

Οι Μηχανές Αναζήτησης (Search Engines) ως μέσα παροχής πληροφοριών

I. Εισαγωγή

Το μέγεθος του Internet αυξάνεται με έντονο ρυθμό τα τελευταία 2-3 χρόνια. Δεκάδες οργανισμοί προσθέτουν την ιστοσελίδα τους κάθε βδομάδα και χιλιάδες σελίδες εκδίδονται καθημερινά. Ο αριθμός των ιστοσελίδων που είναι δωρεάν προσβάσιμες στο Internet υπερβαίνει τα 800 εκατ.¹ Αυτός ο τεράστιος όγκος υλικού καθιστά το Διαδίκτυο το μεγαλύτερο εργαλείο αναζήτησης και πληροφόρησης στα τέλη του 20ου αιώνα, αλλά και στις αρχές του 21ου.

II. Μηχανές Αναζήτησης

1. Ορισμός

Η ανάγκη κατηγοριοποίησης και ευρετηρίασης του εκτεταμένου αριθμού των ιστοσελίδων δημιούργησε τις **μηχανές αναζήτησης (search engines)**. Κατά τους Xie-Wang-Goh² οι μηχανές αναζήτησης είναι εργαλεία αυτοματοποιημένης ανάκτησης για την εύκολη ανεύρεση πληροφοριών μέσω του Internet. Κατά τη Webber³, οι μηχανές αναζήτησης περιλαμβάνουν κάθε βάση δεδομένων που μπορεί να ερευνηθεί και που ευρετηριάζει μια σειρά ιστοσελίδων.

2. Χρησιμότητα

Όπως προκύπτει από τον ορισμό, οι μηχανές αναζήτησης είναι βοηθητικά εργαλεία, χρήσιμα για τη μεταχείριση του Internet, για την αναδίφηση και ανεύρεση των κατάλληλων πληροφοριών που ζητά ο ερευνητής, ως απάντηση στα ερωτήματά του.

¹ Bradley, Phil, "Virtual libraries and Internet searching," *Online & CD-ROM Review*, v.23(1999) no.6, p.353.

² Xie, M., H. Wang and T.N. Goh, "Quality dimensions of Internet search engines," *Journal of Information Science*, v.24 (1998) no.5, p.365.

³ Webber, Sheila, "Search engines and news services: developments on the Internet," *Business Information Review*, v.15 (Dec. 1998) no.4, p. 230.

3. Στοιχεία που τις απαρτίζουν

Σύμφωνα με την ιστοσελίδα της Search Engine Watch (<http://www.searchenginewatch.com>) οι μηχανές αναζήτησης έχουν τρία βασικά στοιχεία. Πρώτον η «αράχνη» (*spider* ή *crawler*) – ένα αυτοματοποιημένο πρόγραμμα που επισκέπτεται κάθε νέα ιστοσελίδα – αναζητά και διαβάζει την ιστοσελίδα. Ακολουθούν οι συνδέσεις (links) με άλλες σελίδες μέσα στη θέση του ιστού (web site)⁴. Οτιδήποτε βρει η «αράχνη» στη νέα ιστοσελίδα, πάει στο δεύτερο στοιχείο της μηχανής αναζήτησης το *ευρετήριο* (*index* ή *catalog*). Πρόκειται για ένα τεράστιο «βιβλίο» που περιέχει ένα αντίγραφο από κάθε ιστοσελίδα, την οποία ανακαλύπτει η «αράχνη». Η «αράχνη» κατά διαστήματα επιστρέφει στην ιστοσελίδα για να ενημερωθεί και να αποτυπώσει τις αλλαγές στη μηχανή αναζήτησης, της οποίας αποτελεί τμήμα.

Το τρίτο στοιχείο είναι το *λογισμικό της μηχανής αναζήτησης* (*search engine software*). Είναι το πρόγραμμα που μετακινείται ανάμεσα στα εκατομμύρια σελίδων που έχουν καταγραφεί στο ευρετήριο, για να βρει εκείνες που ανταποκρίνονται σε μια έρευνα και να τις κατατάξει με τη σειρά την οποία θεωρεί πιο σχετική για να απαντήσει σε κάθε ερώτημα.

Ας σημειωθεί ότι οι μηχανές αναζήτησης ευρετηριάζουν ιστοσελίδες που είναι δωρεάν προσβάσιμες στο κοινό. Επομένως από τις ιστοσελίδες που ευρετηριάζει κάθε μηχανή αναζήτησης απουσιάζουν οι σύνθετες βάσεις δεδομένων, όπως η Amazon.com (<http://www.amazon.com>), ιστοσελίδες που απαιτούν κωδικούς πρόσβασης, αλλά και βάσεις δεδομένων οι οποίες είναι προσβάσιμες με πληρωμή, όπως αυτές κατασκευαστών βάσεων δεδομένων και προμηθευτών (π.χ. Dialog, LEXIS-NEXIS, OCLC First Search, SilverPlatter, STN).

⁴ Για την απόδοση στα ελληνικά αγγλικών όρων χρησιμοποιήθηκε το λεξικό: *Το Λεξικό της Πληροφορικής* / Microsoft Press.- 3η αμερ. Εκδ.- Αθήνα: Κλειδάριθμος, 1998.

Ο Bradley⁵ πολύ καιρία αναφέρει ότι η χρήση της μηχανής αναζήτησης στην έρευνα, μοιάζει με τη χρήση ενός πληροφοριακού βιβλίου. Όπως δεν υπάρχει ένα μόνο πληροφοριακό βιβλίο σε ένα κέντρο πληροφόρησης, το οποίο να παρέχει απαντήσεις σε όλα τα θέματα όλων των επιστημών, έτσι δεν υπάρχει μόνο μια μηχανή αναζήτησης για την αντιμετώπιση κάθε έρευνας.

