



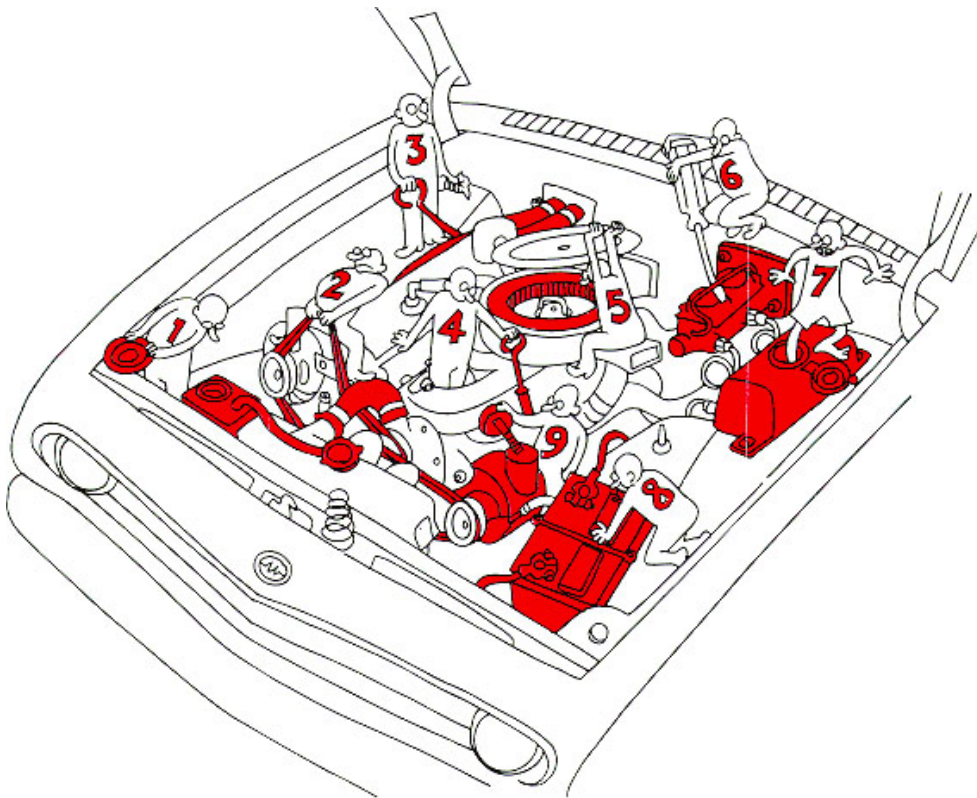
# Τεχνολογίες ψηφιακών βιβλιοθηκών Πως δουλεύουν οι Ψηφιακές βιβλιοθήκες?

Παπαδάκης Γιάννης  
2008-9

Διασκευή παρουσίασης Joe Futrelle  
National Center for Supercomputing Applications – NCSA



# Πώς δουλεύουν οι Ψ.Β?



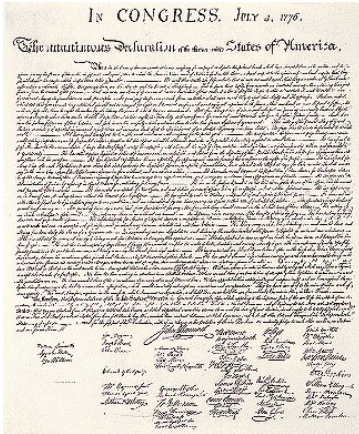
# Πώς δουλεύουν οι Ψ.Β?

