



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΜΕΘΟΔΟΛΟΓΙΑΣ, ΙΣΤΟΡΙΑΣ ΚΑΙ ΘΕΩΡΙΑΣ ΤΗΣ ΕΠΙΣΤΗΜΗΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ
ΤΜΗΜΑ ΦΙΛΟΣΟΦΙΑΣ, ΠΑΙΔΑΓΩΓΙΚΗΣ ΚΑΙ ΨΥΧΟΛΟΓΙΑΣ

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΒΑΣΙΚΗ ΚΑΙ ΕΦΑΡΜΟΣΜΕΝΗ ΓΝΩΣΙΑΚΗ ΕΠΙΣΤΗΜΗ»

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Γνωσιακή Αναζήτηση Πληροφορίας
στον Παγκόσμιο Ιστό:
Μια σύγκριση με άλλες προσεγγίσεις

Βερονίκη Α. Σταμάτη

A.M. 07M19

Επιβλέπουσες Καθηγήτριες
Καράλη Ιζαμπώ
Γρηγοριάδου Μαρία

ΑΘΗΝΑ 2010

Περίληψη

Στην παρούσα διπλωματική συγκρίνονται τρεις προσεγγίσεις στην αναζήτηση πληροφορίας στον Παγκόσμιο Ιστό. Αυτές οι προσεγγίσεις είναι η παραδοσιακή, η σημασιολογική και η γνωσιακή. Οι δύο πρώτες είναι γνωστές πρακτικές στον επιστημονικό χώρο, ενώ η τρίτη είναι σχετικά καινούρια και είναι προϊόν σύγκλισης ανάμεσα στους χώρους της Γνωσιακής Επιστήμης και της Ανάκτησης Πληροφορίας. Σ' αυτό λοιπόν το πλαίσιο ορίζεται η γνωσιακή προσέγγιση, και παρουσιάζονται κι αξιολογούνται συστήματα αναζήτησης που εμπίπτουν στη λογική αυτής της προσέγγισης. Τέλος, γίνεται μία πρόταση για ένα γνωσιακό σύστημα αναζήτησης.

Ευχαριστίες

Καταρχήν οι ευχαριστίες μου δε θα μπορούσαν παρά ν' απευθύνονται στις επιβλέπουσες καθηγήτριές μου, κ. Ιζαμπώ Καράλη και κ. Μαρία Γρηγοριάδου, που με την αδιάλειπτη καθοδήγησή τους και με τις εύστοχες κριτικές τους με βοήθησαν να φέρω στην τωρινή της μορφή την παρούσα εργασία. Θα ήθελα επίσης να ευχαριστήσω τους διευθυντές μου στο Αμερικάνικο Κολλέγιο Ελλάδος, οι οποίοι επί τρία συναπτά έτη διευκόλυναν το ωράριό μου για να μπορέσω να παρακολουθήσω και να ολοκληρώσω το μεταπτυχιακό της Βασικής & Εφαρμοσμένης Γνωσιακής Επιστήμης, επιστέγασμα του οποίου είναι η διπλωματική μου εργασία. Την αγάπη μου έχουν όπως πάντα η οικογένειά μου και οι φίλοι μου, που για έναν ολόκληρο χρόνο υπέμειναν αγόγγυστα τις μακροσκελείς διηγήσεις μου για την πορεία της διπλωματικής μου εργασίας.

Περιεχόμενα

Περίληψη.....	2
Ευχαριστίες.....	3
Περιεχόμενα.....	4
Κεφάλαιο 1: Εισαγωγή.....	6
1.1 Παραδοσιακή αναζήτηση πληροφορίας.....	8
1.2 Σημασιολογική αναζήτηση πληροφορίας.....	9
1.3 Γνωσιακή αναζήτηση πληροφορίας.....	10
Κεφάλαιο 2: Οι κυριότερες τάσεις στη γνωσιακή αναζήτηση πληροφορίας.....	13
2.1 Αναζήτηση εντός πλαισίου.....	14
2.1.1 Σύστημα σημασιολογικής αναζήτησης βασισμένο σε γνωσιακό μοντέλο....	14
2.1.2 Contextual Search Project.....	16
2.2 Εξατομικευμένη αναζήτηση πληροφορίας.....	16
2.2.1 Outride.....	18
2.2.2 Search-In-Synchrony.....	19
2.2.3 OntoSearch_U.....	20
2.3 Έξυπνοι πράκτορες βασισμένοι σε γνωσιακά μοντέλα.....	21
2.3.1 Cognitive web agents.....	22
2.3.2 RASCALLI.....	23
2.4 Συνεργατική αναζήτηση πληροφορίας.....	24
2.4.1 Cobot.....	25
2.5 Γνωσιακές διεπαφές.....	25
2.5.1 Ταξινομητής διαστάσεων.....	26
2.5.2 ConSearch.....	27
2.5.3 Ομαδοποιητής ιστοσελίδων βάσει σχετικότητας.....	27
2.5.4 Media Watch on Climate Change.....	28
2.6 Από την πληροφορία στη γνώση.....	29
2.6.1 Met.a.ware.....	29
2.7 Γνωσιακά vs Σημασιολογικών Συστημάτων Αναζήτησης.....	30
Κεφάλαιο 3: Αποτίμηση των Παραπάνω Τάσεων στο Γνωσιακό Χώρο.....	31
3.1 Μια Θεωρία της Σχετικότητας.....	31

3.2	Αναζήτηση πληροφορίας σε πολυμεσική μορφή.....	33
3.3	Μετρικές για την Αξιολόγηση των Συστημάτων Αναζήτησης.....	34
3.4	Αξιολόγηση των γνωσιακών συστημάτων αναζήτησης που παρουσιάστηκαν .	35
3.4.1	Σύστημα σημασιολογικής αναζήτησης βασισμένο σε γνωσιακό μοντέλο....	35
3.4.2	Contextual Search Project.....	36
3.4.3	Outride.....	36
3.4.4	Search-in-Synchrony.....	37
3.4.5	OntoSearch_U.....	37
3.4.6	Cognitive web agents.....	37
3.4.7	RASCALLI.....	38
3.4.8	Cobot.....	38
3.4.9	Ταξινομητής διαστάσεων.....	38
3.4.10	ConSearch.....	39
3.4.11	Ομαδοποιητής ιστοσελίδων βάσει σχετικότητας.....	39
3.4.12	Media Watch on Climate Change.....	39
3.4.13	Met.a.ware.....	40
	Κεφάλαιο 4: Μια Πρόταση για τη Γνωσιακή Αναζήτηση Πληροφορίας.....	41
4.1	Μια διαφορετική προσέγγιση.....	41
4.2	Η πρότασή μου.....	42
	Κεφάλαιο 5: Συμπεράσματα.....	47
	Βιβλιογραφία.....	50

Γνωσιακή Αναζήτηση Πληροφορίας στον Παγκόσμιο Ιστό:

Μια σύγκριση με άλλες προσεγγίσεις

Κεφάλαιο 1: Εισαγωγή

Η αναζήτηση πληροφοριών στον Παγκόσμιο Ιστό (World Wide Web search) είναι ένα από τα σημαντικότερα προβλήματα του και γι' αυτόν το λόγο προσελκύει το ενδιαφέρον των ερευνητών και των εταιρειών που παράγουν διάφορες εμπορικές εφαρμογές αναζήτησης. Εταιρείες όπως η Google γιγαντώθηκαν κι έφτασαν να κατέχουν σημαίνοντα εμπορικό ρόλο στο χώρο της Πληροφορικής ξεκινώντας από την υλοποίηση μηχανών αναζήτησης πληροφοριών στον Παγκόσμιο Ιστό (World Wide Web – WWW). Χωρίς να παραβλέπουμε την (απόλυτα λογική με εμπορικούς όρους) προσδοκία των διάφορων εταιρειών πληροφορικής να γίνουν «χαλίφηδες στη θέση του χαλίφη» (όπου ο χαλίφης είναι η Google), η αδιάκοπη προσπάθεια ανάπτυξης καλύτερων συστημάτων αναζήτησης αντικατοπτρίζει μια πραγματική ανάγκη για τη βελτίωση των υπαρχόντων, πράγμα που αποδεικνύεται από τις εμπειρίες κάθε χρήστη του Διαδικτύου· οποιοσδήποτε έχει προσπαθήσει να βρει πληροφορίες στον WWW, έχει τουλάχιστον μία φορά εγκαταλείψει απελπισμένος και πονοκεφαλιασμένος την τελικά άκαρπη προσπάθειά του, είτε λόγω μηδενικών ή ελάχιστων αποτελεσμάτων, είτε λόγω υπερβολικά πολλών (π.χ. μερικές χιλιάδες) και άσχετων αποτελεσμάτων, τα οποία μάλιστα εμφανίζονται πρώτα-πρώτα στη λίστα! Εξίσου μεγάλη πρόκληση είναι για τους μηχανικούς λογισμικού η κατασκευή ενός βέλτιστου συστήματος αναζήτησης, καθώς η άναρχη δομή και το πολυποίκιλο περιεχόμενο του Παγκόσμιου Ιστού τον καθιστούν έναν τελείως διαφορετικό χώρο αναζήτησης από τον κατάλογο μιας μηχανογραφημένης βιβλιοθήκης ή έστω, μιας οποιασδήποτε βάσης δεδομένων (Diaz, 2005, σ. 57; Knight & Spink, 2008). Καταλαβαίνει λοιπόν κανείς ότι το συγκεκριμένο πεδίο με τις ιδιομορφίες του είναι πρόσφορο για έρευνα και για εμπορική εκμετάλλευση, τόσο λόγω της ευρείας διείσδυσής του στους χρήστες όσο και λόγω της βελτίωσης που επιδέχονται οι υπάρχουσες λύσεις.

Η έρευνα για την αναζήτηση πληροφορίας στον Παγκόσμιο Ιστό (web search research) είναι μέρος του ευρύτερου χώρου της έρευνας για την Ανάκτηση Πληροφορίας (Information Retrieval – IR) (Knight & Spink, 2008). Σύμφωνα με τους (Mansourian & Madden, 2007) οι μελέτες που γίνονται στο συγκεκριμένο υποχώρο μπορούν να χωριστούν στις ακόλουθες κατηγορίες, βάσει των σημείων όπου εστιάζουν:

1. Έρευνες επάνω σε συμπεριφορά αναζήτησης, όπου εξετάζονται γνωστικά χαρακτηριστικά της συμπεριφοράς διαφόρων κατηγοριών χρηστών όσο αυτοί επιτελούν αναζήτηση πληροφοριών στον Παγκόσμιο Ιστό.
2. Έρευνες επάνω σε τεχνολογίες & εργαλεία αναζήτησης, όπου μελετώνται οι τεχνολογίες και οι μηχανές αναζήτησης πληροφοριών στον Παγκόσμιο Ιστό.
3. Έρευνες επάνω σε περιεχόμενα πηγών (ιστοσελίδων), όπου εξετάζονται οι πηγές που ανακτώνται και οι τρόποι με τους οποίους αναπαρίστανται τα περιεχόμενά τους.

Εκτός από τις τρεις προηγούμενες κατηγορίες μελετών υπάρχει και μία τέταρτη, οι πολυδιάστατες έρευνες, που μελετάνε συνδυαστικά δύο ή τρία ζητήματα από τις προαναφερθείσες κατηγορίες. Η παρούσα διπλωματική κάνει μια κριτική ανασκόπηση ερευνών της δεύτερης κατηγορίας, για το γνωσιακό χώρο όμως.

Γενικά, η αναζήτηση πληροφορίας περιλαμβάνει τη διατύπωση ενός ερωτήματος (επερώτησης αναζήτησης – query) προς άγραν πληροφοριών από μέρος του χρήστη (πρώτο στάδιο). Η επερώτηση είναι δυνατό να εκφράζεται σε φυσική γλώσσα, αν στη μηχανή αναζήτησης διατίθεται διεπαφή φυσικής γλώσσας (Natural Language Processing – NLP), σε κάποια γλώσσα βάσεων δεδομένων (όπως η SQL) (Hu, Chen, Scmalz, & Ritter, 2001; Jin, Ning, Jia, Wu, & Lu, 2008) ή στην «keywordesque», δηλαδή με χρήση ορισμένων λέξεων-κλειδιών (keywords) οι οποίες συμπυκνώνουν τις αναζητούμενες πληροφορίες. Όπως εξηγεί ο (Diaz, 2005) στη διδακτορική του διατριβή, ακολουθούν δύο ακόμη στάδια: εντοπισμός σχετικών ιστοσελίδων και ταξινόμησή τους. Ο εντοπισμός επιτυγχάνεται από προγράμματα λογισμικού που ονομάζονται spiders ή crawlers

και περιδιαβαίνουν τον Παγκόσμιο Ιστό καταγράφοντας τις ιστοσελίδες του και το περιεχόμενό τους. Εφόσον εντοπιστούν οι σχετικές ιστοσελίδες (οι οποίες ενίοτε είναι εκατοντάδες ή χιλιάδες) τα αποτελέσματα ταξινομούνται βάσει κάποιων αλγορίθμων, διαφορετικών για κάθε σύστημα αναζήτησης. Εδώ να προσθέσουμε ένα ακόμη σημαντικό κομμάτι, αυτό της διεπαφής (interface), δηλαδή του τρόπου με τον οποίο παρουσιάζονται οι ιστοσελίδες-αποτελέσματα. Η διεπαφή έχει ιδιαίτερη σημασία καθώς είναι ο τρόπος με τον οποίο τελικά παρουσιάζονται οι πληροφορίες στον άνθρωπο, οπότε μπορούμε να τη θεωρήσουμε ως το τέταρτο στάδιο της αναζήτησης πληροφορίας.

Συνοψίζοντας, λοιπόν, τα στάδια της αναζήτησης πληροφορίας στον WWW είναι:

1. Η διατύπωση της επερώτησης,
2. Ο εντοπισμός των σχετικών ιστοσελίδων/αποτελεσμάτων,
3. Η ταξινόμηση των αποτελεσμάτων, και
4. Η παρουσίαση (διεπαφή) των αποτελεσμάτων στο χρήστη.

Μέσα στα 20 χρόνια ύπαρξης του Παγκόσμιου Ιστού έχουν αναπτυχθεί πάμπολλες τεχνικές και συστήματα αναζήτησης του περιεχομένου του. Στην παρούσα διπλωματική θεωρείται πως τα υπάρχοντα συστήματα αναζήτησης ανήκουν σε μία από τις εξής τρεις κατηγορίες: παραδοσιακά (traditional), σημασιολογικά (semantic) και γνωσιακά (cognitive).

1.1 Παραδοσιακή αναζήτηση πληροφορίας

Η παραδοσιακή αναζήτηση πληροφορίας, υλοποιούμενη από μηχανές αναζήτησης όπως οι Google, Yahoo κ.ά., για τον εντοπισμό σχετικών ιστοσελίδων κάνει απλό ταίριασμα των λέξεων-κλειδιών της επερώτησης με τα δεδομένα της ιστοσελίδας, είτε αυτά βρίσκονται στο σώμα της ιστοσελίδας είτε στη μετά-ετικέτα (metatag) της. Η ανάλυση που γίνεται στην επερώτηση είναι μόνο συντακτική, δηλαδή η επερώτηση κατακερματίζεται στα δομικά στοιχεία (λέξεις) που την αποτελούν, χωρίς να εξετάζεται το εννοιολογικό τους περιεχόμενο.

Φυσικά, με την πάροδο των χρόνων έχουν χρησιμοποιηθεί πολλά ευριστικά τεχνάσματα για τη βελτίωση της ταξινόμησης των αποτελεσμάτων στην παραδοσιακή αναζήτηση πληροφορίας, καθώς ο όγκος των εγγράφων στον WWW αυξάνεται με αλματώδεις ρυθμούς και οι μηχανές αναζήτησης που βασίζονται σε παραδοσιακές μεθόδους αδυνατούν να αντιμετωπίσουν αποτελεσματικά το γεγονός αυτό. Για παράδειγμα, μια τεχνική είναι το φιλτράρισμα της πληροφορίας, όπου οι ιστοσελίδες που επιστρέφονται από τη μηχανή αναζήτησης φιλτράρονται βάσει ορισμένων κριτηρίων (όπως το προφίλ του χρήστη). Άλλο ευριστικό είναι η ταξινόμηση των αποτελεσμάτων βάσει του πλήθους των ιστοσελίδων που έχουν υπερσυνδέσμους προς την εκάστοτε ιστοσελίδα (δηλαδή η αξιοπιστία της ιστοσελίδας στην κοινότητα του Παγκόσμιου Ιστού (web community)) (π.χ. ο περιώνυμος αλγόριθμος PageRank τον οποίο χρησιμοποιεί η Google) ή η αξιολόγηση των ιστοσελίδων από τους χρήστες της συγκεκριμένης μηχανής, όπως αναφέρεται από τους (Wechsler, Baier, Nussbaum, & Baeza-Yates, 2004). Άλλη τεχνική είναι η επέκταση της επερώτησης (query expansion) με προτεινόμενες προς το χρήστη λέξεις-κλειδιά, ούτως ώστε να μειώνεται ο φόρτος του χρήστη στη διατύπωση της επερώτησής του. Και πάλι όμως, η βάση αυτών των τεχνικών είναι η συντακτική ανάλυση επερωτήσεων και περιεχομένων ιστοσελίδων.

