

ΕΠΛ 001:
ΕΙΣΑΓΩΓΗ ΣΤΗΝ
ΕΠΙΣΤΗΜΗ ΤΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ

Μηχανές αναζήτησης

Στόχοι

1

- Να εξηγήσουμε για ποιον λόγο μας είναι απαραίτητες οι μηχανές αναζήτησης στον Παγκόσμιο Ιστό.
- Να περιγράψουμε κάποιους από τους βασικότερους **κανόνες χρήσης** της μηχανής αναζήτησης της Google.
- Να σκιαγραφήσουμε την αρχιτεκτονική μιας μηχανής αναζήτησης.
- Να αναφέρουμε κάποιες από τις σοβαρότερες **προκλήσεις** που πρέπει να αντιμετωπίσουν οι μηχανές αναζήτησης στο μέλλον.

Εντοπισμός χωρίς URL

2

Ο Ανέστης είναι οπαδός της Ανόρθωσης και θέλει να βρει πληροφορίες για την αγαπημένη του ομάδα: στατιστικά για το τρέχον πρωτάθλημα, στατιστικά για προηγούμενα πρωταθλήματα, φωτογραφίες, βίντεο, ιστορικά στοιχεία, κλπ.

Η καλύτερη πηγή αυτών των πληροφοριών είναι, φυσικά, ο επίσημος ιστοτόπος της Ανόρθωσης. Όμως, ο Ανέστης δεν γνωρίζει τον εντοπιστή (URL) αυτού του ιστοτόπου.

Τι γίνεται; Πώς μπορεί ο Ανέστης να εντοπίσει τις επίσημες ιστοσελίδες της ομάδας του;

Εντοπισμός χωρίς URL

3

Θα ήταν ωραίο να υπήρχε ένας κατάλογος (σαν τον τηλεφωνικό), όπου ο Ανέστης να μπορούσε να αναζητήσει τις ιστοσελίδες που τον ενδιαφέρουν.

Όμως αυτό είναι αδύνατον. Το τοπικό τηλεφωνικό δίκτυο είναι μικρό, αλλάζει αργά, και ελέγχεται από κεντρική αρχή. Αντιθέτως, ο Παγκόσμιος Ιστός

- **είναι τεράστιος** (περισσότερες από 18 δισεκατομμύρια ιστοσελίδες την Δευτέρα 27 Απριλίου 2009).
- **αλλάζει διαρκώς** (ο καθένας μπορεί να δημιουργεί/τροποποιεί/ διαγράφει ιστοσελίδες, κάθε στιγμή).
- **δεν ελέγχεται κεντρικά** (κανείς δεν χρειάζεται ειδική άδεια για να δημιουργήσει/τροποποιήσει/διαγράψει τις ιστοσελίδες του).

Μηχανές αναζήτησης

4

Για την διευκόλυνση της αναζήτησης πληροφοριών στον Παγκόσμιο Ιστό, έχουν δημιουργηθεί ειδικά εργαλεία που ονομάζονται **μηχανές αναζήτησης** (*search engines*).

Μια μηχανή αναζήτησης μας επιτρέπει να εκτελούμε αναζητήσεις βάσει κάποιων κριτηρίων. Το βασικότερο είναι οι **λέξεις-κλειδιά** (*keywords*, δηλ. λέξεις που θέλουμε/δεν θέλουμε να περιέχουν οι αναζητούμενες ιστοσελίδες). Άλλα κριτήρια είναι η *γλώσσα* που χρησιμοποιεί η ιστοσελίδα, η *ημερομηνία* δημιουργίας της, κ.ά.

Η δημοφιλέστερη μηχανή αναζήτησης είναι αυτή της εταιρίας Google. Άλλες μηχανές είναι οι Yahoo! Search (Yahoo), Bing (Microsoft), Ask (Ask.com), Wolfram Alpha, κ.ά.

Μηχανές αναζήτησης

5



www.google.com

[Advanced Search](#) [Preferences](#) [Language Tools](#)

[Web](#) | [Images](#) | [Video](#) | [Local](#) | [Shopping](#) | [more](#) ▾

search.yahoo.com

[Options](#) ▾



www.bing.com



[web](#), [images](#), [news](#), [videos](#), [maps](#), and [lots of answers](#).

www.ask.com

Skins  [Advanced](#)



Enter what you want to **calculate** or **know about**:

www.wolframalpha.com

Αναζήτηση κειμένου

6

Για να αναζητήσουμε ιστοσελίδες που περιέχουν κάποια λέξη-κλειδί, δίνουμε απλώς αυτήν τη λέξη:

The Google logo is displayed in its characteristic multi-colored font: 'G' in blue, 'o' in red, 'o' in yellow, 'g' in blue, 'l' in green, and 'e' in red. A small 'TM' trademark symbol is located to the upper right of the 'e'.

