



# Από τον Παγκόσμιο Ιστό στον Σημασιολογικό Ιστό

Αλέξανδρος Βαλαράκος †‡  
(alexv@iit.demokritos.gr)

† Υποψήφιος Διδάκτορας, Τμήματος Μηχανικών Υπολογιστικών και Πληροφοριακών Συστημάτων,  
Παν/μίου Αιγαίου.

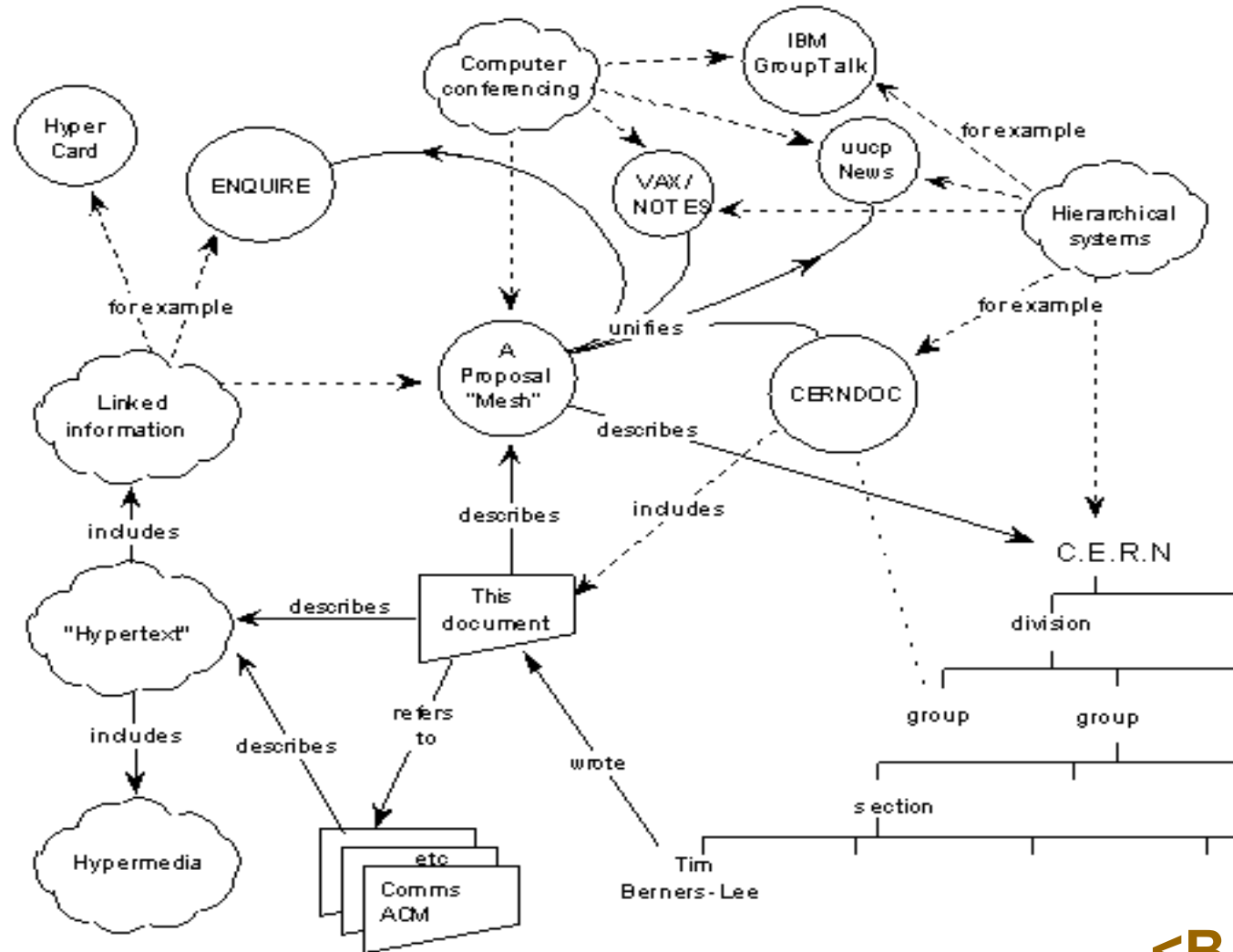
‡ Συνεργαζόμενος Ερευνητής (Υπότροφος) Ινστιτούτου Πληροφορικής και Τηλεπικοινωνιών, Εργαστήριο  
Λογισμικού και Γνώσης, Ε.Κ.Ε.Φ.Ε Δημόκριτος.

