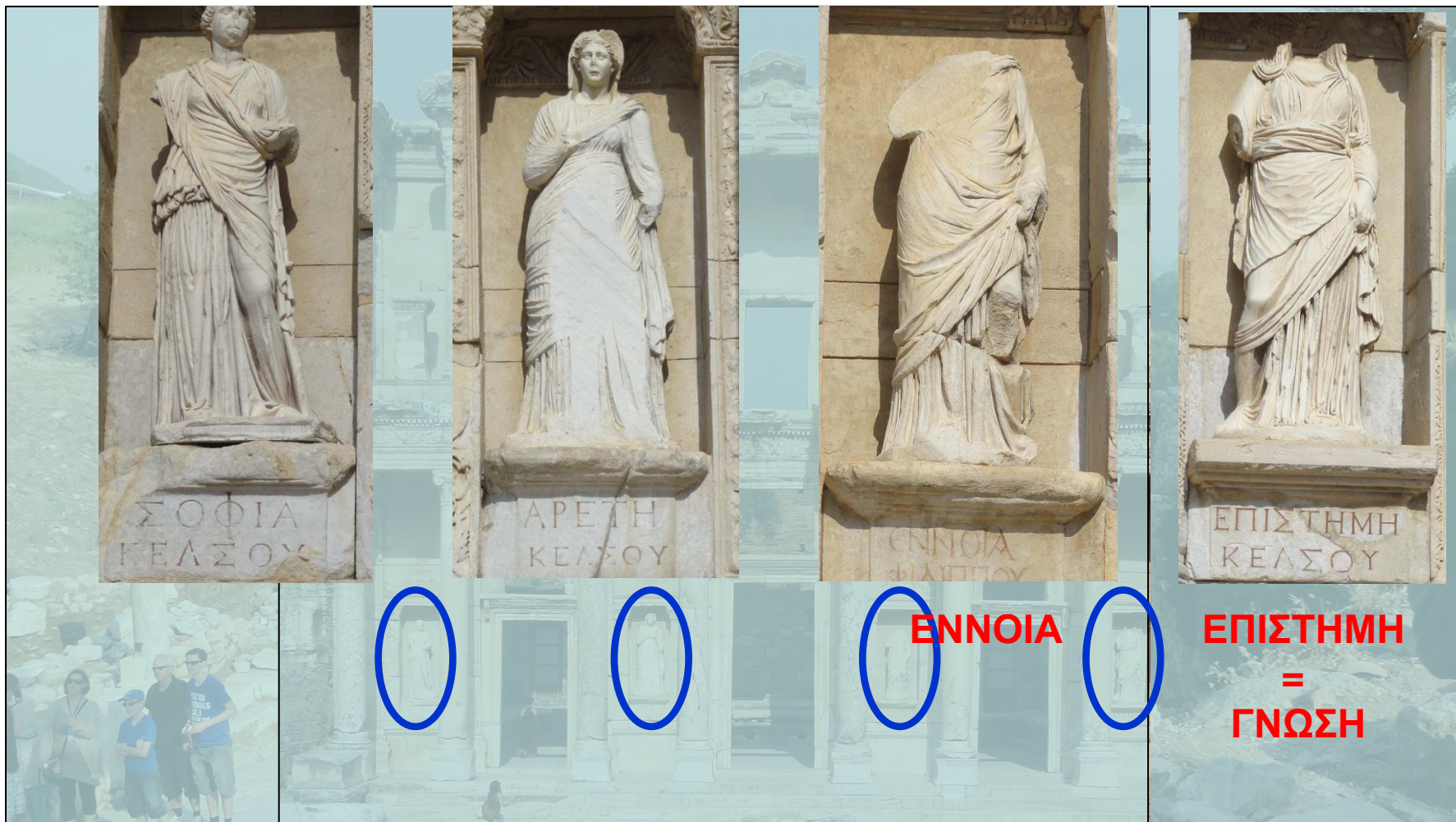

Ο ρόλος των ελεγχόμενων λεξιλογίων και θησαυρών στην οργάνωση της γνώσης

Νικόλαος Μήτρου
Καθηγητής ΕΜΠ

Ιούνιος 2013

Εισαγωγή

Αρχαία Έφεσος



Ορισμοί

Μάθηση και Γνώση

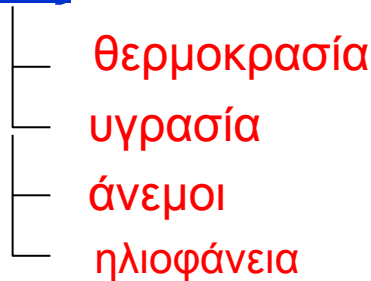
- Ηλεκτρονική μάθηση (e-learning): η μάθηση με ηλεκτρονικά μέσα
- Μάθηση: Η διαδικασία απόκτησης γνώσης
- Γνώση:
 - Στην καθομιλούμενη
 - Φιλοσοφική άποψη
 - Ψυχολογική άποψη
 - Θρησκευτική άποψη
 - Επιστημολογική άποψη

Ορισμοί (2/2)