---

- Συλλογές από ψηφιακά αντικείμενα, τα οποία είναι αποθηκευμένα σε συστήματα αρχείων και βάσεις δεδομένων
- Υπηρεσίες παρέχονται χρησιμοποιώντας πολλών ειδών τεχνολογίες
  - Ανάκτηση πληροφοριών: ευρετηρίαση/indexing, αναζήτηση, κατηγοριοποίηση/classification
  - μεταδεδομένα: σχολιασμός, περιγραφή
  - Περιβάλλον διεπαφής/User Interface: GUI εργαλεία, web-based interfaces, εφαρμογές οπτικοποίησης, ανάδραση χρηστών/user feedback



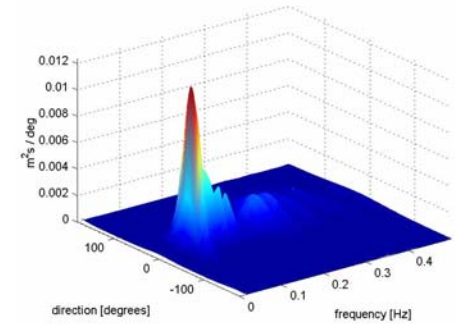
# Τί είδους ψηφιακών αντικειμένων μπορούν να υπάρχουν σε μια Ψ.Β?



έγγραφο



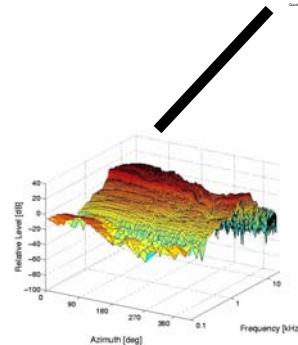
media  
Π.χ., εικόνες



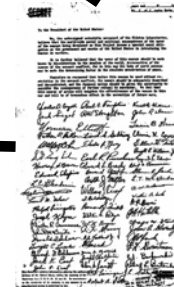
datasets



χάρτες



πολύπλοκα αντικείμενα



# Πώς αναπαρίσταται η πληροφορία?

---

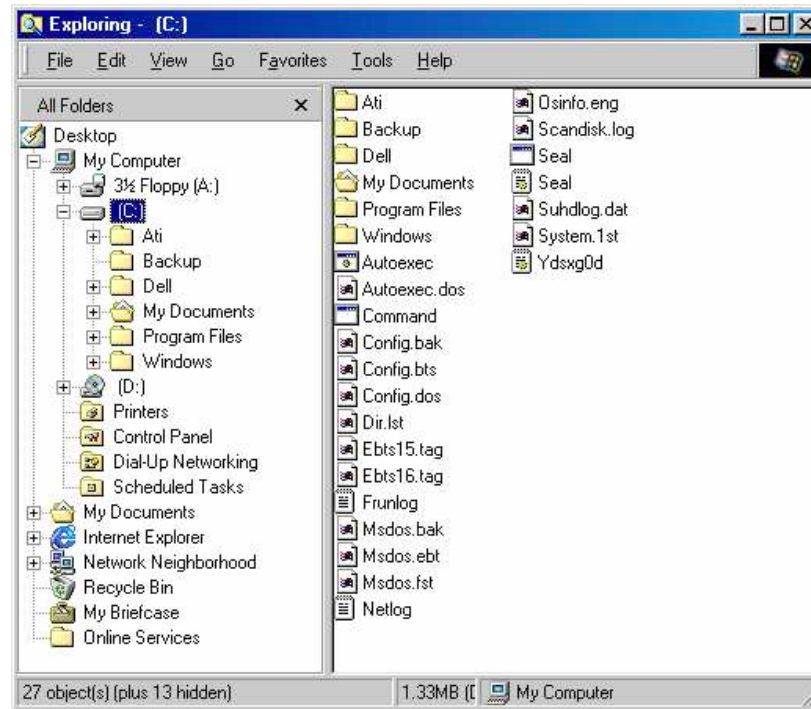
- Με πολλούς τρόπους
- File formats (π.χ., HTML, JPEG, XML, HDF)
- Μοντέλα αναπαράστασης
  - Δομημένα
    - Αριθμητικά δεδομένα (π.χ., πίνακες, grids)
    - Ιεραρχίες
  - Ημι-Δομημένα
    - Markup languages – Γλώσσες σημείωσης
  - Αδόμητα
    - Απλό κείμενο



