

Γενικές Έννοιες Ψηφιακών Αντικειμένων

Επισκόπηση Θεμάτων

- Ψηφιακή αποθήκευση και αναπαράσταση
- Ψηφιακά αντίγραφα
- Αναλυτική ή δομική αναπαράσταση
- Ψηφιακό περιεχόμενο
- Τεκμήριο – πόρος – αντικείμενο
- Περιεχόμενο – είδη και μέσο

Συμβατική ή Ψηφιακή Αποθήκευση

- Συμβατικό \equiv μη ψηφιακό – συχνά αναλογικό
- Τα φυσικά αντικείμενα δεν είναι ψηφιακά
- Το ψηφιακό αντικείμενο είναι πιο δύσκολο «αναγνώσιμα» από ανθρώπους: θέλουν ειδικές συσκευές ανάγνωσης, γνώση του τρόπου ερμηνείας της αναπαράστασης, εξαρτώνται από την τεχνολογία, ...

Παραδείγματα Αποθήκευσης

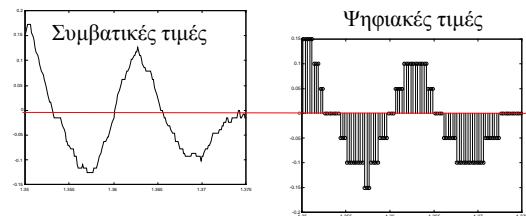
- Παραδείγματα συμβατικής αποθήκευσης
 - Χαρτί, βιντεοκασέτα, κασέτα μουσικής, βινύλιο
- Παραδείγματα ψηφιακής αποθήκευσης
 - Δισκέτα, CD (ακόμα και μουσικής), μνήμη
- Παραδείγματα αποθήκευσης – στο ίδιο μέσο
 - Μια ταινία – είναι συμβατική ή ψηφιακή;

Διαφορά Συμβατικής – Ψηφιακής Αναπαράστασης

- Η συμβατική είναι μια αδιάκοπη διαδοχή τιμών
 - Με συνεχές φάσμα τιμών, πιθανότατα πολυδιάστατες
- Η ψηφιακή είναι μια αλληλουχία τιμών
 - Με διακριτές τιμές (όπως οι ακέραιοι), ανά διάσταση
- Η ψηφιακή αναπαράσταση είναι αλλοιωμένη φυσική αναπαράσταση, σε βαθμό ανάλογο με
 - τη **συχνότητα δειγματοληψίας** (σε κάποια διάσταση)
 - και το **βήμα των τιμών** (στις άλλες διαστάσεις)
 - Αυτά αντιστοιχούν σε ψηφιοποίηση ήχου (χρόνος, συχνότητα, ένταση, ...), εικόνας (μέγεθος, RGB), ...

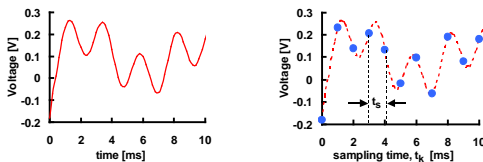
Παράδειγμα Ψηφιοποίησης

Κατά την ψηφιοποίηση, διαφέρει η συχνότητα δειγματοληψίας και το βήμα των τιμών

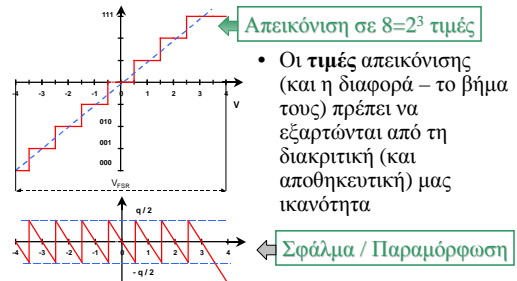


Ψηφιοποίηση και Συχνότητα Δειγματοληψίας

- Η συχνότητα (και άρα ο χρόνος, ή το διάστημα, ή ...) **δειγματοληψίας** πρέπει να εξαρτάται από τη συχνότητα μεταβολής των δεδομένων.



Συνεχές Αναλογικό Σήμα, Ψηφιοποιημένο σε 2^3 τιμές



- Οι **τιμές** απεικόνισης (και η διαφορά – το βήμα τους) πρέπει να εξαρτώνται από τη διακριτική (και αποθηκευτική) μας ικανότητα

Πλεονεκτήματα Ψηφιακής Αναπαράστασης

Η ψηφιακή αναπαράσταση προσφέρει περαιτέρω

- **Αναπαράσταση ή ανάγνωση με ακρίβεια**
 - Ακόμα και με εξοπλισμό «φυσιολογικής» ακρίβειας
 - Χωρίς φυσικούς περιορισμούς, από οσοδήποτε μακριά
- **Αναπαραγωγή χωρίς επιπρόσθετη παραμόρφωση**
 - Εύκολα: εξοπλισμός, διαδικασία, μέσο αναπαραγωγής
 - Είναι τέλειο αντίγραφο, καταλληλότερο για διατήρηση
- **Σύγκριση** – ταυτοποίηση αντιγράφων/αντιτύπων
 - Επιβεβαίωση: π.χ. (7, 15, 3, 28, 12, sum=65)

Αντίγραφο Αντικειμένου

- Στον συμβατικό κόσμο δεν υπάρχουν ακριβή αντίγραφα
 - Όλα είναι διαφορετικά μεταξύ τους (π.χ. φωτεινότητα)
 - Ακόμα και αν οπτικά δεν βλέπουμε τη διαφορά
- Στον ψηφιακό κόσμο δεν υπάρχουν αντίγραφα
 - Υπάρχουν πολλά πρωτότυπα (ακριβώς ίδια)
 - Άρα η φυσική ιδιοκτησία τους έχει μειωμένη αξία
 - Δημιουργεί επιτακτικότερη ανάγκη για προστασία της πνευματικής ιδιοκτησίας – **Copyright**



Αναλυτική ή Δομική Αναπαράσταση

- **Αναλυτική:** περιγράφει επακριβώς την πληροφορία (σε κάποια ανάλυση)
 - Π.χ. με bitmaps τιμών **01111110**
 - Σε μορφότυπα όπως TIFF, WAV, MPG **01100110**
 - Παράγεται από ψηφιοποίηση ή για φυσική απεικόνιση **01111110**
- **Δομική:** περιγράφει ένα τρόπο δημιουργίας της πληροφορίας – γλώσσα περιγραφής
 - Π.χ. κύκλος(), γραμμή(), κείμενο(), ...
 - Σε μορφότυπα όπως EPS, MIDI, διάφορα πολυμέσα
 - Στο τέλος μετατρέπεται σε αναλυτική (με ανάλυση ...)

... 01111110 01100110 ...

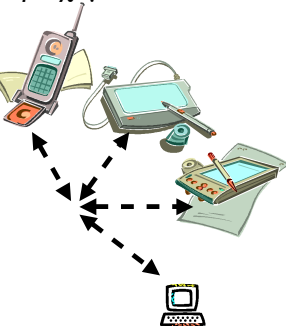
Περιορισμοί Ψηφιακού Αντικειμένου

Το ψηφιακό αντικείμενο εξαρτάται από

- Υπολογιστικό περιβάλλον
 - Unix, Windows, Mac, PDA, MP3-player
- Πρότυπα αναπαράστασης ή ιδιογενείς εφαρμογές
 - .doc, .xls, .pdf, MS Internet Explorer, Netscape
- Το φυσικό μέσο αποθήκευσης
 - Δισκέτες, σκληροί δίσκοι, ταινίες, CD, DVD
 - Πιθανά σφάλματα στην ανάκτησή των ψηφιακών αντικειμένων από το φυσικό μέσο μπορεί να έχουν σοβαρές επιπτώσεις

Ψηφιακό Περιεχόμενο

(Ψηφιακή) Πληροφορία (επεξεργασμένη ή μη) σε οποιοδήποτε υπολογιστικό σύστημα ή διαδικτυακή πλατφόρμα (από τον Ιστό μέχρι τις ασύρματες ή κινητές συσκευές και το βίντεο ευρείας ζώνης)



Παρανοήσεις για Ψηφιακό Περιεχόμενο

- *Το Ψηφιακό Περιεχόμενο είναι πάντα καλύτερο;*
- Ανάλογα με τους στόχους μας:
 - Έχει χρήσιμες ιδιότητες
 - Εύκολη αντιγραφή, με σταθερή ποιότητα
 - Εύκολη επεξεργασία από μηχανές
 - Αλλά είναι πιο δύσκολα «αναγνώσιμο» από ανθρώπους
 - Θέλει ειδικές συσκευές ανάγνωσης, γνώση του τρόπου ερμηνείας του, εξαρτάται από την τεχνολογία, ...
 - Πιθανά σφάλματα στην ανάκτησή του από το φυσικό μέσο μπορεί να έχουν σοβαρές επιπτώσεις

Γιατί «Ψηφιακά Βιβλία»;

Και κάθε είδους **ψηφιακή πληροφορία**

- Κείμενο
 - Εικόνες, κινούμενες εικόνες
 - Ήχος – μουσική
 - Βίντεο
 - Προγράμματα υπολογιστών (λογισμικό)
 - Πολυμέσα, ιστοσελίδες, ..., συνδυασμοί
- Πως θα είναι τα «ψηφιακά άπαντα του Καβάφη»;

Ηλεκτρονικό Βιβλίο

Είναι κάθε βιβλίο που μπορεί να φορτωθεί και διαβαστεί με υπολογιστή ή εξειδικευμένο μηχάνημα



Η Ψηφιακή Βιβλιοθήκη δεν έχει (συνήθως) ηλεκτρονικά βιβλία, αλλά ηλεκτρονική πληροφορία

Τι είναι Ψηφιακό «Τεκμήριο»

- Ηλεκτρονική έκδοση ενός εντύπου
- Οποιοσδήποτε «πόρος» σε ψηφιακή μορφή
 - Γεννημένο ψηφιακά
 - Ψηφιοποιημένο από συμβατικό πρωτότυπο
 - Και μεταδίδεται ηλεκτρονικά ...
- Δεν είναι συμβατικό αντικείμενο
 - Δεν υπόκειται στους φυσικούς περιορισμούς του
 - Δεν έχει ψηφιακή υπόσταση, αλλά bits και bytes

Βασικές Έννοιες – Αντικείμενα

Από το γενικότερο προς το ειδικότερο

- (Πληροφοριακό) Αντικείμενο
 - Με «οτιδήποτε» (μαύρο κουτί) ιδιότητες συμπεριφοράς
- (Πληροφοριακός) Πόρος
 - Χρήσιμο, αναζητήσιμο και προσδιορισμένο αντικείμενο
- (Πληροφοριακό) Τεκμήριο
 - Μεταφέρει πληροφορία: ήχος, βίντεο, τεκμήριο
- Κείμενο
 - Πρωταρχικά λέξεις για ανάγνωση (στατική μορφή)
- Διαχωρίζονται από τον τρόπο «χρήσης»

Πόρος (Resource) – κατά RFC 2396

- A resource can be anything that has identity. Familiar examples include an electronic document, an image, a service (e.g., "today's weather report for Los Angeles"), and a collection of other resources. Not all resources are network "retrievable"; e.g., human beings, corporations, and bound books in a library can also be considered resources. The resource is the conceptual mapping to an entity or set of entities, not necessarily the entity which corresponds to that mapping at any particular instance in time. Thus, a resource can remain constant even when its content--the entities to which it currently corresponds---changes over time, provided that the conceptual mapping is not changed in the process.

Ο Πόρος (Resource) – Περιληπτικά

- Ένας πόρος προσδιορίζεται με ένα URI
 - URI = Uniform Resource Identification
 - [απόλυτο-URI | σχετικό-URI] ["#" id-αποσπάσματος]
 - URL = Uniform Resource Location \subseteq URI
- Ο πόρος είναι διαφορετικός από τις οντότητες που προσδιορίζονται σε κάποια συγκεκριμένη στιγμή
 - Π.χ. το σημερινό τοπικό δελτίο καιρού
- Ο πόρος που προσδιορίζεται από ένα URI μπορεί να είναι **αφηρημένος**
 - Π.χ. μη αναζήτησιμος δικτυακά (όπως ένας άνθρωπος)

Οι Ψηφιακοί Πόροι Μπορεί να Είναι – Περιέχουν ...

- Απλές, οικείες οντότητες



- Περιπλοκά, σύνθετα, δυναμικά αντικείμενα



Ψηφιακός Πόρος – «Ηλεκτρονικό» Δημοσίευμα

Οτιδήποτε προσφέρει πληροφορία στον ψηφιακό κόσμο

- Αντικείμενα με ψηφιακή «υπόσταση»
 - Αρχεία διάφορων μορφών
 - Εφαρμογές λογισμικού
 - Ψηφιακές υπηρεσίες
- Είδη ηλεκτρονικών δημοσιευμάτων (αντικειμένων – πόρων):
- Μορφοποιημένα (σταθερά) – π.χ. σελίδες βιβλίων, εικόνες
 - Μορφοποιησιμα (μεταβλητά) – π.χ. αρχεία επεξεργαστή κειμένου, πίνακες αριθμητικών δεδομένων
 - Δυναμικά, δημιουργούμενα κατά την ανάκτηση τους (από προγράμματα)

Μέσο Αντικειμένου και Βιβλιοθήκη

- Έντυπο
 - Σε χαρτί απαραίτητα, π.χ. «βιβλίο»
 - Συμβατικό (ή Παραδοσιακό)
 - Σε χαρτί ή άλλο μέσο που έχει φυσική υπόσταση
 - Ψηφιακό (ή Ηλεκτρονικό)
 - Που διακινείται ηλεκτρονικά
 - Π.χ. το «CDROM» τι είναι;
-
- Αντίστοιχα Βιβλιοθήκες συμβατικές, ψηφιακές ...

Ορισμός και Σημασία των Ψηφιακών Βιβλιοθηκών

Επισκόπηση Θεμάτων

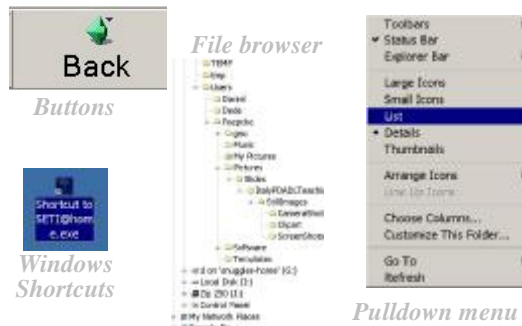
- Ψηφιακό περιεχόμενο σήμερα
- Τι είναι «Ψηφιακή Βιβλιοθήκη»
- Εφαρμογές Ψηφιακών Βιβλιοθηκών
- Οργάνωση Ψηφιακών Βιβλιοθηκών
- Συστατικά Ψηφιακής Βιβλιοθήκης
- Ψηφιακό τεκμήριο
- Παραδείγματα Ψηφιακής Βιβλιοθήκης
- Προκλήσεις Ψηφιακών Βιβλιοθηκών

Καθημερινή Πρόσβαση στο Ψηφιακό Περιεχόμενο

- Με σταθερές ή φορητές συσκευές
- Με ενσύρματη ή ασύρματη σύνδεση



Με Απλές Γνώριμες Διεπαφές



Η Νέα Διάσταση των Τεχνολογικών Επενδύσεων

Οικονομικός Ταχυδρόμος, Τεύχος 52 (5238), 28 Δεκεμβρίου 2002

... που θα συμβάλουν στην κατανόηση από τις επιχειρήσεις του πραγματικού νοήματος της χρήσης των ψηφιακών τεχνολογιών, που δεν είναι τίποτ' άλλο παρά η αποτελεσματική υποβοήθηση της διαχείρισης της γνώσης και της οργάνωσης της δημιουργικής εργασίας.

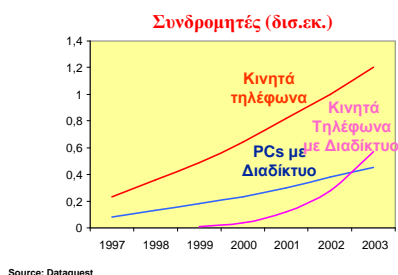
Αλέξανδρος-Ανδρέας Κύρτσης

Αν. Καθηγητής Κοινωνιολογίας στο Τμήμα Οικονομικών Επιστημών του ΕΚΠΑ

Εξάπλωση Ψηφιακών Βιβλιοθηκών

- Με την ραγδαία εξέλιξη της τεχνολογίας των πολυμέσων, δημιουργούνται ψηφιακές βιβλιοθήκες εικόνων, ήχου και βίντεο. Μπορούμε ακόμα να δούμε σαν βιβλιοθήκες τις συλλογές που είναι αποθηκευμένες σε βάσεις δεδομένων, βάσεις γνώσεως, βάσεις κειμένων ή στο WWW
- Την ίδια ώρα, ένα συνεχώς αυξανόμενο ποσοστό του προϋπολογισμού των βιβλιοθηκών αφιερώνεται σε ηλεκτρονικές υπηρεσίες, είτε αυτές είναι CD-ROMs, online κατάλογοι κοινής πρόσβασης ή online βάσεις δεδομένων

Ρυθμός Υιοθέτησης της Τεχνολογίας



Μεγαλύτερες Κατηγορίες Αγορών στο Ηλεκτρονικό Εμπόριο

- à Βιβλία και Πληροφορία (9.2 εκ. χρήστες)
- à CD's και Videos (7.2 εκ. χρήστες)
- à Υπολογιστές (5.4 εκ. χρήστες)
- à Ρουχισμός (4.5 εκ. χρήστες)
- à Λογισμικό (4.0 εκ. χρήστες)

Από πάνω από 28 εκατομμύρια αγορές μέσα στο 1999
Commercenet/Nielsen Media Research (1999)

Οικονομική Σημασία

- Κατάσταση 2001:
 - 4.000.000 εργαζόμενοι
 - 412 B Euro, 5% του ΑΕΠ της ΕΥ.
- Ρυθμός αύξησης:
 - Ως και 20% κάθε χρόνο
 - 1.000.000 νέοι εργαζόμενοι για 2002 – 2005!
- Το περιεχόμενο παίζει σημαντικό ρόλο!

Τι είναι «Βιβλιοθήκη»

- Τι είναι «Βιβλιοθήκη»
 - Χώρος συγκέντρωσης, φύλαξης, διατήρησης, περιφρούρησης, οργάνωσης και διάθεσης της πληροφορίας.
- Τι είναι «Ψηφιακή Βιβλιοθήκη»
 - Συγκέντρωση (ή εξασφάλιση πρόσβασης) πληροφορίας (περιεχόμενο)
 - Σε ψηφιακά μέσα
 - Οργανωμένο για εύκολη πρόσβαση
 - Δεν είναι η «Αυτοματοποιημένη βιβλιοθήκη»

Τι είναι «Ψηφιακή Βιβλιοθήκη»

- Οι πηγές πληροφοριών που μπορούν να προσπελαστούν μέσω του Διαδικτύου αποτελούν συστατικά μιας ψηφιακής βιβλιοθήκης. Αυτές περιλαμβάνουν βιβλία, περιοδικά, εφημερίδες, καταλόγους τηλεφώνων, ηχητικές και μουσικές εγγραφές, εικόνες, βίντεοκλιπ, επιστημονικά δεδομένα και προσωπικές πληροφοριακές υπηρεσίες όπως κάποιες αναφορές σχετικά με τα αποθέματα μιας επιχείρησης. Αυτές οι πληροφοριακές πηγές, όταν συνδεθούν ηλεκτρονικά μέσω ενός δικτύου, αναπαριστούν σημαντικούς συντελεστές μιας αναδυόμενης, διεθνώς προσβάσιμης ψηφιακής βιβλιοθήκης.
 - Digital Library Initiative

Τι είναι «Ψηφιακή Βιβλιοθήκη»

- Μια ψηφιακή βιβλιοθήκη είναι ένα καταναμημένο τεχνολογικό περιβάλλον που ελαττώνει δυναμικά τα εμπόδια της δημιουργίας, διανομής, διαχείρισης, αποθήκευσης, ολοκλήρωσης και επαναχρησιμοποίησης των πληροφοριών από ιδιώτες ή ομάδες ερευνητών

– Source Book on Digital Libraries, Edward A. Fox

Τι είναι «Ψηφιακή Βιβλιοθήκη»

- Οι ερευνητές των οποίων ο τομέας δραστηριότητας τους είναι οι βάσεις δεδομένων ή η ανάκτηση πληροφοριών, βλέπουν τις ψηφιακές βιβλιοθήκες ως ένα σύνολο ενοποιημένων βάσεων δεδομένων
- Οι ασχολούμενοι με βιβλιοθήκες, βλέπουν τις ψηφιακές βιβλιοθήκες ως την συνεχής τάση προς την αυτοματοποίηση και εξέλιξη των φυσικών βιβλιοθηκών. Παροτρύνονται νέες πληροφοριακές πηγές, νέοι τρόποι απόκτησης γνώσης, νέοι μέθοδοι αποθήκευσης και ασφάλειας των πληροφοριών και νέες προοπτικές στην ταξινόμηση και οργάνωση των καταλόγων της βιβλιοθήκης
 - ["Digital Libraries: Issues and Architectures"](#), Peter J. Nörnberg, [Digital Libraries Conference 1995](#)

Τι είναι «Ψηφιακή Βιβλιοθήκη»

- Οι *Ψηφιακές Βιβλιοθήκες* ασχολούνται με τη **δημιουργία** και τη **διαχείριση** πληροφοριακών πόρων, τη **μεταφορά** πληροφορίας ανάμεσα σε καθολικά δίκτυα και την αποτελεσματική **χρήση** αυτής της πληροφορίας από ένα μεγάλο εύρος χρηστών
 - International Journal on Digital Libraries, 1997

Τι είναι «Ψηφιακή Βιβλιοθήκη»

"... Ένα διαχειριζόμενο περιβάλλον πολυμεσικού υλικού σε ψηφιακή μορφή, σχεδιασμένο χάριν των χρηστών του, δομημένο έτσι ώστε να διευκολύνει την πρόσβαση στα περιεχόμενά του, και εφοδιασμένο με βοηθήματα για την πλοήγηση του παγκόσμιου δικτύου ... με χρήστες και υλικό πλήρως καταναεμημένα, αλλά διαχειριζόμενα σαν ένα συνεκτικό σύνολο."

–Collier, Mel *International Symposium on Research, Development, and Practice in Digital Libraries 1997*

Τι είναι «Ψηφιακή Βιβλιοθήκη»

«...η έννοια της "ψηφιακής βιβλιοθήκης" δεν ισοδυναμεί απλά με μια ψηφιοποιημένη συλλογή με εργαλεία διαχείρισης πληροφορίας. Αφορά μάλλον ένα **περιβάλλον** το οποίο ενώνει συλλογές, υπηρεσίες και ανθρώπους για την υποστήριξη ολόκληρου του κύκλου ζωής της δημιουργίας, διάχυσης, χρήσης και συντήρησης δεδομένων, πληροφορίας και γνώσης.»

Santa Fe planning workshop on distributed knowledge work environments, Digital Library Initiative (DLI), March 1997

Τι είναι «Ψηφιακή Βιβλιοθήκη»

- "...the concept of a "digital library" is not merely equivalent to a digitized collection with information management tools. It is rather an environment to bring together collections, services, and people in support of the full life cycle of creation, dissemination, use, and preservation of data, information, and knowledge."

Santa Fe planning workshop on distributed knowledge work environments, Digital Library Initiative (DLI), March 1997

Τι είναι «Ψηφιακή Βιβλιοθήκη»

- Digital Libraries are organized collections of digital information. They combine the structure and gathering of information, which libraries and archives have already done, with the digital representation that computers have made possible

– Lesk, 1997

Τι είναι «Ψηφιακή Βιβλιοθήκη»

- Digital Libraries are organizations that provide resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities

– DLF, 1999

Τι είναι «Ψηφιακή Βιβλιοθήκη»

1. Digital Libraries are a set of electronic resources and associated technical capabilities for creating, searching, and using information. ... that are an extension and enhancement of information storage and retrieval systems that manipulate digital data in any medium ... The content of digital libraries includes data, [and] metadata ...
2. Digital Libraries are constructed, collected, and organized, by (and for) a community of users, and their functional capabilities support the information needs and uses of that community

– Borgman, 1999

Τι είναι «Ψηφιακή Βιβλιοθήκη»

- Σύμπλοκο υπολογιστικών συστημάτων, λογισμικού και δικτύωσης
- που διαχειρίζεται ψηφιοποιημένες πληροφορίες διαφορετικής προέλευσης, περιεχομένου, μορφοτύπου και δυνατοτήτων περαιτέρω επεξεργασίας.
- Παρέχει με ολοκληρωμένο τρόπο, στο χρήστη
- δυνατότητες πρόσβασης, αναζήτησης παρουσίασης και αξιοποίησης πληροφοριών
- από διασυνδεδεμένα υποσυστήματα εγκατεστημένα σε διαφορετικούς γεωγραφικούς τόπους

Τι είναι «Ψηφιακή Βιβλιοθήκη»

- ένα σύνολο υπηρεσιών το οποίο αναφέρεται στην πρόσκτηση, αποθήκευση, πρόσβαση και διακίνηση πληροφορίας και υλικού το οποίο είναι σε ψηφιακή μορφή



Τι είναι «Ψηφιακή Βιβλιοθήκη»

- Συλλογή εγγράφων οργανωμένων σε ηλεκτρονική μορφή, διαθέσιμα στο Διαδίκτυο ή σε CDROM

Διαδίκτυο

Παροδική(?) διαθεσιμότητα

Χαμηλή(?) ταχύτητα

Καθημερινή ανανέωση

Απεριόριστη χωρητικότητα

Αμελητέο κόστος

CDROM

Συνεχής διαθεσιμότητα

Υψηλή ταχύτητα

Αδύνατη ανανέωση

Μικρή χωρητικότητα

Κόστος (κυρίως) διανομής

Ορισμοί Ψηφιακής Βιβλιοθήκης

- Υπάρχουν πολλοί ορισμοί για την ψηφιακή βιβλιοθήκη, κανείς δεν δίνει την απόλυτη έννοια του όρου, ούτε όμως αναρεί άλλους. Τελικά, ο σωστός ορισμός της ψηφιακής βιβλιοθήκης είναι συνδυασμός όλων αυτών των απόψεων

Υπάρχουν πολλές σχετικές δραστηριότητες όπως:

- Βάση δεδομένων με πολυμέσα
- Περιβάλλον ανάκτησης πληροφοριών
- Online αποθετήρια πληροφοριών
- Ηλεκτρονικό σύστημα βιβλιοθήκης
- Σύστημα επεξεργασίας κειμένων και πολυμέσων
- Ο Παγκόσμιος Ιστός (WWW)
- Διασυνδεδεμένες Πληροφοριακές Υπηρεσίες

Τι είναι και τι δεν είναι σε καθένα από αυτά;

Τι είναι οι «Ψηφιακές Βιβλιοθήκες»

- Είναι **Βάσεις Δεδομένων** που
 - Έχουν **ψηφιακά αντικείμενα**
 - Είναι περισσότερο **καταναμημένες**
 - Έχουν περισσότερη **ανομοιογένεια**
 - Έχουν λιγότερες δυνατότητες **αναζήτησης**
 - Έχουν λιγότερες **ενημερώσεις**

Τι είναι οι «Ψηφιακές Βιβλιοθήκες»

- Είναι **μηχανές αναζήτησης** που
 - Έχουν επίσημη διαδικασία **καταχώρησης**
 - Έχουν **έλεγχο ποιότητας** στις προσκτήσεις
 - Δεν περιέχουν **παραπλανητικά** μεταδεδομένα
 - Όπως και το <http://www.google.com>
 - Είναι λιγότερο **καταναμημένες**
 - Έχουν περισσότερα **μεταδεδομένα**

Ομοιότητες Διαδικασιών και Αρχιτεκτονικής Ψηφιακών Βιβλιοθηκών και Ιστού

Αντιστοίχιση κύριων σημείων:

- Πόροι (αντικείμενα): Αρχεία, CGI, MIME-Types
- Αναγνωριστικά – ονοματολογία: URLs
- Υπηρεσίες πόρων: εξυπηρετητές HTTP
- Ανακάλυψη πόρων: web crawlers, Ιστός
- Συλλογές: ιστότοποι, αυτοσχεδιασμοί
- Πολιτικές / κανόνες (περιορισμών) πρόσβασης: Λίστες ελέγχου πρόσβασης, συνθηματικές λέξεις, αυτοσχεδιασμοί

Ο Ιστός δεν είναι Ψηφιακή Βιβλιοθήκη

- Η πληροφορία στον Ιστό
 - Δεν είναι (αρκετά) δομημένη
 - Έχει απλοϊκή διεπαφή (μηχανές αναζήτησης)
 - Είναι δύσκολο να ανακτηθεί (πλήρως)
 - Είναι παροδική
 - Δεν έχει διαδικασία πρόσκτησης
 - Και συχνά είναι παραπλανητική
 - The Italian Republic of West Cyprus
 - <http://www.turcoman.binternet.co.uk/italian-republic-west-cyprus.htm>
- Αλλά συχνά μια Ψηφιακή Βιβλιοθήκη είναι προσπελάσιμη από τον Ιστό

Ανάμεσα σε Ιστό και ΒΔ



Γιατί «Ψηφιακές» Βιβλιοθήκες

- Για μεγάλη ποσότητα πληροφοριών
 - Που δεν μπορεί να περιέχεται σε ένα σύστημα
- Για άμεση (και απομονωμένη) πρόσβαση
 - Σε καταναμημένη ή μη πληροφορία
- Για μείωση κόστους
 - Εκδόσης, πρόσβασης, συντήρησης, ...
- Για νέα λειτουργικότητα
 - Ενσωμάτωση νέων (και μελλοντικών) μορφοτύπων

Στόχοι Μετατροπής Βιβλιοθήκης σε Ψηφιακή

- Εντατικοποίηση χρήσης «υλικού - προσβάσεων»
- Βελτιστοποίηση δαπάνης για υλικό-πρόσβαση
- Ταχεία απορρόφηση τεχνολογίας (προσωπικό - χρήστες)
- Εξασφάλιση εναλλακτικών λύσεων, σε
 - Τεχνολογική υποδομή (CD-ROM - INTERNET)
 - Πηγές (εκδότες - ενδιάμεσοι)
 - Υλικό φορέα (έντυποι - ηλεκτρονικοί τίτλοι)
 - Προέλευση υλικού (εθνική - διεθνής παραγωγή)

Η Υβριδική Βιβλιοθήκη: καθ' οδόν προς την Ψηφιακή

- Το ενδιάμεσο βήμα ανάμεσα στην συμβατική και την ψηφιακή, όπου οι ηλεκτρονικές και οι έντυπες πηγές χρησιμοποιούνται παράλληλα
- «Σχεδιασμένη να συνενώσει ποικίλες τεχνολογίες από διαφορετικές πηγές στο καθημερινό περιβάλλον λειτουργίας μιας βιβλιοθήκης»

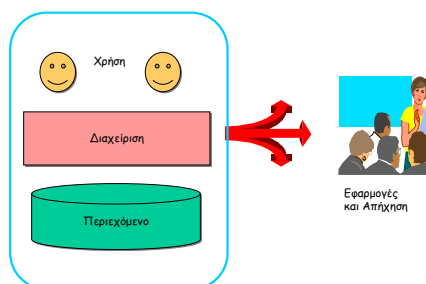


Πεδία Εφαρμογής Ψηφιακών Βιβλιοθηκών

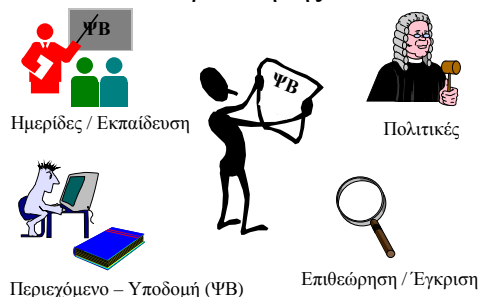
Η ψηφιακή βιβλιοθήκη μπορεί να σχετίζεται: ...

