



Πανεπιστήμιο Αιγαίου

Βασικές Δεξιότητες Πληροφορικής και Επικοινωνίας

Αναζήτηση στον
Παγκόσμιο Ιστό

Μιχάλης Βαΐτης
Τμήμα Γεωγραφίας



Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.

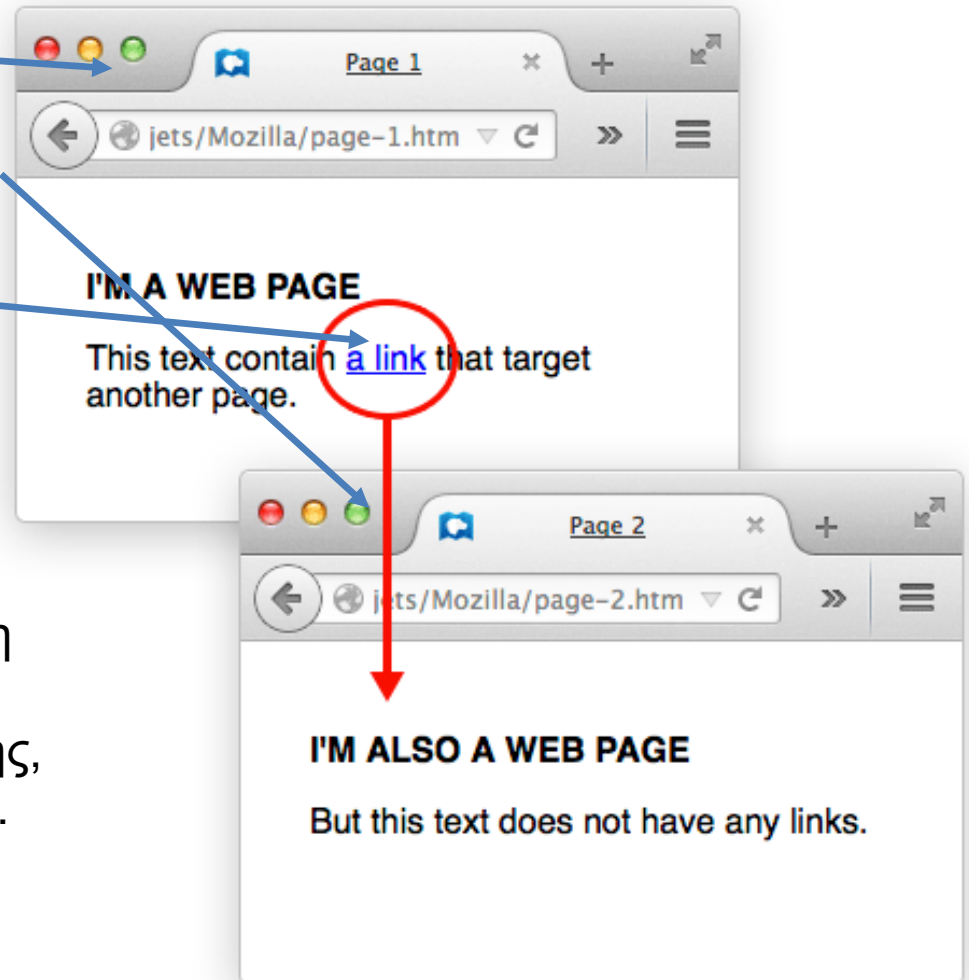


Ο Παγκόσμιος Ιστός (world wide web)

Τα συστατικά

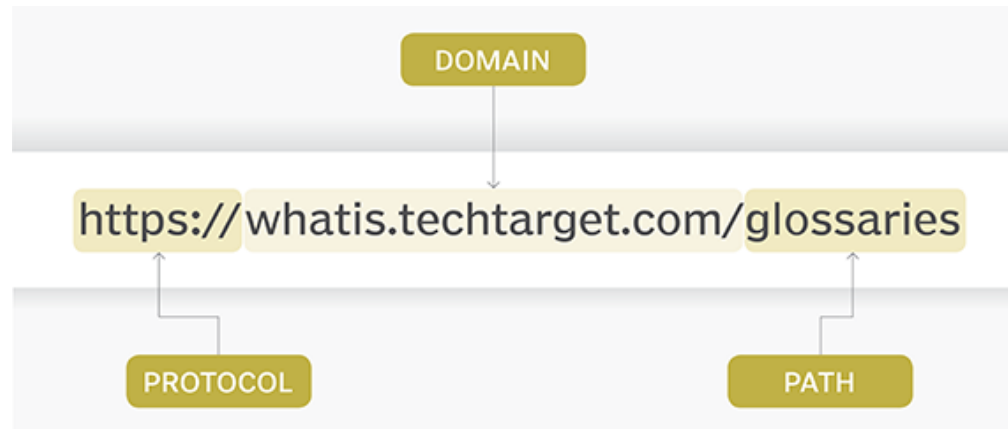
- Ιστοσελίδα (Webpage)
- Υπερσύνδεσμος/Σύνδεσμος
- Άγκυρα (Anchor)