- Παγκόσμιος Ιστός (ΠΙ)
  - Από τον ΠΙ στο Σημασιολογικό Ιστό (ΣΙ)
  - Σημασιολογικός Ιστός
  - Αρχιτεκτονική Σημασιολογικού Ιστού
  - Τεχνολογίες Σημασιολογικού Ιστού
  - Εφαρμογές
  - Προκλήσεις
-

- Ο Tim Berners-Lee εφευρίσκει τον Παγκόσμιο Ιστό (1989)
- M. Andreessen εφευρίσκει το MOSAIC (1993)
  - γραφικός φυλλομετρητής (browser)
  - η πρώτη ελεύθερη “killer App”
  - Netscape browser

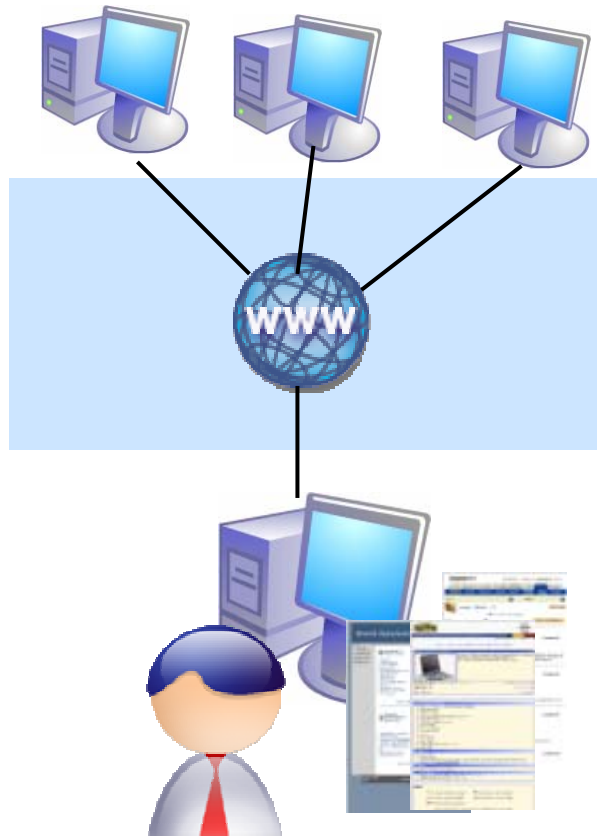


# Ο Παγκόσμιος Σημασιολογικός Ιστός του Berners Lee



[\*\*<BACK>\*\*](#)