III. Αναζήτηση

1. Έναρξη της αναζήτησης και σημεία που πρέπει να ληφθούν υπόψη

Όπως αναφέρουν οι Kassel⁶ και Kassler⁷ οι στόχοι και οι ανάγκες της έρευνας προσδιορίζουν τη μηχανή αναζήτησης που θα χρησιμοποιήσει ο ερευνητής. Ένα αρχικό ερώτημα που πρέπει να απαντηθεί είναι αν ο ερευνητής θα επιλέξει τις παραδοσιακές βάσεις δεδομένων (cd-rom, online, OPAC) ή αν θα χρησιμοποιήσει το Internet. Στη δεύτερη περίπτωση πρέπει να σκεφθεί την κατάλληλη στρατηγική έρευνας (χρήση λέξεων, φράσεων, συνωνύμων που θα περιγράφουν το θέμα). Επομένως, απαιτείται να γνωρίζει τον τρόπο αναζήτησης και λειτουργίας των μεγαλύτερων μηχανών αναζήτησης, επισκεπτόμενος τις αντίστοιχες σελίδες τους. Η γνώση του περιεχομένου των μηχανών αναζήτησης θα τον βοηθήσει στην απόφαση για τη χρήση της καταλληλότερης για την έρευνά του. Τα πρώτα αποτελέσματα της αναζήτησης ίσως τον οδηγήσουν σε επαναπροσδιορισμό της στρατηγικής έρευνας μέσα από πιο συγκεκριμένους όρους. Μάλιστα στις μεγάλες μηχανές αναζήτησης με εκτενή ευρετήρια, αυξάνονται οι πιθανότητες να βρεθούν απαντήσεις που δε σχετίζονται με το αρχικό ερώτημα. Καθώς καμιά μηχανή αναζήτησης δεν είναι ακριβώς η ίδια με κάποια άλλη, ενώ σε πολλές από αυτές υπάρχει ένα ποσοστό αλληλοκάλυψης, τα αποτελέσματα ποικίλλουν. Η ίδια έρευνα μπορεί να επαναληφθεί σε άλλη μηχανή αναζήτησης για σύγκριση των αποτελεσμάτων ή να χρησιμοποιηθούν metasearch engines, οι οποίες εκτελούν μian έρευνα ταυτόχρονα σε πολλές μηχανές αναζήτησης, π.χ. η Dogpile: <http://www.dogpile.com>.

⁵ Bradley (o.a.), p. 354.

⁶ Kassel, Amelia, "Internet power searching: finding pearls in a zillio grains of sand," *Information Outlook*, v.3 (Apr. 1999) no.4, p.30.

2. Βασικές υποδείξεις, κοινά χαρακτηριστικά των μηχανών αναζήτησης

- Αρκετές μηχανές αναζήτησης επιτρέπουν έρευνα με τελεστές Boole (AND, OR, AND NOT).
- Άλλες μηχανές αναζήτησης απαιτούν χρήση εισαγωγικών σε φράσεις.
- Άλλες επιτρέπουν τη χρήση αποκοπής μιας λέξης για να περιληφθούν παραλλαγές της, παράγωγα, διαφορετικοί τύποι ρήματος ή ουσιαστικών, ενικός / πληθυντικός, κλπ.
- Συνήθως οι μηχανές αναζήτησης δε λαμβάνουν υπόψη λέξεις όπως άρθρα, συνδέσμους, αλλά και κοινές λέξεις, π.χ. Internet, web.
- Σε άλλες, η χρήση κεφαλαίων ή πεζών γραμμάτων ή συνδυασμός τους, διαφοροποιεί την έρευνα.
- Σε περιπτώσεις που η μηχανή αναζήτησης επιτρέπει προχωρημένη έρευνα (advanced search) είναι δυνατή η χρήση γνωστικών κλάδων για να εντοπιστούν ακριβέστερες ιστοσελίδες.
- Η γλώσσα αναζήτησης, ο τίτλος, η ημερομηνία, η αναζήτηση σε URL (Uniform Resource Locator: Ενιαίος Εντοπιστής Πόρων) είναι άλλα από τα χαρακτηριστικά των μηχανών αναζήτησης.

Επειδή οι μηχανές αναζήτησης πολλαπλασιάζονται συνεχώς σε αριθμό, ποικίλλουν σε μέγεθος, ποιότητα και δυνατότητες, οι κατασκευαστές τους προσπαθούν να βρουν στοιχεία που να τις κάνουν ανταγωνιστικότερες και ελκυστικότερες για τον ερευνητή. Έτσι μπορούν εναλλακτικά να προσθέσουν μια βάση δεδομένων με κατηγοριοποίηση θεμάτων μαζί με ιστοσελίδες, πληροφορίες για εταιρείες, διαφημίσεις, προτροπές προς τους αναγνώστες να ψηφίσουν για τις σελίδες που βλέπουν, να βρουν πληροφορίες για τον αριθμό των ερευνητών που έχουν επισκεφθεί τη συγκεκριμένη σελίδα.⁸

⁷ Kassler, Helene, "Competitive intelligence on the Internet: going for the gold," *Information Outlook*, v.4 (Feb. 2000) no.2, p.41.