- **Έννοια (concept)**: στοιχειώδης ή σύνθετη μονάδα σκέψης (unit of thought), αναφερόμενη σε μια συγκεκριμένη κατηγορία (class) οντοτήτων, υπαρκτών ή αφηρημένων (αντικειμένων, ιδεών, ...) ή γεγονότων του υπαρκτού κόσμου. Κάθε επιμέρους οντότητα ή γεγονός αποτελεί στιγμιότυπο ή επεισόδιο (instance) της έννοιας
- **Γνώση (knowledge)**: Το σύνολο των εννοιών που έχουμε σχηματίσει, οργανωμένο σε ένα δομημένο σύστημα με ιεραρχικές ή άλλες σχέσεις μεταξύ τους, των λογικών προτάσεων που μπορούν σχετικά να διατυπωθούν, καθώς και των γνωστών στιγμιοτύπων αυτών

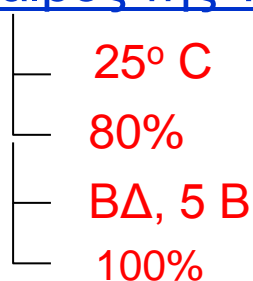
Έννοια

καιρός



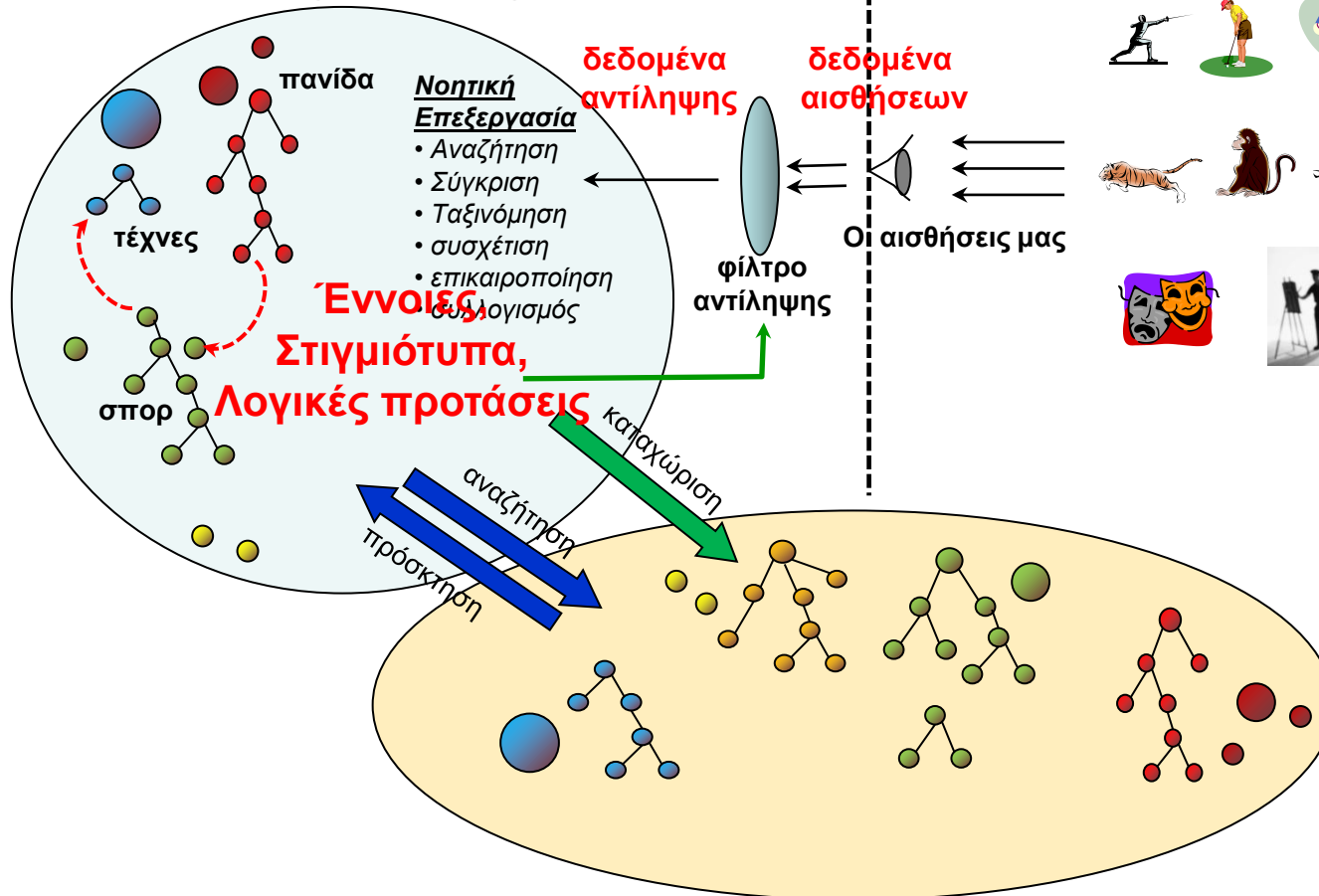
Στιγμιότυπο (instance)

