

## Περιεχόμενο Μεταδεδομένων

- 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

## Στόχοι Μεταδεδομένων

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

## Στόχοι Μεταδεδομένων (Τεκμηρίων)

- 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 όλες τους πόρους από όλους
    - (εν δυνάμει...)
  - Έμφαση στην χρήση – πελάτη
  - Υπερπληθώρα πληροφοριών
    - Ποσότητα αντί για ποιότητα
  - το φαινόμενο του Google (νέα προέλευση μεταδεδομένων)

## Παραγωγή 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

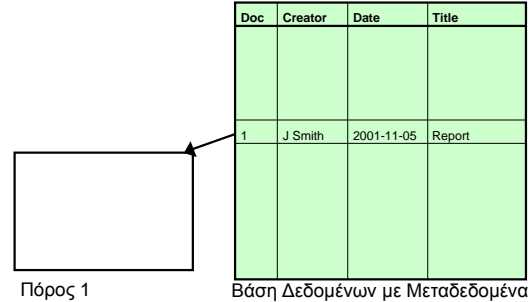
## Παραγωγή Μεταδεδομένων (2)

- 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

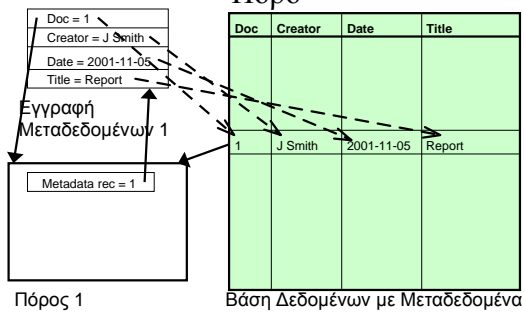
- Δημιουργούνται σαν εγγραφές «βάσης δεδομένων»
  - Πιθανώς απομακρυσμένη (remote) βάση δεδομένων
- Συνδεδεμένα με τους πόρους
- Εμφυτευμένα στους πόρους
  - Εξαρτώνται από την μορφή του πόρου
  - Μπορούν τα μεταδεδομένα να εξαχθούν από τον πόρο;
- Υιοθετείται η προσέγγιση που δίνει τη μεγαλύτερη ευελιξία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

## Μεταδεδομένα σαν Εγγραφές «Βάσης Δεδομένων»



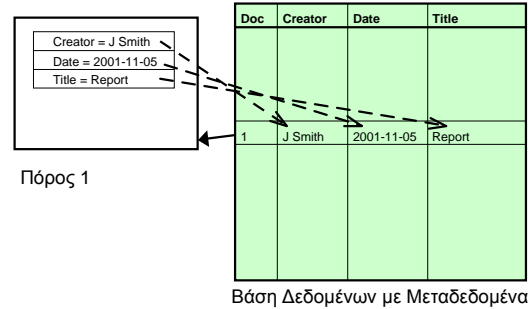
## Μεταδεδομένα Συνδεδεμένα με τον

### Πόρο

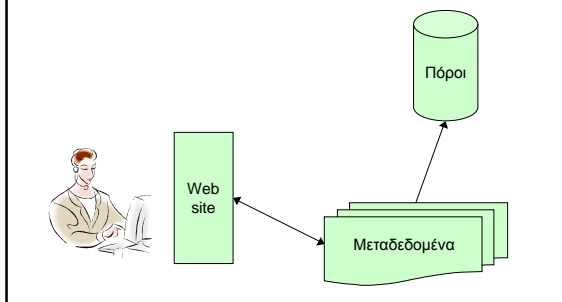


## Μεταδεδομένα Εμφυτευμένα στον

### Πόρο



## Ανακάλυψη Πόρων με Μεταδεδομένα



## Χρήση Μεταδεδομένων

- Λογισμικό: Αναζήτηση και παρουσίαση μεταδεδομένων + δεδομένων
- Λογισμικό OPAC
  - Προσδιορισμοί στα μεταδεδομένα(= δεδομένα);
- Λογισμικό Ψηφιακής Βιβλιοθήκης
  - Προσδιορισμοί στα μεταδεδομένα
  - Λειτουργίες στα δεδομένα
    - Π.χ. ταύτιση φράσεων / εικόνας / ήχου
    - Θεματική σύγκριση
      - Π.χ. διακοπές στη Χαβάη / ηλιοβασιλέμα / βάλς