1.2 Σημασιολογική αναζήτηση πληροφορίας

Η σημασιολογική αναζήτηση πληροφορίας, υλοποιούμενη από μηχανές αναζήτησης όπως οι Powerset, Cognition κ.ά., προχωράει ένα βήμα παραπέρα στην ανάλυση της επερώτησης: σύμφωνα με τον ορισμό που παρατίθεται στη Wikipedia (http://en.wikipedia.org/wiki/Semantic_search), εξάγει (είτε μέσω ειδικών εργαλείων αναζήτησης είτε μέσω των μετά-δεδομένων (metadata) στη μετά-ετικέτα, που έχουν συνταχθεί από το δημιουργό της ιστοσελίδας) τις έννοιες που περιέχονται σε κάθε ιστοσελίδα. Ταυτόχρονα, λαμβάνει τις λέξεις-κλειδιά της επερώτησης και προσπαθεί να βρει το νόημά τους βάσει και των συμφραζόμενων. Για να γίνει αυτό, χρησιμοποιούνται οντολογίες όπου κάθε λέξη

αντιστοιχίζεται σε μία ή περισσότερες έννοιες. Μπορεί επίσης να χρησιμοποιούνται δεδομένα εκφρασμένα σε σημασιολογικές γλώσσες περιγραφής για να προσαυξηθεί η πληροφορία που προκύπτει από την παραδοσιακή αναζήτηση (Guha, McCool, & Miller, 2003). Εφαρμόζοντας λοιπόν κάποιους συμπερασμούς στην επερώτηση επιλέγονται πρώτες προς εμφάνιση οι ιστοσελίδες που είναι νοηματικά εγγύτερες (σχετικότερες) προς την επερώτηση του χρήστη. Μετά τη δημοσίευση ενός επιδραστικού άρθρου από το (Zadeh, 2004), πλέον χρησιμοποιείται και fuzzy (δηλαδή όχι δίτιμη) λογική για την περιγραφή οντολογιών ή τη διατύπωση επερωτήσεων (ενδεικτικά, (Kim & Cho, 2007; Jin, Ning, Jia, Wu, & Lu, 2008)) στη σημασιολογική αναζήτηση. Για μια ανασκόπηση και κατηγοριοποίηση των σημαντικότερων σημασιολογικών συστημάτων αναζήτησης πληροφορίας στον Παγκόσμιο Ιστό, μπορεί κανείς να ανατρέξει στους (Wei, Barnaghi, & Bargiela, 2008).

1.3 Γνωσιακή αναζήτηση πληροφορίας

Η «γνωσιακή» αναζήτηση πληροφορίας καταρχήν δεν είναι καθιερωμένος επιστημονικός όρος. Όπου χρησιμοποιείται (ως “cognitive IR”), χρησιμοποιείται από επιστήμονες εκτός του Γνωσιακού Χώρου (συνήθως από το χώρο της Πληροφορικής, π.χ. (Spink & Cole, 2005)), κατά κάποιον τρόπο καταχρηστικά ως προς το τι θα έπρεπε να σημαίνει βάσει των κριτηρίων της Γνωσιακής Επιστήμης. Λίγοι είναι οι ερευνητές ((Dhir & Lee, 2009; Dimitrova M. , 2003)) που τον χρησιμοποιούν αναφερόμενοι σε τεχνικές αναζήτησης πληροφορίας που λαμβάνουν υπόψη στοιχεία από τη Γνωσιακή Επιστήμη. Αν θέλουμε βέβαια να είμαστε απόλυτα συνεπείς με τη Γνωσιακή Επιστήμη, οποιαδήποτε τεχνολογία χαρακτηρίζεται ως «γνωσιακή» θα πρέπει να κάνει τα πράγματα όπως ο ανθρώπινος νους. Αυτό όμως είναι ένα πολύπλευρο πρόβλημα και για την ώρα δεν υπάρχει ολοκληρωμένη λύση του. Σαν γνωσιακούς επιστήμονες, εμάς μας ενδιαφέρει το πώς ανακτώνται οι σχετικές πληροφορίες από το μυαλό ενός έμπειρου χρήστη ή ενός ειδήμονα της ανάκτησης γνώσης, όπως είναι π.χ. ένας βιβλιοθηκάριος· σαν τεχνολόγους, μας ενδιαφέρει πώς είναι δυνατόν οι μηχανές

αναζήτησης που θα φτιάξουμε να φέρνουν μπροστά στον οποιοδήποτε χρήστη τα αποτελέσματα που θέλει στη μορφή που τα αντιλαμβάνεται καλύτερα, όσο ασαφής κι αν είναι η επερώτησή του, ακριβώς όπως θα έκανε ένας καλός φίλος που γνωρίζει τις προτιμήσεις του χρήστη ή ένας έμπειρος βιβλιοθηκάριος. Ωστόσο, όπως εξηγεί κι ο (Saracevic, 2007, σ. 55), ένα πολύπλοκο σύνολο από παράγοντες εμπλέκεται στον καθορισμό της σχετικότητας των πληροφοριών προς τις γνωστικές ανάγκες των χρηστών. Μιλώντας με ρεαλιστικούς όρους, η γνωσιακή αναζήτηση πληροφορίας σήμερα μπορεί να αφορά τους τρόπους με τους οποίους εντοπίζεται και ανασύρεται η πληροφορία από τον WWW (σπανιότερο) ή (και) τη διεπαφή μέσω της οποίας αυτή η πληροφορία παρουσιάζεται στο χρήστη (συνηθέστερο). Είναι σε εξέλιξη πολλές ερευνητικές προσπάθειες για γνωστική μοντελοποίηση της συμπεριφοράς αναζήτησης πληροφορίας των χρηστών οι οποίες χρησιμοποιούνται ως έμπνευση για νέα συστήματα αναζήτησης που ακολουθούν αυτήν τη λογική (ενδεικτικά, (Schoelles, Gray, Veksler, & Gamard, 2006; Gwizdka, 2009)· για μια συνολική ανασκόπηση της έρευνας στον τομέα, βλ. (Spink & Cole, 2005; Spink & Zimmer, 2008)). Σ' αυτό το πλαίσιο λοιπόν, ως γνωσιακή νοούμε καταρχήν την αναζήτηση πληροφορίας στον WWW η οποία υλοποιείται με εργαλεία που λαμβάνουν υπόψη γνωσιακές θεωρίες για τη διατύπωση επερωτήσεων και την αναγνώριση/κωδικοποίηση/παγίωση/ανάσυρση πληροφορίας και τη συμπεριφορά των χρηστών κατά την αναζήτησή της (Aula, 2005, σ. 2), ώστε ενίοτε να μπορεί να προχωρήσει (και) σε εξατομίκευση αυτών των εργαλείων.

Στην ουσία, το πρόβλημα της γνωσιακής αναζήτησης πληροφορίας είναι το άλμα από την ανάκτηση πληροφορίας (information) στην εύρεση γνώσης (knowledge). Η λέξη «γνώση» χρωματίζεται με το ιδιαίτερο είδος πληροφορίας που αποκτούν οι άνθρωποι μέσα από την αντίληψη και την εμπειρία τους κι ο ιδανικός στόχος είναι η κατασκευή συστημάτων αναζήτησης που όχι απλά θα ανακτούν πληροφορία αλλά θα απαντούν σε ερωτήσεις (Zadeh, 2004), όπως κάνουν οι άνθρωποι μεταξύ τους στην καθημερινότητα. Αυτό που περιγράφηκε στην προηγούμενη παράγραφο είναι η υλοποίηση ενός συστήματος που θα μπορεί να βρίσκει με γνωστικά εύληπτο τρόπο πληροφορία από τον Παγκόσμιο Ιστό. Όμως

μετά από τη διαδικασία αυτής καθεαυτής της αναζήτησης και της εξόρυξης πληροφορίας, υπάρχει το κομμάτι του σχηματισμού της γνώσης. Νωρίτερα δόθηκε ένας ορισμός για την αναζήτηση πληροφορίας και αναφέρθηκαν τα τέσσερα στάδια που την αποτελούν. Η αντιμετώπιση της αναζήτησης πληροφορίας στον WWW με λογική Γνωσιακής Επιστήμης δεν μπορεί παρά να περιλαμβάνει κι ένα τελευταίο στάδιο, που θα υλοποιεί αυτό το άλμα από την πληροφορία στη γνώση.

Με την προσθήκη αυτού του τελευταίου βήματος, τα στάδια αναζήτησης πληροφορίας διαμορφώνονται τελικά στα εξής πέντε:

1. Διατύπωση επερώτησης,
2. Εντοπισμός σχετικών ιστοσελίδων/αποτελεσμάτων,
3. Ταξινόμηση αποτελεσμάτων,
4. Παρουσίαση (διεπαφή) αποτελεσμάτων στο χρήστη, και
5. Μετάβαση από την πληροφορία στη γνώση.

Το ιδανικό λοιπόν στην αναζήτηση πληροφορίας είναι να παρουσιάζεται στο χρήστη η γνώση που θέλει στη μορφή που θέλει και που είναι καταλληλότερη για τις δικές του ανάγκες· αυτό εμπεριέχει:

- Εντοπισμό των χαρακτηριστικών που ο χρήστης θεωρεί σημαντικά της ποιότητας των ιστοσελίδων.
- Ορισμό των κριτηρίων μέτρησης της ποιότητας των χαρακτηριστικών αυτών.
- Εξατομίκευση των αποτελεσμάτων με παρουσίαση στο χρήστη πρώτων αυτών που ταιριάζουν στις γνωστικές του ανάγκες.

Κεφάλαιο 2: Οι κυριότερες τάσεις στη γνωσιακή αναζήτηση πληροφορίας

Σ' αυτό το κεφάλαιο παρουσιάζονται οι κυριότερες τάσεις στη γνωσιακή αναζήτηση πληροφορίας, όπως έχουν διαμορφωθεί σε πραγματικά συστήματα που έχουν υλοποιηθεί για ερευνητικούς ή εμπορικούς σκοπούς την τελευταία δεκαετία. Ορισμένα απ' αυτά τα συστήματα συνδυάζουν δύο ή περισσότερες τάσεις, αλλά η κατηγοριοποίησή τους γίνεται βάσει της επικρατέστερης τάσης που φαίνεται να τα διαμορφώνει.

Οι τάσεις αυτές είναι:

- Η αναζήτηση εντός πλαισίου (contextual web search),
- Η εξατομικευμένη αναζήτηση (personalized web search),
- Οι γνωσιακοί πράκτορες (cognitive web agents),
- Η συνεργατική αναζήτηση (collaborative web search),
- Η κατασκευή γνωσιακών διεπαφών (cognitive interfaces), και
- Η μετά-γνωστική υποστήριξη των χρηστών (meta-cognitive tools).

Οι διαφορές μεταξύ αυτών των τάσεων έγκεινται στο στάδιο της αναζήτησης (όπως αυτά αναφέρθηκαν στο Κεφάλαιο 1) στο οποίο υλοποιούνται, και στις μεθοδολογίες που χρησιμοποιούνται.

Όμως, και μέσα σε καθεμιά απ' αυτές τις τάσεις το κάθε σύστημα διαφέρει από τα υπόλοιπα της κατηγορίας του. Αυτό έχει να κάνει με τις γνωσιακές θεωρίες/μοντέλα όπου στηρίζεται και με τις μεθοδολογίες/τεχνολογίες πάνω στις οποίες υλοποιείται το κάθε σύστημα.

2.1 Αναζήτηση εντός πλαισίου

Ο σκοπός για τον οποίο αναζητείται η πληροφορία γενικά ποικίλλει από χρήστη σε χρήστη και σχετίζεται με το πλαίσιο (context) στο οποίο ψάχνει κανείς πληροφορίες. Το πλαίσιο περιλαμβάνει και το χρήστη (ενδιαφέροντα, μόρφωση κι όποιο άλλο χαρακτηριστικό τον προσδιορίζει σαν άτομο) και το έργο της αναζήτησης (σκοπό, χρήση κ.ό.κ.) και το σύστημα αναζήτησης, κι είναι σημαντικό να ληφθεί υπ' όψη προκειμένου να αυξηθεί η σχετικότητα των αποτελεσμάτων με τον αρχικό στόχο του χρήστη (Lawrence, 2000). Οι μέθοδοι με τις οποίες συλλέγονται τα απαραίτητα δεδομένα για την κατασκευή του πλαισίου σε κάθε περίπτωση ποικίλλουν από σύστημα σε σύστημα.

Η ανάγκη για υλοποίηση αναζήτησης εντός πλαισίου εκ μέρους των συστημάτων αναζήτησης, δηλαδή αναζήτησης που λαμβάνει υπόψη όχι μόνο την επερώτηση παρά και τα συμπαρομαρτούντα – προφίλ χρήστη, σκοπός, γνωστικό πεδίο αναζήτησης κ.ό.κ. –, έχει προκύψει εντονότερα από τη δεκαετία του '90 κι έπειτα (Spink & Cole, 2005, σ. 193). Μια πολύ σύντομη παρουσίαση των προσεγγίσεων επάνω στο ζήτημα του πλαισίου στο χώρο της ανάκτησης πληροφορίας βρίσκεται στην εργασία των (Freund & Toms, 2005, σσ. 4-5).

Τα γνωσιακά συστήματα αναζήτησης που θα παρουσιαστούν ακόλουθα και τα οποία εμπίπτουν στην παρούσα κατηγορία, υλοποιούν τα τέσσερα πρώτα στάδια της αναζήτησης πληροφορίας (επερώτηση-εντοπισμός-ταξινόμηση-παρουσίαση σχετικών αποτελεσμάτων).

2.1.1 Σύστημα σημασιολογικής αναζήτησης βασισμένο σε γνωσιακό μοντέλο

Ένα γνωσιακό μοντέλο απόκτησης πληροφορίας το οποίο βασίζεται σε εργασίες των (Guilford, 1967; Perkins, 1998; Gardner, 1999) ακολουθούν οι (Wechsler, Baier, Nussbaum, & Baeza-Yates, 2004) στην κατασκευή ενός συστήματος αναζήτησης πληροφορίας στον Παγκόσμιο Ιστό. Η δική τους μηχανή, σε αντίθεση με πολλές άλλες που ουσιαστικά αναδιατάσσουν τα αποτελέσματα των Google, Yahoo κ.ό.κ., εξαρχής κάνει αυτόνομα τις αναζητήσεις της βασιζόμενη

σε οντολογίες. Οι οντολογίες αυτού του συστήματος είναι μέρος του CYC Project, το οποίο είναι μια προσπάθεια για έξυπνο προγραμματισμό μηχανών, με εκ μέρους τους δυνατότητα εκμάθησης γνώσης και συλλογισμού με αυτήν. Για την περιγραφή των εγγράφων του WWW μέσω οντολογιών, οι (Wechsler, Baier, Nussbaum, & Baeza-Yates, 2004) ακολουθούν μια προσέγγιση σύμφωνα με την οποία κάθε κλάση μιας μοναδικής οντολογίας δείχνει προς μια συλλογή εγγράφων με παρεμφερές περιεχόμενο. Για την εξαγωγή των εννοιών από τα έγγραφα και τη συσχέτιση των εγγράφων προς τις οντολογίες χρησιμοποιούν ως επεξεργαστή φυσικής γλώσσας μια βιβλιοθήκη της Java, το Stand-Alone Gazetteer.

Το γνωσιακό μοντέλο που χρησιμοποιούν περιλαμβάνει πέντε μέρη: αντίληψη, παρατήρηση, ερμηνεία, κατανόηση και αναστοχασμό. Καθένα από τα μέρη μοντελοποιείται με ένα στοιχείο (module). Αρχικά (πρώτο μέρος) ο χρήστης εισάγει κάποιες λέξεις-κλειδιά, χωρίς να επιβαρύνεται με το γνωστικό φόρτο του πώς ακριβώς να διατυπώσει την επερώτησή του· με τον ίδιο τρόπο π.χ. που θα ρωτούσε έναν βιβλιοθηκάριο «θέλω ένα βιβλίο με πληροφορίες για τα αυτοκίνητα» γράφει τη λέξη «αυτοκίνητο». Έπειτα (δεύτερο μέρος) γράφει μια σύντομη παράγραφο με το συγκεκριμένο πεδίο του ενδιαφέροντός του σε σχέση με τις λέξεις-κλειδιά (όπως θα έκανε και με το βιβλιοθηκάριο, προκειμένου να διευκρινίσει το ερώτημά του), οπότε χτίζεται ένα προφίλ χρήστη βάσει των εισαγόμενων δεδομένων. Έτσι, το σύστημα μπορεί να ερμηνεύσει (τρίτο μέρος) το νοηματικό πλαίσιο μέσα στο οποίο εμφανίζονται οι λέξεις-κλειδιά. Για την παρουσίαση των αποτελεσμάτων (τέταρτο μέρος) εμφανίζεται ο υπογράφος (subgraph) της οντολογίας που είναι νοηματικά εγγύτερος στην πληροφοριακή ανάγκη του χρήστη. Πατώντας πάνω σε κάθε κόμβο του υπογράφου εμφανίζονται οι αντίστοιχες πηγές (ιστοσελίδες) με τις σχετικές πληροφορίες. Το πέμπτο μέρος στάδιο είναι ο «αναστοχασμός» του συστήματος πάνω στην ποιότητα της δουλειάς του, πράγμα το οποίο κρίνεται από το πόσους από τους προτεινόμενους κόμβους εξερευνά ο χρήστης, οπότε το σύστημα μπορεί να προχωρήσει σε ενημέρωση του προφίλ του χρήστη για βελτίωση μελλοντικών

αναζητήσεων. Το σύστημα αυτό υλοποιεί τα τέσσερα πρώτα στάδια της αναζήτησης πληροφορίας, όπως αυτά αναφέρονται στο 1^ο Κεφάλαιο.

