

ΠΑΙΔΑΓΩΓΙΚΗ ΑΞΙΟΠΟΙΗΣΗ ΤΩΝ ΤΠΕ

Μηχανές αναζήτησης
Όλα όσα θέλατε να μάθετε και
δεν τολμούσατε να ρωτήσετε



Terra Computerata

Διαδίκτυο - εν αρχή ην το χάος

- Βιβλιοθήκη ή αποθήκη ψηφιακών αντικειμένων;
- Υπάρχει καταλογράφηση;
- **Πύλες** οργανώνουν θεματικά – ανθρώπινος παράγων
- **Μηχανές αναζήτησης**
- Αυτόματοι κατάλογοι robots-crawlers
- **Μεταμηχανές**

Ζουσάμι άνοιξε

Και τω κρούοντι ανοιγήσεται...



Terra Computerata @ www.netschoolbook.gr

Μηχανές αναζήτησης

3 επίπεδα επεξεργασίας

Συλλογή - crawling

Καταλογράφηση – ευρετήριο - indexing

Λογισμικό αναζήτησης – searching

Ιδιαίτερα χαρακτηριστικά κάθε μηχανής

Terra Computerata @ www.netschoolbook.gr

Crawling

Ο Crawler:

- Είναι ένα πρόγραμμα που επισκέπτεται εκατομμύρια σελίδων web
- ακολουθεί όλα τα links κάθε σελίδας που επισκέπτεται

Terra Computerata @ www.netschoolbook.gr

Indexing

- Τα περιεχόμενα κάθε σελίδας αναλύονται για να καθοριστεί τι θα καταγραφεί σε ένα ευρετήριο (index)
 - Διάφορες λέξεις εξάγονται από τις επικεφαλίδες ή από κάποια πεδία που λέγονται **meta tags**
- Οι παραπάνω λέξεις αποτελούν τα δεδομένα που θα καταγραφούν σε μία βάση δεδομένων (Indexing database).
 - Google: αποθηκεύει όλη ή τμήμα κάθε σελίδας που επισκέπτεται (cache) καθώς και πληροφορίες για τη σελίδα. Αυτό είναι χρήσιμο όταν η αρχική σελίδα δεν υπάρχει πλέον

Terra Computerata @ www.netschoolbook.gr

Αρχές ευρετηρίασης

- Τίτλος ιστοσελίδας
- Κείμενο επικεφαλίδας στις πρώτες παραγράφους
(πιο πιθανό το θέμα να εμφανίζεται στην αρχή παρά στη μέση ή στο τέλος μιας ιστοσελίδας)
- Λέξεις κλειδιά στον κώδικα HTML – search engine spamming
- Επαναλαμβανόμενες λέξεις

Terra Computerata @ www.netschoolbook.gr

Searching (αναζήτηση)

- Όταν ο χρήστης χρησιμοποιεί μία μηχανή αναζήτησης, σχηματίζει ένα **ερώτημα (query)**, το οποίο αποτελείται από **λέξεις κλειδιά**
- Στη συνέχεια η μηχανή αναζήτησης ψάχνει στο **ευρετήριο** για να βρει τις σελίδες web που ταιριάζουν όσο το δυνατόν περισσότερο στα κριτήρια του χρήστη.
 - Παρουσιάζει **κατάλογο** αυτών των σελίδων καθώς και μία μικρή περίληψη κάθε σελίδας
- Η χρησιμότητα μίας μηχανής αναζήτησης εξαρτάται από το **πόσο σχετικά** είναι τα αποτελέσματα αναφορικά με το ερώτημα του χρήστη

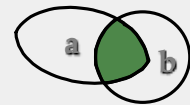
Terra Computerata @ www.netschoolbook.gr

Οι μηχανές αναζήτησης όπως και οι βάσεις δεδομένων χρησιμοποιούν τους **Τελεστές Bool** για να εκτελέσουν μία αναζήτηση π.χ.
οικογενειακή **and**
βία **not** σεξουαλική

Terra Computerata @ www.netschoolbook.gr

Λογικοί Τελεστές Bool

AND



OR



NOT



Προβλήματα

- Πλήθος αποτελεσμάτων
- Σχετικότητα
- Διαφορετικά αποτελέσματα

Terra Computerata @ www.netschoolbook.gr

Ποιος πληρώνει;

- Η διαφήμιση
- Συστηματική καταγραφή προτιμήσεων
- Πανεπιστήμια Αμερικής

Terra Computerata @ www.netschoolbook.gr

Διαφορές μεταξύ μηχανών

Ως προς:

- Τον αριθμό των αποτελεσμάτων
- Την ενημέρωσή τους
- Τη σειρά παρουσίασης τους στη λίστα
- Το ίδιο το περιεχόμενο (είδος ψηφιακών αντικειμένων – πλήρες κείμενο – pdf)

Terra Computerata @ www.netschoolbook.gr

Ζητήματα που ανακύπτουν

- Σχέσεις με κρατικούς φορείς
- Εταιρείες πληρώνουν για να εμφανίζονται πρώτες στις λίστες

Terra Computerata @ www.netschoolbook.gr

Η περίπτωση google

Pagerank

Αριθμός δεσμών ως δείκτης αξίας κάθε σελίδας

Ψήφοι από σημαντικές σελίδες έχουν μεγαλύτερο βάρος

Σύνθετες μέθοδοι δυσκολεύουν την ανθρώπινη παρέμβαση

Ναι στις διαφημίσεις αλλά δεν πωλούνται θέσεις παρουσίασης
Google-bombing

Terra Computerata @ www.netschoolbook.gr

Μέγεθος web και μηχανές αναζήτησης

- Μία μηχανή αναζήτησης (Google) μπορεί να λαμβάνει και 200 εκατομμύρια ερωτήματα κάθε ημέρα
- Έχει πληροφορίες (Ιούνιος 2006),
 - για 8.05 δισεκατομμύρια σελίδες web
 - 1.3 δισεκατομμύρια εικόνες
 - Περισσότερο από 1 δισεκατομμύριο Usenet messages
 - Σύνολο 10.4 δισεκατομμύρια πληροφορίες!

