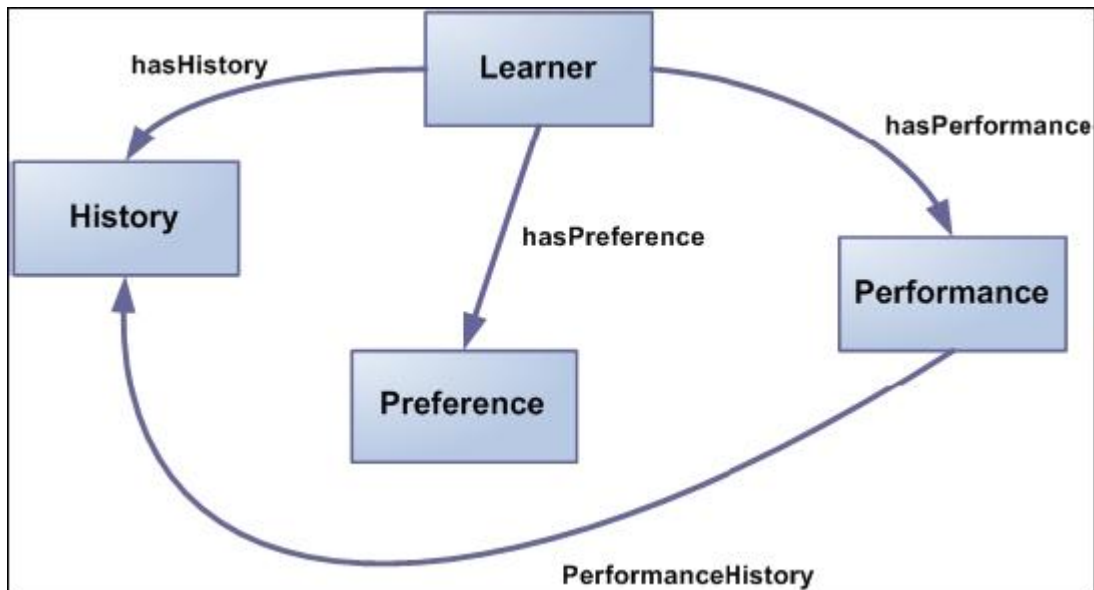
Η κλάση Concept χρησιμοποιείται για να περιγράψει βασικές πληροφοριακές οντότητες από την περιοχή της γνώσης που συνδέονται με τα Learning Resources. Concept και Learning Resource συνδέονται με την ιδιότητα dc:subject. Οι έννοιες αυτές μπορούν να περιοριστούν από τη σχέση is Prerequisite For εάν οι προϋποθέσεις οριστούν στο επίπεδο του Concept ή στο επίπεδο του LearningResource.



Σχήμα 9. Web Mining (b)

Πηγή: Stratakis et al(2003),Ontology Infrastructure, IEEE INTELLIGENT SYSTEMS

Επίσης, οι πληροφορίες σχετικά με τους εκπαιδευόμενους π.χ. τα ενδιαφέροντά τους, την απόδοσή τους σε διαφορετικές σειρές μαθημάτων, τους στόχους τους, τις προτιμήσεις τους, κλπ. θα πρέπει να οργανωθούν σε μια οντολογία. Το εννοιολογικό μοντέλο (conceptual model) για το προφίλ ενός Learner φαίνεται στο παρακάτω Σχήμα . Αυτό περιλαμβάνει την κλάση Learner η οποία συνδέεται με άλλες κλάσεις όπως, Performance, Preference και History. Η κλάση History περιλαμβάνει εκείνα τα LearningResources τα οποία έχουν προσπελαστεί και άρα μελετηθεί από τον εκπαιδευόμενο. Η κλάση Performance συνδέεται τόσο με την κλάση Learner όσο και με την κλάση History και απεικονίζει την απόδοση του εκπαιδευόμενου με βάση το υλικό που έχει μελετήσει, ενώ η κλάση Preference δηλώνει τις προτιμήσεις του.



Σχήμα 10. Το εννοιολογικό μοντέλο του Learner.

Πηγή: Erdmann M. & Studer R.: Ontologies as Conceptual Models for XML Documents, Proceedings of the 12th Workshop of Knowledge Acquisition Modeling Managing (KAW' 99), Bauff, Canada, October 1999.

3.4 ΔΙΑΔΙΚΑΣΙΑ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΣΕ E-LEARNING ΠΕΡΙΒΑΛΛΟΝΤΑ

3.4.1 Συλλογή των Δεδομένων

Το πρώτο στάδιο στη διαδικασία εξόρυξης γνώσης από δεδομένα του Web είναι η συλλογή των δεδομένων, ή αλλιώς η καταγραφή και η καταχώρηση πληροφοριών που αφορούν στους χρήστες. Η συλλογή των δεδομένων αποτελεί τη βάση της διαδικασίας εξόρυξης γνώσης αφού τα τελικά αποτελέσματα εξαρτώνται τόσο από τη διαθεσιμότητα των

στοιχείων όσο και από την ποιότητά τους. Υπάρχουν πολλοί τρόποι για να συλλεχθούν στοιχεία για τους χρήστες μιας e-learning εφαρμογής. Τα στοιχεία αυτά μπορούν να παρασχεθούν είτε ρητά (ζητώντας τα από το χρήστη), είτε έμμεσα (καταγράφοντας τη συμπεριφορά και ακολουθώντας τα «ίχνη» του στο μονοπάτι πλοήγησης που επιλέγει). Συγκεκριμένα, συνήθως συγκεντρώνονται οι ακόλουθοι τύποι στοιχείων:

Δεδομένα που εισάγουν ρητά οι χρήστες

- Πληροφορίες προφίλ. Ο χρήστης εισάγει ρητά στο σύστημα δεδομένα που αφορούν στα δημογραφικά στοιχεία του, καθώς και πληροφορίες για το εκπαιδευτικό του υπόβαθρο, τις δεξιότητές του, τις ανάγκες και τις προτιμήσεις του αναφορικά με τη χρήση του συστήματος ή τα περιεχόμενα και τη δομή της εκπαιδευτικής διαδικασίας. Τα δεδομένα αυτής της κατηγορίας συγκεντρώνονται συνήθως χρησιμοποιώντας ένα ηλεκτρονικό ερωτηματολόγιο το οποίο εμφανίζεται σε κάποιο στάδιο της διαδικασίας εγγραφής του χρήστη στο σύστημα και αποτελεί την προσωπική του καρτέλα (ή εγγραφή/record). Είναι σκόπιμο, οι χρήστες να έχουν δυνατότητα πρόσβασης και επεξεργασίας των στοιχείων αυτών ώστε να είναι συνεχώς ενημερωμένο το σύστημα και να μπορεί να παίρνει τις κατάλληλες αποφάσεις κατά τη διαδικασία της εξατομίκευσης. Μέρος του προφίλ αποτελούν και τα δεδομένα που εισάγει ο χρήστης κατά τη διαδικασία πιστοποίησής του κάθε φορά που θέλει να χρησιμοποιήσει το σύστημα, δηλαδή το όνομα (login) και ο κωδικός πρόσβασής του (password). Με αυτό τον τρόπο αναγνωρίζεται μονοσήμαντα, καταγράφεται η συμπεριφορά του και το ιστορικό του στο σύστημα και ενημερώνεται το προφίλ του ώστε να εφαρμοστούν με μεγαλύτερη ακρίβεια οι τεχνικές εξατομίκευσης.

- Δεδομένα αξιολόγησης του συστήματος. Ορισμένα e-learning συστήματα που υποστηρίζουν εξατομίκευση περιλαμβάνουν στα δεδομένα που εισάγουν οι χρήστες και τις αξιολογήσεις τους για την εκπαιδευτική διαδικασία που εφαρμόζεται στο σύστημα ή για επιμέρους συστατικά της. Η μορφή τους μπορεί να είναι βαθμολόγηση σε κλίμακα Likert ή σχόλια ελεύθερου κειμένου σε forums και bulletin boards του συστήματος (στην τελευταία

περίπτωση, οι πληροφορίες δεν είναι άμεσα χρησιμοποιήσιμες από το σύστημα και απαιτείται η παρέμβαση του ανθρώπινου παράγοντα ώστε να εξαχθούν u960 ποιοτικά συμπεράσματα).

Δεδομένα που συγκεντρώνει το σύστημα για το χρήστη

1. Αρχεία logs

Στα αρχεία αυτά, τα οποία κρατούνται στους εξυπηρετητές, καταγράφονται λεπτομέρειες που αφορούν σε κάθε αίτηση για εξυπηρέτηση όλων των ιστοσελίδων που βρίσκονται αποθηκευμένες στο συγκεκριμένο εξυπηρετητή. Συγκεκριμένα ένα αρχείο log περιέχει τη διεύθυνση IP του χρήστη, τη χρονική στιγμή που παρέλαβε την αίτηση για μια συγκεκριμένη ιστοσελίδα, τη διάρκεια παραμονής του σε αυτή, το URL της ιστοσελίδας από όπου πραγματοποιήθηκε η αίτηση και ένα σύνολο από επιπλέον στοιχεία που ενδέχεται να είναι εξίσου χρήσιμα ανάλογα με το σχεδιασμό των αλγορίθμων εξατομίκευσης που θα εφαρμοστούν(Luotonen, 1995). . Στη συνέχεια αναλύονται τα πεδία των αρχείων log ως προς τις ουσιαστικές πληροφορίες που μπορεί να παρέχουν για την παρακολούθηση της συμπεριφοράς των επισκεπτών στις σελίδες ενός web site, με δεδομένο τον τρόπο συνεργασίας των servers με τους browsers.

- Host / IP address. Πρόκειται για ένα εξαιρετικά ενδιαφέρον πεδίο. Πολλοί web servers έχουν την δυνατότητα να μεταφράζουν τις διευθύνσεις IP σε πλήρη ονόματα host (π.χ. το 150.140.141.183 στο www.ceid.upatras.gr) Με βάση το πλήρες domain και το hostname προκύπτουν ένα σύνολο από πληροφορίες που περιλαμβάνουν δημογραφικά στοιχεία που αφορούν είτε τη χώρα προέλευσης του χρήστη είτε το είδος του δικτύου στο οποίο ανήκει (αν, για παράδειγμα, η κατάληξη είναι .edu υποδηλώνει πρόσβαση από εκπαιδευτικό ίδρυμα της Αμερικής, ενώ το .gov δηλώνει πρόσβαση από εργαζόμενους σε κυβερνητικούς οργανισμούς).Εκτός από τέτοιου

είδους στοιχεία, το συγκεκριμένο πεδίο θα μπορούσε θεωρητικά να αναγνωρίζει ένα συγκεκριμένο χρήστη και τις "κινήσεις" του σε ένα αρχείο log, θεωρώντας ότι κάθε εγγραφή από μια συγκεκριμένη IP αναφέρεται σε ένα συγκεκριμένο χρήστη. Επιπλέον, μια απλή μέτρηση των μοναδικών IP στις εγγραφές του αρχείου log για μια χρονική περίοδο μπορεί να αποκαλύψει το πλήθος των διαφορετικών χρηστών που επισκέφτηκαν το site. Με τον τρόπο όμως που λειτουργεί στην πράξη στις μέρες μας το web και οι δυο ισχυρισμοί αναιρούνται.

- Αναγνώριση της ταυτότητας του χρήστη(ident). Το πεδίο αυτό χρησιμοποιείται στις περιπτώσεις που για την είσοδο στα περιεχόμενα μιας σελίδας απαιτείται πιστοποίηση της αυθεντικότητας του χρήστη και περιέχει τον κωδικό αναγνώρισης του. Στην πράξη όμως πρόκειται για ένα πεδίο που δεν χρησιμοποιείται (παραμένει κενό) καθώς προκαλεί προβλήματα ασφαλείας.
- Χρονική σφραγίδα ημερομηνίας και ώρας (date / time stamp). Προσδιορίζει την χρονική στιγμή που ο εξυπηρετητής παρέλαβε μια αίτηση (η οποία μπορεί να αφορά μια ιστοσελίδα , ένα εκτελέσιμο πρόγραμμα κλπ.). Επειδή ενδέχεται ανάλογα με το web server να διαφοροποιείται το τι ακριβώς καταγράφεται στο πεδίο αυτό (για παράδειγμα, να καταγράφει τη χρονική στιγμή που εξυπηρετήθηκε η αίτηση και όχι αυτή που παραλήφθηκε) καλό είναι να επιβεβαιώνεται από τις κατά περίπτωση προδιαγραφές.
- Μέθοδος / ομοιόμορφος εντοπιστής πόρων / πρωτόκολλο. Το πρώτο μέρος του πεδίου αυτού αναφέρεται στη μέθοδο που χρησιμοποιήθηκε για τη συγκεκριμένη αίτηση. Πιθανές τιμές του πεδίου είναι: *Get* (αίτηση αναγνώρισης ενός αντικειμένου ή μιας ιστοσελίδας από τον εξυπηρετητή), *Post* (αποστολή πληροφορίας στον εξυπηρετητή). Γεγονός που συνήθως συνεπάγεται την προώθηση μιας αίτησης από τον server για εκτέλεση κάποιου προγράμματος. Και *Head* (αίτηση ανάγνωσης του αρχείου επικεφαλίδας ενός αντικειμένου ή μιας ιστοσελίδας. Το δεύτερο μέρος του πεδίου αναφέρεται σε ένα στατικό αρχείο του τοπικού συστήματος αρχείων ή σε κάποιο

εκτελέσιμο πρόγραμμα το οποίο ζήτησε να εκτελέσει ο χρήστης. Το τελευταίο μέρος του πεδίου αυτού αναφέρεται στην έκδοση του πρωτοκόλλου που χρησιμοποιείται (συνήθως HTTP).

- Κατάσταση (status). Στο πεδίο αυτό ο server καταγράφει έναν κωδικό κατάστασης (status code) ο οποίος σχετίζεται με κάθε συγκεκριμένη αίτηση που εξυπηρετήθηκε. Για παράδειγμα κωδικοί από το 200 έως το 299 υπονοούν επιτυχή συναλλαγή, κωδικοί από το 300 έως το 399 υπονοούν κάποιας μορφής δρομολόγηση, κωδικοί από το 400 έως το 499 υπονοούν ανεπιτυχή συναλλαγή (για παράδειγμα, ο κωδικός 404 υποδεικνύει ότι το αιτούμενο αντικείμενο δεν βρέθηκε) και κωδικοί από 500 έως 599 υπονοούν κάποιου είδους πρόβλημα με τον εξυπηρετητή.
- Μέγεθος (bytes). Στο πεδίο αυτό καταγράφεται ο αριθμός ψηφιολέξεων (bytes) που επιστράφηκε σαν αποτέλεσμα της εξυπηρέτησης μιας αίτησης .
- Αναφορές αντικειμένου (referrer). Αναφέρεται στο URL στο οποίο βρισκόταν ο χρήστης όταν έστειλε την τρέχουσα αίτηση προς εξυπηρέτηση στον server. Έτσι, η τιμή του πεδίου αυτού είναι ένα URL που μπορεί να περιέχει και ένα query string στις περιπτώσεις που χρησιμοποιήθηκε κάτι τέτοιο (όταν για παράδειγμα, ο χρήστης κάνει μια αναζήτηση με βάση λέξεις-κλειδιά). Ενδέχεται επίσης το πεδίο αυτό να είναι κενό , όπως συμβαίνει στην περίπτωση που ο χρήστης δεν ακλούθησε κάποιο link (και έτσι στάλθηκε η αίτηση), αλλά πληκτρολόγησε ένα συγκεκριμένο URL ή χρησιμοποίησε ένα bookmark. Ο λόγος που προστέθηκε το πεδίο αυτό στο HTTP 1.0 ήταν για να μπορεί, για παράδειγμα, να μετρηθεί η αποτελεσματικότητα ενός διαφημιστικού banner.
- Πράκτορας (user agent). Προσδιορίζει το λειτουργικό σύστημα και τον browser που χρησιμοποιεί ο χρήστης.

Τα αρχεία log αποτελούν την βασική πηγή δεδομένων clickstream, ενώ παράλληλα η ανάλυσή τους για την εξαγωγή συμπερασμάτων αναφορικά με την συμπεριφορά του χρήστη

παραμένει μια σημαντική πρόκληση. Αν και για κάθε εξυπηρέτηση αίτησης που ολοκληρώνει ο server δημιουργείται μια νέα εγγραφή, οι εγγραφές που <<δημιουργούνται>> από τον ίδιο χρήστη στην πλειονότητα των περιπτώσεων δεν καταγράφονται διαδοχικά στο αρχείο log, αλλά βρίσκονται διασκορπισμένες. Αυτό συμβαίνει επειδή ένας server συνήθως εξυπηρετεί ταυτόχρονα πολλούς χρήστες (τόσους όσοι οι ταυτόχρονοι επισκέπτες του/των web sites που φιλοξενεί), με αποτέλεσμα να μην είναι απλή η ανάκτηση και η μελέτη του μονοπατιού που ακολούθησε ένας συγκεκριμένος χρήστης στις σελίδες ενός web site.

2. Cookies

Ένα cookie είναι μια μεταβλητή, την τιμή της οποίας μπορεί να αποθηκεύσει μια ιστοσελίδα στον υπολογιστή ενός επισκέπτη της, ώστε να την ανακτήσει από εκεί όταν ο χρήστης την ξανά επισκεφτεί χρησιμοποιώντας το ίδιο PC. Για παράδειγμα στα cookies οφείλεται το ότι, αφού συμπληρώσει ο χρήστης για πρώτη φορά το password που επιθυμεί, σε επόμενες επισκέψεις του συμπληρώνεται από μόνο του αφού δώσει username. Τα cookies δεν είναι τίποτα περισσότερο από από δεδομένα σε μορφή απλού κειμένου, και το μόνο site που μπορεί να προσπελάσει-διαβάσει ένα cookie είναι το site που το δημιούργησε και το αποθήκευσε. Τα cookies δεν είναι προγράμματα και δεν μπορούν από μόνα τους να συγκεντρώσουν πληροφορίες. Τα αρχεία αυτά (η χρήση των οποίων συχνά συνδυάζεται με τα αρχεία log) είναι δυνατό να περιέχουν ποικίλες πληροφορίες, συμπεριλαμβανομένων των σελίδων που ο χρήστης έχει επισκεφθεί, των κωδικών πρόσβασης που έχει εισαγάγει, κλπ. Ένα cookie αποθηκεύει τις πληροφορίες στον τοπικό δίσκο του χρήστη (κάτι που συνήθως συμβαίνει την πρώτη φορά που ζητάει εξυπηρέτηση ο χρήστης) και εκεί έχει πρόσβαση ο εξυπηρετητής σε κάθε μελλοντική αίτηση για εξυπηρέτηση που θα παραλάβει από το συγκεκριμένο χρήστη, επιτρέποντας με αυτό το «διαφανή» τρόπο την αυτόματη αναγνώριση του χρήστη από το σύστημα. Αξίζει όμως να σημειωθεί ότι είναι εύκολο οι χρήστες να απενεργοποιήσουν τα cookies από αντίστοιχη ρύθμιση του browser (φυλλομετρητή).

Άλλα δεδομένα.

Εκτός από τα δεδομένα που αφορούν στο χρήστη σε μια τυπική διαδικασία εξατομίκευσης χρησιμοποιούνται επίσης:

- Το περιεχόμενο του site. Τα δεδομένα αυτής της κατηγορίας αφορούν στο περιεχόμενο του site δηλαδή στο θεματικό του, στα μαθήματα από τα οποία αποτελείται, στις ασκήσεις και στα παραδείγματα, κλπ.
- Η δομή του site. Τα δεδομένα αυτής της κατηγορίας αφορούν στον τρόπο που είναι οργανωμένο το περιεχόμενο του site. Μπορεί να είναι είτε στοιχεία που χρησιμοποιούνται μέσα σε μια ιστοσελίδα (π.χ. ετικέτες HTML ή XML), είτε στοιχεία που διασύνδεουν ένα site (π.χ. υπερσύνδεσμοι).

3.4.2 Καθαρισμός των Δεδομένων & Μοντελοποίηση του Χρήστη

Μετά την ολοκλήρωση της συλλογής, ακολουθεί η φάση του καθαρισμού των δεδομένων κατά την οποία αφαιρούνται όλες οι περιττές πληροφορίες ώστε στη συνέχεια να χρησιμοποιηθούν τα «καθαρά» δεδομένα για την κατασκευή ή την ενημέρωση των προφίλ των χρηστών (users' profiles). Το προφίλ κάθε χρήστη αναπαρίσταται εσωτερικά με βάση κάποιο μοντέλο (model) το οποίο επιλέγεται κατάλληλα ώστε να επιτρέψει την αποδοτικότερη εφαρμογή των κανόνων εξόρυξης γνώσης που θα χρησιμοποιηθούν σε τελικό στάδιο για την εξατομίκευση της ηλεκτρονικής μαθησιακής εμπειρίας.

Η διαδικασία καθορισμού του μοντέλου αυτού είναι γνωστή σαν μοντελοποίηση χρήστη ή user modeling και αποτελεί μια ώριμη επιστημονική περιοχή που έχει να επιδείξει ένα σημαντικό σύνολο τεχνικών, οι οποίες έχουν αξιολογηθεί στην πράξη. Οι πιο γνωστές

από αυτές τις τεχνικές είναι οι πιθανοτικές, οι βασισμένες σε κανόνες, τα νευρωνικά δίκτυα, οι βασισμένες σε στερεότυπα και οι επαγωγικοί κανόνες. Τα προφίλ που προκύπτουν περιέχουν το σύνολο των πληροφοριών που έχουν συλλεχθεί για κάθε χρήστη και κρατούνται σε μια βάση δεδομένων, η οποία και ενημερώνεται συνεχώς με τις νέες πληροφορίες που καταγράφονται κατά την παρακολούθηση των ενεργειών κάθε χρήστη στο σύστημα (Luotonen A:1994).

3.4.3 Ανάλυση των Δεδομένων

Μετά από τη φάση της συλλογής των δεδομένων και του καθαρισμού τους, ακολουθεί η περαιτέρω επεξεργασία τους προκειμένου να εντοπιστούν πρότυπα ή μοτίβα (patterns) συμπεριφορών. Έτσι, το επόμενο βήμα είναι η ανάλυση (analysis) των διαθέσιμων στοιχείων και η εφαρμογή τεχνικών Web Mining. Στις πιο συχνά χρησιμοποιούμενες τεχνικές περιλαμβάνονται οι:

- Κατηγοριοποίηση (classification). Εφαρμόζεται στο σύνολο των δεδομένων ώστε να εξαχθούν οι κλάσεις των χρηστών με βάση προαποφασισμένες ετικέτες (class labels). Για παράδειγμα, έστω ότι θέλουμε να δημιουργήσουμε τις κλάσεις των χρηστών που παρακολουθούν τη θεματική ενότητα Microsoft Excel. Με τη βοήθεια δένδρων αποφάσεων (decision trees) μπορούμε να βρούμε ότι το προφίλ αυτών των χρηστών είναι: άτομα ηλικίας μικρότερης των 30 ετών, ανύπαντρα και χωρίς παιδιά, απόφοιτοι τριτοβάθμιας εκπαίδευσης που χρησιμοποιούν ηλεκτρονικό υπολογιστή στην εργασία τους. Αυτές οι πληροφορίες μπορούν να χρησιμοποιηθούν αργότερα για να προτείνουμε συγκεκριμένο εκπαιδευτικό πλάνο σε κάθε χρήστη. Η κατηγοριοποίηση ανήκει στις προγνωστικές τεχνικές (predictive web mining). Για να λειτουργήσει απαιτείται ένα σύνολο δεδομένων εκπαίδευσης (training data set) ώστε να αναγνωριστούν τα ιδιαίτερα χαρακτηριστικά κάθε κλάσης και στη

συνέχεια να ταξινομηθούν τα νέα αντικείμενα με βάση τις τιμές τους στα χαρακτηριστικά αυτά. Οι πιο γνωστοί αλγόριθμοι που χρησιμοποιούνται είναι: η απλοϊκή κατηγοριοποίηση κατά Bayes (naive Bayes), ο K-πλησιέστερος γείτονας (K-nearest neighbor ή KNN), τα δέντρα αποφάσεων (decision trees), τα νευρωνικά δίκτυα (neural networks), κλπ.

- Συσταδοποίηση ή ομαδοποίηση (clustering). Η λογική της τεχνικής είναι ο διαχωρισμός των χρηστών σε ομάδες με βάση το ότι χρήστες που ανήκουν στην ίδια ομάδα είναι όσο το δυνατόν περισσότερο όμοιοι, ενώ χρήστες διαφορετικών ομάδων είναι όσο το δυνατόν πιο διαφορετικοί. Η ομοιότητα μετριέται χρησιμοποιώντας μια συνάρτηση ομοιότητας (similarity function) που εφαρμόζεται στα προφίλ των χρηστών. Αφού δημιουργηθούν οι ομάδες, είναι στη συνέχεια δυνατό να τους αποδοθούν ετικέτες ανάλογα με το είδος της κάθε ομάδας και να χρησιμοποιηθούν οι ετικέτες αυτές σαν βάση για την εφαρμογή κατηγοριοποίησης (με αυτό τον τρόπο χρησιμοποιούμε την ομαδοποίηση σαν ένα αρχικό βοηθητικό στάδιο πριν εφαρμόσουμε κατηγοριοποίηση, στις περιπτώσεις που δεν είναι εύκολο να προκαθορίσουμε τις κλάσεις). Εκτός από την ομαδοποίηση των χρηστών είναι δυνατή η εφαρμογή του clustering με στόχο το σχηματισμό ομάδων σελίδων. Αυτό σημαίνει ότι σελίδες που αντιμετωπίζονται με παρόμοιο τρόπο από τους χρήστες ή περιέχουν παρόμοιο περιεχόμενο ομαδοποιούνται στο ίδιο cluster. Πρέπει να σημειωθεί ότι είναι απαραίτητο να επιλέγονται τεχνικές που επιτρέπουν την αλληλοεπικάλυψη των clusters (overlapping), καθώς στις περισσότερες περιπτώσεις ένα χρήστης ή μια σελίδα ανήκουν σε περισσότερα από ένα clusters. Αρκετοί αλγόριθμοι ομαδοποίησης είναι διαθέσιμοι: η ιεραρχική συσσωρευτική συσταδοποίηση (hierarchical agglomerative clustering ή HAC) , ο k-means, οι αυτό- οργανωμένοι χάρτες χαρακτηριστικών (self-organizing maps ή SOMs) με πιο γνωστούς τους χάρτες Kohonen.

- Κανόνες συσχέτισης (association rules). Χρησιμοποιούνται για να εντοπίσουν ομοιότητες μεταξύ διαφορετικών τύπων δεδομένων και να επιτευχθεί καλύτερη κατανόηση της συμπεριφοράς των χρηστών. Οι κανόνες αυτοί συνδέουν ένα ή περισσότερα (αρχικά μη συσχετιζόμενα) γεγονότα και ανακαλύπτουν σχέσεις που δεν μπορούν (εύκολα) να

προβλεφθούν. Ένας τέτοιος κανόνας έχει τη μορφή A.B και υποδηλώνει μια σχέση εξάρτισης ανάμεσα σε δύο (ξένα) σύνολα στοιχείων A και B δηλαδή υπολογίζει την πιθανότητα να συμβεί το B με δεδομένο ότι έχει συμβεί το A. Για παράδειγμα, σε ένα e-learning σύστημα ένας κανόνας συσχέτισης θα μπορούσε να είναι ότι το «20% των χρηστών που μελετούν το μάθημα Microsoft Windows 2000 επιλέγουν στη συνέχεια να μελετήσουν το μάθημα Microsoft Word 2000», ή ότι το «50% των χρηστών που επισκέφθηκαν τις σελίδες της βοήθειας ανήκουν στην ηλικιακή ομάδα 25 30». Ο πιο γνωστός αλγόριθμος για την ανακάλυψη κανόνων συσχέτισης είναι ο Apriori και οι παραλλαγές του, ενώ υπάρχουν και προσεγγίσεις που χρησιμοποιούν Bayesian Network. Μια επέκταση των κανόνων συσχέτισης αποτελεί η τεχνική ανακάλυψης ακολουθιακών μοτίβων ή sequential pattern discovery η οποία ενσωματώνει την έννοια του χρόνου στα patterns. Ένα παράδειγμα θα μπορούσε να είναι της μορφής (C)(A,B) όπου A , B , C είναι σελίδες ενός e-learning site και σημαίνει ότι οι εκπαιδευόμενοι που επισκέπτονται τη σελίδα C κάποια στιγμή στη συνέχεια ζητούν να δουν τις σελίδες A και B (είναι δυνατόν να παρεμβληθούν και άλλες σελίδες μεταξύ των συγκεκριμένων αιτήσεων). Οι περισσότερες υλοποιήσεις αυτής της τεχνικής βασίζονται σε αλγορίθμους που μοιάζουν με τον Apriori, καθώς και σε μοντέλα Markov (Seidner R and N. Tindall: 2003)

3.4.4 Ολοκλήρωση των Δεδομένων

Τα συμπεράσματα και η γνώση που εξορύσσεται με βάση μια ή περισσότερες από τις παραπάνω τεχνικές χρησιμοποιούνται σαν είσοδος στην επόμενη φάση της διαδικασίας εξατομίκευσης που είναι γνωστή σαν ολοκλήρωση δεδομένων, καθώς συνδυάζει το σύνολο των πληροφοριών που είτε συλλέχθηκαν είτε παράχθηκαν στην πορεία. Ανάλογα με την κατά περίπτωση τεχνική (ή τις τεχνικές) εξατομίκευσης που εφαρμόζει ένα e-learning σύστημα, ενδέχεται να χρησιμοποιεί σαν είσοδο και πληροφορίες που σχετίζονται με τη

δομή και την τοπολογία του εκπαιδευτικού υλικού, το ακριβές περιεχόμενο ή το σκοπό μιας ιστοσελίδας, το πάτημα κάποιου κουμπιού, την επιλογή κάποιου συγκεκριμένου συνδέσμου, το χρόνο παραμονής σε μια ιστοσελίδα και γενικότερα δεδομένα που σχετίζονται με τη συμπεριφορά των χρηστών σε επίπεδο ακολουθίας κλικ (clickstream) σε συνδυασμό όμως με τη σημασιολογία κάθε ενέργειας στα πλαίσια του συγκεκριμένου συστήματος. Σε αυτές τις περιπτώσεις, εκτός από το λογικό επακόλουθο της αύξησης της πολυπλοκότητας του συστήματος, όσο και των απαιτήσεών του σε πόρους, για την εξαγωγή συμπερασμάτων, είναι απαραίτητη η συμβολή εκπαιδευτών οι οποίοι μπορούν να αντιληφθούν και να επεξηγήσουν τη μαθησιακή συμπεριφορά των χρηστών και να προτείνουν το είδος της κατάλληλης εξατομικευμένης προσαρμογής.