2.1.2 Contextual Search Project

Η αφορμή για το σύστημα που ανέπτυξαν οι (Freund & Toms, 2005) ήταν η βελτίωση του συστήματος ανάκτησης εγγράφων που χρησιμοποιείται από συμβούλους πώλησης που απευθύνονται σε μηχανικούς λογισμικού. Τα έγγραφα αυτά βρίσκονται σε διάφορες τοποθεσίες, συμπεριλαμβανομένου του WWW. Εκκινώντας από τη μελέτη της συμπεριφοράς των εμπειρογνομόνων οι (Freund & Toms, 2005) ανέπτυξαν ένα εμπειρικό μοντέλο αναζήτησης για το συγκεκριμένο γνωστικό τομέα. Στην προκειμένη περίπτωση το πλαίσιο αφορά το είδος των χρηστών (σύμβουλοι πωλήσεων), το γνωστικό πεδίο αναζήτησης (εξειδικευμένο λογισμικό), το έργο αναζήτησης (παροχή συμβουλευτικών υπηρεσιών) και το είδος της πληροφορίας (παρουσιάσεις για πελάτες, τεχνικά εγχειρίδια κ.λπ.). Το σύστημα των (Freund & Toms, 2005) ήταν σε φάση υλοποίησης κατά τη δημοσίευση του (Contextual search: from information behavior to information retrieval) αλλά η σχεδιαστική πρόθεσή τους ήταν να συλλέγονται άμεσα από τους χρήστες τα απαραίτητα δεδομένα για το έργο αναζήτησης, να εφαρμόζονται τεχνικές μηχανικής μάθησης για την κατηγοριοποίηση των εγγράφων και οι σχέσεις που συνιστούν το πλαίσιο να ενσωματώνονται στον ίδιο τον αλγόριθμο εντοπισμού και ταξινόμησης κι όχι στην παρουσίαση των αποτελεσμάτων. Αυτό είναι σημαντικό να τονιστεί, καθώς σε αρκετά συστήματα, ακόμη και σ' αυτά που χαρακτηρίζονται γνωσιακά, παρακάμπτεται η εφαρμογή των τεχνικών στον αλγόριθμο αυτό καθεαυτό (ο οποίος εντοπίζει τις σχετικές πληροφορίες στο WWW) και γίνεται εφαρμογή τους κατευθείαν στην παρουσίαση των αποτελεσμάτων.

2.2 Εξατομικευμένη αναζήτηση πληροφορίας

Τι είναι η εξατομικευμένη αναζήτηση πληροφορίας; Μπορεί οι πληροφορίες στο Διαδίκτυο να είναι διαθέσιμες σε όλους τους χρήστες, αλλά ανάλογα με τις προτιμήσεις, τις εμπειρίες, το μορφωτικό επίπεδο κ.ά., δηλαδή ανάλογα με την

προσωπικότητα του καθενός, διαφορετικές ιστοσελίδες με κατά βάση ίδιες πληροφορίες μπορεί να απαντούν στις γνωστικές ανάγκες διαφορετικών χρηστών. Για τη βελτίωση των αποτελεσμάτων που προκύπτουν από μια μηχανή αναζήτησης ώστε να ταιριάζουν καλύτερα στις ανάγκες του κάθε χρήστη χρησιμοποιείται ευρέως η τεχνική της εξατομίκευσης των αναζητήσεων, όπου με τη χρήση ενός προφίλ χρήστη τα αποτελέσματα αναδιατάσσονται σύμφωνα με τις μακροπρόθεσμες ή βραχυπρόθεσμες προτιμήσεις του καθενός. Οι πληροφορίες που χρησιμοποιούνται για την κατασκευή του προφίλ μπορεί να συλλέγονται άμεσα ή έμμεσα και οι τεχνολογίες που χρησιμοποιούνται για την αναπαράστασή του περιλαμβάνουν σημασιολογικά δίκτυα/οντολογίες, λέξεις-κλειδιά με βάρη ή κανόνες συσχέτισης, σύμφωνα με τους (Gauch, Speretta, Chandramouli, & Micarelli, 2007). Μιλώντας για συλλογή δεδομένων προκειμένου να χτιστεί το προφίλ του χρήστη, ως «άμεση» χαρακτηρίζεται όταν οι χρήστες ρητά δίνουν πληροφορίες (για τα δημογραφικά τους, τις σπουδές τους, τα ενδιαφέροντά τους κ.λπ.), ενώ ως «έμμεση» χαρακτηρίζεται όταν αυτές οι πληροφορίες αλιεύονται από το σύστημα αναζήτησης, χωρίς ο χρήστης να χρειάζεται να ασχοληθεί με την παροχή τους γι' αυτόν το σκοπό.

Όπως αναφέρουν οι (Daoud, Tamine-Lechani, & Boughanem, 2008, σ. 57), «η εξατομικευμένη αναζήτηση πληροφορίας είναι ένα υπό-πεδίο της αναζήτησης εντός πλαισίου, που εστιάζει στο υπόβαθρο, τα ενδιαφέροντα και τις προτιμήσεις του χρήστη ως τους κύριους παράγοντες του πλαισίου». Γενικά, η εξατομίκευση των αναζητήσεων (ενδεικτικά, (Lieberman, Fry, & Weitzman, 2001; Pitkow, και συν., 2002; Guha, McCool, & Miller, 2003; Kim & Cho, 2007; Dhir & Lee, 2009; Jiang & Tan, 2009)) χρησιμοποιείται κατά κόρον για τη βελτίωση των αποτελεσμάτων μιας αναζήτησης. Ωστόσο, οι τεχνικές που χρησιμοποιούνται σε ορισμένα τέτοια συστήματα μπορούν να χαρακτηριστούν γνωσιακές μόνο με μια πολύ ευρεία έννοια ή ακόμη κι αυτός ο χαρακτηρισμός μπορεί να ποικίλει ανάλογα με το σύστημα όπου χρησιμοποιείται η συγκεκριμένη τεχνική, εφόσον συχνά δε βασίζονται στιβαρά πάνω σε κάποια θεωρία για τη λειτουργία του νου κατά το έργο της αναζήτησης πληροφορίας. Για λεπτομερέστερες ανασκοπήσεις των σημαντικότερων τάσεων στην εξατομικευμένη αναζήτηση πληροφορίας στον

Παγκόσμιο Ιστό, μπορεί κανείς να ανατρέξει στους (Keenoy & Levene, 2005; Jeevan & Padhi, 2006; Gauch, Speretta, Chandramouli, & Micarelli, 2007).

Η εξατομίκευση αναφέρεται εδώ, παρόλο που σαν τεχνική χρησιμοποιείται και στην παραδοσιακή και στη σημασιολογική αναζήτηση πληροφορίας, επειδή η υλοποίησή της προϋποθέτει παρακολούθηση της συμπεριφοράς αναζήτησης του χρήστη. Αυτή η παρακολούθηση μπορεί να εξαντλείται σε απλή περιπτωσιολογία – όπου βέβαια δε μιλάμε για γνωσιακή προσέγγιση της εξατομίκευσης, π.χ. (Lieberman, Fry, & Weitzman, 2001) – ή να βασίζεται σε συγκεκριμένες θεωρίες για τη λειτουργία του νου, πράγμα που προκύπτει από τη μελέτη των γνωστικών διαστάσεων του έργου της αναζήτησης πληροφορίας στον Παγκόσμιο Ιστό. Προφανώς η γνωσιακή αναζήτηση πληροφορίας επεκτείνει μεθόδους και χρησιμοποιεί κι άλλα εργαλεία της σημασιολογικής αναζήτησης, θέμα που θα συζητηθεί στο τέλος του κεφαλαίου αυτού.

Στο παρελθόν η αναπαράσταση των προφίλ χρηστών βασιζόταν σε μάλλον απλοϊκές μεθόδους, όπως τα διανύσματα από λέξεις-κλειδιά. Στα εξατομικευμένα γνωσιακά συστήματα που θα παρουσιαστούν, για την αναπαράσταση των προφίλ χρηστών χρησιμοποιούνται οντολογίες πεδίου και βάσεις γνώσης, η συλλογή των δεδομένων γίνεται με άμεσο κι έμμεσο τρόπο είτε από πράκτορες, είτε μέσα από ιστορικά αναζητήσεων είτε από ενέργειες του ίδιου του χρήστη. Η εξατομίκευση γίνεται στο τρίτο στάδιο της αναζήτησης πληροφορίας (ταξινόμηση αποτελεσμάτων).

2.2.1 Outride

Ένα πρώιμο σύστημα εξατομικευμένης αναζήτησης πληροφορίας είναι το Outride των (Pitkow, και συν., 2002). Οι δημιουργοί του ακολούθησαν τη λογική της Θεωρίας Θήρευσης της Πληροφορίας (Information Foraging Theory), των (Pirolli & Card, 1999). Κατ' αρχήν, ο χρήστης διατυπώνει την επερώτησή του και το Outride κάνει προσαύξηση της επερώτησης (query augmentation), δηλαδή την εμπλουτίζει με όρους που την αποσαφηνίζουν/εξειδικεύουν. Η προσαύξηση γίνεται βάσει του προφίλ χρήστη, που περιέχει δημογραφικά, ιστορικό αναζητήσεων, ιστορικό επισκέψεων, προοριζόμενες χρήσεις της πληροφορίας

κ.ά.. Στη συνέχεια γίνεται επεξεργασία των αποτελεσμάτων (result processing) που επιστρέφονται από μια εμπορική μηχανή αναζήτησης (εν προκειμένω, την Google), δηλαδή βάσει του προφίλ χρήστη τα αποτελέσματα φιλτράρονται. Η συλλογή των δεδομένων για την κατασκευή του προφίλ γίνεται με έμμεσο τρόπο, παρακολουθώντας το χρήστη κατά τις περιηγήσεις του στον WWW. Το σύστημα εμφανίζεται ως διεπαφή ενσωματωμένη στο πρόγραμμα-φυλλομετρητή.

2.2.2 Search-In-Synchrony

Οι (Dhir & Lee, 2009) έχουν δημιουργήσει το Search-In-Synchrony, ένα πολλαπλών-χρηστών περιβάλλον (multi-user environment) εξατομίκευσης της αναζήτησης πληροφορίας. Γι' αυτόν το σκοπό χρησιμοποίησαν έναν αναγνωριστικό πράκτορα (reconnaissance agent), το User Profile Model (UPM), που ενσωματώνει εκ των προτέρων γνώση για την IR συμπεριφορά και τις γνωστικές στρατηγικές αναζήτησης του χρήστη, με σκοπό τη μίμησή τους. Γενικά μιλώντας, η δουλειά των αναγνωριστικών πρακτόρων παρομοιάζεται από τους (Lieberman, Fry, & Weitzman, 2001) με αυτή ενός φίλου του χρήστη, ο οποίος γνωρίζοντας τα ενδιαφέροντα του χρήστη και παρατηρώντας τις ιστοσελίδες που επισκέπτεται τον «δρομολογεί» κατευθείαν στο μονοπάτι περιήγησης που πιθανόν θα διάλεγε κι ο ίδιος ο χρήστης. Οι προτεινόμενες από τον πράκτορα ιστοσελίδες συνδέονται σημασιολογικά με αυτήν στην οποία βρίσκεται ο χρήστης. Προφανώς το UPM δεν είναι στατικό· παρακολουθεί τις μεταβολές στο περιεχόμενο του WWW και στην αλληλεπίδραση του χρήστη μ' αυτόν και διακρίνει τις συμπεριφορές των διαφόρων χρηστών του (δεδομένου ότι είναι ένα πολλαπλών-χρηστών περιβάλλον). Η εκπαίδευσή του γίνεται με αλγορίθμους μηχανικής μάθησης (Lerner & Lawrence, 2001; Williams & Ren, 2001) και προσεγγίσεις νευρωνικών δικτύων (Liu & Zhang, 2001), με σκοπό την ενσωμάτωση γνωστικών ικανοτήτων στον πράκτορα.

Η συλλογή των στοιχείων που επιτρέπουν στο σύστημα να αρχικοποιήσει το προφίλ του χρήστη γίνεται και με άμεσο και με έμμεσο τρόπο. Το σύστημα χρησιμοποιεί μια βάση δεδομένων με το ιστορικό των αναζητήσεων του χρήστη. Για την κατασκευή της βάσης αρχικά συλλέγονται κάποια δεδομένα: όταν ο

χρήστης κάνει την επερώτησή του σε κάποια εμπορική μηχανή αναζήτησης χαρακτηρίζει τα αποτελέσματα ως ενδιαφέροντα ή μη πατώντας ένα κατάλληλο κουμπί σε μια γραμμή εργαλείων του προγράμματος-φυλλομετρητή που χρησιμοποιεί για να περιηγηθεί στον WWW (άμεση συλλογή). Τα HTML περιεχόμενα των ιστοσελίδων-αποτελεσμάτων αναλύονται και αποθηκεύονται στη βάση μαζί με την επερώτηση, το URL της ιστοσελίδας και το όνομα χρήστη. Μετά από προεπεξεργασία κατασκευάζεται με στατιστικές μεθόδους το UPM, ένας 2-διάστατος πίνακας με τις συχνότητες εμφάνισης των σημαντικότερων λέξεων στις διάφορες ιστοσελίδες (πράγμα που θεωρείται πως αντιπροσωπεύει τα μακροπρόθεσμα ενδιαφέροντα του χρήστη). Για την ενημέρωση του UPM που χρησιμοποιείται για την επαναταξινόμηση των ιστοσελίδων που προκύπτουν ως αρχικά αποτελέσματα αναζήτησης από την εμπορική μηχανή αναζήτησης, ώστε πρώτες να εμφανίζονται οι ιστοσελίδες που αναμένεται να είναι πλησιέστερες στα ενδιαφέροντα του χρήστη.

2.2.3 **OntoSearch_U**

Οι οντολογίες πεδίου (domain ontologies) είναι σημασιολογικά δίκτυα που αναπαριστούν έννοιες για ένα συγκεκριμένο κι αρκετά στενό γνωστικό πεδίο, ενώ οι βάσεις γνώσης περιέχουν τα λεγόμενα γεγονότα, δηλαδή αλήθειες ή πεποιθήσεις εκφρασμένες με απλές προτάσεις. Οι οντολογίες πεδίου χρησιμοποιούνται κατά κόρον για την αναπαράσταση γνώσης. Οι (Jiang & Tan, 2009) τις εκμεταλλεύονται για την υλοποίηση του OntoSearch_U, ενός συστήματος εξατομικευμένης αναζήτησης στο Σημασιολογικό Ιστό (Semantic Web), όπου το προφίλ χρήστη αναπαρίσταται ως οντολογία των ενδιαφερόντων του. Για την ακρίβεια, χρησιμοποιούν μια οντολογία πεδίου για να παράξουν μια υβριδική (και μοναδική για κάθε χρήστη) «οντολογία χρήστη», όπου οι έννοιες της οντολογίας «βαρύνονται» με τέτοιο τρόπο ώστε να αντανakλούν τα ενδιαφέροντα του χρήστη. Η εκάστοτε ενεργοποίηση των κόμβων της οντολογίας χρήστη (δηλαδή τα βάρη που αποδίδονται στις έννοιες της οντολογίας και που αντιπροσωπεύουν τα τρέχοντα ενδιαφέροντα του χρήστη) γίνεται βάσει της θεωρίας διαδιδόμενης ενεργοποίησης (Spreading Activation Theory) (Anderson,

1983) που στη Γνωσιακή Επιστήμη είναι πολύ δημοφιλής για την επεξεργασία πληροφορίας από τη μνήμη. Κάθε προσωρινή ενεργοποίηση συνδυάζεται με τα προηγούμενα ενδιαφέροντα του χρήστη στο γνωστικό πεδίο για να δημιουργηθούν οι εξατομικευμένες υπηρεσίες.

Για την εκπαίδευση του μοντέλου ο χρήστης δηλώνει άμεσα ποιες ιστοσελίδες θεωρεί σχετικές. Η οντολογία χρήστη αξιοποιείται για την επαναταξινόμηση των αποτελεσμάτων της αναζήτησης, τα οποία έχουν προκύψει από κάποια εμπορική μηχανή αναζήτησης. Η πρωτοτυπία του OntoSearch_U είναι ότι αξιοποιεί τα σημασιολογικά (semantics) των εγγράφων (με τα οποία είναι χαρακτηρισμένα) και, χρησιμοποιώντας την οντολογία χρήστη, συμπεραίνει τη σημασιολογία των επερωτήσεων, χωρίς ο χρήστης να δηλώνει σαφώς τις έννοιες που περιέχονται στο ερώτημά του.

2.3 Έξυπνοι πράκτορες βασισμένοι σε γνωσιακά μοντέλα

Σύμφωνα με τον ορισμό που δίνεται από το (Wooldridge, 2001, σ. 15) «Οι πράκτορες (agents) είναι υπολογιστικά συστήματα που βρίσκονται σε κάποιο περιβάλλον, και που είναι ικανά για αυτόνομη δράση σ' αυτό το περιβάλλον προκειμένου να εκπληρώσουν τους στόχους του σχεδιασμού τους». Οι πράκτορες του Παγκόσμιου Ιστού (web agents) είναι πράκτορες που δρουν στο Διαδίκτυο βοηθώντας τους χρήστες να ικανοποιήσουν τις πληροφοριακές τους ανάγκες. Αυτοί οι πράκτορες συνήθως εξειδικεύονται σε κάποια εφαρμογή του WWW (μηχανές αναζήτησης, ηλεκτρονικό ταχυδρομείο, ηλεκτρονικό εμπόριο κ.λπ.). Στη συνέχεια θα παρουσιαστούν συστήματα αναζήτησης που χρησιμοποιούν πράκτορες οι οποίοι έχουν κατασκευαστεί βάσει γνωσιακών μοντέλων. Ο ρόλος – και η πρωτοτυπία – των συγκεκριμένων πρακτόρων έγκειται πρωταρχικά στο ότι αυτοί – κι όχι κάποια εμπορική μηχανή αναζήτησης – εντοπίζουν τις ιστοσελίδες που περιέχουν πληροφορίες σχετικές προς τις ανάγκες του χρήστη.