# Πώς αποθηκεύονται τα ψηφιακά αντικείμενα?

- Σε συστήματα αρχείων και βάσεις δεδομένων

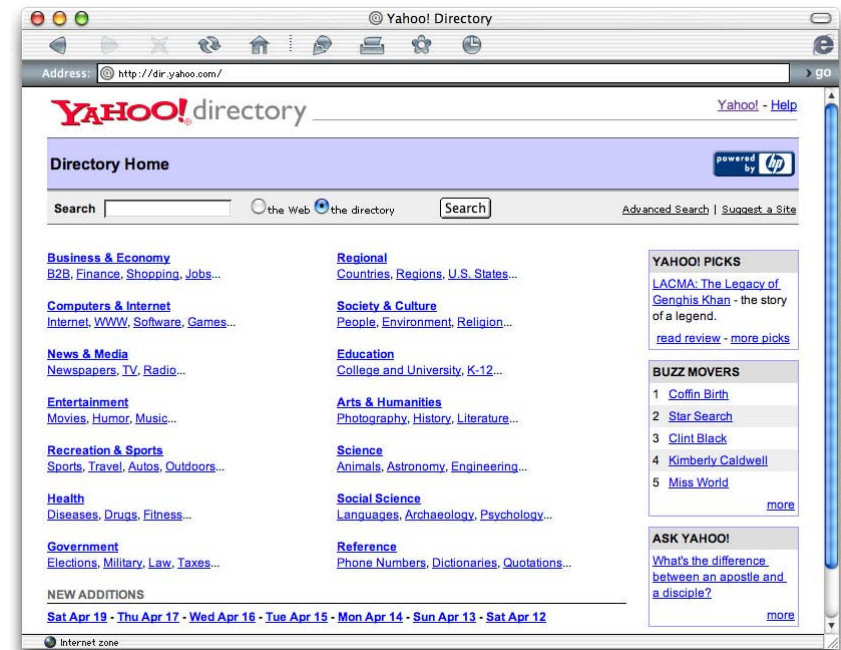
Θυμίζει τίποτα?



# Πώς βρίσκουν οι χρήστες ψηφιακά αντικείμενα?

- Μέθοδος 1: πλοήγηση-browsing
  - Τα αντικείμενα συνδέονται μεταξύ τους χρησιμοποιώντας κάποια δομή π.χ., hyperlinks, directories
  - Οι χρήστες χρησιμοποιούν κάποια διεπαφή για να πλοηγήσουν τη δομή