Ο σύνδεσμος περιέχει τη διεύθυνση της σελίδας προορισμού, χωρίς να την εμφανίζει στο χρήστη. Αντ' αυτής, εμφανίζεται το κείμενο της άγκυρας.



URL (Uniform Resource Locator)

ΒΑΣΙΚΗ ΔΟΜΗ



ΠΛΗΡΗΣ ΔΟΜΗ

`protocol://username:password@domain:port/
path?query#fragment_id`

Για το http πρωτόκολλο, το τελευταίο τμήμα (/path...) προσδιορίζει την ιστοσελίδα που παρέχεται από τον web server που προσδιορίζεται από το domain και το port.

Hyperlink

HREF → **H**ypertext **R**eference

`<a href="http://www.tinydancinghorse.com">Tiny dancing horse</a>`

Target Link Anchor Text

Αναζήτηση στον Παγκόσμιο Ιστό

Αναζήτηση στο web

- Θεματικοί κατάλογοι / Πύλες (portals)
- Μηχανές αναζήτησης (search engines)
- Μηχανές μετα-αναζήτησης (metasearch engines)
- Εξειδικευμένες μηχανές αναζήτησης

Θεματικοί κατάλογοι / Πύλες

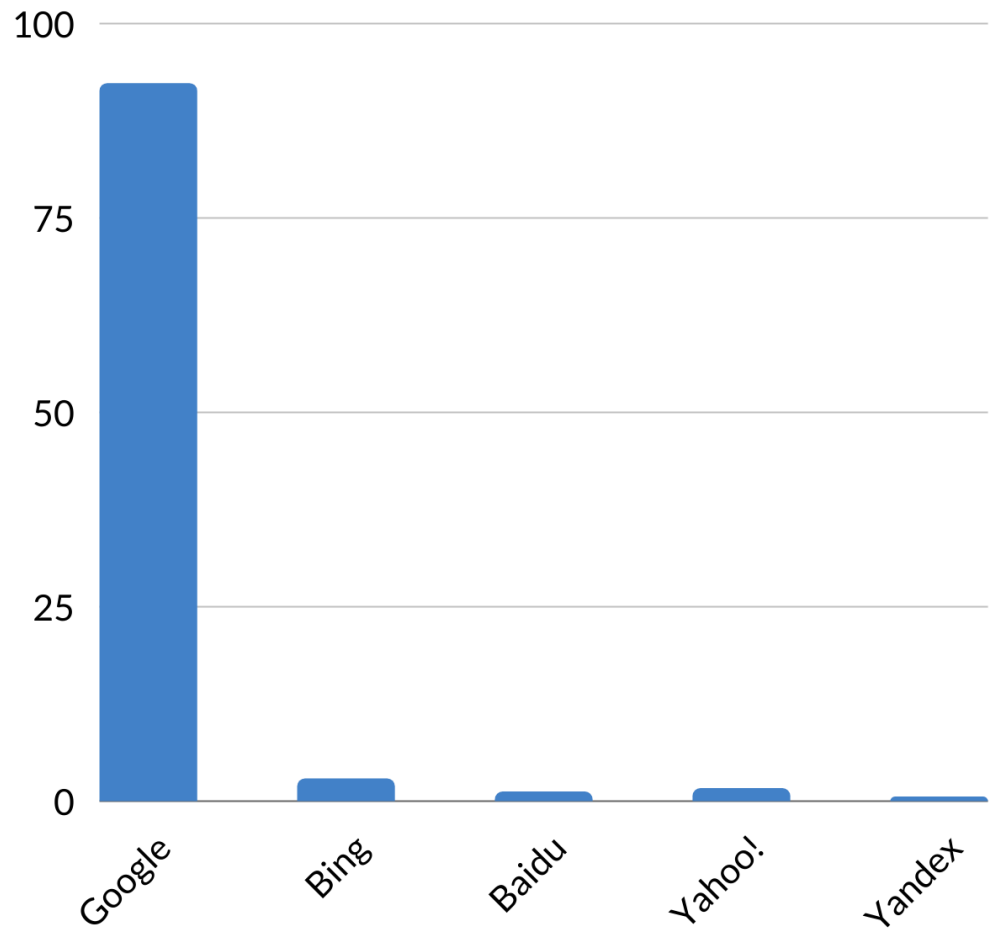
Πύλη (Portal) στον web

- Λειτουργεί ως «είσοδος» (όπως δηλώνει και το όνομά της)
- Συγκεντρώνει περιεχόμενο/πληροφορίες από πολλές πηγές
- Δίνει πρόσβαση σε άλλες ιστοσελίδες
- Απευθύνεται στο γενικό κοινό ή σε χρήστες με ειδικά ενδιαφέροντα

Μηχανές αναζήτησης (search engines)

Μηχανές αναζήτησης

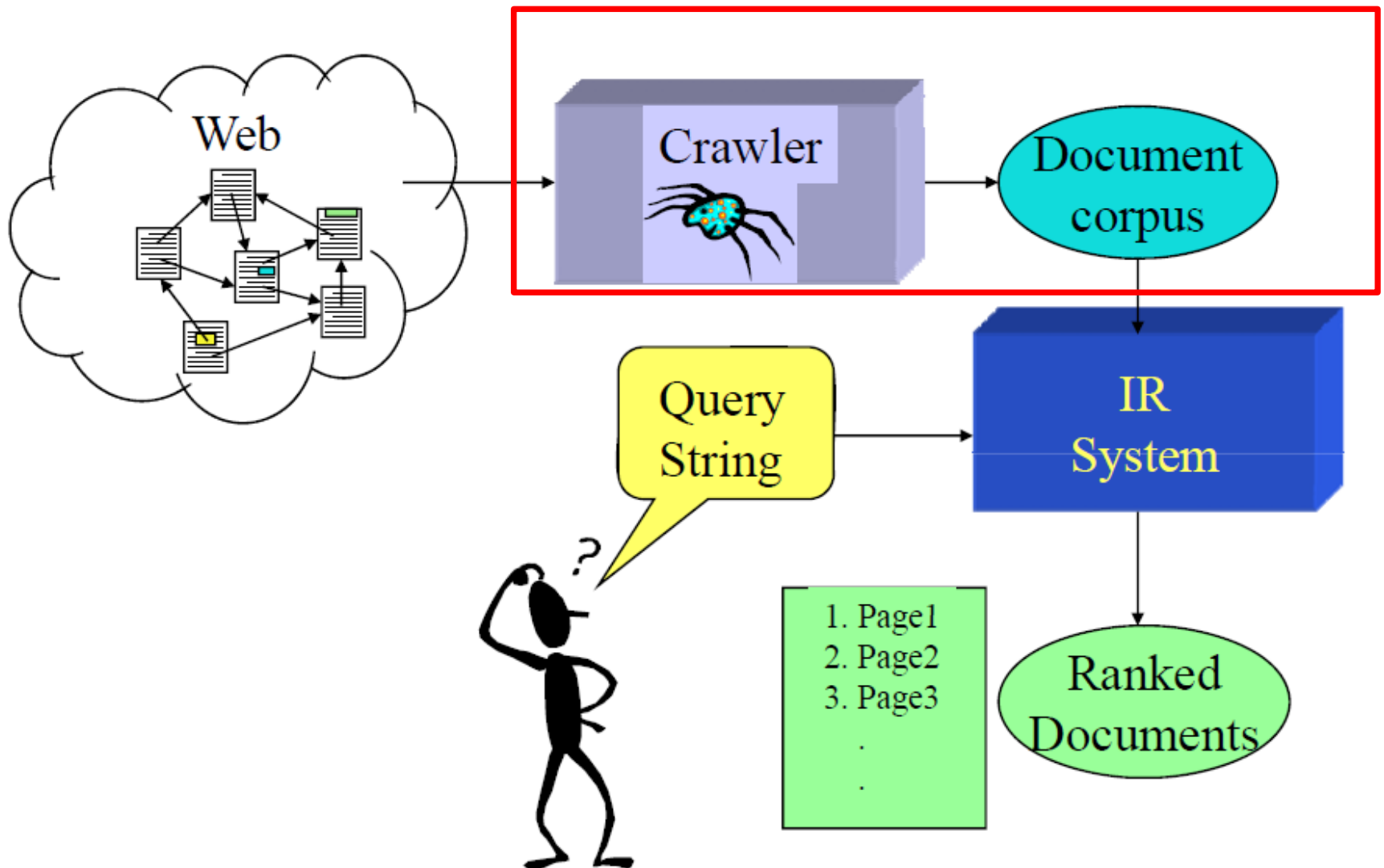
- Google
- Bing
- Yahoo!
- Baidu (Κίνα)
- Sogou (Κίνα)
- Yandex (Ρωσία)
- ...