Ο καιρός της Τετάρτης



Η διαδικασία της μάθησης

Ο προσωπικός «συσσωρευτής»
της γνώσης

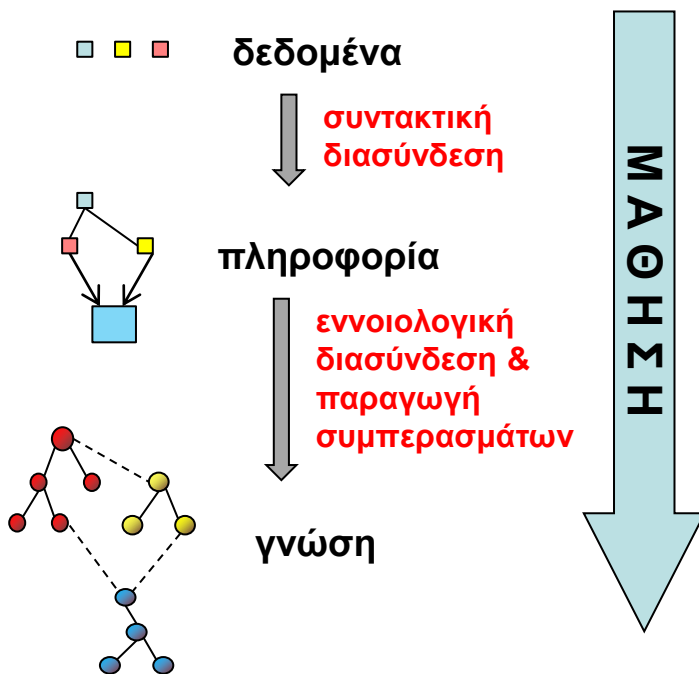


Ο παγκόσμιος «συσσωρευτής» της γνώσης

Η διαδικασία της μάθησης (συνεχ.)

- Μια διαρκής, αμφίδρομη επικοινωνία με τον περιβάλλοντα κόσμο και τον παγκόσμιο συσσωρευτή γνώσης
- Τα δεδομένα των αισθήσεων (sensed data) φιλτράρονται από τα «φίλτρα» πρόσληψης (perceived data) και, μετά από επεξεργασία, ενσωματώνονται στο ήδη υπάρχον γνωστικό οικοδόμημα ως
 - νέες έννοιες και οντότητες/στιγμιότυπα αυτών (concepts & instances)
 - επικαιροποίηση υφιστάμενων εννοιών
 - επικαιροποίηση των ταξινομικών-εννοιολογικών δένδρων (της δομής, εν γένει)
 - συσχετίσεις εννοιών και οντοτήτων
 - ...
- Αντίστροφη της παραπάνω συνθετικής διαδικασίας (επαγωγής) είναι η αναλυτική διαδικασία της παραγωγής συμπερασμάτων και νέας γνώσης από την ήδη υπάρχουσα, μέσω συλλογισμού
- Η νέα γνώση καταχωρίζεται στον παγκόσμιο συσσωρευτή για κοινή χρήση

Δεδομένα – πληροφορία – γνώση



- Κρίσιμη στη μαθησιακή διαδικασία είναι η διάκριση των όρων (και ο σταδιακός μετασχηματισμός από τον πρώτο στον τελευταίο):
δεδομένα → πληροφορία → γνώση
(data → information → knowledge)

Παράδειγμα:

Δεδομένα

Θερμοκρασίες

00:00	22°C
06:00	25°C
12:00	28°C

Υγρασίες

00:00	56 %
06:00	65 %
12:00	77 %

Άνεμοι

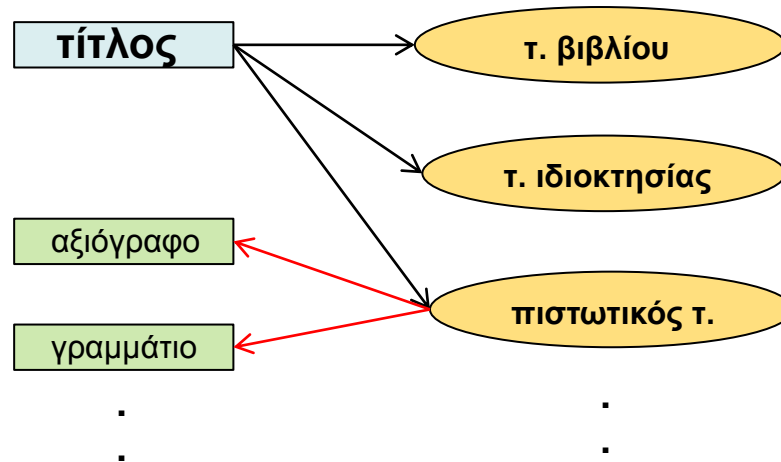
00:00	ΒΔ 4B
06:00	ΒΔ 5B
12:00	Δ 6.5B

↓
Πληροφορία “Ο καιρός της Τετάρτης: ...”

↓
Γνώση “Ο συνδυασμός υψηλών πιέσεων στην περιοχή των Βαλκανίων με τις χαμηλές πιέσεις στη Μεσόγειο ...”

Λέξεις και Έννοιες

- Για την επικοινωνία οι έννοιες διατυπώνονται (κωδικοποιούνται) με λέξεις
- Η αντιστοιχία δεν είναι αμφιμονοσήμαντη
 - Ομόγραφα (ομώνυμα): μία λέξη για πολλές έννοιες
 - Συνώνυμα : πολλές λέξεις για την ίδια έννοια



Μηχανική οργάνωση & αναζήτηση πληροφορίας

Πώς λειτουργεί:

- Τα έγγραφα (ή **κελύφη** πληροφορίας) επισημειώνονται ή αλλιώς συνοδεύονται από μετα-δεδομένα (metadata)
 - Από τον ίδιο το δημιουργό
 - ✓ τίτλος - συγγραφέας
 - ✓ πίνακας περιεχομένων
 - ✓ περίληψη
 - ✓ λέξεις –κλειδιά για το θεματικό περιεχόμενο
 - ✓ ...
 - Από ειδικούς (π.χ. Βιβλιοθηκονόμους)
 - ✓ βιβλιογραφικές εγγραφές
 - ✓ Βιβλιογραφικοί κατάλογοι
 - Από μηχανές (με επεξεργασία του σώματος της πληροφορίας)
 - ...
- Δημιουργούνται δεικτοδοτικοί κατάλογοι (indexes), με αφετηρίες τα μεταδεδομένα, οι οποίοι και χρησιμοποιούνται στην αναζήτηση

Μηχανική οργάνωση & αναζήτηση πληροφορίας (συνεχ.)