2.3.1 Cognitive web agents

Δουλειές των (Dimitrova, Barakova, Lourens, & Radeva, 2004; Dimitrova, Wagatsuma, & Yamaguchi, 2007) προτείνουν και υλοποιούν μια ανθρώπινου τύπου μνήμη για την κατασκευή ενός πράκτορα του Παγκόσμιου Ιστού, καθώς κάτι τέτοιο δίνει στον πράκτορα τη δυνατότητα εκμάθησης από την εμπειρία. Στην κατασκευή του πράκτορα χρησιμοποιήθηκαν Hebbian κανόνες μάθησης (Dimitrova, Barakova, Lourens, & Radeva, 2004) και η θεωρία της Theta Phase Precession για την επεισοδική μνήμη, προκειμένου να γίνει η μοντελοποίηση της αυτοβιογραφικής μνήμης (Dimitrova, Wagatsuma, & Yamaguchi, 2007) και συνεπακόλουθα, για την εκμάθηση επεισοδίων περιήγησης. Η θεωρία της Theta Phase Precession υποθέτει την εμπλοκή του ενδορρινικού φλοιού, του ιππόκαμπου και προμετωπιαίων περιοχών για την κωδικοποίηση, παγίωση και ανάκληση αυτοβιογραφικών αναμνήσεων. Ο λόγος για τον οποίο οι ερευνητές ενδιαφέρθηκαν για την υλοποίηση μιας ανθρώπινου τύπου μνήμης είναι για την καλύτερη πρόβλεψη, εκ μέρους του πράκτορα, των πληροφοριακών αναγκών των χρηστών. Τα συνδεδειστικά μοντέλα που κατασκεύασαν οι (Dimitrova, Barakova, Lourens, & Radeva, 2004; Dimitrova, Wagatsuma, & Yamaguchi, 2007) χρησιμοποιήθηκαν για την κατασκευή πρακτόρων που συνδυάζουν κομμάτια γνώσης από ιστοσελίδες που έχουν επισκεφθεί, ούτως ώστε οι εμπειρίες της επίσκεψής τους να χρησιμοποιηθούν σε επόμενες περιηγήσεις τους στον Παγκόσμιο Ιστό για να βελτιωθεί η εμπειρία της αναζήτησης. Εδώ, η έμφαση δίνεται όχι στην αποθήκευση/εκμάθηση εκ μέρους του πράκτορα της ακριβούς σειράς των ιστοσελίδων τις οποίες επισκέφθηκε, αλλά στην αποθήκευση/εκμάθηση των εποικοδομητικών θραυσμάτων γνώσης μέσα στις ιστοσελίδες. Η βεβαιότητα για την ορθότητα της προσέγγισης των ερευνητών ενισχύθηκε από ένα πείραμα αναζήτησης στον WWW, όπου οι ενέργειες των χρηστών αρχικά ήταν τυχαίες αλλά τελικά είχαν νόημα (εμφανή στόχο), κατασκευάζοντας στην πορεία (κι όχι αναπαράγοντας από μνήμης) ένα μονοπάτι αναζήτησης που οδηγούσε στον τελικό στόχο/αναζητούμενη πληροφορία.

2.3.2 RASCALLI

Επίσης προσανατολισμένο στην ανάπτυξη πρακτόρων που ζουν και δρουν μέσα στο Διαδίκτυο είναι το έργο RASCALLI (Responsive Artificial Situated Cognitive Agents Living and Learning in the Internet) των (Krenn, 2008; Krenn & Schollum, 2008), όπου για την κατασκευή ορισμένων από τους πράκτορες αξιοποιείται το μνημονικό μοντέλο DUAL/AMBR. Αυτό το μοντέλο διακρίνει μεταξύ μνήμης εργασίας και μακροπρόθεσμης μνήμης¹ και για όσους από τους Rascalli πράκτορες ο νους τους υλοποιείται βάσει του DUAL/AMBR, μπορεί κανείς να πει ότι είναι γνωσιακοί πράκτορες. Εκτός από την εξέταση εγγράφων του Παγκοσμίου Ιστού, οι Rascalli έχουν πρόσβαση σε βάσεις γνώσης, αλλά επίσης συλλέγουν γνώση για τις προτιμήσεις των χρηστών τους. Έτσι μαθαίνουν και προσαρμόζουν τη συμπεριφορά τους σύμφωνα με τις ανάγκες των χρηστών τους και κάθε Rascallo (ενικός για το Rascalli) είναι πλήρως εξατομικευμένος για κάποιον χρήστη. Οι χρήστες κάνουν ερωτήσεις και αναζητούν πληροφορίες μέσω κατάλληλων διεπαφών και τα αποτελέσματα παρουσιάζονται από τους Rascalli με γνωστικά εύληπτο τρόπο (μέσω ενός ενσωματωμένου διαλογικού πράκτορα (embodied conversational agent) ή μέσω υπερσυνδέσμων προς ιστοσελίδες με ενδιαφέρουσες πληροφορίες). Στους (Krenn & Schollum, 2008) οι Rascalli δραστηριοποιούνται σε ένα συγκεκριμένο γνωστικό πεδίο (τη μουσική) αλλά μπορούν να χρησιμοποιηθούν γενικά σε διάφορα γνωστικά πεδία. Το πρότζεκτ Rascalli υλοποιεί και τα τέσσερα στάδια της αναζήτησης πληροφορίας βασιζόμενο σε γνωσιακά εργαλεία· από τον εντοπισμό των σχετικών ιστοσελίδων έως τη διεπαφή χρήστη-υπολογιστή, είναι στιβαρά θεμελιωμένο σε γνωσιακές έρευνες. Οι τεχνολογίες που χρησιμοποιεί είναι ενίοτε σημασιολογικές (RDFS/OWL²) και ενίοτε καθαρά γνωσιακές (DUAL/AMBR).

¹ Για περισσότερες λεπτομέρειες σχετικά με την αρχιτεκτονική του μοντέλου DUAL/AMBR μπορεί κανείς να ανατρέξει στους (Kostadinov, Petkov, & Grinberg, 2008; Kostadinov & Grinberg).

² Γλώσσες περιγραφής δεδομένων για το Σημασιολογικό Ιστό.

2.4 Συνεργατική αναζήτηση πληροφορίας

Μια διαφορετική τάση στη γνωσιακή αναζήτηση πληροφορίας είναι η συνεργατική αναζήτηση. Κατά κάποιον τρόπο, μπορεί κανείς να πει ότι αυτή η τάση είναι η αντίρροπη της εξατομίκευσης, υπό την έννοια ότι δεν εστιάζει στον έναν χρήστη, αλλά στο σύνολο. Στην εξατομίκευση το σύστημα αδιαφορεί για τους υπόλοιπους χρήστες (γι' αυτό συχνά «τρέχει» στον Η/Υ του χρήστη κι όχι σε κάποιον εξυπηρετητή (server)) και οι συστάσεις που κάνει προς τον εκάστοτε χρήστη είναι «sur-mesure» βάσει των αναγκών και χαρακτηριστικών του συγκεκριμένου χρήστη. Στη συνεργατική αναζήτηση, από την άλλη, το σύστημα λαμβάνει υπόψη τις ενέργειες και των υπολοίπων χρηστών που είναι συνδεδεμένοι σε αυτό και που συνιστούν τον «κοινωνικό περίγυρο» (όπως χαρακτηριστικά αναφέρουν οι (Sahay, Venkatesh, & Ram, 2009)) του εν λόγω χρήστη στο ηλεκτρονικό περιβάλλον του συστήματος, προκειμένου βέβαια να εξυπηρετήσει καλύτερα το χρήστη. Οι χρήστες συζητούν μεταξύ τους και πραγματοποιούν αναζητήσεις υποβοηθούμενοι από κάποιον πράκτορα.

Στη συνεργατική αναζήτηση πληροφορίας μπορεί να υπάρχει χρηστο-κεντρική (user oriented) ή συστημο-κεντρική (system oriented) προσέγγιση. Μπορεί δηλαδή να προσφέρεται μια διεπαφή που επιτρέπει στο χρήστη να επικοινωνεί άμεσα με άλλους χρήστες, να παρακολουθεί τις ενέργειες αναζήτησης των άλλων και να αποφασίζει αυτός ο ίδιος ποιες ιστοσελίδες που χρησιμοποιούν οι άλλοι χρήστες μπορεί να είναι χρήσιμες και γι' αυτόν (χρηστο-κεντρική προσέγγιση), ή αυτή η παρακολούθηση να γίνεται αυτόματα από το σύστημα και μ' έναν τρόπο διάφανο προς το χρήστη (συστημο-κεντρική προσέγγιση).

Το γνωσιακό σύστημα που θα παρουσιαστεί ακολουθεί τη συστημο-κεντρική προσέγγιση, χρησιμοποιεί σημασιολογικές τεχνολογίες όπως έξυπνους πράκτορες και οντολογίες, και καλύπτει τα τέσσερα πρώτα στάδια της αναζήτησης πληροφορίας, μιας και δε βασίζεται σε κανένα άλλο σύστημα αναζήτησης για τη λειτουργία του.

2.4.1 Cobot

Οι (Sahay, Venkatesh, & Ram, 2009) έχουν μια διαφορετική προσέγγιση στην αναζήτηση πληροφορίας· εμπνέονται από μελέτες κοινωνικής ψυχολογίας που επιβεβαιώνουν πως η συνεργατική αναζήτηση πληροφορίας βελτιώνει τη γνωστική απόδοση. Η αναζήτηση πληροφορίας στον Παγκόσμιο Ιστό είναι ως επί το πλείστον μια μοναχική δραστηριότητα, αλλά το σύστημά τους, το Cobot, είναι μια διαλογική πλατφόρμα, όπου οι χρήστες συζητούν μεταξύ τους (και κατ' επέκταση εκφράζουν και τις ερωτήσεις τους σε φυσική γλώσσα) κι ο πράκτορας του συστήματος κατανοεί (μέσω εννοιών που εκφράζονται σε κατάλληλες οντολογίες) τις συζητήσεις. Ο πράκτορας βασίζεται στις εμπειρίες του για να βελτιώσει τις επιδόσεις του όχι ταιριάζοντας τις ερωτήσεις αυτές καθ' εαυτές σε ιστοσελίδες, αλλά κατανοώντας για ποιο είδος ερωτήσεων περιέχει κατάλληλες απαντήσεις μια ιστοσελίδα. Κατ' αυτόν τον τρόπο ο πράκτορας κατασκευάζει μοντέλα αλληλεπίδρασης χρηστών, διατηρεί σημασιολογικά δίκτυα που αντιπροσωπεύουν τη μνήμη του και βρίσκει πληροφορίες σχετικές με τις τρέχουσες συζητήσεις είτε ψάχνοντας στο WWW είτε ανατρέχοντας σε παρελθοντικές συζητήσεις.

Για την υλοποίηση της μνήμης του συστήματος οι (Sahay, Venkatesh, & Ram, 2009) ακολουθούν το μοντέλο “Socio-Semantic Conversation Net”, μια δυναμική δομή δεδομένων που παρέχει αναπαράσταση, ενημέρωση και επαναχρησιμοποίηση της γνώσης και της εμπειρίας του χρήστη.

2.5 Γνωσιακές διεπαφές

Όπως συζητήθηκε στο Κεφάλαιο 1, γνωσιακή αναζήτηση πληροφορίας μπορεί να υλοποιείται κι από συστήματα που το μόνο γνωσιακό κομμάτι που έχουν είναι η διεπαφή τους, δηλαδή ο τρόπος με τον οποίο παρουσιάζεται το σύστημα και τα αποτελέσματα της αναζήτησης στο χρήστη, αφήνοντας άορατο προς το χρήστη τον τρόπο με τον οποίο το σύστημα επιτελεί αυτά που επιτελεί. Συχνά χρησιμοποιείται η τεχνική της ομαδοποίησης (clustering), όπου τα αποτελέσματα της αναζήτησης παρουσιάζονται ομαδοποιημένα με έναν ή περισσότερους

τρόπους ούτως ώστε ανάμεσα στις πολλές προτεινόμενες ιστοσελίδες ο χρήστης να εντοπίζει γρηγορότερα τις σχετικές με το δικό του πλαίσιο αναζήτησης.

Αυτή η προσέγγιση κατά κάποιον τρόπο επιεικώς χαρακτηρίζεται γνωσιακή, καθώς αφορά μόνο ένα από τα στάδια της αναζήτησης πληροφορίας. Είναι όμως συνήθης, και σ' ό,τι αφορά την παρουσίαση των αποτελεσμάτων είναι συνεπής με τη γνωσιακή θεώρηση για την πρόσληψη πληροφορίας από το χρήστη.

2.5.1 Ταξινομητής διαστάσεων

Μια προσέγγιση στην ανάπτυξη διεπαφής αναζήτησης, η οποία ακολουθεί τη γνωσιακή γραμμή σκέψης, είναι της (Dimitrova M. , 2003), όπου χρησιμοποιούνται ευριστικά από τη Γνωσιακή Επιστήμη. Συγκεκριμένα, ανέπτυξε έναν ταξινομητή ιστοσελίδων που τις κατηγοριοποιεί βάσει δύο διαστάσεων: του επιπέδου εμπειρογνωμοσύνης και της ποσότητας λεπτομερειών που περιέχουν. Το αποτέλεσμα είναι πως οι χρήστες εντοπίζουν ευκολότερα, ανάμεσα στα αποτελέσματα, τις ιστοσελίδες των οποίων η γλώσσα περιγραφής είναι καταλληλότερη για τις δικές τους γνωστικές ανάγκες.

Οι μετρικές που χρησιμοποιήθηκαν γι' αυτές τις διαστάσεις βασίζονται σε συχνότητες εμφάνισης (αγγλικών) λέξεων μακρύτερων από ορισμένο μήκος γραμμάτων, οι οποίες (συχνότητες) έχουν μετρηθεί στα πλαίσια γνωσιακών ερευνών. Αλλά κι οι ίδιες οι διαστάσεις αντιστοιχούν στα χαρακτηριστικά που έπειτα από έρευνες συμπεριφοράς χρηστών προέκυψε πως οι ίδιοι αντιλαμβάνονται ως σημαντικά για την ποιότητα ενός εγγράφου. Μέσω πειραμάτων με απλούς χρήστες και αξιολόγησης του ταξινομητή από ειδικό στο σχεδιασμό ιστοσελίδων προέκυψε ότι η κατηγοριοποίηση των ιστοσελίδων βάσει ποσότητας λεπτομερειών ήταν επιτυχέστερη (δηλαδή πλησιέστερη στη γνώμη των χρηστών) απ' ό,τι η κατηγοριοποίησή τους βάσει επιπέδου εμπειρογνωμοσύνης. Επίσης, λειτουργούσε καλύτερα για μακροσκελείς παρά για συνοπτικές ιστοσελίδες. Παρά ταύτα, σε αρκετά μεγάλο βαθμό ο ταξινομητής ανταποκρίνεται σε γνωσιακά χαρακτηριστικά του έργου της αναζήτησης πληροφορίας σε ό,τι αφορά το στάδιο της παρουσίασης των αποτελεσμάτων στο χρήστη.

Ο ταξινομητής λοιπόν, όσο απλός κι αν είναι στην υλοποίηση και στη σύλληψή του, είναι ωστόσο συνεπής με μελετημένα γνωστικά φαινόμενα και μπορεί να θεωρηθεί ως εργαλείο γνωσιακής αναζήτησης πληροφορίας.

2.5.2 ConSearch

Το ConSearch (Lee & Lieberman, 2004) είναι μια διεπαφή αναζήτησης που εμπνέεται από τις δουλειές των (Marchionini, 1995) και (Ellis & Haugan, 1997) για την αναζήτηση πληροφορίας, που υποθέτουν έξι κύρια στάδια στην αναζήτηση πληροφορίας: (1)εκκίνηση, (2)καθήλωση, (3)επισκόπηση, (4)ξεχώρισμα, (5)επόπτευση, και (6)απόσπαση πληροφορίας. Το ConSearch αποπειράται να καλύψει τα 3 από τα 6 στάδια (επισκόπηση, ξεχώρισμα & απόσπαση πληροφορίας) που συνιστούν και μεγάλο γνωστικό φόρτο για τους χρήστες, ενώ οι συμβατικές μηχανές αναζήτησης καλύπτουν συνήθως μόνο 1 ή 2 απ' αυτά. Η κύρια ιδέα είναι να παρουσιάζονται ομαδοποιημένα ανά έννοιες τα αποτελέσματα της αναζήτησης. Γι' αυτόν το σκοπό χρησιμοποιούνται η OpenMind και το ConceptNet, μια βάση γνώσης κι ένας επεξεργαστής φυσικής γλώσσας αντίστοιχα. Η OpenMind περιέχει προτάσεις κοινής λογικής που εκφράζουν γεγονότα. Το ConceptNet επεξεργάζεται το περιεχόμενο των ιστοσελίδων που προκύπτουν ως σχετικές από κάποια εμπορική μηχανή αναζήτησης και το αντιστοιχίζει σε έννοιες χρησιμοποιώντας έναν αλγόριθμο διαδιδόμενης ενεργοποίησης (Anderson, 1983).

2.5.3 Ομαδοποιητής ιστοσελίδων βάσει σχετικότητας

Έχοντας κατά νου πως η σχετικότητα μιας ιστοσελίδας με την πληροφοριακή ανάγκη ενός χρήστη είναι μια πολύ-τιμη ποσότητα, οι (Desai & Spink, 2005) ανέπτυξαν έναν αλγόριθμο που ομαδοποιεί τα αποτελέσματα αναζήτησης τα οποία προκύπτουν από μια εμπορική μηχανή αναζήτησης (AlltheWeb.com) σε 3 ομάδες: σχετικά, μερικώς σχετικά, και άσχετα. Το πρόβλημα των εμπορικών μηχανών αναζήτησης είναι πως η κατάταξη των αποτελεσμάτων δε συμπίπτει με την αίσθηση των χρηστών για τη σχετικότητά τους. Γι' αυτό οι (Desai & Spink, 2005) κατασκεύασαν μια διεπαφή που μπορεί να ενσωματωθεί σε οποιοδήποτε εμπορικό σύστημα αναζήτησης, κι η οποία διευκολύνει και τους αρχάριους και

τους πεπειραμένους χρήστες στη διάκριση των σχετικότερων αποτελεσμάτων. Τα ευριστικά που χρησιμοποιήσαν βασίζονται σε συντακτικά στοιχεία του εγγράφου (συχνότητα εμφάνισης των λέξεων-κλειδιών σε διάφορα σημεία, θέση των λέξεων-κλειδιών μέσα στο έγγραφο), όμως το σημείο εκκίνησης των ερευνητών είναι η θεωρία των (Sperber & Wilson, 1986), η οποία μεταξύ άλλων αναγνωρίζει περισσότερες από δύο τιμές για τη σχετικότητα.