Π.χ.,



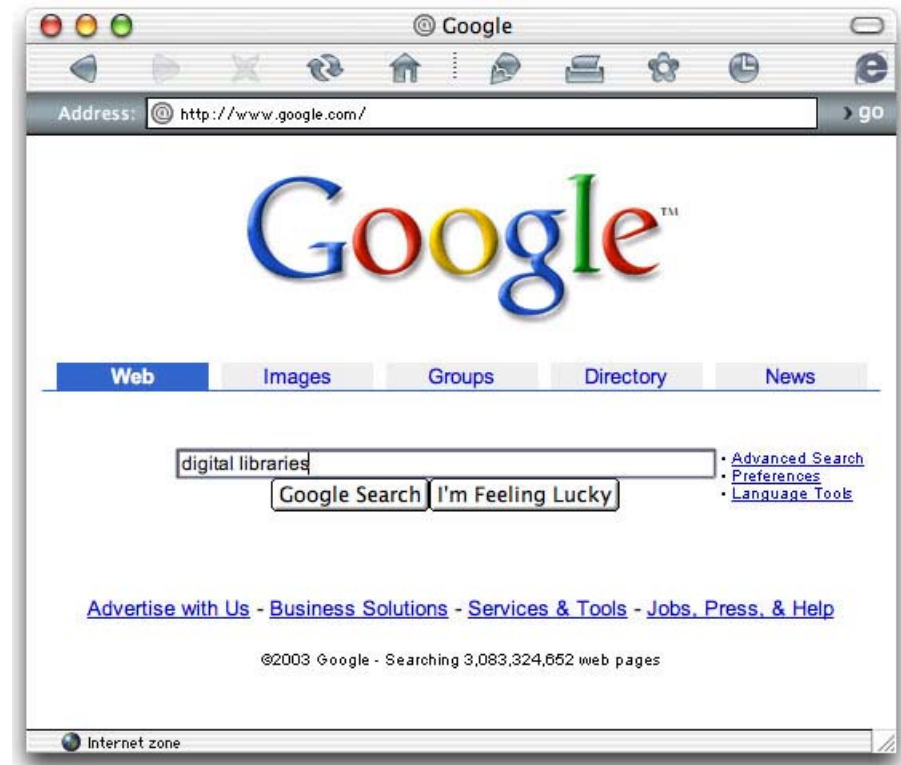


# Πώς αλλιώς βρίσκουν οι χρήστες ψηφιακά αντικείμενα?

- **Method 2: Αναζήτηση**

- Χρήστες εκφράζουν τις πληροφοριακές τους ανάγκες σε ερωτήματα
- Οι ψ.β. βρίσκουν αντικείμενα που ικανοποιούν τα ερωτήματα

Π.χ.,





# Πώς δουλεύει η πλοήγηση?

- Τα αντικείμενα οργανώνονται προς πρόσβαση
  - **Linking**: αναφορές αντικειμένων σε αντικείμενα (βιβλιογραφία)
  - **Classification**: Ομαδοποίηση αντικειμένων σε θεματικές περιοχές
  - **Recommendation**: οργάνωση σύμφωνα με τις προτιμήσεις των χρηστών
- Οι χρήστες διασχίζουν την οργάνωση



# Τεχνολογίες για browsing

---

- **To web**

- Η HTML υποστηρίζει hyperlinked έγγραφα με ενθυλακωμένα πολυμέσα (εικόνες, animations, movies, ήχους, applets)
- Το HTTP υποστηρίζει κατανεμημένη πρόσβαση σε έγγραφα υπερμέσων
- Τα Websites μπορούν να σχεδιαστούν έτσι ώστε να παρέχουν θεματικά ευρετήρια indexes (π.χ., Yahoo directory)
- Faceted categorization (Hearst, 2002)
- Τα αποτελέσματα μηχανών αναζήτησης συνδέουν σχετικά έγγραφα



# Σχετικά με κατηγοριοποίηση/classification

---

- Σε παραδοσιακές βιβλιοθήκες, **classification** γίνεται χειρονακτικά από ειδικούς σύμφωνα με αυστηρά καθορισμένα συστήματα (π.χ., **Dewey Decimal**)
- Φανταστείτε να προσπαθήσει κανείς κάτι ανάλογο για τις  $> 10^{10}$  σελίδες του Web



× **100**

**Library of Congress**



# Πώς δουλεύει η αναζήτηση?

- Η εξέταση κάθε αντικειμένου στη συλλογή δεν είναι αποδοτική, οπότε κατασκευάζεται ένα ευρετήριο/index το οποίο και ψάχνεται



# Πώς δουλεύει η ευρετηρίαση για ψηφιακά αντικείμενα?

- **Ανάλογα με το είδος των αντικειμένων**

<b>Κείμενο</b>	<b>Inverted files</b>
<b>Αριθμητικά</b>	<b>Βάσεις δεδομένων</b>
<b>Πολυμέσα</b>	<b>Τεχνικές βασισμένες στο περιεχόμενο</b>
<b>Αλλα (π.χ., επιστημονικά)</b>	<b>Εξειδικευμένες τεχνικές</b>

# Τα πάντα ευρετηριάζονται?

- Συνήθως όχι, συχνά μόνο μεταδεδομένα ευρετηριάζονται
- Τα μεταδεδομένα περιγράφουν τα δεδομένα

δεδομένο



Δεν ευρετηριάζεται

μεταδεδομένα

“Zoey ο σκύλος”  
Οκτώβριος 2000

ευρετηριάζεται



# Τα μεταδεδομένα είναι σημαντικά!

---

- Οντως, η ανάπτυξη υψηλής ποιότητας μεταδεδομένων αποτελεί βασική απαίτηση των ψηφιακών βιβλιοθηκών



Χρειαζόμαστε υψηλής  
ποιότητας μεταδεδομένα !