Indexing-searching με τη λέξη-κλειδί «τίτλος»

Κατάλογος αναζήτησης

keyword	URL
τίτλος	www.legal.gr/doc1.pdf
	www.sports.gr/doc2.html
	www.library.gr/catalog/book3

ΔΙΑΤΑΓΕΣ ΠΛΗΡΩΜΗΣ - ΠΙΣΤΩΤΙΚΟΙ ΤΙΤΛΟΙ

A. ΕΙΡΗΝΟΔΙΚΕΙΟ ΑΘΗΝΩΝ

1. ΥΛΙΚΗ ΑΡΜΟΔΙΟΤΗΤΑ: Μέχρι και δώδεκα χιλιάδες (12.000) ΕΥΡΩ.
2. ΤΟΠΙΚΗ ΑΡΜΟΔΙΟΤΗΤΑ: Ισχύουν, σύμφωνα με το άρθρο 591 παρ. 1 ΚΠολΔ, οι γενικές διατάξεις (άρθρα 22-44 ΚΠολΔ). Η κατοικία του οφειλέτη καθού ή έστω η κατοικία ενός από τους οφειλέτες - καθών, δεν αποκλείεται, όμως, η θεμελίωση της κατά τόπον αρμοδιότητας σε κάποια ειδική δικαιοσύνη. Έτσι, ο τίτλος είναι ειδική πληρωτέα η συναλλαγματική θεμελιώνει την κατά άρθρο 33 ΚΠολΔ εκπληρώσεως της παροχής, ο τόπος εκδόσεως της συναλλαγής κατά το ίδιο άρθρο δικαιοσύνη του τόπου καταστάσεως της δικαιοσύνης.

Ψηφιακή Βιβλιοθήκη



INTERNET



indexer

Αναζήτηση: τίτλος



“Ο τίτλος στην τακτική Χρησιμότητα”, Αλεξανδροπούλου



Ιστότοπος εφημερίδας

Ο τίτλος θα κριθεί στη Μαδρίτη!

Η Μπαρτσελόνα δευτερεύει στη Ρεάλ να σηκώνει κούπα στο «Palau Blaugrana» (73-62) κι έτσι ο τίτλος θα κριθεί στη Μαδρίτη (19/06).



Βιβλιογραφικός κατάλογος



Μηχανική αναζήτηση πληροφορίας - τα προβλήματα

Google

τίτλος

12.500.000 results !
(όχι συνώνυμα)

Web Images More Search tools

About **12,500,000 results** (0.25 seconds)

Θα θέλαμε:

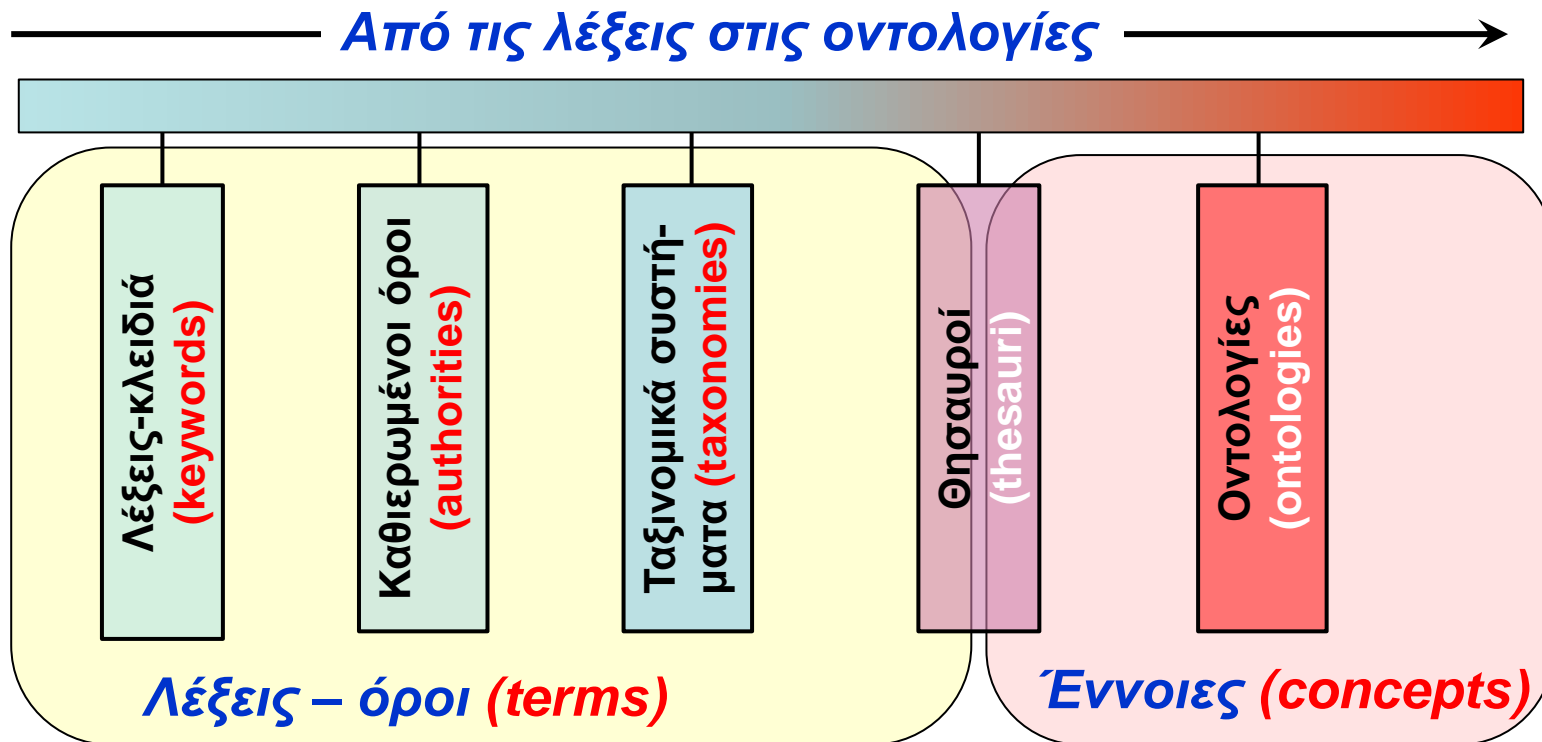
- Όλα όσα ζητούμε (πληρότητα – **recall**)
- Μόνον όσα ζητούμε (ακρίβεια – **precision**)
- ... άρση της πολυσημίας λόγω των ομογράφων
- ... εύρεση των συνωνύμων