- Αρχεία
- Βιβλιοθήκες
- Μουσεία
- Αποθήκες
- Προϊόντα καταστημάτων
- Εμπόριο, επιχειρήσεις
- Κρατικές υπηρεσίες
- Εκπαίδευση εξ αποστάσεως
- Μέσα Μαζικής Ενημέρωσης

Σχεδιασμός Ψηφιακής Βιβλιοθήκης



Δημιουργία Τοπικής Ψηφιακής Βιβλιοθήκης

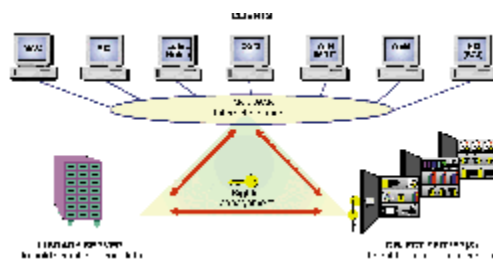


Νομική, Οικονομική, Κοινωνική Οργάνωση

- Η ψηφιακή βιβλιοθήκη σχετίζεται με μια πολύ μεγαλύτερη νομική, οικονομική, κοινωνική οργάνωση.
- Η νομική διάσταση έγκειται στην κωδικοποίηση της δυναμικής εξέλιξης της ευρύτερης οικονομικής και κοινωνικής οργάνωσης. Οι σχετικές περιοχές της νομικής περιλαμβάνουν τα δικαιώματα πνευματικής ιδιοκτησίας, απόδοσης και άλλων θεμάτων όπως κυριότητα, δυσφήμιση, δίκαιο επικοινωνιών και διεθνές δίκαιο

Παραδείγματα Ψηφιακών Βιβλιοθηκών

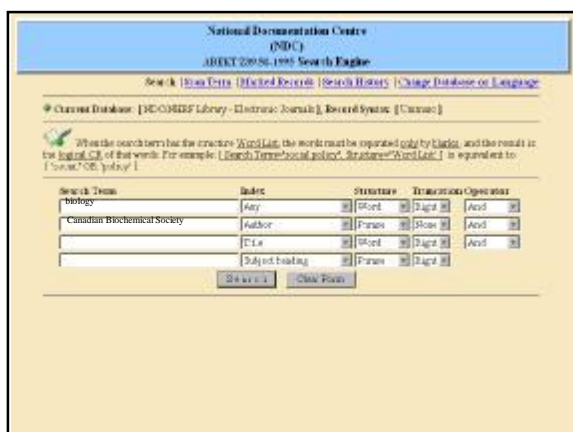
Η Ψηφιακή Βιβλιοθήκη



Παράδειγμα Ψηφιακής Βιβλιοθήκης με Περιοδικά

Από το ΕΚΤ, με χρήση ΑΒΕΚΤ / Ζ39.50

- Υποστηρίζει αναζήτηση
 - Με διάφορα κριτήρια
- Δίνει πρόσβαση στα περιεχόμενα
 - Με τρόπο που εξαρτάται από τα περιεχόμενα (αντικειμενοστραφή).

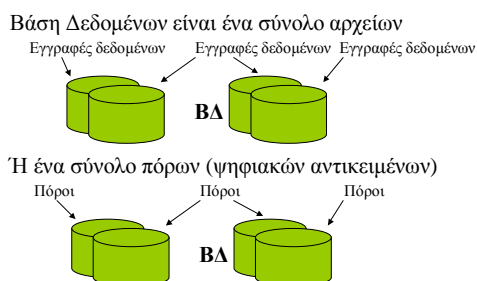


Γενικές Έννοιες Ψηφιακών Βιβλιοθηκών

Κατανόηση των Εννοιών μιας Ψηφιακής Βιβλιοθήκης

- Τα συστατικά μιας ψηφιακής βιβλιοθήκης ονομάζονται «ψηφιακά αντικείμενα» (digital objects) ή «πόροι» (resources)
- Αυτά αποθηκεύονται σε «αποθετήρια δεδομένων» (repositories)
- Προσδιορίζονται από τα «αναγνωριστικά» (handles ή identifiers)
- Οι πληροφορίες που είναι αποθηκευμένες σε ένα ψηφιακό αντικείμενο αποτελούν το «περιεχόμενο» (content)
- Αυτό διαχωρίζεται σε «δεδομένα» (data) και πληροφορίες για αυτά, τις «ιδιότητες» (properties) ή μεταδεδομένα

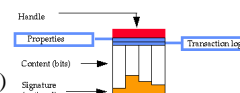
Βάσεις Δεδομένων / Αντικειμένων



Αντικείμενα Ψηφιακής Βιβλιοθήκης

Σε μια ψηφιακή βιβλιοθήκη οι πληροφορίες καταχωρούνται ως «ψηφιακά αντικείμενα», με

- Περιεχόμενο (content)
- Ιδιότητες (Properties)
- Αναγνωριστικό (Handle)
- Ψηφιακή υπογραφή (signature)
- «Σημειωματάριο» όλων των δραστηριοτήτων (transaction log)



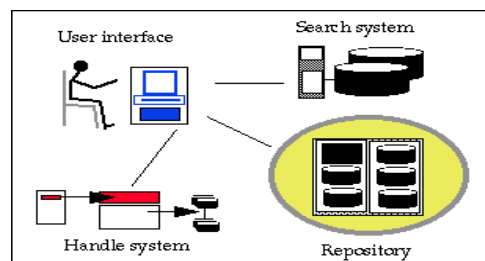
Βασικές Συνιστώσες Ψηφιακής Βιβλιοθήκης

Οι συνιστώσες εκφράζουν (συνεργαζόμενα) τμήματα με ανεξάρτητους ρόλους και σχεδιασμούς

Οι «βασικές» συνιστώσες της δομής μιας τυπικής ψηφιακής βιβλιοθήκης είναι συχνά θέμα οπτικής γωνίας:

- Διεπαφές με τον χρήστη (**User interfaces**)
- Αποθετήριο δεδομένων (**Repository**)
- Σύστημα αναγνώρισης (ή προσδιορισμού ή ονοματολογίας – **Handle system**)
- Σύστημα αναζήτησης (**Search system**)

Βασικές Συνιστώσες Ψηφιακής Βιβλιοθήκης – Σχηματικά



Άλλες Συνιστώσες Ψηφιακής Βιβλιοθήκης

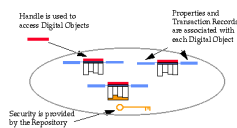
- Πύλη πρόσβασης (gateway – portal)
- Σύνολο εξωτερικών online πηγών
- Στρατηγική διατήρησης, αλλά και συντήρησης
- Πακέτο ηλεκτρονικών εργαλείων (διαχείρισης ...)

Αποθηκευμένα Αντικείμενα;

- Σε μια ψηφιακή βιβλιοθήκη, ότι είναι αποθηκευμένο δεν ανακτάται πάντα όπως είναι.
- Ο χρήστης γίνεται δέκτης του αποτελέσματος της εκτέλεσης ενός προγράμματος πάνω στο αποθηκευμένο αντικείμενο.
 - Αυτό μπορεί να είναι απλό, όπως ένα πρόγραμμα μεταφοράς αρχείων (file transfer program)
 - Ή κάτι σαφώς πιο πολύπλοκο (π.χ. να βάζει υδατογραφήματα στις εικόνες).

Αποθετήρια Δεδομένων (repositories)

- Τα ψηφιακά αντικείμενα απεικονίζουν πολύτιμες πνευματικές ιδιοκτησίες
- Η μορφή τους μέσα στο αποθετήριο περιέχει πληροφορίες που επιτρέπει τη διαχείρισή τους, εντός του οικονομικο-κοινωνικού συστήματος
- Το αποθετήριο διατηρεί αυτές τις πληροφορίες και εξασφαλίζει την απαραίτητη ασφάλεια



Αναγνωριστικά

Τα αναγνωριστικά αποτελούν ζωτικής σημασίας δομικό υλικό μιας ψηφιακής βιβλιοθήκης. Η ύπαρξή τους διευκολύνει:

- Τον προσδιορισμό ψηφιακών αντικειμένων
- Την δήλωση της πνευματικής κυριότητας των ψηφιακών αντικειμένων
- Την καταγραφή των αλλαγών ιδιοκτησίας

Είναι απαραίτητα για παραπομπές, για ανάκτηση πληροφοριών και χρησιμοποιούνται ως σύνδεσμοι (links) ανάμεσα στα διάφορα αντικείμενα

Συλλογή

- Η συμβατική συλλογή αποτελείται από διακριτά φυσικά αντικείμενα
- Η ιδιοκτησία και η φυσική εγγύτητα δεν είναι πλέον προαπαιτούμενα για την πρόσβαση στο ψηφιακό περιεχόμενο.
 - Τοπική διάθεση
 - Απομακρυσμένη πρόσβαση
 - Ελεύθερα διαθέσιμο περιεχόμενο
 - Περιεχόμενο για το οποίο έχουμε δικαίωμα χρήσης

Νέα Έννοια “Συλλογής”

- Σημαντικό ερώτημα: πώς η κάθε βιβλιοθήκη ορίζει τη συλλογή της
 - Τι συλλέγει
 - Με ποιο μηχανισμό (αντιγραφή, παραπομπή, ...)
 - Πώς το διατηρεί (για μελλοντική εκμετάλλευση)

Λειτουργίες Συμβατικής Βιβλιοθήκης (1)

- Υπηρεσίες “τεχνικής” φύσεως
 - Πρόσκτηση (συλλογή)
 - Οργάνωση (καταλογογράφηση και ταξινόμηση)
 - Διατήρηση
- Υπηρεσίες χρήστη
 - Δανεισμός
 - Πρόσβαση, ανάκτηση και παρουσίαση
 - Ανταπόκριση σε ερωτήσεις παραπομπών, παραγωγή περιλήψεων
 - Ενημέρωση κοινότητας χρηστών

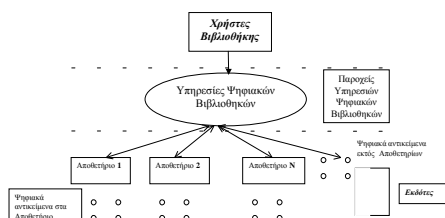
Λειτουργίες Συμβατικής Βιβλιοθήκης (2)

- Σχέσεις με βιβλιοθήκες
 - Συνεργασία μεταξύ βιβλιοθηκών
 - Διαδανεισμός
 - Κοινή πρόσκτηση και αποθήκευση
 - Συνεργατική καταλογογράφηση
 - Ενώσεις και διεθνείς οργανισμοί

Υπηρεσίες της Ψηφιακής Βιβλιοθήκης

- Πρόσκτηση υλικού
- Καταλογογράφηση
- Αποθήκευση και Προστασία υλικού
- Αναζήτηση και Εύρεση (όπως σε OPAC, ...)
- Διακίνηση υλικού (rights management)

Ψηφιακή Βιβλιοθήκη – Σχηματικά



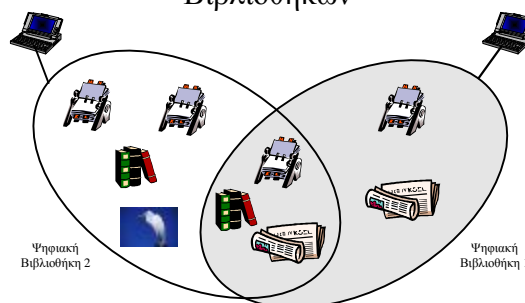
Διαλειτουργικότητα και Μεταδεδομένα

Επισκόπηση Θεμάτων

- Διαλειτουργικότητα
- Ο Ρόλος των Μεταδεδομένων σε Ψηφιακές Βιβλιοθήκες
- Εισαγωγή στα Μεταδεδομένα
- Dublin Core
- Επιλογή Μεταδεδομένων

Διαλειτουργικότητα

Διαλειτουργικότητα Ψηφιακών Βιβλιοθηκών



Διαλειτουργικότητα – Ορισμός

- Η «*Διαλειτουργικότητα*» ορίζεται ως η ικανότητα μεταφοράς και χρησιμοποίησης της πληροφορίας με ένα ομοιογενές και αποτελεσματικό τρόπο μεταξύ διαφόρων οργανισμών σε επίπεδο συστημάτων πληροφορικής
- Η Βρετανική πρωτοβουλία του e-Government χαρακτηρίζει την Διαλειτουργικότητα ως τον ακρογωνιαίο λίθο της γενικής στρατηγικής της Κοινωνίας της Πληροφορίας

Διαλειτουργικότητα Σημαίνει ...

- Τυποποίηση σε όλα τα επίπεδα των εφαρμογών ώστε να υπάρχει προς τα μέσα και προς τα έξω από τον Οργανισμό δυνατότητα σαφούς και ξεκάθαρης επικοινωνίας με ενιαία άποψη.
- Δυνατότητες ολοκλήρωσης της πληροφορίας, εξασφαλίζοντας
 - Ολοκληρωμένα και ενταγμένα σε ενιαίο μοντέλο δεδομένα.
 - Συμβατότητα στην μορφή (format) των δεδομένων.
 - Συμβατότητα στην σημασία των δεδομένων.
 - Συμβατότητα στην ονομασία των δεδομένων.
 - Ενιαίο αποθετήριο μεταδεδομένων (metadata repository).

Οδηγίες Διαλειτουργικότητας ...

- Στα πλαίσια κομβικών έργων, αλλά και γενικά όπου απαιτείται η ενοποίηση πολλών ετερογενών υποσυστημάτων πρέπει να αποφεύγονται λύσεις «point-to-point integration» που οδηγούν σε μεγάλο αριθμό από «απο-έργα» διογκώνοντας το σχετικό προϋπολογισμό και καθυστερώντας την περάτωση των έργων και να δίνεται υψηλή προτεραιότητα στην ενοποίηση μέσα από τη χρήση προτύπων όπως η γλώσσα XML και τα Web Services, με ιδιαίτερη έμφαση στον έγκαιρο καθορισμό της δομής των διακινούμενων μεταξύ των υποσυστημάτων δεδομένων.

Υλοποίηση έργων ηλεκτρονικής διακυβέρνησης: Στόχοι, Σχεδιαστικές Αρχές και Τεχνολογικές Επιλογές

Άλλα Παραδείγματα Διαλειτουργικότητας

- Αυτοκίνητα – χειρισμός και κανόνες οδήγησης
- Ηλεκτρικό ρεύμα και συσκευές
- Ασύρματες επικοινωνίες και κινητά τηλέφωνα
- Δίκτυα υπολογιστών (TCP/IP)
- Συστήματα και εφαρμογές υπολογιστών
- Εφαρμογές σε βιβλιοθήκες (MARC, Z39.50)

Διαλειτουργικότητα σε ...

- Ονοματολογία
 - CNRI Handles
 - PURLs
 - DOIs
 - URNs
- Μεταδεδομένα
 - MARC21
 - UNIMARC
 - Dublin Core
- Μοντέλα πόρων
 - HTML
 - PDF
 - WebDAV
- Ομόσπονδο ψάξιμο
 - Z39.50
 - SRU, SRW
- Υπηρεσίες και πρωτόκολλα
 - Dienst

Στρώματα Διαλειτουργικότητας

Ομοιογένεια διεπαφής ανά στρώμα:

Υπηρεσίες ενδιάμεσων:

Σύνδεσμοι, Αναζήτηση + Αποτελέσματα

Μεταδεδομένα

Μοντέλα δεδομένων (πόρων)

Διαλειτουργικότητα – Επίπεδα

Είναι απαραίτητη εξαιτίας της ανομοιογένειας και αναλύεται σε:

- Πολυγλωσσική (διεθνοποιημένη)
- Διαπεδική (πολυθεματική)
- Οργανωτική (κανόνες, διαδικασίες πρόσβασης)
- Σημασιολογική
 - περιεχόμενο: δεδομένα + μεταδεδομένα
 - Κανόνες γραφής πεδίων
- Συντακτική (τεχνική)

Συντακτική Διαλειτουργικότητα

Η γλώσσα έκφρασης του περιεχομένου. Πρέπει

- Να είναι επαρκώς πλούσια
- Να έχει μονοσήμαντες έννοιες
- Να είναι γνωστή ή εύκολη

Παραδείγματα σε χρήση:

- ASCII
- ISO2709
- XML
 - RDF

Σημασιολογική Διαλειτουργικότητα

Το **περιεχόμενο** που περιγράφεται. Εξαρτάται από τη θεματική περιοχή και τους στόχους μας:

- Τι πεδία / υποπεδία χρειάζονται
- Πώς συνδέονται
- Αν / πώς επαναλαμβάνονται
- Είναι υποχρεωτικά;
- Τι ακριβώς περιέχει κάθε πεδίο
 - Καθιερωμένοι όροι
 - **Κανόνες:** π.χ. Όνομα = επίθετο ή ονοματεπώνυμο;

Προϋποθέσεις Διαλειτουργικότητας

- Ωριμη τεχνολογία
- Κρίσιμη μάζα
 - Μικρό κόστος εγκατάστασης
 - Μεγάλη βάση εγκατάστασης
- Βασική προϋπόθεση: ίδιοι στόχοι ή χρήση
- Εξασφάλιση διαλειτουργικότητας
 - Προέλευση από 1 κατασκευαστή (και μιμητές)
 - Σύγκλιση προδιαγραφών πολλών κατασκευαστών

Παράδειγμα Διαλειτουργικότητας σε Αναζήτηση (με Μεταδεδομένα)

- Σύμφωνα με την ελάχιστη κοινή δομή, π.χ.
 - Συλλογή με **Βιβλία**
 - TITLE, AUTHOR, PUBLISHER, ABSTRACT
 - Συλλογή με **Διατριβές**
 - TITLE, AUTHOR, SUMMARY
 - Συλλογή με **Τραγούδια**
 - TITLE, CREATOR, VERSE
- Κοινά ανά 2; Ανά 3;
 - ABSTRACT = SUMMARY;
 - AUTHOR = CREATOR;

Ανάγκες Ψηφιακών Πόρων

- Η διαλειτουργικότητα μεταδεδομένων απαιτεί
 - Χρήση κανόνων περιεχομένου και προτύπων
 - Σαφήνεια για τους πόρους που περιγράφονται
 - Π.χ. εργασία, παρουσίαση, έκφραση, επιμέρους θέματα
- Οι πραγματικοί πόροι είναι πιο περίπλοκοι από τα (στατικά) αντικείμενα τύπου «τεκμηρίων»
 - Τα χαρακτηριστικά των πόρων αλλάζουν με το χρόνο
 - Οι ενέργειες (των μεσαζόντων) προκαλούν αλλαγές
 - Π.χ. αύξηση της αναγνωσιμότητας, σημειώσεις, ...

Κουλτούρα Διαλειτουργικότητας (– Κείμενο Στρατηγικής για την ΚτΠ – 2004)

Η επιλογή τεχνικών προδιαγραφών διαλειτουργικότητας, βασίζεται, μεταξύ άλλων:

- Υποστήριξη τους από την αγορά (μειώνοντας το αντίστοιχο κόστος και ρίσκο)
- Δυνατότητα κλιμάκωσής τους (όγκος δεδομένων, αριθμός συναλλαγών, πλήθος χρηστών)
- Ανοικτή αρχιτεκτονική και διασύνδεσή τους (υιοθετώντας ανοικτά και ευρέως αποδεκτά πρότυπα, που είναι καταγεγραμμένα και διαθέσιμα στο ευρύ κοινό)

Ιδιογένεια και Διαλειτουργικότητα

- Τα εμπορικά προϊόντα προσπαθούν να «δεσμεύσουν» την αγορά, συχνά δεν ακολουθούν τυχόν πρότυπα, και δημιουργούν ιδιογενείς λύσεις
 - Απευθυνόμενοι στους πολλούς και μεμονωμένους χρήστες
- Μικρότεροι «παίκτες» στην αγορά δεν επιβιώνουν χωρίς διαλειτουργικότητα
 - Απευθυνόμενοι στους «συνεργαζόμενους» χρήστες
- Το ΕΛ/ΛΑΚ δεν έχει στόχο κέρδους και δέσμευσης χρηστών
 - Και απευθύνεται πρωτίστως σε προχωρημένους χρήστες

Ελεύθερο Λογισμικό ή Λογισμικό Ανοικτού Κώδικα (ΕΛ/ΛΑΚ)

Αλήθειες για το ΕΛ/ΛΑΚ:

- Οι μέθοδοι ανάπτυξης του αποτελούν και την εγγύηση για τη δημιουργία λογισμικού υψηλής ποιότητας
- Υπάρχει επαρκής υποστήριξη από δίκτυα χρηστών και πληροφοριών
- Οι άδειες για το ΕΛ/ΛΑΚ έχουν στο επίκεντρό τους το χρήστη, ενώ η επιλογή τους δεν είναι πάντα απλή υπόθεση
- Δεν είναι σε όλες τις περιπτώσεις πιο φθηνό από το ιδιογενές («ιδιοκτησιακό») λογισμικό

Ο Ρόλος των Μεταδεδομένων σε Ψηφιακές Βιβλιοθήκες

Αναγκαιότητα Μεταδεδομένων

- Ο Παγκόσμιος Ιστός είναι περιβάλλον

- αχανές
- ετερογενές
- κατανεμημένο
- δυναμικά μεταβαλλόμενο



- Πώς ενοποιείται η πρόσβαση σε πηγές ψηφιακής βιβλιοθήκης;
- Τα **μεταδεδομένα** είναι κλειδί για τη λύση: είναι η ταυτότητα του Ψηφιακού Αντικειμένου

Μεταδεδομένα

- Τα μεταδεδομένα είναι **δομημένη** πληροφορία που περιγράφει, εξηγεί, εντοπίζει ή διευκολύνει την ανάκτηση, τη χρήση ή τη διαχείριση ενός πληροφοριακού πόρου.
- Συχνά καλούνται «**δεδομένα για άλλα δεδομένα**» ή «πληροφορία για άλλη πληροφορία»

Εμφανίσεις Μεταδεδομένων

- Τρεις συνηθισμένες εμφανίσεις των μεταδεδομένων:
 - Συστήματα διαχείρισης βάσεων δεδομένων
 - Σχήματα των σχεσιακών βάσεων δεδομένων
 - Παραδοσιακοί κατάλογοι βιβλιοθήκης
 - Παγκόσμιος Ιστός

MARC και Μεταδεδομένα



MARC

← Καταλογογράφηση

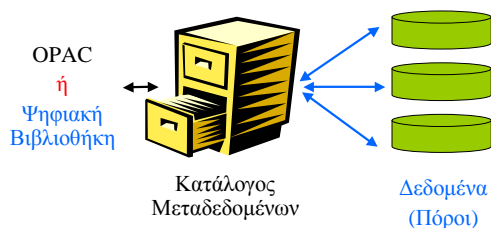
Τα μεταδεδομένα είναι για τους (ψηφιακούς) πόρους
ότι το MARC για τα παραδοσιακά τεκμήρια



→ Μεταδεδομένα

← Καταλογογράφηση

Σχέση: Μεταδεδομένα ↔ Δεδομένα



Χειρισμός Δεδομένων και Μεταδεδομένων

Διαφορές στο χειρισμό:

- **Δεδομένα** – τα αντικείμενα που αναζητούμε
 - Χωρίς αυτά, τα μεταδεδομένα δεν έχουν λόγο ύπαρξης
- **Μεταδεδομένα** – στοιχεία περιγραφής των δεδομένων
- Το λογισμικό της Ψηφιακής Βιβλιοθήκης στηρίζεται στη **διάκριση** Δεδομένων και Μεταδεδομένων
- Π.χ., σε μία Ψηφιακή Βιβλιοθήκη:
 - Δεδομένα = αντικείμενα: άρθρα, αναφορές, βίντεο, ...
 - Μεταδεδομένα: περιγραφές με: ΤΙΤΛΟΣ, ΘΕΜΑ, ...