## Ποιοι Χρησιμοποιούν Μεταδεδομένα

- Άνθρωποι
  - Ιδιοκτήτες που διαχειρίζονται τους πόρους τους
  - Ερευνητές που αναζητούν πόρους
  - Υπηρεσίες τρίτων (π.χ. βιβλιοθηκών, κέντρων τεκμηρίωσης)
- Λογισμικό
  - «Αναζήτησης – Συγκέντρωσης» (π.χ. μηχανές αναζήτησης)
  - «Πύλες» (portals) που παρουσιάζουν το «τοπίο» στο χρήστη
  - «Μεσάζοντες» (brokers) που εκτελούν εργασίες επερωτήσεων για λογαριασμό του χρήστη

## Πώς Μοιραζόμαστε Μεταδεδομένα

- Πώς ένας παροχέας δεδομένων φτιάχνει εγγραφές μεταδεδομένων διαθέσιμες σε μια κοινά κατανοητή μορφή;
- Πώς ένας παροχέας υπηρεσιών λαμβάνει αυτές τις εγγραφές μεταδεδομένων από τους παροχείς δεδομένων

## Κοινή Κατανόηση Μεταδεδομένων

- Τα μεταδεδομένα σαν γλώσσα
  - Οι εγγραφές μεταδεδομένων είναι σαν σύνολα δηλώσεων
- Η αποτελεσματική μετάδοση πληροφορίας προϋποθέτει συμφωνία σε:
  - Σημασιολογία (semantics): η σημασία των όρων
    - Π.χ. "cat", "to sit", "mat"
  - Δομή (structure): η βαρύτητα της διάταξης των όρων
    - Π.χ. πρόταση (στα Ελληνικά...): υποκείμενο -> ρήμα -> αντικείμενο
  - Συντακτικό (syntax): οι κανόνες της έκφρασης και οι χαρακτήρες
    - "The cat sat on the mat."

## Συναίνεση στα Μεταδεδομένα

- Μια κοινωνία περιγραφής πόρων ορίζεται από τη συναίνεση στις συμβάσεις
- Συναίνεση στο **συντακτικό**
  - Χρήση της XML
- Συναίνεση στη **σημασιολογία** των όρων
  - Η σημασία των στοιχείων / ιδιοτήτων
    - Με μοναδικά ονόματα μέσω των XML namespace
- Συναίνεση στην ερμηνεία της **δομής**
  - Χρήση ενός XML DTD/Schema πρότυπο της κοινότητας

## Οι Λειτουργίες (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 συμφοραζόμενα από τους πόρους

## Οι Λειτουργίες (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 υπομνηματίζουν
  - Να επανα-εξάγουν συμφοραζόμενα

## Ανακάλυψη Πόρων και Μεταδεδομένα

- Οι χρήστες των πόρων μπορεί να θέλουν να
  - Αναζητήσουν περιγραφές σε διαφορετικούς παροχείς
  - Συγκρίνουν/συνδυάσουν περιγραφές από διαφορετικούς παροχείς
- Οι παροχείς των πόρων μπορεί να θέλουν να
  - Διαδώσουν περιγραφές ευρέως
  - Μοιραστούν περιγραφές με άλλους προμηθευτές και υπηρεσίες
  - Περιγράψουν συσχετίσεις μεταξύ πόρων
- Οι υπηρεσίες τρίτων μπορεί να θέλουν να
  - Στηρίξουν «υπηρεσίες» πάνω σε περιγραφές προετοιμασμένες από άλλους
  - Σχολιάσουν περιγραφές προετοιμασμένες από άλλους

## Ανακάλυψη Πόρων και Μεταδεδομένα

- Τα μεταδεδομένα για ανακάλυψη πόρων πρέπει να
  - Χρησιμοποιούνται πέρα από την κοινωνία δημιουργίας τους
  - Συνδυάζονται με μεταδεδομένα από άλλες κοινωνίες
- Τα μεταδεδομένα συναθροίζονται ή αναζητούνται μεταξύ των διαφορετικών τοποθεσιών
  - Πρόκληση της «σημσιολογικής διαλειτουργικότητας»