ΚΕΦΑΛΑΙΟ 4

ΕΦΑΡΜΟΓΕΣ ΕΞΟΥΥΞΗΣ ΧΡΗΣΗΣ ΚΑΙ ΕΞΑΤΟΜΙΚΕΥΣΗΣ

4.1 ΕΞΑΤΟΜΙΚΕΥΣΗ ΣΤΟ ΠΕΡΙΒΑΛΛΟΝ ΤΟΥ WEB

Πριν αρχίσει η ανάλυση των επιμέρους θεμάτων, είναι χρήσιμο να δοθούν ορισμένες διευκρινήσεις σχετικά με την έννοιες εξατομίκευση(personalization), προσαρμογή(customization), και εξατομικευμένο περιβάλλον πληροφόρησης (Personalized Information Environment).

Εξατομικεύω = προσαρμόζω (κάτι) σε κάθε ατομική περίπτωση. Ετυμολογικά σημαίνει δράση, ενέργεια, αναφορά σε επίπεδο ατόμων και όχι στο σύνολο, όχι σε ένα ή συγκεκριμένο άτομο, αλλά «κατ' άτομο», στον καθένα χωριστά.

Προσαρμόζω = τροποποιώ, μεταβάλλω κάτι ώστε να εναρμονιστεί (με ορισμένη κατάσταση).

Προσαρμογή ορίζεται ως η ικανότητα που έχουν οι χρήστες να ελέγξουν τον τρόπο με τον οποίο το περιεχόμενο τακτοποιείται στη σελίδα τους, ενώ εξατομίκευση είναι η διαδικασία συλλογής από ένα portal, των συνηθειών και προτιμήσεων του χρήστη, προκειμένου να δημιουργηθεί ένα περιβάλλον πληροφοριών που προσαρμόζεται στις ιδιαίτερες προτιμήσεις του και εστιάζει στην αποτελεσματικότερη ανάκτηση.

Εξατομίκευση στο χώρο του διαδικτύου είναι η διαδικασία δυναμικής προσαρμογής των υπηρεσιών ενός συστήματος προκειμένου να ικανοποιήσει ή να υποστηρίξει τις ιδιαίτερες προτιμήσεις και τους συγκεκριμένους στόχους ενός χρήστη ή μιας ομάδας ατόμων .

Συχνότατα στις εφαρμογές εξατομικευμένου περιβάλλοντος συναντούμε τη χρήση του όρου customization ο οποίος χρησιμοποιείται ως συντομογραφία και για τις δύο έννοιες.

Εξατομικευμένες υπηρεσίες αρχικά εμφανίστηκαν στο web για να κρατήσουν όσο το δυνατόν περισσότερο ένα επισκέπτη στο site. Σήμερα στοχεύουν στην ταχύτερη και ακριβέστερη παράδοση των πληροφοριών στον επισκέπτη με στόχο το κέρδος.

Η εξατομίκευση αποτελεί τεχνολογία κλειδί σε διάφορες πτυχές του ηλεκτρονικού εμπορίου, στη διαφήμιση και προώθηση προϊόντων, στη διαχείριση του περιεχομένου ιστοσελίδων, στη διαχείριση γνώσης. Παρά το ότι σε κάθε μία από τις περιοχές εφαρμογής της η εξατομίκευση απαιτεί ειδικές ρυθμίσεις όσον αφορά το περιβάλλον αλληλεπίδρασης και τη διαδικασία συλλογής δεδομένων, οι βασικές τεχνικές του έχουν σημαντικές ομοιότητες.

Τα περισσότερα συστήματα εξατομίκευσης στο web μπορούν να κατηγοριοποιηθούν ως εξής:

1. Manual decision rule systems: Οι διαχειριστές του συστήματος (system administrators) καθορίζουν τους κανόνες είτε από τα στοιχεία που συγκεντρώνονται μέσω μίας διαδικασίας εγγραφής, είτε από το ιστορικό του session. Οι κανόνες χρησιμοποιούνται για να προσαρμόσουν το περιεχόμενο που αποστέλλεται από τον server σε κάθε χρήστη

2. Collaborative filtering systems: Συστήματα που παίρνουν σαν είσοδο ρητή πληροφορία με τη μορφή προτιμήσεων ή βαθμολογήσεων που δίνουν οι χρήστες και μέσω μιας μηχανής, επιστρέφουν πληροφορία που προβλέπεται ότι προσεγγίζει τις προτιμήσεις συγκεκριμένων χρηστών. Δηλαδή το σύστημα ταιριάζει τις συμπεριφορές των χρηστών και προτείνει με βάση τις συναλλαγές. Αν βρει άλλους χρήστες με τις ίδιες καταναλωτικές συμπεριφορές τότε ελέγχει τι δεν καταναλώθηκε και γίνεται σύσταση.

3. Content-based filtering agents: Στηρίζονται στο content similarity (ομοιότητα περιεχομένου) μεταξύ των κειμένων στο web και του προφίλ που δημιουργείται άμεσα ή έμμεσα από τους χρήστες. Συσχετίζονται τα κείμενα με βάση τις προτιμήσεις των χρηστών και αφού φιλτραριστεί η πληροφορία προτείνεται στο χρήστη με ακρίβεια αυτό που θέλει.

4.2 ΠΑΡΑΓΩΓΗ ΕΞΑΤΟΜΙΚΕΥΣΗΣ

Η διαδικασία εξατομίκευσης είναι όταν περνάμε στην τελική φάση στην οποία παράγετε για κάθε χρήστη μια διαφορετική εξατομικευμένη εμπειρία μάθησης με βάση τις συγκεκριμένες προτιμήσεις και τις ανάγκες του, ώστε κάθε φορά που επιστρέφει ο χρήστης στο σύστημα, αυτό να προσαρμόζεται κατά το δυνατό στο δικό του προφίλ.

Αυτό το γεγονός αυξάνει τη φιλικότητα του συστήματος προς το χρήστη κάνοντας την εκπαιδευτική διαδικασία αποδοτικότερη και διευκολύνοντας την απόκτηση νέων δεξιοτήτων, παρέχει αμεσότητα στην αλληλεπίδραση, αυξάνει την εμπιστοσύνη των χρηστών προς το σύστημα και βελτιώνει τη συνολική εμπειρία της ηλεκτρονικής μάθησης. Ενδεικτικές προσαρμογές που μπορούν να πραγματοποιηθούν σε ένα e-learning σύστημα ώστε να επιτευχθεί εξατομίκευση περιλαμβάνουν:

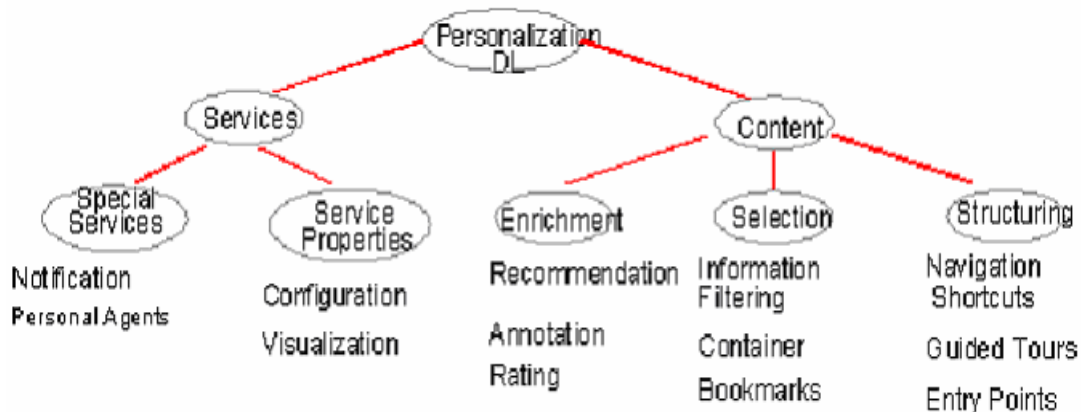
- Διαφοροποιήσεις στο εκπαιδευτικό περιεχόμενο: εισαγωγή/αφαίρεση θεματικών ενοτήτων, τμημάτων ή παραγράφων, προαιρετικές εξηγήσεις ή αναλυτικές πληροφορίες, εξατομικευμένες συστάσεις για μελέτη άλλων θεματικών ενοτήτων, κλπ.

- Διαφοροποιήσεις στη δομή του εκπαιδευτικού υλικού: τρόπος ταξινόμησης, κρύψιμο, προσθήκη, αφαίρεση ή σήμανση υπερσυνδέσμων, κλπ.

- Διαφοροποιήσεις στον τρόπο παρουσίασης του εκπαιδευτικού περιεχομένου: με περισσότερες ή λιγότερες εικόνες, με ή χωρίς ήχο, με ή χωρίς βίντεο, με μεγαλύτερη ή μικρότερη γραμματοσειρά, χρωματισμένο κείμενο ή όχι, κλπ.

4.2.1 Ταξινόμηση Μεθόδων Εξατομίκευσης

Το παρακάτω σχήμα αφορά μία ταξινόμηση των μεθόδων εξατομίκευσης



Σχήμα 11. Μέθοδοι εξατομίκευσης.

ΠΗΓΗ: Denise A. Roberts, User Interfaces and personalization, June 4, 2003 INFO653
Digital libraries Spring 2003

Αναλύοντας το διακρίνουμε:

1. Μεθόδους εξατομίκευσης που αφορούν υπηρεσίες βιβλιοθηκών (services) οι οποίες διακρίνονται σε:
 - a. Εξατομίκευση που αφορά ειδικές υπηρεσίες (personalization of special services) όπως υπομνήσεις (notification), προσωπικούς πράκτορες (personal agents). και
 - b. Εξατομίκευση ιδιοτήτων των υπηρεσιών (personalization of Service Properties) όπως. Διαμόρφωση (Configuration) και οπτικοποίηση (visualization).
2. Μεθόδους εξατομίκευσης που αφορούν το περιεχόμενο των βιβλιοθηκών (content) οι οποίες διακρίνονται σε:

a. Εξατομίκευση όσον αφορά τον εμπλουτισμό του περιεχομένου (Enrichment). Προσφέρεται για τη διευκόλυνση λήψης ατομικών αποφάσεων αναφορικά με τη συλλογή περιεχομένου (content selection) και χρήση περιεχομένου (content use), σχόλια από ειδικούς προς συγκεκριμένες ομάδες χρηστών, αξιολογήσεις, συστάσεις, σχολιασμοί.

b. Εξατομίκευση όσον αφορά την επιλογή του περιεχομένου(selection). Αφορά μεθόδους φιλτραρίσματος της πληροφορίας (information filtering) βασισμένες στις προτιμήσεις και ικανότητες των χρηστών. Οι περισσότερες από αυτές τις μεθόδους βασίζονται στα χαρακτηριστικά μοντελοποίησης χρηστών, (ενδιαφέροντα, προτιμήσεις, δεξιότητες) Επίσης αφορά μεθόδους για ενεργή συλλογή συναφούς υλικού σε bookmark λίστες και information containers (αποθήκες πληροφοριών) και

c. Εξατομίκευση αναφορικά με ατομικές πληροφορίες για τη διάρθρωση του περιεχομένου (individual information content structuring). Εδώ περιέχονται navigation Shortcuts (πληροφορίες σχετικά με την πορεία πλοήγησης ανάμεσα στους συνδέσμους, προκειμένου να οδηγηθείς στο συγκεκριμένο υλικό), guided tours (οδηγίες για την πορεία του χρήστη προς τη συγκεκριμένη πληροφορία) και entry points (σημεία εισόδου).

4.2.2 Προ-επεξεργασία των Δεδομένων

Όπως έχει αναφερθεί προηγουμένως η όλη διαδικασία εξατομίκευσης διακρίνεται στις ενέργειες που θα γίνουν offline και αυτές που θα γίνουν online. Στις πρώτες εντάσσονται εκτός από την κατασκευή της e-learning οντολογίας, η προ-επεξεργασία των δεδομένων και οι ενέργειες εξόρυξης των δεδομένων χρήσης. Η προ-επεξεργασία των δεδομένων έχει σαν αποτέλεσμα τη δημιουργία δομών όπως είναι το ιστορικό με τις αλληλεπιδράσεις των χρηστών με το web site ώστε να υπολογισθούν σημαντικά σημασιολογικά στοιχεία της δραστηριότητας των χρηστών που θα χρησιμοποιηθούν στο στάδιο εξόρυξης. Έχοντας στη διάθεσή μας τα προ-επεξεργασμένα δεδομένα είναι δυνατή στη συνέχεια η εκτέλεση μιας

σειράς από τεχνικές εξόρυξης δεδομένων. Στην προτεινόμενη προσέγγιση εστιάζουμε στην ανακάλυψη κανόνων συσχέτισης (association rules mining) και χρησιμοποιούμε τον αλγόριθμο (Apriori Barron, B. B:2002)

Η πιο σημαντική πηγή δεδομένων clickstreams είναι ο web server στον οποίο βρίσκεται το e-learning site. Ο web server καταγράφει σε πραγματικό χρόνο την αλληλεπίδραση των χρηστών με το browser με τη μορφή αρχείων server logs. Γενικά, ένα αρχείο log καταγράφει την κίνηση και τη συμπεριφορά πλοήγησης των χρηστών στις σελίδες ενός δικτυακού τόπου (Mobasher, 2005). Το συγκεκριμένο σύστημα χρησιμοποιεί τα logs αρχεία που περιγράφουν τη δραστηριότητα πλοήγησης των χρηστών και πιο συγκεκριμένα, κάθε φορά που ο web server εξυπηρετεί μια αίτηση, μια νέα εγγραφή προστίθεται στο αρχείο log με λεπτομερείς πληροφορίες σχετικά με το τι ζητήθηκε, πότε, από ποιόν, σε ποια σελίδα βρισκόταν ο χρήστης όταν το ζήτησε, το εάν η αίτηση εξυπηρετήθηκε με επιτυχία ή υπήρχαν προβλήματα από την πλευρά του server και ένα σύνολο από επιπλέον στοιχεία που εξαρτώνται από τις ρυθμίσεις του server και τη μορφοποίηση των αρχείων log που κρατάει (Austen, G:1998)..

Στην ουσία, αυτά τα αρχεία συμπυκνώνουν όλες τις πληροφορίες που σχετίζονται με τη χρήση του e-learning site από τους χρήστες. Σε αυτό το στάδιο, τα server logs καθαρίζονται σύμφωνα με τα αρχεία του web site. Αυτή η εργασία περιλαμβάνει την αφαίρεση όλων των πλεοναζουσών αναφορών. Επίσης, απαιτούνται λεπτομερείς πληροφορίες σχετικά με τη δομή του site προκειμένου να καθοριστούν ποιες προσβάσεις αρχείων σελίδων αντιστοιχούν σε ένα μοναδικό παράθυρο φυλλομετρητή και πιο συγκεκριμένα ποιο περιεχόμενο αντιστοιχεί στο αίτημα κάθε χρήστη.

Οι βασικές ενέργειες για την προ-επεξεργασία των log δεδομένων είναι οι ακόλουθες(Cooley et al., 1999):

- Καθαρισμός των δεδομένων. Σε αυτό το βήμα εξετάζονται τα διαθέσιμα δεδομένα και απομακρύνονται όλες οι άσχετες ή πλεονάζουσες εγγραφές. Αυτές συνήθως προέρχονται από το γεγονός ότι το HTTP πρωτόκολλο απαιτεί ξεχωριστά αιτήματα στο server για κάθε αρχείο εικόνας, βίντεο, ήχου, κλπ. που είναι ενσωματωμένο στη web page. Έτσι τα αιτήματα για αρχεία γραφικών, βίντεο, ήχου, CGI scripts, κλπ. εντοπίζονται σε αυτό το στάδιο και απομακρύνονται με βάση την κατάληξή τους π.χ. .gif, .jpg, .avi, κλπ.. Επειδή όμως η απομάκρυνση αυτών των δεδομένων μπορεί να οδηγήσει σε απώλεια χρήσιμης πληροφορίας ιδιαίτερα σε sites με πολυμεσικό περιεχόμενο, τα αρχεία αυτά μπορεί να παραμείνουν στο σύνολο δεδομένων και να συμπεριληφθούν στα αντίστοιχα pageviews σε επόμενο βήμα της προ-επεξεργασίας. Επιπλέον, φιλτράρονται οι εγγραφές που αφορούν σε αιτήματα που δεν έχουν ικανοποιηθεί π.χ. HTTP απαντήσεις λάθους, εγγραφές από spiders ή crawlers, κλπ. (Tan & Kumar, 2002).

- Αναγνώριση των χρηστών. Αυτό το βήμα υλοποιήθηκε εύκολα αφού στο συγκεκριμένο e-learning site οι χρήστες δεν είναι ανώνυμοι αλλά είναι απαραίτητη η χρήση login και password μέσω μηχανισμού εγγραφής (registration) για να μπορέσει ένας εκπαιδευόμενος να παρακολουθήσει την εκπαιδευτική διαδικασία.

- Αναγνώριση των sessions. Το επόμενο βήμα είναι να αναγνωριστούν τα sessions από τα clickstreams κάθε χρήστη. Η πρόκληση είναι να αναγνωρίσουμε την αρχή και το τέλος ενός session. Αυτό σε ένα σύστημα e-learning είναι αρκετά δύσκολο αφού οι εκπαιδευόμενοι είναι δυνατόν να πλοηγούνται σε άλλα sites αναζητώντας και συλλέγοντας σχετικές πληροφορίες με το μάθημά τους, ενώ το session τους στο συγκεκριμένο e-learning site είναι ακόμα «ανοικτό». Επίσης, ένα learning session μπορεί να καλύπτει περισσότερες μέρες με διαφορετικές προσβάσεις, ενώ πολλές σελίδες του δημιουργούνται δυναμικά με αιτήματα scripts π.χ. σελίδες τεστ. Συνεπώς το heuristic που συνήθως χρησιμοποιείται και ορίζει ένα χρονικό κατώφλι γνωστό και ως page viewing time (Shahabi, et al., 1997) π.χ. στα 30 λεπτά (Cooley et al., 1999) δεν μπορεί να εφαρμοστεί για να διαχωρίσει το clickstream σε sessions. Για το σκοπό αυτό τα sessions αναγνωρίζονται με βάση τις σελίδες

περιεχομένου που ο χρήστης επιλέγει να δει στο site (content pages) και τα μονοπάτια που ακολούθησε για να φτάσει σε αυτές τις σελίδες (navigational pages). Χρησιμοποιούμε την ακόλουθη μέθοδο για να αναγνωρίσουμε τα sessions. Το reference length που εξετάζει το χρόνο που ο εκπαιδευόμενος ξοδεύει στο site. Αν αυτός ο χρόνος είναι μεγαλύτερος από ένα συγκεκριμένο όριο (threshold) τότε θεωρείται ότι η σελίδα περιέχει χρήσιμο περιεχόμενο και ταξινομούνται στις content pages διαφορετικά ταξινομούνται στις navigational pages ενώ θεωρούμε ότι η τελευταία σελίδα που ο χρήστης επισκέπτεται είναι πάντα μια σελίδα περιεχομένου.

- Αναγνώριση των pageviews. Το βήμα αυτό εξαρτάται σημαντικά από τη δομή και την οργάνωση του e-learning site, καθώς επίσης και από το περιεχόμενο των σελίδων και τη γνώση του πεδίου όπως αυτό καθορίστηκε κατά τη σχεδίαση και την υλοποίησή του. Το συγκεκριμένο site αποτελείται από ένα frame οπότε η έννοια του pageview ταυτίζεται με αυτή της σελίδας (HTML αρχείο). Επίσης, η σημασία κάθε pageview είναι διαφορετική αφού αυτά που αντιστοιχούν σε content pages είναι σημαντικότερα από αυτά που αντιστοιχούν σε navigational pages. Για το λόγο αυτό καταγράφεται με κάθε pageview ένας αριθμός από χαρακτηριστικά όπως: το pageviewID (συνήθως το URL αντιπροσωπεύει μοναδικά το pageview), η διάρκεια, το είδος του pageview (content page, navigational page, κλπ.) και άλλα metadata όπως ιδιότητες ή χαρακτηριστικά του περιεχομένου.

- Αντιστοίχιση των εγγραφών log σε συγκεκριμένες e-learning δραστηριότητες. Αυτό το βήμα έχει σαν αποτέλεσμα μια ακολουθία από δραστηριότητες του εκπαιδευόμενου της μορφής: *LogiTestList-Submission Test1-TestList View Message...*

- Συμπλήρωση των μονοπατιών. Το βήμα αυτό αφορά στη συμπλήρωση των μονοπατιών που ο εκπαιδευόμενος ακολούθησε κατά την αλληλεπίδρασή του με το e-learning site. Συχνά παρατηρείται το φαινόμενο να λείπουν από τα αρχεία log εγγραφές για αιτήσεις σελίδων τις οποίες ο χρήστης ζήτησε να επισκεφθεί κατά την πλοήγησή του. Οι αιτήσεις αυτές δεν καταγράφονται στο server επειδή λόγω του caching εξυπηρετούνται από τους proxy servers ή από την τοπική cash στο browser του χρήστη. Το heuristic που

χρησιμοποιούμε για τον εντοπισμό των «χαμένων» αιτήσεων ώστε να συμπληρωθεί το ακριβές μονοπάτι που ακολούθησε ο χρήστης κατά την πλοήγησή του επιλέγει κάθε φορά το συντομότερο μονοπάτι (δηλαδή αυτό που αποτελείται από τις λιγότερες μεταβάσεις μεταξύ των σελίδων).

- Ομαδοποίηση των sessions πρόσβασης κάθε εκπαιδευομένου με στόχο τον καθορισμό των learning sessions. Στο βήμα αυτό εντοπίζουμε και ομαδοποιούμε τα sessions πρόσβασης ενός χρήστη με στόχο τον καθορισμό των learning sessions που έχουν σημασιολογική αξία. Έτσι επιτυγχάνεται σε μεγαλύτερο βαθμό διάσπαση μέσα στο session και παράγονται δυναμικές ομαδοποιήσεις (clusters) με ιδιαίτερη σημασία.

- Ολοκλήρωση με άλλα δεδομένα για τους εκπαιδευόμενους ή για ομάδες εκπαιδευόμενων. Τέλος, στο βήμα αυτό συνδυάζουμε τα δεδομένα από τα καθαρισμένα clickstreams με υπάρχοντα δεδομένα για τους εκπαιδευόμενους. Αυτά μπορεί να προέρχονται από τα προφίλ τους, στοιχεία από την αξιολόγησή τους, κλπ. Για παράδειγμα, συνδυάζοντας τις βαθμολογίες που σχετίζονται με την ολοκλήρωση συγκεκριμένων δραστηριοτήτων με την ακολουθία των γεγονότων που οδήγησαν σε αυτές τις δραστηριότητες μπορούν να αναγνωριστούν χρήσιμα πρότυπα ώστε να αναγνωριστούν ακολουθίες δραστηριοτήτων που οδηγούν σε καλά αποτελέσματα και άλλες που οδηγούν σε ότι και τόσο καλά.

Υπάρχουν, όμως, κάποια τεχνικά εμπόδια που πρέπει να υπερπηδηθούν στη φάση της προετοιμασίας και προ-επεξεργασίας δεδομένων. Έχουν εφαρμοστεί πολλές τέτοιοι μέθοδοι για το σκοπό αυτό και η αίσθηση και η αίσθηση είναι ότι όσο πιο ακριβής είναι μια μέθοδος τόσο περισσότερο παραβιάζει την ιδιωτικότητα του χρήστη. Αν υποθεθεί ότι ένας μοναδικός χρήστης χαρακτηρίζεται από ένα ζεύγος IP διεύθυνση/agent, αυτό δεν ισχύει πάντα γιατί πολλοί χρήστες ίσως χρησιμοποιούν τον ίδιο υπολογιστή για την πρόσβαση στον ιστό ενώ και ο ίδιος χρήστης μπορεί να προσπελαύνει τον ιστό από διάφορους υπολογιστές. Ένας ενσωματωμένος αριθμός συνόδου που απαιτεί δυναμικούς κόμβους

ιστού, μπορεί να αναγνωρίσει τους διάφορους χρήστες από το ίδιο ζεύγος IP/Agent αλλά αποτυγχάνει για τον ίδιο χρήστη από διαφορετικές IPs.

Τα cookies και οι πράκτορες λογισμικού επιτυγχάνουν και τους δύο άνω στόχους αλλά δεν είναι πολύ καλά αποδεκτά από τους περισσότερους χρήστες.

4.2.3 Εξατομικευμένη Έξοδος

Έπειτα από τη συλλογή των κατάλληλων δεδομένων εισόδου, την αποθήκευσή τους χρησιμοποιώντας την κατάλληλη αναπαράσταση και την ανάλυσή τους για την εξαγωγή δευτερευόντων συμπερασμάτων μένει να εξερευνηθεί και να αποφασιστεί το είδος των προσαρμογών που πρέπει να αναπτύξει ένας κόμβος ιστού για την εξατομίκευσή του. Οι προσαρμογές αυτές μπορούν να λάβουν χώρα σε διαφορετικά επίπεδα:

- *Περιεχόμενο*: Τυπικές εφαρμογές τέτοιων προσαρμογών περιλαμβάνουν προαιρετικές επεξηγήσεις και πρόσθετες πληροφορίες, προσωποποιημένες συστάσεις, θεωρητικές παρουσιάσεις και άλλα.

- *Δομή*: Αναφέρεται σε αλλαγές στη δομή συνδέσεων εγγράφων υποκειμένου ή στην παρουσίασή τους.

- *Παρουσίαση και μορφή των media*: Το περιεχόμενο της πληροφορίας παραμένει το ίδιο αλλά η μορφή του και ο τρόπος εμφάνισης αλλάζουν.

4.3 ΠΑΡΑΔΕΙΓΜΑΤΑ ΕΞΑΤΟΜΙΚΕΥΣΗΣ

Οι ιστοσελίδες προκειμένου να αντεπεξέλθουν στην ανταγωνιστικότητα προσπάθησαν να προσελκύσουν ή να κρατήσουν τους χρήστες τους προσφέροντας υπηρεσίες που θα εξασφάλιζαν γρήγορη, ασφαλή και αποτελεσματική αναζήτηση.

Η εξατομίκευση μπορεί να επιτευχθεί χρησιμοποιώντας διάφορους τύπους “MyLibrary”. «Mylibrary» είναι μία έννοια η οποία προσδιορίζει κάτι πολύ οικείο σε σχέση με τη βιβλιοθήκη. Τη βιβλιοθήκη μου. Κάτι που το ξέρω καλά, που μπορώ να το φτιάξω απ’ την αρχή που μπορώ να έχω μέσα τα δικά μου πράγματα που μπορώ να το αλλάζω, που θα είναι ο χώρος μου τον οποίο εάν θέλω μπορώ να μοιραστώ με κάποιους άλλους(Denice A:2003).