Λειτουργίες Λογισμικού ΨΒ

- Ψ.Β.: δεδομένα + μεταδεδομένα (>0 επίπεδα)
- Λογισμικό: Αναζήτηση και παρουσίαση μεταδεδομένων + δεδομένων
- Λογισμικό Ψηφιακής Βιβλιοθήκης
 - Προσδιορισμοί στα μεταδεδομένα
 - Λειτουργίες στα δεδομένα
 - Π.χ. ταύτιση φράσεων / εικόνας / ήχου
 - Θεματική σύγκριση
 - Π.χ. διακοπές στη Χαβάη / ηλιοβασιλέμα / βάλς
- Τετριμμένη Ψ.Β.: μόνο μεταδεδομένα
 - Τα δεδομένα δεν είναι διαθέσιμα ψηφιακά

(Μετά)? Δεδομένα σε OPAC

- Τα ίδια αντικείμενα μπορεί να είναι δεδομένα ή μεταδεδομένα, ανάλογα με το σύστημα αναφοράς
- Οι εγγραφές στο OPAC με: ΤΙΤΛΟΣ, ΘΕΜΑ, ...
 - Με σύστημα αναφοράς την βιβλιοθήκη (ή την ψηφιακή βιβλιοθήκη)
 - Είναι μεταδεδομένα (τα δεδομένα είναι τα βιβλία, ...)
 - Με σύστημα αναφοράς το OPAC
 - Είναι δεδομένα (αλλά μπορεί να τα καλούμε «μεταδεδομένα» μιας τετριμμένης ψηφιακής βιβλιοθήκης)

Σπουδαιότητα Μεταδεδομένων

- Σε μια παραδοσιακή βιβλιοθήκη
 - Παράγονται – αναγνωρίζονται (π.χ. τίτλος, συγγραφέας)
 - Συγκεντρώνονται (στον κατάλογο, ή στα ράφια ανά ομάδες)
 - Χρησιμοποιούνται στην αναζήτηση – από τον κατάλογο αλλά και από τα βιβλία
- Σε ηλεκτρονικές πηγές (π.χ. Διαδίκτυο)
 - Είναι ο τρόπος για να βρούμε την πληροφορία!
 - Όπως και τυχόν τεχνικές πάνω στο πλήρες περιεχόμενο.

Ανταλλαγή (Μετά)? Δεδομένων

Τα δεδομένα (και τα μεταδεδομένα) τα

- Χρησιμοποιούμε (εσωτερικά στην εφαρμογή)
- Αποθηκεύουμε (σε αρχεία)
- Μετατρέπουμε, σε άλλες μορφές [εύκολα;]
- **Μεταφέρουμε – Δίνουμε – Ανταλλάσσουμε**
 - Με αποθήκευση σε αρχεία (και φύλαξη για μετά)
 - Μέσω πρωτοκόλλων (μεταξύ εφαρμογών)

Ανταλλαγή και Πρότυπα

Κάθε ανταλλαγή χρειάζεται πρότυπα:

- Για εικόνα, ήχο, βίντεο, κλπ
- Για γραφήματα, πίνακες
- Για μορφοποιημένα κείμενα
- Για κωδικοποίηση γραμμάτων / γραμματοσειρές
- Για βάσεις δεδομένων
- Για **μεταδεδομένα**
- Για πρωτόκολλα επικοινωνίας
- ...

Πρότυπα και Συστάσεις

Για να είναι τα μεταδεδομένα χρήσιμα και επωφελή είναι ουσιαστικό

- Η Σύνταξη, Σημασιολογία και Δομή να **συμμορφώνονται** με τα πρότυπα
- Να απεικονίζουν την **ουσία** των πηγών (την πληροφορία που έχουν να δώσουν)

Πρώιμη έκφραση των προτύπων, για χρήση και κριτική, είναι οι **συστάσεις** – RFC (Request For Comment), που χρησιμοποιούνται παρόμοια

Αναγκαιότητα Προτύπων Μεταδεδομένων

- Χρειάζονται για να έχουμε κοινή αντίληψη των δεδομένων που περιγράφονται με αυτά
- Μας προσφέρουν μεγαλύτερη **δομή**
- ⇒ Μας περιορίζουν στην **ευελιξία**
- ⇒ Επεξεργάζονται ευκολότερα **μηχανικά**
- Είναι αναγκαιότερα σε **ψηφιακά** αντικείμενα
 - Αφού στα συμβατικά έχουμε αισθητήρια αντίληψη

Πόσα Πρότυπα Μεταδεδομένων

- Χρειάζονται **κοινά** καταναμημένα μοντέλα μεταδεδομένων
- Δεν υπάρχει **ένα** μοναδικό διεθνές πρότυπο για μεταδεδομένα, γιατί:
 - Χρειαζόμαστε διαφορετικά **επίπεδα πολυπλοκότητας**, από πλούσιες μέχρι απλές περιγραφές
 - **Υπάρχουν** κάμποσα σχήματα μεταδεδομένων, για διαφορετικά επίπεδα και απαιτήσεις
 - **Επεκτείνουμε** τα υπάρχοντα πρότυπα
- Πότε πρέπει να αναζητούμε **άλλο** πρότυπο – ή σύσταση;

Προβληματισμοί Επιλογής Μεταδεδομένων

- Τα σωστά μεταδεδομένα εξαρτώνται από:
 - Θέμα και διαλειτουργικότητα
 - Π.χ. GIS, βιβλιοθήκη
 - Τύπο υλικού
 - Π.χ. βιβλίο, CD, WWW, βίντεο
 - Απαιτήσεις αναζήτησης – επένδυση καταλογογράφησης
 - Καθιερωμένα πρότυπα

Παρανοήσεις για Μεταδεδομένα

- **Τα μεταδεδομένα έχουν αυστηρό συντακτικό;**
- Τα μεταδεδομένα έχουν αυστηρή **σημασιολογία**
- Και πιθανώς προτεινόμενο ή υποχρεωτικό συντακτικό, π.χ.
 - MARC: ISO2709, αλλά και MARC-XML ή ASCII
 - DC: RDF/XML (2 χρόνια μετά ... τον ορισμό του DC)

Ο Μύθος της Κατηγοριοποίησης

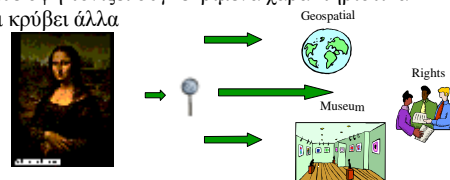
...there is no classification of the universe that is not fictional and conjectural.

...δεν υπάρχει κατηγοριοποίηση του σύμπαντος που να μην είναι φανταστική και αυθαίρετη

– Jorge Luis Borges

Φακοί και Όψεις

- Όλες οι κατηγοριοποιήσεις παρέχουν και **πρέπει** να παρέχουν μια μεροληπτική όψη της πραγματικότητας
- Κάθε όψη τονίζει συγκεκριμένα χαρακτηριστικά και κρύβει άλλα



Σχήμα (Schema)

Σχήμα έχουν:

- Βάσεις δεδομένων, XML, RDF, μεταδεδομένα

Σχήμα είναι ο ορισμός της δομής των δεδομένων που περιγράφουν, δηλαδή

- Των «πεδίων» που χρησιμοποιούνται
- Των τύπων των τιμών τους
- Των κανόνων και περιορισμών που ισχύουν

Σχήματα Μεταδεδομένων

- Τα σχήματα μεταδεδομένων είναι σύνολα στοιχείων μεταδεδομένων, σχεδιασμένα για ένα ειδικό σκοπό, όπως περιγραφή συγκεκριμένου τύπου πληροφοριακών πόρων.
- Ο ορισμός ή σημασία των στοιχείων καλείται «σημασιολογία του σχήματος».
- Οι τιμές που δίνονται στα στοιχεία των μεταδεδομένων είναι το περιεχόμενο.

Συστήματα Οργάνωσης Γνώσης στο Ψηφιακό Περιβάλλον

- Τα σχήματα οργάνωσης πληροφορίας και διευκόλυνσης διαχείρισης γνώσης μπορούν να βελτιώσουν την οργάνωση των ψηφιακών βιβλιοθηκών και να διευκολύνουν την πρόσβαση στο περιεχόμενό τους
- Λειτουργούν ως γέφυρες μεταξύ των αναγκών πληροφόρησης των χρηστών και του υλικού μιας συλλογής αλλά ποικίλουν ως προς την περιπλοκότητα, τη δομή και τη λειτουργία
 - Λίστες καθιερωμένων όρων (π.χ. λεξικά)
 - Σχήματα ταξινόμησης (π.χ. LCSH)
 - Λίστες όρων και σχέσεων τους (π.χ. θησαυροί)

Εισαγωγή στα Μεταδεδομένα

Περιεχόμενο Μεταδεδομένων (Ποιοτικά)

Τα μεταδεδομένα:

- Είναι **συμπυκνωμένη** πληροφορία
- **Προσαρμοσμένη** στις συνήθειες αναζήτησης
- Που **δεν εξάγεται** από τα δεδομένα
- Ή είναι **δαπανηρό** να εξαχθεί

Αλλά πληροφορίες που δεν συμπυκνώνονται πολύ, εξάγονται πάντα από τα δεδομένα: π.χ. για ταύτιση προσώπου

Περιεχόμενο Μεταδεδομένων (Αναλυτικά)

- Περιγραφή περιεχομένου
- Τεχνικά στοιχεία αναπαράστασης
- Εσωτερική δομή αντικείμενου
- Σχέσεις με άλλα αντικείμενα
- Πληρότητα / αυθεντικότητα αντικείμενου
- Τρόποι χρήσης – δικαιώματα
- Στοιχεία μη τετριμμένης ανεύρεσης / Ειδικών
- Πληροφορίες συντήρησης
- Διαχειριστική πληροφορία

Μεταδεδομένα Ψηφιακής Βιβλιοθήκης

- Περιγραφικά: αναζήτηση-εύρεση
- Δομικά: αποθήκευση-παρουσίαση
- Διοικητικά: διαχείριση-συντήρηση
- Όροι & Προϋποθέσεις
- *Συγγραφέας, τίτλος, λέξεις-κλειδιά, έκδοση, θέματα*
- *Συστατικά μέρη, σχέσεις συστατικών, πλοήγηση, είδη αντικειμένων, κλπ.*
- *Ημερομηνία δημιουργίας, τελευταία τροποποίηση, υπεύθυνος, οδηγίες χειρισμού*
- *Δικαιώματα χρήσης, πρόσβασης, εκτύπωσης, αντιγραφής, επαναχρησιμοποίησης, αναδημοσίευσης; κόστος*

Στόχοι Χρήσης Μεταδεδομένων

- Ανακάλυψη πόρων
- Διαχείριση αντικειμένων
- Διαχείριση δικαιωμάτων
- Αξιολόγηση περιεχομένου
- Ταξινόμηση του υλικού
- Ασφάλεια και αυθεντικοποίηση χρηστών
- Κατάσταση αρχειοθέτησης
- Προέλευση πόρων
- Δομή και ολοκλήρωση πόρων

Στόχοι Χρήσης Μεταδεδομένων (Πόρων)

- Οργάνωση και συσχέτιση (μη δομημένων, σύνθετων και ενδεχόμενα μεγάλων) αρχείων
- Περιγραφή και διαχείριση πνευματικών δικαιωμάτων
- Προσδιορισμός εκδόσεων
- Πιστοποίηση αυθεντικότητας
- Περιγραφή κατάστασης
- Γεωγραφική τοποθέτηση
- Έλεγχος και περιγραφή διαδικασιών
- Απεικόνιση δυναμικής πληροφορίας

Δημιουργία Μεταδεδομένων Σήμερα

- «Δημιουργία μεταδεδομένων» = η τέχνη γνωστή από παλιά σαν «καταλογογράφηση»
- Η ψηφιακή εποχή χαρακτηρίζεται από
 - Παροχή των πόρων από τους δημιουργούς / ιδιοκτήτες
 - Και όχι από (ή συγχρόνως με) μεσάζοντες
 - Απομονωμένη πρόσβαση σε όλους τους πόρους από όλους
 - (εν δυνάμει...)
 - Έμφαση στον χρήστη – πελάτη
 - Υπερπληθώρα πληροφοριών
 - Ποσότητα αντί για ποιότητα
 - το φαινόμενο του Google (νέα προέλευση μεταδεδομένων)

Διαδικασίες Παραγωγής Μεταδεδομένων (1)

- Από εξειδικευμένο προσωπικό
 - Από δημιουργούς / ιδιοκτήτες του πόρου
 - Από υπηρεσίες τρίτων (π.χ. βιβλιοθηκονόμους)
- Από ειδική εφαρμογή επεξεργασίας των δεδομένων (παράγωγα μεταδεδομένα μόνο)
 - Μηχανές ευρετηρίασης
 - Από το περιεχόμενο του πόρου, και δεδομένα εξυπηρετητών
 - Αυτόματα ή ημιαυτόματα

Διαδικασίες Παραγωγής Μεταδεδομένων (2)

- Όλα τα παραπάνω
 - Ανάλογα με το πεδίο
 - ή σε συνδυασμό σε κάθε πεδίο
- Η δημιουργία (και διατήρηση) καλής ποιότητας μεταδεδομένων δεν είναι φθηνή
 - Προκύπτουν θέματα δικαιωμάτων για τα μεταδεδομένα, όπως και για τους ίδιους του πόρους

Αποθήκευση Μεταδεδομένων

- Δημιουργούμενα σαν εγγραφές «βάσης δεδομένων»
 - Πιθανώς απομακρυσμένη (remote) βάση δεδομένων
- Συνδεδεμένα με τους πόρους
- Εμφυτευμένα στους πόρους
 - Εξαρτώνται από την μορφή του πόρου
 - Μπορούν τα μεταδεδομένα να εξαχθούν από τον πόρο;
- Υιοθετείται η προσέγγιση που δίνει τη μεγαλύτερη ευελιξία
 - Μπορεί να χρειάζεται να «εμφανίζει» διαφορετικά υποσύνολα σαν «τα πλήρη μεταδεδομένα» σε διάφορα περιβάλλοντα

Διαφορετικές Προσδοκίες από τα Μεταδεδομένα (1)

- Διαφορετικές «παραλλαγές» των μεταδεδομένων εξυπηρετούν διαφορετικούς σκοπούς
 - απλά, γενικά ή πλούσια, ειδικά
 - Δημοσιευμένα ευρέως ή χρησιμοποιούμενα σε μια κοινωνία ή χρησιμοποιούμενα τους δημιουργούς / ιδιοκτήτες του πόρου
- Ο Ιδιοκτήτης / διαχειριστής / παροχέας θέλει να
 - Καθιερώσει έλεγχο των πόρων
 - Διαχειριστεί πόρους (καθώς περνά ο καιρός)
 - Εκθέτει ή προάγει πόρους
 - Να επιτρέπει και να ελέγχει την πρόσβαση / χρήση των πόρων
 - Να εξάγει συμφραζόμενα από τους πόρους

Διαφορετικές Προσδοκίες από τα Μεταδεδομένα (2)

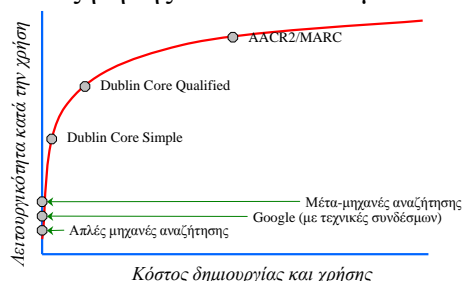
- Ο τελικός χρήστης θέλει να
 - Εντοπίσει
 - Αναγνωρίσει
 - Επιλέξει
 - Λάβει και χρησιμοποιήσει
 - Ερμηνεύσει
- Οι υπηρεσίες τρίτων (Third party service) μπορεί να θέλουν
 - Να Εκθέτουν ή προάγουν
 - Να επιτρέπουν και να ελέγχουν την πρόσβαση / χρήση
 - Να υπομηματίζουν
 - Να επανα-εξάγουν συμφραζόμενα

Συγκρίσεις και Τάσεις στα Μεταδεδομένα

Προκλήσεις και Τάσεις στα Μεταδεδομένα

- «Απλού και ευρύ» και «πλούσιου αλλά συγκεκριμένου» (απλότητα ↔ λειτουργικότητα)
- «Γενικού» και «προσαρμοσμένου» (διαλειτουργικότητα ↔ επεκτασιμότητα)
- «Αναγνώσιμου από ανθρώπους» και «επεξεργάσιμου από μηχανές» (σε δημιουργία και χρήση)
- «Κατανομή» και «Συγκέντρωση» σε ένα μέρος
- Αυτόματη παραγωγή ή με ανθρώπινη παρέμβαση

Κόστος και Λειτουργικότητα Αναζήτησης και Μεταδεδομένων



Τι Δεν Είναι Καλό στο Μοντέλο του MARC

- Ακριβό
 - Περίπλοκο (ακόμα και για τον αρχικό του στόχο;)
 - Επαγγελματική μεσολάβηση (υποθέτει ότι μία μόνο επαγγελματική κοινότητα έχει την απαιτούμενη γνώση)
- Μονολιθικό
 - Ένα μέγεθος ταιριάζει για όλα
 - Απεικονίζει την κεντροποιημένη του προέλευση
- Προκαταλαμβάνει προς φυσικά τεχνουργήματα
 - Περιγράφει προκαθορισμένους πόρους
 - Ελλιπής κάλυψη της εξέλιξης πόρων και άλλων σχέσεων των πόρων
- Άγγλο-κεντρικό

Μειονεκτήματα του MARC

- Μη σχεσιακό, δόμηση με περιορισμούς
- Δύσκολοι χαρακτηρισμοί συχνά καταγράφονται σε λάθος θέση
- Επεκτείνεται, αλλά τοπικά, περιορισμένα και αδόμητα
- Χρειάζεται αλλαγές στα δεδομένα για να παρακολουθεί την εξέλιξη
 - Π.χ. CDROM -> μεγάλο ή μικρό!
- Κοστίζει \$50 / δημιουργία πλήρης εγγραφής

Dublin Core

*Επειδή είναι
απλό...
... αλλά και
αποτελεσματικό
....*



Χαρακτηριστικά του «Dublin Core»

- «Dublin Core» σημαίνει «Dublin, Ohio» / OCLC
- Πρωτοβουλία για να βελτιώσει την ανακάλυψη πόρων στο Διαδίκτυο
- Κοινός παρονομαστής για επικοινωνία / διαλειτουργικότητα
- Ένασμα για σύγκλιση των προτύπων
- Επεκτάσιμο, για να καλύψει τις επιπρόσθετες ανάγκες ανακάλυψης πόρων των διαφορετικών εφαρμογών / περιοχών

Στόχοι του Dublin Core

- Απλότητα δημιουργίας και διατήρησης
 - Μη ειδικοί να δημιουργούν περιγραφικές εγγραφές για αποτελεσματική ανάκτηση σε δικτυωμένο περιβάλλον
- Κοινά κατανοητή (διαθεματική) σημασιολογία
 - Σύγκλιση κοινών, περισσότερο γενικών στοιχείων
 - Αυξημένη ορατότητα και προσβασιμότητα
 - Κατάλληλο και για τον μη ειδικό της αναζητητής
 - Τον «ψηφιακό τουρίστα»

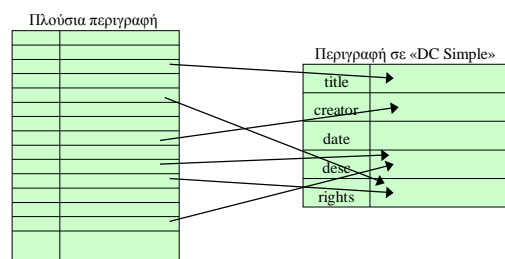
Το Πρότυπο Dublin Core

- Προτυποποίηση
 - ISO 15836-2003
 - US: NISO Z39.85-2001
 - Ευρώπη: αναγνώριση από το CEN/ISSS Workshop Agreement 13874-2000

Χρήση του Dublin Core

- Είναι βασικός πυρήνας στοιχείων
- Δεν είναι υποκατάστατο σε πλουσιότερα περιγραφικά πρότυπα
- Παρέχει 15 «παράθυρα» από πλουσιότερη περιγραφή πόρων
 - Φαερώνει πλούσιες περιγραφές σε απλή μορφή
 - Σημασιολογικά σταυροδρόμια, αντιστοιχίσεις σε υπάρχοντα δεδομένα

Περιγραφή σε Dublin Core



Το Dublin Core είναι περισσότερο αντικείμενο εξαγωγής – παρά δημιουργίας

Συναίνεση και Dublin Core

- Δημιουργήθηκε σαν συναίνεση από μία διεθνή, διαθεματική ομάδα που αντιπροσωπεύει
 - Βιβλιοθήκες και Πληροφόρηση
 - Πληροφορική
 - Κωδικοποίηση κειμένου (TEI)
 - Μουσεία
 - Σχετικούς επιστημονικούς τομείς
- Τα στοιχεία του αναπαριστούν μια ευρεία διαθεματική συναίνεση
- Διεθνείς συναντήσεις, ημερίδες, κλπ για ανάπτυξη συναίνεσης

Dublin Core και Διεθνοποίηση

- Διεθνούς εμβέλειας
 - 26+ Γλώσσες
 - Συντονισμένες προσπάθειες
 - XML / RDF - WWW
- Τεχνικές προκλήσεις διεθνοποίησης
 - Πολυ-γλωσσική και Πολυ-πολιτισμική φύση του σύμπαντος της ηλεκτρονικής πληροφορίας

Χρήση και Κατηγορίες Στοιχείων Dublin Core

- Το Dublin Core έχει 15 στοιχεία. Κάθε ένα από αυτά είναι προαιρετικό και επαναλαμβανόμενο
- Τα 15 στοιχεία χωρίζονται σε 3 κατηγορίες:
 - Περιεχόμενο:
 - Περιγράφουν το αντικείμενο
 - Πνευματική Ιδιοκτησία:
 - Περιγράφουν το copyright και τη δημιουργία
 - Στιγμιότυπο:
 - Περιγράφουν την εισαγωγή και διαχείριση

Τα 15 Στοιχεία του Dublin Core

- Τίτλος / **Title**
- Θέμα / **Subject**
- Περιγραφή / **Description**
- Πηγή / **Source**
- Γλώσσα / **Language**
- Σχέση / **Relation**
- Κάλυψη / **Coverage**
- Δημιουργός / **Creator**
- Εκδότης / **Publisher**
- Συντελεστής / **Contributor**
- Δικαιώματα / **Rights**
- Ημερομηνία / **Date**
- Τύπος / **Type**
- Μορφότυπο / **Format**
- Αναγνωριστικό / **Identifier**

Εξειδίκευση Στοιχείων στο Dublin Core

- Το απλό Dublin Core, «Simple DC»
- Το εξειδικευτικό Dublin Core, «Qualified DC»
 - Μερικά στοιχεία έχουν ένα περιορισμένο σύνολο από καθιερωμένες εξειδικεύσεις και ιδιότητες
- Κάθε στοιχείο μπορεί να έχει άτυπες εξειδικεύσεις
 - Κάνει την σημασία ενός στοιχείου στενότερη ή πιο ειδική
 - Δεν την τροποποιεί ή επεκτείνει, σε καμία περίπτωση
 - Π.χ. TITLE.TRANSLATED
- Αν δεν καταλαβαίνουμε κάποιο εξειδικευτή, μπορούμε με ασφάλεια να τον αγνοήσουμε!

Συσχέτιση Dublin Core – UNIMARC

Π.χ. ... DC:Creator ...

700 \$a Personal Name - Primary Intellectual Responsibility, ή αν πάνω από ένας:

701 \$a Personal Name - Alternative Intellectual Responsibility

710 \$a Corporate Body Name - Primary Intellectual Responsibility, ή αν πάνω από ένας:

711 \$a Corporate Body Name - Alternative Intellectual Responsibility

USMARC + Dublin Core

http://www.dlib.vt.edu/~paul/ndltd_metadata.htm

Παραδείγματα:

Title	245\$a	
Title.X-Notes	500	Πηγή σημειώσεων τίτλου
Creator.PersonalName	100\$a	Author, personal
Creator.Address	-	
Creator.X-Institution	710\$a	Corporate Name
Description.Abstract	520\$a	Περιγραφή ή θέση (URL)
Format	856\$q	Transfer mode
Rights	540\$a / 856\$su	Terms governing use / URL

Παρανοήσεις για Μεταδεδομένα

- *Η πιο λεπτομερής περιγραφή μεταδεδομένων είναι πάντα καλύτερη;*
- Συνήθως ισχύει για συμβατικά τεκμήρια
 - Που δεν διαθέτουμε **άλλη πληροφορία** για αυτά
- Αλλά θα γίνουν και περισσότερα **λάθη** στην περιγραφή – θέλει ειδική **εκπαίδευση** και μεγάλο **κόστος** επένδυσης
- Δυσκολεύει τη **διαλειτουργικότητα**

Οι Χαμένες Λεπτομέρειες

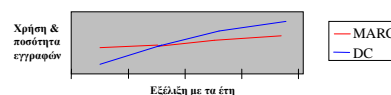
- Το Dublin Core είναι νεότερο πρότυπο από το MARC
- Το Dublin Core έχει μεγαλύτερη λεπτομέρεια όπου χρειάζεται – σε ελάχιστα πεδία
- Το Dublin Core έχει λιγότερη λεπτομέρεια στα πεδία που
 - Δεν έχουν πια πολλή χρησιμότητα / χρήση
 - Είναι ακριβό να συντηρούνται
 - Δεν αντιμετωπίζονται ενιαία από όλους

MARC ή Dublin Core

Κριτήριο	MARC	Dublin Core
Κόστος καταλογογράφησης	Μεγάλο	Κλιμακωτό
Ακρίβεια περιγραφής	Μεγάλη	Κλιμακωτή
Εξέλιξη	Απαραίτητη	Προαιρετική
Διαλειτουργικότητα	Η μικρότερη	Κλιμακωτή
Χρόνος / ακρίβεια εκμάθησης	Μεγάλος / χαμηλή	Μικρός / υψηλή

Παραλληλισμός MARC – Dublin Core

- Παράδειγμα Επέκτασης
 - Ελληνικά και ξένα διδακτορικά:
 - Π.χ. Ημερομηνία απονομής (ή αναγνώρισης).
 - Σε MARC: \$9xx [Δεν είναι προβλέψιμο / μοναδικό!]
 - Σε DC: Date.award
- Τι είναι καλύτερο – τι προοπτικές έχουν;



Επιλογή Μεταδεδομένων

Προσαρμογή Μεταδεδομένων

- Τα μεταδεδομένα **προσαρμόζονται** σε πολλαπλές παραλλαγές / ποικιλίες
 - Αποτυπώνοντας την **ανομοιομορφία** των δημιουργών και διατηρητών μεταδεδομένων
 - Προσφέροντας για κάθε **κοινότητα** (συγκεκριμένο χώρο) εξειδικευμένη λειτουργικότητα, δημιουργία, διαχείριση, πρόσβαση, ...
- Τηρώντας τα αντίστοιχα **πρότυπα** ...
- Μεγαλύτερη εξειδίκευση = μεγαλύτερη λειτουργικότητα = λιγότερη διαλειτουργικότητα!