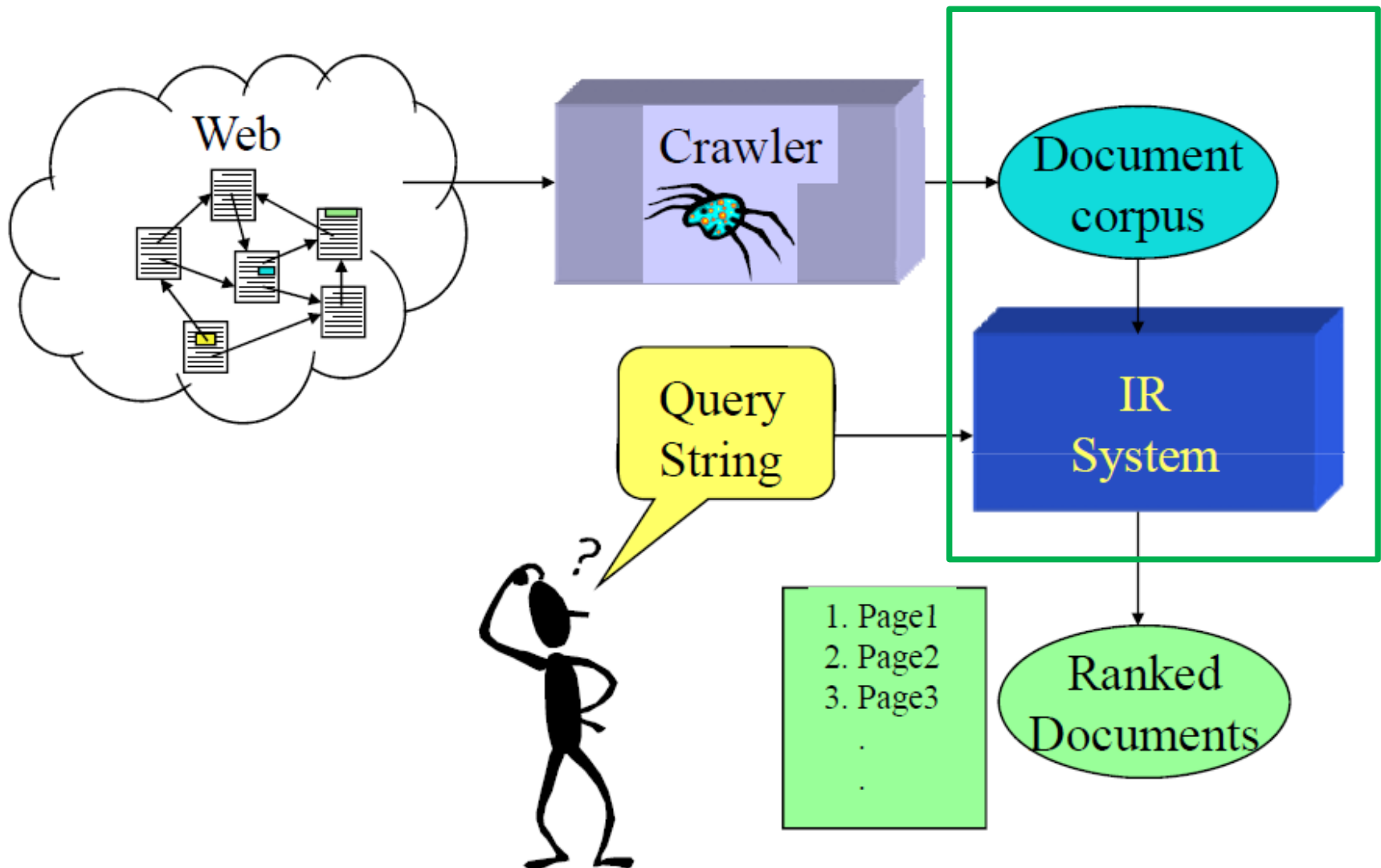
Στάδια για την παροχή υπηρεσιών αναζήτησης στον Παγκόσμιο Ιστό