Η έννοια του my χρησιμοποιείται πολύ σε εφαρμογές που στοχεύουν σε εξατομικευμένες υπηρεσίες, αρχικά στο ηλεκτρονικό εμπόριο και στη συνέχεια σε πολλές υπηρεσίες οι οποίες υφίστανται μέσω του Διαδικτύου. Επομένως αφού MyYahoo, MyCNN, γιατί όχι και Mylibrary.

Η έννοια Mylibrary χρησιμοποιείται ως ομπρέλα για να περιγράψει το σύνολο των υπηρεσιών customization και personalization στο χώρο των ψηφιακών βιβλιοθηκών. Σε άλλες περιπτώσεις η υπηρεσία αυτή εμφανίζεται ως MyGateway, MyWelch, MyUcla, MyDspace κλπ. Ουσιαστικά δεν πρόκειται ούτε για συγκεκριμένο μοντέλο ούτε για συγκεκριμένο λογισμικό αλλά είναι η εννοιολογική απόδοση της παροχής υπηρεσιών εξατομίκευσης από τις βιβλιοθήκες(Denice A:2003)..

Οι βιβλιοθήκες με τη συνεργασία καινοτόμων βιβλιοθηκάρων οδηγήθηκαν στην υιοθέτηση της ιδέας «Mylibrary» γιατί οι χρήστες προτιμούν να έχουν εξατομικευμένη πρόσβαση στο υλικό των βιβλιοθηκών, να προσαρμόζουν τις διαθέσιμες πηγές στις δικές τους προσωπικές απαιτήσεις και στα ιδιαίτερα ενδιαφέροντά τους. , να έχουν τη δυνατότητα

να ψάχνουν τις τοπικές βάσεις δεδομένων και τον παγκόσμιο ιστό με περισσότερη λειτουργικότητα.

Εξατομικευμένες υπηρεσίες βιβλιοθηκών με το πρόθεμα «MyLibrary» έχουν υλοποιήσει αμερικάνικα κυρίως πανεπιστήμια και ερευνητικά ινστιτούτα. Οι υλοποιήσεις αυτές, παρά το γεγονός ότι ο στόχος τους είναι ίδιος, ποικίλουν και όσον αφορά το επίπεδο εξατομίκευσης και όσον αφορά τον τρόπο υλοποίησης. Το κάθε έργο(project) χρησιμοποίησε διαφορετικά εργαλεία ανάπτυξης και περιβάλλοντα, άλλες γλώσσες προγραμματισμού, άλλο λογισμικό κλπ. Η κάθε εφαρμογή στηρίχτηκε σε δικό της κώδικα. Μερικά έργα βεβαίως ενδιαφέρονται να μοιράσουν τον κώδικά τους (VCU, UW, NC State) και κάποια ήδη το έχουν κάνει.

Χρησιμοποιώντας το «MyLibrary», οι βιβλιοθηκάριοι και οι χρήστες μπορούν να φτιάξουν μια demo βιβλιοθήκη με προϊόντα που θα είχαν ενδιαφέρον γι' αυτούς. Επίσης οι χρήστες εργαζόμενοι από το σπίτι ή αυτοί που έχουν εξ' αποστάσεως πρόσβαση, μπορούν εύκολα να έχουν πρόσβαση σε οποιοδήποτε ηλεκτρονικό άρθρο σε περιοδικά και σε όλους τους web συνδέσμους που χρειάζονται χωρίς κανένα κόστος μετακίνησης.

Οι υποστηριχτές και πρωτεργάτες των εφαρμογών αυτών συμπεραίνουν ότι οι περισσότεροι που χρησιμοποιούν υπηρεσίες MyLibrary είναι πολύ ενθουσιασμένοι, επειδή μπορούν να κάνουν τις επιλογές τους, μπορούν να έχουν πρόσβαση σ' αυτές εξαιρετικά εύκολα, από οπουδήποτε, και επιπροσθέτως είναι προσβάσιμα από μία τοποθεσία.

Όμως η Susan Gibbons αναφέρει ότι οι χρήστες επηρεασμένοι από τις αυξημένες εξατομικευμένες υπηρεσίες εμπορικών εφαρμογών όπως Amazon.com κλπ φτάνουν στις σελίδες των βιβλιοθηκών με μεγαλύτερες προσδοκίες οι οποίες δεν ικανοποιούνται λόγω του στατικού χαρακτήρα αυτών των υπηρεσιών που προσπαθούν να ανταποκριθούν στις ποικίλες ανάγκες τους με ένα απλό «one – size-fits - all» interface (James C. F: 1999).

Σε πολλές αναφορές επισημαίνεται ότι ο αριθμός των χρηστών που μπαίνουν στη διαδικασία να δημιουργήσουν προφίλ σε Mylibrary συστήματα είναι πολύ χαμηλός. Σπάνια υπερβαίνουν το 10% της πιθανής κοινότητας χρηστών. Ακολουθώς αναφέρονται πιθανοί λόγοι που οδηγούν στη χαμηλή χρηστικότητα.

- Απαιτείται ενεργός συμμετοχή των χρηστών στην προσπάθεια να δημιουργήσουν ένα λογαριασμό το οποίο συνεπάγεται να αφιερώσουν χρόνο.

- Οι χρήστες δεν αντιλαμβάνονται εξ' αρχής τα οφέλη έτσι ώστε να το χρησιμοποιούν συχνά .

- Οι χρήστες ανησυχούν για το πως θα χρησιμοποιηθούν τα προσωπικά τους δεδομένα.

- Τα «Mylibrary» συστήματα στηρίζουν τη δημιουργία του προφίλ σε ένα συγκεκριμένο θέμα ή μια θεματική περιοχή. Αυτό διευκολύνει την εξειδικευμένη έρευνα (μεταπτυχιακούς φοιτητές ή τους καθηγητές ή ερευνητές) αλλά όχι τους προπτυχιακούς φοιτητές οι οποίοι σε μία πανεπιστημιακή κοινότητα αποτελούν την πλειοψηφία.

- Πολλές εφαρμογές στηρίζουν την εξατομίκευση κυρίως στο ψηφιακό τους υλικό με αποτέλεσμα κάτι τέτοιο να μη λειτουργεί ενθαρρυντικά στην προσέλκυση ερευνητών από το χώρο των ανθρωπιστικών και κοινωνικών επιστημών όπου το μεγαλύτερο μέρος των πηγών αποτελείται από συμβατικό υλικό.

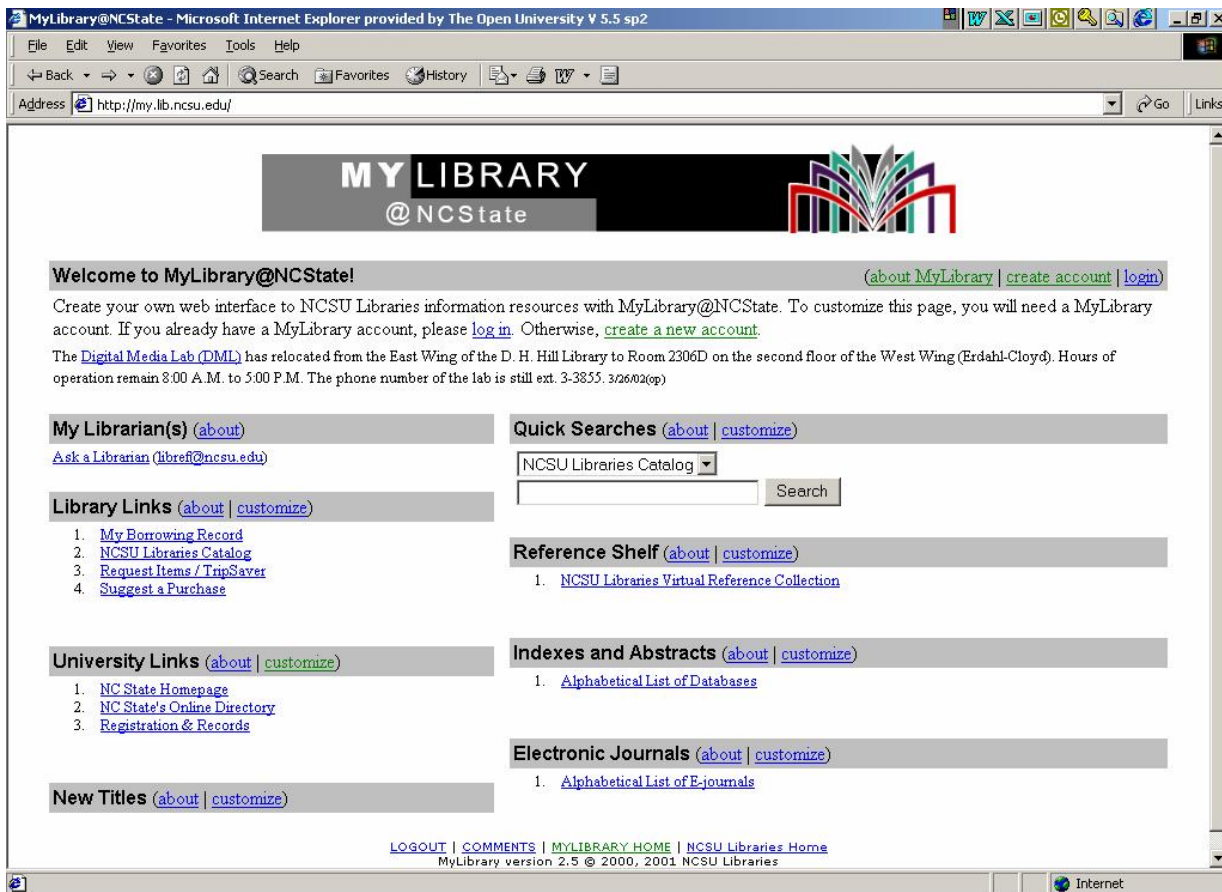
- Η χρήση εξατομικευμένων υπηρεσιών ποικίλει ανάλογα και με το επίπεδο εξοικείωσης των χρηστών με τέτοια συστήματα. Εάν οι χρήστες εκπαιδεύονται τότε είναι πιο εύκολο γι' αυτούς να τα χρησιμοποιούν πιο συχνά. Σε έρευνα η οποία αναζητούσε μεταξύ άλλων και τους λόγους της χαμηλής χρηστικότητας εξατομικευμένων υπηρεσιών στο London Business School, University of Hertfordshire, και London School of Economics, διαπιστώθηκε ότι μία τέτοια υπηρεσία δεν είναι εύκολο να χρησιμοποιηθεί από ανθρώπους οι οποίοι δεν έχουν εκπαιδευτεί για τη χρήση της (Anne G:2001).

- Πολλές εφαρμογές εξατομίκευσης ψηφιακών βιβλιοθηκών Mylibrary έχουν εστιάσει να εφαρμόσουν βασική εξατομίκευση και στοιχειώδη συστήματα σύστασης (recommended

systems). Αυτές οι εφαρμογές εξατομίκευσης δεν δίνουν αξία στις ψηφιακές βιβλιοθήκες και βεβαίως δεν οδηγούν τις ψηφιακές βιβλιοθήκες στο επόμενο βήμα .

Το 1998 North Carolina State University, προέβη στη δημιουργία του MyLibrary@NCState. Ο Eric Lease Morgan, Network Technologies development librarian στο NCState. ένας από τους πρωτεργάτες της υλοποίησης του Mylibrary@NCState παρομοιάζει το Mylibrary@NCState με το My yahoo, το οποίο αντί για καιρό και ωροσκόπια επιτρέπει στους χρήστες να επιλέξουν τις αγαπημένες τους συνδρομές σε βάσεις δεδομένων, ηλεκτρονικά περιοδικά κλπ.. Το Mylibrary@NCState προσφέρει εκτεταμένες δυνατότητας έρευνας με επιλογές από έναν μεγάλο αριθμό βάσεων δεδομένων και μηχανών αναζήτησης. Κίνητρα-Στόχοι του Mylibrary@NCState(Anne G:2001):

- Ø Περιορισμός της πληθώρας των διαθέσιμων στο διαδίκτυο πληροφοριών.
- Ø Δυνατότητα από πλευράς των χρηστών, δημιουργίας των προσωπικών τους σελίδων, επιλέγοντας τους τίτλους ή την ποσότητα των πληροφοριών σύμφωνα με τις επιθυμίες τους.
- Ø Χρήση των πόρων της βιβλιοθήκης από τους χρήστες της, μέσα από ένα εξατομικευμένο Website interface.



Εικόνα 1. My library

Το κύριο χαρακτηριστικό του είναι η ύπαρξη τμημάτων (sections). Οι χρήστες έχουν τη δυνατότητα να προσαρμόζουν (customize) κάθε τμήμα, να δηλώνουν Bookmarks, να αποθηκεύουν προηγούμενες έρευνες, καθώς επίσης να καθορίζουν τη μορφή εμφάνισης των σελίδων (χρώμα, διάταξη κλπ.). Κάθε τμήμα έχει μία περιοχή βοήθειας εξηγώντας στους χρήστες το ρόλο του. Υπάρχουν σύνδεσμοι 'live chat' για συνομιλία με το βιβλιοθηκάριο σε πραγματικό χρόνο.

Ø MyLibrarian section: Στο τμήμα αυτό περιέχεται το χαρακτηριστικό "Ask a Librarian" με το οποίο ο χρήστης έχει τη δυνατότητα να κάνει chat με τους

βιβλιοθηκονόμους σε πραγματικό χρόνο. Επίσης παρέχονται πληροφορίες σχετικά με subject Specialists μαζί με ένα “Frequently Asked Questions τμήμα”

Ø Library Links section: Το τμήμα αυτό περιλαμβάνει συνδέσμους με τη βιβλιοθήκη του Πανεπιστημίου (κατάλογος δανεισμός τον κατάλογο, αιτήσεις).

Ø University Links section: Εδώ περιέχονται σύνδεσμοι σε web sites και πόρους του NCState, την κεντρική ιστοσελίδα του NCState, τα πανεπιστημιακά τμήματα και ένας σύνδεσμοι που επιτρέπει στους χρήστες να βλέπουν τα μαθήματα και τους βαθμούς τους.

Ø Quick Search Section: Το τμήμα αυτό δίνει τη δυνατότητα στο χρήστη να κάνει Keyword αναζήτηση (από ένα search box) σε διαφορετικές μηχανές αναζήτησης οι οποίες περιλαμβάνουν τον κατάλογο της βιβλιοθήκης, ένα λεξικό, ένα θησαυρό και το διαδίκτυο μέσω της μηχανής αναζήτησης του Google.

Ø Reference Shelf Section: Ο χρήστης έχει τη δυνατότητα να πάρει στοιχεία από κάθε θεματική περιοχή και να τα προσθέσει σ’ αυτό το τμήμα. Επίσης του δίνεται η δυνατότητα να μετακινήσει στοιχεία τα οποία δεν χρειάζεται και να προσαρμόσει αυτά που θεωρεί απαραίτητα.

Ø New Titles Section. Το σύστημα δίνει τη δυνατότητα στο χρήστη , εφόσον το δηλώσει να παίρνει μηνύματα για εκθέσεις, νέους τίτλους κλπ . Οι χρήστες δημιουργούν ένα Current Awareness Manager προφίλ και αυτό τους δίνει δυνατότητα να ενημερώνονται, μέσω τακτικών μηνυμάτων, για προσκτήσεις υλικού, χρησιμοποιώντας το LCC εδραιωμένο στο προφίλ τους

Ø Indexes & Abstracts Section: Εδώ είναι διαθέσιμη μία αλφαβητική λίστα ηλεκτρονικών βάσεων δεδομένων

Ø Message of the Day section: Εδώ ο χρήστης διαχειρίζεται τα μηνύματα του.

Ø Sandbox Area: Εδώ δίνεται η δυνατότητα σε χρήστες χωρίς username και log-on να αλληλεπιδράσουν με το σύστημα ως φιλοξενούμενοι, να αλλάξουν πράγματα και να το γνωρίσουν καλύτερα χωρίς βεβαίως να μπορούν να σώσουν τις αλλαγές ή τις έρευνές τους.

4.3.1 Ηλεκτρονικές Βιβλιοθήκες

Οι παραδοσιακές βιβλιοθήκες ακολουθώντας και αυτές τη ραγδαία εξέλιξη της τεχνολογίας, άρχισαν να εκσυγχρονίζονται ολοένα και περισσότερο. Οι πληροφορίες αποθηκεύονται πλέον σε CD-ROM, δημιουργούνται ηλεκτρονικοί κατάλογοι χρηστών, ενώ αρχίζει η κατασκευή ηλεκτρονικών ευρετηρίων και βάσεων δεδομένων. Με την εμφάνιση προγραμμάτων επεξεργασίας και διάδοσης ηλεκτρονικής εικόνας, ήχου και βίντεο, κάθε μορφή πληροφορίας μπορεί πλέον να αποθηκευτεί ηλεκτρονικά.

Οι ηλεκτρονικές βιβλιοθήκες σήμερα έχουν εφαρμογή σε πάρα πολλούς τομείς, επιστημονικούς και μη, ενώ παρατηρείται μια συνεχής διεύρυνση του συνόλου των πεδίων δραστηριοποίησής τους.

Οι Ηλεκτρονικές βιβλιοθήκες λειτουργούν σε διαφορετική βάση και φιλοσοφία από τα site εύρεσης. Οι Ηλεκτρονικές Βιβλιοθήκες είναι εστιασμένες σε συγκεκριμένα πεδία έρευνας, τα οποία γνωρίζει ο χρήστης και όταν επιλέγει να τις χρησιμοποιήσει είναι εστιασμένος και ξέρει και που και πως θα ψάξει.

Αυτή η πρόσβαση συνήθως δεν είναι ελεύθερη και ο χρήστης πρέπει να καταβάλλει κάποιο κόστος χρήσης αυτών των υπηρεσιών. Ωστόσο όλα σχεδόν τα Πανεπιστήμια είναι εγγεγραμμένα σε αυτές τις υπηρεσίες και κάποιος χρήστης που συνδέεται στο δίκτυο μέσω των πόρων του Πανεπιστημίου μπορεί να τις χρησιμοποιήσει με μηδενικό κόστος. Προκειμένου να γίνει δυνατός αυτός ο τύπος ελεύθερης πρόσβασης, το λογισμικό του ιδρύματος που παρέχει την ηλεκτρονική βιβλιοθήκη, ελέγχει τη διεύθυνση IP του χρήστη που έχει συνδεθεί σε αυτή. Εάν αυτή η διεύθυνση ανήκει στο σύνολο των διευθύνσεων IP που έχουν χορηγηθεί στο Πανεπιστήμιο, η εφαρμογή επιτρέπει την ανάκτηση ηλεκτρονικών άρθρων σε ψηφιακή μορφή, ενώ στην αντίθετη περίπτωση ο χρήστης θα πρέπει να συνδεθεί με κάποιο όνομα και κάποιο κωδικό πρόσβασης. Συγχρόνως οι ηλεκτρονικές βιβλιοθήκες θα

του βγάλουν ακριβές αποτελέσματα, μια και θα του δώσουν τη δυνατότητα να ψάξει με συγκεκριμένες λέξεις κλειδιά και δίνοντας συγκεκριμένα δεδομένα εύρεσης όπου η ανάκτηση ερευνητικών εργασιών και τεχνικών αναφορών οι οποίες στις πιο πολλές περιπτώσεις είναι αποθηκευμένες σε αρχεία τύπου PDF .Ο συνδυασμός συνειδητοποιημένου ερευνητή και εστιασμένου site, δίνει καλύτερα αποτελέσματα από τα διάφορα site εύρεσης.

4.3.2 Τομείς Λειτουργίας Ηλεκτρονικών Βιβλιοθηκών

Πολλοί είναι οι τομείς, επιστημονικοί και μη, για τους οποίους έχουν αναπτυχθεί συστήματα ηλεκτρονικών βιβλιοθηκών. Οι κυριότερες περιοχές δραστηριοποίησής τους επικεντρώνονται στις παρακάτω:

- Διαχείριση γνώσης - περιεχομένου.

Πρόκειται για βιβλιοθήκες που διατηρούν μεγάλο όγκο πληροφοριών διαφόρων τομέων και παρέχουν πρόσβαση σ' αυτές.

- Σχολική επικοινωνία, εκπαίδευση και έρευνα.

Βιβλιοθήκες που αναπτύσσονται προκειμένου να εξυπηρετήσουν τις ανάγκες της εκπαίδευσης για έρευνα και πληροφόρηση. Περιλαμβάνουν ηλεκτρονικές εφημερίδες, διάφορες δημοσιεύσεις, βιβλία και γενικά εκπαιδευτικό υλικό σε ηλεκτρονική μορφή. Επίσης, προωθούν την ηλεκτρονική μάθηση (e-learning) μέσω σχετικών ηλεκτρονικών προγραμμάτων που παρέχουν.

Το πρόβλημα των ηλεκτρονικών βιβλιοθηκών αναφέρεται ότι λειτουργούν μόνο συνδρομητικά είτε επί πληρωμή είτε για ακαδημαϊκές κοινότητες, είτε για βιβλιοθήκες. Οι

βιβλιοθήκες αυτές δεν είναι προσβάσιμες από το σύνολο των χρηστών. Παρακάτω παραθέτουμε κάποιες ηλεκτρονικές βιβλιοθήκες αναφέροντας και μια μικρή περιγραφή για κάθε μια από αυτές(ΕΚΠΑ, 2006):

1. MATHSCINET

Δυνατότητα αναζήτησης σε βιβλία, περιοδικά και πρακτικά συνεδρίων γύρω από τα Μαθηματικά και άλλες σχετικές επιστήμες. Παρέχονται τα βιβλιογραφικά στοιχεία για άρθρα περιοδικών, συνεδρίων και βιβλίων καθώς και τις περιλήψεις αυτών, ενώ σε αρκετές περιπτώσεις διατίθεται το πλήρες κείμενο του άρθρου.

2. EVIDENCE BASED MEDICINE REVIEWS - COCHRANE LIBRARY

Η βάση παρέχει πρόσβαση σε περιοδικά, άρθρα, περιλήψεις και άλλες ηλεκτρονικές πηγές σχετικά με την τεκμηριωμένη ιατρική και την κλινική ιατρική. Συγκεκριμένα περιλαμβάνει τις εξής επιμέρους βάσεις:

- I. EBM Reviews - ACP Journal Club
- II. EBM Reviews - Cochrane Central Register of Controlled Trials.
- III. EBM Reviews - Cochrane Database of Systematic Reviews.
- IV. EBM Reviews - Database of Abstracts of Reviews of Effects

3. JSTOR

Προσφέρει τη δυνατότητα ανάκτησης σαρωμένων σελίδων και τευχών περιοδικών που έχουν εκδοθεί, καλύπτοντας μεγάλη ποικιλία θεμάτων. Συγκεκριμένα παρέχουμε πρόσβαση στις συλλογές Arts & Sciences I, II, III. Για περισσότερες πληροφορίες σχετικά με την τρέχουσα συλλογή του JSTOR πατήστε.

4.SCOPUS: Παρέχει πρόσβαση στη βάση δεδομένων του Elsevier με δυνατότητα για ταυτόχρονη αναζήτηση σε περισσότερους από 14.000 τίτλους επιστημονικών περιοδικών

καθώς και στο Διαδίκτυο. Επίσης στην εμφάνιση αποτελεσμάτων παρέχει τη δυνατότητα για αυτόματη σύνδεση με το πλήρες κείμενο του άρθρου, αν ο τίτλος ανήκει στις συνδρομές του Συνδέσμου Ελληνικών Ακαδημαϊκών Βιβλιοθηκών.

5. WILSON WEB: Παρέχει πρόσβαση σε 3 επιμέρους βάσεις:

I. Art Index και Art Index Retrospective: 55 χρόνια δημοσιεύσεων σχετικών με την Τέχνη με αναφορές περιοδικών, επετηρίδων και δελτίων μουσείων.

II. Art Museum Image Gallery: πλούσια ψηφιακή πηγή εικόνων τέχνης και σχετικών πολυμέσων από τις συλλογές διακεκριμένων μουσείων.

III. Wilson OmniFile Select Full Text: πλήρες κείμενο σε περισσότερους από 1600 τίτλους περιοδικών.

6. ΖΕΦΥΡΟΣ: Ο Ζέφυρος είναι μια πύλη πρόσβασης για τον εντοπισμό των διαθέσιμων βιβλίων, περιοδικών και άλλου υλικού στους καταλόγους των Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Ο Ζέφυρος δημιουργήθηκε από την Υπηρεσία Τεχνολογιών και Πληροφόρησης της Βιβλιοθήκης του Πανεπιστημίου Κρήτης. Βασικός άξονας κατά τον σχεδιασμό και την υλοποίησή του ήταν να αποτελέσει ένα ομοιογενές περιβάλλον για την αναζήτηση και ανάκτηση βιβλιογραφικών πληροφοριών από βάσεις δεδομένων των Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Παράλληλα δόθηκε βαρύτητα στην ανάπτυξη προηγμένων υπηρεσιών προς τους επισκέπτες, όπως δυνατότητα δημιουργίας προσωπικής σελίδας για αποθήκευση αναζητήσεων και εγγραφών κ.α.

ΚΕΦΑΛΑΙΟ 5

ΕΡΓΑΛΕΙΑ ΕΞΑΤΟΜΙΚΕΥΜΕΝΩΝ ΥΠΗΡΕΣΙΩΝ ΣΤΟ ΔΙΑΔΙΚΤΥΟ

Ο Berners-Lee έλεγε ότι: «Έχω ένα όραμα για τον ιστό... και περιλαμβάνει δυο μέρη. Στο πρώτο ο ιστός γίνεται ένα πολύ ισχυρό μέσο για τη συνεργασία των ανθρώπων. Πάντοτε φανταζόμουν τον πληροφοριακό χώρο σαν κάτι το οποίο ο καθένας έχει άμεση και ενστικτώδη πρόσβαση και όχι μόνο για περιήγηση αλλά και για να δημιουργήσει. Στο δεύτερο μέρος οι συνεργασίες επεκτείνονται στους υπολογιστές. Μηχανές γίνονται ικανές να αναλύσουν όλες τις πληροφορίες του ιστού και να δώσουν τις κατάλληλες πληροφορίες στους χρήστες του. Όταν δημιουργηθεί ο σημασιολογικός ιστός το διαδίκτυο θα διαχειρίζεται τα πάντα και μέσα από τη πληροφόρηση την οποία θα δίνει οι άνθρωποι θα εμπνέονται και θα δημιουργούν» (Berners-Lee T, 2006:36)

Με τη παραπάνω ρήση ο Berners-Lee ξεκινά την αναφορά στο βιβλίο του «*Weaving the Web: The Past, Present and Future of the World Wide Web by its Inventor*» για τη μελλοντική μορφή του Ίντερνετ, δηλώνοντας μ' αυτό τον τρόπο την ανάγκη, των ανθρώπων που μελετούν και ασχολούνται με το διαδίκτυο να υπάρξει σύντομα, μια βελτίωση της λειτουργίας του αλλά και του τρόπου με τον οποίο βοηθά τον σύγχρονο άνθρωπο, δίνοντας του την πληροφόρηση που του αρμόζει και μπορεί να τον οδηγήσει σε σωστά και όχι χαλεπά μονοπάτια.

Προκειμένου στο συγκεκριμένο κεφάλαιο ν' αναλύσουμε τη σημερινή λειτουργία του διαδικτύου, τις τεχνικές αναζήτησης, τους λόγους που οδηγούν στη συγκέντρωση σε πολλές περιπτώσεις άτοπων και μη χρήσιμων πληροφοριών και γενικά να κάνουμε μια ουσιαστική περιγραφή διαφόρων μεγάλων ιστοτόπων, αξιολογώντας τη σημερινή τους λειτουργία, η οποία οδηγεί σε πολλές περιπτώσεις στο παράδοξο, θα κάνουμε μια περιγραφή της έναρξης

αυτής της λειτουργίας, την αρχική της φιλοσοφία και πως αυτή οδηγήθηκε σε λαθεμένους δρόμους, τους δρόμους της εξόρισης πολυποίκιλων και πολλές φορές ανούσιων πληροφοριών για το χρήστη.

Συγκεκριμένα η αρχική προσπάθεια πλοήγησης των χρηστών στο διαδίκτυο ήταν ανεπιτυχής. Η επίλυση της και η μετέπειτα βελτιωμένη μορφή της σε σχέση με την αναζήτηση, ήταν απόρροια της ανάπτυξης των εργαλείων αναζήτησης. Συγκεκριμένα αυτές ήταν: 1. Η Yahoo, 2. Η AltaVista, 3. Η Infoseek, 4. Η Excite, 5. Ο Lycos, 6. Το Hotbot, 7. Η Google.

Όμως η συνεχής και χωρίς έλεγχο αύξηση των δικτυακών τόπων, οδήγησε στην αύξηση του όγκου του υλικού που ήταν αποθηκευμένο στον ιστό, οπότε και στην αδύνατη εύρεση των κατάλληλων κάθε φορά πληροφοριών. Αυτό ήταν αποτέλεσμα κυρίως της μη οργανωμένης αναζήτησης με τη βοήθεια συγκεκριμένων εφαρμογών. Όπως αναφέρει Prettejohn, M το 1996 στο βιβλίο του «*The First Year: August 1995-August 1996*» το 1995 υπήρχαν περίπου 10.000 δικτυακοί τόποι, ενώ 8 χρόνια αργότερα ο αριθμός τους θα ξεπερνούσε τα 35. εκατομμύρια (Prettejohn M, 1996:24).