IV. Περιγραφή των κυριότερων μηχανών αναζήτησης

Σύμφωνα με στοιχεία που προέρχονται από την ιστοσελίδα της Search Engine Showdown (<http://www.notess.com/search>) του Greg Notess – αρθρογράφου, βιβλιοθηκονόμου πληροφόρησης και βοηθού καθηγητή στο Montana State University - κατά την επίσκεψη της σελίδας του στις 12-4-2000⁹, οι τρεις μεγαλύτερες μηχανές αναζήτησης είναι οι **Fast, Northern Light και AltaVista** (στοιχεία του Ιανουαρίου 2000). Η σύγκριση με στοιχεία του αντίστοιχου μήνα του 1999 δίνουν την πρώτη σειρά στη **Northern Light** και ακολουθούν οι **AltaVista** και **HotBot**. Τα ενδεικτικά αυτά στοιχεία αποκαλύπτουν την ταχύτητα με την οποία μεταβάλλεται το μέγεθος των μηχανών αναζήτησης, την ευκολία με την οποία δημιουργούνται νέες (FAST) καθώς και το ευμετάβολο και ασταθές της διατήρησης στην πρώτη θέση.

Σύμφωνα πάντα με τα τελευταία στοιχεία του Greg Notess, που αναφέρουν ημερομηνία 21-2-2000, οι μηχανές αναζήτησης κατατάσσονται ως εξής σε σχέση με το μέγεθός τους.

1.	FAST	280-308 εκατ. σελίδες
2.	Northern Light	219-241
3.	AltaVista	197-217
4.	Goggle	140-154
5.	Excite	135-149
6.	Answers (Inktomi)	73-81
7.	iWon Inktomi	43-48
8.	Yahoo!'s Inktomi	58-64
9.	AOL Inktomi	61-67
10.	Lycos	41-45
11.	Infoseek	48-52

⁸ Webber, (ο.α.) p.230.

⁹ Αναφέρεται η ημερομηνία επίσκεψης στις ιστοσελίδες που χρησιμοποιήθηκαν, επειδή μέχρι την ημερομηνία έκδοσης του παρόντος άρθρου ίσως έχουν προστεθεί νέες μηχανές αναζήτησης ή έχουν τροποποιηθεί τα δεδομένα.

12.	HotBot	36-39
13.	WebCrawler	5-6

Μια απλή γνωριμία με ορισμένες από τις μηχανές αναζήτησης θα μας βοηθήσει στην επιλογή της κατάλληλης που θα προτείνουμε στον ερευνητή ανάλογα με τις απαιτήσεις της έρευνάς του. Για την επιλογή, κριτήριο δεν αποτελεί μόνο το μέγεθος της μηχανής αναζήτησης, αλλά και η ποιότητά της. Εκτός από την προαναφερθείσα ιστοσελίδα του G.Notes, χρησιμοποιήθηκαν αναφορές από την ιστοσελίδα της Search Engine Watch (<http://www.searchenginewatch.com>) του Danny Sullivan, δημοσιογράφου και συμβούλου Internet.

1. **FAST** (<http://www.alltheweb.com>)

Η FAST, από τις δημοφιλέστερες μηχανές αναζήτησης, δημιουργήθηκε από μια νορβηγική εταιρεία την Fast Search and Transfer το Μάιο του 1999, που έτρεχε σε υπολογιστές Dell.

Ακολουθώντας το επιχειρηματικό μοντέλο της Inktomi (βλ. πιο κάτω), θα αρχίσει να διαθέτει τη FAST και σε άλλες μηχανές αναζήτησης, όπως έχει ήδη κάνει η Lycos με το προχωρημένο επίπεδο έρευνάς της (Lycos Pro). Η FAST ξεκίνησε με 80 εκατ. URLs, έφθασε στα 200 εκατ. τον Αύγουστο του 1999 και τον Ιανουάριο του 2000 στα 300 εκατ. Η εταιρεία σχεδιάζει να προσφέρει μια βάση δεδομένων 400 εκατ. σελίδων μέσα στον Απρίλιο του 2000. Στοχεύει να ευρετηριάσει όλο το web που διατίθεται ελεύθερα στο κοινό χωρίς χρέωση.

Το Δεκέμβριο του 1999 η FAST πρόσθεσε τη σελίδα για προχωρημένη αναζήτηση (advanced search) που προσφέρει αναζήτηση με τελεστές Boole οδηγούμενους από μενού (menu driven) σε αντίθεση με την απλή FAST. Επίσης προσφέρει αναζήτηση κατά πεδίο (field) δηλαδή τίτλο, σύνδεση*, URL, κείμενο, γλώσσα (25 γλώσσες) και περιορισμό κατά πεδίο ορισμού (domain, π.χ. .edu, .com).

Στα δυνατά σημεία της FAST συγκαταλέγονται το μέγεθος, η ταχύτητα απόκρισης και το νεαρό της «ηλικίας» της, δεν ευρετηριάζει μικρές λέξεις (stop words) και παρέχει νέα επίπεδα στην προχωρημένη έρευνα. Ενώ στις αδυναμίες της περιλαμβάνονται η έλλειψη συνεχούς ανανέωσης της βάσης δεδομένων και η έλλειψη τελεστών Boole στην αναζήτηση, στην αποκοπή, στους τελεστές γειννίας (proximity operators) κατά την απλή αναζήτηση.

2. **Northern Light** (<http://www.northernlight.com>)

Είναι μια άλλη δημοφιλής μεταξύ των ερευνητών και από τις μεγαλύτερες, μηχανή αναζήτησης. Ξεκίνησε τον Αύγουστο του 1997 και το Νοέμβριο του 1999 ξεπέρασε τα 300 εκατ. ιστοσελίδες. Η Northern Light προσφέρει μια μεγάλη βάση δεδομένων με ιστοσελίδες, καθώς και πρόσβαση σε πλήρες κείμενο από τις 5600 δημοσιεύσεις της Special Collection της, με χρέωση από \$1-\$4 για κάθε τεκμήριο. Μερικές από τις εκδόσεις αυτές, όπως αναφορές από εταιρείες (MarkIntel, Find/SVP, WEFA, Investext) δεν είναι διαθέσιμες από άλλους εμπορικούς οίκους.