2.5.4 Media Watch on Climate Change

Το Media Watch on Climate Change (Hubmann-Haidvogel, Scharl, & Weichselbraun, 2009) είναι ένα εξειδικευμένο σύστημα αναζήτησης για κλιματολογικά θέματα, όμως ενσωματώνει πολλές ενδιαφέρουσες τεχνολογίες και λογικές της Γνωσιακής Επιστήμης, οι οποίες μπορούν ν' αποτελέσουν την πλατφόρμα για την ανάπτυξη γενικότερου σκοπού συστημάτων αναζήτησης. Το Media Watch on Climate Change χρησιμοποιεί «πολλαπλές όψεις» (multiple views): εκτός από τα κλασικά παράθυρα με αυτήν καθεαυτή την ιστοσελίδα και σχετικές ιστοσελίδες, στη διεπαφή περιλαμβάνονται:

- Σημασιολογικοί χάρτες που δείχνουν, σαν σε γεωγραφικό χάρτη, διάφορες κλιματολογικές έννοιες και το πόσο πολλές (όρη) ή πόσο λίγες (κοιλιάδες) ιστοσελίδες σχετικές με την κάθε έννοια έχουν εντοπιστεί,
- Οντολογίες των κλιματολογικών εννοιών, που εμφανίζονται με την κλασική μορφή των σημασιολογικών δικτύων, και
- «Σύννεφα ετικετών» (cloud tags), όπου οι έννοιες εμφανίζονται ως λέξεις (ετικέτες) με αυτές που στο σημασιολογικό χάρτη είναι τα όρη να εμφανίζονται με μεγαλύτερη γραμματοσειρά και μαύρα γράμματα, και τις «υψομετρικά χαμηλότερες» έννοιες να εμφανίζονται με μικρότερη γραμματοσειρά και γκριζα γράμματα.

Όπως αναφέρουν οι (Hubmann-Haidvogel, Scharl, & Weichselbraun, 2009, σσ. 1814-1815) από προηγούμενες έρευνες έχει προκύψει ότι με τις πολλαπλές αναπαραστάσεις της πληροφορίας, οι οποίες υλοποιούνται από τις διαφορετικές «όψεις», καθώς και με την ενημέρωση όλων των υπολοίπων όψεων κάθε φορά

που υπάρχει κάποια αλλαγή σε κάποια από αυτές, διευκολύνεται το έργο της αναζήτησης πληροφορίας στον WWW.

2.6 Από την πληροφορία στη γνώση

Όπως αναφέρθηκε στο Κεφάλαιο 1, μια πλήρως γνωσιακή προσέγγιση στην αναζήτηση πληροφορίας θα πρέπει να υλοποιεί και το άλμα από την πληροφορία στη γνώση. Το σύστημα που παρουσιάζεται στη συνέχεια, αν κι επικεντρωμένο σε συγκεκριμένο γνωστικό πεδίο, είναι μια ολοκληρωμένη προσπάθεια για την υλοποίηση αυτού του άλματος.

2.6.1 Met.a.ware

Μια εφαρμογή που ασχολείται με τη μετά-γνωστική υποστήριξη των χρηστών για να μπορέσουν να αξιολογήσουν καλύτερα τα αποτελέσματα που τους επιστρέφει η εκάστοτε μηχανή αναζήτησης και να ελέγξουν το επίπεδο κατανόησής τους, είναι το met.a.ware, που αναπτύχθηκε από τους (Stadtler & Bromme, 2008). Μέσα από πολλές έρευνες έχει προκύψει ότι υπάρχει σημαντική συσχέτιση αυτών των μετά-γνωστικών διαδικασιών με την απόκτηση γνώσης. Το met.a.ware επιτρέπει στο χρήστη να αποθηκεύει τις πληροφορίες που βρίσκει σε διάφορες ιστοσελίδες στον Παγκόσμιο Ιστό και να τις αξιολογεί επιτόπου εισάγοντάς τες σε πεδία/οντολογικές κατηγορίες που αντιπροσωπεύουν το θέμα της αναζήτησης. Οι οντολογικές κατηγορίες επιλέχτηκαν μετά από συνεργασία με ειδικούς στο θέμα της αναζήτησης (εδώ, η χοληστερίνη) και με ανασκόπηση της σχετικής βιβλιογραφίας. Εμπλεκόμενος σε μια τέτοια διαδικασία αξιολόγησης και κατηγοριοποίησης, ουσιαστικά, της πληροφορίας, ο χρήστης καθοδηγείται τελικά από την αναπαράσταση που δίνει ο ίδιος στην πληροφορία, για να οξύνει τις ουσιαστικότερες πλευρές της και να αμβλύνει τις πιο επουσιώδεις, ώστε να αναδείξει και να παγιώσει στο νου του τη γνώση που προκύπτει από αυτό το «ξεκαθάρισμα». Αυτό το αποτέλεσμα συνάδει με τη θεωρία της Αναπαραστασιακής Καθοδήγησης (Representational Guidance) του Suthers στο χώρο της εκπαιδευτικής ψυχολογίας.

2.7 Γνωσιακά vs Σημασιολογικών Συστημάτων Αναζήτησης

Μέχρι στιγμής είναι προφανές πως τα γνωσιακά συστήματα αναζήτησης πληροφορίας στον Παγκόσμιο Ιστό χρησιμοποιούν πολλές σημασιολογικές τεχνολογίες, όπως είναι τα σημασιολογικά δίκτυα ή οι γλώσσες στις οποίες εκφράζονται τα δεδομένα (π.χ. RDFS, OWL και οι διάφορες παραλλαγές τους) κ.λπ. Αυτό που τα διαφοροποιεί από τα καθαρά σημασιολογικά συστήματα είναι η πηγή της έμπνευσης των κατασκευαστών τους και ο προσανατολισμός τους προς μια γνωσιακά εμπειριστατωμένη θεώρηση των πληροφοριακών συστημάτων, που περιλαμβάνει συστήματα που αυτά προσαρμόζονται στη (φυσιολογική) γνωστική ροή των χρηστών κι όχι το αντίθετο, δηλαδή οι χρήστες να χρειάζεται να προσαρμόζουν τις γνωστικές τους λειτουργίες στις απαιτήσεις των συστημάτων. Βέβαια, ο άνθρωπος αλληλεπιδρά με το οποιοδήποτε περιβάλλον μέσα στο οποίο βρίσκεται, κι η γνωστική του κατάσταση φυσιολογικά μεταβάλλεται και διαμορφώνεται απ' αυτήν την αλληλεπίδραση, όμως συχνά στα πληροφοριακά συστήματα (υποκατηγορία των οποίων είναι οι μηχανές αναζήτησης) η αλληλεπίδραση είναι σημαντικά ξένη προς τον ανθρώπινο τρόπο λειτουργίας. Στη σημασιολογική αναζήτηση γίνεται ταίριασμα των εννοιών της επερώτησης προς τις έννοιες των ιστοσελίδων, ενώ στη γνωσιακή αναζήτηση γίνεται συνολικότερο ταίριασμα των πληροφοριακών αναγκών του χρήστη προς τις ιστοσελίδες που τις ικανοποιούν.

Η γνωσιακή θεώρηση των πληροφοριακών συστημάτων δεν είναι αυθαίρετη, αλλά προκύπτει από μελέτες της συμπεριφοράς των χρηστών κατά τη γενικότερη αλληλεπίδρασή τους με τον Η/Υ και κατά την επιτέλεση συγκεκριμένων έργων αναζήτησης πληροφορίας. Όλα τα συστήματα που παρουσιάστηκαν χαρακτηρίστηκαν γνωσιακά επειδή βασίζονται σαφώς σε κάποια θεωρία ή σε κάποιο μοντέλο, εντός του χώρου της Γνωσιακής Επιστήμης, για τη λειτουργία του νου κατά το έργο της αναζήτησης πληροφορίας.

Κεφάλαιο 3: Αποτίμηση των Παραπάνω Τάσεων στο Γνωσιακό Χώρο

Σ' αυτό το κεφάλαιο γίνεται μια αποτίμηση των γνωσιακών συστημάτων που παρουσιάστηκαν. Η αποτίμηση αφορά και τις επιδόσεις του κάθε συστήματος ξεχωριστά, και τα προβλήματα που εμπλέκονται σ' αυτήν τη διαδικασία. Έτσι, θα ακολουθήσουν τέσσερις ενότητες: η πρώτη αφορά την έννοια της σχετικότητας των αποτελεσμάτων/ιστοσελίδων που προσφέρονται ως λύσεις στις γνωστικές ανάγκες ενός χρήστη. Η δεύτερη αφορά την ικανότητα των συστημάτων αναζήτησης να βρίσκουν πληροφορίες που αναπαρίστανται με πολυμεσική μορφή. Η τρίτη αφορά τις μετρικές των επιδόσεων των συστημάτων αναζήτησης. Η τελευταία προχωράει σε αξιολόγηση των επιδόσεων των συγκεκριμένων γνωσιακών συστημάτων, βάσει των γραφόμενων των ίδιων των δημιουργών τους.

3.1 Μια Θεωρία της Σχετικότητας

Σ' αυτό το σημείο πρέπει να εξηγήσουμε ένα πρόβλημα εγγενές στο σχεδιασμό των συστημάτων αναζήτησης, πράγμα που συνιστά μια γενική κριτική και των γνωσιακών συστημάτων αναζήτησης που παρουσιάστηκαν. Ήδη από την εισαγωγή αναφέρθηκαν τα στάδια της αναζήτησης πληροφορίας, ένα εκ των οποίων είναι ο εντοπισμός των σχετικών ιστοσελίδων. Αυτό μας εισάγει σε μια θεμελιώδη έννοια, της οποίας ο ορισμός είναι το Ιερό Γκράαλ πολλών ερευνών από διάφορους επιστημονικούς κλάδους: της έννοιας της σχετικότητας (relevancy).

Στη σύγχρονη επιστήμη η σχετικότητα αντιμετωπίζεται ακόμη όπως ο Θεός από τους θεολόγους: όπως αναφέρει ο (Saracevic, 1975, σ. 90) η σχετικότητα δεν ορίζεται και δε χρειάζεται να οριστεί, διότι είναι μια οικουμενικά και διαισθητικά κατανοητή έννοια· ωστόσο μπορούμε να διερευνήσουμε τις ιδιότητές της, τις οποίες παραθέτει ο ίδιος (Saracevic, 2007, σ. 13). Από τις πολλές θεωρίες που έχουν προταθεί για την έννοια της σχετικότητας, μία που είναι πλησίον σε μια γνωσιακή θεώρησή της οδήγησε σε ένα πρακτικό αποτέλεσμα: η «Θεωρία της

Σχετικότητας», που πρωτοδημοσιεύτηκε το 1986 (Sperber & Wilson, 1986) και πήρε την τελική της μορφή 18 χρόνια μετά (Wilson & Sperber, 2004). Αυτή η θεωρία βασίζεται στην υπόθεση πως «...ένα καίριο χαρακτηριστικό της επικοινωνίας είναι η έκφραση και η αναγνώριση προθέσεων [...] σχετική πληροφορία είναι η πληροφορία που αξίζει να προσλάβει κανείς» (Saracenic, 2007, σ. 17) και πρεσβεύει πως όσο μεγαλύτερη η σχετικότητα, τόσο μεγαλύτερος ο θετικός γνωστικός αντίκτυπος και τόσο μικρότερος ο γνωστικός φόρτος για το άτομο. Ένα σημαντικό σημείο αυτής της θεωρίας είναι η παραδοχή πως η σχετικότητα δεν είναι μια κατάσταση «άσπρο-μαύρο», αλλά έχει πολλά επίπεδα. Στα πλαίσια των πληροφοριακών συστημάτων, ο (White, 2007; White, 2007) εφήρμοσε τη Θεωρία της Σχετικότητας για τη δημιουργία μιας δισδιάστατης μετρικής που αντιπροσωπεύει το θετικό αντίκτυπο και το γνωστικό φόρτο και που μετράει τη λειτουργική σχετικότητα ενός συστήματος αναζήτησης. Αυτή η αντίληψη για τις μετρικές είναι πλησιέστερη στο δικό μου ορισμό των γνωσιακών συστημάτων, ο οποίος δόθηκε στο Κεφάλαιο 1.

Εν πάση περιπτώσει, ένας παρατηρητικός αναγνώστης θα έχει ήδη διακρίνει πως τα περισσότερα γνωσιακά συστήματα αναζήτησης που παρουσιάστηκαν δεν εντοπίζουν από μόνα τους τις σχετικές ιστοσελίδες, παρά ασχολούνται με το τρίτο και το τέταρτο στάδιο της αναζήτησης πληροφορίας, ήτοι την ταξινόμηση και την παρουσίαση των αποτελεσμάτων, πράγμα που εισάγει από μόνο του μια προβληματική. Αυτό οφείλεται στο ότι για διάφορους λόγους (οικονομικούς, χρονικούς κ.ά. όπως επισημαίνεται στον (Diaz, 2005)) η δουλειά αυτών των συστημάτων βασίζεται πάνω στις εμπορικές μηχανές αναζήτησης, δηλαδή το πρώτο στάδιο υλοποιείται από τις τελευταίες ενώ τα υπόλοιπα στάδια από τα συστήματα που παρουσιάστηκαν. Τα κριτήρια όμως που χρησιμοποιούν οι εμπορικές μηχανές αναζήτησης για τον εντοπισμό «σχετικών» ιστοσελίδων βασίζονται στο συντακτικό ταίριασμα επερώτησης-ιστοσελίδας, συν μερικά ευριστικά που όμως δεν ξεφεύγουν από τη λογική του συντακτικού ταιριάσματος, λογική που είναι ο πυρήνας της παραδοσιακής αναζήτησης πληροφορίας. Χωρίς να αμφισβητείται η αξία της προσπάθειας, αρκετά από αυτά για την ώρα ακολουθούν μάλλον απλουστευμένα μοντέλα της σχετικότητας (όσο είμαστε σε

θέση να κρίνουμε από εμπορικές μηχανές των οποίων ο πυρήνας λειτουργίας κρατείται ως επτασφράγιστο μυστικό για ευνόητους λόγους), αλλά όπως υπογραμμίζει ο (Saracevic, 2007, σ. 30) «...τα συστήματα αναζήτησης πληροφορίας έχουν ασχοληθεί με το πιο [...] απλοποιημένο μοντέλο και έκφραση της σχετικότητας [...] τώρα που έχουν ταυτοποιηθεί πολύ πολυπλοκότερα μοντέλα και εκφάνσεις της σχετικότητας, [...] η πρόκληση να μεταφραστούν σε έρευνα και πρακτική έχει αυξηθεί πάρα πολύ». Και το θετικό είναι πως στην πρόκληση της επίλυσης του «δύσκολου προβλήματος του υπολογισμού [...] της σχετικότητας» (Zemirli, Tamine-Lechani, & Boughanem, 2007) ανταποκρίνονται όλο και περισσότεροι ερευνητές.

3.2 Αναζήτηση πληροφορίας σε πολυμεσική μορφή

Ένα άλλο σημαντικό σημείο στο οποίο υστερούν τα γνωσιακά συστήματα αναζήτησης που παρουσιάστηκαν, είναι το γεγονός πως δεν προσφέρουν καμία δυνατότητα αναζήτησης πληροφορίας σε πολυμεσική μορφή – δηλαδή πληροφορίας που αναπαρίσταται όχι (μόνο) με κείμενο, αλλά με εικόνα, γραφικά, ήχο ή/και βίντεο. Φυσικά, αυτό δεν αποτελεί κριτική μόνο για τα υπό εξέταση συστήματα· είναι ένα γενικότερο πρόβλημα ακόμη και των εμπορικών μηχανών αναζήτησης, όπως αποδεικνύεται μέσα από την έρευνα των (Tjondronegoro & Sprink, 2008). Το πρόβλημα αυτό πηγάζει από το γεγονός πως η γλώσσα περιγραφής των πηγών πληροφορίας (εν προκειμένω, των πολυμέσων) δεν είναι κατάλληλη για τις υπάρχουσες μηχανές αναζήτησης. Οπωσδήποτε, δεδομένου αυτού του προβλήματος, συνεχώς αναπτύσσονται νέες τεχνικές (όπως η βασισμένη-στο-περιεχόμενο ανάσυρση εικόνας (content-based image retrieval)). Όμως, όπως επισημαίνουν οι (Picard, Cord, & Revel, 2008), αυτή και άλλες τεχνικές δεν είναι κατάλληλες για έναν τόσο άναρχο και ευρέως περιεχομένου χώρο, όπως ο Παγκόσμιος Ιστός. Πάντως, υπάρχουν αρκετά παραδοσιακά και σημασιολογικά συστήματα αναζήτησης πολυμέσων, ενώ τα γνωσιακά υστερούν σ' αυτόν τον τομέα. Εν πάση περιπτώσει, η πολλαπλότητα των μορφών

αναπαράστασης της πληροφορίας συνεισφέρει στην πρόσληψή της κι απαλύνει το γνωστικό φόρτο των χρηστών, κι αυτό είναι ένα σημείο το οποίο ακόμη δεν καλύπτεται από τα προαναφερθέντα συστήματα ή από οποιοδήποτε άλλο σύστημα που έχει υποπέσει στην αντίληψή μου και που θα μπορούσε να χαρακτηριστεί γνωσιακό.