Η δαιδαλώδης κατάσταση της περιόδου 1995-2005 οδήγησε στη σημερινή δυσχερή κατάσταση, κάτω από την οποία λειτουργεί, αναπτύσσεται και λειτουργεί το διαδίκτυο. Η εστιασμένη προσπάθεια για βελτίωση στην εύρεση δεδομένων, είχε τελικά τα αντίθετα αποτελέσματα, μια και το διαδίκτυο βομβαρδίστηκε πολύ γρήγορα με αμέτρητες πληροφορίες, οι οποίες δίνονταν στο βωμό του ανταγωνισμού και της προσπάθειας για διαφοροποίηση. Σήμερα η κατάσταση αυτή θεωρείται ανεξέλεγκτη (Prettejohn M, 1996:24).

Μέσα στη δεκαετία του 1990, η οποία θεωρείται η περίοδος των μεγάλων ανακατατάξεων αλλά και των μεγάλων αποκλίσεων από την αρχική σκέψη και προσπάθεια ανάπτυξης του διαδικτύου, εμφανίστηκαν τα πρώτα εργαλεία αναζήτησης που στις

περισσότερες περιπτώσεις ήταν το αποτέλεσμα της εργασίας ερευνητών και φοιτητών μεγάλων πανεπιστημιακών ιδρυμάτων των ΗΠΑ, οι οποίοι επιδίωκαν με αυτό τον τρόπο να ενισχύσουν την ερευνητική τους προσπάθεια (Βλέπε Πίνακα 1).

Πίνακας 1: Έτος Έναρξης λειτουργίας των σημαντικότερων εργαλείων αναζήτησης και οργανισμός από τον οποίο ξεκίνησε η προσπάθεια.

ΕΡΓΑΛΕΙΑ ΑΝΑΖΗΤΗΣΗ	ΕΤΟΣ ΕΝΑΡΞΗΣ ΛΕΙΤΟΥΡΓΙΑΣ	ΟΡΓΑΝΙΣΜΟΣ ΑΠΟ ΤΟΝ ΟΠΟΙΟ ΞΕΚΙΝΗΣΕ Η ΠΡΟΣΠΑΘΕΙΑ
Lycos	1994	Carnegie Mellon University
WebCrawler	1994	University of Washington
Yahoo	1994	Stanford University
AltaVista	1995	Digital Equipment Corporation
Excite	1995	Excite Inc
Infoseek	1995	Ιδιωτική Προσπάθεια
HotBot	1996	Wired Ventures
Google	1998	Stanford University

Πηγή: Sullivan, D.(2003), Searches per Day,
<http://searchenginewatch.com/reports/article.php/2156461>

5.1 TO YAHOO

Η αρχική φιλοσοφία του Yahoo, το οποίο δημιουργήθηκε το 1994 από τους David Filo και Jerry Yang υποψήφιους διδάκτορες στο Stanford University, αναφερόταν στη δημιουργία καταλόγων με αγαπημένους συνδέσμους επικοινωνίας ώστε να διευκολύνεται η πλοήγηση των χρηστών στον ιστό.

Στη συνέχεια οι κατάλογοι αυτοί απέκτησαν κατηγορίες και υποκατηγορίες δημιουργώντας τη βασική δομή του Yahoo, αλλά και ένα ιστότοπο ο οποίος έδινε τη δυνατότητα στο χρήστη να πλοηγηθεί ευκολότερα και γρηγορότερα σε βασικούς καταλόγους εύρεσης. Ουσιαστικά οι χρήστες ήταν αυτοί που δημιουργούσαν το κατάλογο, με βάση τις προτιμήσεις τους και τις προτεραιότητες τις οποίες είχαν σε σχέση με τις διαδικασίες εύρεσης που ακολουθούσαν. Αρχικά η χρήση γινόταν μόνο από τους ιδρυτές και τους φίλους τους, αλλά στη συνέχεια η φήμη του εξαπλώθηκε στην ευρύτερη κοινότητα των χρηστών του διαδικτύου.

Η βάση της πληροφόρησης της yahoo, ήταν η δυνατότητα αξιολόγησης άλλων ιστοσελίδων αλλά και η διευρυμένη παρουσίαση καταλόγων, οι οποίοι είχαν ενδιαφέρον για τους χρήστες, δίνοντας τους τη δυνατότητα να ερευνήσουν σ' αυτούς και να βρουν πιο εύκολα τα στοιχεία τα οποία έψαχναν. Η yahoo εκμεταλλεύτηκε την έλλειψη πληροφόρησης και τη μετέτρεψε, σε έναν «βομβαρδισμό» ανεξέλεγκτων και πολλές φορές χωρίς έλεγχο πληροφοριών (Yahoo, 2007).

Η στήριξη της Yahoo από την εταιρία επενδύσεων χαρτοφυλακίου Sequoia Capital, έδωσε τη δυνατότητα στους ιδρυτές της να τη μετατρέψουν σε μια δυναμική, συνεχώς αναπτυσσόμενη και κερδοφόρα επιχείρηση η οποία δραστηριοποιείται σήμερα σ' όλο τον κόσμο, έχοντας βάσεις και δίκτυο σ' όλες τις ηπείρους. Η Yahoo προσπάθησε να αυξήσει τη

δυναμική της όχι βελτιώνοντας τη δράση της στο τομέα της πληροφόρησης, αλλά παρέχοντας περισσότερες υπηρεσίες στους χρήστες της προκειμένου να καταφέρει να τους εγκλωβίσει στις τάξεις της.

Η κίνηση της αυτή την χαρακτηρίζει στον ευρύτερο ανταγωνισμό μέχρι και σήμερα. Η στρατηγική που ακολούθησε εστιάσθηκε στην οριζόντια ολοκλήρωση, εξαγοράζοντας μπρος και πίσω εταιρίες με παράπλευρες δραστηριότητες, οι οποίες τη βοήθησαν να αυξήσει την γκάμα των υπηρεσιών της ενισχύοντας τη με υπηρεσίες ηλεκτρονικού ταχυδρομείου, με οργάνωση θεματικών ενοτήτων, με υπηρεσίες φιλοξενίας ιστοσελίδων κλπ. (Παπαδάκης, 2001:10-100).

Οι κινήσεις της τη κατέστησαν ανταγωνιστική, δεν της έδωσαν όμως τη δυνατότητα να εστιάσει και να βελτιώσει τη δυναμική εύρεσης την οποία και παρείχε στους χρήστες της. Η Yahoo από τη στιγμή της ίδρυσης της προχωράει στη συνεχή αύξηση των παρεχόμενων υπηρεσιών της, χωρίς να εστιάζει στη βελτίωση των δυνατοτήτων εύρεσης που δίνει. Οι πρόσφατες κινήσεις της στη παροχή υπηρεσιών αναφέρονται στην παροχή στους χρήστες της προσωπικού ημερολογίου, αλλά και άλλων δυνατοτήτων διαχείρισης του χρόνου τους και της καθημερινότητας τους. Τέλος η Yahoo τα τελευταία χρόνια δρα και στο τομέα, του ηλεκτρονικού εμπορίου δίνοντας τη δυνατότητα στους χρήστες της να προβούν σ' αγορές μέσω του διαδικτύου αλλά και να κάνουν ένα μεγάλο εύρος συναλλαγών (Yahoo, 2007).

Αντίθετα με το Yahoo όλες οι ανταγωνίστριες ιστοσελίδες λειτούργησαν ως μηχανές αναζήτησης, δεν αποτέλεσαν ένα απλό πίνακα περιεχομένων αλλά μια αράχνη, η οποία έψαχνε και ψάχνει στον ιστό και διαμορφώνει ευρετήριο με τις λέξεις κλειδιά που βρίσκει. Σαν αποτέλεσμα συνήθως επιστρέφει πολύ περισσότερες θέσεις στις οποίες υπάρχουν λέξεις που ταιριάζουν με τα κριτήρια της αναζήτησης σε σχέση με τον τρόπο που λειτουργεί το Yahoo. Το πρόβλημα αναφέρεται στο ότι δεν υπάρχει κάποια λογική και στοιχειώδης επιμέλεια, οπότε πολλές φορές ο χρήστης χάνεται ανάμεσα στις χιλιάδες των πληροφοριών

ενώ οι περισσότερες από αυτές δε σχετίζονται με το θέμα το οποίο τον ενδιαφέρει. Με βάση τα παραπάνω κατανοούμε ότι το φαινόμενο της υπερπληροφόρησης σε σχέση με τις υπάρχουσες δυναμικές ιστοσελίδες εστιάζεται σε δυο πεδία αναφοράς(Λέανδρος Ν,2005:88)

1. Στην λειτουργία ως καταλόγου και στην απλή παρουσίαση των ιστοσελίδων που υπάρχουν στο Ίντερνετ και θα μπορούσαν να βοηθήσουν, χωρίς όμως να δίνονται εγγυήσεις.

2. Στη μη καλή γνώση του χρήστη των δυνατοτήτων εύρεσης, οπότε και στη μη σωστή χρήση λέξεων κλειδιών. Αυτό έχει ως αποτέλεσμα την άσκοπη εισροή πληροφοριών αλλά και τους συμβιβασμούς που μπορεί και προτίθεται να κάνει ο χρήστης.

Ο παρακάτω πίνακας είναι ενδεικτικός σε σχέση με τον όγκο των αναζητούμενων πληροφοριών από τους χρήστες, ανά ημέρα τη περίοδο 2003. Ο πίνακας αυτός παρατίθεται προκειμένου να υποδείξει το προβληματισμό σε σχέση με την υπερπληροφόρηση μιας και ο αριθμός εισερχόμενων χρηστών ανά ημέρα είναι τεράστιος, πράγμα που καθιστά ουσιαστικό το πρόβλημα καθώς οι περισσότεροι δεν καταλήγουν στο να βρουν αυτό που πραγματικά τους οδήγησε να ανατρέξουν στο Ίντερνετ.

Πίνακας 2. Στατιστικά Στοιχεία για τα εργαλεία αναζήτησης (Ιανουάριος 2003).

ΕΡΓΑΛΕΙΑ ΑΝΑΖΗΤΗΣΗ	ΩΡΕΣ ΑΝΑΖΗΤΗΣΗΣ ΑΝΑ ΜΗΝΑ (ΣΕ ΕΚΑΤΟΜΜΥΡΙΑ)	ΛΕΠΤΑ ΑΝΑΖΗΤΗΣΗΣ ΑΝΑ ΗΜΕΡΑ (ΣΕ ΕΚΑΤΟΜΜΥΡΙΑ)	ΑΡΙΘΜΟΣ ΑΝΑΖΗΤΗΣΕΩΝ ΑΝΑ ΗΜΕΡΑ (ΣΕ ΕΚΑΤΟΜΜΥΡΙΑ)
Google	18,7	37	112

AOL Search	15,5	31	93
Yahoo	7,1	14	42
MSN Search	5,4	11	32
Ask Jeeves	2,3	5	14
InfoSpace	1,1	2	7
Alta Vista	0,8	2	5
Overture	0,8	2	5
Netscape	0,7	1	4
Earthlink	0,4	1	3
Looksmart	0,2	0	1
Lycos	0,2	0	1
Σύνολο	53,2	106	319

Πηγή: Sullivan, D.(2003), Searches per Day,
<http://searchenginewatch.com/reports/article.php/2156461>

Με βάση το παραπάνω πίνακα κατανοούμε το μέγεθος του προβλήματος μια και η ίδια έρευνα οδηγήθηκε στο συμπέρασμα μέσα από ερωτηματολόγιο το οποίο διανεμήθηκε στους χρήστες των 12 αυτών ιστοσελίδων, ότι περίπου το 55,8% είχε βρει ακριβώς ή σχεδόν ακριβώς αυτό το οποίο ζητούσε μέσα από το διαδίκτυο. Οι περισσότεροι δε από αυτούς θεωρούσαν την μη εύρεση δικό της άγνοια σε σχέση με τις διαδικασίες εύρεσης και όχι πρόβλημα των ίδιων των ιστοσελίδων, οι οποίες τους βομβάρδιζαν με ανούσιες

πληροφορίες, χωρίς όμως να τους βοηθούν να εστιάσουν σε μια θεματολογία σχετική με το αντικείμενο της έρευνας τους.

Ανακεφαλαιώνοντας και σε σχέση με τη σημερινή χρήση των δικτυακών τόπων αλλά και την εισροή χιλιάδων πληροφοριών μέσα από τις καθημερινές ευρέσεις που γίνονται, μπορούμε να πούμε ότι ο τεράστιος όγκος πληροφοριών του Internet σε συνδυασμό με την ελάχιστη οργάνωση που υπάρχει στο Δίκτυο καθιστά πολύ δύσκολη και χρονοβόρα την εύρεση των πληροφοριών ή των εγγράφων που αναζητά ο χρήστης. Για την επίλυση αυτού του προβλήματος έχουν αναπτυχθεί διάφορες μέθοδοι, οι οποίες με βάση και τη παραπάνω ανάλυση καταλήγουμε ότι είναι ανεπαρκείς.

Οι δύο πιο δημοφιλείς είναι οι indexes (ευρετήρια) και οι search engines (μηχανές αναζήτησης). Τα ευρετήρια παρουσιάζουν έναν εξαιρετικά δομημένο τρόπο ανεύρεσης πληροφοριών, ουσιαστικά επιτρέπουν στο χρήστη να πλοηγηθεί στις πληροφορίες μέσω κατηγοριών όπως τέχνη, υπολογιστές, ψυχαγωγία, αθλήματα, κ.ο.κ. Χρησιμοποιώντας τον Web browser αν γίνει κλικ σε μία κατηγορία θα οδηγηθεί ο χρήστης σε μία σειρά υποκατηγοριών. Στην κατηγορία αθλήματα για παράδειγμα θα οδηγηθείτε σε διάφορες κατηγορίες όπως ποδόσφαιρο, μπάσκετ, βόλεϊ, πόλο, κ.ο.κ. Ανάλογα με το μέγεθος του ευρετηρίου μπορεί να υπάρχουν διάφορα επίπεδα υποκατηγοριών(Reid R,2004:20-25).






Όταν μπει ο χρήστης στην υποκατηγορία που έχει επιλέξει θα παρουσιαστεί ένας κατάλογος με τα σχετικά έγγραφα. Για να δει τα εν λόγω έγγραφα αρκεί να κάνει κλικ στα links που αντιστοιχούν. Το Yahoo! Όπως προαναφέραμε αποτελεί το μεγαλύτερο και δημοφιλέστερο ευρετήριο στο Internet. Το Yahoo και τα άλλα ευρετήρια επιτρέπουν στο χρήστη να ψάξει, πληκτρολογώντας λέξεις οι οποίες περιγράφουν τις πληροφορίες που αναζητά. Εν συνεχεία παίρνει ένα σύνολο αποτελεσμάτων τα οποία αποτελούν links σε έγγραφα που πληρούν τα κριτήρια της αναζήτησης. Για να ανακτήσει την πληροφορία αρκεί να κάνει κλικ σε κάποιο link.

Μία ακόμη δημοφιλής μέθοδος αναζήτησης πληροφοριών στο Internet είναι η χρήση των search engines οι οποίες παλαιότερα αποκαλούνταν και Web crawlers ή spiders. Οι μηχανές αναζήτησης λειτουργούν διαφορετικά από τα ευρετήρια. Πρόκειται ουσιαστικά για μαζικές βάσεις δεδομένων οι οποίες καλύπτουν ευρείες περιοχές του Internet. Οι μηχανές αναζήτησης δεν παρουσιάζουν τις πληροφορίες με ιεραρχικό τρόπο. Αντιθέτως, η αναζήτηση γίνεται όπως και σε μία βάση δεδομένων, πληκτρολογώντας λέξεις - κλειδιά οι οποίες περιγράφουν τις πληροφορίες που θέλει ο χρήστης.

5.1.1 Το Προφίλ της Yahoo

Εικόνα 2. Το προφίλ Yahoo

Ελληνικό Yahoo!: Εργαλεία Επικοινωνίας

-  **Yahoo! Mail** - Διαβάστε και στείλτε emails από οπουδήποτε στον κόσμο με το Yahoo! Mail. Μπορείτε να έχετε πρόσβαση ακόμα και στον POP3 mail λογαριασμό σας, μέσω του Web.
-  **Yahoo! Ημερολόγιο** - Κάνει εύκολη την οργάνωση της ημέρας σας με τις υπενθυμίσεις, τις e-mail προσκλήσεις, και τη λίστα εκκρεμοτήτων.
-  **Yahoo! Ατζέντα** - Οργανώστε τις διευθύνσεις των φίλων σας. Μπορείτε να μεταφέρετε όλες τις πληροφορίες για τους φίλους σας από το Microsoft Outlook, από το Netscape Address Book ή από το Palmtop σας.
-  **Yahoo! Messenger** - Στείλτε άμεσα μηνύματα στους φίλους σας ή στους συνεργάτες σας. Σας στέλνει υπενθυμίσεις σε συνεργασία με το Yahoo! Ημερολόγιο, ώστε να μη χάνετε κανένα σημαντικό ραντεβού.
-  **Yahoo! Σημειώσεις** - Αποθηκεύστε όλες τις σημειώσεις σας στο Web. Τώρα μπορείτε να έχετε πρόσβαση στις σημειώσεις σας από οπουδήποτε, οποιαδήποτε στιγμή.

YAHOO

- **Web**
- Images
- Video
- Local
- Shopping
- more
 - Answers
 - Audio
 - Directory
 - Jobs
 - News
 - All Search Services Options

- [Advanced Search](#)
- [Preferences](#)

- [Advertising Programs](#)
- [About This Page](#)

[Yahoo!](#)

• Start typing to see suggestions.

Showing results containing:

Explore concepts: +

• **Search In:**  the Web  pages in German

sec

1. Περιεχόμενα: Παγκοσμιοποίηση, Αριστερά και Περιεκτική Δημοκρατία

Η εγκληματική οικολογία της Νέας Τάξης. Η άνοδος και η πτώση του οικολογικού κινήματος ... Η παγκοσμιοποίηση και ο νέος ανορθολογισμός ...

www.inclusivedemocracy.org/fotopoulos/greek/grbooksglobal/contents.htm - 24k -

Cached

2. Άρθρα σε άλλα περιοδικά και εφημερίδες του Τάκη Φωτόπουλου

3. παγκοσμιοποίηση | Re-public : re-imagining democracy

...φύλο και νέα μέσα, τέχνη και δημοκρατία, χρόνος/πολιτική, ... οικολογία. παγκοσμιότητα. πνευματικά δικαιώματα. πολιτισμικές διαφορές. σοσιαλδημοκρατία ...

www.re-public.gr/?cat=2 - 32k - Cached

4. Νέα Τάξη Πραγμάτων -Παγκοσμιοποίηση

... αδιάλλακτος, μισαλλόδοξος, ακροδεξιός, ρατσιστής και έχει καταρασθεί τη λέξη ανοχή" ... και φορέας της προσαρμογής σε αυτή την παγκοσμιοποίηση, χρησιμοποιώντας ...

www.agiasofia.com/turco/turco4.html - 122k - Cached

5. Καρότο και Μαστίγιο

... νομών: τοπικισμός vs. παγκοσμιοποίηση; Impossible is nothing ... Blog Action Day II: Εσωτερική οικολογία. Blog Action Day I: Το Περιβάλλον τα χει μ' άλλον ...karotomastigio.wordpress.com - 65k - Cached

6. Η ΟΙΚΟΛΟΓΙΑ ΑΠΟΚΤΑ

Η πολιτική οικολογία μπορεί να συμπληρώσει και να στηρίξει τις προσπάθειες που ... τη σχέση της Ελλάδας με την παγκοσμιοποίηση και το τι είδους πολιτική θέλουμε. ...

www.ecogreens.gr/arxeio/nea26.htm - 27k - Cached

7. MONTHLY REVIEW

Αναλύσεις και εξελίξεις. Νέα του MR. Σύνδεσμοι. Θεματικό ευρετήριο. Το δίκτυο του MR: Οικολογία ... Πολιτική | Παγκοσμιοποίηση | Οικολογία | Περιβάλλον ...
monthlyreview.gr/antilogos/greek/kathgories/...?a=oikologia - 32k - Cached

8. A.K.O.A - ΟΡΓΑΝΩΣΗ ΘΕΣΣΑΛΟΝΙΚΗΣ

Η στρατηγική του Yahoo την οποία και αναλύσαμε διεξοδικά σε προηγούμενες ενότητες φαίνεται και μέσα από τα παραπάνω εικονίδια, τα οποία και αντιγράψαμε από το site της εταιρίας. Συγκεκριμένα αρχικά βλέπουμε ότι ακολουθεί μια στρατηγική παροχών και επικοινωνίας η οποία δίνει τη δυνατότητα στο χρήστη με μια απλή εγγραφή να έχει δικό του email, δικό του ημερολόγιο γενικά δικό του χώρο δράσης.

Η στρατηγική αυτή βοηθά τη yahoo να προωθεί τις υπηρεσίες της και κυρίως τους καταλόγους της. Η δυναμική εύρεσης μέσα από τους καταλόγους, είναι αποτελεσματική αφού βοηθά τους χρήστες να εστιάσουν την εύρεση τους σε συγκεκριμένα πεδία, βάση συγκεκριμένων καταλόγων εύρεσης, τους οποίους βλέπουμε στο δεύτερο εικονίδιο. Όταν όμως η εύρεση είναι εκτός καταλόγων και αναφέρεται πάλι στην αναζήτηση αόριστων εννοιών όπως είναι η παγκοσμιοποίηση, η διαδικασία είναι παρόμοια με τη Google, χιλιάδες σελίδες, και στις πρώτες θέσεις υποκειμενικές θέσεις εφημερίδων, κομμάτων και γενικότερα φορέων που δε μπορούν να δώσουν τη σωστή πληροφόρηση στο χρήστη.

Το φαινόμενο της υπερπληροφόρησης, αντιμετωπίζεται καλύτερα από τη yahoo, μια και σ' επίπεδο εμπορικών και αγοραστικών αναγκών δίνει τη δυνατότητα καλύτερης εύρεσης μέσα από τους καταλόγους της, από την άλλη στην εύρεση γενικών εννοιών οι κατάλογοι δε μπορούν να βοηθήσουν, οπότε και πάλι ο χρήστης πελαγοδρομεί ανάμεσα σε χιλιάδες και πολλές φορές ανούσιες πληροφορίες.

5.2 TO GOOGLE

Μία ακόμη δημοφιλής μέθοδος αναζήτησης πληροφοριών στο Internet είναι η χρήση των search engines οι οποίες παλαιότερα αποκαλούνταν και Web crawlers ή spiders. Οπού όταν λεμέ μηχανές αναζήτησης είναι οι Lycos, Excite και AltaVista Και φυσικά εννοούμε την πιο διάσημη μηχανή αναζήτησης του internet το google. Τον Απρίλιο του 2006 οι αναζητήσεις στο google έφτασαν στο 50% οι λόγοι που έγινε τόσο επιτυχημένο και ευρέως διαδεδομένο το google είναι ότι

- Το περιβάλλον του internet είναι σαφές και απλό.
- Οι σελίδες φορτώνονται στιγμιαία.
- Η θέση ανάμεσα στα αποτελέσματα αναζητήσεις δεν πωλείτε.
- Η διαφήμιση πρέπει να είναι σχετική με το περιεχόμενο.

Το χαρακτηριστικό που έχει κάνει τόσο δημοφιλές το google είναι η αποτελεσματικότητα που εμφανίζει στη δημιουργία του ευρετηρίου και η ταχύτητα και οι ακρίβεια των αποτελεσμάτων του, όσοι έχουν συνηθίσει να χρησιμοποιούν το google δεν θα ανεχόταν να περιμένουν περισσότερο από 2 sec για να εμφανίσει το αποτέλεσμα μιας αναζητήσεις.

Το περιβάλλον του google που έχει ρυθμιστεί στο πρόγραμμα περιηγήσεις του χρήστη και στην έκδοση για κάθε γεωγραφική περιοχή οπού έχει πραγματοποιηθεί η σύνδεση άρα στην περίπτωση μας θα εμφανιστεί η ελληνική έκδοση του google με περιβάλλον στα ελληνικά, βεβαία μπορούμε να κάνουμε κάποιες ρυθμίσεις για να αλλάξουμε την γλώσσα αν επιθυμούμε και να εμφανίζει τα αποτελέσματα σε όλες τις γλώσσες. Οι αναζητήσεις γίνονται με 5 κύριους τρόπους παγκόσμιο ιστό, εικόνες, ομάδες, κατάλογο και ειδήσεις.

Οι περισσότεροι χρήστες κάνουν αναζήτηση στον παγκόσμιο ιστό, το γεγονός είναι ότι οι μηχανές αναζήτησης προβάλλουν μια πληθώρα αφιλτράριστων πληροφοριών πρέπει να

γνωρίζει τα εργαλεία και τις τεχνικές που σου προσφέρει το google για να γίνει πιο επιτυχής η αναζήτηση. Τα εργαλεία αυτά λέγονται τελεστές και είναι

- Τα εισαγωγικά « » με τα εισαγωγικά δείχνουμε ότι αναζητούμε μια συγκεκριμένη φράση με την συγκεκριμένη σειρά όπως την έχουμε γράψει. Π.χ. επιτροπή "αγροτικής πολιτικής" ότι αναζήτηση μας προβάλλει πρέπει η αγροτική πολιτική να είναι μαζί.

- Ένα ακόμη τρόπος είναι με το **MEION** (-) όπου η αναζήτηση γίνεται χωρίς να περιλαμβάνει την συγκεκριμένη λέξη αλλά την ευρύτερη έννοια. Αν βάλουμε την λέξη θάλασσα τότε έχουμε 2.550.000 αποτελέσματα αλλά αν βάλουμε θάλασσα - Ελλάδα έχουμε 77

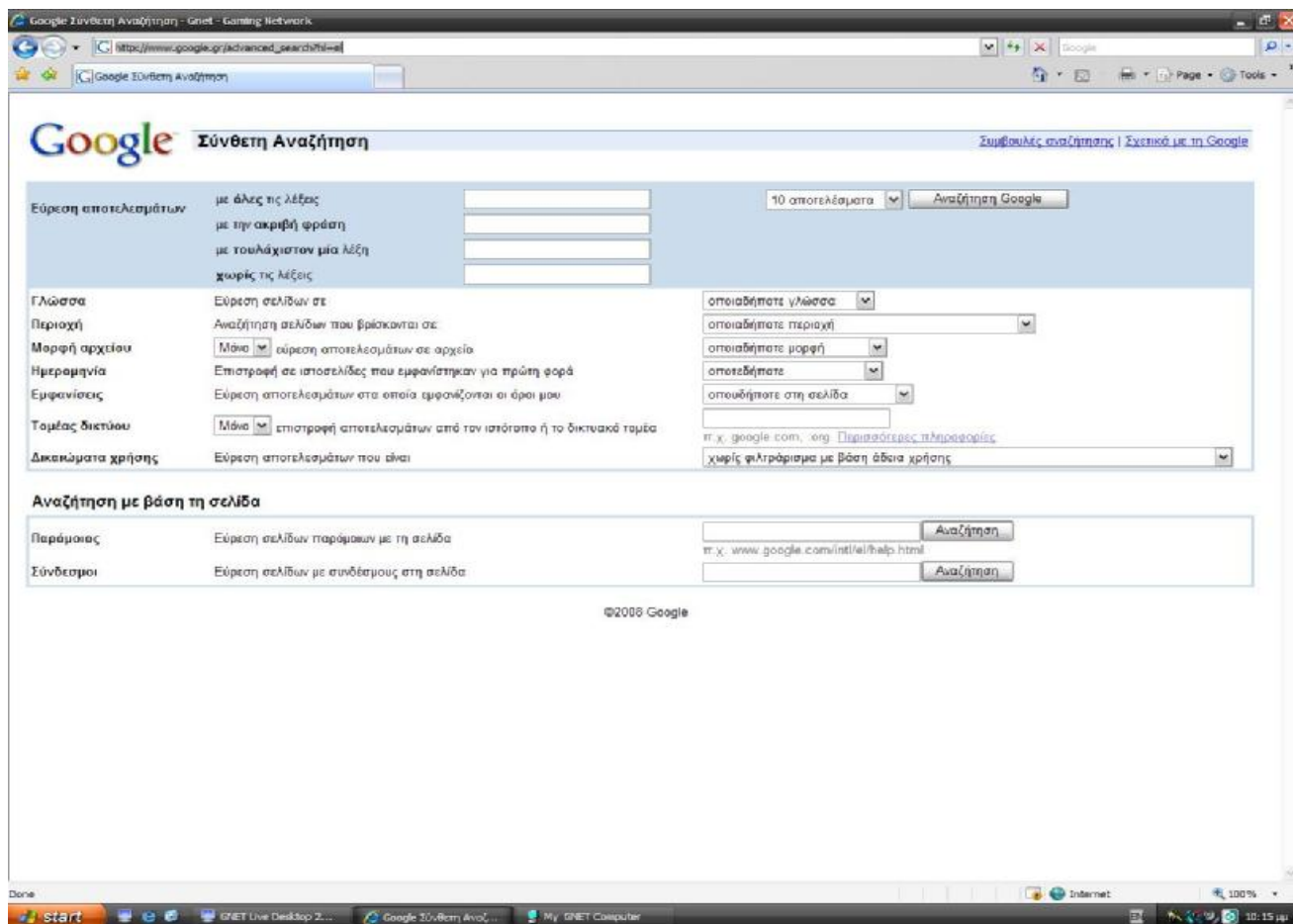
- Ο (*) μας δίνει την δυνατότητα να βάλουμε ανάμεσα σε δυο λέξεις το αστεράκι και το αποτέλεσμα που μας δίνει στην αρχή θα έχει την πρώτη λέξη και μετά από μια η περισσότερες λέξεις την δεύτερη. πχ αν γράψουμε μάζεμα * ήλιων μας δίνει τα εξής αποτελέσματα 1) μάζεμα των ελιών 2) μάζεμα εργαλεία για της ελιές και πολλά άλλα.