Η Northern Light υποστηρίζει αναζήτηση με τελεστές Boole (AND, OR, AND NOT) με χρήση παρενθέσεων αλλά και εισαγωγικών για αναζήτηση με φράσεις. Επιτρέπει την αποκοπή με αστερίσκο * για απεριόριστο αριθμό χαρακτήρων και το σημείο % αντιπροσωπεύει ένα απλό χαρακτήρα. Περιλαμβάνει ενικό και πληθυντικό για τη λέξη αναζήτησης, καθώς και πολλαπλή αναζήτηση με όρους και με προσδιορισμό πεδίων (fields) έρευνας (τίτλος, URL, κείμενο, εταιρεία). Παρέχει περιορισμούς κατά χρονολογία, είδος εγγράφου, έκδοση, θέμα, γλώσσα, πεδίο ορισμού.

Στα δυνατά σημεία της περιλαμβάνονται το μεγάλο και αυξανόμενο μέγεθος της βάσης δεδομένων, τα πλούσια χαρακτηριστικά έρευνας, η Special Collection, οι News Databases και οι Custom Search Folders, ενώ στις αδυναμίες της συγκαταλέγονται η εμφάνιση μόνο 10 αποτελεσμάτων για κάθε αναζήτηση, η έλλειψη τελεστών γειννίας και η έλλειψη συνδέσεων (links).

3.

AltaVista (<http://www.altavista.com>)

Η AltaVista εξακολουθεί να κατατάσσεται στις τρεις μεγαλύτερες μηχανές αναζήτησης. Δημιουργήθηκε το Δεκέμβριο του 1995 από την εταιρεία Digital, συνέχισε υπό την ηγεσία της Compaq, η οποία αγόρασε τη Digital το 1998 και κατέληξε να λειτουργεί ως ανεξάρτητη εταιρεία ελεγχόμενη από τη CMGI.

Είναι ιδιαίτερα δημοφιλής, εξαιτίας της περιεκτικής κάλυψης και της ευρείας κλίμακας δυναμικών εντολών αναζήτησης. Τους τελευταίους μήνες πρόσθεσε περιεχόμενα πύλης (portal content) και ένα διεθνές είδωλο (international mirror)¹⁰. Περιλαμβάνει απλή και προχωρημένη έρευνα (simple search, advanced search). Τα αποτελέσματα προέρχονται από τη βάση δεδομένων της AltaVista με εκατομμύρια ευρετηριασμένων σελίδων, τη Suggested Relevant Searches (με εξειδικευμένες φράσεις), τη RealNames Internet Keywords, την Ask Jeeves, την Open Directory και από τρέχουσες ειδήσεις.

Είναι κατάλληλη για εξειδικευμένες έρευνες με ευέλικτους τρόπους αναζήτησης. Προσφέρει έρευνα σε φυσική γλώσσα. Χρησιμοποιούνται απλοί ή πολλαπλοί όροι ή φράσεις με εισαγωγικά, επιτρέπεται η χρήση τελεστών Boole (AND, OR, AND NOT αλλά και NEAR), παρενθέσεων, αποκοπή με αστερίσκο * για 0-5 επιπλέον χαρακτήρες. Διαθέτει αναζήτηση σε πεδία (τίτλος, εικόνα, διαθέτης, σύνδεση, κείμενο, πεδίο ορισμού, URL). Υπάρχουν περιορισμοί κατά γλώσσα και χρονολογία. Αγνοούνται οι κοινές λέξεις. Κατά την εμφάνιση τους κατατάσσονται πρώτα τα αποτελέσματα από τη βάση δεδομένων της Ask Jeeves, ενώ η σχετικότητα καθορίζεται σύμφωνα με τη θέση των όρων μέσα στην ιστοσελίδα, την γειννίαση πολλαπλών όρων αναζήτησης, ενώ στην advanced search τα αποτελέσματα δεν παρουσιάζονται με κάποια σειρά, αλλά μόνο κατά θέση ιστού (site).

Πλεονεκτήματά της είναι το μέγεθος της βάσης δεδομένων, τα δικά της χαρακτηριστικά έρευνας, η διεθνής κάλυψη, η διασύνδεση (interface) και η διαχείριση των ξένων γλωσσών, η ευρετηρίαση του πλήρους κειμένου των ιστοσελίδων, η δυνατότητα παροχής ελεύθερης μετάφρασης για τις γλώσσες Αγγλικά, Γαλλικά, Γερμανικά, Ιταλικά,

¹⁰ Notess, Greg, "Internet search engine update," Online, Mar. 2000: <http://www.onlineinc.com/onlinemag/OL2000/engine3.html>

Ισπανικά, Πορτογαλικά. Μειονεκτήματά της είναι τα ασυνεπή αποτελέσματα, η εμφάνιση μόνο 10 αναφορών κάθε φορά, η έλλειψη κατάταξης με άλλο τρόπο εκτός από το βαθμό σχετικότητας, ο μεγάλος αριθμός των αποτελεσμάτων.

4. Google (<http://www.google.com>)

Τέταρτη σε κατάταξη μεγέθους είναι η Google. Πρόκειται για μια από τις νεότερες μηχανές αναζήτησης, καθώς επισήμως «βγήκε στον αέρα» το Σεπτέμβριο του 1999. Η δημοτικότητά της έγκειται στο γεγονός ότι κατατάσσει τις ιστοσελίδες με βάση την ανάλυση των συνδέσεων (links). Αυτό μπορεί να αποβεί χρήσιμο στην ανεύρεση κατάλληλων θέσεων (sites), ως απάντηση σε γενικές ερωτήσεις, π.χ. cars.