Dublin Core – Επεκτασιμότητα

- Η αρχιτεκτονική του Dublin Core υποστηρίζει πιο εξελιγμένες λύσεις μεταδεδομένων
 - Αλλά προσοχή: προτιμάμε επέκταση του DC ή επιλογή άλλων πλουσιότερων σχημάτων;
- Έχουμε επεκτάσεις του Dublin Core σε εξειδικευμένους τομείς (domain-specific)
- Ομάδες εργασίας (του DCMJ) αναπτύσσουν *Προφίλ Εφαρμογών (application profiles)* για ειδικούς τομείς

Τι είναι το Προφίλ Εφαρμογής

- Ένα σχήμα μεταδεδομένων που ενσωματώνει ένα **σύνολο στοιχείων** από ένα ή περισσότερα **σύνολα στοιχείων μεταδεδομένων** (ή λεξιλογίων ή χώρους ονομάτων – namespaces)
- Ένα σύνολο από **πολιτικές** που ορίζουν πώς τα στοιχεία πρέπει να εφαρμόζονται στο πεδίο της εφαρμογής
- Ένα σύνολο **οδηγιών** που ξεκαθαρίζουν τις πολιτικές που αφορούν τα στοιχεία

Λεξιλόγια και Προφίλ Εφαρμογών

- Τα **Λεξιλόγια δηλώνουν** όρους και ορισμούς
 - Λεξιλόγιο Dublin Core = Πρότυπο Dublin Core
- Τα **Προφίλ Εφαρμογών επανα-χρησιμοποιούν** όρους από ένα ή περισσότερα λεξιλόγια
 - Μπορεί να συνδυάζει όρους από πολλαπλά λεξιλόγια
 - Μπορεί να προσαρμόζει (εξειδικεύει) ορισμούς για τοπικούς σκοπούς ή εξειδικευμένους τομείς (domain-specific)
 - Όλοι οι όροι πρέπει να ορίζονται στα λεξιλόγια
 - Μπορεί να περιλαμβάνει τοπικά ορισμένα λεξιλόγια

Χρησιμοποιώντας Dublin Core με Άλλα Λεξιλόγια

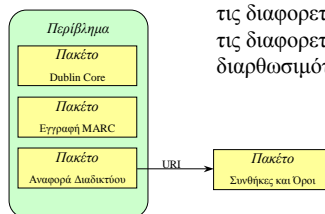
Εξειδικευμένα *προφίλ εφαρμογών* μπορεί να:

- Χρησιμοποιούν τα **γενικά στοιχεία** του Dublin Core
- Χρησιμοποιούν στοιχεία από άλλο πρότυπο, πιο **ειδικό στον τομέα εφαρμογής**
- Να **εξειδικεύσουν (περιορίσουν) τους καθιερωμένους ορισμούς** των στοιχείων του Dublin Core για ειδικές τοπικές χρήσεις
- Να **εφεύρουν και ορίσουν τοπικά στοιχεία** έξω από το πλαίσιο των υφιστάμενων προτύπων

Απόσπασμα Πολλαπλού Λεξιλογίου

```
xmlns:DC="http://purl.org/dc/elements/1.1/"
xmlns:co="http://purl.org/rss/1.0/modules/company/"
<DC:publisher>The O'Reilly Network</DC:publisher>
<DC:creator>Rael Dornfest</DC:creator>
<DC:rights>Copyright ©1999; 2000 O'Reilly & Associates, Inc.</DC:rights>
<DC:date>2000-01-01T12:00+00:00</DC:date>
<DC:description>XML is placing increasingly heavy loads on the existing technical infrastructure of the Internet.</DC:description>
<co:name>XML.com</co:name>
<co:market>NASDAQ</co:market>
<co:symbol>XML</co:symbol>
```

Η Ιδέα του Πλαισίου Warwick (1996)



- Πακέτα μεταδεδομένων για τις διαφορετικές χρήσεις από τις διαφορετικές κοινότητες – διαρθρωσιμότητα!

Παρανοήσεις για Dublin Core (1)

- *Το Dublin Core είναι άλλο ένα πρότυπο περιγραφής ή ανταλλαγής μεταδεδομένων, και πρέπει να επιλέξουμε αν θα το προτιμήσουμε;*
- Το Dublin Core είναι η **σημσιολογική τομή** των (κυριότερων) **προτύπων** (που χρησιμοποιούνται σήμερα)
- Μπορεί να χρησιμοποιηθεί ως βάση στοιχείων σε «**προφίλ εφαρμογών**», για ευκολότερη (προσχεδιασμένη) διαλειτουργικότητα

Παρανοήσεις για Dublin Core (2)

- Το Dublin Core δεν είναι μορφότυπο **ανταλλαγής** μεταδεδομένων (αποθηκευμένα π.χ. σε MARC)
- Μπορούμε να απεικονίσουμε («μετατρέψουμε») πολλά πρότυπα σε Dublin Core (**export DC**)
- Δεν μας ενδιαφέρει να «μετατρέψουμε» από Dublin Core πίσω σε άλλα πρότυπα (**import DC**)
- **Δεν καταργούμε** τις αρχικές (λεπτομερείς) εγγραφές – δημιουργούμε **καινούργιες** (συνοπτικές) για άμεση χρήση
- Το Dublin Core είναι σχεδιασμένο για εύκολη / **προφανή** απεικόνιση (άλλων προτύπων) μεταδεδομένων σε αυτό
- Το Dublin Core έχει περισσότερο σχέση με τη **διαλειτουργικότητα** παρά με τα μεταδεδομένα

Και το Z39.50

- **Δεν είναι** πρότυπο περιγραφής και ανταλλαγής δεδομένων ή μεταδεδομένων
- Είναι **πρωτόκολλο** που περιλαμβάνει:
 - Διαπραγμάτευση δυνατοτήτων
 - Συμφωνία σε προφίλ δεδομένων (π.χ. **BIB-1**)
 - Διαδικασίες επικοινωνίας
 - Διατύπωση ερωτήσεων
 - Παρουσίαση αποτελεσμάτων
- Σημαντικό για διαλειτουργικότητα

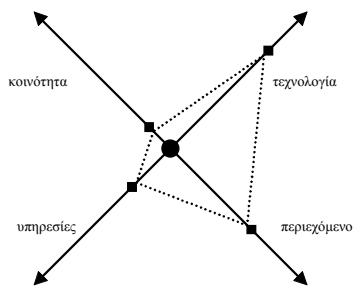
Προοπτικές Ψηφιακών Βιβλιοθηκών

Εξέλιξη, τάσεις, χρεώσεις

Επισκόπηση Θεμάτων

- Σχεδίαση Ψηφιακών Βιβλιοθηκών
- Η αλλαγή με το χρόνο στις Τεχνολογίες
- Διαφοροποίηση Ψηφιακών Βιβλιοθηκών
- Εξέλιξη στις Ψηφιακές Βιβλιοθήκες
- Πρόσβαση – χρεώσεις Ψηφιακών Βιβλιοθηκών

Χώρος Σχεδίασης Ψηφιακών Βιβλιοθηκών



Ψυχολογία και Ψηφιακή Βιβλιοθήκη

- Ανθρώπινοι Παράγοντες / Γνωστική Εργονομία
- Οπτική Επεξεργασία Πληροφορίας
- Ψυχολογία του Υπερκειμένου
- Κοινωνική Ψυχολογία

Δημιουργία Ψηφιακής Βιβλιοθήκης

- Προσδιορισμός της ομάδας χρηστών
- Προσδιορισμός των πηγών που περιέχουν τις πληροφορίες που μας ενδιαφέρουν
- Διαπραγμάτευση όρων και συνθηκών με τους εκδότες
- Δημιουργία ευρετηρίου
- Υπηρεσίες, όπως «αναζήτηση» (Search) και «φυλλομέτρηση» (Browse)

Δημιουργία Ψηφιακής Βιβλιοθήκης

- Δημιουργία υπηρεσιών αλληλεπίδρασης με το χρήστη
 - Όροι και συνθήκες
 - Εξακρίβωση ταυτότητας (authentication)
 - Πληρωμή (billing)
 - Εμφάνιση

Τεχνολογικές Εξελίξεις

- Υπολογιστική ισχύ
 - Επεξεργαστής και μνήμη – RAM
- Μέσα αποθήκευσης
 - Σκληροί δίσκοι, DAT, CD-ROM, DVD
- Επικοινωνίες και δίκτυα
 - Καλώδια, οπτικές ίνες, συναφείς τεχνολογίες
- Ψηφιακό περιεχόμενο
 - Γεννημένο ψηφιακά, ψηφιοποιημένο
- Τεχνικές ψηφιοποίησης

Τεχνολογικές Απαιτήσεις Ψηφιακών Βιβλιοθηκών

- Υποδομή πρόσβασης
 - Δίκτυα υψηλής ταχύτητας, μαζική αποθήκευση, επεξεργασία
 - Πρόσβαση από παντού (γραφείο, σπίτι, αυτοκίνητο)
- Οργάνωση πρόσβασης στο περιεχόμενο
 - Ευρετηρίαση και αναζήτηση
 - Διάθεση επισκοπήσεων και «πρεμιέρον»
- Διεπαφές του χρήστη
 - Γραφικές διεπαφές
 - Πολλαπλοί τρόποι, κινητοί
- Τεχνολογία λογισμικού
 - Γρήγορη κατασκευή πρωτοτύπων, επαναληπτικός σχεδιασμός
 - Διαλειτουργικότητα και ομόσπονδες αρχιτεκτονικές

Τεχνολογία Ψηφιακών Βιβλιοθηκών

- Τεχνολογίες καταλύτες
 - Γραφικό περιβάλλον (GUI)
 - Παγκόσμιος Ιστός (WWW)
 - Προγράμματα πλοήγησης (browsers)
 - Ραγδαία αύξηση ψηφιακού περιεχομένου
- Σημερινά επιτεύγματα της τεχνολογίας
 - Υποδομή Διαδικτύου – Παγκόσμιου Ιστού
 - Αρκετά προχωρημένη έρευνα
 - Εξελιγμένα πρωτότυπα Ψηφιακών Βιβλιοθηκών
 - Εργαλεία λογισμικού

Ρυθμός Αλλαγής

- Διαρκής μεταβατική περίοδος (ταχύτατες αλλαγές - ρευστότητα κατάστασης - ανίχνευση τάσεων)
- Η τεχνολογία αλλάζει γρήγορα
 - π.χ., ο νόμος του Moore: 2πλασιασμός ισχύος /18μήνες
- Η σημερινή αλλαγή της τεχνολογίας της πληροφορίας δεν είναι άνευ προηγουμένου
 - Π.χ. 1844-1874: 1Gm καλωδίων, 20000 πόλεις, 150Mm υποβρύχιων καλωδίων ...

Εξέλιξη Συνθηκών στην Επιστήμη της Πληροφορίας

- Αλλαγή μορφής βιβλιοθήκης (τερματικοί σταθμοί- ράφια)
- Διαφοροποίηση υποδείγματος λειτουργίας
 - Πρόσβαση έναντι πρόσκτησης (Access vs acquisition)
 - Βάσεις δεδομένων – μεταδεδομένα
- Προσαρμογή δεξιοτήτων βιβλιοθηκονόμων
 - Αναζήτηση από online πηγές: cybrarian, data miners
- Αλλαγή ρόλων: εκδότες –(βιβλιοθήκες)– χρήστες
- Ανάγκη αξιοποίησης νέων δυνατοτήτων

Αλλαγή Ψηφιακών Βιβλιοθηκών

- Οι άνθρωποι **ευαισθητοποιούνται** για την διανοητική ενέργεια και το χρόνο τους
- Η τεχνολογία αλλάζει γρήγορα αλλά οι άνθρωποι και οι οργανισμοί αλλάζουν **αργά**
- Αλλά απαιτείται και **οργανωτική αλλαγή**
- Βασικές Λειτουργίες Βιβλιοθήκης
 - Ανάπτυξη / Ενημέρωση / Συντήρηση συλλογών
 - Πρόσβαση (καταλόγων, παραπομπών)
 - Διαχείριση (διαδικασιών, πόρων)
 - Αλλά και **αξιοποίηση των ανθρώπων**

Αποθήκευση

(cpi = χαρακτήρες ανά κυβική ίντσα)

- 3000 πX Πλάκες αργίλου 1 cpi
- 1450 μX Τυπωμένη σελίδα 500 cpi
- 1990+ Οπτικοί δίσκοι 125,000,000,000 cpi
- Σήμερα 1000 βιβλία 300 σελίδων ανά CD-ROM

Υπολογισμοί

(ips = εντολές το δευτερόλεπτο)

- 5000 πX Άβακας 2-4 ips
- 1945 μX Υπολογιστής 100's ips
- 1960+ Υπολογιστής 100,000's ips
- 1970+ Υπολογιστής (MFLOPS) 1,000,000's ips
- 1980+ Υπολογιστής 10,000,000's ips
- 1990+ Υπολογιστής (GFLOPS) 1,000,000,000's ips

Μετάδοση Πληροφορίας

(wpm = λέξεις το λεπτό)

- 4000 πX Αγγελιοφόρος .01 wpm
- 1844 μX Τηλέγραφος 50-60 wpm
- 1980+ Καλώδιο/ίνα 1,000,000,000 wpm
- 1990+ Ίνα 100,000,000,000 wpm
- «Within a decade or so we will be able to send all human knowledge past your house in a few seconds»
 - Eric Sumner, 1990 (President IEEE)

Ανθρώπινη Επεξεργασία

(wpm = λέξεις το λεπτό, bpg = bits τη ματιά)

- 4000 πX Γραφή Γλώσσα 300 wpm
- Σήμερα Γραφή Γλώσσα 300 wpm
- 4000 πX Οπτικές Εικόνες 100,000,000 bpg
- Σήμερα Οπτικές Εικόνες 100,000,000 bpg
- 4000 πX Ομιλούμενη Γλώσσα 120 wpm
- Σήμερα Ομιλούμενη Γλώσσα 120 wpm

Βασικές Λειτουργίες Βιβλιοθήκης

- Ανάπτυξη συλλογών
 - Συντήρηση
 - Πρόσβαση
 - Καταλόγων
 - Αναφορών
 - Διαχείριση
 - Διαδικασιών
 - Πόρων
- Υπάρχουν **άνθρωποι** στις Ψηφιακές Βιβλιοθήκες

Πολυπλοκότητα

- Τα ανθρώπινα συστήματα (περιλαμβάνοντας και τις βιβλιοθήκες) είναι από τη φύση τους περίπλοκα
- Η οργάνωση και διαχείριση πολύπλοκων συστημάτων δεν είναι προφανής
- “Perhaps networks just at the phase transition, just poised between order and chaos, are best able to carry out ordered yet flexible behaviors.” Kauffman, p. 90
- Δεν είναι προβλέψιμες οι εκπλήξεις στα περίπλοκα συστήματα

Ποικιλία

- Ακραίες τάσεις
 - Μεγάλη διαφοροποίηση αναγκών και προδιαγραφών
 - Υπέρ-εξαιτομίκευση
 - Παγκόσμια πρόσβαση
 - Προτυποποίηση
 - Έχει απόδοση λόγω κλίμακας
 - Στοχεύει να ελαττώσει την ανομοιογένεια
- Όσο πιο κοντά στη βασική υποδομή, τόσο μεγαλύτερη ευκαιρία να ελαττωθεί η ποικιλία.
- Όσο πιο κοντά στον καθένα, τόσο μεγαλύτερη ευκαιρία να αυξηθεί η ποικιλία.

Σύγκριση

- Praktiker → Elephant
- Wall Mart → Nordstrom
 - Μεγάλη ποικιλία προϊόντων
 - Ευρύ κοινό
 - Μεγάλη ποσότητα
 - Τιμή και απόδοση
- Ταξιδιωτικός πράκτορας → Travelocity
 - Εξειδικευμένα προϊόντα
 - Περιορισμένο κοινό
 - Μικρή ποσότητα
 - Εξυπηρέτηση, ατμόσφαιρα

Σύγκριση Διάθεσης Βιβλίων

Βιβλιοπωλείο της γειτονιάς	Borders	Amazon
Εξειδικευμένα προϊόντα	Μεγαλύτερη ποικιλία προϊόντων	Μεγάλη ποικιλία προϊόντων
Περιορισμένο κοινό	Ευρύτερο κοινό	Ευρύ κοινό
Μικρή ποσότητα	Μεγαλύτερη ποσότητα	Μεγάλη ποσότητα
Εξυπηρέτηση, ατμόσφαιρα, διευκόλυνση	Τιμή, ατμόσφαιρα	Διευκόλυνση, τιμή, συστάσεις

Σύγκριση

- | | |
|--|---|
| <u>Ερευνητική Βιβλιοθήκη</u> | <u>Ψηφιακή Βιβλιοθήκη</u> |
| <ul style="list-style-type: none"> • Ευρεία γκάμα προϊόντων (περικεκτικότητα) • Περιορισμένο κοινό (γεωγραφικά) • Μεγάλος όγκος • Υπηρεσία αναζήτησης (αλλά όχι 24/7), βάθος | <ul style="list-style-type: none"> • Στενή γκάμα προϊόντων (ως τώρα) • Ευρύ κοινό (?) (καθολικά) • Μεγάλος όγκος • Διευκόλυνση (24/7) |

Εξέλιξη και Αφανισμός

- Η εξέλιξη φέρνει (ανταγωνισμό και) αφανισμό;
- Άλογο και άμαξα ⇒ αυτοκίνητο
 - Αυτοκίνητο + αεροπλάνο
 - Ραδιόφωνο + τηλεόραση
 - Εφημερίδα + (ραδιοφωνική / τηλεοπτική) μετάδοση νέων
 - Παρατηρούμε ότι η πληροφορία και οι υπηρεσίες είναι **συμπληρωματικά** – όχι ανταγωνιστικά

Εξέλιξη Ψηφιακών Βιβλιοθηκών με το Χρόνο

- Πρώτα βήματα
 - Περιορισμένες υλοποιήσεις Ψηφιακών Βιβλιοθηκών τοπικού χαρακτήρα
- Επόμενα βήματα
 - Παγκόσμια διασύνδεση
 - Των τοπικών ή νέων υλοποιήσεων
 - Κύριο πρόβλημα: διαλειτουργικότητα
 - Βασικότερη προσέγγιση: μεταδεδομένα

Ψηφιακές Βιβλιοθήκες και Νεωτερισμοί

Οργάνωση Ψηφιακού Περιεχομένου

- Το ψηφιακό περιεχόμενο
 - Σε βιβλιοθήκες, αρχεία, μουσεία, ...
- Έχει τις ίδιες αρχές οργάνωσης.
 - Με διαφορετικά πρότυπα μεταδεδομένων.
- Και το χειρίζομαστε με τον ίδιο τρόπο.

Βιβλιοθήκες, Ψηφιακές ή Μη

Έχουν **όμοιες** λειτουργίες:

- Επιλογή υλικού βιβλιοθήκης (προσκτήσεις)
- Κανόνες τήρησης μεταδεδομένων
- Κατηγοριοποίηση υλικού
- ...

Βιβλιοθήκες, Ψηφιακές και Μη

Έχουν **παρόμοιες** λειτουργίες:

- Πολιτικές πρόσβασης / αντιγράφων
- Επικύρωση / επιλογή / έλεγχος ποιότητας
- Ύπαρξη μεσαζόντων
- Διατήρηση (π.χ., πόρων αλλά και δημόσιων κλειδιών)
- Διαχείριση ρυθμίσεων (configuration management)
- Όροι χρήσης
- ...

Βιβλιοθήκες, Ψηφιακές μόνο: Διαφορές

- Ψηφιακό περιεχόμενο
- Στηρίζονται στον Παγκόσμιο Ιστό
- Υποστηρίζουν διευρυμένη, ταχύτερη πρόσβαση
- Κατάργηση φυσικών περιορισμών
- Έχουν νέα μεταδεδομένα (και πρότυπα)
- Περισσότερα συστήματα διαχείρισης γνώσης
 - Οντολογίες, ...
- Επιτρέπουν αυτόματη υποστηρικτική επεξεργασία

Βιβλιοθήκες, Ψηφιακές μόνο: Νεωτερισμοί

- Νέοι τύποι επαναχρησιμοποίησης και συμμετοχής
- Συνεισφορές από τους χρήστες
- Εικονικές κοινότητες και συνεργασίες
- Άμεση υποστήριξη σε δημιουργία και χρήση (όλος ο κύκλος ζωής της πληροφορίας)
- Συνεργατικό φιλτράρισμα, καταλογογράφηση, απάντηση σε ερωτήσεις
- Βιβλιοθήκες ανοικτού κώδικα

Προκλήσεις Ψηφιακών Βιβλιοθηκών

- Μεγάλες ποσότητες δεδομένων
- Αποθήκευση (Storage)
- Συμπίεση (Compression)
- Εργαλεία και περιβάλλοντα
- Πρότυπα
- Ανοικτά συστήματα

Προκλήσεις Ψηφιακών Βιβλιοθηκών Περιεχομένου

- Ταξινόμηση και ευρετηρίαση (Classification και indexing)
- Αναζήτηση (Searching)
- Προστασία περιεχομένου, τήρηση δικαιωμάτων
- Ανάκτηση πληροφοριών (Information retrieval)
- Παράδοση του περιεχομένου (Content delivery)
- Παρουσίαση του περιεχομένου (Presentation)
- Πολυγλωσσικότητα (Multilinguality)
 - Διαλειτουργικότητα γλωσσών (interoperability).
 - Διεθνοποίηση (internationalization).
 - Τοπικοποίηση (localization).

Προκλήσεις Ψηφιακών Βιβλιοθηκών Διαχείρισης

- Διαχειριστικά – Οργανωτικά – Διοικητικά ζητήματα (Administration)
- Εμπιστευτικότητα και επιλεκτικότητα
 - Κρυπτογράφηση, υδατογραφήματα
 - Κρυπτογραφημένη μεταφορά
 - Επαλήθευση ταυτότητας
 - Ψηφιακές υπογραφές
 - Ασφάλεια
 - Περιορισμοί πρόσβασης

Προκλήσεις Ψηφιακών Βιβλιοθηκών Διεπαφής – Ολοκλήρωσης

- Φιλικότερες διεπαφές χρήστη (User interface)
 - Παρουσιάζουν την πληροφορία.
 - Ενεργοποιούν την επιπλέον λειτουργικότητα.
 - Επιτρέπουν εξατομίκευση
 - Επιτρέπουν προσωπικές σημειώσεις.
- Ενσωμάτωση – ενοποίηση – ολοκλήρωση λειτουργιών και υπηρεσιών
- Διαλειτουργικότητα Ψηφιακών Βιβλιοθηκών

Διάθεση Ψηφιακού Υλικού

- Τέσσερις κατηγορίες με βάση
 - Το πού βρίσκεται η πληροφορία
 - Τι δέσμευση αναλαμβάνει η βιβλιοθήκη για τη διαχείριση και μακροχρόνια συντήρησή του:

• Αποθηκευμένα – <i>Archived</i>	Τοπικά	Συντήρηση
• Αντιγραμμένα – <i>Mirrored</i>	Τοπικά	Καμία
• Διασυνδεδεμένα – <i>Linked</i>	Απομακρυσμένα	Καμία
• Εξυπηρετούμενα – <i>Served</i>	Απομακρυσμένα	Συντήρηση
- Είναι δυνατή η αλλαγή κατηγορίας με βάση τις ανάγκες των χρηστών, την ανταπόκριση του δικτύου, τα πνευματικά δικαιώματα, τη σπουδαιότητα του υλικού.

Τύποι Πρόσβασης

- Απομακρυσμένη
 - Ελεύθερη
 - Μέσω εξουσιοδότησης (login/password, IP)
- Τοπική
 - Στους χώρους της βιβλιοθήκης
 - Σε συγκεκριμένα μηχανήματα χρηστών

Νέες Πρακτικές Προσκτήσεων

- Άδειες χρήσεις (license agreements)
- Πολιτικές (policies)
- Κοινοπραξίες (consortia)
- Και νέοι τρόποι χρέωσης

Χρέωσεις

- Δωρεάν
- Ετήσια (ή μηνιαία) συνδρομή
- Χρέωση ανάλογα με την υπηρεσία
 - Π.χ. ανάκτηση (αριθμός άρθρων), αναζήτηση (χρόνος)
- Συνδυασμοί τρόπων χρέωσης ή δυνατότητα επιλογής τρόπου χρέωσης
 - Π.χ. Μερική πρόσβαση χωρίς χρέωση
 - Χρέωση σε αποστολή τυπωμένου αντιγράφου

Οικονομία και Ψηφιακή Βιβλιοθήκη

- Διερεύνηση της οικονομίας της κλίμακας
- Κατανόηση της εικονικής τιμής των ψηφιακών αντικειμένων
- Πρόβλεψη για ηλεκτρονικό χρήμα (E-Cash) / Μικρο-συναλλαγές (Micro-Transactions)

Διάκριση Τιμών

- Πρέπει να γίνει κατάλληλη **διάκριση τιμών** των πληροφοριακών πόρων
 - Πρέπει να προσδιορίσουμε πώς διαφορετικές κατηγορίες χρηστών αποτιμούν διαφορετικές υπηρεσίες
 - Χρειάζεται να προσδιορίσουμε αποτελεσματικές πολιτικές τιμών για ακαδημαϊκά αγαθά
 - Μέσω πειραμάτων μάρκετινγκ και οικονομετρικές αναλύσεις

Οικονομική Ανάλυση

- Χρειάζεται οικονομική **ανάλυση** για να κατανοήσουμε το χώρο των πιθανών βιώσιμων **μοντέλων**
- Όπου μοντελοποιούνται οι **οικονομικές επιπτώσεις** των διαφορετικών μορφών απόσβεσης κόστους
- Μερικά αξιοπρόσεκτα θέματα:
 - Ποιος πληρώνει (συγγραφέας, αναγνώστης, οργανισμοί)
 - Πώς χρεώνονται (με βάση τη χρήση, το ίδιο για όλους)
 - Πώς εξετάζεται το υλικό (τεχνικές αναφορές, πρακτικά συνεδρίων, με κρίση ...)
 - Βαθμός συσσώρευσης (άρθρα, περιοδικά, συλλογές).
 - Όροι και συνθήκες χρήσης (μόνο ο ενδιαφερόμενος, άδεια για όλο τον οργανισμό, απεριόριστη επαναδιαανομή).

Σημεία Προσοχής Ψηφιακών Βιβλιοθηκών

- Ευκολότερος σφετερισμός περιεχομένου.
- Ρευστότερη αξιοπιστία λειτουργίας.
- Αμφίβολη μονιμότητα συνδέσμων.
- Ανάγκη αυτοματοποίησης, λόγω κλίμακας.
- Εναλλακτικές δυνατότητες.
- Διαφορετικοί μηχανισμοί και πολιτικές.

Ψηφιακά Αντικείμενα

(Πληροφοριακοί Πόροι)

Επισκόπηση Θεμάτων

- Αρχιτεκτονικές ψηφιακών αντικειμένων
- Διαχείριση πρόσβασης

Ανομοιομορφία Μέσων

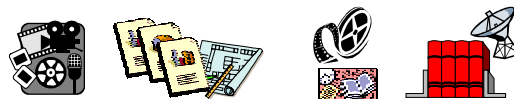
- Το περιεχόμενο (ψηφιακά αντικείμενα) είναι «απλώς bits», κατά κάποιο τρόπο
 - Έτσι έχουμε Ψηφιακές Βιβλιοθήκες «πολυμέσων» και «μικτών» μέσων
- Προκλήσεις στις Ψηφιακές Βιβλιοθήκες
 - Ομοιομορφία διαχείρισης, ανακάλυψης, πρόσβασης, ...
 - Ελεγκτασιμότητα για διάφορα μέσα, κοινότητες, χρήσεις, ...

Το Ψηφιακό Περιεχόμενο Μπορεί να Αναπαριστά ...