- Για διάστημα αριθμών 2000...2008 π.χ. ποσά παιδιά γεννήθηκαν από 2000...2008

- Αν θέλουμε να δούμε τι σημαίνουν τα αρχικά NOK γραφούμε **define.. NOK** και βγαίνει ναυτικός όμιλος Καστοριάς

- Και τέλος αν βάλουμε μέσα στην αναζήτηση αριθμητικές πράξεις βγαίνει το αποτέλεσμα. Πχ $4*5+4$ μας δίνει $(4 * 5) + 4 = 24$.

Εικόνα 3.1 Σύνθετη αναζήτηση στο Google



Έκτος από τους τελεστές υπάρχει και η σύνθετη αναζήτηση πατώντας αυτό το κουμπί ανοίγει ένα παραθύρι όπου μας δίνετε η δυνατότητα για αναζήτηση σε ποια γλώσσα του έγγραφου , μορφή του αρχείου, ο τομέας δικτύου, που ανήκει. Στην σύνθετη αναζήτηση δεν υπάρχουν τελεστές υπάρχει η επιλογή χωρίς τις λέξεις επίσης μπορούμε να ελέγξουμε την θέση των λέξεων στο εσωτερικό του έγγραφου η της σελίδας που είναι ή που ανοίγουμε.

Οι μηχανές αναζήτησης ανανεώνονται συνέχεια ώστε να παρουσιάζουν τις πιο ενημερωμένες πληροφορίες και περιέχουν τεράστιες ποσότητες δεδομένων. Οι μηχανές αναζήτησης εξάγουν και καταλογοποιούν τις πληροφορίες με διαφορετικό τρόπο μεταξύ τους. Ορισμένες τοποθετούν σε κατάλογο κάθε λέξη που συναντούν σ' ένα έγγραφο, ενώ άλλες για παράδειγμα βάζουν στον κατάλογο μόνο 100 λέξεις-κλειδιά του εγγράφου. Άλλες

τοποθετούν σε κατάλογο το μέγεθος του εγγράφου, τον τίτλο, τον υπότιτλο κ.ο.κ. Επιπροσθέτως κάθε μηχανή αναζήτησης παρουσιάζει τα αποτελέσματα με διαφορετικό τρόπο. Ορισμένες "ζυγίζουν" τα αποτελέσματα για να δείξουν την συνάφεια των εγγράφων, άλλες δείχνουν τις πρώτες προτάσεις του εγγράφου, ενώ άλλες εμφανίζουν τον τίτλο του εγγράφου καθώς και το URL.

Στο Internet υπάρχουν αρκετές μηχανές αναζήτησης και ευρετήρια, κάθε ένα εκ των οποίων έχει ορισμένα πλεονεκτήματα και μειονεκτήματα. Όταν αναζητά ο χρήστης μία πληροφορία και για να καλύψει το μεγαλύτερο δυνατό τμήμα του Δικτύου θα πρέπει να χρησιμοποιήσει αρκετές μηχανές αναζήτησης. Όμως η διαδικασία είναι αρκετά χρονοβόρα και για το λόγο αυτό έχει αναπτυχθεί ειδικό λογισμικό, ονόματι meta-search, το οποίο αυτοματοποιεί την αναζήτηση.

Συγκεκριμένα αν έχει ο χρήστης εγκαταστήσει ένα τέτοιο πρόγραμμα στον υπολογιστή και δώσει τα στοιχεία της αναζήτησης, το πρόγραμμα στέλνει αυτόματα την αναζήτηση σε αρκετές μηχανές αναζήτησης και ευρετήρια ταυτόχρονα, επεξεργάζεται τα αποτελέσματα και τα παρουσιάζει στον υπολογιστή.

Οι παραπάνω περιγραφές μας παρουσιάζουν περιεκτικά τη σημερινή λειτουργία των βασικών ιστοσελίδων, αναπτύσσοντας το σύστημα βάση του οποίου βοηθούν το χρήστη να κάνει εύρεση. Τα συστήματα αυτά παρουσιάζουν αδυναμίες, διότι αντί να βοηθούν το χρήστη και να τον οδηγούν σ' αυτό που ψάχνει, τον βομβαρδίζουν με πολλές και ανούσιες σε πολλές περιπτώσεις πληροφορίες, οι οποίες δε του χρειάζονται, οπότε και τον αποπροσανατολίζουν.

Αφού ο ενδιαφερόμενος συσσωρεύσει ένα μεγάλο όγκο πληροφοριών και διαβάσει έναν αριθμό κειμένων, είναι πιθανό να κρατήσει ορισμένες δικτυακές διευθύνσεις στα «αγαπημένα του». Αυτό θα τον βοηθήσει να παρακολουθεί αυτές τις ιστοσελίδες και στο

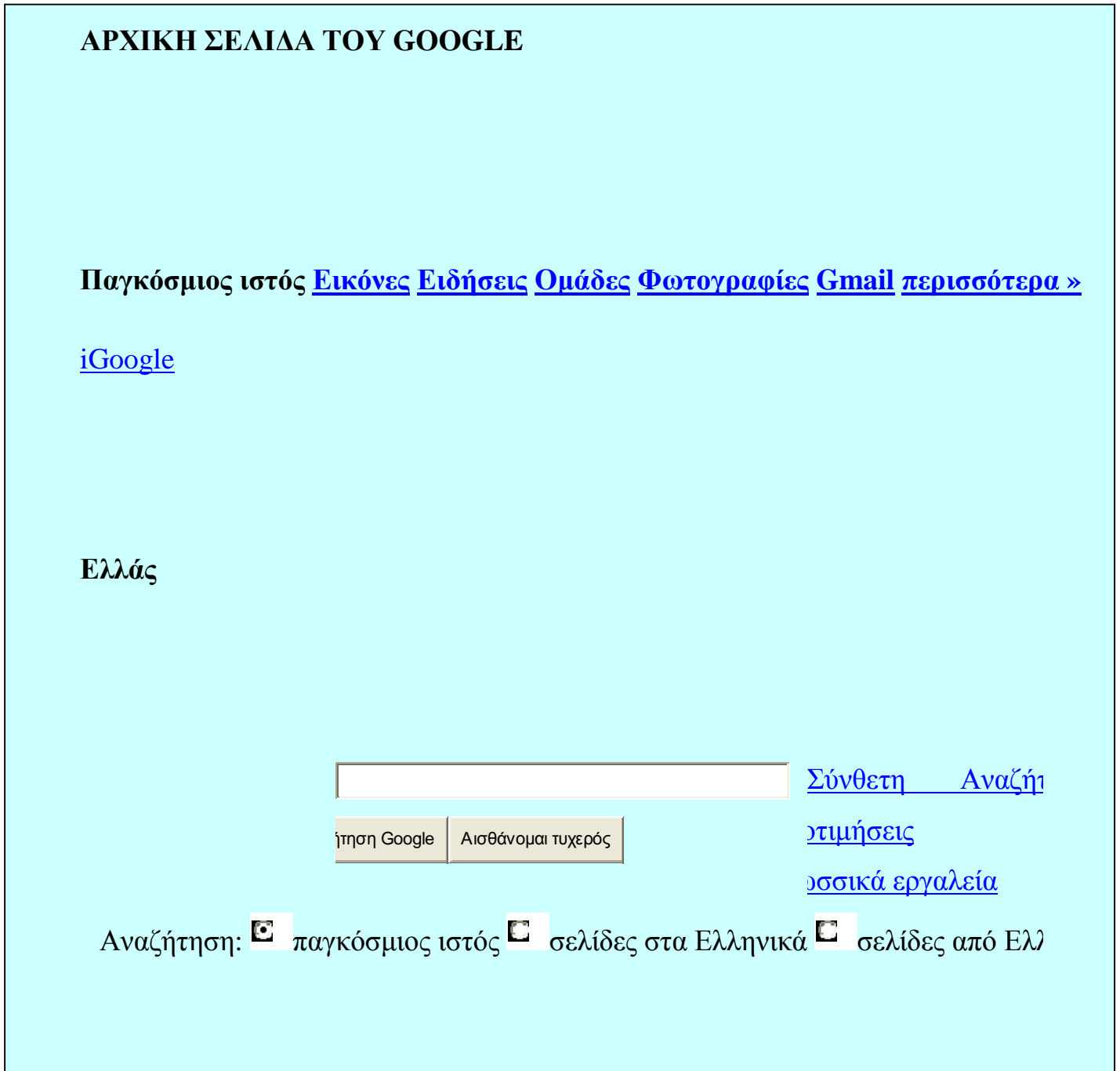
μέλλον, το πρώτο πρόβλημα εδώ αναφέρεται στην υποκειμενικότητα των ιστοσελίδων σε σχέση με τις απόψεις που εκφέρουν. Το δεύτερο πρόβλημα αναφέρεται στην αύξηση των ιστοσελίδων που ελλοχεύει το κίνδυνο στο εγγύς μέλλον, ο χρήστης να μη μπορεί πια να τις παρακολουθεί χάνοντας έτσι σημαντικές εξελίξεις, αλλά και πληροφορίες που πιθανόν να του ήταν χρήσιμες. Το τρίτο και τελευταίο πρόβλημα αναφέρεται στη γλώσσα, μια και πολλές από τις ιστοσελίδες θα είναι στα αγγλικά, πράγμα που σημαίνει ότι κάποιοι εκ των χρηστών, δε θα μπορούν να τις διαβάσουν(Λέανδρος Ν,2005:94-96).

Με βάση τα δυο πονήματα καταλήγουμε ότι υπάρχει μεγάλος αριθμός δυσκολιών προερχόμενος από την υπερπληροφόρηση αλλά και μεγάλη αδυναμία των υπαρχουσών ιστοσελίδων να την αντιμετωπίσουν, δίνοντας καλύτερη και πιο ουσιαστική ενημέρωση στους χρήστες.

Παρακάτω θα προχωρήσουμε παρουσιάζοντας παραδείγματα από τα πέντε πιο δημοφιλή site προκειμένου να κάνουμε μια εμπειριστατωμένη και ουσιαστική επιβεβαίωση του προβλήματος της υπερπληροφόρησης, κάνοντας από τη μια, μια αξιολόγηση της ευρύτερης διαδικασίας εύρεσης που τα χαρακτηρίζει και από την άλλη χρησιμοποιώντας το παράδειγμα του Λέανδρου Ν προκειμένου να δείξουμε και στην πράξη το πώς ο χρήστης συγκεντρώνει χιλιάδες πληροφορίες οι οποίες αντί να τον βοηθούν τον μπερδεύουν και τον αποπροσανατολίζουν από το αρχικό αντικείμενο εύρεσης του.

5.2.1 Το Προφίλ του Google

Εικόνα 3.2 Το προφίλ Google



Με βάση το παραπάνω εικονίδιο το οποίο και εκμαιεύσαμε από την αρχική σελίδα του Google, μπορούμε να πούμε ότι η διαδικασία εύρεσης στη συγκεκριμένη ιστοσελίδα είναι αρκετά απλή. Ο χρήστης από τη πρώτη στιγμή που εισέρχεται στο περιβάλλον της Google,

του γίνεται κατανοητό ότι το Google, αποτελεί μια απλή στη χρήση της ιστοσελίδα η οποία εστιάζει καθαρά στη διαδικασία αναζήτησης, την οποία και τοποθετεί στην αρχική της σελίδα. Ο χρήστης μπορεί απευθείας να θέσει υπό εύρεση τη λέξη κλειδί που ψάχνει είτε ακολουθώντας την απλή διαδικασία, είτε αν έχει περισσότερα δεδομένα να προβεί σε σύνθετη αναζήτηση, τοποθετώντας στα ζητούμενα κελιά, περισσότερες λέξεις κλειδιά σε σχέση με την εύρεση που κάνει, όπως ημερομηνίες συγκεκριμένες, όνομα συγγραφέα κ.λ.π.

Η Google επενδύει στη βελτίωση της εύρεσης δεδομένων από τους χρήστες, προσπαθώντας να καταπολεμήσει το φαινόμενο της υπερπληροφόρησης, προάγοντας υπηρεσίες όπως το Google Desktop και το Google searcher. Η πρώτη υπηρεσία δίνει τη δυνατότητα στο χρήστη να έχει μόνιμα το εικονίδιο search της Google στο Desktop του ώστε να κάνει απευθείας αναζητήσεις από τον υπολογιστή του, η υπηρεσία αυτή προσφέρει στο χρήστη ταχύτητα εύρεσης και καλύτερη ποιότητα αποτελεσμάτων.

Το Google Searcher δίνει στον ερευνητή τη δυνατότητα εύρεσης στοιχείων σ' ένα βελτιωμένο περιβάλλον το οποίο περιλαμβάνει άρθρα σε μορφή acrobat αλλά και ολόκληρα βιβλία σε συνεργασία με διάφορους οργανισμούς όπως το Amazon. Η υπηρεσία αυτή δίνει μεγαλύτερη δυναμική στο χρήστη αφού δε διαβάζει τυπικές περιλήψεις ή μικρά κομμάτια από ένα ολόκληρο άρθρο, μην έχοντας ξεκάθαρη εικόνα από το αποτέλεσμα μιας αναζήτησης, αντίθετα έχει πρόσβαση σ' ολόκληρα κείμενα, τα οποία μπορεί να μελετήσει και να καταλήξει σ' ασφαλή συμπεράσματα σε σχέση με την έρευνα του.

Η Google παρότι έκανε αρκετές κινήσεις βελτίωσης στο τομέα της εύρεσης δεν παύει να έχει προβλήματα τα οποία αναφέρονται στη ποικιλότητα των πληροφοριών, οι οποίες τις περισσότερες φορές δε δίνουν ουσιαστική ενημέρωση στο χρήστη αλλά τον παραπληροφορούν και τον αποπροσανατολίζουν. Το παράδειγμα εύρεσης σε σχέση με τη παγκοσμιοποίηση και την οικολογία αποτελεί χαρακτηριστικό παράδειγμα, σε σχέση με το άνωθεν σκεπτικό μας. Παρακάτω αρχικά παραθέτουμε τα αποτελέσματα από την έρευνα.

Παγκόσμιος ιστός

Αποτελέσματα 1 - 10 από περίπου 90.200 για **Παγκοσμιοποίηση και οικολογία**.
(0,25 δευτερόλεπτα)

1) [/ Ελληνική Δημοκρατία - Υπουργείο Εξωτερικών \](#)

Η **παγκοσμιοποίηση** της **οικολογίας** βρίσκει την έκφρασή της μέσα από τις πολυπληθείς περιβαλλοντικές συνθήκες που συμφωνήθηκαν από τα μέσα του '80 και μετά ..

old.mfa.gr/greek/foreign_policy/environment/omilia.html - 35k

- [Προσωρινά αποθηκευμένη](#) - [Παρόμοιες σελίδες](#)

2) [Τα πολιτικά ρεύματα στην Οικολογία « Το Μανιτάρι του Βουνού](#)

Το μέγεθος και η **παγκοσμιοποίηση** της **οικολογικής** κρίσης, στην αρχή περιθωριακά και στη συνέχεια πιο έντονα, άρχισε από τα μέσα της δεκαετίας του '70 να ..

manitaritoubounou.wordpress.com/2007/11/18/Τα-πολιτικά-ρεύματα-στην-Οικολογία/ - 78k –

- [Προσωρινά αποθηκευμένη](#) - [Παρόμοιες σελίδες](#)

3) [Άρθρα σε άλλα περιοδικά και εφημερίδες του Τάκη Φωτόπουλου](#)

Η ανάγκη για ένα ριζοσπαστικό κίνημα κατά της **παγκοσμιοποίησης**(Ουτοπία, Εξαρτημένη ανάπτυξη και περιβάλλον (**Οικολογία** και περιβάλλον, www.inclusivedemocracy.org/fotopoulos/greek/grvarious/grvar.htm - 51k - [Προσωρινά αποθηκευμένη](#) - [Παρόμοιες σελίδες](#)

4) [Παγκοσμιοποίηση. Είστε υπέρ ή κατά; \[Αρχείο\] - Ελεύθερη σκέψη ...](#)

Από την άλλη όμως έχουμε την **παγκοσμιοποίηση** των ανθρωπίνων δικαιωμάτων και των δημοκρατικών αξιών, την **παγκοσμιοποίηση της οικολογίας** και την ανησυχία για ..

www.antidogma.gr/vb/archive/index.php?t-2477.html - 33k –

[Προσωρινά αποθηκευμένη](#) - [Παρόμοιες σελίδες](#)

5) [ΕΠΙΣΤΗΜΗ ΚΑΙ ΚΟΙΝΩΝΙΑ / Μεταπτυχιακό Σεμινάριο Πολιτικής ...](#)

Παγκοσμιοποίηση και οικολογία • Οι μετασχηματισμοί του διοικητικού κράτους
• Πολυεπίπεδη διακυβέρνηση και εθνικό κράτος .Οργανωμένο έγκλημα και σύγχρονη ..

www.media.uoa.gr/sas/pages/seminar.html - 12k -

[Προσωρινά αποθηκευμένη](#) - [Παρόμοιες σελίδες](#)

6) [Βιβλιοπαρουσίαση Ευώνυμος Οικολογική Βιβλιοθήκη](#)

Τα προβλήματα τα οποία παραπέμπουν στο φαινόμενα της παραπληροφόρησης, τα οποία και εντοπίζουμε από την εύρεση που κάναμε στη Google είναι τα ακόλουθα:

1. Η εύρεση μας έδωσε σχετικές 90.200 ιστοσελίδες. Σε καμία περίπτωση ένας χρήστης δε θα κοιτάξει στο σύνολο των ιστοσελίδων. Αυτό σημαίνει ότι δε θα έχει τη δυνατότητα να έχει πρόσβαση σ' όλη την πληροφόρηση αλλά μόνο σ' ένα μέρος αυτής, που μπορεί και να μην τον καλύψει, η μπορεί να μη του δώσει το περιθώριο, να μελετήσει άρθρα ή κείμενα από τα οποία θα μπορούσε να αποκομίσει από μια σχετική πρόταση μέχρι και ένα ολοκληρωμένο κείμενο. Το Google βάζει τις σελίδες σε σειρά προτεραιότητας, όμως στη πραγματικότητα το κάνει αυτό με δικά του υποκειμενικά κριτήρια χωρίς να δίνει στο χρήστη τη δυνατότητα να επιλέξει ο ίδιος το τι θέλει και τι όχι να μελετήσει. Η προσπάθεια για πληρότητα στη πληροφόρηση έχει τα αντίθετα αποτελέσματα, δίνοντας στο χρήστη μια πληροφόρηση άτοπη και όχι ουσιαστική, η οποία σε καμία περίπτωση δε μπορεί να τον βοηθήσει χρησιμοποιώντας την προς δικό του συμφέρον.

2. Η εύρεση μας έφερε στη πρώτη σελίδα τους ακόλουθους φορείς πληροφόρησης:

α Το Υπουργείο Εξωτερικών.

α Απόψεις Οικολόγων.

α Το site του Συνασπισμού.

α Απόψεις από εφημερίδες όπως τον Ριζοσπάστη.

Καμία από τις παραπάνω πηγές δε μπορεί να θεωρηθεί ως έγκυρη. Συγκεκριμένα το κόμμα του Συνασπισμού και οι εφημερίδες θα δίνει υποκειμενικές απόψεις, οι οποίες θα εκφράζονται υπό το πρίσμα των κομματικών τους συμφερόντων και απόψεων, το ίδιο ισχύει και για τις απόψεις των οικολογικών οργανώσεων. Η πιο έγκυρη πηγή θεωρείτε το Υπουργείο Εξωτερικών η οποία βγαίνει και πρώτη στη σειρά προτεραιότητας, μια και εκεί ο ενδιαφερόμενος θα μπορέσει να βρει στοιχεία σχετικά με την ελληνική και Ευρωπαϊκή Νομοθεσία. Με βάση τα παραπάνω κατανοούμε ότι ο χρήστης από τις πρώτες σελίδες που του δίνει η Google προς εύρεση, μόνο η πρώτη θα μπορούσε να αξιωθεί ως αξιόπιστη πηγή, ενώ στη συνέχεια θα πρέπει να συνεχίσει το ψάξιμο στις υπόλοιπες σελίδες, που θεωρητικά είναι δευτερεύοντος σημασίας και σπουδαιότητας. Παρακάτω σας παρουσιάζουμε δυο μηχανές αναζήτησης και με ποιον τρόπο λειτουργούν.

5.3 TO MSN

Εικόνα 4. Το προφίλ του MSN

MSN SEARCH

- [Español](#)
- [RSS](#)
- [Web](#)
- [MSN](#)
- [Images](#)
- [News](#)
- [Maps](#)
- [Shopping](#)
- [Video](#)
- Αρχή φόρμας

Search the Web

Search:

Search Web

Popular Searches:

- [Brunch recipes](#)
- [USC vs. Illinois](#)
- [Baby New Year](#)

Live Search	Αργική σελίδα	
	Hotmail	Είσοδος
	Spaces	
	MSN	

ΠΑΓΚΟΣΜΙΟΠΟΙΗΣΗ

Αναζήτηση



Μόνο [Ελληνικά](#)



Μόνο από Ελλάδα

- [Σύνθετη](#)
- [Επιλογές](#)

Αποτελέσματα Web 1-10 από 98

Δείτε επίσης:

-

ο Περισσότερα ▼

ο

§ [Ακαδημαϊκά θέματα](#)Beta

§ [Τροφοδοσίες](#)Beta

§ [Spaces](#)

§ [Επεξεργασία μακροεντολών](#)

§ [Εύρεση μακροεντολών](#)

[Άρθρα σε άλλα περιοδικά και ...](#)

... Σεπτέμβριος 2005) [κατηγορία: **ΠΑΓΚΟΣΜΙΟΠΟΙΗΣΗ** ... 26/11/1995)
[κατηγορία: **ΟΙΚΟΛΟΓΙΑ ΚΑΙ ...**

ο www.democracynature.org/fotopoulos/greek/grvarious/grvar.htm

ο • [Αποθηκευμένη σελίδα](#)

• [Various texts by Takis Fotopoulos](#)

... Σεπτέμβριος 2005) [κατηγορία : **ΠΑΓΚΟΣΜΙΟΠΟΙΗΣΗ** ... 26/11/1995)
[κατηγορία: **ΟΙΚΟΛΟΓΙΑ ΚΑΙ ...**

ο www.inclusivedemocracy.org/fotopoulos/grvar.htm

ο • [Αποθηκευμένη σελίδα](#)

Το MSN Search λειτουργεί όπως το AOL ουσιαστικά δίνει καταλόγους εύρεσης όπως το yahoo, ενώ συγχρόνως παρέχει επικοινωνιακά μέσα όπως free email αλλά και παροχές οργάνωσης, στους χρήστες όπως ημερολόγιο, agenda κλπ. Πέρα από τα παραπάνω χρησιμοποιεί και μια συγκεκριμένη μηχανή αναζήτησης σε συνεργασία με τη google η οποία δίνει λιγότερα αποτελέσματα, τα οποία προέρχονται και πάλι σε μια υποκειμενική

επιλογή ανάμεσα στα αποτελέσματα της Google. Η επιλογή αυτή γίνεται με τη βοήθεια της μηχανής αναζήτησης που χρησιμοποιεί η MSN. Η υποκειμενική επιλογή μπορεί και πάλι να παραπληροφορήσει αντί να ενημερώσει το κοινό, μη δίνοντας του τη κατάλληλη ενημέρωση.

5.4 TO AOL

Εικόνα 5. Το προφίλ του AOL

- [AOL](#)
- | [My AOL](#)
- | [Mail](#)
- | [Get AOL Toolbar](#)
- | [Make AOL Your Home Page](#)

- [Search History](#) |
- [Sign In](#)

[Get Help X](#)

<a

href="["
target="_top">Sign in](http://my.screenname.aol.com/_cqr/login/login.psp?sitedomain=search.aol.com&authLev=1&seamless=y&-lang=en&locale=us&siteState=OrigUrl%3Dhttp%253A%252F%252Fsearch.aol.com%252F%252Fwebhome&loginId=)

AOL Search Logo

Enhanced by Google

- Web

- [Shopping](#)
- [Images](#)
- [Video](#)
- [News](#)
- [Local](#)
- [more »](#)

Η συγκεκριμένη βάση συνδυάζει στοιχεία της yahoo και στοιχεία της Google. Από τη μια παρέχει δυνατότητες επικοινωνίας στους χρήστες του, όπως παρέχει και η yahoo, ενώ από την άλλη λειτουργεί ως site εύρεσης όπως η google, η δυναμική του και στους δυο τομείς είναι μικρότερης εμβέλειας.

5.5 META ΜΗΧΑΝΕΣ ΑΝΑΖΗΤΗΣΗΣ & ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ


Οι μετά-μηχανές αναζήτησης είναι μηχανές αναζήτησης οι οποίες χρησιμοποιούν άλλες μηχανές αναζήτησης, προκειμένου να βρουν πληροφορίες στην ερώτησή του χρήστη. Η ίδια η μεταμηχανή δεν διαθέτει δική της βάση δεδομένων ούτε Web Robot. Βασίζεται αποκλειστικά στις μηχανές αναζήτησης με τις οποίες συνεργάζεται. Γενικά, πλεονεκτεί έναντι μιας απλής μηχανής στο ότι χρησιμοποιεί τα καλύτερα αποτελέσματα από πολλές μηχανές και δεν στηρίζεται μόνο σε μία (Εργαστήριο Εφαρμογών Πληροφορικής στα ΜΜΕ, 2007).

Η λειτουργία τους αναπτύσσεται ως εξής. Οι META μηχανές αναζήτησης, στέλνουν την ερώτηση σε περισσότερες από μια μηχανές αναζήτησης, μερικές φορές μετά από επιλογή του χρήστη. Τα αποτελέσματα συγχωνεύονται και παρουσιάζονται σε μια λίστα,

αφού αφαιρεθούν τα διπλά, ή παρουσιάζονται σαν ξεχωριστές λίστες ανά μηχανή αναζήτησης. Παραδείγματα μεταμηχανών παρουσιάζουμε στο παρακάτω πίνακα(Βιβλιοθήκη Κολεγίου Αθηνών, 2007):

Πίνακας 3. Μετά-Μηχανές Αναζήτησης

ΜΕΤΑΜΗΧΑΝΗ	ΔΡΑΣΤΗΡΙΟΤΗΤΑ
	<p>Μεταμηχανή αναζήτησης με παράλληλη αναζήτηση στις κυριότερες μηχανές. Διαθέτει πολλούς τρόπους αναζήτησης μεταξύ των οποίων με Boolean τελεστές. Διαθέτει Advanced Search.</p>
	<p>Χρησιμοποιεί πολλές μηχανές αναζήτησης όπως Google, Yahoo, Ask Jeeves, About, FindWhat, LookSmart, κλπ. Διαθέτει Advanced Search.</p>
	<p>Το Mamma αναζητά στις μεγαλύτερες μηχανές αναζήτησης. Τα αποτελέσματα τα παρουσιάζει σε μια ενιαία μορφή κατά πηγή και ποσοστό συνάφειας με την ερώτηση.</p>
	<p>Μεταμηχανή που αναζητά μέσω του Google, Yahoo!, Ask Jeeves, About, Looksmart και άλλων μηχανών αναζήτησης. Μπορείτε να ψάξετε χωριστά για εικόνες, ήχο, βίντεο και ειδήσεις.</p>

	<p>Γρήγορη μεταμηχανή με δυνατότητα απλής και σύνθετης αναζήτησης. Μεταξύ των διαφόρων επιλογών που έχει μπορεί κανείς να δει, χωρίς να φύγει από τη σελίδα των αποτελεσμάτων, μέρος της ιστοσελίδας κάθε συνδέσμου.</p>
---	--

Πηγή: Βιβλιοθήκη Κολεγίου Αθηνών(2007),Μετά Μηχανές Αναζήτησης, <http://www.haef.gr/gre/libraries/websearch/index.html> (Ανακτήθηκε 11-2-08)

Παρακάτω αναλύουμε κάποιες από τις ποιες γνωστές μεταμηχανές, δίνοντας διάφορα στοιχεία για την ιστορία τους, τη λειτουργία τους αλλά και για τη χρήση τους(Εργαστήριο Εφαρμογών Πληροφορικής στα ΜΜΕ, 2007).