## Μεταδεδομένα Ανακάλυψης Πόρων

- Τα μεταδεδομένα τυπικά καλύπτουν
  - Περιγραφή του **περιεχομένου** του πόρου
    - Τι είναι;
    - Μπορεί να περιλαμβάνει κάποια περιγραφή του περιβάλλοντος
  - Περιγραφή της **μορφής** του πόρου
    - Πώς είναι κατασκευασμένο;
  - Περιγραφή της **χρήσης** του πόρου
    - Τι εργαλεία χρειάζομαι για να το χρησιμοποιήσω;
    - Έχω τη δυνατότητα να το χρησιμοποιήσω;

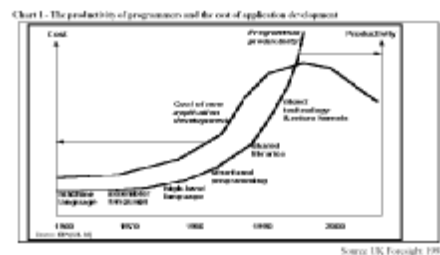
## Απλά ή Σύνθετα Μεταδεδομένα

- Αντίστοιχα ερωτήματα σε άλλες περιοχές
  - Π.χ. γλώσσες επεξεργαστών / υπολογιστών
- Τα **Απλούστερα** είναι
  - Ευκολότερα σε **δημιουργία** και **χρήση**
    - Με καλύτερη κατανόηση και λιγότερα λάθη
  - Με λιγότερη **λειτουργικότητα**
- Πρέπει να κατανοήσουμε **που** χρειάζονται
  - Και τους στόχους / μοντέλο χρήσης / κόστος

## Πηγές Πολυπλοκότητας

- Εσωτερική, οφείλεται στα δεδομένα
  - Π.χ. τύποι υλικού
- Διαδικαστική, οφείλεται στον τρόπο περιγραφής
  - Π.χ. ο συσχετισμός των εγγραφών
- Ιστορική, οφείλεται σε απαρχαιωμένες πληροφορίες, περιγραφές και εργαλεία
  - Π.χ. περιγραφή εξωφύλλου ή δεσμάτος

## Κόστος Τεχνολογίας



## Πρότυπα ... Μεταδεδομένων

- IAFA templates
- WWW semantic header
- URS (Uniform Resource Citation)
- OCLC InterCat project
- TEI (Text Encoding and Interchange)
- Search engine meta tags
- Resource Description Framework
- EAD (Encoding Archival Description)
- GILS (Govt Information Locator Service)
- Federal Geographic Data Committee
- Museum Educational Site Licensing Project
- Dublin Core

## Μεταδεδομένα: ποικιλία προσεγγίσεων και προτύπων

- USMARC / AACR<sub>x</sub>
- CIMI
- Visual Resources Association
- IAFA/ROADS
- Z39.50 Tag Set G
- SOIF
- MCORE
- DDI
- EPA

## Πρότυπα Μεταδεδομένων

- Βιβλιοθηκονομία
  - ISBD
    - CCF, USMARC, UNIMARC, UKMARC, ...
  - RFC 1807 – από επιστήμονες υπολογιστών
  - BIB1 (Z39.50)
  - Dublin Core
- Αρχαιονομία
  - EAD (περιγράφεται με XML)
- Μουσεία
  - CIDOC

## 1994+: ο Ιστός ως ένα Μεγάλο Πληροφοριακό Σύστημα

- Μηχανές αναζήτησης – αλλά με γνωστούς περιορισμούς:
  - Η κάλυψη ευρετηρίων είναι αποσπασματική, ή απλά άγνωστη
  - πάρα πολλές προσβάσεις (hits), και λίγη ακρίβεια
  - διαφημιστές πληρώνουν για τα κορυφαία σημεία
  - Spamming (ψευδείς λέξεις κλειδιά για να προσελκύσει τις προσβάσεις)
  - Στοιχεία αλλάζουν διεύθυνση ή εξαφανίζονται
  - ολόκληρες περιοχές Ιστού εξαφανίζονται