*Tim Berners-Lee, Information Management: A Proposal, CERN March 1989*

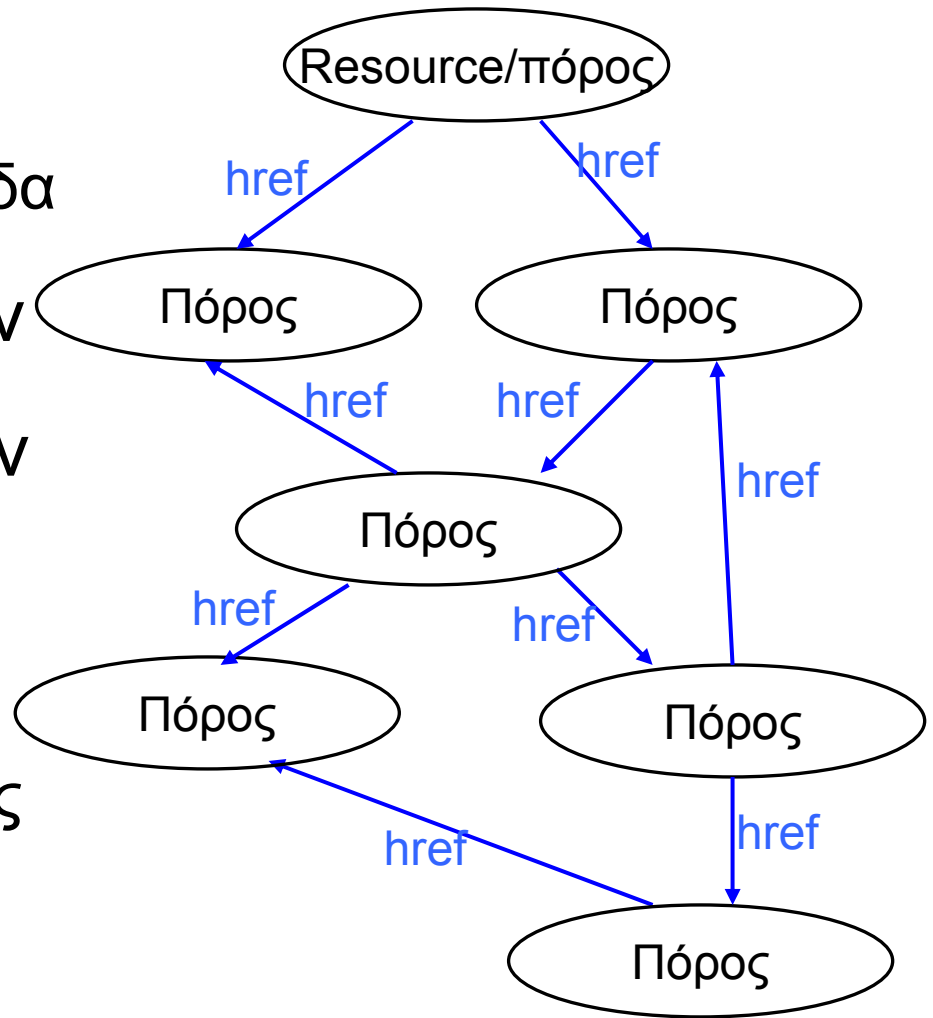


- Μετάδοση πληροφορίας από:
  - διαφορετικά λειτουργικά συστήματα
  - διαφορετικές μορφές κειμένων
- Αποθήκευση πληροφορίας
- Συσχέτιση πληροφορίας
- Εμφάνιση πληροφορίας
  - Αναγνώσιμη από τον άνθρωπο
- Ευκολότερη πλοήγηση και αναζήτηση πληροφορίας
- Χρήση υπερκειμένων (hypertext)
- Κατανεμημένο

- Καθολικό & αποκεντρωμένο πληροφοριακό σύστημα
- Πλουραλιστικό
  - [πηγές/απόψεις  $> 1$ ] για μία πληροφορία
- Άναρχο
  - όχι συγκεκριμένος τρόπος παρουσίασης πληροφορίας
  - αριθμός θεμάτων  $\rightarrow \infty$
- Ταχέως αναπτυσσόμενο
- Τεράστιο όγκο πληροφοριών (*δισ. ιστοσελίδες*)

# Ο Παγκόσμιος Ιστός τώρα: χαρακτηριστικά

- Ψηφιακή βιβλιοθήκη
  - υπερκείμενα = ιστοσελίδα
- Πλατφόρμα εφαρμογών
- Πλατφόρμα πολυμέσων
- Ένα ονοματισμένο σχήμα
  - κατευθυνόμενος γράφος
  - μοναδική ταυτότητα υπερκειμένων



# Τι βλέπει η μηχανή και τι ο άνθρωπος

## Ο τελικός χρήστης: πάντα ο άνθρωπος!

### Άνθρωπος



### Books

34% Off Books over \$25

Hello, Eftimios Mavrikas. We have [Book Recommendations](#) for you. (If you're not Eftimios Mavrikas, [click here](#).)

### Great News from Hogwarts



Harry Potter fans, rejoice! J.K. Rowling has announced that [Harry Potter and the Half-Blood Prince](#), Book 6 in her wildly popular [Harry Potter series](#), will be released on July 16, 2005.