MetaCrawler: Η MetaCrawler είναι πρώτη μεταμηχανή αναζήτησης. Ξεκίνησε την λειτουργία της τον Ιούλιο του 1995 ως πτυχιακή εργασία ενός φοιτητή του Πανεπιστημίου της Ουάσινγκτον. Έγινε γρήγορα δημοφιλής στους χρήστες του διαδικτύου μόλις άνοιξε τις πύλες της στο κοινό. Τον Φεβρουάριο του 1997 εξαγοράστηκε από την εταιρεία παροχής υπηρεσιών Internet Go2Net. Η MetaCrawler ανακτά τα αποτελέσματά της από τις υπηρεσίες αναζήτησης AltaVista, Infoseek, WebCrawler, Excite, Lycos, Google, GoTo, Direct Hit, Look Smart, Real Names, About.com και Thunderstone. Η MetaCrawler δεν δέχεται την χρήση λογικών συμβόλων (+, -) ούτε την λογική των τελεστών. Διαθέτει ένα περιβάλλον για σύνθετες αναζητήσεις. Παρ' όλα αυτά, είναι μία από τις καλύτερες μεταμηχανές αναζήτησης στο διαδίκτυο.

Highway61: Η Highway61 είναι μία από τις παλαιότερες και πιο γνωστές μεταμηχανές αναζήτησης στο διαδίκτυο. Ξεκίνησε το 1996 και ανακτά τα αποτελέσματά της από τις υπηρεσίες αναζήτησης Yahoo!, Excite, Lycos, Infoseek και WebCrawler. Το κενό ανάμεσα

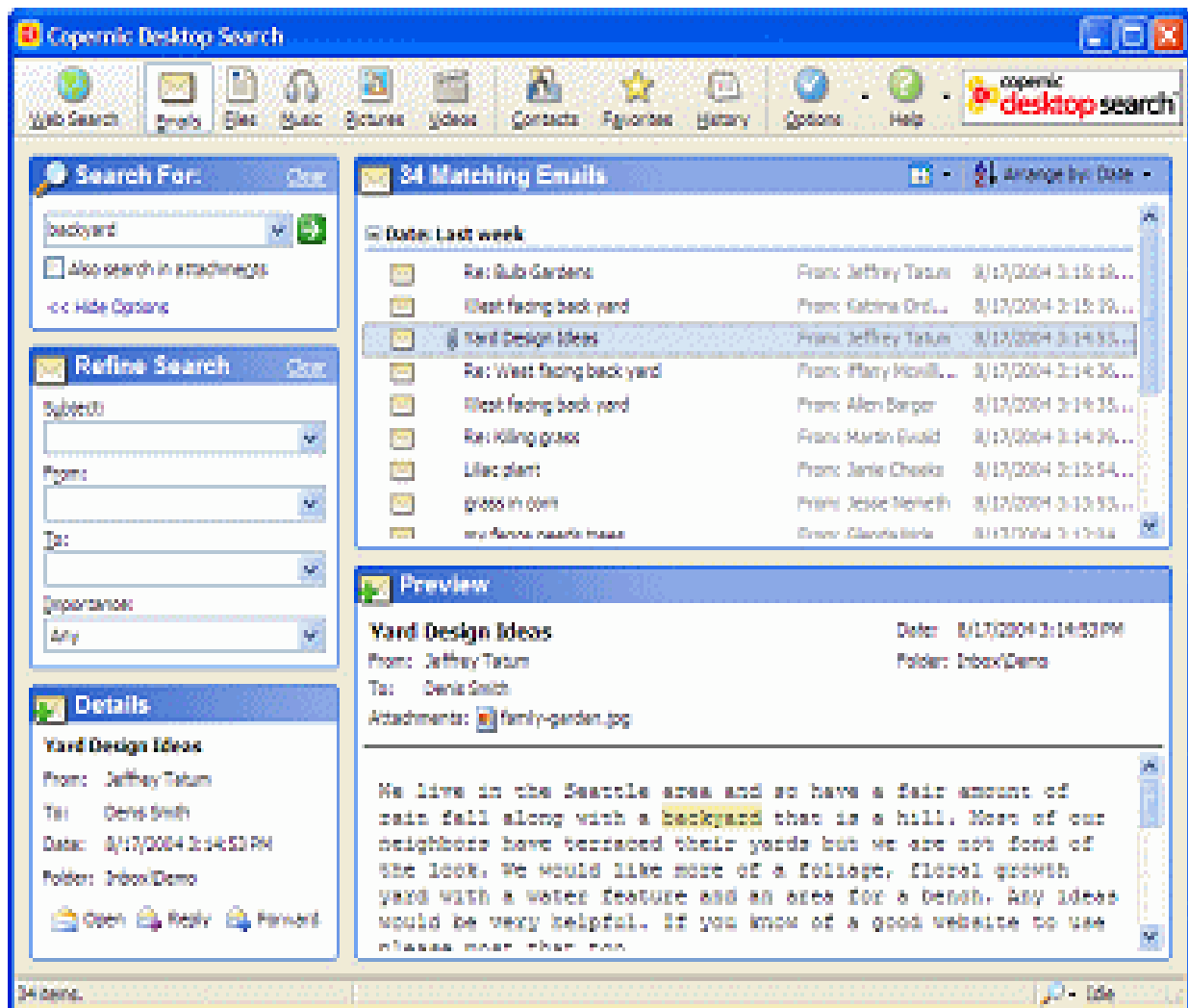
στις λέξεις-κλειδιά το αντιλαμβάνεται ως τον τελεστή or, δέχεται όμως και την χρήση του τελεστή and. Έχει σίγουρα το πιο διασκεδαστικό περιβάλλον χάρη στο χιούμορ της ιστοσελίδας της, δεν είναι όμως ότι καλύτερο θα περίμενε κανείς από μία μεταμηχανή αναζήτησης.

5.5.1 Το Πρόγραμμα Κοπέρνικος

Η Copernic είναι μία μεταμηχανή αναζήτησης, την οποία ο χρήστης πρέπει να κατεβάσει από το Internet και να την εγκαταστήσει στον υπολογιστή του. Λειτουργεί μόνο με την εφαρμογή πλοήγησης Internet Explorer (όχι Netscape Navigator). Η βασική εκδοχή της Copernic είναι δωρεάν και ανακτά τα αποτελέσματά της από τις υπηρεσίες αναζήτησης AltaVista, Excite, HotBot, Infoseek, Lycos, WebCrawler και Yahoo!. Δέχεται την λογική των τελεστών στην διερεύνηση των αποτελεσμάτων της πρώτης αναζήτησης. Μπορεί να μην είναι γνωστή, είναι όμως μία από τις καλύτερες μεταμηχανές αναζήτησης και ο χρήστης μπορεί να την κατεβάσει και να την εγκαταστήσει εύκολα στον υπολογιστή του (Εργαστήριο Εφαρμογών Πληροφορικής στα ΜΜΕ, 2007).

Η ανανεωμένη έκδοση του δημοφιλούς προγράμματος, επιτρέπει ο χρήστης να κάνει αναζήτηση στους τοπικούς σκληρούς δίσκους. Το πρόγραμμα δίνει τη δυνατότητα αναζήτησης αρχείων, μηνυμάτων ηλεκτρονικού ταχυδρομείου, εικόνες, αρχεία πολυμέσων και πολλά άλλα (Γιαννακίδης Α, 2007).

Εικόνα 6. Το πρόγραμμα Κοπέρνικος



Το πρόγραμμα δημιουργεί βάση δεδομένων στο δίσκο, με την οποία αργότερα μπορεί να δουλέψει ο χρήστης πολύ άνετα. Η αναζήτηση μπορεί να γίνει με λέξεις-κλειδιά, και για αυτό δεν είναι απαραίτητο να εκτελεστεί το πρόγραμμα. Αυτό εγγράφεται στην αυτόματη εκκίνηση και για αυτό είναι πάντα διαθέσιμο. Το Copernic Desktop Search μπορεί επίσης να εισάγει στη βάση του το ιστορικό του browser, τους συνδέσμους του φακέλου "Αγαπημένα", καθώς και τις επαφές του βιβλίου διευθύνσεων.

5.5.2 Λειτουργία Βάσεων Δεδομένων

Στις περισσότερες Βάσεις η αναζήτηση μπορεί να γίνει με τη χρήση ΛΕΞΕΩΝ-ΚΛΕΙΔΙΑ. Ο λόγος που χρησιμοποιούμε λέξεις και όχι μια φράση που περιγράφει το θέμα μας είναι ότι πρέπει να «μεταφράσουμε» το θέμα σε γλώσσα που να καταλαβαίνει η βάση. Πριν την αναζήτηση ορίζουμε τις λέξεις κλειδιά που θα χρησιμοποιήσουμε, αναζητώντας όλα τα συνώνυμα και τις σχετικές με το θέμα μας λέξεις. Στις περισσότερες Βάσεις επιτρέπεται επίσης η αναζήτηση με το όνομα συγγραφέα, τον τίτλο βιβλίου ή /και άρθρου ή με λέξεις που περιέχονται σε τίτλο/ περίληψη/ κείμενο.

ΔΙΑΧΕΙΡΙΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

Αν τα αποτελέσματα της αναζήτησης δεν είναι ικανοποιητικά ως προς το πλήθος των αποτελεσμάτων ή τη σχετικότητα των αναφορών, μπορούμε να τροποποιήσουμε την αναζήτησή μας. Οι περισσότερες Βάσεις μας δίνουν τη δυνατότητα να περιορίσουμε τα αποτελέσματα ως προς τη χρονολογία έκδοσης των αναφορών, τον τύπο του υλικού και τη γλώσσα του υλικού.

Στις περισσότερες Βάσεις δίνεται η δυνατότητα επιλογής συγκεκριμένων αναφορών μέσα από τα αποτελέσματα, των οποίων μπορούμε να ορίσουμε τη μορφή με την οποία θα εμφανιστούν (απλή αναφορά της δημοσίευσης, πλήρη αναφορά, εμφάνιση της περίληψης της δημοσίευσης). Δίνεται επίσης η δυνατότητα, τόσο για όλα τα αποτελέσματα αναζήτησης όσο και για τα επιλεγμένα να αποθηκευτούν, να τυπωθούν ή και να σταλούν με το ηλεκτρονικό ταχυδρομείο (Presno C:2000)

ΒΟΗΘΕΙΑ

Σε όλες τις Βάσεις Δεδομένων υπάρχει διαθέσιμη «βοήθεια» (HELP) η οποία παρέχει πληροφορίες σχετικές με τους τρόπους αναζήτησης στη Βάση, τους λογικούς τελεστές που

μπορούν να χρησιμοποιηθούν και τη διαχείριση των αποτελεσμάτων της αναζήτησης.

ΣΤΡΑΤΗΓΙΚΗ ΑΝΑΖΗΤΗΣΗΣ

1. Επιλέγουμε την κατάλληλη Βάση Δεδομένων
2. Ορίζουμε το θέμα και βρίσκουμε τις κατάλληλες λέξεις-κλειδιά Ορισμένες Βάσεις δίνουν τη δυνατότητα χρήσης «Θησαυρού»(Thesaurus) για αναζήτηση λέξεων σχετικών με ένα θέμα.
3. Αποφασίζουμε πώς θα συνδυάσουμε τους όρους αναζήτησης χρησιμοποιώντας τους λογικούς τελεστές.
4. Χρησιμοποιούμε, όπου είναι απαραίτητο, τη δυνατότητα αποκοπής για αναζήτηση με ρίζα λέξεων.
5. Εξετάζουμε τα αποτελέσματα και τροποποιούμε την αναζήτηση, αν χρειάζεται.
6. Επιλέγουμε, αποθηκεύουμε, τυπώνουμε ή στέλνουμε τα αποτελέσματα.
7. Κρατάμε σημειώσεις και τα αποτελέσματα αναζητήσεων για μελλοντική χρήση και αποφυγή επαναληπτικής αναζήτησης.
8. Για την πληρέστερη κάλυψη του θέματός μας μπορούμε να επαναλάβουμε την ίδια διαδικασία σε περισσότερες από μία Βάσεις.

Είναι αλήθεια άλλωστε πως σε συγκεκριμένες διδακτικές διαδικασίες όπου χρειάζεται να προστρέξει κανείς σε βιβλιογραφική αναζήτηση ή να αποκτήσει άμεσα σημαντικό όγκο εξειδικευμένων πληροφοριών το διαδίκτυο παρέχει απίθανες δυνατότητες. Σε κάθε όμως περίπτωση τίθεται επιτακτικά το ζήτημα της εγκυρότητας και της αξιοπιστίας των «πηγών» του διαδικτύου όπου μια γενική νομοθεσία φαίνεται πως θα αργήσει πολύ ακόμα. Πολλοί μελετητές πιστεύουν πως υπάρχει μεγάλη ανάγκη για την ανάπτυξη της κριτικής σκέψης των μαθητών, ακόμη και αν αυτό χρειαστεί να λάβει το χαρακτήρα υποχρεωτικής εκπαίδευσης (Kling, 1996). Η ανάγκη για κριτική σκέψη είναι μεγάλη τόσο στο στάδιο της αναζήτησης των πληροφοριών, όσο και στο στάδιο της επιλογής και κατάταξης τους.

5.6 ΠΑΡΑΔΕΙΓΜΑΤΑ CASE STUDIES

Στη συγκεκριμένη ενότητα θα μελετήσουμε συγκεκριμένα παραδείγματα με τη μορφή case studies, προκειμένου να αξιολογήσουμε συγκεκριμένους μηχανισμούς. Δηλαδή συγκεκριμένα:

1. MyYahoo: Μέσα από το MyYahoo ο χρήστης καλείται να καθορίσει τις προτιμήσεις του. Συγκεκριμένα:

1. Χρηματιστήριο.
2. Πρόγνωση καιρού και αθλητικά.
3. Προτίμηση διαφόρων ιστοσελίδων από 1-3 ως βασικές.

Το περιεχόμενο των σελίδων αυτών στη συνέχεια θα προσαρμόζεται αυτόματα, ανταποκρινόμενο όμως πάντα στις επιλογές που έχουν γίνει. Εκτός από το περιεχόμενο της κάθε ιστοσελίδας, ο χρήστης μπορεί ακόμα να προσωποποιήσει και την εμφάνιση της. Επίσης αξίζει να σημειωθεί ότι μέρος των στοιχείων που εμφανίζονται στον κάθε χρήστη είναι δυνατό να προσωποποιηθούν αυτόματα, ανάλογα με τα δεδομένα που αυτός έχει εισάγει για τον εαυτό του. Έτσι, η πρόγνωση του καιρού μπορεί να αρχίζει με την πόλη που έχει δηλώσει ο χρήστης ως τόπο διαμονής.

Άλλα πλεονεκτήματα που δίνει το συγκεκριμένο εργαλείο στο χρήστη είναι:

1. Yahoo Companion: Το συγκεκριμένο είναι μια μικρή μπάρα που ενσωματώνεται στο βοηθό πλοήγησης (browser) και με τη βοήθεια της ο χρήστης αποκτάει πρόσβαση στα κυριότερα χαρακτηριστικά του Yahoo! χωρίς να είναι αναγκαίο να βρίσκεται στη σελίδα του Yahoo. Η συγκεκριμένη προσαρμόζεται στις ανάγκες και προτιμήσεις που έχει ήδη καθορίσει ο χρήστης.

2. Inside Yahoo Search: χρησιμοποιεί πληροφορίες που έχει συλλέξει για ένα

χρήστη κατά την εγγραφή του ώστε να προσωποποιήσει τα αποτελέσματα των αναζητήσεων που αυτός πραγματοποιεί. Έτσι, όταν ο χρήστης ψάχνει στο διαδίκτυο για εστιατόρια, το Inside Yahoo! Search με βάση τον ταχυδρομικό κώδικα που αυτός έχει δηλώσει, μπορεί να του προτείνει εκείνα τα εστιατόρια που βρίσκονται κοντά στην περιοχή που μένει.

Με βάση την αξιολόγηση τα προβλήματα του συγκεκριμένου εργαλείου είναι τα ακόλουθα:

1. Πολλές φορές ο σχεδιασμός του συστήματος γίνεται με τέτοιο τρόπο που υπάρχει πρόβλημα ταχύτητας και αποτελεσματικότητας του δικτυακού τόπου μια και αριθμός των χρηστών που χρησιμοποιούν τις δυνατότητες προσωποποίησης του αυξάνεται συνεχώς.

2. Υπάρχει πρόβλημα μυστικότητας και ασφάλειας (privacy and security) των προσωπικών δεδομένων που εμπιστεύεται ο χρήστης στους δημιουργούς του δικτυακού τόπου. Τα στοιχεία αυτά οφείλονται να χρησιμοποιηθούν με ιδιαίτερη λεπτότητα και προσοχή, ενώ όποια διαφορετική αντιμετώπιση μπορεί να έχει καταστροφικές συνέπειες.

3. Υπάρχει πρόβλημα ευελιξίας και πρόβλημα διαμεσολάβησης.

4. Έξυπνη Δικτυακή Αναζήτηση: Σε περίπτωση που η περιγραφή του χρήστη είναι ασαφής, τα αποτελέσματα δεν θα είναι αρκετά ικανοποιητικά και κατά πάσα πιθανότητα θα αποπροσανατολίσουν το χρήστη. Παράλληλα, τα συμβατικά προγράμματα αναζήτησης δεν λαμβάνουν καθόλου υπόψη τους ότι οι δικτυακές σελίδες είναι σχεδιασμένες σαν hypertext (υπερ-κείμενο) που σημαίνει ότι οι σύνδεσμοι (links) ανάμεσα στις σελίδες αλλά και η ίδια η δομή των διαφόρων δικτυακών τόπων περιλαμβάνουν σημαντικές πληροφορίες.

Η έξυπνη δικτυακή αναζήτηση προσπαθεί να κάνει χρήση αυτών και άλλων παρόμοιων χαρακτηριστικών ώστε να διευκολύνει το έργο του χρήστη, μαθαίνοντας στην ουσία από προηγούμενα παραδείγματα και χρησιμοποιώντας αυτή τη γνώση για να αυξήσει την

ταχύτητα και την ακρίβεια της μελλοντικής αναζήτησης πληροφοριών στο Διαδίκτυο. Σε επίπεδο αξιολόγησης έχουν παρουσιαστεί διαφορετικές απόψεις για τον τρόπο με τον οποίο μπορεί να επιτευχθεί η Έξυπνη Δικτυακή Αναζήτηση. Κοινό σημείο των περισσότερων είναι ότι οι ευφυείς πράκτορες μπορούν να παρέχουν ολοκληρωμένη βοήθεια στους χρήστες, μόνο αν τους «ακολουθούν» και εργάζονται παράλληλα με αυτούς καθόλη τη διάρκεια της αναζήτησης που αυτοί διενεργούν.

5.6.1 Το Amazon & άλλα παραδείγματα Δικτυακής Αναζήτησης

AMAZON: Εδώ και 550 εκατομμύρια χρόνια, στην κάμβρια παλαιοντολογική περίοδο, σημειώθηκε μία έξαρση ανάπτυξης. Για άγνωστους λόγους, τα είδη διακλαδώθηκαν σε πολυκυτταρικό και πολλαπλασιάστηκαν με μία εκπληκτική ταχύτητα. Δημιουργήθηκαν εξαιρετικά πολλές νέες μορφές ζωής, όπως και εξαλείφθηκαν άλλες. Σήμερα το Internet περνάει την κάμβρια περίοδο του. Ένα λίθος και θα βρεθούμε και εμείς στο σωρό των εξαλειφθέντων ειδών. Αν όμως επιβιώσουμε, μας δίνεται η ευκαιρία να γράψουμε μία υποσημείωση στην ιστορία των “επιχειρήσεων”.

Κάπως έτσι περιγράφει την οπτική του για το σημερινό Internet ο Τζεφ Μπέζος, ιδρυτής και πρόεδρος του amazon.com, του μεγαλύτερου βιβλιοπωλείου on-line αλλά και του πιο δημοφιλούς εμπορικού τόπου για καταναλωτές στο Internet. Μέσα σε τέσσερα μόλις χρόνια η Amazon πήγε από το γκαράζ του Μπέζος, τον Ιούλιο του 1995 που άνοιξε τις εικονικές πόρτες του το amazon.com, σε τέσσερα επιβλητικά κτίρια στο Σιατλ. Τον πρώτο χρόνο της λειτουργίας της η Amazon είχε ένα εκατομμύριο καταγεγραμμένους πελάτες. Σήμερα οι πελάτες της ξεπερνούν τα 12 εκατομμύρια σε 160 χώρες! Το ίδιο εντυπωσιακός είναι και ο τζίρος της Amazon: για το 1997 ήταν 149 εκατομμύρια δολάρια, δέκα φορές

μεγαλύτερος από τον τζίρο του αμέσως επόμενου ανταγωνιστή της, Barnes and Noble, ενώ για το 1998 έφτασε τα 610 εκατομμύρια

Δηλαδή αύξηση 313% σε ένα χρόνο), έναντι 20 εκατομμυρίων δολαρίων του Barnes and Noble.

Σήμερα ο τόπος της Amazon δεν πουλά μόνο βιβλία. Πουλά CD, DVD, καταναλωτικά ηλεκτρονικά, είδη σπιτιού και δώρα, ενώ πρόσφατα πρόσθεσε και δημοπρασίες on-line, όπου οποιοσδήποτε μπορεί να πουλήσει οτιδήποτε. Το περίεργο και το ενδιαφέρον όμως είναι ότι, ενώ η Amazon έχει κερδίσει τον πόλεμο σχεδόν από την πρώτη μέρα της λειτουργίας της, μέχρι σήμερα... δεν έχει κερδίσει ούτε δραχμή. Χάνει συνεχώς χρήματα, αλλά η μετοχή της -μπήκε στο Χρηματιστήριο της Νέας Υόρκης το 1997 με αξία 1,5 δολάρια- έχει φτάσει τα 74 δολάρια! Είναι ένα φαινόμενο που, αν το Harvard Business School δεν έχει συμπεριλάβει ακόμη στις επιχειρηματικές περιπτώσεις που μελετά, σίγουρα θα το κάνει σύντομα

Στο τέλος του 1994, τότε που ο Μπέζος άφησε τη δουλειά του οικονομικού αναλυτή στη Νέα Υόρκη, για να ξεκινήσει κάτι καινούριο στο Web, τα πράγματα ήταν πολύ διαφορετικά. Ένα από το παρατσούκλια του WWW ήταν “World Wide Wait”, λόγω της μεγάλης αναμονής και καθυστέρησης που σημειωνόταν σε ένα δίκτυο που είχε σχεδιαστεί για άλλη και μικρότερη χρήση. Προκειμένου λοιπόν να πουλήσει οτιδήποτε σε ένα αργό διαδικτυο, ο Μπέζος έπρεπε να προσπαθήσει να κάνει την αγορά από ευχάριστη έως πολύ διασκεδαστική. Να προσφέρει δηλαδή αξία στον πελάτη, που θα αντισταθμίσει τις όποιες καθυστερήσεις του δικτύου. Έτσι επένδυσε με έμφαση στον τομέα “φροντίδα του πελάτη”. Τον φρόντισε με ένα νέο τρόπο, που, όπως φαίνεται, ήταν αποτελεσματικός.

Γιατί όμως πηγαίνουν και ξαναπηγαίνουν οι πελάτες στην Amazon; πηγαίνουν διότι έχει μεγάλη ποικιλία προϊόντων, διότι η πλοήγηση στον τόπο είναι εύκολη, διότι ο τόπος

ανανεώνεται προσφέροντας νέες υπηρεσίες και διότι τους αρέσει η εξαιρετική χρήση του e-mail για Αποστολή πληροφοριών και αμφίδρομη επικοινωνία (δηλαδή για marketing και εξυπηρέτηση πελατών). Πηγαίνουν, τέλος, διότι η Amazon έχει έναν πολύ καλό τρόπο να προτείνει προϊόντα χωρίς πιεστικό τρόπο. Η εμπειρία του shopping στο amazon είναι, με άλλα λόγια, εύκολη και ευχάριστη. Ο πελάτης μπορεί να μπει οποτεδήποτε και να βρει ότι εμφανίστηκε στο διάστημα που έλειψε. Ακόμη όμως και αν δεν μπει, η Amazon τον θυμάται μέσω του e-mail και του στέλνει μηνύματα με παραπομπές που του εξάπτουν την περιέργεια. Όλα αυτά όμως μπορούν να αναγράφουν από τους ανταγωνιστές. Τι είναι αυτό που κάνει τη διαφορά;

Η δυνατότητα του amazon.com να είναι πρωτοπόρο, να ανακαλύπτει νέους τρόπους επικοινωνίας και να βρίσκεται πάντα μπροστά από τον ανταγωνισμό. Πώς ακριβώς συμβαίνει αυτό; “Το ξέρω ότι ακούγεται γελοίο απλά” λέει ο Μπέζος “αλλά ρωτάμε τους πελάτες μας τι θέλουν”. Αυτό το “ρωτάμε” όμως δεν είναι απλό σημαίνει ότι προσπαθούν να ενθαρρύνουν την επικοινωνία με e-mail, ότι αναλύουν πολύ προσεκτικά την αγοραστική συμπεριφορά των πελατών τους και τη συμπεριφορά τους κατά την απλή περιήγηση του τόπου. Είναι μία υποδειγματική χρήση του marketing 1:1, όπου η εταιρεία απευθύνεται άμεσα στον πελάτη και αυτός νιώθει ότι τον μεταχειρίζονται ως μοναδική και ξεχωριστή οντότητα. Όλη αυτή η ανάλυση συμπεριφοράς με ένα διάλογο που διεξάγεται “σταγόνα-σταγόνα”, γίνεται με έναν τρόπο που δεν ενοχλεί τους πελάτες. Έτσι, η εταιρεία χρησιμοποιεί αυτά τα δεδομένα για να “εφεύρει πράγματα που υποψιαζόμαστε πως θέλει ο κόσμος” συνεχίζει ο Μπέζος “όπως αγορές με ένα κλικ, λίστες best-seller και ομάδες ειδικού ενδιαφέροντος, π.χ., ποια βιβλία και CD είναι δημοφιλή στο Yale”.

Η Amazon εμφανίζεται επίσης πανταχού παρούσα. Επεκτείνει την παρουσία της στο Διαδίκτυο με το πρόγραμμα των “συνεργατών”, όπου σε έναν τόπο μπορεί να υπάρχει μία παραπομπή προς το amazon.com για συγκεκριμένα βιβλία που σχετίζονται με τον τόπο και που, αν αγοραστούν, δίνουν ένα ποσοστό στον τόπο από όπου προήλθε η πώληση.

Με όλα αυτά δεν θα πρέπει να προκαλεί έκπληξη το γεγονός ότι το μαγαζί της Amazon δεν χάνει πελάτες. Το αντίθετο μάλιστα ένα ποσοστό 65% των πελατών είναι επαναληπτικό, ποσοστό εξαιρετικά υψηλό γι' αυτόν το χώρο. Γιατί άλλωστε οι πελάτες να πανε αλλού. Για τα λίγα σεντς που ίσως να είναι φθηνότερο ένα βιβλίο δεν είναι διατεθειμένοι να χάσουν αυτήν τη σχέση που έχουν καλλιεργήσει. Είναι σαν έναν ευτυχισμένο γάμο, σίγουρα υπάρχουν παντού πειρασμοί, αλλά αξίζει τον κόπο να αναστατώσει κανείς μία καλή σχέση και μία σταθερή ισορροπία για κάτι άγνωστο?

Σαφώς λοιπόν η Amazon είχε τη σωστή στρατηγική, σαφώς είχε τις σωστές, φιλοδοξίες και με αυτές κατέκτησε ένα ζωτικό χώρο στο internet, από όπου μπόρεσε να επεκταθεί.

Όπως είπαμε όμως και παραπάνω, είχε από την αρχή και ορισμένα σημαντικά ατού με τα οποία κέρδισε τη μάχη

1. Το σωστό προϊόν. Δεν χρειάζεται να δοκιμάσεις, να αγγίξεις η να μυρίσεις ένα βιβλίο για να το αγοράσεις. Το προϊόν είναι όσο στάνταρτ γίνεται. Δεν είναι κεράσια Βοδενών.