3.3 Μετρικές για την Αξιολόγηση των Συστημάτων Αναζήτησης

Σε κάθε έρευνα που παρουσιάζει ένα καινούριο σύστημα αναζήτησης πληροφορίας (πρέπει να) υπάρχει κι ένα κομμάτι που με κάποιον τρόπο μετρά τις επιδόσεις του συστήματος. Για ν' αξιολογήσει κανείς ένα σύστημα πρέπει να το συγκρίνει με κάτι· αυτό το κάτι μπορεί να είναι κάποιο(α) μέτρο(α)· μπορεί να είναι ένα άλλο σύστημα που χρησιμοποιείται ως σημείο αναφοράς. Στη μεν περίπτωση, το πρόβλημα είναι πως οι μετρικές που χρησιμοποιούνται είναι αρκετά αυθαίρετες, υπό την έννοια πως κατά καιρούς οι ερευνητές χρησιμοποιούν διάφορα μέτρα για διαφορετικές διαστάσεις των συστημάτων αναζήτησης. Η επιλογή των μετρικών έχει να κάνει με την έννοια και τις διαστάσεις της σχετικότητας, προβλήματα για τα οποία οι επιστήμονες δεν έχουν καταλήξει σε κοινή γραμμή πλευσης, όπως εξηγήθηκε στην ενότητα 3.1. Στη δε περίπτωση, το σύστημα που χρησιμοποιείται ως σημείο αναφοράς είναι κάποια παραδοσιακή (π.χ. Google ή Yahoo) ή σημασιολογική (π.χ. Lucene) μηχανή αναζήτησης, όμως και πάλι, δεν έχει καθοριστεί από την επιστημονική κοινότητα ένα συγκεκριμένο σύστημα αναζήτησης ως σημείο αναφοράς.

Κατά καιρούς έχουν προταθεί διάφορα μέτρα αξιολόγησης, όπως: ακρίβεια (precision), ανάκληση (recall), αποτελεσματικότητα (efficiency), χρησιμότητα (utility), ικανοποίηση χρήστη (user satisfaction), εμπιστοσύνη χρήστη (user confidence), πληροφορισιμότητα (informativeness) κ.ά. Συχνότερα χρησιμοποιούμενα από αυτά είναι η ακρίβεια (οριζόμενη ως το πλήθος των σχετικών αποτελεσμάτων προς το συνολικό πλήθος των αποτελεσμάτων που

επιστράφηκαν από μια αναζήτηση) και η ανάκληση (οριζόμενη ως το πλήθος των σχετικών αποτελεσμάτων που επιστράφηκαν από μια αναζήτηση προς το συνολικό πλήθος των σχετικών αποτελεσμάτων που υπάρχουν). Παρά την ευρεία χρήση αυτών των μετρικών η αξιολογητική τους ικανότητα έχει αμφισβητηθεί από πολλούς ερευνητές για πολλούς λόγους, με τον κυριότερο – κατά τη γνώμη μου – να είναι το γεγονός πως προϋποθέτουν μια δίτιμη σχετικότητα, όταν σύμφωνα με τους (Sperber & Wilson, 1986) η σχετικότητα έχει πολλές τιμές.

Εν πάση περιπτώσει το ζήτημα της αξιολόγησης των συστημάτων αναζήτησης είναι από μόνο του μια διπλωματική και δεν είναι δυνατόν να αποτελέσει αντικείμενο περαιτέρω ανάλυσης στην παρούσα. Έτσι, στην επόμενη ενότητα θα γίνει μια αξιολόγηση των συστημάτων βάσει των γραφόμενων των ίδιων των δημιουργών τους, όπου υπάρχει κάτι τέτοιο. Αυτό, επειδή ορισμένα από τα συστήματα είναι σε φάση μερικής ή ολικής υλοποίησης, κι ενίοτε δεν υπάρχει κανένα πείραμα που να μετράει τις επιδόσεις τους.

3.4 Αξιολόγηση των γνωσιακών συστημάτων αναζήτησης που παρουσιάστηκαν

3.4.1 Σύστημα σημασιολογικής αναζήτησης βασισμένο σε γνωσιακό μοντέλο

Για τη μέτρηση της επίδοσης του συστήματός τους, οι (Wechsler, Baier, Nussbaum, & Baeza-Yates, 2004) το σύγκριναν με το Google σε ένα πρωτόλειο πείραμα. Θέλησαν να διαπιστώσουν αν με μία ασαφή, κακοδιατυπωμένη επερώτηση (για μια ωστόσο συγκεκριμένη, στο μυαλό του χρήστη, πληροφοριακή ανάγκη) θα είχαν ικανοποιητικά αποτελέσματα. Στο μεν Google από τα 44 πρώτα αποτελέσματα (υπό μορφή συνδέσμων) μόνο τα 6 ήταν σχετικά (αποτελεσματικότητα 13,6%), ενώ στο σύστημά τους ο γράφος αποτελεσμάτων που παρουσιάστηκε περιείχε 18 συνδέσμους με 6 σχετικά αποτελέσματα (αποτελεσματικότητα 33%). Εκ πρώτης όψεως το σύστημα

υπερτερεί του Google, αλλά προφανώς απαιτούνται αναλυτικότερα πειράματα για την ορθή αξιολόγηση του συστήματος.

3.4.2 Contextual Search Project

Όπως εξηγούν κι οι ίδιες οι (Freund & Toms, 2005) το σύστημά τους ήταν σε φάση υλοποίησης κατά τη δημοσίευση του άρθρου τους, αλλά είχαν την πρόθεση, μόλις ολοκληρωνόταν η υλοποίηση, να διενεργήσουν πειράματα αξιολόγησης της επίδοσής του μετρώντας την ανάκληση, την ακρίβεια, την αποτελεσματικότητα, την ικανοποίηση των χρηστών, τη χρηστικότητα, τη συνεισφορά στην ολοκλήρωση του έργου αναζήτησης και την ποιότητα των αποτελεσμάτων (Freund & Toms, 2005, σ. 11).

3.4.3 Outride

Για τη μέτρηση της επίδοσης του Outride επιλέχθηκαν ως κριτήρια ο χρόνος που χρειάστηκε για να ολοκληρωθεί ένα έργο αναζήτησης και το πλήθος των ενεργειών εισόδου ενός χρήστη (δηλαδή το πόσες φορές χρησιμοποίησε το πληκτρολόγιο ή το ποντίκι για να εισάγει δεδομένα ή να μεταβεί σε κάποια ιστοσελίδα). 48 συμμετέχοντες με διαφορετικά επίπεδα εμπειρίας χρήσης του Διαδικτύου διεκπεραίωσαν από 12 έργα αναζήτησης στο Outride και σε μία από τις εξής εμπορικές μηχανές αναζήτησης: AOL Search, Excite, Google, Yahoo. Από τις μετρήσεις προέκυψαν τα εξής ενδιαφέροντα στοιχεία, ανεξαρτήτως επιπέδου εμπειρίας χρηστών:

- Κατά μέσο όρο ο χρόνος ολοκλήρωσης ενός έργου αναζήτησης ήταν μικρότερος στο Outride απ' ό,τι σε οποιοδήποτε από τα άλλα συστήματα.
- Στο Outride το πλήθος των ενεργειών εισόδου εκ μέρους του χρήστη ήταν μικρότερο απ' ό,τι σε οποιοδήποτε από τα άλλα συστήματα.

Βάσει, λοιπόν, των δύο μέτρων που επιλέχθηκαν για την αξιολόγησή του, το Outride υποσκελίζει τις εμπορικές μηχανές αναζήτησης (Pitkow, και συν., 2002, σσ. 53-54). Εδώ να σημειωθεί πως το Outride εξαγοράστηκε το Σεπτέμβριο του 2001 από την Google.

3.4.4 Search-in-Synchrony

Για τη μέτρηση της επίδοσης του Search-in-Synchrony επιλέχθηκαν ως κριτήρια ο χρόνος που χρειάστηκε για να ολοκληρωθεί ένα έργο αναζήτησης και το πλήθος των ενεργειών εισόδου ενός χρήστη (δηλαδή το πόσες φορές χρησιμοποίησε το ποντίκι για να μεταβεί σε κάποια ιστοσελίδα). Ως συμμετέχοντες χρησιμοποιήθηκαν 10 ερευνητές που έπρεπε να βρουν επαρκείς πληροφορίες για 4 θέματα αναζήτησης. Το έργο αναζήτησης πραγματοποιήθηκε στο Search-in-Synchrony και στη μηχανή αναζήτησης Google. Από τους μέσους όρους προέκυψε πως σε κάθε περίπτωση η αναζήτηση μέσω του Search-in-Synchrony είναι ταχύτερη και λιγότερο κοπιώδης απ' ό,τι μέσω του Google. Οι (Dhir & Lee, 2009) δεν έχουν όμως πραγματοποιήσει περαιτέρω στατιστικές αναλύσεις που να δείχνουν εάν αυτές οι διαφορές στους μέσους όρους είναι στατιστικά σημαντικές.

3.4.5 OntoSearch_U

Οι επιδόσεις του OntoSearch_U μετρήθηκαν σε 2 πειράματα, όπου συμμετείχαν 10 άτομα. Κάθε άτομο έκανε 2 επερωτήσεις στο OntoSearch_U, τη μία για την εκπαίδευση του συστήματος (στη δημιουργία της οντολογίας χρήστη) και την άλλη για τον έλεγχο του. Στη συνέχεια αξιολογήθηκε από κάθε χρήστη η ακρίβεια (precision) των 5 και 10 πρώτων ιστοσελίδων που επιστρέφονταν για συγκεκριμένα θέματα αναζήτησης από το OntoSearch_U, το OntoSearch (ίδιο με το OntoSearch_U αλλά χωρίς δυνατότητα εξατομίκευσης), το OntoSearch_C (ίδιο με το OntoSearch_U αλλά μόνο με εκμάθηση εννοιών και χωρίς εκμάθηση σχέσεων για την κατασκευή του προφίλ χρήστη), το Google Directory και το Lucene. Μέσα από t-test επιβεβαιώθηκε η υπόθεση πως, συγκρινόμενο με αυτά τα συστήματα, το OntoSearch_U βελτιώνει την εμπειρία αναζήτησης στο Σημασιολογικό Ιστό (Jiang & Tan, 2009, σσ. 2803-2804, 2805-2807).

3.4.6 Cognitive web agents

Η πρόταση των (Dimitrova, Barakova, Lourens, & Radeva, 2004; Dimitrova, Wagatsuma, & Yamaguchi, 2007) είναι συνεπής με επικρατείς θεωρίες για τη λειτουργία της αυτοβιογραφικής μνήμης, κι από πειράματα και δοκιμές που

έκαναν οι ερευνητές για τον έλεγχο του μοντέλου τους φαίνεται πως η συμπεριφορά του προτεινόμενου πράκτορα προσομοιάζει σε αυτήν των χρηστών κατά την αναζήτηση πληροφορίας μέσω μηχανών αναζήτησης, ούτως ώστε ο πράκτορας μπορεί να υποστηρίξει κάποια εξατομικευμένη διεπαφή αναζήτησης. Το πρόβλημα είναι πως οι ερευνητές δεν έχουν αναπτύξει κάποιο ολοκληρωμένο σύστημα αναζήτησης, ούτε καν έχουν ενσωματώσει τον πράκτορα τους σε κάποιο υπάρχον· φαίνεται πως είχαν κατά νου να ενσωματωθεί στο σύστημα WebPersonae των (McGowan, Kushmerick, & Smyth, 2002), το οποίο κάνει εξατομίκευση προφίλ χρήστη δημιουργώντας πολλαπλές «περσόνες» που αντιπροσωπεύουν τα ενδιαφέροντα ενός (ίδιου) χρήστη πάνω σε διαφορετικά γνωστικά αντικείμενα. Η εργασία τους έχει προοπτικές ανάπτυξης μεν, είναι αποσπασματική δε, οπότε είναι ουσιαστικά αδύνατον να την αξιολογήσει κανείς εφόσον δεν υλοποιείται σε κάποιο σύστημα.

3.4.7 RASCALLI

Στους (Krenn, 2008; Krenn & Schollum, 2008) δε γίνεται μέτρηση της επίδοσης κάποιου συστήματος αναζήτησης που χρησιμοποιεί τους πράκτορες Rascalli, διότι είναι ακόμη σε εξέλιξη ως ερευνητικό έργο.

3.4.8 Cobot

Το σύστημα Cobot των (Sahay, Venkatesh, & Ram, 2009) φαίνεται μια ολοκληρωμένη πρόταση για τη συνεργατική αναζήτηση πληροφορίας, όπου οι ερευνητές ασχολούνται με όλα τα θέματα· από την αναπαράσταση γνώσης έως την εκπαίδευση του συστήματος. Είναι όμως ακριβώς αυτό· μια θεωρητική πρόταση που δεν έχει πλήρως υλοποιηθεί, οπότε δεν μπορεί ν' αξιολογηθεί μέσω πειραματισμού.

3.4.9 Ταξινομητής διαστάσεων

Ο ταξινομητής διαστάσεων της (Dimitrova M. , 2003) είναι μια ελπιδοφόρα προσπάθεια της οποίας όμως οι επιδόσεις, όταν δημοσιεύτηκε το άρθρο, δεν είχαν ακόμη μετρηθεί βάσει πειραμάτων.

3.4.10 ConSearch

Μέχρι σήμερα οι (Lee & Lieberman, 2004) δεν έχουν κάνει κάποιο πείραμα που να αξιολογεί την επίδοση του ConSearch, πράγμα που δικαιολογείται απ' το ότι το ConSearch αναπτύχθηκε στα πλαίσια ενός μεταπτυχιακού μαθήματος κι απ' ό,τι έχει υποπέσει στην αντίληψή μου δεν έχει υπάρξει κάποια εξέλιξη στην ανάπτυξή του.

3.4.11 Ομαδοποιητής ιστοσελίδων βάσει σχετικότητας

Για να διαπιστώσουν εάν οι εκτιμήσεις του αλγορίθμου τους για τη σχετικότητα των αποτελεσμάτων συμπίπτουν με τις εκτιμήσεις των χρηστών, οι (Desai & Spink, 2005) ζήτησαν από 5 φοιτητές να αξιολογήσουν τη σχετικότητα των αποτελεσμάτων για μια συγκεκριμένη επερώτηση, για την οποία η μηχανή αναζήτησης AlltheWeb.com επέστρεψε 98 αποτελέσματα. Οι εκτιμήσεις των χρηστών και οι εκτιμήσεις του αλγορίθμου για τις ομάδες των «σχετικών» και «άσχετων» αποτελεσμάτων συμφωνούσαν περισσότερο, απ' ό,τι οι εκτιμήσεις τους για τα «μερικώς σχετικά» αποτελέσματα. Πάντως, με αυτήν την ομαδοποίηση, ο αλγόριθμος εξυπηρετεί και τους χρήστες που είναι ειδικοί και τους χρήστες που είναι αρχάριοι σε κάποιο γνωστικό πεδίο, κατευθύνοντάς τους γρηγορότερα στην ομάδα των ιστοσελίδων που ικανοποιούν καλύτερα τις γνωστικές τους ανάγκες.

3.4.12 Media Watch on Climate Change

Το Media Watch on Climate Change είναι μια διεπαφή αναζήτησης για εξειδικευμένο θέμα (το κλίμα) και είναι μάλλον ανούσιο να συγκριθεί με κάποια γενικού σκοπού μηχανή αναζήτησης. Εν πάση περιπτώσει οι (Hubmann-Haidvogel, Scharl, & Weichselbraun, 2009) δεν έχουν μέχρι στιγμής μετρήσει με κάποιον τρόπο τις επιδόσεις του συστήματός τους, πράγμα που αποτελεί ένα μειονέκτημα ως προς την αξιοπιστία του· η πίστη γι' αυτήν μένει ν' αποδειχθεί επιστημονικά στο μέλλον.

3.4.13 Met.a.ware

Το met.a.ware, στη μορφή με την οποία παρουσιάζεται από τους (Stadtler & Bromme, 2008), προσφέρει όχι απλά δυνατότητα αναζήτησης αλλά και μετά-γνωστική υποστήριξη για ένα μόνο εξειδικευμένο θέμα αναζήτησης (για τη χοληστερίνη) σε χρήστες με λίγες ή μέτριες γνώσεις για τη χοληστερίνη. Για τη μέτρηση της αποτελεσματικότητάς του, σε 118 φοιτητές δόθηκαν 15 προεπιλεγμένοι ιστότοποι από τους οποίους έπρεπε να βρουν απαντήσεις για μια ιατρική περίπτωση. Οι συμμετέχοντες κατανεμήθηκαν τυχαία σε έξι ομάδες, τέσσερις εκ των οποίων χρησιμοποιούσαν το met.a.ware αλλά λάμβαναν διαφορετική μετά-γνωστική υποστήριξη απ' αυτό, μία κρατούσε σημειώσεις σε χαρτί και μία κρατούσε σημειώσεις σε απλό κειμενογράφο στον Η/Υ. Ως μέτρα της επίδοσης του συστήματος χρησιμοποιήθηκαν η κατανόηση του θέματος, η αξιολόγηση των πηγών πληροφορίας και η τεκμηρίωση των αποφάσεων των συμμετεχόντων σχετικά με τις απαντήσεις όπου κατέληξαν για το θέμα αναζήτησης (τη χοληστερίνη).

Από τη στατιστική ανάλυση των αποτελεσμάτων προέκυψε πως σε σχέση με τις ομάδες ελέγχου οι χρήστες του met.a.ware απέκτησαν περισσότερες γνώσεις επί του θέματος, καλύτερη αξιολόγηση των πηγών πληροφορίας αλλά όχι ουσιαστικότερη τεκμηρίωση των αποφάσεών τους. Για περισσότερες λεπτομέρειες σχετικά με τα αποτελέσματα και πιθανές εξηγήσεις, μπορεί κανείς να ανατρέξει στους (Stadtler & Bromme, 2008, σσ. 727-734).

Κεφάλαιο 4: Μια Πρόταση για τη Γνωσιακή Αναζήτηση Πληροφορίας

Σ' αυτό το κεφάλαιο πρώτα παρουσιάζεται μια διαφορετική προσέγγιση στη γνωσιακή αναζήτηση πληροφορίας κι έπειτα γίνεται μια δική μου πρόταση για το ζήτημα, παρουσιάζοντας μια πιθανή αρχιτεκτονική για ένα γνωσιακό σύστημα αναζήτησης πληροφορίας στον Παγκόσμιο Ιστό.