# Περισσότερα για μεταδεδομένα

---

- Τα μεταδεδομένα είναι δομημένα
  - Αυξάνουν με την πολυπλοκότητα των δεδομένων και όχι με την ποσότητά τους
- Τα μεταδεδομένα συνδέονται με τα δεδομένα
  - Είναι ενθυλακωμένα στα δεδομένα, ή
  - Κάνουν (κάπως, π.χ. URI) αναφορές στα δεδομένα
- Τα μεταδεδομένα πρέπει να είναι μεταφέρσιμα
- *Η σύνταξη και η σημασία των μεταδεδομένων πρέπει να καθορίζονται από τις κοινότητες που θα τα χρησιμοποιήσουν*



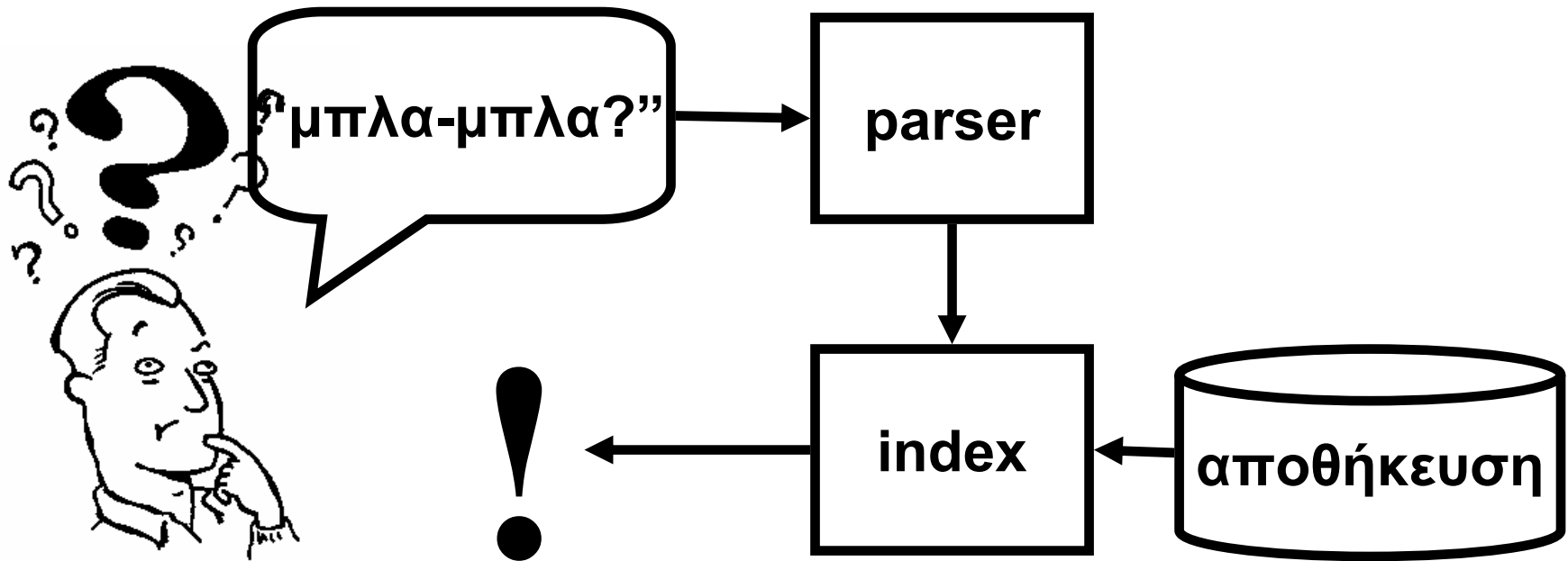
# Τί είδους search queries υπάρχουν?