- Απλές, οικείες οντότητες



- Περίπλοκα, σύνθετα, δυναμικά αντικείμενα



Αρχιτεκτονικές Ψηφιακών Αντικειμένων - Πόρων

- Νέες μορφές περιεχομένου θα συνεχίζουν να αναπτύσσονται
- Νέοι τρόποι συσσώρευσης και δόμησης περιεχομένου θα συνεχίζουν να αναπτύσσονται
- Η «ύπαρξη της βιβλιοθήκης» κυρίως σημαίνει καθιέρωση αρχών κανονικοποίησης πάνω σε ανόμοιους πόρους

Ψηφιακό Αντικείμενο και Αποθετήριο

- Το ψηφιακό αντικείμενο
 - Περιβλήμα για συγκέντρωση ψηφιακού υλικού
 - Τρόπος ομοιομορφου χειρισμού πολύπλοκων τύπων
 - Γενικός μηχανισμός ελεγκτασιμότητας (δεδομένων)
- Το αποθετήριο
 - Επίπεδο υπηρεσιών για τα "περιεχόμενα" ψηφιακά αντικείμενα
 - Διαχείριση του κύκλου ζωής των αντικειμένων
 - Ασφαλές περιβάλλον για να τρέχει φορητός κώδικας

Προκλήσεις Γεωγραφικής Πληροφορίας

Cataloging – Metadata Creation

Metadata

```

<ENTITY % geographic-coordinate "(#PCDATA)">
<!-- a geographic latitude in degrees north of the equator or
geographic longitude in degrees east of the Greenwich
meridian, e.g. "-41.025" -->
<ELEMENT west_bounding_coor %geographic-coordinate>
<ELEMENT east_bounding_coor %geographic-coordinate>
<ELEMENT south_bounding_coor %geographic-coordinate>
<ELEMENT north_bounding_coor %geographic-coordinate>
<ELEMENT measurement_begin_date %calendar-date>

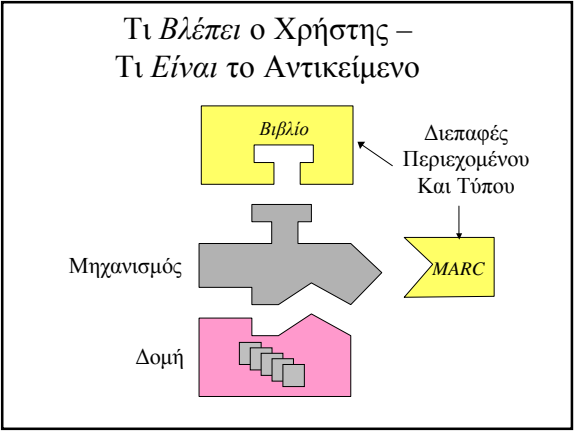
```

Georeferencing by placename and by spatial footprint

Translation needed between placenames - locations

Search Engines

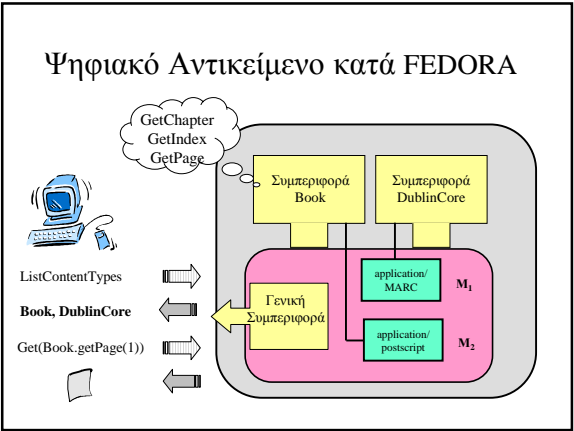
Where is ...?
What's there?
What happened there?



Το Ψηφιακό Περιεχόμενο Είναι ...

- Διακρίνεται από το τι μπορεί να **κάνει** (όχι από το πώς έχει κατασκευαστεί)

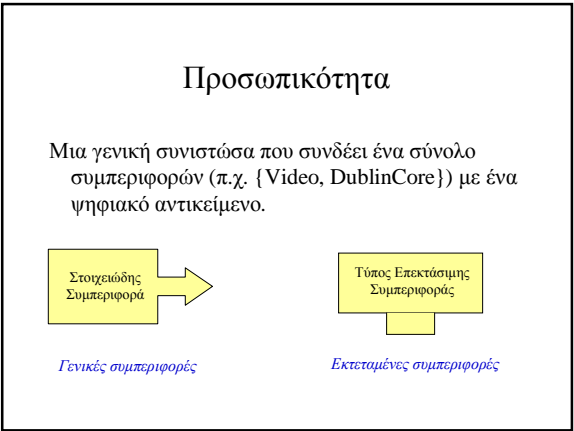
getSection
getArticle
getTrack
getLabel
getChapter
getPage
getFrame
getLength



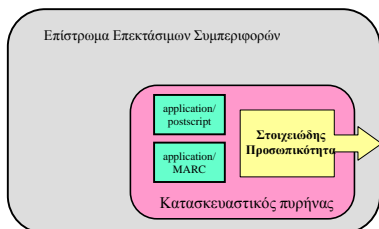
Τύπος Κήρυκα

Τύπος **Κήρυκα** (Disseminator type)

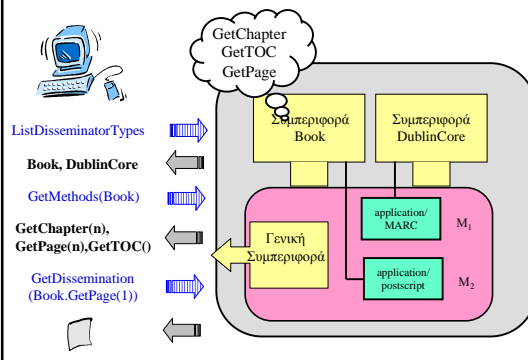
- Παρέχει το **σύνολο από συμπεριφορές** που περιγράφει επίσημα την λειτουργικότητα οποιασδήποτε αντίληψης του περιεχομένου, καθολικής ή τοπικής (εξαρτώμενης από την κοινότητα)
- Π.χ. { Video, DublinCore, Experiment, ... }



Προσωπικότητες Ψηφιακού Αντικειμένου



Οι Πελάτες Επικοινωνούν με Γενικές Αιτήσεις



Δομή Κλασικού Τεκμηρίου

- Λογική δομή πόρου τύπου βιβλίου
 - Μερικές φορές και παρόμοια φυσική (αρχαία) δομή



Στόχοι Ψηφιακού Αντικειμένου

- **Κατανομή** – ψηφιακού περιεχομένου και υπηρεσιών
- **Σταθερότητα** διεπαφής – ψηφιακών αντικειμένων
- **Διαλειτουργικότητα** – για ψηφιακά αντικείμενα και αποθετήρια
- **Επεκτασιμότητα** – φυσικά εξελισσόμενο σύστημα τύπων
- **Ευελξία** – Ανάπτυξη οδηγούμενη από την κοινότητα
- **Ασφάλεια** – διαχείριση δικαιωμάτων και έλεγχος πρόσβασης
- **Διατήρηση** – μακροβιότητα των ψηφιακών αντικειμένων

Διαχείριση Πρόσβασης Ψηφιακών Αντικειμένων

Πολιτικές Διάθεσης Περιεχομένου

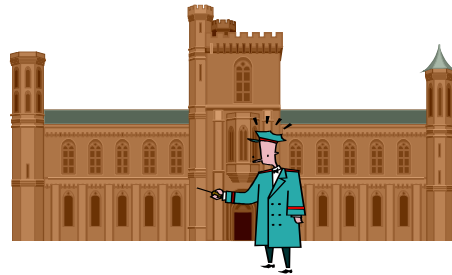
Πολιτικές διάθεσης περιεχομένου με εμπορική αξία και διαδικασίες χρέωσης:

- **Αυστηρότητα εφαρμογής**
 - Διασφαλίζει τα έσοδα
 - **Φιλικότητα προς τους χρήστες**
 - Χαλαρότητα μέτρων
 - Χάσιμο κερδών
 - Προσέλκυση περισσότερων χρηστών
- Όπως και με την αγορά του λογισμικού!

Διαχείριση Πρόσβασης

- ... Έλεγχος πρόσβασης με βάση την συμπεριφορά
- Η πολυπλοκότητα του προβλήματος της διαχείρισης της πρόσβασης αντικατοπτρίζει την πολυπλοκότητα των ψηφιακών αντικειμένων
 - Ο βαθμός λεπτομέρειας των γνωστών τεχνικών διαχείρισης πρόσβασης (π.χ. αρχεία και κατάλογοι αρχείων) δεν ταιριάζουν με τον έλεγχο πρόσβασης με βάση την συμπεριφορά

Επιβολή Πολιτικής Γενικού Σκοπού



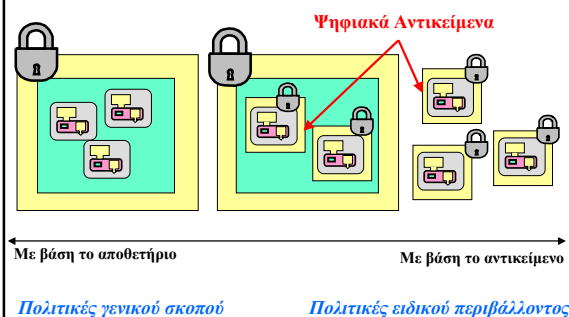
Επιβολή Πολιτικής Ειδικού Περιβάλλοντος



Περιορισμοί των Παραδοσιακών Μηχανισμών Ελέγχου Πρόσβασης

- Περιορισμένη εκφρασσιμότητα πολιτικών
- Σταθερό σύνολο αφαιρέσεων
 - Τα αντικείμενα είναι αρχεία, κατάλογοι, ...
 - Οι ενέργειες είναι ανάγνωση, αλλαγή, εκτέλεση, εκτύπωση, ...
- Δεν επεκτείνεται εύκολα για σύνθετες ή λεπτομερείς πολιτικές

Αλληλουχία Επιβολής Πολιτικής

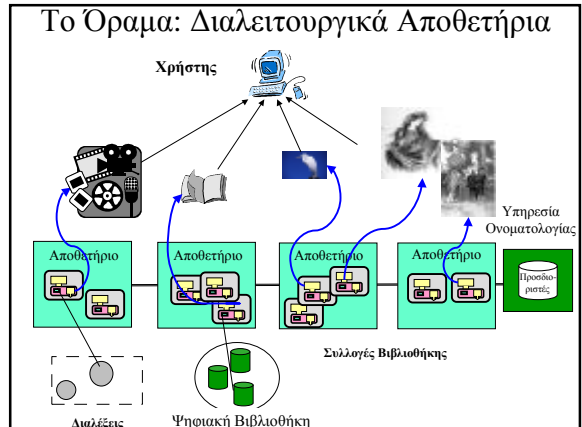


Επιβολή Πολιτικής

- Η σημασιολογία της γλώσσας των πολιτικών πρέπει να συμβαδίζει με τη σημασιολογία της συμπεριφοράς των πραγματικών οντοτήτων
- Ασφαλής επιβολή λεπτομερών και ειδικών πολιτικών
- Επεκτασιμότητα για πολιτικές και μηχανισμούς επιβολής
- Υποστήριξη για μεταφεριμότητα και φορητό υπολογισμό (επιβολή πολιτικών σε μη έμπιστους κινητούς πράκτορες)
- Αποκεντρωμένη διαχείριση πολιτικών

Παραδείγματα: Πολιτικές Ειδικού Περιβάλλοντος

- Διάλεξη: εκπαίδευση από απόσταση
 - "Οι απλοί επισκέπτες μπορούν να δουν τις διαφάνειες 1-10 της διάλεξης 1, αλλά όχι άλλες διαφάνειες ή το βίντεο της διάλεξης."
 - "Οι φοιτητές δεν μπορούν να δουν το βίντεο της διάλεξης 2 αν δεν αποστείλουν πρώτα την άσκηση της διάλεξης 1."
- Βιβλίο: ψηφιοποίηση Βιβλιοθήκης
 - "Πριν τη λήξη του copyright στις 1/1/2004, οι φοιτητές έχουν πρόσβαση στα κεφάλαια 1-6 και το ευρύ κοινό στις σελίδες 1-20 του κεφαλαίου 1. Μετά τη λήξη, όλοι οι χρήστες έχουν πρόσβαση σε όλες τις σελίδες όλων των κεφαλαίων"
- Χαρτοφύλακας τεχνολογιών: επιχειρησιακή στρατηγική
 - "Οι διευθυντές μπορούν να δουν προδιαγραφές προϊόντων μόνο μετά την τελική αναφορά του τμήματος Έρευνας και Ανάπτυξης"
 - "Μόνο οι υψηλόβαθμοι μπορούν να τρέξουν τον εξομοιωτή του μεριδίου της αγοράς"



Τεχνολογίες Ψηφιακών Αντικειμένων

Τεχνολογίες Ψηφιακών Αντικειμένων – Επισκόπηση –

- Τεχνολογίες και αναζήτηση κειμένου
 - Μορφότυπα και αποθήκευση
- Αναζήτηση και κατανόηση εικόνων
- Ήχος και αναγνώριση ομιλίας
- Μορφότυπα βίντεο
- Παράδειγμα (CMU)
 - Χαρακτηρισμός βίντεο
 - Οπτικοποίηση χωρικής πληροφορίας

Αναζήτηση σε Αρχεία Κειμένου

- Σειριακή σάρωση (grep): δεν έχει προβλήματα ενημέρωσης, ακατάλληλη για πολύ μεγάλες συλλογές
- Ευρετηρίαση με ανεστραμμένα αρχεία
- Πιθανή συμπίεση λιστών εμφανίσεων, ευρετηρίαση και με τα δύο άκρα, αναζήτηση σε συγκεκριμένα πεδία, διατήρηση πληροφοριών συχνότητας όρων, ...
- Αρχεία υπογραφών: ηλεκτρονικά σημαδεμένες κάρτες
- «Bitmaps»: βέλτιστα για πολύ κοινές λέξεις – βελτιώνουν τα ανεστραμμένα αρχεία
- Περίληψη αποτελεσμάτων, σύμπτυξη ομοειδών
- Κεφαλαία – μικρά, προθέματα, stop λίστες.

Γιατί Όχι Ένα DBMS

Γιατί τα συστήματα ανάκτησης κειμένου δεν χρησιμοποιούν ένα DBMS από κάτω?

- Λίγα αριθμητικά πεδία, και μεγάλος αριθμός άλλων αντικειμένων (κειμένων, εικόνων, ...)
- Ειδικές ανάγκες, όπως ξεφύλλισμα παραπομπών και αναζήτηση με αποκοπή (στα πρώτα αντικείμενα)
- Η είσοδος δεν χωρίζεται καθαρά σε εγγραφές, και μεταβλητός αριθμός αντικειμένων μπορεί να πρέπει να ανακτηθούν
- Κυρίως μόνο πρόσβαση, και σπάνια ενημέρωση
- Παράλληλη αναζήτηση – μόλις έρχεται στη μόδα
- Βαθμολόγηση αποτελεσμάτων (όχι 0/1)

Τι Κάνουν οι Μηχανές Αναζήτησης?

- Πολύ μεγάλα ανεστραμμένα αρχεία και παράλληλες μηχανές αναζήτησης σε πολλά (χιλιάδες) μηχανήματα
- Μεγάλες δευτερεύουσες μνήμες (caches). Μπορεί να ψάχνουν μόνο εκεί, και να αποφύγουν καθυστερήσεις δίσκων
- Είναι διατεθειμένες να δώσουν διαφορετικά αποτελεσμένα, ανάλογα με τι δεδομένα είναι στη δευτερεύουσα μνήμη

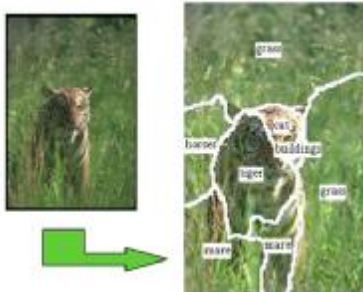
Αναζήτηση Εικόνας

- Ερωτήσεις (Query By Image Content) με βάση το χρώμα, τη βασική δομή (texture) ή κάποιο σχήμα
- Το έγχρωμο ιστόγραμμα είναι ευκολότερο: αλλά τα ηλιοβασιλέματα θα μας παραπλανήσουν
- Κατάτμηση και αναγνώριση (segmentation & labeling)



Κατανόηση Εικόνας

David Forsyth & Jitendra Malik, Berkeley



Λειτουργίες Αναγνώρισης Φωνής

Αναγνώριση συνεχούς ομιλίας, με μεγάλο λεξιλόγιο και ανεξάρτητα του ομιλητή

- Δημιουργεί κείμενο, για να επιτρέψει αναζήτηση βάσει κειμένου σε «προφορικούς» πόρους
- Βελτιώνει το συγχρονισμό ήχου και βίντεο σε σενάρια εκτέλεσης
- Προμηθεύει την πληροφορία την απαραίτητη για κατάτμηση και χρήση πολυμέσων
- Παρέχει διεπαφή ήχου στην Ψηφιακή Βιβλιοθήκη

Πηγές Γνώσης Αναγνώρισης Φωνής



Αναζήτηση Ήχων

- Ομιλία: αναγνώριση ομιλίας, ταύτιση ομιλητή
- Μουσική: τώρα έχουμε λογισμικό “σφύριξε και αναζήτησε”. Π.χ. Bill Birmingham, U of Michigan – Donald Byrd, U. Mass.



Παράδειγμα Ψηφιακής Βιβλιοθήκης (CMU) με Ειδήσεις (1)

Από το CMU, με περιεχόμενο βίντεο

- Δυνατότητα *ομοιότητας* εικόνων
 - Η *ομοιότητα* νοείται διαφορετικά για άλλους ανθρώπους, ή για άλλες καταστάσεις.
- Αναγνώριση φωνής
 - Συμβουλευόμενο κατάλληλες πηγές γνώσης
 - Παράγει το κείμενο της φωνής, για χρήση ...
 - Και παρέχει διεπαφή φωνής

Παράδειγμα Ψηφιακής Βιβλιοθήκης (CMU) με Ειδήσεις (2)

- Εξαγωγή χωρικής πληροφορίας
 - Για κατηγοριοποίηση / αναζήτηση
 - Παρουσίαση σε χάρτη
 - Και διεπαφή αναζήτησης (σε χάρτη)
- Ανάλυση γλώσσας
 - Συντακτική και γραμματική
 - Και σημασιολογική (οντολογίες)
 - Και προχωρημένες διεπαφές.

Ανάκτηση Πληροφορίας

- Μηχανή αναζήτησης SEIDX
 - Μακρινός συγγενής του Lycos
- Εντοπισμός ελεγχόμενων λέξεων, stop-words, συνώνυμων, έξτρα υπολογισμός για «πρωτιά», κανονικοποίηση αριθμού λέξεων, χειρισμός καταλήξεων
- Ευρετηρίαση κρυμμένων εννοιών
 - Σημασιολογική επέκταση ερωτήσεων (με βάση το λεξιλόγιο των άρθρων)
 - Π.χ. accident / crash, bacteria / germ, children / kids

Εικόνες με Παρόμοια Χρώματα



Εικόνες με Παρόμοιες Μορφές



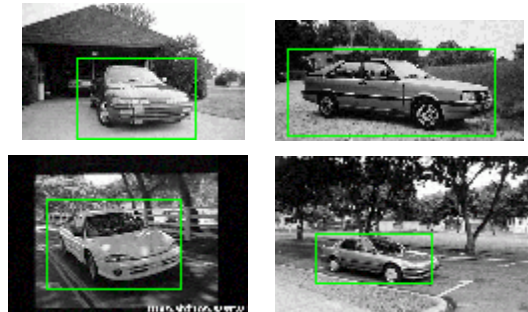
Εικόνες με Παρόμοιο Περιεχόμενο



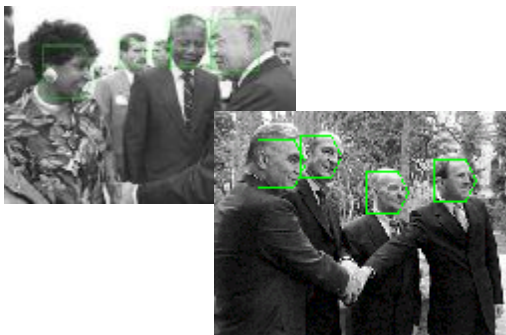
Ανίχνευση Κειμένου



Ανίχνευση Αυτοκινήτου



Ανίχνευση Προσώπου



Αναζήτηση Βίντεο

- Συνδυασμός από:
 - Αναγνώριση υποτίτλων
 - Αναγνώριση ομιλίας
 - Αναγνώριση προσώπων
 - OCR κειμένου στην οθόνη
 - Κάποια αναζήτηση σε εικόνες
- Επίσης, πολλή δουλειά στην παρουσίαση
- Πηγή: Informedia, Howard Wactlar, CMU.

Αυτόματος Χαρακτηρισμός Βίντεο

Scene Cuts			
Camera	Static	Static	Zoom
Objects	Adult Female	Animal	Two adults
Action	Head Motion	Left Motion	None
Captions	[None]	Yellowstone	[None]
Scenery	Indoor	Outdoor	Indoor

Εξαγωγή Χωρικής Πληροφορίας από Άρθρα

- Εξαγωγή **ονομάτων** οντοτήτων για αναγνώριση, κωδικοποίηση και ευρετηρίαση όλων των αναφερόμενων τοποθεσιών σε κάθε άρθρο
- Αυτόματη δημιουργία **χαρτών** που συνοψίζουν και ακολουθούν την εξέλιξη κάθε άρθρου
- Κάνει δυνατή την αναζήτηση σε άρθρα που αφορούν την περιοχή που **οριοθετεί** ο χρήστης



Άλλες Δυνατότητες

- Υπάρχουν πολλά πράγματα στις Ψηφιακές Βιβλιοθήκες σήμερα. Και έπονται περισσότερα: 3-D αντικείμενα, επιστημονικά δεδομένα, λογισμικό, ...
- Όλα αυτά πρέπει να αποθηκευτούν, οργανωθούν και αναζητηθούν
- Στην επόμενη γενιά Ψηφιακών Βιβλιοθηκών:
 - Ιδέες σε πιλοτική εφαρμογή σήμερα
 - Που θα είναι ώριμη τεχνολογία αύριο.

Πρόσφατες Εξελίξεις

- Νέα μοντέλα πόρων
 - Με πολλαπλές μορφές και χρήσεις (π.χ. χρήση σε GIS)
- Συνεργατικό ποιοτικό φιλτράρισμα ή κριτική
- Συμπαγείς συνδέσεις (π.χ. DOI)
- Ερμηνεία εικόνας:
 - Καλύτερη ερμηνεία εικόνας βάσει περιεχομένου
 - Συνδυάζοντας εικόνα και κείμενο
- Αναγνώριση τεκμηρίων (με ανάλυση της διάταξής τους).
- Προσωπικές Βιβλιοθήκες
- Αυτό-διαχειριζόμενα τεκμήρια
- Συλλογές με εξειδικευμένο περιεχόμενο:
 - Εικόνες και συλλογές δεδομένων που συνδέονται με τη βιολογία
 - Προσπάθεια ψηφιοποίησης σπάνιων βιβλίων

Αναζήτηση σε Σελίδα



Αποθήκευση Σημειώσεων σε Ιστοσελίδα



Έμφαση και Σημειώσεις

PDF FILE # 1

FARLAN: A MODEL OF SALES

3. The Analysis

The maximum number of customers a store can sell is $1/\epsilon > 0$. Let $p^* = (1+\epsilon)/2$. If $\epsilon > 0$ is the average cost associated with the number of customers.

PROPOSITION 1: If $p > 0$ then $\mu < \mu^*$.

PROOF: Suppose that all stores were charging a single price p with $\mu > p^*$. Then a slight increase in price by one of the stores would result in a higher profit, and that

If a store's price is changed a slightly lower price $p' < p$ with the same probability with which the other stores charge p , it would lose profits in order to gain a fixed positive amount of profit when the other stores bid. That for small ϵ (μ profit) would be positive, contradicting the assumption of equilibrium.

The number of stores of positive profit in any probability distribution must be finite. If we can find an arbitrarily small ϵ such that $p > p^*$ is charged with probability ϵ , consider what happens if we charge $p - \epsilon$ with the probability with which we used to charge p , and charge p with probability ϵ . The increase in profit will be

(Profit) $\epsilon (1 - p^*)$

Σύνθεση με Μεγέθυνση

VOL. 70 NO. 4

FARLAN: A MODEL OF SALES

equations

$$\pi_i(p) \prod_{j \neq i} (1 - F_j(p))$$

$$= -\pi_j(p) \prod_{k \neq j} (1 - F_k(p))$$

$$\pi_i(p) \prod_{j \neq i} (1 - F_j(p)) = -\pi_j(p) \prod_{k \neq j} (1 - F_k(p))$$

Dividing one equation into the other, we have

$$\frac{(1 - F_i(p))^{-1}}{(1 - F_j(p))^{-1}} = \frac{1 - (1 - F_i(p))^{-1}}{1 - (1 - F_j(p))^{-1}}$$

which implies $F_i(p) = F_j(p)$.

REFERENCES

J. Farrell, D. Foray, and R. Zeckhauser, "Price Variation in Almost Competitive Markets," *J. Econ. May 1979*, 93, 189-211.

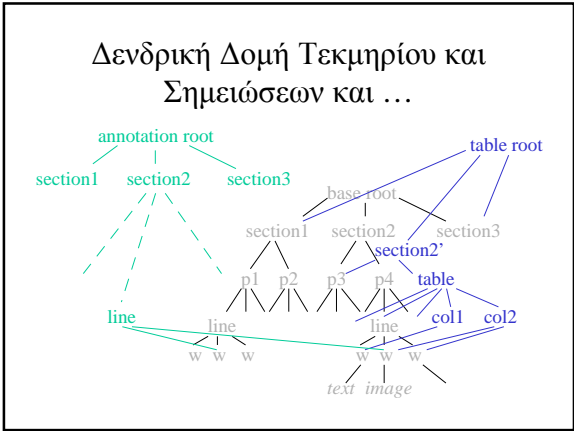
H. Rothstein, "Models of Markets with Imperfect Information: A Survey," *J. Polit. Econ.*, Dec. 1973, 61, 1283-308.

S. Shilony, "The Noisy Monopolist: Imperfect Information, Price Dispersion and Price Discrimination," *Rev. Econ. Stud.*, Oct. 1977, 44, 465-91.