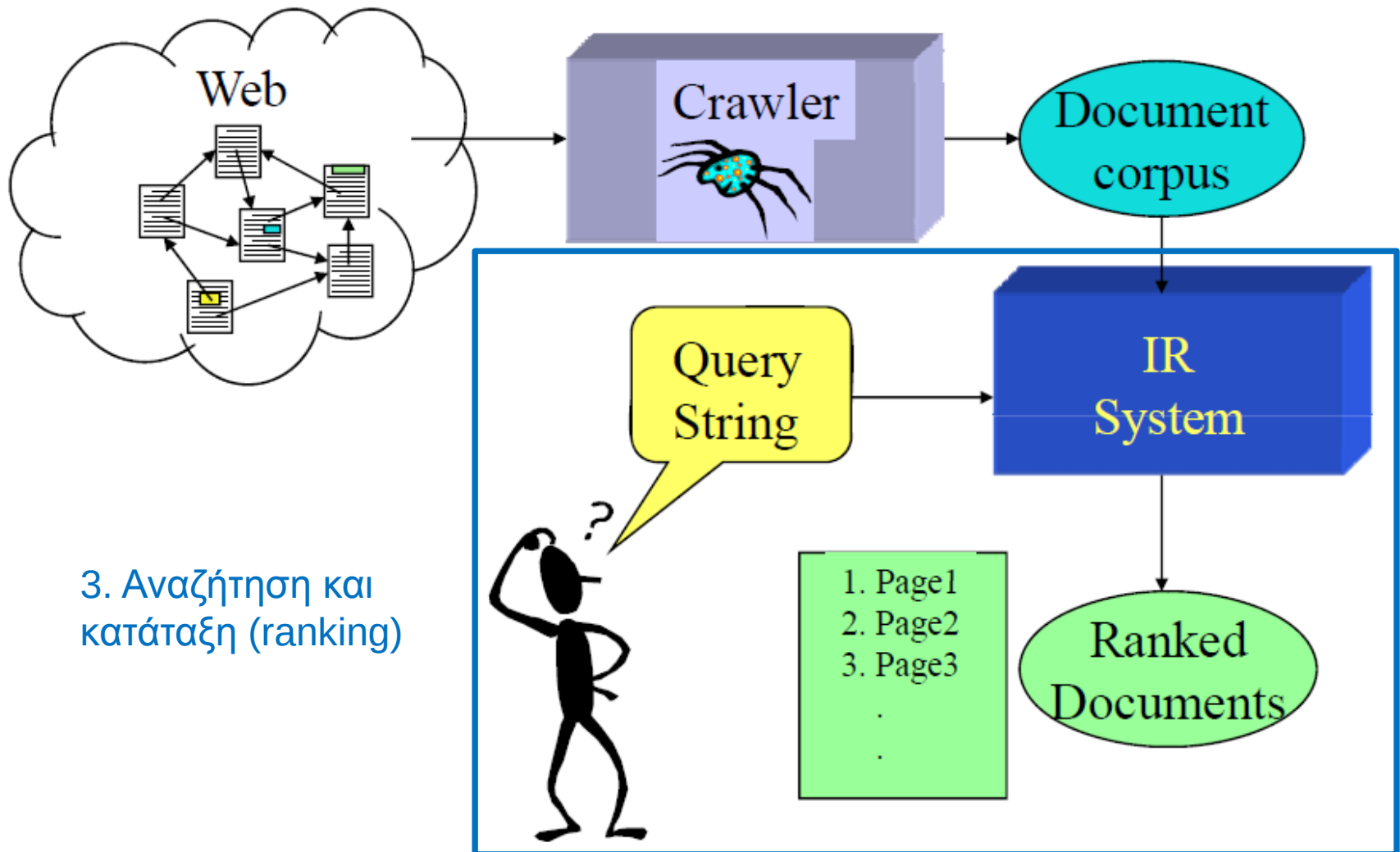
1. Ανίχνευση ιστού (web crawling)
2. Ευρετηρίαση (indexing)
3. Αναζήτηση και κατάταξη (ranking)

1. Ανίχνευση ιστού (web crawling)



2. Ευρετηρίαση (indexing)





Ευρετηρίαση (indexing)