- **Απλό κείμενο/Free text**
  - Π.χ., “Papadakis, digital library”
- **Boolean**
  - e.g., “AUTHOR = papadakis AND TITLE contains “digital library”
- **Example data object**
  - Π.χ., “σαν αυτό



# Πώς επεξεργάζονται τα ερωτήματα?

- Βασικά, μεταγλωττίζονται στο API του συστήματος ευρετηρίασης



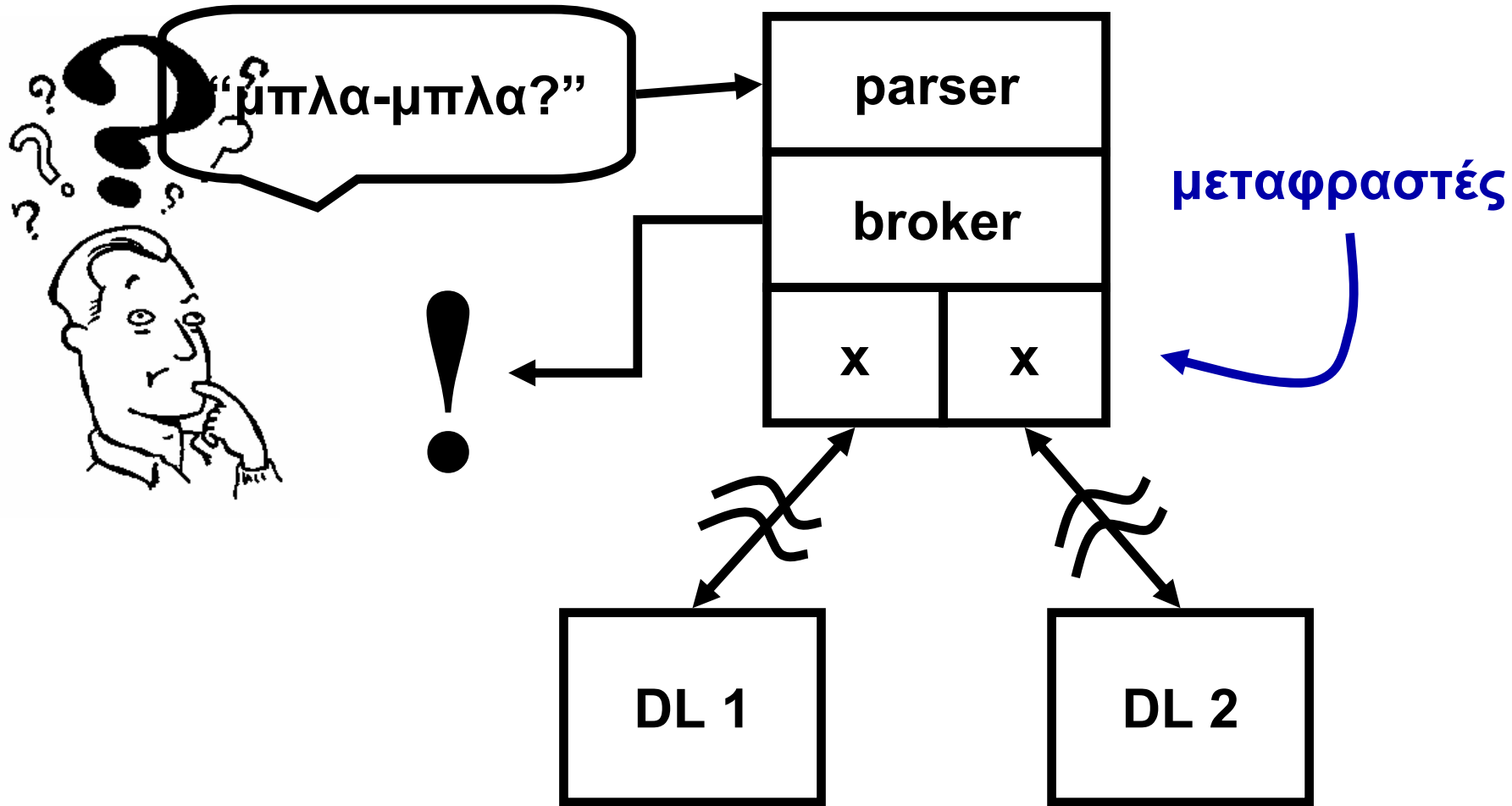
# Τί γίνεται όταν χρειαζόμαστε υλικό από 2 Ψ.Β?