4.1 Μια διαφορετική προσέγγιση

Η γνωσιακή αναζήτηση πληροφορίας προσπαθεί να υλοποιήσει μηχανές αναζήτησης που προσφέρουν διεπαφή και αποτελέσματα αναζήτησης μ' έναν τρόπο «φυσικό» για το χρήστη. Υπάρχουν ωστόσο έρευνες που υποδεικνύουν ότι υπάρχει ένας φυσικότερος τρόπος για την ανεύρεση πληροφορίας στον Παγκόσμιο Ιστό, κι αυτός δεν είναι μέσω των μηχανών αναζήτησης αλλά μέσω της πλοήγησης στον WWW (Juvina I. , Oostendorp, van, & Pauw, 2005). Σύμφωνα μ' αυτήν τη λογική, δεδομένου ότι η φυσική τάση των χρηστών είναι να ξεκινούν από μια οικεία ιστοσελίδα (π.χ. από κάποιον κατάλογο του Ιστού (web directory) ή από κάποια ιστοσελίδα-σελιδοδείκτη (bookmark)) κι ακολουθώντας συνδέσμους να φτάνουν στον προορισμό τους αντί να περιμένουν έτοιμες λίστες ιστοσελίδων προς επίσκεψη, ίσως ο γνωσιακός τρόπος για την αναζήτηση πληροφοριών θα ήταν η προσφορά υποστήριξης πλοήγησης κι όχι μια μηχανή αναζήτησης που επιστρέφει ιστοσελίδες ως αποτελέσματα. Οι (Magoulas, Chen, & Dimakopoulos, 2004) έχουν σχεδιάσει μια τέτοια εξατομικευμένη διεπαφή για καταλόγους του Ιστού, η οποία ποικίλλει ανάλογα με το γνωστικό στυλ κάθε χρήστη. Κάτι τέτοιο κάνουν επίσης οι Letizia και PowerScout των (Lieberman, Fry, & Weitzman, 2001), μόνο που η βάση των δημιουργών τους δεν πατάει σε ευρήματα της Γνωσιακής Επιστήμης, παρά είναι διαισθητική. Σύμφωνα με τη λογική της Γνωσιακής Επιστήμης είναι ορισμένα μοντέλα προσομοίωσης της συμπεριφοράς χρηστών κατά την πλοήγηση στον WWW, όπως το CoLiDeS, το

CoLiDeS+ και το SNIF-ACT³. Αυτά μπορούν να χρησιμοποιηθούν υποστηρικτικά για την πλοήγηση των χρηστών.

Χωρίς να αμφισβητώ την υποστήριξη που προσφέρουν τέτοια συστήματα στους χρήστες, επιμένω στο εξής γεγονός: το πρόβλημα είναι πως με τα δεδομένα στον Παγκόσμιο Ιστό να αυξάνονται αδιάκοπα με ταχύτατους ρυθμούς (λόγω και της ανάπτυξης του Web 2.0), είναι μάλλον αδύνατον για τους χρήστες να αναζητήσουν πληροφορίες με πλοήγηση παρά με χρήση μηχανών αναζήτησης. Κατά τη γνώμη μου, τέτοιες ερευνητικές προσπάθειες είναι χρησιμότερο να ενσωματωθούν σε συστήματα αναζήτησης πληροφορίας στον WWW ως ένα επιπλέον εργαλείο που θα υποστηρίζει το χρήστη άπαξ κι αποφασίσει να επιδοθεί σε πλοήγηση στον Παγκόσμιο Ιστό. Για παράδειγμα, θα μπορούσαν να αξιοποιηθούν ως μια πρόσθετη λειτουργία, η οποία θα ενεργοποιείται αφού το σύστημα αναζήτησης έχει παρουσιάσει κάποια αποτελέσματα στο χρήστη. Πατώντας ο τελευταίος επάνω σε κάποιο αποτέλεσμα, ενδεχομένως να θελήσει να περιηγηθεί σε κάποια από τις υπερσυνδέσεις που περιέχει η τρέχουσα ιστοσελίδα. Σ' αυτό το σημείο θα μπορούσε να «επέμβει» το γνωσιακό μοντέλο, καθοδηγώντας το χρήστη σε ένα συντομότερο και αποδοτικότερο – ως προς την κάλυψη των πληροφοριακών αναγκών του – μονοπάτι πλοήγησης.

4.2 Η πρότασή μου

Σε ένα σημείο του άρθρου τους οι (Wei, Barnaghi, & Bargiela, 2008, σ. 81) αναφέρουν πως «Η Σημασιολογική αναζήτηση είναι ένα διεπιστημονικό πεδίο που φέρνει σε επαφή έρευνα από κοινότητες του IR, του σημασιολογικού ιστού, της μηχανικής μάθησης, της επεξεργασίας φυσικής γλώσσας, της εξαγωγής πληροφορίας κ.λπ.» Δε θα μπορούσα να σκεφτώ τίποτε καλύτερο για μια πρόταση για τη γνωσιακή αναζήτηση πληροφορίας, από το να τους παραφράσω λέγοντας πως η γνωσιακή αναζήτηση πληροφορίας στον Παγκόσμιο Ιστό πρέπει να είναι ένα διεπιστημονικό πεδίο που φέρνει σε επαφή έρευνα από τις

³ Για περισσότερες λεπτομέρειες σχετικά μ' αυτά τα μοντέλα, μπορεί κανείς να ανατρέξει στους (Kitajima, Polson, & Blackmon, 2000), (Juvina & Oostendorp, 2008) και (Pirolli & Fu, 2003) αντίστοιχα.

κοινότητες της Γνωσιακής Επιστήμης κι όλων των άλλων που αναφέρθηκαν κι έχουν να κάνουν με το πληροφοριακό/τεχνολογικό υπόβαθρο. Αυτό μπορεί να περιλαμβάνει τη χρήση γνωσιακών μοντέλων και τεχνολογιών αλλά πάνω από όλα, συνεργασία ανάμεσα στα διάφορα επιστημονικά πεδία που συνθέτουν την κοινότητα της Γνωσιακής Επιστήμης και την κοινότητα της Ανάκτησης Πληροφορίας.

Μιλώντας επί του πρακτέου, θεωρώ πως βασικές τεχνολογίες υλοποίησης ενός γνωσιακού συστήματος αναζήτησης πρέπει να είναι οι οντολογίες και οι πράκτορες οι οποίοι όμως θα βασίζονται σε γνωσιακά μοντέλα.

Οι οντολογίες χρησιμοποιούνται σχεδόν σε όλα τα γνωσιακά συστήματα που παρουσιάστηκαν και φαίνεται πως αποτελούν μια ικανοποιητική προς το παρόν και με προοπτικές βελτίωσης πρόταση για την αναπαράσταση γνώσης και την περιγραφή δεδομένων. Το πλεονέκτημα των οντολογιών είναι πως προσομοιάζουν σε μια μοντελοποίηση των εννοιών η οποία προτιμάται από τους γνωσιακούς επιστήμονες. Είναι κοινή θεώρηση στη Γνωσιακή Επιστήμη πως στον ανθρώπινο νου γίνεται αναπαράσταση των εννοιών μέσω δικτύων, όπου αυτές συνδέονται μεταξύ τους, και η ενεργοποίηση της μιας προκαλεί την ενεργοποίηση και των εγγύτερων προς αυτήν μέσω των συνδέσεών τους. Το πάγιο πρόβλημα βέβαια με τις οντολογίες είναι πως απαιτείται επισήμανση (annotation) των πηγών-δεδομένων, η οποία σε ό,τι αφορά τον WWW είναι ανομοιογενής λόγω του ανομοιογενούς της προέλευσής τους, αλλά και υπάρχουν και αναπτύσσονται εργαλεία αυτόματης επισήμανσης των εγγράφων του Ιστού.

Για την πραγματοποίηση της εξερεύνησης του WWW και του εντοπισμού των σχετικών ιστοσελίδων, οι πράκτορες που βασίζονται σε γνωσιακά μοντέλα συνιστούν μια δυνατή υποψηφιότητα για ένα στιβαρό γνωσιακό σύστημα αναζήτησης πληροφορίας. Μέσα στην τελευταία δεκαετία υπάρχουν πολλές προσπάθειες για μοντελοποίηση της συμπεριφοράς πλοήγησης και αναζήτησης των χρηστών (π.χ. CoLiDeS, SNIF-ACT). Τέτοια μοντέλα μπορούν να ενσωματωθούν στον πυρήνα ενός γνωσιακού πράκτορα, κάνοντάς τον να προσομοιάζει περισσότερο σε έναν εμπειρογνώμονα της αναζήτησης

πληροφορίας του οποίου η εμπειρία μπορεί να χρησιμοποιηθεί για την εξυπηρέτηση του χρήστη.

Πέρα από τις δύο αυτές τεχνολογικές προσεγγίσεις, υπάρχουν ένα πλήθος από τεχνολογίες που αφορούν τη διεπαφή των συστημάτων αναζήτησης, τόσο στο κομμάτι της επερώτησης όσο και στο κομμάτι της παρουσίασης των αποτελεσμάτων. Το ποια προσέγγιση επιλέγεται δεν είναι, κατά τη γνώμη μου, τόσο κρίσιμο, μιας και υπάρχουν πλεονεκτήματα και μειονεκτήματα σε όλες. Το σίγουρο είναι πως οι διεπαφές των δημοφιλέστερων εμπορικών συστημάτων αναζήτησης (λέξεις-κλειδιά για την επερώτηση, και παρουσίαση των αποτελεσμάτων σε λίστα) δεν είναι γνωστικά εύληπτες. Η διατύπωση της επερώτησης σε φυσική γλώσσα ή υπό μορφή στόχου πληροφορίας, και η παρουσίαση των αποτελεσμάτων ομαδοποιημένων σε έννοιες, υφίστανται σε πραγματικά γνωσιακά συστήματα και είναι προτάσεις βιώσιμες και αρκετά ικανοποιητικές.

Πιο συγκεκριμένα, μια προτεινόμενη αρχιτεκτονική ενός γνωσιακού συστήματος αναζήτησης πληροφορίας στον Παγκόσμιο Ιστό θα μπορούσε να δομηθεί σύμφωνα με το ακόλουθο σχήμα: καταρχήν, το σύστημα θα τρέχει ως πρόγραμμα-πελάτης (client) στον υπολογιστή κάθε χρήστη. Ο λόγος γι' αυτήν την επιλογή είναι ότι βρίσκεται σε αντίθεση με τη λογική «ένα μέγεθος για όλους» (“one size fits all”) των περισσότερων εμπορικών συστημάτων αναζήτησης. Εδώ, δίνεται έμφαση στην ατομικότητα του κάθε χρήστη, κι έτσι το σύστημα χρειάζεται να είναι σε στενή συνεργασία με το συγκεκριμένο χρήστη στον οποίο απευθύνεται, αφιερώνοντας τις υπηρεσίες του αποκλειστικά στον τελευταίο.

Στην είσοδο του συστήματος ο χρήστης θα σχηματίζει την επερώτησή του διατυπωμένη σε φυσική γλώσσα. Η επιλογή μιας τέτοιας διατύπωσης, αν και επιβάλλει επιπλέον πολυπλοκότητα στο σύστημα, προτιμάται διότι είναι ανετότερη για το χρήστη. Σημαντικός σ' αυτό το στάδιο είναι ο εντοπισμός της πληροφοριακής ανάγκης του χρήστη. Προκειμένου να γίνει αυτό, χρειάζεται να προσδιοριστεί ο σκοπός της επερώτησης. Αυτό περιλαμβάνει τον προσδιορισμό του υπόβαθρου του χρήστη και του έργου αναζήτησης, δηλαδή του πλαισίου

(context). Η προσέγγιση λοιπόν που προτιμάται είναι αυτή της αναζήτησης εντός πλαισίου.

Η εξαγωγή των εννοιών της επερώτησης μπορεί να γίνει με τη χρήση ενός επεξεργαστή φυσικής γλώσσας (NLP), που εκτελεί το συντακτικό κατακερματισμό της επερώτησης. Για την αναπαράσταση και την οργάνωση των εννοιών προτείνεται να χρησιμοποιηθούν οντολογίες (οργανωμένες όπως π.χ. στο πρότζεκτ CYC)⁴. Για τον προσδιορισμό του πλαισίου μπορεί να χρησιμοποιηθεί ένα προφίλ χρήστη, το οποίο θα ενημερώνεται με έμμεσες μεθόδους σχετικά με τα χαρακτηριστικά και τη συμπεριφορά του χρήστη κατά την αναζήτηση πληροφορίας. Το προφίλ χρήστη χρησιμεύει στην αντιστοίχιση των λέξεων της επερώτησης προς τις έννοιες που είναι πιθανό ν' αντιπροσωπεύουν ακριβέστερα την πληροφοριακή ανάγκη του χρήστη. Για την υλοποίηση του προφίλ μπορεί να χρησιμοποιηθεί ένας πράκτορας βασισμένος σε κάποιο γνωσιακό μοντέλο μνήμης ή/και συλλογιστικής⁵, που παρακολουθεί τις ενέργειες του χρήστη και που περιδιαβαίνει τον Ιστό για να εντοπίσει ιστοσελίδες σχετικές προς τις επερωτήσεις του [χρήστη] (όπως στο πρότζεκτ RASCALLI των (Krenn & Schollum, 2008)).

Με τη χρήση του NLP θα εξάγονται οι έννοιες κι από τις διάφορες ιστοσελίδες. Ο πράκτορας – που είναι στον πυρήνα του συστήματος – θα επιλέγει προς παρουσίαση τις ιστοσελίδες που απαντούν καλύτερα στις πληροφοριακές ανάγκες του χρήστη, κρίνοντας από την εγγύτητα των εννοιών των ιστοσελίδων με τις έννοιες της επερώτησης.

Σε ό,τι αφορά την έξοδο του συστήματος, δηλαδή την παρουσίαση της σχετικής πληροφορίας στο χρήστη, αντί να παρουσιάζονται απευθείας οι ιστοσελίδες σε μορφή λιστών, θα ήταν καλύτερο να παρουσιάζονται ομαδοποιημένες σε έννοιες ούτως ώστε ο χρήστης να εντοπίζει άκοπα την ομάδα των ιστοσελίδων που

⁴ Το πρόβλημα βέβαια στην προκειμένη περίπτωση είναι πως ο χώρος (οι έννοιες) αναζήτησης είναι ευρύς, καθώς το σύστημα θα είναι γενικού σκοπού κι όχι επί ενός περιορισμένου θέματος αναζήτησης. Σε έναν τόσο ευρύ χώρο αναζήτησης, λοιπόν, οι οντολογίες που θα χρησιμοποιηθούν θα πρέπει να καλύπτουν πολλούς τομείς γνώσης. Όχι πως δεν υπάρχουν διάφορα πρότζεκτ που υλοποιούν κάτι τέτοιο, αλλά κατεβαίνοντας επίπεδα στις ιεραρχίες των οντολογιών η γνώση πολλαπλασιάζεται τόσο, που δεν είναι τεχνικά εύκολο να τις χειριστεί κανείς.

⁵ Τέτοια μοντέλα είναι το ACT-R, το DUAL/AMBR κ.ά.

απαντούν στο ερώτημά του (όπως στο ConSearch των (Lee & Lieberman, 2004)). Επίσης ελκυστική θα ήταν η τεχνική των πολλαπλών όψεων (multiple views) που εφαρμόζεται στο σύστημα των (Hubmann-Haidvogel, Scharl, & Weichselbraun, 2009).

Όπως έχει προαναφερθεί, το ιδανικό γνωσιακό σύστημα αναζήτησης θα υλοποιεί και το πέμπτο βήμα της αναζήτησης, τη μετατροπή της πληροφορίας σε γνώση. Πρακτικά όμως, και μιλώντας για ένα σύστημα αναζήτησης γενικού σκοπού, αυτό δεν είναι ακόμη εφικτό. Ο λόγος έχει να κάνει με την ποσότητα της (εξειδικευμένης) γνώσης που θα έπρεπε να έχει ένα τέτοιο σύστημα και με την ποσότητα της οντολογικής κατηγοριοποίησης που θα έπρεπε να κάνει σε κάθε γνωστικό αντικείμενο, προκειμένου να τα καλύψει όλα.

Κεφάλαιο 5: Συμπεράσματα

Η παρούσα εργασία ασχολείται με το πρόβλημα της αναζήτησης πληροφορίας στον Παγκόσμιο Ιστό, και μάλιστα από την οπτική μιας συγκεκριμένης προσέγγισης στο εν λόγω πρόβλημα, η οποία εμπνέεται από το χώρο της Γνωσιακής Επιστήμης.

Οι περισσότεροι ερευνητές, βάσει του τι υλοποιούν τα υπάρχοντα πληροφοριακά συστήματα, αναγνωρίζουν τέσσερα στάδια στην αναζήτηση πληροφορίας: σχηματισμό επερώτησης, εντοπισμό σχετικών ιστοσελίδων, ταξινόμησή τους και παρουσίασή τους. Σ' αυτά προστέθηκε ένα πέμπτο, η μετάβαση από την πληροφορία στη γνώση. Η παραδοχή της ύπαρξης αυτού του σταδίου έχει τη σημασία της, καθώς μια τέτοια διαδικασία επιτρέπει την αφομοίωση της πληροφορίας από το χρήστη ούτως ώστε μέσω της εμπειρίας να τη μετατρέψει σε γνώση. Αυτό το στάδιο γενικά δεν εμφανίζεται σε άλλες δουλειές ή τουλάχιστον όχι μέσα σε ένα τέτοιο ολοκληρωμένο πλαίσιο, ενδεχομένως επειδή τα λεγόμενα μετά-γνωστικά εργαλεία που το υλοποιούν είναι μια σχετικά πρόσφατη ερευνητική τάση και ακόμη κι οι τεχνολογίες που τα υποστηρίζουν είναι σε πρωτόλεια μορφή.