Τα αποτελέσματα της έρευνας προέρχονται από τη βάση δεδομένων της Google, την Open Directory, και τη RealNames. Προσφέρει επίσης εξειδικευμένες ομάδες θεμάτων: μια κυβερνητική βάση δεδομένων, έρευνα στο Stanford University, έρευνα στη Linux και εξειδικευμένη έρευνα για Apple/Macintosh. Χρησιμοποιείται από τη θέση ιστού της Netscape.

Για την έρευνα χρησιμοποιεί το AND, αλλά και τα σημεία + και -, φράσεις σε εισαγωγικά, ενώ δε διαθέτει αποκοπή ή αυτόματη αναζήτηση ενικού / πληθυντικού. Προσφέρει πολλά πεδία έρευνας, βρίσκει με τη σύνδεση (link) υπερκείμενες συνδέσεις στον ακριβή URL ή σε σελίδες όμοιες με τα συνδεόμενα στοιχεία στον URL και αγνοεί τις συχνά εμφανιζόμενες λέξεις. Τα αποτελέσματα κατατάσσονται σύμφωνα με τη σχετικότητα στο ερώτημα, η οποία καθορίζεται από τις συνδέσεις με άλλες σελίδες.

Ανάμεσα στα δυνατά σημεία της Google είναι η σχετικότητα στην απόκριση στην έρευνα που βασίζεται στις συνδέσεις της θέσης (site). Η αρχαιακή κρυφή μνήμη (cache) μπορεί να αναζητήσει μικρές λέξεις. Ως χαρακτηριστικό που τη διαφοροποιεί, είναι ότι παρέχει πρόσβαση σε ιστοσελίδες την ώρα που ευρετηριάζονται και έχει τα καλύτερα κριτήρια που ορίζουν τη σχετικότητα των θεμάτων. Ανάμεσα στα αδύνατα σημεία της είναι τα περιορισμένα πεδία για αναζήτηση και ότι οι συνδεόμενες έρευνες πρέπει να αναγράφονται επακριβώς.

5. **Excite (<http://www.excite.com>)**

Η Excite βρίσκεται ανάμεσα στις δημοφιλείς μηχανές αναζήτησης, αν και είναι από τις μικρότερες. Εμφανίστηκε στο τέλος του 1995. Είναι κατάλληλη για έρευνες σε ευρέα, γενικά θέματα. Προσφέρει όχι μόνο μια μηχανή αναζήτησης, αλλά και ένα κατάλογο (directory) με κατηγορίες.

Η αναζήτηση γίνεται στη δική της βάση δεδομένων, η οποία περιέχει και ειδήσεις, αλλά και σε πληροφοριακές βάσεις δεδομένων, όπως λεξικά και εγκυκλοπαίδειες. Για την έρευνα υποστηρίζει τους τελεστές Boole (AND, OR, AND NOT και το MUST με κεφαλαία), το + και το -, φράσεις με εισαγωγικά. Αγνοεί τις μικρές λέξεις μέσα στη φράση, δε διαθέτει αποκοπή, ούτε έρευνα σε πεδία.

Η Excite αυτόματα περιορίζεται σε έρευνες στην Αγγλική γλώσσα, εκτός αν επιλεγεί μια γλώσσα αναζήτησης. Έχει περιορισμό κατά χώρα και κατά πεδίο ορισμού (domain), όπως χρησιμοποιούνται στις ΗΠΑ. Η κατάταξη των αποτελεσμάτων γίνεται σύμφωνα με τη σχετικότητα τους. Προσθέτει επιπλέον στοιχεία στην αναζήτηση όχι μόνο από τον ιστό (web), αλλά και από τίτλους ειδήσεων, αποτελέσματα αθλημάτων, πληροφορίες για εταιρείες. Παρέχει μια φιλική για το χρήστη θέση (site) για ταξιδιωτικές πληροφορίες και για κρατήσεις αεροπορικών εισιτηρίων.

Από τα πλεονεκτήματα της Excite αναφέρονται η υψηλή σχετικότητα σε δημοφιλή θέματα και ιδιαίτον χαρακτηριστικό είναι οι δυνατότητές της να παρέχει εξατομικευμένες απαντήσεις. Στα μειονεκτήματά της αναφέρονται οι τελεστές Boole που πρέπει να γράφονται με κεφαλαία, είναι μικρή βάση δεδομένων και δεν επιτρέπει την αποκοπή ή την έρευνα σε πεδία.

6. **Inktomi (<http://www.inktomi.com>)**

Η μηχανή αναζήτησης Inktomi ξεκίνησε αρχικά από το UC Berkeley. Αργότερα οι δημιουργοί της σχημάτισαν τη δική τους εταιρεία με το ίδιο όνομα και μαζί το ευρετήριο Inktomi, που αρχικά υποστήριζε τη HotBot.

Σήμερα είναι μια βάση δεδομένων και μηχανή αναζήτησης πίσω από τις γνωστές μηχανές αναζήτησης στο web. Παρέχει δεδομένα για μηχανές αναζήτησης όπως η HotBot, MSN Web Search, Snap, Yahoo!, GoTo¹¹, iWon¹², GeoCities, Anzwers, Canada.com, AOL NetFind. Τον Ιανουάριο του 2000 η Inktomi δημοσίευσε πληροφορίες για τη βάση δεδομένων της WebMap με πάνω από 1 δις τεκμήρια. Η βάση δεδομένων WebMap δεν είναι διαθέσιμη για έρευνα. Αντίθετα οι συνεργάτες της Inktomi επιτρέπεται να έχουν πρόσβαση σε αρκετά μικρότερο αριθμό δεδομένων.