8 hours ago – Η Μπαρτσελόνα δεν επέτρεψε στη Ρεάλ να σηκώσει κούπα στο «Palau Blaugrana» (73-62) κι έτσι ο τίτλος του Πρωταθλητή Ισπανίας θα κριθεί ...

Λύσεις στο πρόβλημα της αναζήτησης

- **Περιορισμός στο θεματικό χώρο αναζήτησης**
 - Στο παράδειγμα “τίτλος”
 - Βιβλιογραφικά: τίτλος (βιβλίου, περιοδικού, ...)
 - Οικονομικά: τίτλος (ιδιοκτησίας, πιστωτικός, ...)
 - Αθλητικά: τίτλος (πρωταθλήματος, ...)
- **Επισημείωση – αρχειοθέτηση - αναζήτηση** των εγγράφων με λέξεις-κλειδιά από **ελεγχόμενα λεξιλόγια** (καθιερωμένους όρους και ταξινομικά σχήματα της βιβλιοθηκονομίας)
- **Έννοιολογική επισημείωση και αναζήτηση** των εγγράφων (semantics) – **αντί για «λέξεις-κλειδιά», «έννοιες-κλειδιά»** – με χρήση **θησαυρών** και **οντολογιών**
web of documents → semantic web

Ελεγχόμενα λεξιλόγια - θησαυροί



Λέξεις-κλειδιά (keywords)

- Λεξιλόγιο ελεύθερο ή ελεγχόμενο, αλλά αδόμητο (flat list)
 - Ελεγχόμενο: περιορισμένος αριθμός καθιερωμένων όρων
- Αυτοδηλούμενη η σημασία των όρων
- Η συνήθης πρακτική στην αυταρχειοθέτηση και αναζήτηση στο σημερινό web of documents

Καθιερωμένοι όροι στη βιβλιοθηκονομία

- Χρήση στη βιβλιοθηκονομία (authority files)
 - Καθιερωμένοι όροι ονομάτων (name authority files)
 - Θεματικοί καθιερωμένοι όροι (subject headings)
- Κωδικοποιημένοι για μηχανική επεξεργασία (MARC), μπορούν να δοθούν και σε μορφή MARCXML ή RDF για χρήση στο web ή ως θησαυροί (π.χ. τα Library of Congress Subject Headings)

The screenshot shows the Library of Congress website interface. At the top, there is a navigation bar with the Library of Congress logo and buttons for 'ASK A LIBRARIAN', 'DIGITAL COLLECTIONS', and 'LIBRARY'. Below this is a breadcrumb trail: 'The Library of Congress > Linked Data Service > LC Subject Headings'. The main heading is 'Title insurance--Law and legislation--United States'. Below the heading, it says 'From Library of Congress Subject Headings'. There are three tabs: 'Details' (selected), 'Visualization', and 'Suggest Terminology'. The 'Details' tab shows the following information:

- Title insurance--Law and legislation--United States**
- URI(s)**
 - > <http://id.loc.gov/authorities/subjects/sh2010116568>
 - > [info:loc.gov/authorities/sh2010116568](http://info.loc.gov/authorities/sh2010116568)
 - > <http://id.loc.gov/authorities/sh2010116568#concept>
- Instance Of**
 - > [MADS/RDF ComplexSubject](#)
 - > [MADS/RDF Authority](#)
 - > [SKOS Concept](#)
- Components**
 - > [Title insurance](#)
 - > [Law and legislation](#)
 - > [United States](#)
- Scheme Membership(s)**
 - > [Library of Congress Subject Headings](#)
- Collection Membership(s)**
 - > [LCSH Collection - Authorized Headings](#)

Ταξινομίες (taxonomies)

- Ελεγχόμενα λεξιλόγια με ιεραρχική δομή
- Συνήθως χωρίς συνώνυμα
- Συχνά με σχόλια
- Γενικές κατηγορίες μικρής ακρίβειας
- Κατάλληλες για αναζήτηση με πλοήγηση
- Παραδείγματα:
 - **Library of Congress Classification (LCC)**
 - **Dewey Decimal Classification (DDC)**