Παρά το γεγονός ότι οι επιστημονικές κοινότητες της Γνωσιακής Επιστήμης και της Ανάκτησης Πληροφορίας (κομμάτι της οποίας είναι η Αναζήτηση Πληροφορίας στον Παγκόσμιο Ιστό) δε βρίσκονται πάντα σε ιδιαίτερα στενή συνεργασία, υπάρχουν συστήματα (ορισμένα εκ των οποίων έχουν ενσωματωθεί σε εμπορικές μηχανές αναζήτησης) που ανήκουν και στους δύο χώρους, αποτελώντας την απόδειξη ότι η γνωσιακή αναζήτηση πληροφορίας δεν είναι χίμαιρα. Τα συστήματα αυτά ακολουθούν τάσεις και τεχνολογίες που δεν είναι αποκλειστικό προνόμιο της γνωσιακής αναζήτησης πληροφορίας, όμως τις ενσωματώνουν σε μια λογική Γνωσιακής Επιστήμης. Αυτό επιτυγχάνεται με την υιοθέτηση συγκεκριμένων γνωσιακών μοντέλων και θεωριών για την αναζήτηση πληροφορίας, που βρίσκονται στην καρδιά της σύλληψης ή/και της υλοποίησης των περιγραφόμενων συστημάτων.

Τα γνωσιακά μοντέλα και οι γνωσιακές θεωρίες πάνω στις οποίες βασίζονται αυτά τα συστήματα αφορούν ενίοτε πολύ γενικές νοητικές διεργασίες, όπως η

μνήμη, κι ενίοτε πολύ συγκεκριμένες νοητικές διεργασίες, όπως αυτές που εμπλέκονται στην αναζήτηση πληροφορίας και στη διάκριση των στοιχείων (chunks) πληροφορίας εκ μέρους ενός χρήστη. Σε ορισμένα από τα γνωσιακά συστήματα η εφαρμογή των θεωριών είναι αποσπασματική, δεδομένου ότι ιδανικά, θα έπρεπε να υλοποιούν και τα πέντε στάδια της αναζήτησης πληροφορίας βασισμένα στην εκάστοτε θεωρία. Αυτό δε συμβαίνει για δύο λόγους:

- Λόγω της αδυναμίας των προτεινόμενων θεωριών, δηλαδή της δυσκολίας να βρεθεί μια καθολικά αποδεκτή περιγραφή του έργου της αναζήτησης και του εντοπισμού πληροφορίας, έτσι όπως το επιτελεί ο ανθρώπινος νους, στο εξαιρετικά δυναμικό περιβάλλον του Παγκόσμιου Ιστού. Αυτό δεν είναι κάτι που αναμένεται να λυθεί σύντομα, καθώς ανήκει στα «δύσκολα προβλήματα» της Γνωσιακής Επιστήμης και απαιτείται πολλή ερευνητική προσπάθεια για την αποσαφήνισή του, ωστόσο υπάρχουν κάποιες καλές προσεγγίσεις στο ζήτημα.
- Για πρακτικούς λόγους που έχουν να κάνουν με το κόστος υλοποίησης ενός τέτοιου ολοκληρωμένου ερευνητικού έργου. Καλώς ή κακώς, η αναζήτηση πληροφορίας στον Παγκόσμιο Ιστό είναι ένα χρυσωρυχείο στο οποίο εμπλέκονται εταιρείες-κολοσσοί με τεράστιες επενδύσεις στην αύξηση της τεχνογνωσίας τους, κι ως συνήθως οι κυβερνητικές χρηματοδοτήσεις δε συναγωνίζονται αυτές τις επενδύσεις.

Όπως και να 'χει το πράγμα, τα γνωσιακά συστήματα αναζήτησης έχουν ελπιδοφόρες επιδόσεις. Χωρίς να παραβλέπονται οι αδυναμίες των γνωσιακών σε σχέση με τα υπόλοιπα συστήματα (κυρίως σε ό,τι αφορά τον εντοπισμό πολυμεσικής πληροφορίας) και με τη δεδομένη δυσκολία της μέτρησης της απόδοσής τους, λόγω της έλλειψης συναίνεσης για το τι πρέπει τελικά να μετράται, προκύπτει ότι, αν μη τι άλλο, δεν είναι χειρότερα από τα παραδοσιακά ή τα σημασιολογικά συστήματα. Εδώ να σημειωθεί η πεποίθηση της γράφουσας πως όποια τεχνολογική προσπάθεια στηρίζεται σε γνωσιακές προσεγγίσεις αργά ή γρήγορα καρποφορεί.

Σε ό,τι αφορά τις εφαρμοζόμενες τεχνολογίες για την υλοποίηση των γνωσιακών συστημάτων αναζήτησης, δεν υπάρχει αυστηρή γραμμή· ανάλογα με την προσέγγιση που επιλέγεται, επιλέγονται και τεχνολογίες υλοποίησης. Γνώμη μου είναι πως οι καταλληλότερες τεχνολογίες είναι αυτές που αξιοποιούν οντολογίες για την αναπαράσταση γνώσης και που βασίζονται σε συνδεδειστικά/υπολογιστικά/δυναμικά μοντέλα για την κατασκευή έξυπνων πρακτόρων του Παγκόσμιου Ιστού.

Εν κατακλείδι, ο χώρος της Αναζήτησης Πληροφορίας στον Παγκόσμιο Ιστό είναι ένας χώρος που, αν και στόχος του είναι να διευκολύνει τους ανθρώπους στο δύσκολο νοητικό έργο του εντοπισμού της πληροφορίας μέσα στον κυκεώνα της διαθέσιμης (πληροφορίας), επί δεκαετίες εξελισσόταν μακριά από το χώρο της Επιστήμης του Νου. Η σύγκλιση αυτών των δύο χώρων, λοιπόν, μόνο θετικά αποτελέσματα μπορεί να έχει.

Βιβλιογραφία

- Anderson, J. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior* , 22, σσ. 261-295.
- Aula, A. (2005). Studying User Strategies and Characteristics for Developing Web Search Interfaces. *Dissertations in Interactive Technology* . Tampere, Finland.
- Daoud, M., Tamine-Lechani, L., & Boughanem, M. (2008). Learning user interests for a session-based personalized search. *ACM International Conference Proceeding Series, Proceedings of the second international symposium on Information interaction in context* , 348, σσ. 57-64.
- Desai, M., & Spink, A. (2005). An algorithm to cluster documents based on relevance. *Information Processing and Management* , 41, σσ. 1035-1049.
- Dhir, C. S., & Lee, S. Y. (2009). Search-In-Synchrony: Personalizing Web Search with Cognitive User Profile Model, Revised. Στο M. Koppen, & e. al (Επιμ.), *Advances in Neuro-Information Processing, 15th International Conference, ICONIP 2008* (σσ. 80-87). Auckland, New Zealand: Springer-Verlag.
- Diaz, A. M. (2005). Through the Google Goggles: Sociopolitical Bias in Search Engine Design.
- Dimitrova, M. (2003). Cognitive modelling and Web search: Some heuristics and insights. *Cognition, Brain, Behaviour* , 7 (3), σσ. 251-258.
- Dimitrova, M., Barakova, E. I., Lourens, T., & Radeva, P. (2004). The web as an autobiographical agent. Στο C. Bussler, & D. Fensel (Επιμ.), *Artificial Intelligence: Methodology, Systems, and Applications* (σσ. 510-519). Springer-Verlag.
- Dimitrova, M., Wagatsuma, H., & Yamaguchi, Y. (2007). Design of Web Agents Inspired by Brain Research. *Proceedings of the 11th East-European Conference on Advances in Databases and Information Systems ADBIS'2007*. Varna, Bulgaria.
- Ellis, & Haugan. (1997). Modelling the Information Seeking Patterns of Engineers and Research Scientists in an Industrial Environment. *Journal of Documentation* , 53 (4), σσ. 384-403.
- Freund, L., & Toms, E. G. (2005). Contextual search: from information behavior to information retrieval. Στο F. Crestani, & I. Ruthven (Επιμ.), *Information Context: Nature, Impact, and Role* (σσ. 59-78). Springer.
- Gardner, H. (1999). *The Disciplined Mind: What All Students Should Understand*. New York: Simon & Schuster.
- Gauch, S., Speretta, M., Chandramouli, A., & Micarelli, A. (2007). User Profiles for Personalized Information Access. Στο P. Brusilovsky, A. Kobsa, & W. Nejdl, *The*

Adaptive Web: Methods and Strategies of Web Personalization. Berlin Heidelberg New York: Springer-Verlag.

Guha, R., McCool, R., & Miller, E. (2003). Semantic Search. *In Proceedings of the 12th International Conference on World Wide Web (WWW)*, (σσ. 700-709). Budapest, Hungary.

Guilford, J. (1967). *The Nature Of Human Intelligence*. New York: Mcgraw-Hill.

Gwizdka, J. (2009). Assessing Cognitive Load on Web Search Tasks. *The Ergonomics Open Journal* , 2, σσ. 114-123.

http://en.wikipedia.org/wiki/Semantic_search. (n.d.). Ανάκτηση Οκτώβριος 19, 2009, από Wikipedia: <http://en.wikipedia.org>

Hu, W.-C., Chen, Y., Scmalz, M. S., & Ritter, G. X. (2001). An Overview of World Wide Web Technologies. *Proceedings of the 5th World Multi-Conference on Systemics, Cybernetics and Informatics, SCI 2001*, (σσ. 356-361). Orlando, Florida.

Hubmann-Haidvogel, A., Scharl, A., & Weichselbraun, A. (2009). Multiple coordinated views for searching and navigating Web content repositories. *Information Sciences* , 30 (1), σσ. 1813-1821.

Jeevan, V., & Padhi, P. (2006). A selective review of research in content personalization. *Library Review* , 55 (9), σσ. 556-586.

Jiang, X., & Tan, A. H. (2009). Learning and inferencing in user ontology for personalized semantic web search. *Learning and Information Sciences* , 179 (16), σσ. 2794-2808.

Jin, H., Ning, X., Jia, W., Wu, H., & Lu, G. (2008). Combining weights with fuzziness for intelligent semantic web search. *Knowledge-Based Systems* , 21, σσ. 655-665.

Juvina, I., & Oostendorp, H. v. (2008). Modeling Semantic and Structural Knowledge in Web Navigation. *Discourse Processes* , 45 (4-5), σσ. 346-364.

Juvina, I., Oostendorp, H., van, K. P., & Pauw, B. (2005). Toward Modeling Contextual Information in Web Navigation. *In CogSci2005, 27th Annual Conference of the Cognitive Science Society*, (σσ. 1078-1083). Stresa, Italy.

Keenoy, K., & Levene, M. (2005). Personalisation of Web Search. Στο B. Mobasher, & S. S. Anand (Επιμ.), *Intelligent Techniques for Web Personalization* (σσ. 201-228).

Kim, K. J., & Cho, S. B. (2007). Personalized mining of Web documents using link structures and fuzzy concept networks. *Applied Soft Computing* , 7 (1), σσ. 398-410.

Kim, K.-J., & Cho, S.-B. (2007). Personalized mining of web documents using link structures and fuzzy concept networks. *Applied Soft Computing* (7), σσ. 398-410.

Kitajima, M., Polson, P. G., & Blackmon, M. H. (2000). A Comprehension-based Model of Web Navigation and Its Application to Web Usability Analysis. *Proceedings of CHI2000* (σσ. 357-373). ACM Press.

Knight, S. A., & Spink, A. (2008). Toward a Web Search Information Behavior Model. Στο A. Spink, & M. Zimmer (Επιμ.), *Web Search, Springer Series in Information Science and Knowledge Management 14* (σσ. 209-234). Berlin Heidelberg: Springer-Verlag.

Kostadinov, S., & Grinberg, M. (n.d.). The Embodiment of a DUAL/AMBR Based Cognitive Model in the RASCALLI Multi-Agent Platform.

Kostadinov, S., Petkov, G., & Grinberg, M. (2008). Embodied Conversational Agent Based on the Dual Cognitive Architecture. Στο J. Cordeiro, F. Joaquim, & S. Hammoudi (Επιμ.). INSTICC Press.

Krenn, B. (2008). Responsive Artificial Situated Cognitive Agents Living and Learning on the Internet. *CogSys 2008*.

Krenn, B., & Schollum, C. (2008). The RASCALLI Platform-For a Flexible and Distributed Development of Virtual Systems Augmented with Cognition. *International Conference on Cognitive Systems (CogSys 2008)*, (σσ. 91-96). Karlsruhe.

Lawrence, S. (2000). Context in Web Search. *IEEE Data Engineering Bulletin* , 23 (3), σσ. 25-32.

Lee, C.-H., & Lieberman, H. (2004). ConSearch: An Concept-Associating Search Interface using Commonsense.

Lerner, B., & Lawrence, N. (2001). A comparison of state-of-the-art classification techniques with application to cytogenetics. *Neural Computing and Applications* , σσ. 39-47.

Lieberman, H., Fry, C., & Weitzman, L. (2001). Exploring the Web with Reconnaissance Agents. *Communications of the ACM* , 44 (8), σσ. 69-75.

Liu, Z., & Zhang, Y. (2001). A competitive neural network approach to Web-page categorization. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* , 9, σσ. 731-741.

Magoulas, G. D., Chen, S., & Dimakopoulos, D. (2004). A Personalised Interface for Web Directories based on Cognitive Styles. *Lecture Notes in Computer Science* (3196), σσ. 159-166.

Mansourian, Y., & Madden, A. D. (2007). Methodological approaches in web search research. *The Electronic Library* , 25 (1).

Marchionini, G. (1995). *Information Seeking in Electronic Environments*. Cambridge University Press.

- McGowan, J. P., Kushmerick, N., & Smyth, B. (2002). Who do you want to be today? Web Personae for personalised information access. *AH 2002*, (σσ. 514-517).
- Perkins, D. N. (1998). What Is Understanding? . Στο M. S. Wiske (Επιμ.), *Teaching For Understanding: Linking Research With Practice* (σσ. 39-57). San Francisco.
- Picard, D., Cord, M., & Revel, A. (2008). Image Retrieval Over Networks: Active Learning Using Ant Algorithm. *IEEE Transactions on Multimedia* , 10 (7), σσ. 1356-1365.
- Pirolli, P., & Card, S. (1999). *Psychological Review* , 106 (4), σσ. 643-675.
- Pirolli, P., & Fu, W. (2003). SNIF-ACT: a model of information foraging on the World Wide Web. *9th International Conference on User Modeling (UM 2003)*, (σσ. 45-54). Johnstown; PA.
- Pitkow, J., Schütze, H., Cass, T., Cooley, R., Turnbull, D., Edmonds, A., και συν. (2002). Personalized search. *Communications of the ACM* , 45 (9), σσ. 50-55.
- Sahay, S., Venkatesh, A., & Ram, A. (2009). Collaborative Information Access: A Conversational Search Approach. *ICCBR Workshop on Reasoning from Experiences on the Web (WebCBR-09)*. Seattle.
- Saracevic, T. (1975). Relevance: A review of the literature and a framework for thinking on the notion in information science. *Journal of the American Society for Information Science and Technology* , 26 (6), σσ. 321-343.
- Saracevic, T. (2007). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology* , 58 (3), σσ. 1915-1933.
- Schoelles, M. J., Gray, W. D., Veksler, V. D., & Gamard, A. (2006). Cognitive modeling of web search. *The 13th Annual ACT-R Workshop*. Carnegie Mellon University, Pittsburgh, PA.
- Sperber, & Wilson. (1986). *Relevance: Communication and Cognition*. Cambridge MA: Blackwell.
- Spink, A., & Zimmer, M. (Επιμ.). (2008). *Web Searching: Multidisciplinary Perspectives*. Springer.
- Spink, A., & Cole, C. B. (2005). *New Directions in Cognitive Information Retrieval*. Springer.
- Stadtler, M., & Bromme, R. (2008). Effects of the metacognitive computer-tool met.a.ware on the web search of laypersons. *Computers in Human Behavior* , 24 (3), σσ. 716-737.

Tjondronegoro, D., & Spink, A. (2008). Web search engine multimedia functionality. *Information Processing and Management* , 44, σσ. 340-357.

Wechsler, K., Baier, J. A., Nussbaum, M., & Baeza-Yates, R. (2004). Semantic Search in the WWW Supported by a Cognitive Model. *In Proceedings of the 5th International Conference on Advances on Web-age Information Management, July 15-17* (σσ. 179-186). Dalian, China: Springer-Verlag.

Wei, W., Barnaghi, P. M., & Bargiela, A. (2008, June). Search with Meanings: An Overview of Semantic Search Systems. *International journal of Communications of SIWN* , 3, σσ. 76-82.

White, H. D. (2007). Combining bibliometrics, information retrieval, and relevance theory, Part 1: First examples of a synthesis: Research Articles. *Journal of the American Society for Information Science and Technology* , 38 (4), σσ. 536-559.

White, H. D. (2007). Combining bibliometrics, information retrieval, and relevance theory, Part 2: Some implications for information science: Research Articles. *Journal of the American Society for Information Science and Technology* , 58 (4), σσ. 583-605.

Williams, A., & Ren, Z. (2001). Agents teaching agents to share meaning. *Proceedings of the 5th International Conference on Autonomous Agents*, (σσ. 465-472).

Wilson, D., & Sperber, D. (2004). Relevance theory. Στο G. Ward, & L. Horn (Επιμ.), *Handbook of Pragmatics* (σσ. 607-632). Oxford: Blackwell.

Wooldridge, M. J. (2001). *An introduction to multiagent systems*. West Sussex, England: John Wiley & Sons Ltd.

Zadeh, L. A. (2004). A note on web intelligence, world knowledge and fuzzy logic. *Data Knowledge & Engineering* , 50 (3), σσ. 291-304.

Zemirli, N., Tamine-Lechani, L., & Boughanem, M. (2007). A Personalized Retrieval Model based on Influence Diagrams. Στο B.-L. Doan, J. M. Joemon, & M. Melucci (Επιμ.), *The CIR 07 Workshop on Context-Based Information Retrieval in conjunction with CONTEXT-07*. Roskilde, Denmark.