- [Pre-order today and save 40%](#)
- [Catch up on the entire series with the hardcover boxed set](#)
- [Enjoy these great reads while you wait](#)

New in Books  
Blink

### Μηχανή (H/Y)

```
...<TR vAlign=top>
  <TD rowSpan=2 vAlign=top><FONT
    face=verdana,arial,Helvetica><B>2.</B><
    /FONT></TD>
  <TD rowSpan=2 vAlign=top><FONT
    size=-1><A
      href="http://www.amazon.com/exec/obidos
        /ASIN/007018724X/103-1468608-
        7602226"><IMG
          border=0 height=90
          src="web_marketing_files/007018724X.01.
            TZZZZZZZ.gif"
          width=60></A></FONT> <BR
        clear=all></TD> ...
```

- Πόσο μακριά είμαι από αυτό που ψάχνω;
- Το μέλλον είναι δύσκολο!:
  - βιβλίο (κείμενο) ή φωτογραφία:  $\sim 1$  Megabyte
  - βιβλιοθήκη του Κογκρέσου (κείμενα):  $\sim 25$  Terabytes
- Εκτιμώμενη ετήσια παγκόσμια παραγωγή δεδομένων:
  - μερικά Exabytes ( $10^{18}$ ), εκ των οποίων:
    - κείμενο: μερικά Terabytes ( $10^{12}$ )
    - εικόνες, ήχος, αριθμητικά δεδομένα: τα υπόλοιπα
- Αναζήτηση;
  - Με λέξεις κλειδιά????

- Ομοιόμορφη αναζήτηση πληροφορίας σε
  - Διαφορετικές (ανόμοιες) βάσεις δεδομένων (Deep Web);
  - Ιστοσελίδες;
- Πρόβλημα: *Σημασιολογική ανομοιογένεια*
- Είναι δυνατή η αυτοματοποίηση διαδικασιών στο «άναρχο» ΠΙ;

- Αδυναμία κατανόησης του περιεχομένου μιας ιστοσελίδας από τον Η/Υ
- Πληροφορία δύσκολα επεξεργάσιμη από τον Η/Υ
- Αναζήτηση στηριζόμενη σε λέξεις κλειδιά
- Αδυναμία αναζήτησης σε ανομοιογενείς και κατανεμημένες πηγές
- Αδυναμία αυτοματοποίησης διαδικασιών

# Ερώτηση στον Παγκόσμιο Ιστό



buy da vinci code in greek bookstore

Αναζήτηση

Σύνθετη Αναζήτηση  
Ρυθμίσεις

Αναζήτηση:  στον ιστό  σελίδες γραμμένες στα Ελληνικά  σελίδες από Ελλάδα

Η λέξη "in" είναι πολύ συνηθισμένη και δεν λήφθηκε υπόψιν στην αναζήτηση σας. [[Λεπτομέρειες](#)]

## Ιστός

Αποτελέσματα 1 - 10 από περίπου 2.630 για **buy da vinci code in greek bookstore**. (0,31 δευτερόλεπτο)

### [Errors in the Da Vinci Code](#)

... **Da Vinci Code** Dan Brown's **The Da Vinci Code** is a ... he weaves throughout **The Da Vinci Code** after allegedly ... Senate, or write extensive **Greek** philosophical treatises ...

[atheism.about.com/b/a/119227.htm](http://atheism.about.com/b/a/119227.htm) - 27k - [Αποθηκευμένη Σελίδα](#) - [Παρόμοιες σελίδες](#)

### [biblio.com - Bookstore information for Cresthouse Books](#)

... [Details] [**Buy Now**]. Dan Brown. **The Da Vinci Code**. Doubleday & Company, 2003. 2003 hardcover edition is in very... [Details] [**Buy Now**]. Cresthouse Books. ...

[iliad.biblio.com/bookstores/cresthouse.html](http://iliad.biblio.com/bookstores/cresthouse.html) - 12k - Συμπληρωματικό αποτέλεσμα - [Αποθηκευμένη Σελίδα](#) - [Παρόμοιες σελίδες](#)

### [GoUK.com Bookstore](#)

... examinations, books, **bookstore**, the **da vinci code**, dan brown ... book value, music book, **buy** book, online ... of literature, world literature, **greek** literature, reader ...