J. Stiglitz, "A Theory of Bargaining," *Stanford Univ. 1977*.

J. Stiglitz, "Bargain and Monopolistical Price Dispersion," *Rev. Econ. Stud.*, 1977, 44, 493-510.

S. Shilony, "Mixed Pricing in Equilibrium Comparison Shopping," *Rev. Econ. Stud.*, July 1979, 46, 543-54.



Άλλες Λειτουργίες Ψηφιακών Βιβλιοθηκών

Προσκτήσεις
Ψηφιοποίηση

Οι Χρήστες θα Μπορούν ... Ανάκτηση

- Να έχουν πρόσβαση σε ομόσπονδα αποθετήρια που περιέχουν όλη τη σχετική πληροφορία για κάποιο επιστημονικό σκοπό
 - Π.χ. όλες οι εμφανίσεις ενός ζώου στην Ευρώπη, παρουσιαζόμενες σαν χάρτης κατανομής ανά χώρα
- Να βρουν τα καλύτερα (ή τα πιο αμφισβητούμενα) άρθρα σε μια περιοχή
- Να επωφελούνται από αυτά που έχουν βρει χρήσιμα οι συνάδελφοί τους

Οι Χρήστες θα Μπορούν ... Συνεργασία

- Να βρίσκουν εύκολα, προσπελαύνουν και σχολιάζουν οποιοδήποτε διαθέσιμο ψηφιακό αντικείμενο (άρθρο περιοδικού, τεχνική αναφορά, σύνολο δεδομένων, χάρτη, σχόλιο) σε οποιαδήποτε μορφή
 - Για να χρησιμοποιηθεί από κριτές - αξιολογητές, συναδέλφους, φοιτητές
- Να βρίσκουν όλα τα σχόλια που έχουν γίνει σε κάποιο προκαταρκτικό ή δημοσιευμένο άρθρο

Οι Χρήστες θα Μπορούν ... Διευκόλυνση

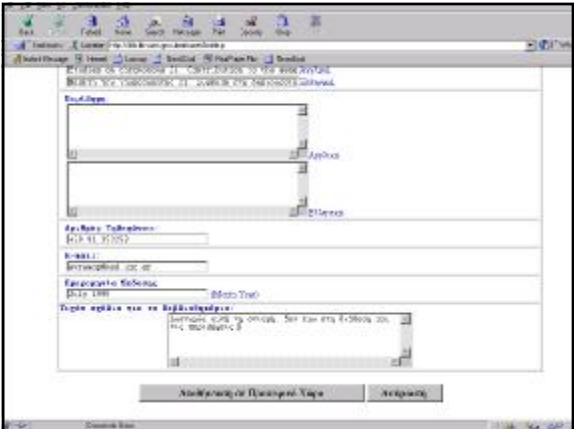
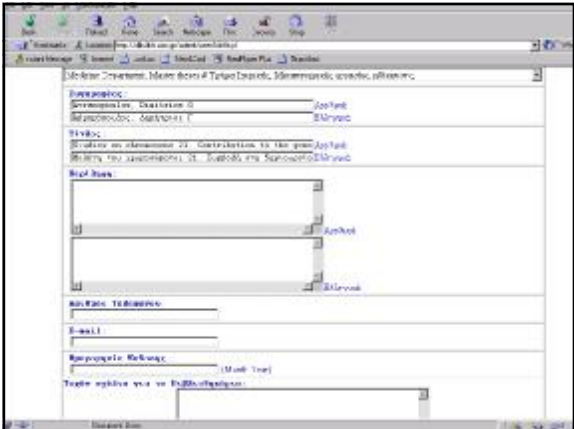
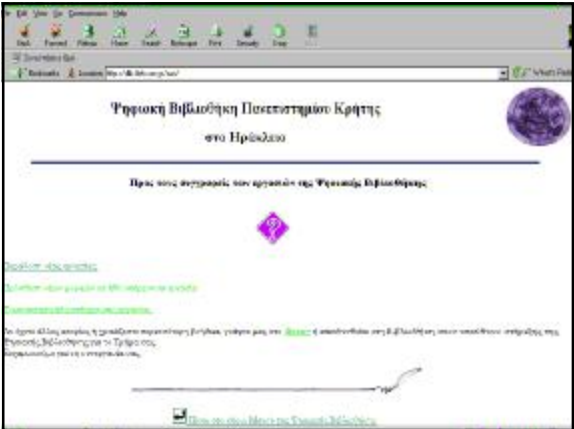
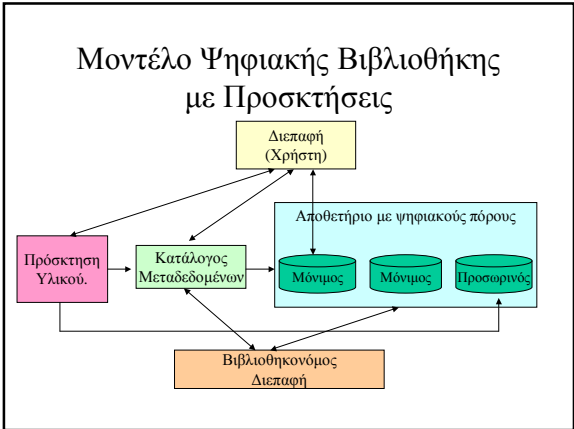
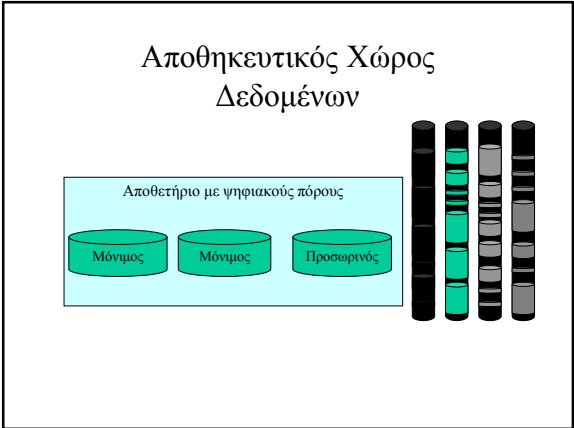
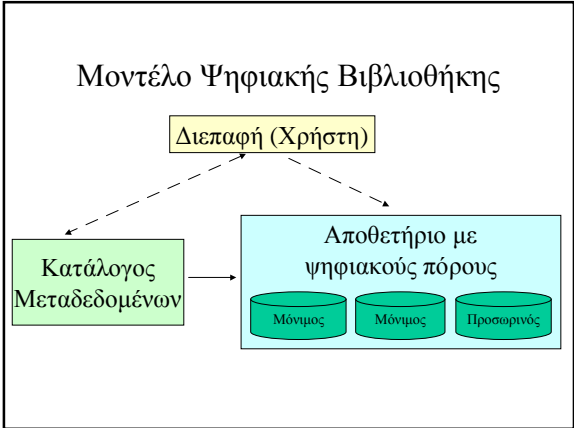
- Να επιλέγουν μια αναφορά σε σαρωμένη εικόνα, και να ξεκινά αναζήτηση για την περιγραφόμενη αναφορά
- Να ορίζουν αυθαίρετες ομάδες καταμεμημένων χρηστών και να τους δίνουν δυνατότητες πρόσβασης σε άρθρα / σχόλια / δεδομένα
- Να έχουν ένα σύστημα που θα υποστηρίζει όλα τα παραπάνω και είναι οικονομικά και τεχνολογικά κλιμακωτό

Χρησιμοποιώντας Νέες Τεχνολογίες για:

- Ανάλυση του περιεχομένου, για πρόσβαση
 - Ειδικά σε εικόνες και βίντεο
- Υποστήριξη του συστήματος
 - Για συμπαγή, διαλειτουργικά, αποδοτικά, διαθέσιμα και ασφαλή συστήματα
- Συνεργατικό χειρισμό πληροφορίας
 - Σημειώσεις (από κριτές, συνεργάτες, άλλους)
 - Καταμεμημένα μοντέλα πόρων και υποστήριξης χρηστών
 - Συνεργατικό φιλτράρισμα για κρίση από αξιολογητές
 - Υπηρεσίες διαχείρισης συλλογών (για υποστήριξη δημιουργίας και συντήρησης «προσωπικών» και «ομαδικών» βιβλιοθηκών χωρίς κόστος)
 - Υποδομή για αυτό-διαχειριζόμενα τεκμήρια
 - Φιλική παρουσίαση πολύπλοκων συνόλων δεδομένων

Πρόσκτηση – Δημοσίευση Υλικού







Ψηφιοποίηση – Πότε;

- Πότε ψηφιοποιούμε;
- Πάντα περιγραφή του υλικού (καταλογογράφηση)



Παρανοήσεις για Ψηφιοποίηση

- *Η ψηφιοποίηση είναι η λύση για να κάνει τους πόρους διαθέσιμους;*
 - Χρειάζεται αναλυτική **περιγραφή** τους
 - Συνήθως πριν την ψηφιοποίηση (καταλογογράφηση)
 - Αρκεί και αυτόματη / μηχανική εξαγωγή της περιγραφής
 - Και κατάλληλη (έστω και μικρού κόστους) **υποδομή** διάθεσης
 - Και πολιτική ενημέρωσης, διατήρησης ..., τους

Κριτήρια Ψηφιοποίησης

- Μεγάλη χρήση – πιθανώς απομακρυσμένη
 - Οικονομικά συμφέρον;
- Διατήρηση του πρωτότυπου
 - Αποφυγή χρήσης πρωτότυπου – αντιοικονομικό αλλά ίσως ιδιαίτερα θεμιτό
- Διατήρηση του περιεχομένου
 - Παρότι τα αντίγραφα είναι λιγότερο σημαντικά από τα πρωτότυπα, αυξάνουν τις πιθανότητες διάσωσης.

Κάνουμε Ψηφιοποίηση;

- Η ψηφιοποίηση είναι χρήσιμη
 - Για μερικές κατηγορίες υλικού
- Αλλά δεν πρέπει να εφησυχάζομαστε
- Πρέπει να γίνεται με καλό σχεδιασμό
 - Σταδιακά επεκτεινόμενη στις διαφορετικές συλλογές
 - Με κατάλληλα κριτήρια επιλογής υλικού / συλλογών
 - Τηρώντας τα αντίστοιχα πρότυπα
 - Και με στρατηγικές διατήρησης
- Μακροπρόθεσμα οφέλη – αλλά με βραχυπρόθεσμους πονοκεφάλους

Αρχιτεκτονικές και Στρατηγικές Ψηφιακών Βιβλιοθηκών

Επισκόπηση Θεμάτων

- Συνηθισμένα συστατικά αρχιτεκτονικής Ψηφιακών Βιβλιοθηκών
- Κλασικά μοντέλα Ψηφιακών Βιβλιοθηκών
 - NCSTRL – DIENST
 - SODA – Έξυπνα αντικείμενα
 - Πύλες, STARTS
- Αρχιτεκτονικές: Επιλογές και Στρατηγικές
- Στρατηγικές Οργάνωσης και Τάσεις
- Καταγραφή Στατιστικών Στοιχείων

Η Αρχιτεκτονική Αφορά

Αρχιτεκτονική = τρόπος σχεδιασμού / κατασκευής

- Αρχιτεκτονική συστημάτων
 - Κατανομή και πρόσβαση σε υπηρεσίες και υπολογισμό
 - Μοντέλα και πρόσβαση χρηστών
- Αρχιτεκτονική δεδομένων
 - Κατανομή και πρόσβαση στο περιεχόμενο
 - Μοντέλα και μορφότυπα δεδομένων
- Η Αρχιτεκτονική θα βοηθήσει
 - Βέλτιστη λειτουργικότητα και διαλειτουργικότητα

Η Αρχιτεκτονική των Ψηφιακών Βιβλιοθηκών Ενδιαφέρει Γιατί ...

- Οι εμπορικές λύσεις είναι περιοριστικές
 - Περιορισμένη ευελιξία
 - Θέματα καθιέρωσης και προτύπων
 - Οι διαδικτυακές εφαρμογές δεν έχουν την αρχιτεκτονική των Ψηφιακών Βιβλιοθηκών
- Ο ίδιος ο Ιστός δεν είναι Ψηφιακή Βιβλιοθήκη
- Πρέπει να διεκδικήσουμε καλύτερες Ψηφιακές Βιβλιοθήκες

Πρόσφατη Εξέλιξη Υποδομής Ιστού

- Uniform Resource Names (URNs)
 - Επιτρέπει μόνιμους, καθολικά μοναδικούς προσδιοριστές
- Extensible Markup Language (XML)
 - Επιτρέπει πολύ δομημένα αντικείμενα και πλούσιες δυνατότητες σύνδεσης (συσχέτισης)
 - <http://www.w3.org/XML>
- Resource Description Framework (RDF)
 - Επιτρέπει πλούσια σημασιολογία μεταδεδομένων για αντικείμενα
 - <http://www.w3.org/RDF>

Χρειαζόμαστε Αρχιτεκτονική Ψηφιακών Βιβλιοθηκών Γιατί ...

- Πλουσιότερο μοντέλο τεκμηρίων – **ψηφιακοί πόροι**
- Μόνιμα, μοναδικά ονόματα – **URNs**
- Καλά-ορισμένες **υπηρεσίες** ψηφιακών βιβλιοθηκών
- Καλύτερες εγκαταστάσεις για **ανακάλυψη πόρων**
- Ευέλικτος ορισμός **συλλογών**
- **Διαχείριση** κατανεμημένων πόρων (περιεχομένου) και υπηρεσιών
- Διαχείριση δικαιωμάτων **πνευματικής ιδιοκτησίας**

Κύρια Θέματα Αρχιτεκτονικής

- Ανοικτή Αρχιτεκτονική
 - Λειτουργικότητα διαρθρωμένη σε σύνολα **καλά ορισμένων υπηρεσιών**
 - Υπηρεσίες προσβάσιμες με **καλά ορισμένα πρωτόκολλα**
- Διαρθρωτή (πτυσσόμενη – modular) αρχιτεκτονική
 - Προάγει τη διαλειτουργικότητα
 - Προσαρμόσιμη για διαφορετικές χρήσεις (ερευνητικές βιβλιοθήκες, ανεπίσημη χρήση του Ιστού)
- Ομοσπονδία
 - Επιτρέπει συνάθροιση σε «λογικές συλλογές»
- Κατανομή
 - Περιεχομένου (αντικειμένων και συλλογών) και υπηρεσιών
 - Διαχείρισης ψηφιακών συλλογών και βιβλιοθηκών

Αρχιτεκτονική και Οργάνωση της Ψηφιακής Βιβλιοθήκης

- Η αρχιτεκτονική μιας Ψηφιακής Βιβλιοθήκης περιλαμβάνει τις βασικές αρχές που πρέπει να υφίστανται στη δομή της
- Η αρχιτεκτονική επισημαίνει τα κύρια στοιχεία διάκρισης της ψηφιακής βιβλιοθήκης, από μια οποιαδήποτε δικτυακή υπηρεσία
- Η αρχιτεκτονική και οργάνωση μιας τυπικής ψηφιακής βιβλιοθήκης συνήθως επιτρέπει σε πολλά υπολογιστικά συστήματα να συνεργάζονται

Μοντέλα Ψηφιακών Βιβλιοθηκών

- Οι Ψηφιακές Βιβλιοθήκες δεν έχουν μία μορφή
- Ανάλογα με το περιεχόμενο, τον τύπο και την προέλευσή του, τους όρους διάθεσής του, ...
- Αντίστοιχα εφαρμόζονται διαφορετικά μοντέλα (αρχιτεκτονικές) για ολόκληρη ή μέρος της Ψηφιακής Βιβλιοθήκης
- Η εσωτερική αρχιτεκτονική της Ψηφιακής Βιβλιοθήκης πρέπει να είναι ανεξάρτητη από το υλικό που είναι αποθηκευμένο σε αυτή

Δομή Πληροφοριών σε Μία Ψηφιακή Βιβλιοθήκη

- Κάθε Ψηφιακή Βιβλιοθήκη αποτελείται από ψηφιακά αντικείμενα (objects) ή πόρους (resources)
- Παραδείγματα ψηφιακών πόρων
 - Κείμενα σε γλώσσα HTML
 - Ψηφιοποιημένα έντυπα
 - Ψηφιακός ή ψηφιοποιημένος ήχος ή βίντεο ή ...
 - Προγράμματα υπολογιστών
 - Πολυμέσα
 - Άλλα ψηφιακά αντικείμενα ή υπηρεσίες

Αρχιτεκτονικές Ψηφιακών Βιβλιοθηκών – Επισκόπηση

- Ενότητα: Αρχιτεκτονικές Ψηφιακών Βιβλιοθηκών
- Αρχιτεκτονική = τρόπος σχεδιασμού / κατασκευής
- Κλασικά μοντέλα Ψηφιακών Βιβλιοθηκών
 - NCSTRL – DIENST
 - SODA – Έξυπνα αντικείμενα
 - Πύλες, STARTS

Μας Ενδιαφέρει η Αρχιτεκτονική των Ψηφιακών Βιβλιοθηκών

- Για να δούμε τι συνιστώσες Ψηφιακών Βιβλιοθηκών μπορούμε να έχουμε
- Δεν αποτελούν πλήρη σύγχρονα συστήματα, αλλά συγκέντρωση πιθανών συνιστωσών

Επισκόπηση Κύριων Σημείων

- Βασικές συνιστώσες
- Μοντέλα – πώς δουλεύουν
- Πού βρίσκεται το περιεχόμενο
- Πώς έχουμε πρόσβαση στο περιεχόμενο
- Πού / πώς υλοποιούνται οι υπηρεσίες
- Διαλειτουργικότητα – συνεργασία
- Κατάλληλες επιλογές
 - Ανάλογα με το περιβάλλον και τη στρατηγική μας

Χρειαζόμαστε Αρχιτεκτονική Ψηφιακών Βιβλιοθηκών

- Για να επιλέγουμε στρατηγική
- Για βέλτιστη λειτουργικότητα και διαλειτουργικότητα

Μοντέλα Ψηφιακών Βιβλιοθηκών

- NCSTRL – DIENST (Cornell University)
 - SODA – Έξυπνα αντικείμενα (Old Dominion University)
 - Πύλες, STARTS (Stanford University)
- Άλλα μοντέλα / προϊόντα:
- GreenStone (NLNZ)
 - DSpace (HP + MIT)
 - Informedia (CMU)
 - ABEKT (EKT)

Διαθεσιμότητα

- Πιλοτικές εφαρμογές, ελεύθερα διαθέσιμες:
 - SODA (ODU), <http://www.cs.odu.edu>
 - STARTS (Stanford), <http://www.stanford.edu>
 - Informedia (CMU), <http://www.cmu.edu>
- Έτοιμα για χρήση, και ελεύθερα διαθέσιμα:
 - DIENST (Cornell – v4 εξελιγμένη: ICS/FORTH)
 - <http://ncstrl.cs.cornell.edu>, <http://www.ics.forth.gr/~dienst>
 - GreenStone (NLNZ), <http://www.greenstone.org>
 - DSpace (HP + MIT), <http://dspace.org>

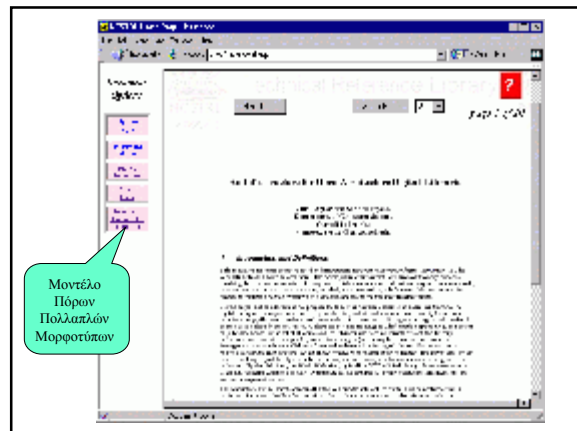
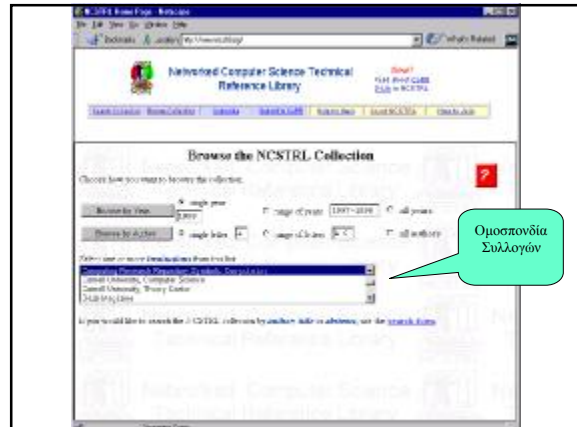
NCSTRL

Το Ομόσπονδο Μοντέλο του NCSTRL

- Έχει κοινό πρωτόκολλο (DIENST)
- Έχει κοινό πρότυπο μεταδεδομένων
- Βασίζεται στην ενεργή συνεργασία των εξυπηρετητών
- Αναπτύχθηκε στο πανεπιστήμιο Cornell για τεχνικές αναφορές στην Επιστήμη των Υπολογιστών
- Έχει τα βασικά στοιχεία της αρχιτεκτονικής ΨΒ
 - Με ένα (από πολλά πιθανά) σημείο εισόδου στο σύστημα

Δημιουργία Ψηφιακής Βιβλιοθήκης

- Δημιουργία αυτόρκους διεθνούς ομόσπονδης ψηφιακής βιβλιοθήκης
- Σύντομα απέκτησε πάνω από 120 κόμβους παγκοσμίως, και έγινε το πρώτο μεγάλο περιβάλλον πειραματισμού
- Κατάλληλο για μελέτη δύσκολων, προβλημάτων του πραγματικού κόσμου: θέματα πολιτικών, ποιότητας υπηρεσιών, ομοσπονδίας εκδοτών, σύστημα δοκιμασίας (Testbed)



DIENST

DIENST Πρωτόκολλο και Σύστημα

- Διαρθρωτό (πτυσσόμενο - modular) σύστημα βασισμένο σε αρχιτεκτονική ανοικτών προτύπων
- Το πρωτόκολλο DIENST περιλαμβάνει:
 - Υποδομή βασισμένη στις υπηρεσίες
 - Μοντέλο δεδομένων (αντικειμένων)

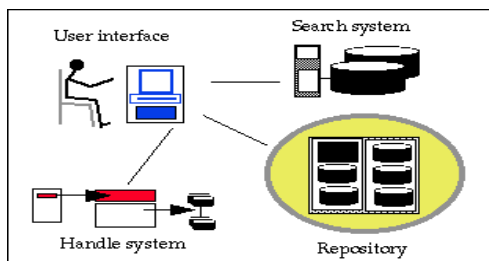
DIENST Η Τεχνική Βάση του NCSTRL

- Υλοποιεί μια αρχιτεκτονική βασιζόμενη σε υπηρεσίες για κατανεμημένες ψηφιακές βιβλιοθήκες
- Πρωτόκολλο και παραδειγματική υλοποίηση
- Δίκτυο υπηρεσιών
- Πρόσβαση μέσω του ξεφυλλιστή Ιστού
- Ομοιόμορφη αναζήτηση σε κατανεμημένα ευρετήρια
- Πρόσβαση σε πόρους σε κατανεμημένα αποθετήρια
- Πρόσβαση σε πολλαπλά μορφότυπα πόρων

Βασικές Συνιστώσες

- Οι συνιστώσες εκφράζουν (συνεργαζόμενα) τμήματα με ανεξάρτητους ρόλους και σχεδιασμούς
- Οι «βασικές» συνιστώσες της δομής μιας τυπικής ψηφιακής βιβλιοθήκης είναι συχνά θέμα οπτικής γωνίας:
- Διεπαφές με τον χρήστη (**User interfaces**)
 - Αποθετήριο δεδομένων (**Repository**)
 - Σύστημα προσδιορισμού (ή ονοματολογίας – **Handle system**)
 - Σύστημα αναζήτησης (**Search system**)

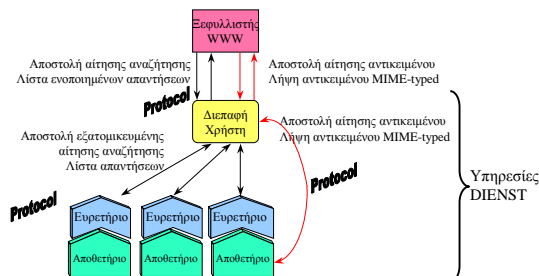
Βασικές Συνιστώσες – Σχηματικά



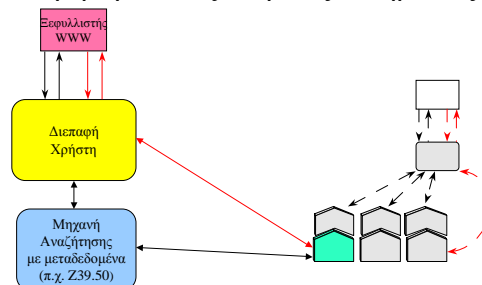
Υποδομή με Βάση τις Υπηρεσίες

- Το πρωτόκολλο των αιτήσεων εξαρτάται από την υπηρεσία
- Ορισμός των υπηρεσιών του πυρήνα
 - Αποθετήριο – πρόσβαση στα αντικείμενα
 - Ευρετήριο – απόκριση σε ερωτήσεις
 - Ονοματολογία – μοναδικά ονόματα (handles)
 - Διεπαφή χρήστη – ανθρωποκεντρικές πύλες
 - Ομοσπονδία – ορισμός συντονισμού των εξυπηρετητών
- Το μοντέλο πόρων (αντικειμένων)

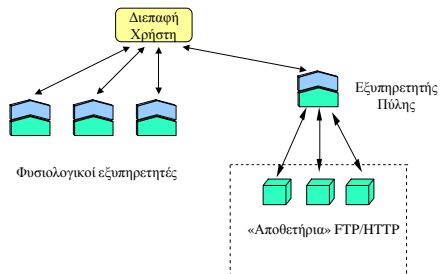
Υπηρεσίες του DIENST



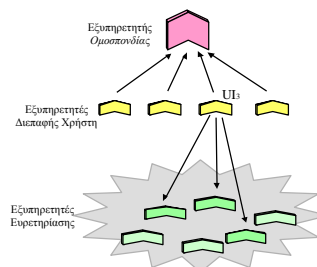
Χρήση από Εξωτερικές Υπηρεσίες



Πρωτόκολλο DIENST – Κατασκευή Πυλών για Μη Συμμορφούμενους

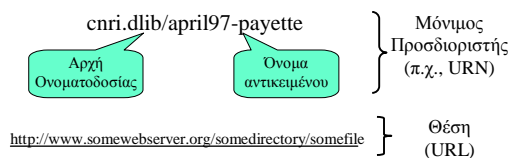


Υπηρεσία Ομοσπονδίας DIENST



Υπηρεσία Ονοματολογίας – CNRI

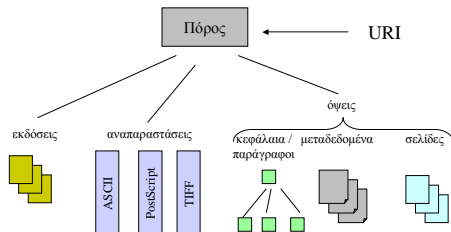
- Πόροι προσδιοριζόμενοι με καθολικά μοναδικά ονόματα
- Τα ονόματα είναι μόνιμα (διαρκή)
- Τα καταχωρημένα ονόματα υποδεικνύουν συγκεκριμένη θέση (URL)



Πρωτοβουλίες Προσδιοριστών

- IETF Uniform Resource Names (URN)
 - Προσδιορισμός του πλαισίου των URN
 - Απαιτήσεις για συστήματα «επίλυσης»
 - Ορισμός συντακτικού
- Υπάρχοντα Συστήματα προσδιοριστών
 - CNRI handles (χρησιμοποιείται στο NCSTRL)
 - PURLs (OCLC)
 - DOIs (τα πιο διαδεδομένα σήμερα)

Το Μοντέλο Δεδομένων του DIENST



Πρωτόκολλο Πόρων του DIENST

- Οι πόροι προσδιορίζονται μέσω των URN τους
- Αιτήσεις υπηρεσιών πόρου
 - Φέρε τα μεταδεδομένα του πόρου
 - Φέρε λίστα με τα μορφότυπα του πόρου
 - Φέρε τον πόρο σε συγκεκριμένο μορφότυπο
 - Φέρε μέρος του πόρου (σελίδα) σε συγκεκριμένο πόρο

Πρωτόκολλο Πόρων DIENST 5.0

- Πιο πολύπλοκο μοντέλο πόρων:
 - Εκδόσεις
 - Ιεραρχικός προσδιορισμός μερών
 - Συνδεδεκά (για πόρους πολλαπλών κομματιών)
- Υπηρεσία αιτήσεων “Δομής”
 - Αποκάλυψη, σε XML, όλης ή σύμπτυξης της δομής ενός πόρου
 - Π.χ., κεφάλαια, παράγραφοι, εικόνες, κλπ.
 - Περιγραφή πολλαπλών όψεων του πόρου
 - Π.χ., βιβλιογραφία, περιεχόμενο, μικρογραφίες (thumbnails)



Βιβλιογραφία

- J. Davis & C. Lagoze, "The Networked Computer Science Technical Report Library," Cornell CS TR96-1595, July 1996 <http://cs-tr.cs.cornell.edu:/Dienst/UI/1.0/Display/ncstrl.cornell/TR96-1595>
- J. R. Davis, D. B. Krafft, & C. Lagoze, "Dienst: Building a Production Technical Report Server," Advances in Digital Libraries, Springer-Verlag, 1995, pp. 211-222

SODA

Έξυπνα Αντικείμενα / Χαζά Αποθετήρια
Smart Objects / Dumb Archives

Χαζά ή Έξυπνα Αντικείμενα

- SODA, dlib (ISSN1082-9873), Vol5, No3, 1999
 - <http://www.dlib.org/dlib/march99/maly/03maly.html>
- Χαζά: Μορφότυπο και Περιεχόμενο
 - Δυνατή η άμεση επεξεργασία
 - Χρειάζεται γνώση του μορφότυπου
- Έξυπνα: Λειτουργίες και Χειρισμός
 - Καλά ορισμένο σύνολο λειτουργιών
- Π.χ. ανάκτηση θέματος από εγγραφή UNIMARC/MARC21

Πλεονεκτήματα – Μειονεκτήματα

- | Χαζά Αντικείμενα | Έξυπνα Αντικείμενα |
|----------------------------------|-------------------------------------|
| • Μεγαλύτερη λειτουργικότητα | • Νέα μορφότυπα |
| • Μικρότερο μέγεθος αντικειμένων | • Ευκολότερη αναβάθμιση αποθετηρίου |
| | • Καθαρότερη χρήση |

Έξυπνα – Χαζά / Αντικείμενα – Αποθετήρια

	Έξυπνα αποθετήρια	Χαζά αποθετήρια
Έξυπνα αντικείμενα	Έξυπνα αντικείμενα σε έξυπνα αποθετήρια (SOSA: Smart objects, Smart Archives)	Έξυπνα αντικείμενα σε χαζά αποθετήρια (SODA: Smart Objects Dumb Archives)
Χαζά αντικείμενα	Χαζά αντικείμενα σε έξυπνα αποθετήρια (DOSA: Dumb Objects Smart Archives) Π.χ. DIENST	Χαζά αντικείμενα σε χαζά αποθετήρια (DODA: Dumb objects Dumb Archives) Π.χ. FTP server

Χαζά → Έξυπνα Αντικείμενα

- Η παραδοσιακή λειτουργικότητα των αποθετηρίων έχει μεταβιβαστεί στα αντικείμενα
- Κάνοντας τα αντικείμενα έξυπνότερα – αυξάνοντας την υπευθυνότητά τους
- Και επιτρέποντας στα αποθετήρια να είναι πιο χαζά – μειώνοντας την υπευθυνότητά τους
- Τα αποθετήρια υπάρχουν για να βοηθούν το χρήστη να εντοπίσει τα αντικείμενα
- Αφού εντοπιστεί ένα αντικείμενο, ο χρήστης αλληλεπιδρά απευθείας με αυτό

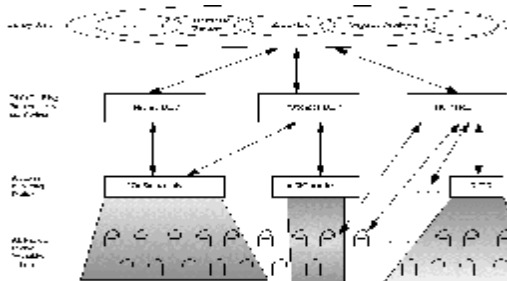
Έξυπνα Αντικείμενα – Άρα ...