## Προς την «Αυτοματοποιημένη» Ψηφιακή Βιβλιοθήκη

- Οι «απλοί αλγόριθμοι» συν την απέραντη υπολογιστική ισχύ συχνά ξεπερνούν την ανθρώπινη νοημοσύνη" (William Arms)
  - Οι βιβλιοθηκονόμοι και τα μεταδεδομένα τους είναι ακριβά
  - Ιδιαίτερη πρόοδος σε αυτόματη ανάλυση αναφορών (σύνδεση στα αντικείμενα που αναφέρονται), ταίριασμα με πρότυπα λέξεων, εξαγωγή περιγραφικών όρων
  - Παράδειγμα: <http://google.com>
- Τα πλήρως αυτοματοποιημένα συστήματα παρέχουν το «Ford μοντέλο T» των πληροφοριών: *στηνές αλλά λειτουργικές λύσεις*

## Η Βιβλιοθήκη είναι τώρα Κατανεμημένη

- Ο Ιστός: χιλιάδες νέων παροχών πληροφοριών κάτω από μια παγκόσμια στέγη
  - χιλιάδες τοπικών συστημάτων «μεταδεδομένων»
  - βιβλιοθήκες, με τα μεταδεδομένα τους, σε μια μειονότητα
  - εξειδικευμένα σχήματα για την περιγραφή των δορυφορικών φωτογραφιών, τηλεοπτικοί συνδετήρες, κυβερνητικά έγγραφα...
- Πώς να ενσωματώσει την πρόσβαση σε αυτήν την ποικιλομορφία;

## Η Ιδέα του Πλαισίου Warwick (1996)

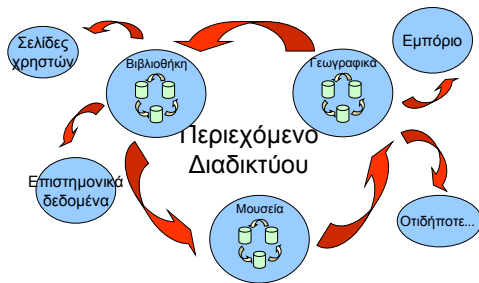


- Πακέτα μεταδεδομένων για τις διαφορετικές χρήσεις από τις διαφορετικές κοινότητες - modularity!

## Η Πρόκληση του WWW στην Παραδοσιακή Καταλογογράφηση

- Κλίμακα
- Μονιμότητα
- Αυθεντικότητα
- Οργανωτικό περιβάλλον
- Κηδεμονικός έλεγχος
- Πολυπλοκότητα περιεχομένου

## Το Περιεχόμενο του Διαδικτύου περιλαμβάνει πολλαπλές Κοινότητες



## Υποθέσεις Εξέλιξης Ιστού

- Υπόθεση «σημασιολογικού Ιστού»
  - Η τραχιά ολοκλήρωση της πρόσβασης μπορεί να επιτευχθεί μέσω των αρχιτεκτονικών προτύπων που επιτρέπουν την ημιαυτόματη, αν όχι τέλεια, συγχώνευση των μεταδεδομένων.
- Υπόθεση «πρωτόγονων μεταδεδομένων»
  - Τα μεταδεδομένα που βασίζονται σε ένα μικρό λεξιλόγιο γενικών όρων χρήσιμων διαθεματικά μπορούν να βελτιώσουν αποτελεσματικά την ακρίβεια της έρευνας.

## Υπόθεση «Συγκομιδής»

- Οι καθολικές επεκτάσιμες Ψηφιακές Βιβλιοθήκες μπορούν να χτιστούν από παροχές υπηρεσιών, που κάνουν συγκομιδή και ολοκλήρωση μεταδεδομένων από ποικιλία παροχών περιεχομένου
  - Προϋποθέτοντας πρότυπα στα σύνολα πεδίων
  - Προϋποθέτοντας μορφές εγγραφών
  - Προϋποθέτοντας συμβάσεις κατασκευής διαθέσιμων εγγραφών και πρωτοκόλλα συγκομιδής

## Υπόθεση «Υποδομής Μητρώου»

- Η συνεκτική ολοκλήρωση των μεταδεδομένων θα απαιτεί μια υποδομή για να γεφυρώσει σημασιολογικές διαφορές με μηχανική επεξεργασία
  - Συμβάσεις για αναπαράσταση στα σύνολα πεδίων
  - Περιγραφή σχέσεων μεταξύ διαφορετικών προτύπων μεταδεδομένων
  - Περιγραφή σχημάτων – βασισμένων σε πρότυπα – που εξειδικεύονται τοπικά για ειδικές χρήσεις