[www.gouk.com/books/bestsellers/bs0.asp](http://www.gouk.com/books/bestsellers/bs0.asp) - 34k - [Αποθηκευμένη Σελίδα](#) - [Παρόμοιες σελίδες](#)

### [Christian Book Reviews](#)

... **The Da Vinci Code** by Dan Brown: **Buy** this book from Amazon ... and save 20%; Our Rating: Mr. **Da Vinci** is thinking ... Our Rating: Recommended; UK customers: **buy** this book. ...

[www.tektonics.org/books/shop02.html](http://www.tektonics.org/books/shop02.html) - 101k - 3 Ιαν 2005 - [Αποθηκευμένη Σελίδα](#) - [Παρόμοιες σελίδες](#)

### [Amazon.com: Books: The Da Vinci Code \[UNABRIDGED\]](#)

... Total List Price: \$70.90 **Buy** Together Today: \$45.94 ... Poetics," his famous study of **Greek** dramatic art ... to Dan Brown's bestseller, "**The Da Vinci Code**," that has ...

[www.amazon.com/exec/obidos/tg/detail/-/0739313126?v=glance](http://www.amazon.com/exec/obidos/tg/detail/-/0739313126?v=glance) - 90k - 2 Ιαν 2005 - [Αποθηκευμένη Σελίδα](#) - [Παρόμοιες σελίδες](#)

### [Amazon.com: Books: Matthew As Story](#)

... Total List Price: \$38.00 **Buy** Together Today: \$26.60. ... Text (The New International **Greek** Testament Commentary ... questions raised by **The Da Vinci Code**: by flanneryfan ...

[www.amazon.com/exec/obidos/tg/detail/-/0800620992?v=glance](http://www.amazon.com/exec/obidos/tg/detail/-/0800620992?v=glance) - 58k - Συμπληρωματικό αποτέλεσμα -

[Αποθηκευμένη Σελίδα](#) - [Παρόμοιες σελίδες](#)

[ [Περισσότερα αποτελέσματα από το www.amazon.com](#) ]

### [DiversityLink Bookstore](#)

... 351%, Harry Potter and the Philosopher's Stone (Ancient **Greek** Edition) by JK ... State Of Fear by Michael Crichton; **The Da Vinci Code** by Dan ... Redeem or **buy** a gift ...

[www.diversitylink.com/.../exec/obidos/tg/browse/-/283155/ref%3Dtab\\_b\\_b\\_3/102-9681143-7118539](http://www.diversitylink.com/.../exec/obidos/tg/browse/-/283155/ref%3Dtab_b_b_3/102-9681143-7118539) - 75k -

[Αποθηκευμένη Σελίδα](#) - [Παρόμοιες σελίδες](#)

### [Collins Booksellers, Online Bookstore](#)

Collins online **bookstore**, **Buy** books online from Collins ... Hardback [ More Info ], The

- Επέκταση του ΠΙ:
  - Κατεύθυνση: δόμηση πληροφορίας
  - Επεξεργάσιμο και
  - ερμηνεύσιμο από Η/Υ
  - Στόχος: αυτοματοποίηση λειτ.
- Τεχνολογίας όχι μόνο για το διαδίκτυο
  - γλώσσες αναπαράστασης
  - εργαλεία επισημείωσης
  - μηχανές συλλογισμού
  - ...



- Σύνδεση πόρων του ΠΙ με καλά ορισμένα *μεταδεδομένα*
- **Μεταδεδομένα<sup>1</sup>**: *Δομημένα* δεδομένα που αφορούν *αντικείμενα* και υποστηρίζουν *λειτουργίες* σχετιζόμενες με αυτά.
  - Αντικείμενα: ενέργειες, πρόσωπα, σχέσεις, περιοχές κτλ.
  - Λειτουργίες: διαχείριση και ανακάλυψη πόρων, χρήση κτλ.