Ανόρθωση

Google Search

I'm Feeling Lucky

[Advanced Search](#)

[Preferences](#)

[Language Tools](#)

Αναζήτηση κειμένου

7

The screenshot shows a Google search interface with the query 'Ανόρθωση'. The search results are displayed in a list format. The first result is from 'Επίσημη Ιστοσελίδα Ανόρθωσης Αμμοχώστου', the second is from 'Anorthosis Famagusta Official ...', and the third is from 'Anόρθωση - Βικιπαίδεια'. Below the text results, there are image results and news results for the same query. The search bar shows the query and the search button. The top navigation bar includes links for Web, Images, Maps, News, Video, Gmail, and more. The search results are sorted by relevance, and the first 10 results are shown.

Η μηχανή μάς επιστρέφει έναν κατάλογο ιστοσελίδων που περιέχουν την λέξη.

διάρκεια αναζήτησης

πόσα αποτελέσματα βρέθηκαν

βλέπουμε τα 10 «σημαντικότερα»

1^ο αποτέλεσμα

2^ο αποτέλεσμα

3^ο αποτέλεσμα

Αναζήτηση κειμένου

8

Όταν δίνουμε περισσότερες λέξεις, χωρισμένες με κενά, η μηχανή καταλαβαίνει ότι αναζητούμε ιστοσελίδες που περιέχουν *καθεμιά* από αυτές τις λέξεις.



Ανόρθωση ποδοσφαιριστές

Google Search

I'm Feeling Lucky

[Advanced Search](#)

[Preferences](#)

[Language Tools](#)

Στο παράδειγμα, ζητάμε ιστοσελίδες που να περιέχουν **και** τη λέξη *Ανόρθωση* **και** τη λέξη *ποδοσφαιριστές*.

Αναζήτηση κειμένου

9

Web Images Maps News Video Gmail more Sign in

Google™ Ανόρθωση ποδοσφαιριστές Search Advanced Search Preferences

Web Results 1 - 10 of about 37,000 for Ανόρθωση ποδοσφαιριστές. (0.07 seconds)

Κατηγορία:Ποδοσφαιριστές Ανόρθωσης - Βικιπαίδεια - [Translate this page]
Σελίδες στην κατηγορία "Ποδοσφαιριστές Ανόρθωσης". Οι ακόλουθες 18 σελίδες είναι σε αυτή την κατηγορία, από 18 συνολικά. ...
el.wikipedia.org/wiki/Κατηγορία:Ποδοσφαιριστές_Ανόρθωσης - 27k - Cached - Similar pages

Sigma Live | ΑΝΟΡΘΩΣΗ: Στην Κύπρο μεθαύριο ο Λίνεν | Sigma Live - [Translate this page]
21 Απρ. 2009 ... Οι ποδοσφαιριστές της Ανόρθωσης επέστρεψαν από γηδες το απόγευμα σε κανονικούς ρυθμούς προπονήσεων, ενόψει του νέου ντέρμπι που τους ...
www.sigmalive.com/simerini/sport+time/cyprus/145179 - 25k - Cached - Similar pages

ΤΟ ΒΗΜΑ ΤΩΝ ΜΑΧΗΤΩΝ - Όλοι βούρκωσαν - [Translate this page]
Αν κάποτε σας φώναζα ήταν για το καλό σας» τους είπε ο Τιμούρ. Πρόσθεσε πως πέρασε υπέροχα στην Ανόρθωση και ευχαρίστησε τους ποδοσφαιριστές για την φετινή ...
www.anorthosis.net/bhma/index.php?mode=thread&id=52783 - 45k - Cached - Similar pages

anorthosis.net :: Γαλανόλευκη Αθλητική Ειδησεογραφία
Επίσημα ... - [Translate this page]
4 Ιαν. 2009 ... Κώστας Σταύρου Πρώην αρχηγός Ανόρθωσης-ποδοσφαιριστής Κώστας Χρυσοδόντας Επιχειρηματίας Λάκης Παπαδόπουλος Επιχειρηματίας ...
www.anorthosis.net/index.php?option=com_content&task=view&id=1120&Itemid=101 - 60k - Cached - Similar pages
[More results from www.anorthosis.net »](#)

Anorthosis F.C. - Σύνδεσμος Παλαιμάχων ποδοσφαιριστών Ανόρθωσης ... - [Translate this page]
06/03/2009 - Σύνδεσμος Παλαιμάχων ποδοσφαιριστών Ανόρθωσης Αμμοχώστου ... τους έφεραν στην μνήμη τους πολλά γεγονότα από τα παλιά χρόνια στην Ανόρθωση. ...