---

- **Είναι πρόβλημα αν έχουν διαφορετικά**
  - Γλώσσες έκφρασης ερωτημάτων
  - Interfaces
  - Πρωτόκολλα πρόσβασης
- **Στρατηγική 1: ομοσπονδία/federation**
  - Επιλογή A: προτυποποίηση των interfaces των Ψ.Β.: OAI-PMH
  - Επιλογή B: διαμεσολάβηση broker για μετάφραση
- **Στρατηγική 2: harvesting**
  - Συλλογή δεδομένων από κατανεμημένες συλλογές
  - Παροχή δυνατότητας ανάκτησης σε κάποιο κεντρικό σημείο



# Κατανεμημένες Ψ.Β: ομοσπονδία



# ομοσπονδία: υπέρ και κατά

---

- **Πλεονεκτήματα**

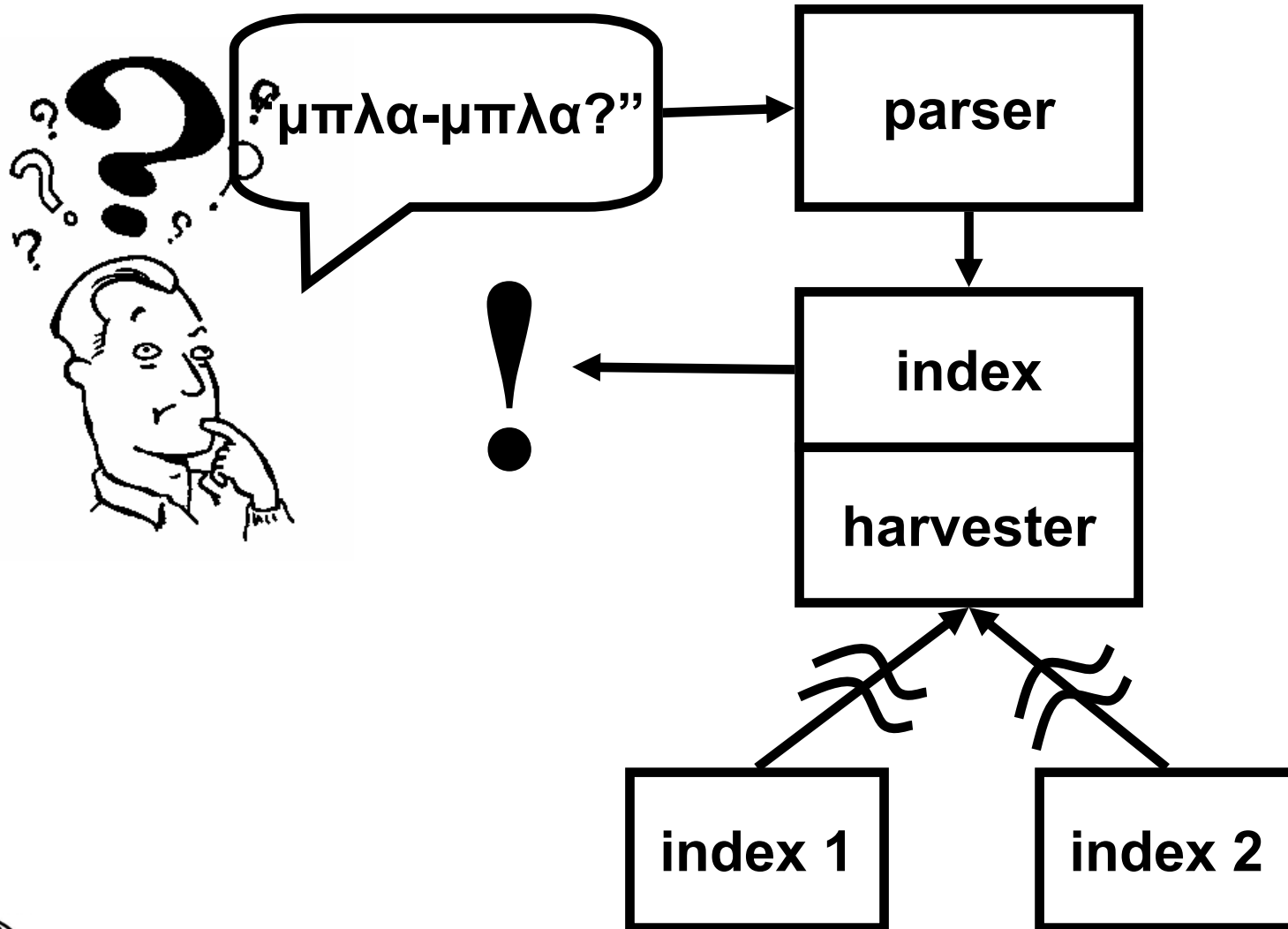
- Οι συλλογές ελέγχουν την ευρετηρίαση: μπορούν να βελτιστοποιήσουν την απόδοση
- Μεγάλη συνέπεια formats, πρωτοκόλλων, σύνταξη ερωτημάτων
- Μικρή καθυστέρηση μεταξύ χρηστών – δεδομένων