(Ιστοσελίδες)		Indexing Items (Όροι)					
		k_1	k_2	...	k_j	...	k_t
D o c u m e n t s	d_1	$c_{1,1}$	$c_{2,1}$	...	$c_{i,1}$	...	$c_{t,1}$
	d_2	$c_{1,2}$	$c_{2,2}$	...	$c_{i,2}$	...	$c_{t,2}$
	...	...	...	...	...	...	...
	d_i	$c_{1,j}$	$c_{2,j}$	...	$c_{i,j}$	...	$c_{t,j}$
	...	...	...	...	...	...	...
	d_N	$c_{1,N}$	$c_{2,N}$	...	$c_{i,N}$	...	$c_{t,N}$

Βαρύτητα του όρου (term/item)
στο έγγραφο (document/ιστοσελίδα)

Μηχανές αναζήτησης

ΙΣΤΟΡΙΑ

[JumpStation](#) (Δεκέμβριος 1993)

Ανίχνευση και ευρετηρίαση μόνο των τίτλων των ιστοσελίδων

[WebCrawler](#) (1994)

Ανίχνευση και ευρετηρίαση όλου του περιεχομένου των ιστοσελίδων

Αρχές λειτουργίας μηχανών αναζήτησης

- Κάθε μηχανή αναζήτησης χρησιμοποιεί το δικό της λογισμικό
 - Δίνει συνεπώς διαφορετικά αποτελέσματα για το ίδιο κριτήριο (π.χ. λέξη-κλειδί, γλώσσα, κλπ.)
- **Χρήση κριτηρίων αναζήτησης**
 - Λέξη-κλειδί, γλώσσα, είδος αρχείου, ημερομηνία
- Δυνατότητα συνδυασμού κριτηρίων
- Χρήση λογικών τελεστών
- Παρουσίαση των αποτελεσμάτων με ιεραρχικό τρόπο (σειρά κατάταξης)

Κριτήρια αναζήτησης

- Βασικό κριτήριο: **λέξη-κλειδί**
- Τα κριτήρια αναζήτησης με **λέξεις-κλειδιά** μπορεί να συνδυαστούν με λογικούς τελεστές ώστε η αναζήτηση να είναι περισσότερο αποτελεσματική.
- Συμπληρωματικά κριτήρια: γλώσσα, είδος αρχείου, ημερομηνία, περιοχή, κλπ.
 - Η σωστή χρήση κριτηρίων είναι απαραίτητη, καθώς επιτρέπει να περιοριστεί το πλήθος των ιστοσελίδων που εντοπίζει η μηχανή χωρίς να χάνονται κάποιες ιστοσελίδες που μας ενδιαφέρουν.

Συνδυασμός κριτηρίων αναζήτησης

- Το βασικό κριτήριο (λέξη-κλειδί) μπορεί να συμπληρωθεί με επιπλέον κριτήρια
- Συμπληρωματικά κριτήρια:
 - Γλώσσα (π.χ. ελληνικά, αγγλικά, κλπ.)
 - Είδος αρχείου (π.χ. ιστοσελίδα, pdf, ppt)
 - Ημερομηνία (πότε εμφανίστηκε η σελίδα)
 - Περιοχή (σε ποια χώρα είναι ο υπολογιστής που περιέχει την αναζητούμενη πληροφορία)
- **Παρατήρηση:** συνήθως η αναζήτηση αφορά ιστοσελίδες, μπορεί όμως να αφορά εικόνες, ιστολόγια, ειδήσεις, κλπ.

Μηχανές μετα-αναζήτησης

- Πραγματοποιούν αναζητήσεις σε άλλες μηχανές αναζήτησης και συνθέτουν τα αποτελέσματα.

Π.χ. www.dogpile.com

Εξειδικευμένες μηχανές αναζήτησης

- Πραγματοποιούν αναζητήσεις συγκεκριμένα θέματα.

Π.χ. scholar.google.com για αναζήτηση επιστημονικών άρθρων

Σημεία προσοχής!

- Ποιος είναι ο συγγραφέας του κειμένου;
- Ποια είναι η ιστοσελίδα που φιλοξενεί το κείμενο; / σε τι κοινό απευθύνεται; / είναι λειτουργική; (οι σύνδεσμοι λειτουργούν)
- Ποια η ημερομηνία του κειμένου;

Θα πρέπει να είναι δυνατή η προσθήκη βιβλιογραφικής αναφοράς στην πηγή.

Τέλος Διάλεξης



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