2. Το σωστό marketing.' Ήταν πραγματικά από τους πρώτους που είχε μία προσωπική σχέση με τον πελάτη, μία σχέση “έναν προς έναν”. Θυμάται και “ήξερε” τι ήθελε ο πελάτης, χωρίς να τον ενοχλεί.

3. Το σωστό περιεχόμενο. Σχόλια και έγκυρες βιβλιοκριτικές, συνεντεύξεις, περιεχόμενα των βιβλίων, σχόλια των αναγνωστών και αξιολογήσεις τους, με την ανέξοδη συμμετοχή των προγενέστερων πελατών. Το περιεχόμενο αυτό βρισκόταν πάντα πολύ κοντά στη δυνατότητα και την απόφαση του πελάτη να αγοράσει.

4. Τα σωστά πλεονεκτήματα. Πουλούσε ένα , προϊόν στο οποίο η ποικιλία και η πληρότητα μετρούν. Είχε, η μπορούσε να βρει, όλα τα βιβλία που κυκλοφορούν, κάτι που είναι πολύ σημαντικό για τον υποψήφιο πελάτη, ειδικά τον επαναληπτικό αγοραστή.

5. Τη σωστή λειτουργικότητα. Για να αγοράσεις από την Amazon, θα έπρεπε η διαδικασία να είναι ευκολότερη από αυτήν του βιβλιοπωλείου. Θα έπρεπε να μπορείς να παραλάβεις γρήγορα όλα τα βιβλία που παραγγέλλεις -όχι μόνο τα πιο δημοφιλή- και να περιμένεις εβδομάδες για τα υπόλοιπα. Επίσης, αν κάτι επρόκειτο να καθυστερήσει, έπρεπε να το ξέρεις από την αρχή και για να μπορείς να αποφασίσεις αν θέλεις να περιμένεις, αλλά και για να αποφύγεις τις δυσάρεστες εκπλήξεις αργότερα. Εδώ η Amazon δεν το κατάφερε πάντα τέλεια, αλλά τα κατάφερε πολύ καλύτερα από τον ανταγωνισμό και έκανε περισσότερα πράγματα σωστά παρά λάθος, πράγμα σημαντικό.

Και το μέλλον? Σε έναν τομέα τόσο καυτό όσο το ηλεκτρονικό εμπόριο, φυσικό είναι οι γνώμες να διχάζονται. Προβλέπεται πως η Amazon θα είναι ζημιογόνα μέχρι το 2000, παρ' όλο που η μετοχή της έχει φτάσει στα ύψη. Από τη μία πλευρά υπάρχουν εκείνοι που λένε πως τώρα πλέον κανείς και τίποτα δεν μπορεί να τη σταματήσει, διότι έχει μεταμορφωθεί σε ένα σημαντικό πόλο συνολικής εμπορικής δραστηριότητας και ο ανταγωνισμός πρακτικά δεν υπάρχει. Από την άλλη υπάρχουν εκείνοι που λένε πως το επιχειρησιακό μοντέλο της φθάνει και, καθώς τώρα μπαίνει σε αγορές με μικρότερα ποσοστά κέρδους, θα αρχίσει να τα βρίσκει δύσκολα. Κατ' ουσία πρόκειται για μία μάχη με το χρόνο προκειμένου να εγκαθιδρυθεί ένα μονοπώλιο αν, δηλαδή, η εταιρεία θα προλάβει να γίνει σχεδόν ο αποκλειστικός, η μάλλον ο προτιμητέος, πόλος μιας ευρύτερης διαδικτυακής καταναλωτικής δραστηριότητας και όχι μόνο λιανικής (κάτι πολύ δύσκολο).

Παρακολουθούμε με αγωνία και ενδιαφέρον. Επειδή, όπως είπαμε, η εξέλιξη στο Διαδίκτυο περνάει την κάμβρια περίοδο, το αποτέλεσμα δεν θα αργήσει να φανεί.

Ως παράδειγμα του κεφαλαίου θα παραθέσουμε το ηλεκτρονικό βιβλιοπωλείο Amazon.com το οποίο πληρεί, αν όχι όλες, τις περισσότερες προϋποθέσεις ενός καλού ηλεκτρονικού καταστήματος.

• **Εισαγωγή στο κατάστημα.** Αρχικά, εμφανίζεται στον χρήστη μια οθόνη καλωσορίσματος. Αν ο χρήστης έχει επισκεφτεί πάλι το site και έχει ήδη δημιουργήσει έναν λογαριασμό στο κατάστημα, τότε μπορεί να εισέλθει χρησιμοποιώντας το προσωπικό του user name και τον δικό του κωδικό πρόσβασης. Σε περίπτωση που είναι η πρώτη φορά που επισκέπτεται το κατάστημα, τότε μπορεί να κάνει εγγραφή (registration) προκειμένου να γίνει πλέον μέλος-πελάτης. Παραθέτουμε μέρος αρχικής σελίδας Amazon.com όπου ο χρήστης ερωτάτε αν έχει επισκεφτεί το κατάστημα για πρώτη φορά.



Εικόνα 7.1 Αρχική σελίδα του Amazon.

Εδώ ο χρήστης εισάγει τον προσωπικό του κωδικό. Πρόκειται για χρήστη ο οποίος έχει ξανά χρησιμοποιήσει το κατάστημα και έχει κάνει εγγραφή (registration) σε αυτό. Παρακάτω βλέπουμε την συγκεκριμένη Φόρμα εγγραφής στο Amazon.



Εικόνα 7.2 Φόρμα εγγραφής κατά την είσοδο σε χρήστη που έχει ξανά επισκεφθεί το κατάστημα Amazon

Στην περίπτωση που ο πελάτης εισέρχεται για πρώτη φορά στο κατάστημα και επιλέξει να κάνει εγγραφή, τότε θα εμφανιστεί η παρακάτω φόρμα, στην οποία θα πρέπει να συμπληρώσει τα απαιτούμενα στοιχεία (user name, password, ονοματεπώνυμο, διεύθυνση κλπ.).



Εικόνα 7.3 Φόρμα εγγραφής κατά την πρώτη είσοδο στο Amazon

- **Πλοήγηση στο κατάστημα.** Μετά την εγγραφή του ή δίνοντας κατευθείαν το αναγνωριστικό και τον κωδικό πρόσβασης, ο πελάτης μπορεί πλέον να πλοηγηθεί στο κατάστημα. Έτσι βλέπει μια λίστα με τα προϊόντα, αλλά και κάποιες ειδικές τιμές που προσφέρονται για συγκεκριμένα προϊόντα, σε προσφορές.

(Στο πάνω μέρος υπάρχουν οι κατηγορίες ενώ κάτω αριστερά ο πελάτης μπορεί να γράψει κάποιο από τα χαρακτηριστικά του προϊόντος που θέλει (π.χ το όνομα του βιβλίου που ενδιαφέρεται)

- **Μηχανισμός αναζήτησης.** Στην περίπτωση που ο χρήστης επιθυμεί να αναζητήσει ένα συγκεκριμένο προϊόν, τότε μπορεί να το αναζητήσει με δύο τρόπους. Με βάση αυτό τον πρώτο τρόπο, που ονομάζεται και "έξυπνη αναζήτηση" (smart search), ο πελάτης αναζητά το προϊόν με συγκεκριμένα κριτήρια, που έχουν οριστεί από τον διαχειριστή του καταστήματος, όπως είναι η κατασκευάστρια εταιρεία, το όνομα, το μοντέλο προϊόντος κ.α. θέτοντας τιμές για κάποια. από αυτά τα κριτήρια., το σύστημα ανακτά μια. λίστα προϊόντων που ανταποκρίνεται σε αυτά. Ο πελάτης ελέγχει την λίστα και επιλέγει το προϊόν που επιθυμεί. Ο δεύτερος τρόπος είναι ανά κατηγορία . Δηλαδή ο χρήστης αναζητά το προϊόν που θέλει βάσει των κατηγοριών που υπάρχουν. Πχ αν θέλει να αγοράσει ένα dvd θα πάει στη κατηγορία dvd μετά κατηγορία έργου, όνομα ηθοποιών κτλ.

- **Καλάθι αγορών.** Επιλέγοντας ο πελάτης κάποιο συγκεκριμένο προϊόν, είτε μέσω της πλοήγησής του στο κατάστημα είτε μέσω του μηχανισμού αναζήτησης, εμφανίζεται μια λεπτομερή περιγραφή του προϊόντος, μαζί με φωτογραφία, τον κωδικό του, καθώς και με την τιμή αγοράς του. Έτσι δίνεται η δυνατότητα στον πελάτη να αποφασίσει αν θα το προσθέσει στο «καλάθι αγορών» ή όχι.



Εικόνα 7.4 Φόρμα συμπλήρωσης προϊόντων στο «Καλάθι αγορών» του Amazon

Αν ο πελάτης αποφασίσει να προσθέσει στο καλάθι αγορών κάποιο προϊόν, τότε εμφανίζεται στην οθόνη του η κατάσταση του καλάθιού, ενημερώνεται δηλαδή για τι προϊόντα έχει βάλει στο καλάθι του, τις τιμές τους, τους κωδικούς τους καθώς και για τη συνολική αξία των αγορών του. Βέβαια, παρέχεται η δυνατότητα στον πελάτη να αφαιρέσει κάποιο προϊόν από το καλάθι του, αν δεν το επιθυμεί πλέον, ή να αλλάξει την ποσότητα παραγγελίας σε κάποιο άλλο.

Παραγγελιοδοσία. Στην περίπτωση που ο πελάτης αποφασίσει να αγοράσει τα προϊόντα που έχει ήδη προσθέσει στο καλάθι αγορών τότε εμφανίζεται πάλι η αναλυτική κατάσταση του καλάθιού, αλλά επιπρόσθετα παρουσιάζεται η φορολόγηση που θα επιβληθεί, ενώ καλείται ο πελάτης να επιλέξει πιστωτική κάρτα για την αγορά του από μια λίστα καρτών που υποστηρίζει το κατάστημα και η οποία έχει καθοριστεί από τον διαχειριστή του καταστήματος. Πρέπει να εισαχθούν τα στοιχεία της πιστωτικής κάρτας, και επιπλέον πρέπει να επιλεγεί και η μέθοδος αποστολής των προϊόντων. Παρέχεται μια λίστα τρόπων παράδοσης προϊόντων, με αναλυτική περιγραφή αυτών (εταιρία αποστολής,

διάρκεια αποστολής και παράδοσης, έξοδα αποστολής). Ζητούνται τα στοιχεία του παραλήπτη, διεύθυνση πόλη, χώρα κτλ

The screenshot shows the Amazon.com shipping address form. At the top, the Amazon logo and navigation links (WELCOME, ADDRESS, ITEMS, WRAP, SHIP, PAY, CONFIRM) are visible. The main heading is "Please enter the shipping address". The form includes fields for Full Name, Address Line 1 (or company name), Address Line 2 (optional), City, State/Province/Region, ZIP/Postal Code, Country (set to United States), and Phone Number. A "Continue" button is located below the fields. Below the "Continue" button, there is a note: "OR, if you're sending items to more than one address, save this address and" followed by an "Add Another Address" button. At the bottom, a small note provides instructions for APO or FPO addresses.

Εικόνα 7.5 Φόρμα συμπλήρωσης παραγγελιοδοσίας στο Amazon

The screenshot shows the Amazon payment method form. It starts with a password field labeled "Enter a password:" and a confirmation field labeled "Type it again:". Below this is the "Select a payment method." section, which includes a note about re-entering credit card numbers and a "We welcome" banner with logos for MasterCard, American Express, Discover, and Visa. The payment method selection is shown in a table:

Payment Method	Credit Card No.	Expiration Date	Name on Credit Card
<input checked="" type="radio"/> Visa		01 2000	
<input type="radio"/> Pay by check or money order (or check funds on account)			

Below the table is a section for "Do you have a Gift Certificate?". It includes a field for "Enter claim code" with an "Enter another" button. A note states: "If your Gift Certificate doesn't cover the cost of your order, we will use the card you enter above for the balance. (You may not pay by check in this case.) Find out more about Gift Certificates." At the bottom, there is a question "Is this your billing address?" with the address: zak Maissis, kononos 31, Athens, 85142, Greece. Two buttons are provided: "No, enter a different one" and "Yes, continue".

Εικόνα 7.5.1 Φόρμα συμπλήρωσης παραγγελιοδοσίας στο Amazon

Η παραγγελιοδοσία, ολοκληρώνεται, μόλις ο πελάτης στείλει την εντολή πληρωμής, έχοντας επιλέξει πιστωτική κάρτα και μέθοδο αποστολής. Τέλος, μπορεί να ελέγξει την κατάσταση της παραγγελίας του.

(επιλογή πιστωτικής κάρτας και συμπλήρωση των στοιχείων του κατόχου π.χ αριθμός, όνομα.)

Συνοπτικά παρακάτω αξιολογούνται μερικά από τα συστήματα που έχουν κατά καιρούς αναπτυχθεί με στόχο την Έξυπνη Δικτυακή Αναζήτηση:

1. Alexa Internet: Ο βοηθός δικτυακής πλοήγησης “Alexa Internet” έχει ως στόχο του να παρέχει την τεχνολογία που θα επιτρέψει στον καθένα να μοιραστεί τη γνώση του με τους άλλους, για το κοινό όφελος όλων. Ο μηχανισμός του Alexa Internet διατρέχει διαρκώς το χώρο του Διαδικτύου και συλλέγει δικτυακούς τόπους που είναι δημοσίως διαθέσιμοι, τους αναλύει, ανακαλύπτει πιθανά μοτίβα (patterns) και αποθηκεύει όσες πληροφορίες τελικά συγκεντρώσει σε γιγαντιαίες βάσεις δεδομένων. Αυτές οι βάσεις δεδομένων περιέχουν, εκτός από τις διευθύνσεις των δικτυακών τόπων που βρέθηκαν, περιγραφές των τόπων αυτών καθώς και αναφορές σε άλλους τόπους με σχετικό περιεχόμενο.

2. ANTWORLD :Χρησιμοποιείται ως μέσω αξιολόγησης ιστοσελίδων. Συγκεκριμένα όταν κάποιος χρήστης επισκεφτεί ένα δικτυακό τόπο, του ζητείται να καταγράψει τις εντυπώσεις του από αυτό τον τόπο. Η καταγραφή των εντυπώσεων μπορεί να γίνει χρησιμοποιώντας μια δεδομένη βαθμολογική κλίμακα -ώστε να βαθμολογηθεί η ποιότητα της σελίδας- αλλά και με την εισαγωγή σύντομου κειμένου που μπορεί να περιέχει κάποια επιπρόσθετα -θετικά ή αρνητικά- σχόλια.

3. WebWatcher: Το σύστημα WebWatcher περιγράφεται από τους δημιουργούς του

σαν ένα εργαλείο που συνοδεύει το χρήστη και του κάνει προτάσεις για ιστοσελίδες που μπορεί να επισκεφτεί στη συνέχεια. Ο χρήστης έχει τη δυνατότητα να επικοινωνήσει με το σύστημα καθώς και να υπάρξει ανάδραση (feedback) μεταξύ τους. Το WebWatcher παρακολουθεί και μαθαίνει από τις ενέργειες των χρηστών, με αποτέλεσμα να προσαρμόζει διαρκώς τη συμπεριφορά του και να βελτιώνει την απόδοση του όταν επισκέφτεται δικτυακούς τόπους που έχει ξαναδεί στο παρελθόν. Ενδιαφέρον παρουσιάζει ο τρόπος με τον οποίο γίνεται η «μάθηση» στο WebWatcher. Ο σκοπός του συστήματος είναι να προτείνει πάντα τον καταλληλότερο σύνδεσμο από αυτούς που βρίσκονται σε μια ιστοσελίδα με βάση τον παρόντα χρήστη, τα ενδιαφέροντα που αυτός έχει εκδηλώσει στην συγκεκριμένη αναζήτηση και φυσικά την σελίδα που κάθε φορά είναι ενεργή.

4. HyperSpace: Το σύστημα HyperSpace είναι ένα εργαλείο για την οπτικοποίηση της «απόστασης» (visualising the distance) ανάμεσα στις διαφορετικές δικτυακές σελίδες που επισκέπτεται ένας χρήστης. Συγκεκριμένα, παρακολουθεί τη συμπεριφορά του χρήστη και τη δομή των σχετικών τόπων που αυτός επισκέπτεται τακτικά. Στη συνέχεια αυτές οι πληροφορίες χρησιμοποιούνται για να κατασκευαστεί ένας τρισδιάστατος χάρτης τμημάτων του Διαδικτύου.

5. Letizia: Το σύστημα Letizia λειτουργεί σαν ένας βοηθός, ο οποίος κάνει προτάσεις στο χρήστη σχετικά με ιστοσελίδες που μπορεί να τον ενδιαφέρουν. Οι προτάσεις αυτές γίνονται με βάση την παρούσα σελίδα που είναι ενεργή καθώς και τη γνώση των λοιπών προτιμήσεων που κατά καιρούς έχει επιδείξει ο χρήστης. Παράλληλα, το σύστημα λαμβάνει υπόψη και άλλες, πιο ευαίσθητες, παραμέτρους σε αυτή την επιλογή των προτεινόμενων τόπων. Επειδή οι χρήστες συνήθως διαβάζουν τις ιστοσελίδες ξεκινώντας από την κορυφή, όταν ένας σύνδεσμος που δεν έχει επιλεγεί, βρίσκεται νωρίτερα σε μια σελίδα (σε σχέση με άλλους συνδέσμους που έχουν ενεργοποιηθεί και βρίσκονται αργότερα στην ίδια σελίδα), ο σύνδεσμος αυτός κρίνεται μειωμένου ενδιαφέροντος. Οι όποιες προτάσεις τελικά επιλεγούν, παρουσιάζονται σε ξεχωριστό παράθυρο, ώστε ο χρήστης να έχει τη δυνατότητα να τις

αγνοήσει αν επιθυμεί κάτι τέτοιο.

WEBSOM: Η μέθοδος WEBSOM επιχειρεί να οργανώσει συλλογές από κείμενα σε έναν οπτικό χάρτη (visual map display) όπου παρόμοια κείμενα βρίσκονται κοντά το ένα στο άλλο. Αυτή η οργάνωση επιτυγχάνεται αυτόματα με τη χρήση ενός αλγορίθμου νευρωνικών δικτύων που ονομάζεται SOM (self-organizing map ή αυτό-οργανωμένος χάρτης; Kohonen 1982). Ο συγκεκριμένος αλγόριθμος μπορεί να τοποθετεί αυτόματα πολυδιάστατα στατιστικά δεδομένα με τέτοιο τρόπο, ώστε δεδομένα που μοιράζονται κοινές ιδιότητες να βρίσκονται σε κοντινή απόσταση.

5.6.2 Πρότυπη Πλατφόρμα Ηλεκτρονικού Καταστήματος

Τα δομικά συστατικά ενός ηλεκτρονικού καταστήματος, έτσι ώστε να έχει η συγκεκριμένη εμπορική πλατφόρμα τις προδιαγραφές και τα χαρακτηριστικά, τη λειτουργικότητα και την αρχιτεκτονική ώστε εκδοθεί αυτό στο διαδίκτυο, είναι τα ακόλουθα:

- **Βάση δεδομένων.** Για την αποθήκευση των δεδομένων του καταστήματος (προϊόντα, περιγραφή προϊόντων και τμημάτων του καταστήματος, στοιχεία πελατών) είναι απαραίτητη η χρήση βάσης δεδομένων, έτσι ώστε να είναι εφικτή η συντήρηση και ανάκτηση πληροφοριών.

- **Δικτυακή υποδομή.** Για την σύνδεση του καταστήματος με το διαδίκτυο και τη λειτουργία του σε αυτό, είναι απαραίτητη η εγκατάσταση κάποιας δικτυακής υποδομής.

- **Πλατφόρμα ηλεκτρονικού εμπορίου.** Για την έκδοση του ηλεκτρονικού καταστήματος στο Internet, είναι απαραίτητη η εγκατάσταση ενός ηλεκτρονικού περιβάλλοντος (πλατφόρμα), το οποίο πρέπει να διαθέτει χαρακτηριστικά που θα επιτρέπουν την υλοποίηση των λειτουργιών ενός ηλεκτρονικού καταστήματος.

- **Σύστημα Πληρωμών.** Για την ολοκλήρωση της λειτουργίας του ηλεκτρονικού καταστήματος, πρέπει να υλοποιηθεί σύστημα ηλεκτρονικών πληρωμών, έτσι ώστε να ολοκληρωθεί ο κύκλος της παραγγελιοδοσίας.

- **Μηχανισμοί Ασφάλειας.** Για την προστασία της ιδιωτικότητας των πελατών και των πληροφοριών που διακινούνται κατά τις ηλεκτρονικές συναλλαγές, πρέπει να υπάρξει ένας μηχανισμός ασφάλειας που θα προσαρμοστεί στο περιβάλλον του ηλεκτρονικού καταστήματος.

Τέλος αναφερόμενοι στους ευφυείς πράκτορες θα πρέπει να πούμε ότι Σύμφωνα με τους Lange και Oshima πράκτορας είναι ένα αντικείμενο λογισμικού (software object) που κινείται σε ένα δεδομένο περιβάλλον εκτέλεσης (execution environment) και διαθέτει υποχρεωτικά μια σειρά από ιδιότητες. Συγκεκριμένα, θα πρέπει να είναι αντιδραστικός (reactive – να αντί λαμβάνεται αλλαγές στο περιβάλλον και να δρα ανάλογα με αυτές), αυτόνομος (autonomous – να διατηρεί τον έλεγχο των ενεργειών του), να οδηγείται από τους στόχους και τα καθήκοντα που του έχουν ανατεθεί (goal driven) και τέλος να ενεργεί χωρίς διακοπές στο χρόνο (temporally continuous). Ακόμα, ένας πράκτορας ενδέχεται να συγκεντρώνει κάποιες από τις ακόλουθες ιδιότητες, σε προαιρετική όμως βάση: να μπορεί να επικοινωνεί με άλλους πράκτορες (communicative), να μπορεί να κινείται ανάμεσα σε hosts (mobile), ή τέλος να μπορεί να προσαρμόζεται ανάλογα με πρότερη εμπειρία (learning). Τα είδη των πρακτόρων είναι τα ακόλουθα:

1. Πράκτορες Συνεργασίας (Collaborative Agents): Οι πράκτορες συνεργασίας δίνουν σημασία στην αυτονομία και τη συνεργασία με άλλους πράκτορες για να εκτελέσουν τις εργασίες που τους αναθέτουν οι «ιδιοκτήτες» (owners) τους. Συνήθως κινούνται σε ανοικτά περιβάλλοντα με πλήθος άλλων πρακτόρων και χρονικούς περιορισμούς (open and time-constrained multi-agent environments). Η λειτουργία τους ενδέχεται να περιλαμβάνει κάποια περιορισμένης μορφής μάθηση, χωρίς αυτό όμως να αποτελεί βασικό κομμάτι της συμπεριφοράς τους. Οι πράκτορες συνεργασίας συχνά

υποχρεώνονται να προχωρήσουν σε διαπραγματεύσεις (negotiations) με άλλους πράκτορες. Σκοπός των διαπραγματεύσεων αυτών είναι να καταλήξουν σε κάποιες από κοινού συμφωνίες που θα συντονίσουν καλύτερα τις ενέργειες τους.

2. Πράκτορες Διαμεσολάβησης: Οι πράκτορες διαμεσολάβησης πρέπει να υποστηρίζουν και να παρέχουν proactive βοήθεια, τυπικά σε άτομα που μαθαίνουν τη χρήση μιας συγκεκριμένης εφαρμογής, όπως είναι ένα λειτουργικό σύστημα. Η μάθηση του πράκτορα μπορεί να γίνει με έναν από τους εξής τρόπους: (i) παρακολουθώντας το χρήστη, (ii) λαμβάνοντας θετική και αρνητική ανάδραση από το χρήστη (iii) λαμβάνοντας ρητές οδηγίες από το χρήστη, και τέλος (iv) ζητώντας συμβουλές από άλλους πράκτορες. Βασική προϋπόθεση για την επιτυχή λειτουργία ενός πράκτορα διαμεσολάβησης είναι η εφαρμογή που χρησιμοποιείται να επιδεικνύει κάποια επαναλαμβανόμενη συμπεριφορά, την οποία να μπορεί να μάθει ο πράκτορας και ακόμα πιο βασικά, αυτή η συμπεριφορά να είναι διαφορετική για κάθε χρήστη – γιατί αλλιώς το πρόβλημα θα μπορούσε να λυθεί από ένα απλό knowledge-based σύστημα.

3. Πληροφοριακοί/Διαδικτυακοί Πράκτορες (Information/Internet Agents): Οι πληροφοριακοί πράκτορες ασχολούνται με τη διαχείριση, την επεξεργασία και την συναρμολόγηση δεδομένων που προέρχονται από πολλές κατανεμημένες πηγές. Η εμφάνιση τους οφείλεται κατά κύριο λόγο στην αρχική ανάγκη εύρεσης ενός τρόπου για να τιθασευτεί η πληθώρα πληροφοριών του Διαδικτύου. Φυσικά, η επιτυχής ανάπτυξη τέτοιων πρακτόρων μπορεί να επιφέρει σημαντικά οικονομικά οφέλη στους δημιουργούς τους

4. Αντιδραστικοί Πράκτορες (Reactive Software Agents): Οι αντιδραστικοί πράκτορες, για τους οποίους χρησιμοποιείται και η ονομασία «αυτόνομοι πράκτορες», αντιπροσωπεύουν μια ειδική κατηγορία πρακτόρων οι οποίοι δεν διαθέτουν εσωτερικά, συμβολικά μοντέλα του περιβάλλοντος τους. Απεναντίας, λειτουργούν με βάση ένα πλήθος ζευγαριών «ερεθίσματος-αντίδρασης». Για κάθε ερέθισμα που δέχονται όντας σε μια συγκεκριμένη κατάσταση (state) επιλέγουν την αντίστοιχη αντίδραση, την οποία και εκτελούν.

5. Κινητοί Πράκτορες (Mobile Agents): Οι κινητοί πράκτορες είναι διεργασίες λογισμικού με την ικανότητα να διατρέχουν (roaming) τον κόσμο του Διαδικτύου, να έρχονται σε επαφή με ξένους hosts, να εκτελούν τις εργασίες που έχουν αναλάβει από τους «ιδιοκτήτες» τους και τελικά να επιστρέφουν έχοντας φέρει σε πέρας όλα τα καθήκοντα τους. Τα καθήκοντα τους μπορούν να περιλαμβάνουν από κρατήσεις αεροπορικών εισιτηρίων μέχρι τη διαχείριση ενός τηλεπικοινωνιακού δικτύου. Οι κινητοί πράκτορες αποκαλούνται «πράκτορες» επειδή είναι αυτόνομοι και συνεργάζονται, αν και διαφορετικά απ' ότι οι πράκτορες συνεργασίας (collaborative agents).

ΚΕΦΑΛΑΙΟ 6

ΣΥΜΠΕΡΑΣΜΑΤΑ & ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ

Με βάση την ανάλυση που έγινε στη παρούσα πτυχιακή μελέτη είδαμε ότι το ίντερνετ έχει μπει πια σε όλες τις εκφάνσεις της ζωής του σύγχρονου ανθρώπου. Ουσιαστικά στο εμπόριο , στη παιδεία στην ευρύτερη λειτουργία των οργανισμών (δημόσιων και ιδιωτικών) το κομμάτι της εξυπηρέτησης και γενικά σε κάθε δραστηριότητα του μέσου ανθρώπου το ίντερνετ δίνει καινοτόμες και σύγχρονες λύσεις οι οποίες απλοποιούν σε σχέση με το παρελθόν τις παραδοσιακές διαδικασίες συναλλαγών.

Με δεδομένη την ανάγκη για πληροφόρηση και το γεγονός ότι το web παρέχει εύκολη πρόσβαση σε μεγάλο όγκο πληροφοριών, το πρόβλημα δεν είναι πλέον να βρεθούν δίοδοι πρόσβασης προς ακόμα περισσότερες πληροφορίες, αλλά ότι όλο και περισσότεροι άνθρωποι με διαφορετικές ανάγκες και προτιμήσεις πλοηγούνται μέσα σε εκτενείς και πολύπλοκες δομές χάνοντας σε πολλές περιπτώσεις το στόχο της αναζήτησης τους.

Εδώ έρχεται η εξατομίκευση να αποτελέσει λύση στο πρόβλημα αυτό του πληροφοριακού υπερφόρτου. Καθώς εξ ορισμού στόχος της είναι να παρέχεται στους χρήστες αυτό που θέλουν ή χρειάζονται , χωρίς να πρέπει να το ζητήσουν ρητά ή να το αναζητήσουν οι ίδιοι.