Class 600 – Technology [\[edit\]](#)

• 600 **Technology (Applied sciences)**

- 600 Technology
- 601 Philosophy & theory
- 602 Miscellany
- 603 Dictionaries & encyclopedias
- 604 Special topics
- 605 Serial publications
- 606 Organizations
- 607 Education, research, related topics
- 608 Invention & patents
- 609 Historical, geographic, persons treatment

• 610 **Medical sciences; Medicine**

- 611 Human anatomy, cytology, histology
- 612 Human physiology
- 613 Personal health & safety
- 614 Incidence & prevention of disease
- 615 Pharmacology and therapeutics
- 616 Diseases
- 617 Surgery & related medical specialties
- 618 Gynecology & other medical specialties

Θησαυροί (thesauri)

Εργαλεία, τα οποία επιτρέπουν θεματική (εννοιολογική) πρόσβαση στην πληροφορία

- Ελεγχόμενα λεξιλόγια
- Ιεραρχική δομή
- Χρήση συνωνύμων
- Συσχετίσεις εννοιών
 - Ευρύτερη
 - Στενότερη
 - Συσχετισμένες
- Σχετικά απλό εννοιολογικό μοντέλο
- Σχετικά υψηλή ακρίβεια
- Υποστηρίζουν εκτεταμένη αναζήτηση και μέσω των συνωνύμων ή των σχετιζόμενων όρων

EuroVoc Multilingual Thesaurus of the European Union

oc homepage > Domains and MT > πιστωτικός τίτλος

Usage:

ch

arch

subject-
sion

alphabetical
st
index

πιστωτικός τίτλος

UF αξιόγραφο
 γραμμάτιο σε διαταγή
 εμπορικό γραμμάτιο
 συναλλαγματική

24 ΔΗΜΟΣΙΟΝΟΜΙΚΑ

MT 2411 νομισματική οικονομία

BT1 λογιστικό χρήμα
 BT2 νόμισμα
 BT3 χρηματαγορά

RT πίστη [2416]
 συναλλαγματικό δίκαιο [2006]

LANGUAGE EQUIVALENTS

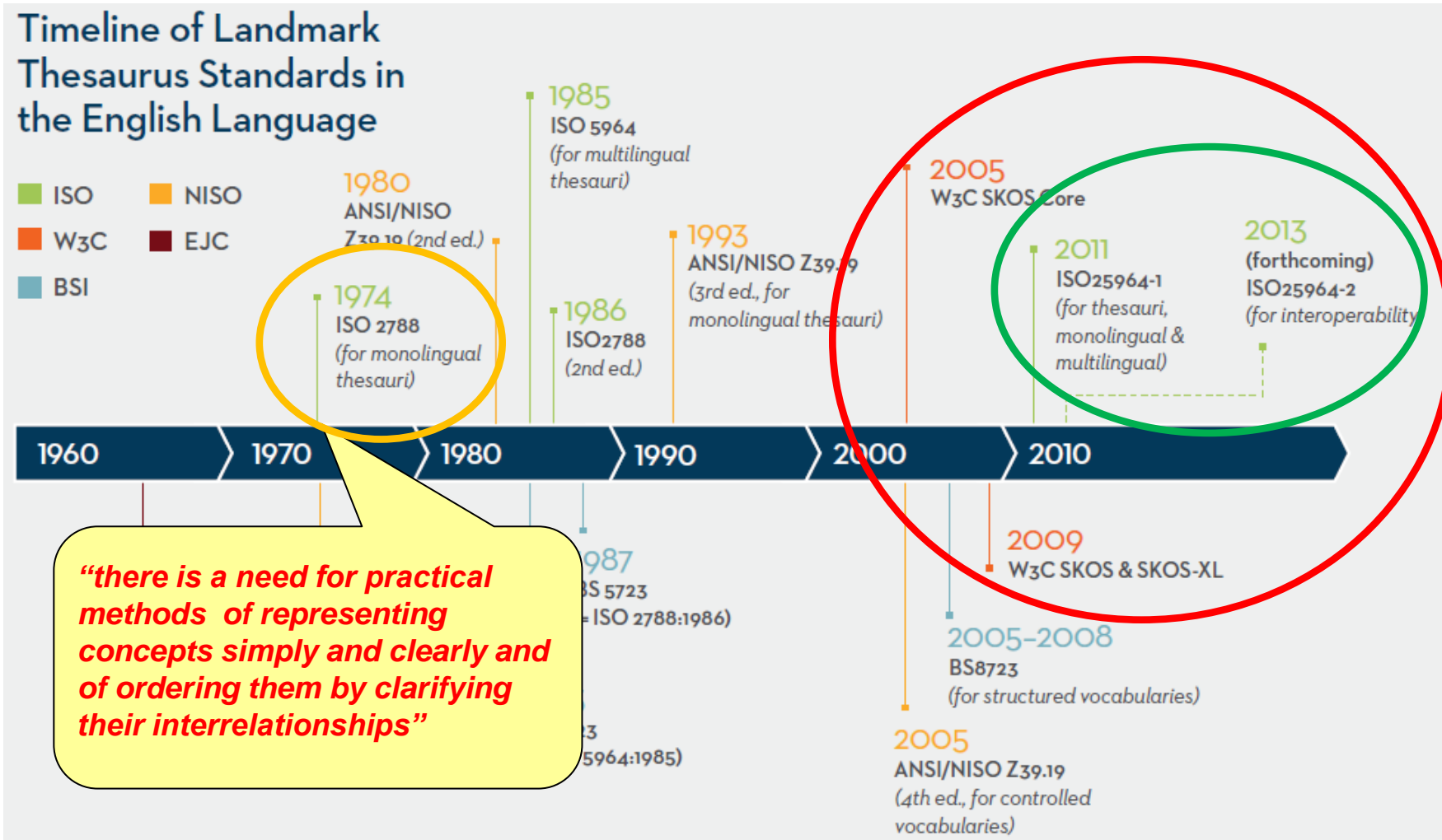
BG	неготируем инструмент
ES	título de crédito
CS	převoditelný cenný papír
DA	gældsbreve
DE	Wechsel
ET	vabalt kaubeldav instrument
EL	πιστωτικός τίτλος
EN	negotiable instrument
FR	titre de crédit
IT	titolo di credito
LV	tirgojams vērtspapīrs
LT	apyvarčioji priemonė
HU	forgatható értékpapír
MT	negotiable instrument (under ti
NL	kredietbrief
PL	weksel
PT	título de crédito
RO	titlu de credit
SK	prevoditeľný cenný papier
SL	prenosni instrument
FI	velkakirja
SV	växel