Η αυστραλιανή μηχανή αναζήτησης Anzwers παρέχει περιεκτικά αποτελέσματα και όλα σχεδόν τα πεδία της προχωρημένης αναζήτησης. Η MSN Advanced Search διαθέτει μεγαλύτερη βάση δεδομένων από την Anzwers και κάποια μοναδικά χαρακτηριστικά. Οι συνεργάτες της Inktomi έχουν πρόσβαση στο κοινό της ευρετήριο, αλλά ξεχωρίζουν μεταξύ τους έχοντας καθένας τα δικά του χαρακτηριστικά. Το ευρετήριο της Inktomi δεν προσεγγίζεται ευθέως, αλλά μόνο μέσω της συνεργατών της.

7. **Lycos (<http://www.lycos.com>)**

Η μηχανή αναζήτησης Lycos είναι από τις πιο παλιές, αφού ξεκίνησε το Μάρτιο του 1994. Ενώ ξεκίνησε ως μηχανή αναζήτησης, από τον Απρίλιο του 1999 κινήθηκε προς την κατεύθυνση του καταλόγου, όπως το Yahoo! Οι κύριες κατηγορίες προέρχονται από την Open Directory, από ειδησεογραφικές βάσεις δεδομένων, από τις Web Pages (η βάση δεδομένων του Lycos). Τον Οκτώβριο του 1998 η Lycos αγόρασε τη HotBot, η οποία συνεχίζει να λειτουργεί αυτόνομα.

Επιτρέπει χρήση πολλαπλών όρων για έρευνα, χρησιμοποιεί τη σύνταξη με + και -, αλλά δεν έχει τελεστές Boole. Επιτρέπει τη χρήση εισαγωγικών για φράσεις, αλλά δεν υποστηρίζει την αποκοπή, ούτε διαθέτει πεδία έρευνας. Στην προχωρημένη έρευνα διατίθενται ορισμένα πεδία για αναζήτηση (τίτλος, διαθέτης, URL). Δεν έχει περιορισμούς,

¹¹ Η μηχανή αναζήτησης GoTo (<http://www.goto.com>) ξεκίνησε το 1997 και πουλά τις λίστες της, δηλαδή οι εταιρείες πληρώνουν για να καταταχθούν σε υψηλότερη κλίμακα ως προς τα αποτελέσματα της έρευνας.

αλλά μόνο στη γλώσσα, υποστηρίζοντας 25 γλώσσες. Η κατάταξη των αποτελεσμάτων γίνεται κατά βαθμό σχετικότητας. Εμφανίζονται μόνο δέκα αναφορές κάθε φορά.

Η Lycos είναι κατάλληλη για απλές έρευνες, είναι εύκολη στη χρήση και επιτρέπει την έρευνα σε φυσική γλώσσα.

8. Go / Infoseek (<http://www.go.com>)

Η Infoseek ξεκίνησε το 1995 και είναι από τις καλά οργανωμένες μηχανές αναζήτησης. Σήμερα αποτελεί τμήμα της πύλης Disney Go Network. Διαθέτει ένα κατάλογο χωριστό από τη μηχανή αναζήτησης, κατηγοριοποιημένο κατά θέμα, επομένως προσφέρεται για εξειδικευμένες, ακριβείς και γρήγορες έρευνες. Όμως το περιορισμένο μέγεθος οδηγεί τον ερευνητή να ψάξει και σε άλλες μηχανές αναζήτησης.

Επιτρέπει έρευνα με πολλαπλούς όρους και με εισαγωγικά, δε διαθέτει τελεστές Boole, αλλά σύνταξη με + και -. Δε διαθέτει αποκοπή, έχει περιορισμένα πεδία αναζήτησης. Η κατάταξη των αποτελεσμάτων γίνεται κατά βαθμό σχετικότητας και κατά θέση (site). Η Infoseek και η Go πρόσθεσαν μια υπηρεσία μετάφρασης τον Απρίλιο του 1999 ανάμεσα στα Αγγλικά, Γαλλικά, Γερμανικά, Ιταλικά, Ισπανικά και Πορτογαλικά.

Στα θετικά της σημεία εντάσσονται η κατάταξη των αποτελεσμάτων κατά θέση και χρονολογία, αλλά και έχει περιεχόμενα πύλης με πλούσιες συμπληρωματικές πηγές. Αρνητικά σημεία της είναι τα λιγότερο δυνατά χαρακτηριστικά αναζήτησης και η μικρή βάση δεδομένων.

9. HotBot (<http://www.hotbot.com>)

Πρόκειται για μια ακόμα γνωστή μηχανή αναζήτησης. Η HotBot ξεκίνησε το Μάιο του 1996 ως είσοδος της Wired Digital στην αγορά των μηχανών αναζήτησης. Τον Οκτώβριο του 1998 η Lycos αγόρασε την Wired Digital και σήμερα η HotBot λειτουργεί ως αυτόνομη

¹² Η μηχανή αναζήτησης iWon (<http://www.iwon.com>) υποστηρίζεται από το αμερικανικό δίκτυο CBS και ξεκίνησε το φθινόπωρο του 1999.

μηχανή αναζήτησης. Η HotBot χρησιμοποιεί την Open Directory¹³. Από τα αποτελέσματα της έρευνας, η πρώτη σελίδα προέρχεται από την υπηρεσία Direct Hit, ενώ τα δευτερεύοντα αποτελέσματα προέρχονται από τη μηχανή αναζήτησης Inktomi. Αντλεί τις πληροφορίες καταλόγου από το πρόγραμμα Open Directory.

Η HotBot προσφέρει πλήρη αναζήτηση με τους γνωστούς τελεστές Boole (AND, OR, AND NOT), με παρενθέσεις, με εισαγωγικά για φράσεις, με χρήση συμβόλων, με σύνταξη με + και -. Στη HotBot λειτουργεί η αποκοπή, ενώ προσφέρει ποικιλία δυνατοτήτων σε αναζήτηση μέσα σε πεδία, π.χ. τίτλος, πεδίο ορισμού (domain), πίνακες, video, audio, χρονολογία. Υπάρχει δυνατότητα περιορισμών χρόνου, γλώσσας, ηπείρου, τοποθεσίας.

