

13/10/20, Week 2, Tuesday

Ανάκτηση Πληροφορίας (Φροντιστήριο)

ΔΕ χρειάζεται αποστολή e-mail για υλοποιητική εργασία.
 Βοηθάει το μάθημα της εξώρυξης δεδομένων
 Όσοι θέλουν θεωρητικό ραφερ, στέλνουν στον Μακρή email
 με τα 3 θέματα που προτιμούν.

Εισαγωγή

email: robolas@ceid.upatras.gr

Ασχολούμαστε με:

- ↳ Αναπαράσταση
 - ↳ Αποθήκευση
 - ↳ Οργάνωση
 - ↳ Πρόσβαση
- } Πληροφοριακών συστημάτων
(έγγραφα, ιστοσελίδες etc)

Βιβλιοθήκες

- Τεράστιος όγκος πληροφοριών
 - ↳ Πως μπορεί να μειωθεί ο χρόνος αναζήτησής τους
 - ↳ Ανάγκη κατασκευής εξειδικευμένων δομών δεδομένων
- ↳ Indexes
 - ↳ Γρήγορη πρόσβαση σε δεδομένα
 - ↳ Επτάχυνση επεξεργασίας ερωτημάτων

Ψηφιακές Βιβλιοθήκες

Gen 1 ⇒ Συγγραφέας / Τίτλος

Gen 2 ⇒ Επικεφαλίδες, λέξεις-κλειδιά, τελευταίες ερωτηράειων

Gen 3 ⇒ GUI, Ηλεκτρονικές φόρμες, hyperlinks

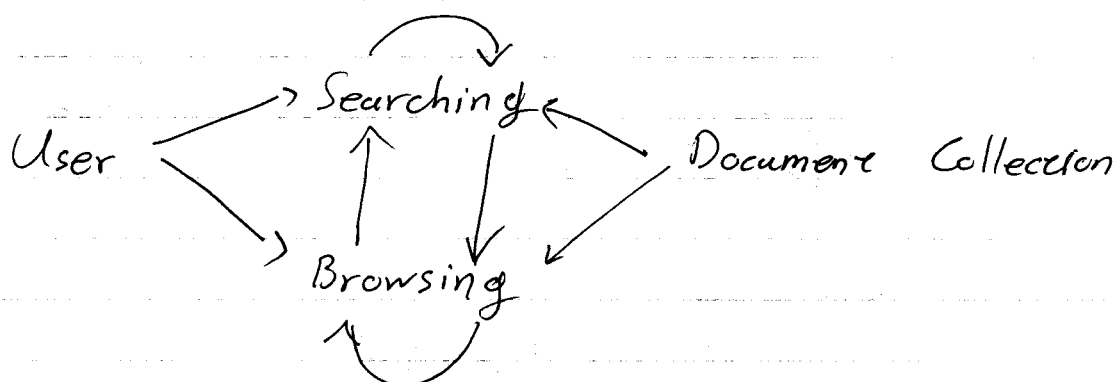
World Wide Web

- ↳ Η μεγαλύτερη πηγή γνώσης στην ιστορία
- ↳ Όχι απλή υπόθεση η εύρεση της σωστής πληροφορίας
- ↳ Έρωτήματα διαφορετικής πολυπλοκότητας
- ↳ Χαοτικό περιβάλλον

Το πρόβλημα της Ανάκτησης πληροφορίας

- ↳ Να ανακτήσουμε τα ζητούμενα έγγραφα, ανακτώντας όσο το δυνατό λιγότερα, μη σχετικά έγγραφα.
- ↳ Κανένα σύστημα δε μπορεί να παρέχει τέλειες απαντήσεις για όλους τους χρήστες

Βιβλίο = { Ricardo Baeza-Yates
 { Bréhier Ribeiro-Neto



Ανάκτηση Δεδομένων

- ↳ Προσδιορισμός εγγράφων μιας συλλογής με λέξεις-κλειδιά
- ↳ Έλα λανθασμένο αντικείμενο = πλήρης αποτυχία
- ↳ Καλά καθορισμένη Δομή-Σηματολογία

Ανάκτηση Πληροφορίας

- ↳ Τα ανακτηθέντα έγγραφα μπορεί να μη περιέχουν όρους του ερωτήματος
- ↳ Τα ανακτηθέντα αντικείμενα μπορεί να είναι ανακριβή
- ↳ Μη καλά δομημένα - φυσική γλώσσα