- Η συσκευασία δεδομένων είναι ανεξάρτητη του αποθετηρίου
- Υποστηρίζονται όλοι οι σημασιολογικοί και συντακτικοί τύποι δεδομένων
- Πολλά αντικείμενα μπορούν να επικοινωνούν μεταξύ τους
- Αντικείμενα ομαδοποιούνται μαζί λογικά – Αρχειοθετημένα και χειριζόμενα σαν ένα αντικείμενο

Έξυπνα Αντικείμενα – Επειδή ...

- Η πληροφορία γεννιέται σε πολλές μορφές
- Διαφοροποιούμενη από τους (σημασιολογικούς) τύπους (αναφορά, λογισμικό, βίντεο, σύνολο δεδομένων, κλπ)
- Δεδομένης της σημασιολογικής αναπαράστασης, διαφοροποιούνται από την (συντακτική) αναπαράσταση (ps, pdf, doc, κλπ)

Υλοποίηση Μοντέλου SODA



Μοντέλο «Κάδων»

Τι Περιέχουν οι «Κάδοι»

- Οι κάδοι (buckets) συγκεντρώνουν πληροφορία σε ένα ή παραπάνω μορφότυπα σχετικά με μία πνευματική ιδιοκτησία
- Με πολλές μορφές / μορφότυπα των ίδιων δεδομένων
- Με άλλα σχετικά και μη παραδοσιακά δεδομένα (υποστηρικτικό υλικό)

Μοντέλο «Κάδων»

- Τα είδη που ομαδοποιούνται (λογικά) σε ένα κάδο
 - Συλλέγονται
 - Αποθηκεύονται
 - Μεταφέρονται σαν μια μονάδα
- Οι κάδοι είναι «Οντοκεντρικές» συσκευασίες
 - Υλοποιώντας τις λειτουργίες για το περιεχόμενό τους
- Αντικείμενα με ίδια μορφότυπα και δομή μπορούν να έχουν ίδια ορισμένους κάδους

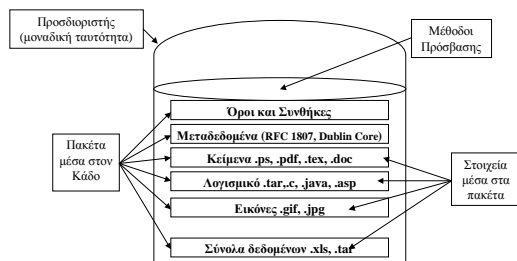
Χαρακτηριστικά Κάδων

- Οι κάδοι έχουν αυθαίρετο μέγεθος
- Με καθολικά μοναδική ταυτότητα
- Οι κάδοι μπορούν και να ξεκινούν ενέργειες, δεν χρειάζεται να περιμένουν παθητικά να κληθούν
- Η επικοινωνία ή αλληλεπίδραση μεταξύ κάδων γίνεται μόνο με καθορισμένες διαδικασίες
- Οι κάδοι μπορεί να καταγράφουν εσωτερικά τις ενέργειές τους («logs»)

Κάδοι Αποτελούμενοι από Πακέτα

- Οι κάδοι έχουν 0 ή παραπάνω συνιστώσες, τα *πακέτα*
- Τα πακέτα μπορεί να αντιστοιχούν σε σημασιολογίες
 - Κείμενο, δεδομένα πειραμάτων, λογισμικό, κλπ.
 - Μεταδεδομένα
 - Όροι και συνθήκες
 - Δείκτες – παραπομπές
- Τα απλά πακέτα μπορεί να ανήκουν σε πολλά είδη

Αρχιτεκτονική Μοντέλου Κάδων



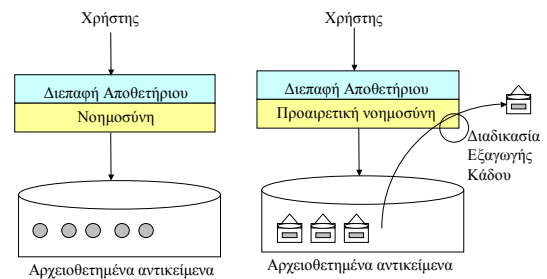
Σύσταση Κάδων

- Οι κάδοι έχουν 0 ή παραπάνω συνιστώσες, τα *πακέτα*
- Κάθε πακέτο περιέχει 1 ή παραπάνω συνιστώσες – *στοιχεία*
- Στοιχείο μπορεί να είναι ένα αρχείο ή σύνδεσμος
- Πακέτο μπορεί να είναι ένας σύνδεσμος σε απομονωμένο κάδο, άλλο πακέτο ή στοιχείο
- Πακέτα και στοιχεία μπορεί να είναι άλλοι κάδοι

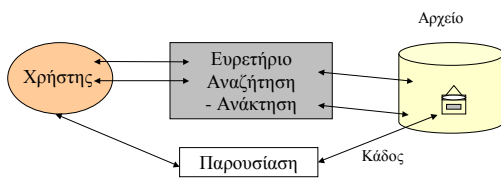
Απαιτήσεις Μοντέλου Κάδων

- Μοναδική ταυτότητα (προσδιοριστικό)
- Απομονωμένα ή πολλαπλά αποθετήρια
- Απομονωμένα – Παγκόσμιος Ιστός με TCP/IP
- Συντονισμός πολλών κάδων – μέσω νοημοσύνης και λειτουργικότητας
- Ξεχωριστοί κάδοι μπορεί να έχουν ειδικούς όρους και συνθήκες

Αποθετήρια Παραδοσιακά και Κάδων



Πρωτόκολλο Κάδων



Εργαλεία Δημιουργίας Κάδων

- Εργαλεία δημιουργίας
 - Μεταδεδομένα
 - Προσθήκη πακέτων
 - Προσθήκη στοιχείων σε πακέτα
 - Επιλογή των σχετικών «συστάδων»
 - Όροι και συνθήκες

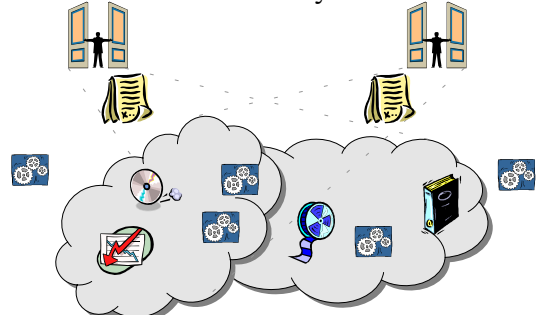
Άλλα Εργαλεία Κάδων

- Εργαλεία διαχείρισης
 - Διεπαφή
 - Επερωτήσεις και ενημέρωση κάδων
- Συσχέτιση κάδων και χρηστών
 - Επιλεκτική διάχυση πληροφορίας (SDI)
 - Εύρεση παρόμοιων εργασιών από διαφορετικούς συγγραφείς
 - «Σάρωση» μεταδεδομένων

Ψηφιακές Βιβλιοθήκες και Πύλες

το STARTS

Οι Ψηφιακές Βιβλιοθήκες σαν Πύλες



Προς Υβριδικές Πύλες

- Οι **Παραδοσιακές** πύλες (π.χ., Yahoo!)
 - Προσφέρουν σύνδεση με περιορισμένη (ή καθόλου) υπευθυνότητα για το τελικό περιεχόμενο
- Οι **Υβριδικές** πύλες
 - Προσφέρουν διαβεβαίωση κάποιου συντονιστικού ρόλου στα συνδεδεμένα αντικείμενα
- **STARTS**: Πρόταση πρωτοκόλλου του Stanford για ανάκτηση και αναζήτηση στο Διαδίκτυο
 - <http://www-db.stanford.edu/~gravano/starts.html>
 - <http://www.cs.cornell.edu/NCSTRL/STARTS/STARTShome.htm>

Το STARTS – Αρχές Σχεδιασμού

- Γνώση των περιορισμών των αυτοσχεδιασμών ανάπτυξης στον Ιστό και στις εμπορικές λύσεις
- Υιοθέτηση προσέγγισης βασισμένη σε υπηρεσίες
 - Ο διαρθρωτός σχεδιασμός αυξάνει την ευελιξία, επεκτασιμότητα, «plug-in/plug-out»
 - Καλά ορισμένες υπηρεσίες με πρωτόκολλα που επιτρέπουν ομοσπονδία και διαλειτουργικότητα
 - Μπορεί να αξιοποιήσει διάφορες τεχνολογίες ή εμπορικό λογισμικό κάτω από τα επίπεδα των υπηρεσιών
- Ακολουθεί τις εξελίξεις στον Ιστό και σε XML και RDF

Το STARTS – Γενικά

STARTS – Stanford Digital Library Project
Π.χ. Alexandria Digital Library

- Πηγές αντικειμένων στο
 - Εσωτερικό δίκτυο
 - Διαδίκτυο
- Περιεχόμενα μορφότυπων
 - Κρυμμένα πίσω από τις διεπαφές αναζήτησης
- Διαφορετικοί αλγόριθμοι και πρωτόκολλα

Αρχιτεκτονική του STARTS

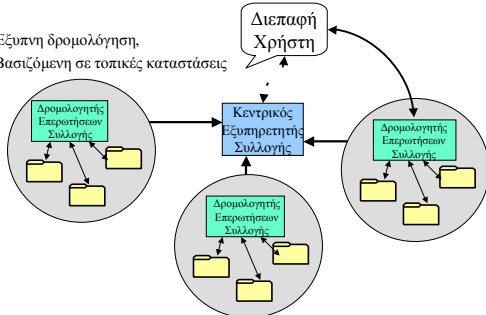
- Σχεδιασμένο για μεγάλο αριθμό πόρων και πηγών
- Είναι γενική μηχανή αναζήτησης
 - Μαζεύει όλη την πληροφορία στο δίκτυο
 - Μαζεύει και μεταδεδομένα για συλλογές
 - Για *μέτα*-αναζήτηση για ανακάλυψη πόρων
 - Δέχεται επερωτήσεις από τους χρήστες, και δίνει αποτελέσματα

Ανακάλυψη Πόρων στο STARTS

- Επερωτήσεις σε πολλαπλές πηγές πόρων
- Επιλογή των καλύτερων πηγών (μικρά σύνολα συλλογών - με *υποσχόμενους* πόρους) για αποτίμηση της επερωτήσεως
- Αποτίμηση της επερωτήσεως σε αυτές τις πηγές
- Συγχώνευση των αποτελεσμάτων των επερωτήσεων από αυτές τις πηγές και αποστολή τους στους χρήστες
- Ανάκτηση των πληροφοριών μέσω συνδέσμων

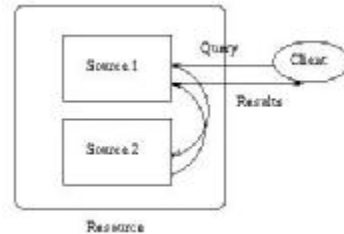
Καταναμημένη Υπηρεσία Συλλογών Ορισμός και Πρόσβαση

Εξυπνη δρομολόγηση,
Βασισμένη σε τοπικές καταστάσεις



Αρχιτεκτονική του STARTS – Σχηματικά

- Ερώτηση πόρου με πολλά μορφότυπα



Μορφότυπα στο STARTS

- Κάθε πόρος αποτελείται από ένα ή παραπάνω μορφότυπα (formats)
- Τα μορφότυπα είναι συλλογές αρχείων
- Τα μορφότυπα μπορεί να είναι μικρά ή μεγάλα
- Εξάγει περιοδικά τη λίστα των μορφότυπων από τους πόρους
- Εξάγει περιοδικά μεταδεδομένα και περιλήψεις περιεχομένου από τα μορφότυπα
- Επερωτήσεις σε ένα μορφότυπο ενός πόρων

Γλώσσα Επερωτήσεων του STARTS – Γενικά Χαρακτηριστικά

- Έκφραση φιλτραρίσματος
 - Λογικής φύσης (boolean)
 - Ορίζει πόρους
- Έκφραση βαθμολόγησης
 - Συνδέει τους πόρους με βαθμολογία
- Καθολικές ρυθμίσεις

Γλώσσα Επερωτήσεων του STARTS – Τελεστές

- L-strings
 - Γλώσσα – χώρα
 - Συμπεριφορά συμβολο-χαρακτήρων (strings)
- Ατομικοί όροι
 - Πεδία
 - Τροποποιητές (modifiers)
- Σύνθετες εκφράσεις φιλτραρίσματος
 - and, or, and-not κλπ
- Σύνθετες εκφράσεις βαθμολόγησης

Συγγώνευση Βαθμολογίας στο STARTS

- Μη κανονικοποιημένη βαθμολογία των πόρων σε κάθε ερώτηση
- Με προσδιοριστικά των μορφότυπων που εμφανίζεται ο πόρος
- Στατιστικά
 - Συχνότητα όρων, Βαρύτητα όρων, Συχνότητα πόρων, Μέγεθος πόρων, Πλήθος πόρων

Εξαγωγή Μεταδεδομένων από Μορφότυπα στο STARTS

- Ιδιότητες ανά μορφότυπο
 - Υποστηριζόμενα πεδία μεταδεδομένων, εύρος βαθμολογίας, υποστήριξη συνδέσεων, ...
- Περίληψη περιεχομένου του μορφότυπου, από
 - Λίστα λέξεων που εμφανίζονται στο μορφότυπο
 - Στατιστικά για κάθε λέξη στη λίστα
 - Συνολικός αριθμός πόρων στη λίστα
 - ...

Μελετάμε Αρχιτεκτονικές

- Για να επιλέξουμε **λογισμικό** ψηφιακών βιβλιοθηκών
- Για να ρυθμίσουμε καλύτερα λογισμικό και επιλογές σχετικές με τα **δεδομένα**
- Για να κάνουμε καλύτερη **χρήση** των δυνατοτήτων των ήδη εγκατεστημένων ψηφιακών βιβλιοθηκών

Κεντρικοποίηση ή Κατανομή

- Κεντρικοποίηση δημιουργίας (και χρήσης)
- Κατανομή χρήσης (και δημιουργίας)
- Κατανομή δημιουργίας & κεντρικοποίηση χρήσης
 - Με συγκομιδή
- Στα μεταδεδομένα – στα δεδομένα;
- Τρόποι ευρετηρίασης – καταναμημένης πληροφορίας

Πλεονεκτήματα – Μειονεκτήματα Κεντρικοποίηση – Κατανομή

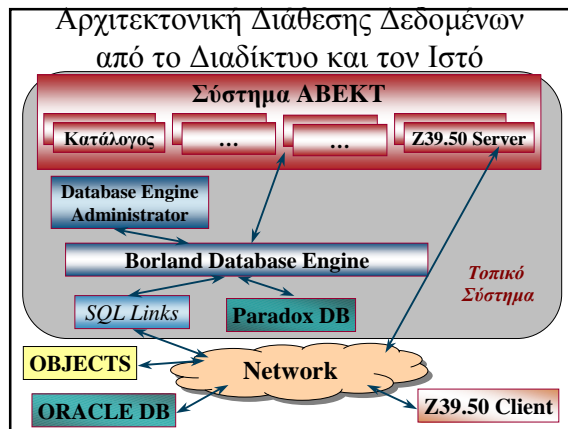
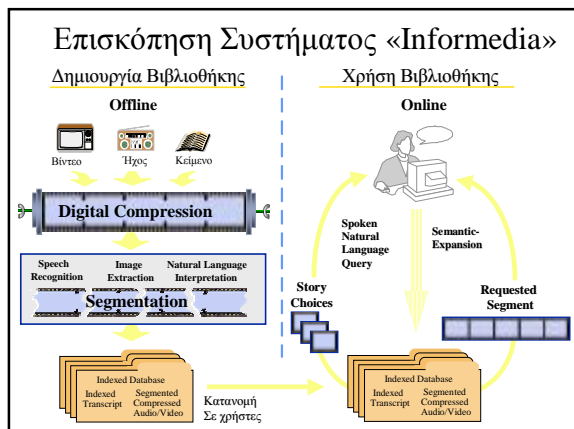
- Καταναμημένη δημιουργία
 - Αυτονομία των επιμέρους βιβλιοθηκών
- Καταναμημένη χρήση
 - Τρόπος ιδιαίτερης προβολής του διαθέτη
 - Εκμετάλλευση των επιμέρους χαρακτηριστικών
 - Αποφεύγουμε περιττές μετατροπές / ομογενοποιήσεις
 - Ανεξάρτητες πολιτικές πρόσβασης ανά πηγή
 - Και διαθεσιμότητα: νομική ή τεχνολογική
 - Δυνατότητα ορισμού προσωπικών συνόλων πηγών
 - Συνεχώς αυξανόμενο πλήθος πηγών
 - Συνεχώς άμεσα ενημερωμένα δεδομένα (π.χ. δανεισμού)
 - Και παροδική μη διαθεσιμότητα (ή καθυστέρηση) πρόσβασης

Αρχιτεκτονικές Τάσεις

- Συνύπαρξη πολλών μοντέλων
 - Ανάλογα με τις απαιτήσεις
 - Αλλά δεν θα συνεργάζονται μεταξύ τους (εύκολα)
- Συγκομιδή μεταδεδομένων
 - Για κεντρικοποιημένη χρήση
 - Αλλά πρέπει να λύσει τα προβλήματα ανομοιογένειας
 - Αρκεί offline

Σημερινά Συστήματα

- SODA, STARTS
 - Είναι μοντέλα, όχι εφαρμογές για χρήση
- NCSTRL/DIENST
 - Σαν εφαρμογή έχει ξεπεραστεί
 - Αλλά έχει ανοικτά πρωτόκολλα
- Οι ιδέες τους έχουν εφαρμοστεί σε νεότερες εφαρμογές
 - GreenStone
 - DSpace



Προβληματισμοί

- Ομοσπονδία – προϋποθέσεις και πλεονεκτήματα
- Προϋποθέσεις κεντροποίησης
 - Δεδομένων
 - Υπηρεσιών
- Ανοικτή αρχιτεκτονική ↔ επικράτηση στην αγορά
- Βαθμός επιθυμητής κατανομής περιεχομένου

Στρατηγικές Οργάνωσης και Τάσεις

Στρατηγική της LC

The Library should selectively adopt the portal model for targeted program areas. By creating links from the Library's Web site, this approach would make available the ever-increasing body of research materials distributed across the Internet. The Library would be responsible for carefully selecting and arranging for access to licensed commercial resources for its users, but it would not house local copies of materials or assume responsibility for long-term preservation.

LC21: Digital Strategy for the Library of Congress, page 5

Γιατί Κατανεμημένες Συλλογές;

- Η κλίμακα που παρέχει ο Παγκόσμιος Ιστός
- Η επικράτηση νέων μοντέλων δημοσίευσης και μεσολάβησης
- Η αυξανόμενη πολυπλοκότητα της χορήγησης άδειας και της διαχείρισης της πρόσβασης
- Η πολυπλοκότητα και η δυναμική φύση του περιεχομένου

Πόσες Ψηφιακές Βιβλιοθήκες

Πότε έχουμε

- Μία, αυτόνομη ψηφιακή βιβλιοθήκη
- Κατανεμημένη ψηφιακή βιβλιοθήκη
- Ομόσπονδες (συνεργαζόμενες) ψηφιακές βιβλιοθήκες
- Διαφορετικές συλλογές σε ψηφιακές βιβλιοθήκες
 - Διαφορετικό περιεχόμενο
 - Διαφορετική πολιτική πρόσβασης
 - Διαφορετικό τρόπο συντήρησης

Καταγραφή Στατιστικών Στοιχείων

Πολύτιμη για καλύτερη χρήση, αξιοπιστία, ...

- Ανακτήσεις σελίδων (hits)
- Πόσοι χρήστες διάβασαν το άρθρο;
- Πρόσβαση στη σελίδα περιεχομένων / πρώτη σελ.
- Οι χρήστες ταυτοποιούνται;
 - Προσωπικά δεδομένα;
 - Τότε, παραλλαγή των μετρήσεων
- Μακροχρόνια αποθήκευση τέτοιων στοιχείων;

Στρατηγικές Οργάνωσης και Τάσεις

Στρατηγική της LC

The Library should selectively adopt the portal model for targeted program areas. By creating links from the Library's Web site, this approach would make available the ever-increasing body of research materials distributed across the Internet. The Library would be responsible for carefully selecting and arranging for access to licensed commercial resources for its users, but it would not house local copies of materials or assume responsibility for long-term preservation.

LC21: Digital Strategy for the Library of Congress, page 5

Στρατηγική της LC



Committee on Information Strategy for the Library of Congress, *LC21: A Digital Strategy for the Library of Congress (2000)*: National Academy Press, Washington, DC, 2000, page 5

Προς μια Εικονική Ζώνη Ελέγχου

Some of the most fundamental aspects of library operations entail the existence of a border, across which objects of information are transferred and maintained. Such a parameter, demarcating a single, distributed digital library (the "control zone"), needs to be created and managed by the academic library community at the earliest opportunity.

*Ross Atkinson
Library Quarterly, 1996*

Γιατί Κατανεμημένες Συλλογές;

- Η κλίμακα του WWW
- Η επικράτηση νέων μοντέλων δημοσίευσης και μεσολάβησης
- Η αυξανόμενη πολυπλοκότητα της χορήγησης άδειας και της διαχείρισης της πρόσβασης
- Η πολυπλοκότητα και η δυναμική φύση του περιεχομένου

Πόσες Ψηφιακές Βιβλιοθήκες

Πότε έχουμε

- Μία, αυτόνομη ψηφιακή βιβλιοθήκη
- Κατανεμημένη ψηφιακή βιβλιοθήκη
- Ομόσπονδες (συνεργαζόμενες) ΨΒ
- Διαφορετικές συλλογές σε ΨΒ
 - Διαφορετικό περιεχόμενο
 - Διαφορετική πολιτική πρόσβασης
 - Διαφορετικό τρόπο συντήρησης

Αξιολόγηση Ψηφιακών Βιβλιοθηκών

Αξιολόγηση Ψηφιακών Βιβλιοθηκών

- Αξιολόγηση και σύγκριση στόχων Ψηφιακών Βιβλιοθηκών
 - Τι σκοπεύει να καλύψει, συγκρινόμενο με το τι κάνουν οι άλλες (Ψηφιακές) Βιβλιοθήκες
- Αξιολόγηση του κατά πόσο ικανοποιούνται οι στόχοι αυτοί, σε τομείς
 - Δεδομένα, μεταδεδομένα, ομάδα χρηστών και κάλυψη
 - Αντικείμενα, συλλογές, τεχνολογία, διαδικασίες, συμπεριφορά χρηστών

Επιδιωκόμενα Δεδομένα και Μεταδεδομένα

- Ποιότητα των επιδιωκόμενων δεδομένων
 - «Λευκή» ή «Γκριζα» βιβλιογραφία
- Λεπτομέρεια μεταδεδομένων
 - Πεδία μεταδεδομένων και κανόνες
 - Χρήση καθιερωμένων όρων, συστημάτων θεματικής ταξινόμησης, θησαυρών
 - Λειτουργικότητα των μεταδεδομένων
 - Ευρετηρίαση και σημεία πρόσβασης
- Ομάδα επιδιωκόμενων χρηστών, όπως
 - Εσωτερικοί χρήστες, ερευνητές
 - Χρήστες σε συγκεκριμένες γεωγραφικές περιοχές

Επιδιωκόμενη Κάλυψη

- Η κάλυψη των δεδομένων – τι θέματα σχεδιάζεται να αντιμετωπίσει, όπως:
 - Θεματική περιοχή
 - Γεωγραφική περιοχή
 - Χρονική κάλυψη
 - Περιορισμοί πρόσβασης:
 - Δημόσιο υλικό
 - Ιδιωτικό υλικό
 - Υλικό προσφερόμενο κάτω από ειδικές συνθήκες (π.χ. ανάλογα με την ταυτότητα του χρήστη, συνδρομή, με πληρωμή ανά χρήση, ...)