- Title
- Creator
- Subject and Keywords
- Description
- Publisher
- Contributor
- Date
- Resource Type
- Format
- Resource Identifier
- Source
- Language
- Relation
- Coverage
- Rights Management

**ISO 15836-2003**

**NISO Z39.85-2001**

**CWA 13874**

---

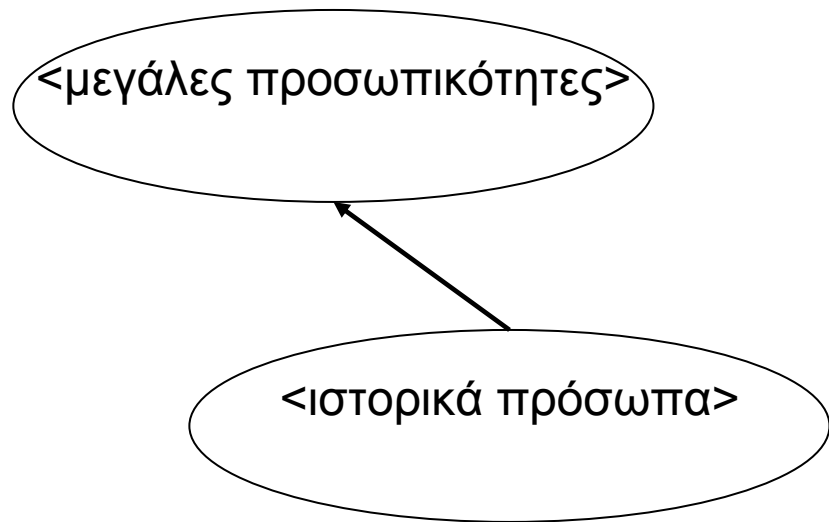
```
<Resource ...this_presentation>  
<Title>Εισαγωγή στον Σημασιολογικό  
  Ιστό</Title>  
<Creator>Αλέξανδρος  
  Βαλαράκος</Creator>  
<Date>19/1/2005</Date>  
<Language>Ελληνική</Language>  
...  
</Resource>
```

# Το μυστικό της επιτυχίας; Μεταδεδομένα!

- χρήση *μεταδεδομένων* για την επισημείωση περιεχομένου
  - *μεταδεδομένα* εκφράζουν το νόημα του περιεχομένου
  - **(ΠΡΟΒΛΗΜΑ)** όμως όλοι πρέπει να μιλάνε την ίδια γλώσσα...
  - **(ΠΡΟΒΛΗΜΑ)** η *αναφορά* σε ένα μεταδεδομένο δεν *αποκαλύπτει το νόημα* του περιεχομένου σε έναν Η/Υ!

# Μυστικό στην επισημείωση: στρωμάτωση μεταδεδομένων!

- Στρωμάτωση (Layering) μεταδεδομένων
  - Ταξινόμηση & σύνδεση μεταδεδομένων



<πόρος ....>

<ιστορικό  
πρόσωπο>M.Αλέξανδρος</ιστορικό  
πρόσωπο></πόρος>

<πόρος ....> <ιστορικό πρόσωπο>  
M.Αλέξανδρος

<έζησε>343-323π.Χ.</έζησε>

<φύλο>αρσενικό</φύλο>

...

</ιστορικό πρόσωπο></πόρος>

- Στρωμάτωση μεταδεδομένων μεταξύ τους
  - Σύνδεσή τους με στόχο να αποκτήσουν νόημα τα μεταδεδομένα!