διάρκεια αναζήτησης

πόσα αποτελέσματα βρέθηκαν

βλέπουμε τα 10 «σημαντικότερα»

1^ο αποτέλεσμα

2^ο αποτέλεσμα

3^ο αποτέλεσμα

Αναζήτηση κειμένου

10

ποδοσφαιριστές της Ανόρθωσης
Google Search I'm Feeling Lucky

ποδοσφαιριστές +της Ανόρθωσης
Google Search I'm Feeling Lucky

"ποδοσφαιριστές της Ανόρθωσης"
Google Search I'm Feeling Lucky

ποδοσφαιριστές -της Ανόρθωσης
Google Search I'm Feeling Lucky

Η μηχανή δεν λαμβάνει υπόψη τις πολύ συνήθεις λέξεις (π.χ. «της»), ούτε τη σειρά των λέξεων.

Αν θέλουμε μια λέξη να ληφθεί οπωσδήποτε υπόψη, προθέτουμε το + .

Αν θέλουμε οι λέξεις να εμφανίζονται διαδοχικά με τη δοσμένη σειρά, τις βάζουμε σε εισαγωγικά.

Αν θέλουμε μια λέξη να μην εμφανίζεται, προθέτουμε το - .

Αναζήτηση κειμένου

11

Για περισσότερες δυνατότητες, μπορούμε να κατευθυνθούμε στη σύνθετη αναζήτηση (*advanced search*).



Use the form below and your advanced search will appear here

Find web pages that have...

all these words:

this exact wording or phrase: [tip](#)

one or more of these words: OR OR [tip](#)

But don't show pages that have...

any of these unwanted words: [tip](#)

Need more tools?

Results per page: ▼

Language: ▼

File type: ▼

Search within a site or domain:

(e.g. youtube.com, .edu)

[+ Date, usage rights, numeric range, and more](#)

Αναζήτηση εικόνων

12

Για να αναζητήσουμε εικόνες, κατευθυνόμαστε στην αντίστοιχη επιλογή.

[Web](#) [Images](#) [Maps](#) [News](#) [Video](#) [Gmail](#) [more](#) ▼

[Sign in](#)



Ανόρθωση ποδοσφαιριστές

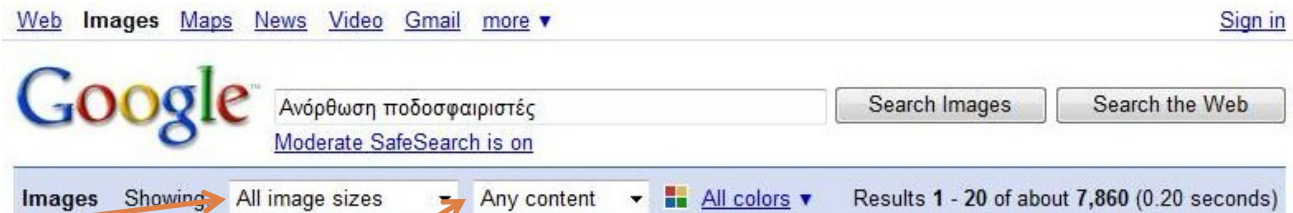
Search Images

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

Αναζήτηση εικόνων

13

Τα αποτελέσματα της αναζήτησης είναι μόνον εικόνες:



επιλογή μεγέθους
(πολύ μεγάλες,
μεγάλες, μικρές, κλπ.)



... της Ανόρθωσης οι ποδοσφαιριστές ...
575 x 805 - 144k - jpg
athkeiaseros.pblogs.gr



Ομόνοια-Ανόρθωση: Για Νέγκρι και ...
447 x 480 - 31k - jpg
www.sigmalive.com
[[More from www.sigmalive.com](#)]



Ανόρθωση: Με 7 απόντες η προπόνηση .
644 x 480 - 34k - jpg
www.sigmalive.com



Δέλλας: "Θέλω τίτλους με Ανόρθωση"
150 x 100 - 12k - jpg
www.sport24.gr

επιλογή είδους
(πρόσωπα, σκίτσα,
φωτογραφίες, κ.ά)



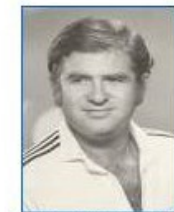
... ποδοσφαιρικό τμήμα της Ανόρθωσης ...
420 x 237 - 16k - jpg
www.sport-fm.gr