Τα αποτελέσματα κατατάσσονται σύμφωνα με τη σχετικότητα σε ομάδες. Μοναδικό χαρακτηριστικό της HotBot, άρα θετικό σημείο είναι ο περιορισμός κατά Page Type. Η επιλογή Top Page περιορίζει τα αποτελέσματα στην κεντρική σελίδα μιας θέσης ιστού. Η Page Depth με τον αριθμό 3 προεπιλεγμένο αναφέρεται στην ιεραρχική δομή υποκαταλόγων και η Personal Page αναφέρεται στη δομή του URL που προσδιορίζει μια σελίδα ως προσωπική.

Οι προχωρημένες δυνατότητες αναζήτησης αποτελούν άλλο θετικό στοιχείο, καθώς και η ευκολία στην εμφάνιση των αποτελεσμάτων. Στις αδυναμίες κατατάσσονται η συστοιχία (clustering) των αποτελεσμάτων, ενώ οι συνδεδεμένες έρευνες πρέπει να διατυπώνονται επακριβώς. Το μέγεθος της βάσης δεδομένων έχει συρρικνωθεί και κατηγορείται ότι της λείπει η ανανέωση.

10. Webcrawler (<http://www.webcrawler.com>)

Η μηχανή αναζήτησης Webcrawler άνοιξε στο κοινό τον Απρίλιο του 1994, ως ερευνητικό πρόγραμμα του University of Washington. Το Μάρτιο του 1995 η America Online την αγόρασε και τη χρησιμοποιούσε στην online υπηρεσία της. Αλλά το Νοέμβριο

¹³ Η Open Directory χρησιμοποιεί εθελοντές εκδότες για να καταλογογραφήσουν το web. Ξεκίνησε τον Ιούνιο του 1998, αγοράστηκε από τη Netscape το Νοέμβριο του 1998 και η εταιρεία απαίτησε ότι καθένας μπορεί να χρησιμοποιήσει τις πληροφορίες μετά από υπογραφή σχετικής σύμβασης, κάτι που πρώτη έκανε η Netscape.

του 1996 η Excite αγόρασε την Webcrawler, η οποία όμως συνεχίζει να λειτουργεί ως ανεξάρτητη μηχανή αναζήτησης.

Έχει μικρό ευρετήριο και σχετίζεται με κοινά θέματα, είναι εύκολη και γρήγορη στη χρήση, καθώς παρέχει αποτελέσματα που δεν επικαλύπτονται από άλλα, λόγω του μικρού μεγέθους της βάσης δεδομένων. Είναι κατάλληλη για γενικές ερωτήσεις.

11. Yahoo! (<http://www.yahoo.com>)

Αν και η Yahoo! δεν είναι ακριβώς μια μηχανή αναζήτησης, περιλαμβάνεται στην παρούσα περιγραφή των κυριότερων ονομάτων. Πρόκειται για το μεγαλύτερο, παλαιότερο – από το 1994 - και πιο γνωστό κατάλογο ιστοσελίδων. Έχει καλή φήμη, επειδή βοηθά τους ανθρώπους να εντοπίσουν εύκολα πληροφορίες. Το μυστικό της Yahoo! είναι οι άνθρωποι ως δημιουργοί του, καθώς 150 εκδότες εργάζονται για την κατηγοριοποίηση του ιστού.

Η Yahoo! απαριθμεί περίπου 1,2 εκατ. θέσεις (sites). Αν δεν βρεθούν στοιχεία στο δικό της κατάλογο, η έρευνα αυτόματα περνά στη μηχανή αναζήτησης Inktomi.

Δεν επιτρέπει πλήρη αναζήτηση Boole, διαθέτει έρευνα με φράσεις και εισαγωγικά, η αποκοπή λειτουργεί αυτόματα για τους αναζητούμενους όρους, ενώ επιτρέπει έρευνα σε δύο πεδία (fields) τον τίτλο και τον URL.

Η παράθεση των αποτελεσμάτων γίνεται σύμφωνα με τις 72 κατηγορίες και υποκατηγορίες της Yahoo! ανάλογα με τη σχετικότητα τους στην απόκριση της έρευνας.

Τα κριτήρια επιλογής των αναγραφών προέρχονται από την προτίμηση των ανθρώπων που το χρησιμοποιούν ως χρήστες. Υπάρχουν παραλλαγές στη Yahoo! σε όλο τον κόσμο, δείγμα της δημοτικότητάς της.

V. Τελευταίες λέξεις

Με τις παραπάνω αναφορές στις μηχανές αναζήτησης έγινε προσπάθεια να παρουσιαστεί η σπουδαιότητα των εργαλείων αυτών του Internet και η χρησιμότητά τους ως εργαλεία πληροφόρησης των χρηστών της βιβλιοθήκης. Έτσι στις καθημερινές δραστηριότητες ενός βιβλιοθηκονόμου πληροφόρησης προστίθενται οι υποδείξεις προς τους

χρήστες της κατάλληλης μηχανής αναζήτησης, πέρα από τις παραδοσιακές βάσεις δεδομένων και τις ακόμα παραδοσιακές ή παρωχημένες, θα λέγαμε, κάρτες δελτιοκαταλόγου.