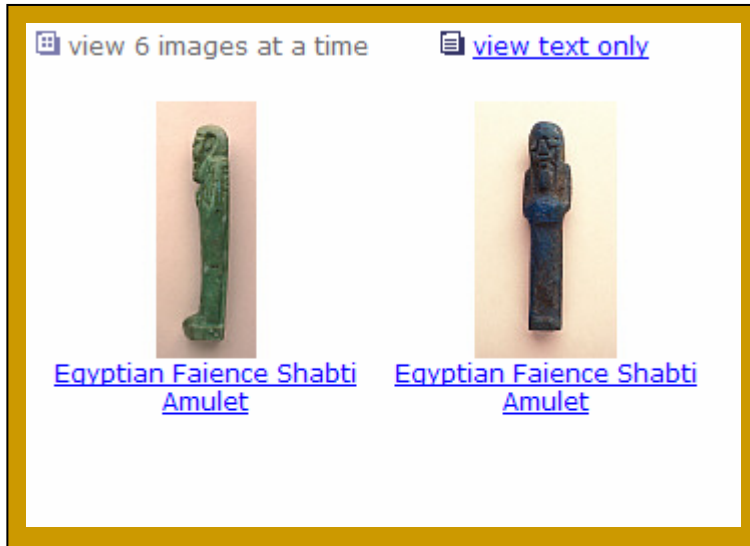
<Αριστοτέλης> είναι <άνθρωπος>

Ο <άνθρωπος> έχει <ποδια>=2

Κτλ.

- Όταν τα μεταδεδομένα *δομούνται* για να αποκτήσουν *νόημα* και να *περιγράψουν* μια *θεματική περιοχή* == **οντολογία**

# Ιστοσελίδες όχι μόνο για τους ανθρώπους



- Ο ΠΙ βασίζεται στο HTML για την στοιχειοθέτηση του περιεχομένου μιας ιστοσελίδας
- Πλέον ο Η/Υ δεν έχει να «μαντέψει» τη σημασιολογία των μεταδεδομένων

```
..<img  
src='../utilitybar/previous10b.gif'  
border='0' alt='/'></td>  
<td width="3"><img  
src='../images/px.gif' width="3"  
height="1" border="0" alt=""/></tr>  
<a href=./egt76.htm>Egyptian  
Faience Shabti Amulet</a>..
```

```
..<exhibit>
```

```
<id>76</id>
```

```
<image><img .../> </image>
```

```
<title>Egyptian Faience Shabti  
Amulet</title>
```