Οντολογίες (ontologies)

- Πλήρη εννοιολογικά μοντέλα σε θεματικές περιοχές ενδιαφέροντος
- Έννοιες, ιδιότητες και στιγμιότυπα (concepts, properties & instances)
- Ιεραρχική δομή εννοιών και ιδιοτήτων
- Επιτρέπουν τους συλλογισμούς
- Επιτρέπουν τη εκτεταμένη σημασιολογική αναζήτηση

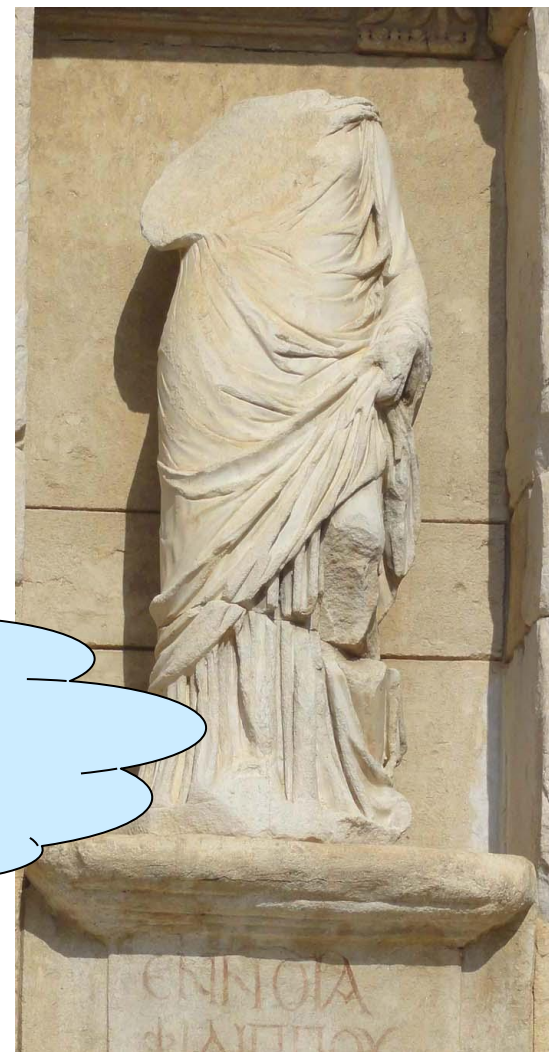
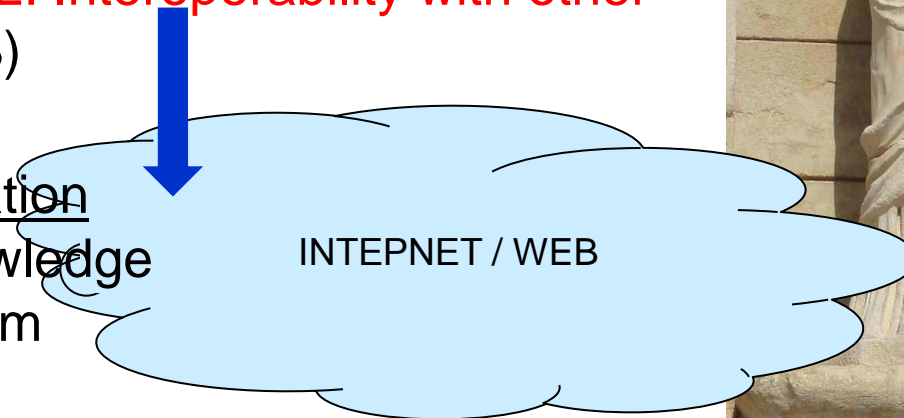
Θησαυροί: εξέλιξη και πρότυπα

S. G. Dextre Clarke, M. Lei Zeng, "From ISO 2788 to ISO25964: The Evolution of Thesaurus Standards towards Interoperability and Data modeling", *SQ Information standards Quarterly*, 2012, vol. 24, issue 1