Γίνεται εμφανές ότι οι απαιτήσεις για τις γνώσεις και δεξιότητες του σημερινού βιβλιοθηκονόμου / ειδικού πληροφόρησης (information specialist, information professional) ή cybretarian ή όπως αλλιώς τείνει να μετονομαστεί, είναι εντονότερες. Χρειάζεται να είναι συνεχώς ανοιχτός σε κάθε αλλαγή και δεκτικός των νέων τάσεων και τεχνολογιών. Επομένως θα συνεχίσει τον πειραματισμό του με τις σημερινές μηχανές αναζήτησης, εγκλιματιζόμενος στις αλλαγές και περιμένοντας τους επιστήμονες να ανακαλύψουν νέα προϊόντα και τρόπους αναζήτησης που φέρνει ο 21ος αιώνας.

Επειδή το μόνο σταθερό είναι η αλλαγή, μέχρι να δει κάτι νεότερο, ο ερευνητής θα συνεχίζει να συλλέγει, αποτιμά, οργανώνει χρήσιμες ιστοσελίδες μέσα από τις μηχανές αναζήτησης και θα εξοικειώνεται με τον τρόπο έρευνας σε αυτές. Επειδή, όπως αποδείχτηκε πιο πάνω, οι μηχανές αναζήτησης εμφανίζονται, αναπτύσσονται, ωριμάζουν, ακμάζουν και ορισμένες από αυτές εξαφανίζονται ή απορροφούνται από άλλες, ο ερευνητής και ο βιβλιοθηκονόμος χρειάζεται να ενημερώνονται για τις τρέχουσες εξελίξεις στο θέμα αυτό.

Κατάλληλες, σωστά και καίρια ενημερωμένες ιστοσελίδες, οι οποίες και χρησιμοποιήθηκαν ευρέως για τις ανάγκες αυτού του άρθρου είναι η Search Engine Showdown (<http://www.notess.com>) Greg Notess, η οποία αποτιμά τα χαρακτηριστικά και την κάλυψη των μηχανών αναζήτησης, παρακολουθεί την εξέλιξή τους και την εμφάνιση νέων και παρέχει πολύτιμα στατιστικά στοιχεία. Μια δεύτερη ιστοσελίδα είναι της Search Engine Watch (<http://www.searchenginewatch.com>) του Danny Sullivan που περιλαμβάνει ειδήσεις, κατατάξεις, βελτιώσεις και εξελίξεις στο σύνολο των μηχανών αναζήτησης, προσφέροντας επίσης ένα δωρεάν newsletter μέσω e-mail.

Βιβλιογραφία

1. Bradley, Phil, "Virtual libraries and Internet searching, *Online & CD-ROM Review*, v.23 (1999), no.6, pp.353-355.

2. Kassel, Amelia, "Internet power searching: finding pearls in a zillion grains of sand," *Information Outlook*, v.3 (Apr. 1999) no.4, pp.28-32.
3. Kassler, Helene, "Competitive intelligence on the Internet: going for the gold," *Information Outlook*, v.4 (Feb. 2000) no. 2, pp.37-42.
4. Notess, Greg at <http://www.notess.com>
5. Sullivan, Danny at <http://www.searchenginewatch.com>
6. Webber, Sheila, "Search engines and news services: developments on the Internet," *Business Information Review*, v.15 (Dec. 1998) no.4 pp.229-237.
7. Xie, M., H. Wang and T.N. Goh, "Quality dimensions of Internet search engines," *Journal of Information Science*, v.24 (1998) no. 5 pp.365-372.
8. Για συμβουλές για τη χρήση μηχανών αναζήτησης ανάλογα με το θεματικό αντικείμενο, βλ. <http://nuevaschool.org/debbie/library/research/adviceengine.html>

Διευθύνσεις των μηχανών αναζήτησης και βάσεων δεδομένων που αναφέρθηκαν

1. AltaVista: <http://www.altavista.com>
2. Amazon.com: <http://www.amazon.com>
3. AOL NetFind: <http://search.aol.com>
4. Ask Jeeves: <http://www.askjeeves.com>
5. Anzwers: <http://www.anzwers.com.au>
6. Business Information Review: http://www.bowker-saur.co.uk/products/catalog/business/bir_j.htm
7. Canada.com: <http://www.canada.com>
8. Dialog: <http://www.dialog.com>
9. Dogpile: <http://www.dogpile.com>
10. Excite: <http://www.excite.com>
11. FAST: <http://www.alltheweb.com>
12. GeoCities: <http://geocities.yahoo.com>
13. Go / Infoseek: <http://www.go.com>

14. GoTo: <http://www.goto.com>
15. Google: <http://www.google.com>
16. HotBot: <http://www.hotbot.com>
17. Information Outlook: <http://www.informationoutlook.com>
18. Inktomi: <http://www.inktomi.com>
19. iWon: <http://www.iwon.com>
20. Journal of Information Science: <http://iis.org.uk/JIS.html>
21. LEXIS-NEXIS: <http://www.lexis-nexis.com>
22. Lycos: <http://www.lycos.com>
23. MSN Web Search: <http://search.msn.com>
24. Netscape Search: <http://search.netscape.com>
25. Northern Light: <http://www.northernlight.com>
26. OCLC First Search: <http://www.oclc.com>
27. Online & CD-ROM Review: <http://www.learned.co.uk/olr/> (η διεύθυνση σύντομα θα αλλάξει, επειδή το περιοδικό αγοράστηκε από τη MCB Press: <http://www.mcb.co.uk>)
28. Open Directory: <http://dmoz.org>
29. RealNames: <http://www.realnames.com>
30. Search Engine Showdown: <http://www.notess.com>
31. Search Engine Watch: <http://www.searchenginewatch.com>
32. SilverPlatter: <http://www.silverplatter.com>
33. Snap: <http://www.snap.com>
34. STN: <http://www.stn.com>
35. Webcrawler: <http://www.webcrawler.com>
36. Yahoo!: <http://www.yahoo.com>