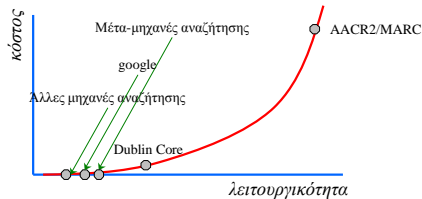
## Μεταδεδομένα: Μέρος μιας Λύσης

- **Δομημένα** δεδομένα που περιγράφουν δεδομένα
  - Βοηθάνε για να επιβάλουμε τάξη στο χάος
  - Κάνουν δυνατή την αυτόματη ανακάλυψη και χειρισμό
- Ποικιλομορφία μεταξύ διαφόρων διαστάσεων:
  - Εξειδίκευση
  - Αποκέντρωση
  - Εκδημοκράτιση

## Προκλήσεις και Τάσεις στα Μεταδεδομένα

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

## Μεταδεδομένα: Μέρος μιας Λύσης



## Προσαρμογή Μεταδεδομένων

- Να προσαρμόζονται σε πολλαπλές παραλλαγές / ποικιλίες μεταδεδομένων
  - Καλύπτοντας ευρεία ποικιλομορφία των δημιουργών και διατηρητών μεταδεδομένων
  - Προσφέροντας για κάθε κοινότητα (συγκεκριμένο χώρο) εξειδικευμένη λειτουργικότητα, δημιουργία, διαχείριση, πρόσβαση
- Τηρώντας Πρότυπα ...

## Προβληματισμοί Μεταδεδομένων

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

## Z39.50

- Δεν είναι πρότυπο περιγραφής και ανταλλαγής δεδομένων
- Είναι **πρωτόκολλο διάθεσης**, και περιλαμβάνει:
  - Διαπραγμάτευση δυνατοτήτων
    - Συμφωνία σε προφίλ δεδομένων (π.χ. **BIB-1**)
  - Διαδικασίες επικοινωνίας
  - Διατύπωση ερωτήσεων
  - Παρουσίαση αποτελεσμάτων

## Πρότυπα Μεταδεδομένων Z39.50

To ANSI/NISO Z39.50-1995 Appendix 3,  
ATR: Attribute Sets, pages 81-83, ορίζει τα:

- Bib-1 Z39.50-attributeSet 1
- Exp-1 Z39.50-attributeSet 2
- Ext-1 Z39.50-attributeSet 3
- CCL-1 Z39.50-attributeSet 4
- GILS Z39.50-attributeSet 5
- STAS Z39.50-attributeSet 6

## Bib-1: Z39.50-attributeSet 1

Έχει «γνωρίσματα» στις εξής κατηγορίες

- Use Attributes (π.χ. Personal name)
- Relation Attributes (π.χ. less than)
- Position Attributes (π.χ. first in field)
- Structure Attributes (π.χ. phrase)
- Truncation Attributes (π.χ. Right Truncation)
- Completeness Attributes (π.χ. incomplete subfield)

## BIB-1 Use Attributes

|                 |   |                      |      |
|-----------------|---|----------------------|------|
| Personal name   | 1 | Thematic-number      | 1030 |
| Corporate name  | 2 | Material-type        | 1031 |
| Conference name | 3 | Doc-id               | 1032 |
| Title           | 4 | Host-item            | 1033 |
| Title series    | 5 | Content-type         | 1034 |
| Title uniform   | 6 | Anywhere             | 1035 |
| ISBN            | 7 | Author-Title-Subject |      |
| ISSN            | 8 |                      | 1036 |

## Πρότυπο Μεταδεδομένων RFC 1357 και RFC 1807

Χρησιμοποιήθηκε για το DIENST

- Απλά
- Ορισμένα από Επιστήμονες Πληροφορικής
- Με συντακτικό που επίσης ορίζει το πρότυπο
- TITLE (ΤΙΤΛΟΣ)
- AUTHOR (ΣΥΓΓΡΑΦΕΑΣ)
- ABSTRACT (ΠΕΡΙΛΗΨΗ)
- SUBJECT (ΘΕΜΑ)
- DATE (ΗΜΕΡΟΜΗΝΙΑ)