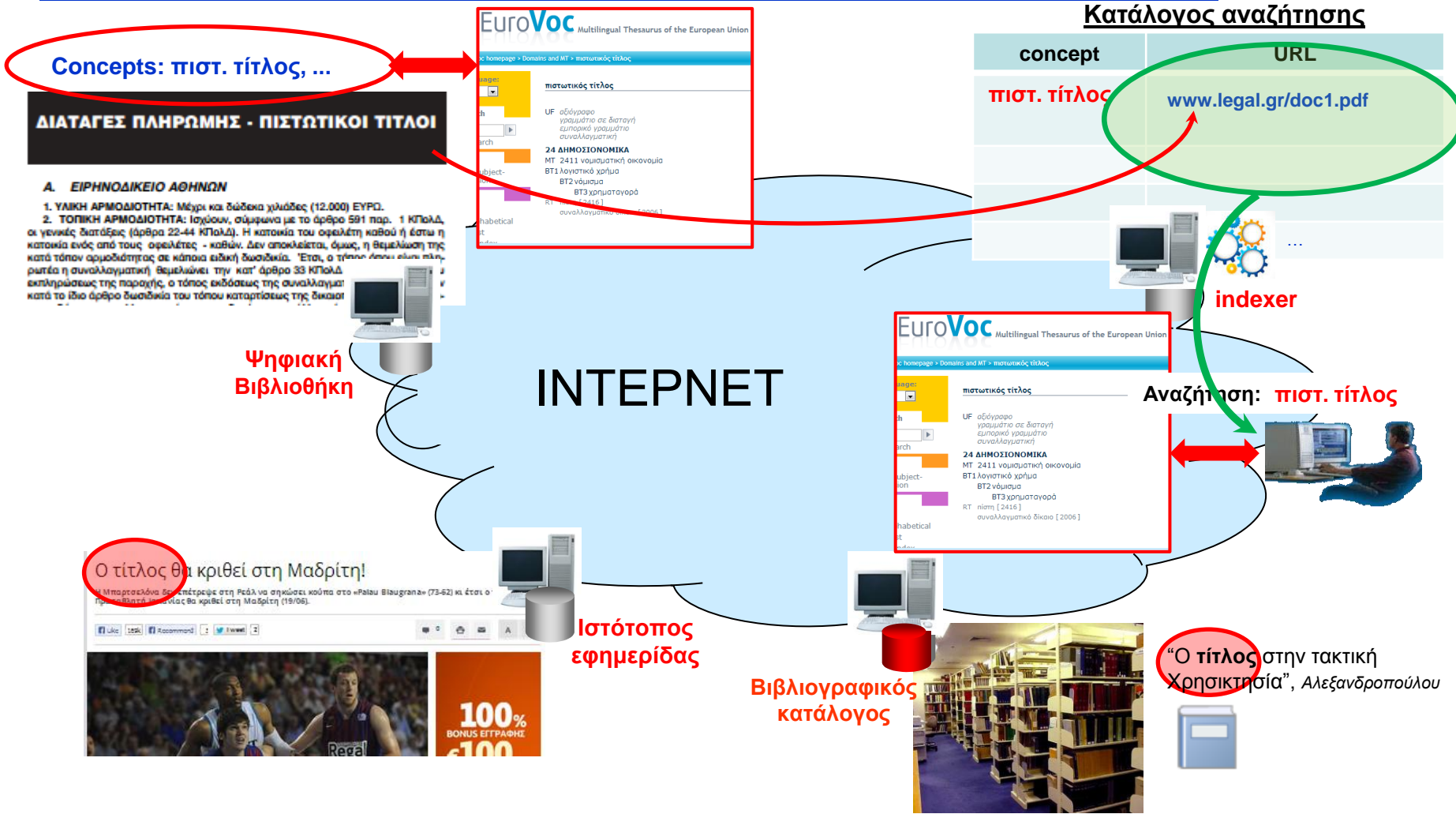
Τα πρότυπα ISO και W3C

- ISO 25964-1 Information and Documentation – Thesauri and interoperability with other vocabularies – **Part 1: Thesauri for Information retrieval** (2011)
- ISO 25964-2 Information and Documentation – Thesauri and interoperability with other vocabularies – **Part 2: Interoperability with other vocabularies** (2013)
- W3C Recommendation SKOS Simple Knowledge Organization System Reference (2009)
(για τη δημοσίευση εννοιο-κεντρικών λεξιλογίων – θησαυρών στο Web)



Χρήση θησαυρών

Indexing-searching με τη έννοια-κλειδί «πιστωτικός τίτλος»



Συμπεράσματα

- Keyword lists, authority files, taxonomies, thesauri, ... είναι ελεγχόμενα λεξιλόγια, τα οποία περιορίζουν την ασάφεια και αυξάνουν την ακρίβεια στην επισημείωση και την αναζήτηση της πληροφορίας
- Όταν αυτά είναι και «εννοιοκεντρικά», όπως οι σύγχρονοι θησαυροί των προτύπων ISO 25964, δίνουν τη δυνατότητα σημασιολογικής οργάνωσης και αναζήτησης της πληροφορίας και αυξάνουν τόσο την ακρίβεια, όσο και την πληρότητα της αναζήτησης
- Διεθνώς, παρατηρείται εργώδης προσπάθεια για τη σημασιολογική διασύνδεση δεδομένων και πληροφορίας (Linked Open Data initiatives) με τη χρήση πολύγλωσσων θησαυρών και οντολογιών
- Στην προσπάθεια συμμετέχουν εθνικές βιβλιοθήκες και άλλα μεγάλα ανοικτά «αποθετήρια γνώσης» (wikipedia, ...)
- Πρέπει να ενεργοποιηθεί στην κατεύθυνση αυτή και η Ελληνική Γραμματεία