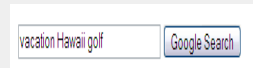
Terra Computerata @ www.netschoolbook.gr

Αναζήτηση πληροφοριών

- Αναζήτηση κειμένου
- Εικόνας
- Μηνύματος σε Newsgroup
- Δορυφορικής φωτογραφίας (google earth)

Terra Computerata @ www.netschoolbook.gr

Αναζητώντας κείμενο στο Google -1



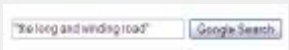
Γίνετε όσο το δυνατόν **σαφέστεροι** στις πληροφορίες που αναζητάτε. Οι τρεις όροι σημαίνουν αναζήτηση και με vacation και Hawaii και gold



Το google **διαγράφει** από τους όρους που του δίνετε κάποιες **κοινές λέξεις**, όπως *from, the*, καθώς και μερικά απομονωμένα γράμματα και αριθμούς. Αν θέλετε να συμπεριληφθούν στην αναζήτηση πρέπει να προτάξετε το +. Η να βάλετε όλη την πρόταση σε διπλά εισαγωγικά

Terra Computerata @ www.netschoolbook.gr

Αναζητώντας κείμενο στο Google -2



Αν θέλετε να αναζητήσετε μια **συγκεκριμένη φράση** πρέπει να την περιλάβετε σε εισαγωγικά



Αν θέλετε να **αποκλείσετε** κάποια λέξη βάζετε το - προ αυτής. Π.χ. αν ψάχνετε για το ψάρι bass (πέρκα) και όχι για τον μουσικό όρο πρέπει να γράψετε όπως φαίνεται αριστερά

Terra Computerata @ www.netschoolbook.gr

Αναζητώντας κείμενο στο Google -3



Αν επιλέξετε το κουμπί I'm feeling Lucky το google θα σας πιάσει στην **πρώτη επιλογή** από τα αποτελέσματα που επιστρέφει.

Terra Computerata @ www.netschoolbook.gr

Αναζητώντας επιστημονική γνώση



Google scholar μπορεί να ανακαλύψει επιστημονικά άρθρα

Terra Computerata @ www.netschoolbook.gr

Αναζητώντας πληροφορίες σε ομάδες συζητήσεων



Αναζήτηση σε **newsgroups** των ορών alpha romeo brea. Το google βρήκε σχετικές πληροφορίες στο newsgroup alt.autos.ferrari

Terra Computerata @ www.netschoolbook.gr

Προσφερόμενες υπηρεσίες από το Google



Terra Computerata @ www.netschoolbook.gr

Κάτσε κάτω από την μπάρα!

Toolbars google - yahoo

Terra Computerata @ www.netschoolbook.gr

Multimedia

- <http://www.alltheweb.com/>
- <http://www.altavista.com/> images, audio,video
- <http://www.ditto.com/>
- <http://images.google.com>
- <http://www.kazaa.com/>
- <http://multimedia.lycos.com/>
- <http://www.musicrobot.com/>
- <http://www.music-finder.net/>
- <http://www.picsearch.com/>
- <http://www.singingfish.com/> (audio/video)

Terra Computerata @ www.netschoolbook.gr

Προκλήσεις για τις μηχανές αναζήτησης

- Το web μεγαλώνει ταχύτερα από ό,τι οποιαδήποτε τεχνολογία μηχανών αναζήτησης μπορεί να κατατάξει
- Πολλές σελίδες web αλλάζουν συχνά, συνεπώς πρέπει οι μηχανές αναζήτησης να τις επισκεπτονται περιοδικά
- Ως επί το πλείστον, οι ερωτήσεις αφορούν σε λέξεις κλειδιά (keywords) και τα αποτελέσματα έχουν πολλά **false positives**
- Για τους δικτυακούς τόπους που δημιουργούνται δυναμικά μπορεί να είναι δύσκολο ή και αδύνατο να τους καταγράψει ένας web crawler
- Πολλοί δικτυακοί τόποι, που δημιουργούνται δυναμικά, δεν καταγράφονται από μηχανές αναζήτησης αυτό το φαινόμενο είναι γνωστό ως το **αόρατο web (invisible web)**
- Μερικές μηχανές **δεν** ταξινομούν τα αποτελέσματα βάσει της σχετικότητας με τις ερωτήσεις του χρήστη, αλλά ανάλογα με τα χρήματα που έχει πληρώσει το site
- Μερικά site μετέρχονται τεχνασμάτων, ώστε η μηχανή αναζήτησης να τα εμφανίζει ως έχοντά υψηλή συσχέτιση

Στο εργαστήριο

- Εντοπίστε λέξεις –κλειδιά
- Καταγράψτε αριθμό αποτελεσμάτων – σχετικότητα σελίδων – βαθμό ικανοποίησης – τυχόν προβλήματα (ένα μέλος παρουσιάζει τα συμπεράσματα στην ολομέλεια)
- Επιλέξτε τη σύνθετη αναζήτηση και δοκιμάστε να βρείτε α) εικόνες π.χ. για τον 2^ο Παγκόσμιο πόλεμο
β) αρχεία παρουσιάσεων γ) βίντεο
- Εγκατάσταση toolbar

Terra Computerata @ www.netschoolbook.gr

webligraphy

- <http://searchenginewatch.com>

Terra Computerata @ www.netschoolbook.gr