- **Μειονεκτήματα**

- Οι συλλογές πρέπει να ελέγχουν την ευρετηρίαση: δεν μπορεί να εγγυηθεί η ομοσπονδία για όλες τις συλλογές QOS
- Οι εφαρμογές – χρήστη τείνουν να γίνουν μονολιθικές (δύσκαμπτες)



# Κατανεμημένες Ψ.Β : harvesting





# Harvesting: υπέρ και κατά

---

- **Πλεονεκτήματα**

- Αρχιτεκτονική τριών στρωμάτων
- Ο Harvester ελέγχει την ευρετηρίαση → συνεπής τρόπος έκφρασης ερωτημάτων
- Μειωμένος φόρτος στις συλλογές
- Ευκολία στην ενσωμάτωση συλλογών

- **Μειονεκτήματα**

- Καθυστέρηση μεταξύ πελατών και δεδομένων
- Αντιγραφή μεταδεδομένων: Προβληματικό caching



# Πώς διαφέρει μια Ψ.Β. Από το WWW?

---

- Η λέξη – κλειδί είναι **οργάνωση**
  - Το WWW σα σύνολο δεν έχει πραγματική οργάνωση
- Κάποιες μετα-μηχανές αναζήτησης (Yahoo, Lycos) προσπαθούν να θέσουν το Web κάτω από ένα οργανωτικό πλαίσιο
  - Παρολαυτά, το μόνο που υποστηρίζεται είναι αναζητήσεις με λέξεις – κλειδιά (π.χ., Altavista)



# Πώς διαφέρει μια Ψ.Β. Από το WWW?

---

- **Μια βασική διαφορά είναι ποιός ελέγχει την είσοδο δεδομένων στο σύστημα**
  - Οι περισσότερες μετα-μηχανές αναζήτησης αναζητούν τα ψηφιακά αντικείμενα
    - *Lycos is short for *Lycosidae lycosa* (the «wolf spider»), which pursues its prey and does not build a web (Mauldin, IEEE Expert, 1/97)*
  - Κάποιες (Yahoo) έχουν ανθρώπους για επιθεώρηση και ταξινόμηση
- **Γενικά, οι Ψ.Β. Έχουν μεγαλύτερο έλεγχο, ενώ επίσης απευθύνονται σε πιο συγκεκριμένο κοινό**



# Τέλος παρουσίασης

---

- Απορίες;