Επομένως, καθώς το web μεγαλώνει εκθετικά και εξαπλώνεται η χρήση του , όλο και περισσότερες εφαρμογές (ηλεκτρονικά καταστήματα, πληροφοριακές πύλες, συστήματα εκπαίδευσης από απόσταση κ.α.) ενσωματώνουν υπηρεσίες εξατομίκευσης, με σκοπό να βελτιώσουν την απόδοσή τους και να ανταποκριθούν σε ατομικές ανάγκες και προτιμήσεις, να αυξήσουν την ικανοποίηση των χρηστών τους , να αποκτήσουν πιστούς πελάτες, και να αναπτύξουν σχέσεις έναν προς έναν.

Στην εργασία αυτή περιηγηθήκαμε στις νέες προοπτικές που προσφέρει η εξατομίκευση στο χώρο των συστημάτων ηλεκτρονικής μάθησης (e- learning),

ηλεκτρονικού εμπορίου (e commerce) , και επιχειρηματικής ευφυΐας (business intelligence). Όπως επίσης και σε πολλές από τις εφαρμογές και τις υπηρεσίες που ενσωματώνουν τεχνικές εξατομίκευσης. Περιγράψαμε τεχνολογίες και δεδομένα που χρησιμοποιούνται από τέτοιες υπηρεσίες , παρουσιάσαμε τον τρόπο παροχής δυναμικών υπηρεσιών εξατομίκευσης χρησιμοποιώντας τεχνικές web mining και αναφερθήκαμε σε μελλοντικές τάσεις και σε προβλήματα και περιορισμούς που εμποδίζουν ή καθυστερούν την εξάπλωση και την εξέλιξη της.

Αναφέραμε μέσα από παραδείγματα τρόπους με τους οποίους τα συστήματα μπορούν να εκμεταλλευτούν τις πληροφορίες που αποθηκεύονται στα προφίλ των χρηστών εφαρμόζοντας τεχνικές εξόρυξης ιστού με στόχο την ανάλυση και την κατανόηση της πλοηγικής συμπεριφοράς τους.

Είδαμε μέσα από τον απλό ιστό πληροφοριών, με την ενσωμάτωση ενός ακόμη επιπέδου ευφυΐας και την πρόσθεση νοήματος στις πληροφορίες , πως δημιουργείται ο σημασιολογικός παγκόσμιος ιστός πληροφοριών. Και ακόμη εξετάσαμε την εφαρμογή της εξατομίκευσης ,δοκιμάζοντάς την στα e-learning συστήματα. Εν ολίγοις αναπτύξαμε την περιοχή των τεχνολογιών εξατομίκευσης και τις νέες προοπτικές που διαμορφώνονται για την ανάπτυξη μακροπρόθεσμων σχέσεων εμπιστοσύνης και ικανοποίησης ανάμεσα στα ηλεκτρονικά συστήματα και τους χρήστες τους.

Μελετήσαμε λοιπόν στα κεφάλαια της πτυχιακής πολλές σύγχρονες θεωρίες οι οποίες είχαν να κάνουν κυρίως με τη δυναμική που πρέπει να έχει το ίντερνετ, προκειμένου ο κάθε χρήστης να νιώθει άνετα κατά την πλοήγησή του σε αυτό. Οι σύγχρονες θεωρίες που έχουν να κάνουν με την οντολογία ή με την προσωποποίηση του ίντερνετ, αλλά και με όλες τις αλληλεπιδραστικές (interactive) λειτουργίες του είναι απαραίτητες πια στην εποχή μας διότι μέσω αυτών ο χρήστης νιώθει καλύτερα κατά την πλοήγησή του , αισθάνεται πιο ασφαλής και γενικά μειώνονται οι όποιες ανησυχίες που γεννιούνται κατά το διάστημα της παρουσίας του σε κάποια συγκεκριμένη ιστοσελίδα.

Πρέπει να γίνει κατανοητό ότι το διαδίκτυο στην πολύχρονη ιστορία την οποία έχει, μπόρεσε να μετασχηματιστεί από ένα ερευνητικό πρόγραμμα που αναφερόταν μόνο σε χρήσεις της Ακαδημαϊκής κοινότητας της Αμερικής σε ένα παγκόσμιο επικοινωνιακό εργαλείο το οποίο αλλάζει τα επίπεδα της επικοινωνίας και της ψυχαγωγίας του σύγχρονου ανθρώπου. Η μετάβασή του αυτή η οποία δεν μπορεί να θεωρηθεί ως προσχεδιασμένη πρέπει σήμερα να αναγνωριστεί από όλους τους φορείς , κυβερνήσεις, εταιρίες και άλλους οργανισμούς κερδοσκοπικούς και μη κερδοσκοπικούς , προκειμένου να αναπτύξουν δράσεις τέτοιες ώστε ο χρήστης να πλοηγείται πιο αποτελεσματικά στις διάφορες ιστοσελίδες. Ο ενθουσιασμός και η δημιουργικότητα των πολιτών του δικτύου θα πρέπει να προωθείται ώστε πια ο χρήστης να νιώθει ότι δεν ελέγχεται από το ίντερνετ αλλά το ελέγχει. Όταν μια ιστοσελίδα διατηρεί αρχείο με τα προσωπικά δεδομένα του χρήστη, με τις επιθυμίες του , τις προτιμήσεις του και γενικότερα με τις όποιες απαιτήσεις έχει κάθε φορά που έμπαινε σε μια ιστοσελίδα τότε αυτόματα μειώνεται ο άσκοπος χρόνος πλοήγησης , αυξάνεται η ικανοποίηση , ο χρήστης γίνεται πιο πιστός προς το διαδίκτυο. Το ίντερνετ σήμερα έχει ανάγκη αυτή τη πιστότητα μιας και είναι πολλοί αυτοί που το κρίνουν αρνητικά οπότε και αυτό μειώνει την εμπιστοσύνη των απλών χρηστών.

Οι σύγχρονες όμως τεχνολογικές λύσεις που έχουν αναπτυχθεί στο ίντερνετ οι οποίες αυτές μεταφράζονται σε καινοτόμες νέες υπηρεσίες από το απλό ηλεκτρονικό ταχυδρομείο (e-mail) μέχρι και τη φυλλομέτρηση, την εξατομίκευση και την προσωποποίηση επέφεραν την αλλαγή τοποθετώντας σήμερα το ίντερνετ σε ένα από τα βασικά εργαλεία του χρήστη για κάθε είδους χρήστη, όποιο κι να είναι το αντικείμενό του αλλά και όπου και να βρίσκεται τη στιγμή που επιλέγει να μπει σε μια ιστοσελίδα.

Όσοι ελέγχουν τη χρήση και τη δυναμική του ίντερνετ σήμερα θα πρέπει να καταλάβουν ότι περιορίζοντας τους ανασφαλούς τρόπους πλοήγησης αυξάνοντας τις υπηρεσίες που παρέχονται , στηρίζοντας τις δράσεις όπου και αυτές να αναφέρονται, επενδύοντας σε υπηρεσίες που δίνουν προσωπικότητα στο διαδίκτυο, καταφέρνουν να το

κάνουν ακόμα πιο ισχυρό , πιο λειτουργικό και πιο αποτελεσματικό ικανοποιώντας τις επιθυμίες και τις ανάγκες όλων των χρηστών.

Το ίντερνετ δεν πρέπει να λειτουργεί μόνο για κάποιες ομάδες, δεν πρέπει να περιθωριοποιεί, δεν πρέπει να τρομάζει, πρέπει να είναι όργανο της νέας οικονομίας , αλλά και της εκπαίδευσης χωρίς σύνορα, μεταφέροντας τη γνώση σε όλες τις χώρες του κόσμου, ακόμα και τις πιο απομακρυσμένες.

Κλείνοντας θα πρέπει να πούμε ότι όλες οι δράσεις που καταγράφηκαν στη συγκεκριμένη εργασία μπορούν πραγματικά να βοηθήσουν το ίντερνετ, να αποκτήσει τα χαρακτηριστικά που προαναφέραμε και να γίνει ένας τόπος συγκέντρωσης αλλά και ένα προσωπικό εργαλείο χρήσης για όλους τους κατοίκους όλων των χωρών του πλανήτη μας.

ΚΕΦΑΛΑΙΟ 7

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Anctil E., Beheshti J., (2004), "Open Source Integrated Library Systems", McGill University, GSLIS, 2004
2. Barreto C.M.,(2002), A survey on metadata for describing and retrieving Internet recourses, *World Wide Web* 1, pp.221-24
3. Barrett, N.,(1998), *The State of the Cybernation*. London: Kogan Page
4. Barron, B. B. (2002). "Distant and Distributed Learners are Two Sides of the Same Coin". *Computers in Libraries*, vol. 22, no. 1 (January 2002), pp. 24-28
5. Berners-Lee, T. (2006), *Weaving the Web: The Past, Present and Future of the World Wide Web by its Inventor*. UK: Orion Business Books.
6. Berners-Lee, T., Hendler, J., & Lassila, O., (2001). *The Semantic Web*, *Scientific American*, 285,5,pp. 34-43.
7. Blom, J.,(2000). *Personalization - A Taxonomy*. *Proceedings of the CHI 2000 Workshop on Designing Interactive Systems for 1-to-1 Ecommerce*. New York, ACM Press, 313-314
8. Brusilovsky, P. (1996). *Methods and Techniques of Adaptive Hypermedia*. *User Modeling and User-Adapted Interaction*, 6, 2/3, 87-12
9. Clarke, T and Pitelis, C.,(1995),*The Political Economy of Privatization*" London and New York: Routledge,pp.100-102
10. Clarke, T.,(1994),*International Privatisation: Strategies and Practices* Berlin and New York: Walter de Gruyter
11. Cooley et al., (1997), *Mathematics Subject Classification*, Springer Berlin Heidelberg

12. De Bra and Calvi L.,(2000), Towards a Generic Adaptive Hypermedia System. Proceedings of the Second Workshop on Adaptive Hypertext and Hypermedia, pp. 5-11
13. Decker S.,(2002),The Semantic Web- on the respective Roles of XML and RDF., <http://www.ontoknowledge.org/oil/download/IEEE00.pdf>
14. Dogac A., (2004), "Semantically Enriched Web Services for the Travel Industry", September 2004 issue of Sigmod Record. (Also available at <http://www.srdc.metu.edu.tr/webpage/projects/satine/publications/Dogac.pdf>)
15. Dolog et al., (2004), Electronic Communication on Diverse Data- The Role of an Object-Oriented CIDOC Reference Model, CIDOC'98 Conference, Melbourne, Australia
16. Erdmann M& Studer R.,(1999), Ontologies as Conceptual Models for XMLDocuments, Proceedings of the 12th Workshop of Knowledge Acquisition Modeling and Managing (KAW' 99), Bauff, Canada
17. Fensel D.,(2004), Adding Semantics to the Web,<http://www.ontoknowledge.org>
18. Fensel, D., Editor (2002). Spinning the Semantic Web: Bringing the World Wide Web to its Full Potential. The MIT Press.
19. Germà B.,(2006), "The coining of `privatization´and Germany's National Socialist Party", Journal of Economic Perspectives, 20,3. 187-194
20. Gotschall, M. (2000). E-learning strategies for executive education and corporate training. Fortune, 141(10), S5 S59.
21. Guarino G.,(1998), An Environment for Merging and Testing Large Ontologies. In Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR-2000), Breckenridge, Colorado,. San Francisco: Morgan Kaufman, pp.14-18
22. Hand et al.,(2001), Customer Oriented Internet Personalized, Industrial Management and Data System, Vo. 105, Issue.1, pp.26-44
23. Kevin R.,(2006),Privatization and the Federal Government: An Introduction", Report from the Congressional Research Service, pp.24-28

24. Kleinberg G.,(1999), Sharing information in web communities,EHESS
25. Kondo, H.,(2004), An Implementation of an intelligent tutoring system (ITS) on the World-Wide Web (WWW). Educational Technology Research 19, 1,pp. 35-44
26. Kosala G & Blockeel P.,(2000), Intelligent agents for web-based tasks: An advice-taking approach, Scientific American
27. Luotonen A.,(1994),The World-Wide-Web. Communications of the ACM, 37,8 :76-82
28. Luotonen G.,(1995), The World-Wide-Web. Communications of the ACM, 37,8, 76-82, August 1994.
29. Lytras M. (2005) “Semantic Web and Information Systems: An Agenda Based on Discourse with Community Leaders”, in International Journal on Semantic Web and Information Systems, Inaugural Issue, 1(1), pp: i-xii
30. Meyen, E.L. (2000). Using technology to move research to practice: The Online Academy. Their World 2000. New York: National Center for Learning Disabilities.
31. Missikoff, M. et al (2003). Harmonise-Towards Interoperability in the Domain. In Proc. of Information and Communication Technologies 2003, Helsinki: Springer, pp.58-66
32. Mizoguchi et al.,(1995), The Semantic Web, Scientific American, 285, 5, pp.34-43
33. Neumann, G. and Zirvas, J.: 1998, SKILL - A scallable internet-based teaching and learning system. Proceedings of WebNet'98, World Conference of the WWW, Internet, and Intranet, Orlando, FL, pp. 688-693
34. Nilson M.,(2001), The Semantic Web: How RDF will change learning technology *standards*, Center for User-Oriented IT-design, Royal Institute of Technology, Stockholm, online at: <http://cid.nada.kth.se/pdf/CID-157.pdf>

35. Papanikolaou K., (2003) Personalising the interaction in a Web-based Educational Hypermedia System: the case of INSPIRE. User Modeling and User Adapted Interaction 13,3, 213-267.

36. Pearce, M(1997), The Interactive Book of Internet, Macmillan Technical Publishing, Indianapolis.

37. Presno C (2000),Taking the byte out of Internet Anxiety: Instructional techniques that reduce computer / internet anxiety in the classroom, Journal of Educational Computing Research, Vol. 18(2), 147-161

38. Prettejohn, M.(1996). The First Year: August 1995-August 1996.Netcraft

39. Reid, R.(2004),Architects of The Web:1000 Days that Build The Future of Business, New York: John Wiley

40. Roberts, L.(1997),Multiple Computer Networks and InterComputer Communication”, in Proceedings of ACM Symposium on Operating System Principles Gatlinburg, Tennessee.

41. Seidner R and N. Tindall.,(2003), Interactive debug requirements. In M.S. Johnson, editor, Proceedings of the Software Engineering Symposium on High-Level Debugging, p.p 9-22

42. Srivastava, J., Cooley , R., Deshpande , M., & Tan , P.-N. (2000). Web usage Mining: Discovery and Applications of Usage Patterns form Web Data.1(2), 12-23.

43. Stratakis et al(2003),Ontology Infrastructure, IEEE INTELLIGENT SYSTEMS

44. Sullivan, D.(2003), Searches per Day, <http://searchenginewatch.com/reports/article.php/2156461>

45. Urdan, T. A., & Weggen C., (2000). Corporate e-learning: Exploring a new frontier. WR Hambrecht + Co. Ανακτήθηκε στις 30-10-2008 απο<http://www.wrhambrecht.com/research/coverage/elearning/ir/ir>

46. Van Heijst et al., (2003), OWL Web Ontology Language Overview, <http://www.w3.org/TR/owl-features/>

47. W3C.,(2005),Το W3C Αναπτύσσει Πρότυπα και Οδηγίες του Παγκοσμίου Ιστού, Ανάκτηση στις 30-10-2008 από <http://www.w3c.gr>
48. Weber, G. (1999) Adaptive learning systems in the World Wide Web. In: J. Kay (ed.) (Proceedings of 7th International Conference on User Modeling, UM99, Banff, Canada, June 20-24, 1999) Wien: SpringerWienNewYork, pp. 371-378
49. Wentling, T.L et al (2000). e-learning-A review of literature. Ανακτήθηκε στις 30-10-2008 από <http://learning.ncsa.uiuc.edu/papers/elearnlit.pdf>
50. Yahoo (2007), <http://docs.yahoo.com/info/pr/milestones.html> (Last Available 1-12-07)
51. Winter Ken.,(1999), Mylibrary can help your library”. American Libraries, August,pp. 65 -67 p
52. Αλεξάκη Σ.,(2002) Αποθήκευση Μεταδεδομένων RDF για Πύλες Κοινοτήτων Διαδικτύου, , Πανεπιστήμιο Κρήτης
53. Βασίλης Τ(2007),Σημασιολογικός Ιστός, Multi-Mine
54. Βιβλιοθήκη Κολεγίου Αθηνών(2007),Μετά Μηχανές Αναζήτησης, <http://www.haef.gr/gre/libraries/websearch/index.html> (Ανακτήθηκε 11-2-08)
55. ΕΚΠΑ(2006),Ηλεκτρονικές Βιβλιοθήκες, <http://www.lib.uoa.gr/engines>, (Ανακτήθηκε 1-12-07)
56. Εργαστήριο Εφαρμογών Πληροφορικής στα ΜΜΕ(2007), Οι Μεταμηχανές Αναζήτησης, http://pacific.jour.auth.gr/online_learning.htm(Ανακτήθηκε 12-2-08)
57. Λέανδρος, Ν.(2005), Το διαδίκτυο Ανάπτυξη και Αλλαγή, Αθήνα: Εκδόσεις Καστανιώτη
58. Μάρκελλος Κ.,(2006), Εξατομίκευση Ηλεκτρονική Μάθηση, Πανεπιστήμιο Πατρών, σελ 1-4
59. Παπαδάκης, Κ(2001),Στρατηγική των Επιχειρήσεων, Αθήνα, Εκδόσεις Μπένοσ
60. Denise A. Roberts.,(2003),User Interfaces and personalization, June 4, 2003 INFO653 Digital libraries Spring

61. Winter Ken.,(1999),“Mylibrary can help your library, American Libraries, August, pp. 65 -67
62. James C. French.,(1999), Personalized Information environments-Lib Magazine
63. Anne G.,(2001),The Headline Personal Information Environment” D-Lib Magazine March, Volume 7 Number 3
64. Austen, G. (1998). What is my core business, Ανάκτηση στις 15-10-2008 απο <http://www.lib.pku.edu.cn/98conf/paper/a/GaynorAusten.htm>

ΚΕΦΑΛΑΙΟ 8

ΠΑΡΑΡΤΗΜΑ

ΛΙΣΤΑ ΣΧΗΜΑΤΩΝ

Σχήμα 1. Η off- line φάση μιας διαδικασίας εξατομίκευσης βασισμένης σε δεδομένα χρήσης web που χρησιμοποιεί για ανακάλυψη προτύπων την τεχνική της εξαγωγής κανόνων συσχέτισης.	17
Σχήμα 2. Η on- line φάση μιας διαδικασίας εξατομίκευσης βασισμένης σε δεδομένα χρήσης web, που στοχεύει στον υπολογισμό και την αποστολή εξατομικευμένων συστάσεων.	19
Σχήμα 3. Σημασιολογικός παγκόσμιος ιστός πληροφοριών.	23
Σχήμα 4 Τα επίπεδα της αρχιτεκτονικής του Σημασιολογικού Παγκόσμιου Ιστού Πληροφοριών.	25
Σχήμα 5. Ταξινόμηση των οντολογιών σύμφωνα με τον Guarino.	32
Σχήμα 6. Οι κατηγορίες του Semantic Web Mining	34
Σχήμα 7. Συσχέτιση προτύπων E- learning μεταδεδομένων	48
Σχήμα 8. Web Mining (a)	53
Σχήμα 9. Web Mining (b)	55
Σχήμα 10. Το εννοιολογικό μοντέλο του Learner	56
Σχήμα 11. Μέθοδοι Εξατομίκευσης	70

ΛΙΣΤΑ ΠΙΝΑΚΩΝ

Πίνακας 1.	Έτος έναρξης λειτουργίας των σημαντικότερων εργαλείων αναζήτησης και οργανισμοί απ' τους οποίους ξεκίνησε η προσπάθεια.	89
Πίνακας 2.	Στατιστικά στοιχεία για τα εργαλεία αναζήτησης (Ιανουάριος 2003)	92
Πίνακας 3.	Μεταμηχανές αναζήτησης	117

ΛΙΣΤΑ ΕΙΚΟΝΩΝ

Εικόνα 1.	My Library	81
Εικόνα 2.	Το προφίλ του Yahoo	96
Εικόνα 3.1	Σύνθετη αναζήτηση στο Google	103
Εικόνα 3.2	Το προφίλ του Google	106
Εικόνα 4.	Προφίλ του MSN	111
Εικόνα 5.	Προφίλ του AOL	115
Εικόνα 6.	Το πρόγραμμα Κοπέρνικος	120
Εικόνα 7.1	Αρχική σελίδα του Amazon	130
Εικόνα 7.2	Φόρμα εγγραφής κατά την είσοδο σε χρήστη που έχει ξανά επισκεφτεί το κατάστημα Amazon	131
Εικόνα 7.3	Φόρμα εγγραφής κατά την πρώτη είσοδο στο κατάστημα του Amazon	131

Εικόνα	7.4 Φόρμα συμπλήρωσης προϊόντων στο «καλάθι αγορών» του Amazon	133
Εικόνα	7.5 Φόρμα συμπλήρωσης παραγγελιοδοσίας στο Amazon	134
Εικόνα	7.5.1 Φόρμα συμπλήρωσης παραγγελιοδοσίας στο Amazon	134

ΓΛΩΣΣΑΡΙ

BROWSER: φυλλομετρητής ή διαφυλλιστής. Λογισμικό που εγκαθίσταται στον υπολογιστή του χρήστη(client side) και χρησιμοποιείται κατά την πλοήγηση στο web. Επικοινωνεί με τους web servers και εμφανίζει τα περιεχόμενα των ιστοσελίδων (κείμενα , γραφικά, video, κλπ.) στην οθόνη του χρήστη.

SERVER: Εξυπηρετητής ή διακομιστής. Το σύστημα που διαχειρίζεται και διαμοιράζει τα δεδομένα στους πελάτες . Οι servers και οι clients είναι συνήθως συνδεδεμένοι μέσω ενός δικτύου. Ο web server είναι ένας υπολογιστής που διαθέτει κατάλληλο λογισμικό ώστε να φιλοξενεί και να διαμοιράζεται με τους πελάτες web σελίδες και web εφαρμογές.

PROXY SERVER: Εξυπηρετητές που χρησιμοποιούνται για την μείωση του κυκλοφοριακού φόρτου στο δίκτυο και του φόρτου των web servers. Τοποθετούνται σε κομβικά σημεία του δικτύου , αποθηκεύοντας τοπικά σελίδες και αναλαμβάνουν να εξυπηρετούν οι ίδιοι τις αιτήσεις για σελίδες που διαθέτουν.

HTML: Hypertext Markup Language. Η γλώσσα συγγραφής web σελίδων , ώστε να μπορούν να διαβαστούν από τους φυλλομετρητές (browsers).

HTTP: Hypertext Transfer Protocol. Πρωτόκολλο μεταφοράς υπερκειμένου που ορίζει πως ένα html αρχείο μεταφέρεται από τον εξυπηρετητή server στον πελάτη (client) πάνω από τον παγκόσμιο ιστό.

HYPERTEXT: Υπερκείμενο. Οποιοδήποτε κείμενο περιέχει συνδέσμους που παραπέμπουν αλλού.

URL: Universal Resource Locator. Η διεύθυνση ενός αρχείου στο internet. Η μοναδική διεύθυνση ιστοσελίδων και άλλων πόρων στο web.

INTERNET: Διαδίκτυο. Τεράστιος αριθμός διασυνδεδεμένων υπολογιστών και δικτύων που επικοινωνούν με το πρωτόκολλο TCP/IP.

NETWORK: Δίκτυο. Η διασύνδεση και επικοινωνία δυο ή περισσότερων υπολογιστών.

TCP: Transmission Control Protocol. Πρωτόκολλο ελέγχου επικοινωνίας .

IP: Internet Protocol. Πρωτόκολλο πρόσβασης στο internet, ο μοναδικός για κάθε υπολογιστή αριθμός αναγνώρισής του.

TCP/IP: Το σύνολο των κανόνων με τους οποίους επικοινωνούν οι υπολογιστές στο internet.

RDF: Resource Description Framework. Είναι ένα μοντέλο δεδομένων που προσθέτει σημασιολογία στα δεδομένα ,τα οποία περιγράφει μέσω δηλώσεων (statements) . Αυτές οι δηλώσεις είναι κατανοήσιμες από τον υπολογιστή και οι μηχανές αναζήτησης μπορούν να χρησιμοποιήσουν αυτή την πληροφορία. Μπορεί η σύνταξη του να παρέχεται από την XML, χωρίς όμως αυτή να είναι συστατικό της RDF.

XML: eXtensible Markup Language. Η XML σχεδιάστηκε να ικανοποιήσει πολλές ανάγκες δίνοντας στα έγγραφα ένα μεγαλύτερο επίπεδο προσαρμοστικότητας στο στυλ και τη δομή από αυτό που υπήρχε παλαιότερα στην HTML. Η XML προσφέρει στους σχεδιαστές της HTML τη δυνατότητα να προσθέτουν περισσότερα στοιχεία στη γλώσσα. Δεν αναφέρεται μονάχα στους σχεδιαστές του web αλλά σε οποιονδήποτε ασχολείται με εκδόσεις. Στην πραγματικότητα, η XML είναι markup γλώσσα για έγγραφα που περιέχουν δομημένες

πληροφορίες. Markup γλώσσα είναι ένας μηχανισμός που καθορίζει δομές σε ένα έγγραφο. Η XML είναι κάτι περισσότερο από markup language είναι metalanguage, δηλαδή μια γλώσσα που χρησιμοποιείται για να καθορίσει νέες markup γλώσσες.

WEB: Παγκόσμιος ιστός πληροφοριών, World Wide Web (www). Η πιο δημοφιλής υπηρεσία του Διαδικτύου λόγω της εύκολης χρήσης και του πλήθους των πληροφοριών που παρέχει. Τα δεδομένα που είναι διαθέσιμα στο web, έχουν την μορφή ιστοσελίδων (web pages), που είναι αποθηκευμένες σε ένα σύνολο διασυνδεδεμένων εξυπηρετητών (web servers). Οι χρήστες κατά την πλοήγησή τους χρησιμοποιούν ένα πρόγραμμα φυλλομετρητή (web browser). Ο browser αναλαμβάνει να εντοπίσει τον εξυπηρετητή που διαθέτει την ιστοσελίδα που ζήτησε να δει ο χρήστης και να τη εμφανίσει με κατάλληλο τρόπο στην οθόνη του υπολογιστή του.

CLICKSTREAM: Μια ακολουθιακή σειρά από αιτήσεις για pageviews.

PAGEVIEW: Ένα σύνολο από αρχεία (ιστο) σελίδων που συνεισφέρουν στο συνολικό περιεχόμενο του παραθύρου του browser, όπως εμφανίζεται σε ένα συγκεκριμένο client, μια δεδομένη χρονική στιγμή.

SEMANTIC WEB: Βλέπε 'Σημασιολογικός Ιστός'.

WEB MINING: Η χρήση τεχνικών ανάκτησης δεδομένων (data mining) για την ανακάλυψη και εξαγωγή πληροφοριών από έγγραφα και υπηρεσίες web. Διακρίνεται –ανάλογα με το κομμάτι του web που εξετάζει- σε web content mining, web structure mining και web usage mining.

ΕΞΑΤΟΜΙΚΕΥΣΗ (1): Οποιαδήποτε ενέργεια που προσαρμόζει τις πληροφορίες ή τις υπηρεσίες που παρέχονται από ένα website στη γνώση που αποκτιέται από την πλοηγική συμπεριφορά και τα ατομικά ενδιαφέροντα των χρηστών του, σε συνδυασμό με το περιεχόμενο και τη δομή του website.

ΕΞΑΤΟΜΙΚΕΥΣΗ (2): Η διαδικασία συγκέντρωσης και αποθήκευσης πληροφοριών αναφορικά με τους χρήστες ενός website, η ανάλυση των πληροφοριών αυτών και, με βάση την ανάλυση, η αποστολή σε κάθε χρήστη της σωστής πληροφορίας στο σωστό χρόνο.

ΣΗΜΑΣΙΟΛΟΓΙΚΟΣ ΙΣΤΟΣ: Ξεκίνησε σαν ένα project με σκοπό να δημιουργηθεί ένα παγκόσμιο μέσο για ανταλλαγή πληροφοριών αποδίδοντας σημασία (semantics) στα έγγραφα του web, με ένα τρόπο που να είναι κατανοητός από τις μηχανές. Στις μέρες μας και υπό την καθοδήγηση του εμπνευστή του Tim Berners-Lee και του World Wide Web Consortium, ο σημασιολογικός ιστός επεκτείνει τις δυνατότητες του web με ένα νέο σύνολο από πρότυπα, γλώσσες σήμανσης και εργαλεία επεξεργασίας.

ΣΥΝΟΛΟΣ SERVER (server session ή επίσκεψη): Μια συλλογή από clicks χρηστών προς ένα συγκεκριμένο server κατά τη διάρκεια μιας συνόδου χρήστη.

ΣΥΝΟΛΟΣ ΧΡΗΣΤΗ (user session): Ένα πεπερασμένο σύνολο από clicks ενός χρήστη (clickstream) προς έναν ή περισσότερους web servers.