... της Ανόρθωσης αναδημοσιεύουμε ...
410 x 230 - 22k - jpg
www.shootandgoal.com
[[More from www.shootandgoal.com](#)]



Πήρε Τικίνο η Ανόρθωση
227 x 314 - 95k - jpg
www.sport-fm.com.cy
[[More from www.sport-fm.com.cy](#)]



1973 - 1980 (ΑΣΙΛ): Ποδοσφαιριστής ...
300 x 400 - 42k - jpg
asil-fc.com

Άλλες ειδικές αναζητήσεις

14

Παρομοίως, μπορούμε να εκτελέσουμε άλλες ειδικές αναζητήσεις: χάρτες, ειδήσεις, βίντεο, κ.ά.

[Web](#) [Images](#) [Maps](#) [News](#) [Video](#) [Gmail](#) [more ▼](#) [Sign in](#)

Google™
Image Search

Ανόρθωση ποδοσφαιριστές

Search Images

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

Δοκιμάστε τα!

Αρχιτεκτονική

Η εργασία μιας μηχανής αναζήτησης οργανώνεται σε τρία στάδια:

1. **ανίχνευση** (*crawling*): Ένα ειδικό πρόγραμμα (*crawler*) επισκέπτεται διαδοχικά δισεκατομμύρια ιστοσελίδες, ακολουθώντας απλώς κάθε συνδέσμο που συναντά.
2. **ευρετηρίαση** (*indexing*): Τα περιεχόμενα των διαβασμένων ιστοσελίδων αναλύονται, ώστε να δημιουργηθεί ένα **ευρετήριο** (*index*) που διευκολύνει την συσχέτιση λέξεων με ιστοσελίδες.
3. **αναζήτηση** (*searching*): Με λέξεις-κλειδιά και άλλα στοιχεία, ο χρήστης σχηματίζει ένα **ερώτημα** (*query*) και το στέλνει στη μηχανή. Η μηχανή ψάχνει στο ευρετήριο τις ιστοσελίδες που ταιριάζουν περισσότερο με το ερώτημα, και επιστρέφει έναν κατάλογο αυτών.

Αρχιτεκτονική

Η χρησιμότητα μίας μηχανής αναζήτησης καθορίζεται από το πόσο σχετικά προς το ερώτημα του χρήστη είναι τα αποτελέσματα που επιστρέφει.

Κατά κανόνα, μια λέξη ή φράση που αναζητείται από τον χρήστη μπορεί να βρίσκεται σε χιλιάδες ή και εκατομμύρια σελίδες. Πριν επιστρέψει αυτές τις σελίδες ως αποτελέσματα, η μηχανή πρέπει να τις κατατάξει βάσει κάποιας βαθμολογίας που να υποδεικνύει πόσο συναφής είναι καθεμιά από αυτές με το ερώτημα του χρήστη.

Προκλήσεις

- Ο Ιστός μεγαλώνει ταχύτερα από ό,τι μπορεί να τον κατατάξει οποιαδήποτε τεχνολογία αναζήτησης.
- Πολλές ιστοσελίδες αλλάζουν διαρκώς, με αποτέλεσμα οι μηχανές αναζήτησης να πρέπει να τις επισκέπτονται περιοδικά.
- Οι περισσότερες ερωτήσεις αφορούν σε λέξεις-κλειδιά και πολλά από τα αποτελέσματα δεν είναι σχετικά με αυτό που πραγματικά ψάχνει ο χρήστης (*false positives*).
- Πολλές ιστοσελίδες δημιουργούνται δυναμικά (δηλ. από λογισμικό---π.χ. ως απάντηση σε κάποια αναζήτηση). Αυτές είναι δύσκολο να καταγραφούν από crawler. Πολλές δεν καταγράφονται ποτέ. Το αποτέλεσμα λέγεται **αόρατος ιστός** (*invisible web*).

Προκλήσεις

- Κάποιες μηχανές δεν ταξινομούν τα αποτελέσματα βάσει της σχετικότητας με τις ερωτήσεις του χρήστη, αλλά ανάλογα με το πόσα χρήματα έχει πληρώσει ο κάθε ιστοτόπος.
- Κάποιοι ιστοτόποι μετέρχονται τεχνασμάτων ώστε η μηχανή αναζήτησης να τους εμφανίζει ως έχοντες υψηλή συσχέτιση με κάποια ερωτήματα, ενώ στην πραγματικότητα δεν έχουν.