Μεμονωμένα Αντικείμενα (σε Δεδομένα ή Μεταδεδομένα)

- Διεθνοποίηση και πολυγλωσσικότητα σε κείμενο
- Τα μορφότυπα του περιεχομένου
- Τα πρότυπα που ακολουθεί το περιεχόμενο
- Διαλειτουργικότητα
- Πληρότητα του περιεχομένου
- Λειτουργίες που προσφέρει το περιεχόμενο
- Αποδοχή – χειρισμός εναλλακτικών μορφότυπων
- Εσωτερική δομή του περιεχομένου

Περιεχόμενο Ολόκληρης της Συλλογής

- Λειτουργίες που προσφέρονται στο περιεχόμενο, όπως συσχετίσεις με σχολιασμό και συστάσεις
- Λειτουργίες που προσφέρονται αθροιστικά στο περιεχόμενο, όπως αναζήτηση, ξεφύλλισμα και στατιστικά
- Υποστήριξη για συνεργατική δουλειά στα αντικείμενα
- Διαλειτουργικότητα της συλλογής με άλλες συλλογές
- Το μέγεθος της συλλογής – τι μέρος των αντικειμένων είναι διαθέσιμα

Τεχνολογία (Αόρατη στους Χρήστες)

- Η αρχιτεκτονική, βαθμός κατανομής
- Αρχιτεκτονική και οργάνωση του αποθετηρίου
- Πρότυπα που χρησιμοποιούνται στις τεχνολογίες
- Διαθεσιμότητα και επαναχρησιμοποίηση λογισμικού
- Διήθηση (filtering) πληροφορίας
- Εξόρυξη κειμένου
- Απόδοση και αποδοτικότητα
- Μέθοδοι και λειτουργικότητα συντήρησης αντικειμένων

Διαδικασίες της Ψηφιακής Βιβλιοθήκης

- Ικανότητα και ροή εργασιών νέων προσκτήσεων
- Βαθμός αυτοματισμού εργασιών
- Διαδικασίες διατήρησης, συχνότητα ενημέρωσης
- Διαδικασίες διαχείρισης χρηστών
- Διαδικασίες κρίσης και ελέγχου ποιότητας
- Πρότυπα των διαδικασιών διαχείρισης
- Διαχείριση δικαιωμάτων
- Έλεγχος του αποθετηρίου

Συμπεριφορά και Ικανοποίηση των Χρηστών

- Διεπαφή, και η λειτουργικότητά της
 - Τοπικοποίηση και διεθνοποίηση της διεπαφής
 - Υποστήριξη για διεπαφή γραφική και με εντολές
- Εξατομίκευση της διεπαφής
- Φιλικότητα, με λογικές διαδικασίες, ευκολία μάθησης και χρήσης
- Αποτελεσματικότητα στα αποτελέσματα, και ικανοποίηση του χρήστη
- Πληροφορίες και στατιστικά χρήσης της Ψηφιακής Βιβλιοθήκης
 - Δυναμικά
 - Βασίζόμενα σε ανάλυση «log»

Σύγκριση Λογισμικού Ψηφιακών Βιβλιοθηκών

- Για Ψηφιακές Βιβλιοθήκες
 - DIENST
 - Z39.50 + OPAC + URL
 - Harvesting – Open Archive Initiative
 - Μηχανές αναζήτησης (π.χ. Google, AltaVista)
 - RDBMS
 - Ιδιογενές (proprietary) σύστημα
 - XML/XQL
 - GreenStone / DSpace

Κριτήρια Σύγκρισης Λογισμικού Ψηφιακών Βιβλιοθηκών (1)

- Μεταδεδομένα και λεπτομέρειες στην αναζήτηση
- Ανάκτηση και παρουσίαση πληροφορίας
- Τρόποι ιεράρχησης – βαθμολόγησης αποτελεσμάτων
- Ευκολία εγκατάστασης και ρύθμισης
- Προσαρμοστικότητα της διεπαφής
- Είναι γενικής χρήσεως (για νέες συλλογές)
- Χειρισμός πολυγλωσσικών δεδομένων
- Πολυγλωσσική διεπαφή

Κριτήρια Σύγκρισης Λογισμικού Ψηφιακών Βιβλιοθηκών (2)

- Κατανεμημένη λειτουργία
- Ενημέρωση δεδομένων (κεντροποιημένη ή κατανεμημένη)
- Συμπεριφορά σε προβλήματα δικτύου
- Δυνατότητα λειτουργίας σε ομοσπονδία (κεντροποιημένη ή κατανεμημένη)
- Ταυτόχρονη αναζήτηση σε πολλές πηγές
- Ευκολία υλοποίησης και ελεύθερος κώδικας
- Ανοικτά πρωτόκολλα
- Εξάρτηση από υλικό και Λειτουργικό Σύστημα

Αυτοματοποίηση – Εξέλιξη και Ρόλοι

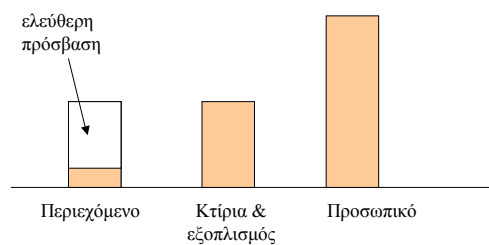
Επισκόπηση Θεμάτων

- Αυτοματοποίηση Διαδικασιών
 - Για χειρισμό μεγάλου όγκου υλικού
- Εξέλιξη και Προοπτικές στην Ανακάλυψη της Πληροφορίας
 - Πόσο βοηθούν ή είναι αναγκαίες οι παραδοσιακές μέθοδοι οργάνωσης και ανακάλυψης της πληροφορίας;
- Ο Ρόλος του Επιστήμονα της Πληροφορίας

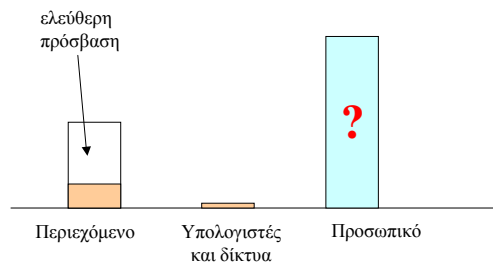
Αυτοματοποίηση Διαδικασιών

Για χειρισμό μεγάλου όγκου υλικού

Το Κόστος της Παραδοσιακής Βιβλιοθήκης



Το Κόστος της Ψηφιακής Βιβλιοθήκης



Το Κόστος των Βιβλιοθηκών και Δημοσιεύσεων

- Το κόστος προσωπικού είναι κυρίαρχο στις βιβλιοθήκες και δημοσιεύσεις
- Μεγάλες μειώσεις στο κόστος απαιτούν διαφορετική χρήση του προσωπικού.
- Μπορούμε να φτιάξουμε βιβλιοθήκες υψηλής ποιότητας με πολύ χαμηλότερο κόστος, με δημιουργική χρήση της τεχνολογίας;

Ικανότητες Βιβλιοθηκονομίας

- Οι άνθρωποι είναι ικανότεροι για κρίση, κατανόηση, διάκριση, κλπ:
 - Επιλογή
 - Κατάλογογράφηση, ευρετηρίαση
 - Αναζήτηση πληροφορίας
 - Αξιολόγηση πληροφορίας
 - Υπηρεσία αναφοράς (reference)
- Μπορούν οι υπολογιστές να παρέχουν ισοδύναμες ποιοτικά υπηρεσίες;

Αυτοματοποίηση στην Ψηφιακή Οργάνωση της Πληροφορίας

Και στην Επιστήμη της Πληροφορίας

Οι διαδικασίες τώρα μπορεί να είναι

- Χειροκίνητες
 - Τις διεκπεραιώνουν πάντα οι άνθρωποι
- Ημιαυτόματες
 - Χρειάζονται συνδυασμό υπολογιστή και ανθρώπου
- Αυτόματες
 - Επαναλαμβάνονται φθηνά, με νεότερες τεχνολογίες

Παραδείγματα Αυτοματοποίησης

- Αυτόματη ευρετηρίαση Lycos, Infoseek, Altavista, Google, ...
- Αντιστοίχιση ερωτήσεων Διανυσματικές μέθοδοι
- Ιεράρχηση Lycos, Google, ...
- Σύνδεση παραπομπών OpenURL
- Δημιουργία συλλογών CiteSeer
- Εξαγωγή μεταδεδομένων Informedia (video), ...

Ο Ρόλος της Υπολογιστικής Ισχύος

- Σύμφωνα με το νόμο του Moore
 - Η υπολογιστική ισχύς διπλασιάζεται κάθε 18 μήνες
 - Αυξάνει 100 φορές σε 10 χρόνια
 - Αυξάνει 10000 φορές σε 20 χρόνια!
- Απλοί αλγόριθμοι αλλά με τεράστια υπολογιστική ισχύς μπορούν συχνά να αντικαταστήσουν την ανθρώπινη δημιουργική (έξυπνη) δραστηριότητα

Παράδειγμα Εκμετάλλευσης Υπολογιστικής Ισχύος

- Παράδειγμα: η δημιουργία του προγράμματος πρωταθλητή στο σκάκι (Deep Thought → Deep Blue).
 - Μέσοι παίκτες σκακιού
 - Απλοί αλγόριθμοι δενδρικής αναζήτησης
 - Πάρα πολύ γρήγορο υπολογιστικό υλικό.
- ... αλλά παρόλα αυτά, επιδιώκουμε καλές παρτίδες σκάκι με ανθρώπινους αντιπάλους

Το Κόστος της Αυτοματοποίησης

- Η εταιρία του Google (Μάρτιος 2000)
 - Δέχεται 5,500,000 αναζητήσεις καθημερινά
 - Απασχολεί 85 ανθρώπους (οι μισοί τεχνικοί, 14 με διδακτορικό σε υπολογιστικό αντικείμενο)
 - Διατηρεί 2,500 PCs με Linux και 80Tb αποθηκευτικό χώρο
- Η Internet Archive (Μάρτιος 2000)
 - Απασχολεί 7 ανθρώπους με υποστήριξη από την Alexa

Αυτοματοποιήσιμες Υπηρεσίες

Αυτοματοποιήσιμες υπηρεσίες σε πλήρη κείμενα

- Διαδικασίες καταχώρησης προσκλήσεων
- Ευρετηρίαση περιεχομένου
- Κατηγοριοποίηση περιεχομένου
- Εξαγωγή κειμένου και δομής
- Εξαγωγή εμφανίσιμης: γραμματοσειράς, μεγέθους γραμμμάτων, ...
- Περίληψη περιεχομένου

Μη Αυτοματοποιήσιμες (Χειροποίητες) Υπηρεσίες

Η Κωδικοποίηση της Πληροφορίας (π.χ. καταλογογράφηση – κατηγοριοποίηση), παρότι βασική και επίπονη εργασία, δεν γίνεται αυτόματα σε περιπτώσεις με λίγα ή ελλιπή στοιχεία. Π.χ.

- Σύντομες προτάσεις για χρηματοδότηση, χωρίς πλήρη περιγραφικά κείμενα
- Ανταλλαγές μέρους μόνο των πληροφοριών
 - Για διαπραγματεύσεις πάνω σε ανταλλαγή πληροφοριών
 - Για προϊόντα προς πώληση (π.χ. ταινίες)
- ...

Εξέλιξη και Προοπτικές στην Ανακάλυψη της Πληροφορίας

Πόσο βοηθούν ή είναι αναγκαίες οι παραδοσιακές μέθοδοι οργάνωσης και ανακάλυψης της πληροφορίας;

Πρόκληση: Ανακάλυψη Πληροφορίας

- Τα παραδοσιακά συστήματα ανακάλυψης πληροφορίας χρησιμοποιούν καταλόγους ή ευρετήρια δημιουργημένα από καλά εκπαιδευμένο προσωπικό, και άρα είναι ακριβά
- Ποιες είναι οι νέες (αυτοματοποιημένες, κατά προτίμηση) τάσεις στην ανακάλυψη της πληροφορίας; Πως συγκρίνονται τα αυτοματοποιημένα συστήματα ευρετηρίασης με τους συμβατικούς καταλόγους και ευρετήρια;

Εξέλιξη της Αναζήτησης

- 1970-1995: ψηφιακά ευρετήρια για έντυπες εκδόσεις:
 - OPACs
 - Υπηρεσίες περίληψης και ευρετηρίασης (INSPEC, Chemical Abstracts, Medline, etc.)
 - Science Citation Index
- 1993-...: ψηφιακά ευρετήρια για εκδόσεις πλήρους (ψηφιακού) κειμένου:
 - Ιστός (Infoseek, Google, κλπ)

Ψηφιακά Ευρετήρια για Έντυπες Εκδόσεις (1970-1995)

- Περιορισμοί του περιβάλλοντος εφαρμογής:
 - Το πλήρες περιεχόμενο δεν ήταν ψηφιακό
 - Ο υπολογιστικός χρόνος ήταν ακριβός (\$300/ώρα)
- Μέθοδος εφαρμογής:
 - Προσεκτική επιλογή υλικού για ευρετηρίαση ή καταλογογράφηση
 - Εγγραφές με μεταδεδωμένα υψηλής ποιότητας: κανόνες καταλογογράφησης, ελεγχόμενα λεξικά, καθιερωμένοι όροι, κλπ.
 - Εκπαιδευμένο προσωπικό: σύνταξη περίπλοκων ερωτήσεων, κλπ
- Αποτέλεσμα: μεγάλο κόστος

Μελέτες της Αποτελεσματικότητας της Αναζήτησης

- **The Cranfield Experiments**, Cyril W. Cleverdon, Cranfield College of Aeronautics, 1957 -1968
- **SMART System**, Gerald Salton, Cornell University, 1964-1988
- **TREC**, Donna Harman, National Institute of Standards and Technology (NIST), 1992 –

Παράδειγμα: τα Δεδομένα στο Πείραμα II στο Cranfield

Ερώτημα προς διερεύνηση: *Αυξάνουν την αποτελεσματικότητα τα χειροποίητα μεταδεδομένα;*

- Συγκριτική αποδοτικότητα συστημάτων ευρετηρίασης σε 18000 δημοσιεύσεις και αναφορές αεροναυτικής
- Πολύ προσεκτική ευρετηρίαση:
 - Χειροκίνητη προετοιμασία 4 ευρετηρίων (UDC, αλφαβητική θεματική ευρετηρίαση, σχήματα, συντεταγμένες)
- Αποτέλεσμα: παρόμοια ποιότητα (εύρεσης)

Αυτοματισμός Μεταδεδομένων

- Τεράστια χρηματικά ποσά ξοδεύονται σε μεταδεδομένα (και εκπαίδευση, πρότυπα, συνέδρια, ... μεταδεδομένων)
- Και η κυριότερη υπηρεσία ανακάλυψης πληροφορίας για ψηφιακά κείμενα (ο Ιστός) είναι εξ' ολοκλήρου αυτοματοποιημένη!
 - Μήπως τα ευρετήρια και οι κατάλογοι (που δημιουργούνται από ειδικούς επαγγελματίες) δημιουργούνται για να αντιμετωπίσουμε τους περιορισμούς του παρελθόντος;
- Με το πλήρες κείμενο ψηφιακό, γίνεται πολύ εύκολα:
 - Αυτόματη ευρετηρίαση
 - Περιήγηση από τους χρήστες

Κριτήρια Παραδοσιακής Ιεράρχησης Αποτελεσμάτων

- Οι περισσότερες μηχανές αναζήτησης ιεραρχούν με πρώτα τα αποτελέσματα που
 - Περιλαμβάνουν μεταδεδομένα με τους ζητούμενους όρους
 - Περιλαμβάνουν εμφανίσεις των πιο σπάνιων λέξεων της γλώσσας
 - Έχουν πολλαπλές εμφανίσεις των ζητούμενων όρων
- Η προσέγγιση αυτή αντικατοπτρίζει το πώς γινόταν η αντίστοιχη χειροκίνητη διαδικασία.

Εμφανίζονται Νέοι Τρόποι Ιεράρχησης Αποτελεσμάτων

Που δεν ήταν δυνατό να γίνουν χειροκίνητα

- Το Google, <http://www.google.com>
 - Ιεράρχηση μεγαλύτερη αν περισσότερες σελίδες έχουν σύνδεση σε αυτή
 - Συνδέσεις από σελίδες με μεγαλύτερη ιεράρχηση έχουν μεγαλύτερη βαρύτητα
- Οι μετα-μηχανές αναζήτησης
 - Συνδυάζουν τις ιεραρχήσεις των επιμέρους μηχανών που καλούν (με μη προφανή τρόπο).

Αδυναμίες της Αναζήτησης στον Ιστό

Που είναι ένα μεγάλο μέρος εφαρμογής αναζήτησης

- Η επιλογή πόρων είναι αυθαίρετη
 - Τα ευρετήρια δεν είναι «εκλεπτυσμένα»
 - Δεν υπάρχει καθιέρωση όρων
 - Η ανίχνευση διπλο-εγγράφων είναι αβέβαιη
 - Αμφίβολη ακρίβεια στην ανακάλυψη.
- Αλλά για πολλούς χρήστες ο Ιστός έχει αντικαταστήσει ακριβές υπηρεσίες αναζήτησης με μεταδεδομένα δημιουργούμενα χειροποίητα.

Ανακάλυψη της Πληροφορίας: 1991 και 2001

	1991	2001
Περιεχόμενο	έντυπο	online
Υπολογισμοί	ακριβοί	φθηνοί
Επιλογή περιεχομένου	επιλεκτική	εξονυχιστική
Δημιουργία ευρετηρίου	χειροποίητη	αυτόματη
Συχνότητα	εφάπαξ	μηνιαία
Λεξιλόγιο / Λεξικό	ελεγχόμενο	μη ελεγχόμενο
Αποτελέσματα ερωτήσεων	boolean	ιεραρχημένα
Χρήστες	εκπαιδευμένοι	αδασείς

Επιπτώσεις των Ψηφιακών Τεχνολογιών

- Έρχεται Αυτοματισμός ή Ολοκληρωτική αλλαγή;
 - Στον κύκλο της πληροφορίας κατά την δημοσίευση
 - Στα συστήματα γνώσης
- Π.χ. στα (ηλεκτρονικά) περιοδικά:
 - Παροχή των άρθρων ηλεκτρονικά: άμεσα, ...
 - Νέα μοντέλα (ανοικτής) πρόσβασης (π.χ. BioMed Central)
- Η ανακάλυψη της πληροφορίας σίγουρα αλλάζει
- Τα συστήματα γνώσης θα παραμείνουν τα ίδια;
 - Και τα επεκτείνουμε – ή τα χρειαζόμαστε μόνο για να αντιμετωπίσουμε τους περιορισμούς του παρελθόντος;

Σημερινά «Ψηφιακά» Συστήματα Γνώσης

- Λίστες όρων
 - Αρχεία καθιερωμένων όρων, γλωσσάρια, λεξικά, γεωγραφικά λεξικά (Authority Files)
- Κατηγοριοποιήσεις
 - Θεματικές επικεφαλίδες (Subject Headings)
 - Συστήματα ταξινόμησης/κατηγοριοποίησης – ταξινομίες (Classification Systems)
- Λίστες συσχετίσεων (τώρα είναι υλοποιήσιμα)
 - Θησαυροί (Thesaurus)
 - Σημασιολογικά δίκτυα (Semantic Networks)
 - Οντολογίες (Ontologies)

Τρέχουσες Τάσεις Πληροφόρησης

- Απευθυνόμαστε σε κοινούς ανθρώπους – όχι σε εξειδικευμένους επιστήμονες της αναζήτησης
- Διατίθεται λιγότερο κόστος για δημιουργία και χρήση της πληροφορίας
- Χρειάζεται μικρότερη λεπτομέρεια σε περιγραφή και ανάκτηση
- Απαιτείται διαλειτουργικότητα με άλλες πληροφορίες

Το Μέλλον στα Συστήματα Γνώσης

- Τα «Αρχεία Καθιερωμένων Όρων», η «Θεματική Ταξινόμηση» ...είναι θεμελιώδη εργαλεία για περιγραφή συμβατικών τεκμηρίων
- Χρειάζονται στο απώτερο μέλλον;
 - Η αυτοματοποίηση συνεπάγεται νέες διαδικασίες
 - Έστω και αν εφαρμόζεται μόνο πάνω στην περιληψη!
- Προσωρινά:
 - κείμενο ⇒ όροι ⇒ καθιερωμένοι όροι ⇒ θέματα ⇒ ...

Προβληματισμοί Αυτοματοποίησης σε Ψηφιακούς Πόρους

- Η αυτοματοποίηση γίνεται με ψηφιακή επεξεργασία
 - Αν έχουμε μόνο μεταδομένα και όχι το ψηφιακό τεκμήριο;
- Είναι ποιοτικά ισοδύναμη με χειροποίητες διαδικασίες;
 - Εξάγει και τα διαφορετικά σημεία πρόσβασης (π.χ. δημιουργός, κατηγοριοποίηση) και μεθόδους πρόσβασης (π.χ. συνώνυμα);
- Είναι απαραίτητη για καλά αποτελέσματα;
 - Αν δεν υπάρχει, πώς υποκαθίσταται στην αναζήτηση;
- Είναι οικονομικά συμφέρουσα;
 - Είναι σαν χειροποίητη ή σαν βιομηχανική παραγωγή;
- Εξασφαλίζουμε διαλειτουργικότητα με τα συμβατικά τεκμήρια (περιγραμμένα με χειροποίητες διαδικασίες);

Διαφαίνονται Νέοι Τρόποι Επεξεργασίας

- Μέθοδοι Έμπειρων συστημάτων / Τεχνητής νοημοσύνης
- Που εφαρμόζονται πάνω στα ψηφιακά τεκμήρια
 - Ή σε μέρος τους: περιλήψεις, τίτλους, μεταδεδομένα, ...
- Με διεπαφές περίπλοκες αλλά με λειτουργικότητα
- Που χρησιμοποιούν (εσωτερικά) τα συστήματα γνώσης
 - Όπως Θησαυροί, Οντολογίες, ...
 - Χωρίς να είναι ορατά στον τελικό χρήστη

Αυτοματοποίηση Ανάλογα με τα Διαθέσιμα Κονδύλια

- Όταν υπάρχουν διαθέσιμα κονδύλια
 - Οι ερευνητικές βιβλιοθήκες, με πληροφοριακές υπηρεσίες με εμπορικά πρότυπα, παρέχουν άριστες υπηρεσίες με μεγάλο κόστος σε λίγους εκλεκτούς
 - Οι αυτοματοποιημένες ψηφιακές βιβλιοθήκες είναι μακριά από το να παρέχουν τις εξατομικευμένες υπηρεσίες αναφοράς που είναι διαθέσιμες σε καλά οργανωμένα πανεπιστήμια
- Αλλά με ανεπαρκή κονδύλια (ως συνήθως!)
 - Η λύση είναι η αυτοματοποίηση ...

Η Βιβλιοθήκη «Model T»

Το Ford Model T, με τη μαζική παραγωγή, έφερε το αμάξι στις μάζες ...

- Οι αυτοματοποιημένες ψηφιακές βιβλιοθήκες, με υλικό ελεύθερης πρόσβασης, μπορούν ήδη να παρέχουν καλές υπηρεσίες σε χαμηλό κόστος
- Στο μέλλον, οι ψηφιακές βιβλιοθήκες με ελεύθερη πρόσβαση θα μπορούν να παρέχουν εξειδικευμένη ιατρική, νομική, ακαδημαϊκή, επιστημονική και άλλη πληροφόρηση σε όλους!

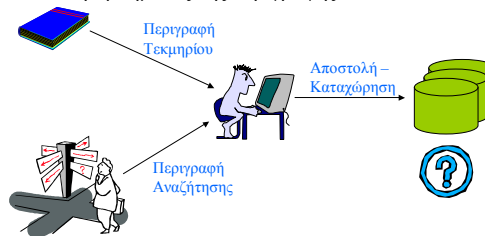
Ο Ρόλος του Επιστήμονα της Πληροφορίας

Ρόλοι κατά την Ψηφιακή Εξέλιξη

- Ο Ρόλος της Βιβλιοθήκης εξαρτάται
 - Από το ρόλο που θέλει και επιζητεί
 - Από τον τρόπο που τον κατοχυρώνει στην συνείδηση των χρηστών της
- Ο Επιστήμονα της Πληροφορίας
 - Αναπροσαρμόζει το ρόλο του για να ανταποκριθεί στις τρέχουσες ανάγκες του χώρου του

Καταχώρηση Περιγραφής

- Καταλογογράφηση και αναζήτηση – είναι όψεις του προβλήματος της περιγραφής



Κατηγοριοποίηση ↔ Αναζήτηση

- Είναι όψεις του ίδιου προβλήματος
 - Της κωδικοποίησης/περιγραφής της πληροφορίας
 - Ενός τεκμηρίου – πόρου
 - Ενός ερωτήματος
- Γίνεται πιο φανερό στον ψηφιακό κόσμο
 - Όχι μόνο επειδή γίνονται καλύτερα από τον ίδιο άνθρωπο
- Μπορεί να αποθηκευτούν (καταχώρηση)
 - Σε ένα κατάλογο τεκμηρίων – πόρων
 - Στο προφίλ χρήστη, σε σύστημα επιφυλακής (alert)
- Και να αναζητηθούν – αντίστοιχα
- Αλλά τα **ερωτήματα** δεν μπορεί να είναι πάντα τυποποιημένα, προαποφασισμένα και αποθηκευμένα!

Κωδικοποίηση της Πληροφορίας

- Η κωδικοποίηση της πληροφορίας (π.χ. καταλογογράφηση – κατηγοριοποίηση) είναι επίπονη εργασία – και δουλειά του Επιστήμονα της Πληροφορίας
 - Αλλά γίνεται σε όλο και περισσότερες περιπτώσεις
 - Περιγραφή πόρων (τεκμηρίων) – και αναζητήσεις
 - Προτάσεις για χρηματοδότηση (προς αξιολόγηση, ...)
 - Προφίλ ανθρώπων (π.χ. σε φόρμες ερωτηματολογίων)
 - Στατιστικές μελέτες (και επεξεργασία αποτελεσμάτων)
 - ... μικρές (κυρίως) και μη καθολικές προσπάθειες
 - Επειδή διευκολύνει τη διαχείριση της πληροφορίας

Πότε Κάνουμε Χειροποίητα Κωδικοποίηση της Πληροφορίας;

- Δεν ανταγωνιζόμαστε τις αυτοματοποιημένες υπηρεσίες, για ότι μπορεί να γίνει αυτόματα
- Αλλά επιβλέπουμε και ρυθμίζουμε τη διαδικασία
- Η κωδικοποίηση της πληροφορίας γίνεται χειροποίητα όταν
- Απαιτείται αυξημένη ποιότητα
 - Και υπάρχουν τα διαθέσιμα κονδύλια
 - Δεν μπορεί να γίνει αυτόματα
 - Λόγω ανεπαρκών δεδομένων ή τεχνολογιών

Εξέλιξη Αυτοκινήτων και Οδηγών

- Αυτοκίνητα: πρόβλεψη (του 1900): ≤100000, γιατί τόσοι μπορούσαν να πληρώνουν οδηγούς!
- Σήμερα είναι εύκολο να οδηγούμε αυτοκίνητο
- Και έτσι έχουμε καλύτερες μεταφορές
 - Και περισσότερα αυτοκίνητα
- Και έχουμε πολλούς κλάδους επαγγελματιών
- Όπως οδηγούς σε διάφορες κατηγορίες οχημάτων
 - Όπως ταξί, φορτηγά, λεωφορεία, ...
- Και παρεμφερή επαγγέλματα: σε αντιπροσωπείες, μηχανικούς, βενζινοπώλες, δασκάλους οδήγησης ...

Επιστήμονας της Πληροφορίας ↔ Οδηγός

- Επιστήμονας της Πληροφορίας: ο οδηγός των οχημάτων προς την πληροφορία
- Κάποτε λίγοι μπορούσαν να πληρώνουν για πληροφορία
- Σήμερα είναι εύκολο να ψάχνουμε μόνοι μας
- Και έτσι έχουμε καλύτερη εξυπηρέτηση – πληροφόρηση
- Και πολλούς επαγγελματίες της πληροφορίας
- Σε διάφορες κατηγορίες εργασιών – και άλλα παρεμφερή επαγγέλματα

Και ο Επιστήμονας της Πληροφορίας;

- Όπως οι υπολογιστές (που μπορούν να κάνουν πολλούς υπολογισμούς) δεν εκτοπίζουν τους λογιστές
 - Αλλά τους αλλάζουν τον τρόπο δουλειάς
 - Ωστε να ανταποκρίνονται σε πιο περίπλοκες καταστάσεις
 - Χρησιμοποιώντας πάντα τον υπολογιστή σαν εργαλείο
- Έτσι και οι διαδικτυακοί υπολογιστές (που μπορούν να ανταλλάσσουν – εξασφαλίζουν πρόσβαση στην πληροφορία) δεν εκτοπίζουν τους Επιστήμονες της Πληροφορίας, όπως τους βιβλιοθηκονόμους
- Σημασία έχει ποιος παίρνει τις αποφάσεις και αναθέτει εργασία στους υπολογιστές!

«Ψηφιακός Βιβλιοθηκονόμος»;

- **Οργανώνει** την ψηφιακή πληροφορία (αντικείμενα) και την ψηφιακή βιβλιοθήκη
- **Κατηγοριοποιεί** την ψηφιακή πληροφορία
- Καθορίζει τους τρόπους της **παρουσίασης** της ψηφιακής πληροφορίας
- Η ψηφιακή πληροφορία έχει μικρότερο κόστος
 - Και έτσι παράγεται περισσότερο υλικό
 - Που έχει μεγαλύτερη ανομοιογένεια ποιότητας