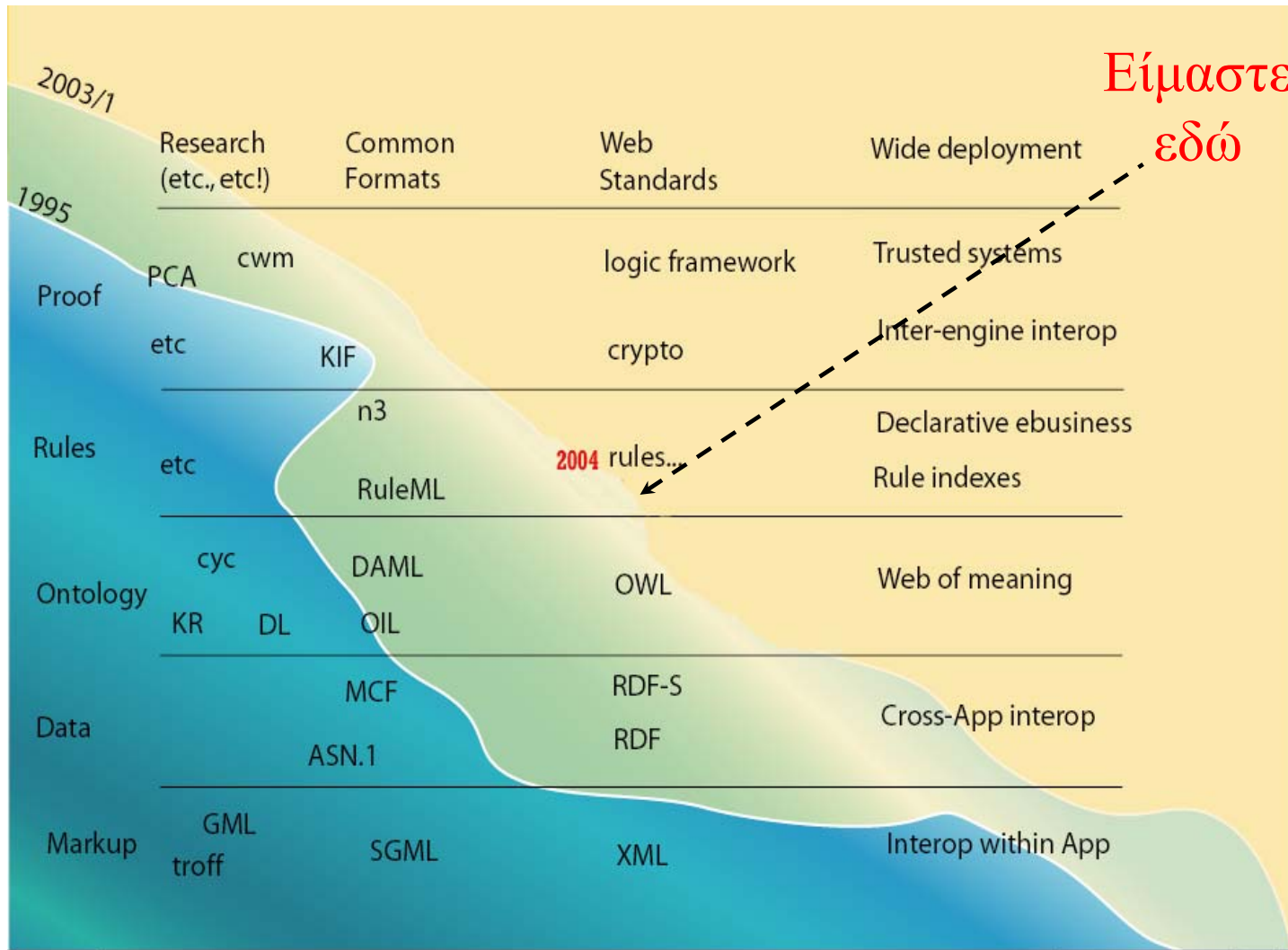
```
</exhibit>..
```



**Οντολογία**

- Προσδιορισμός της γλώσσας αναπαράστασης δεδομένων
  - τυποποίηση σημασιολογίας μεταδεδομένων
- Προσδιορισμός και ανάπτυξη τεχνολογίας υποδομής του ΣΙ
  - προς την πραγματοποίηση του ΣΙ
- Ανάπτυξη πρότυπων εφαρμογών
  - αξιολόγηση υπάρχουσας τεχνολογίας
  - προσδιορισμός αδυναμιών
  - ανάδειξη νέων αναγκών - τεχνολογιών

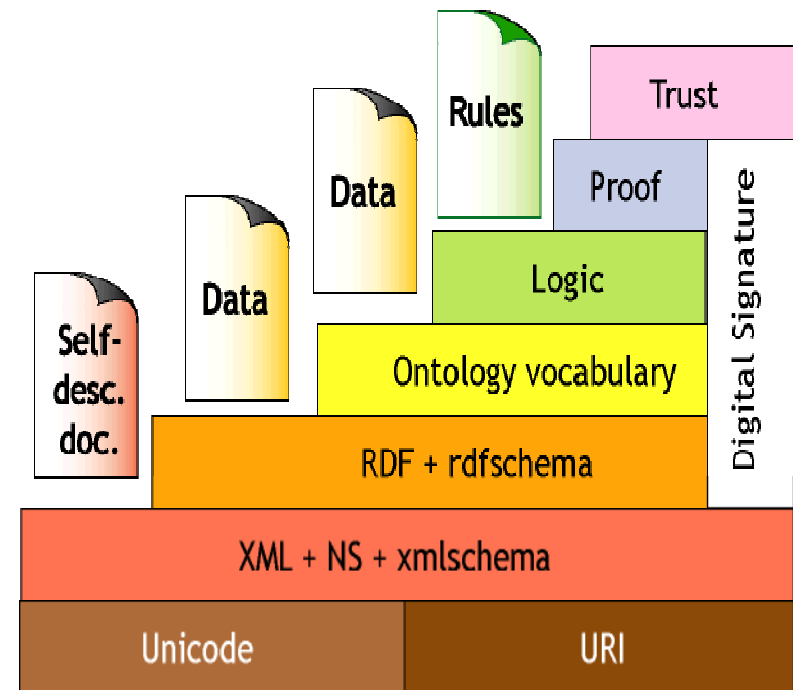
# Το σημασιολογικό κύμα



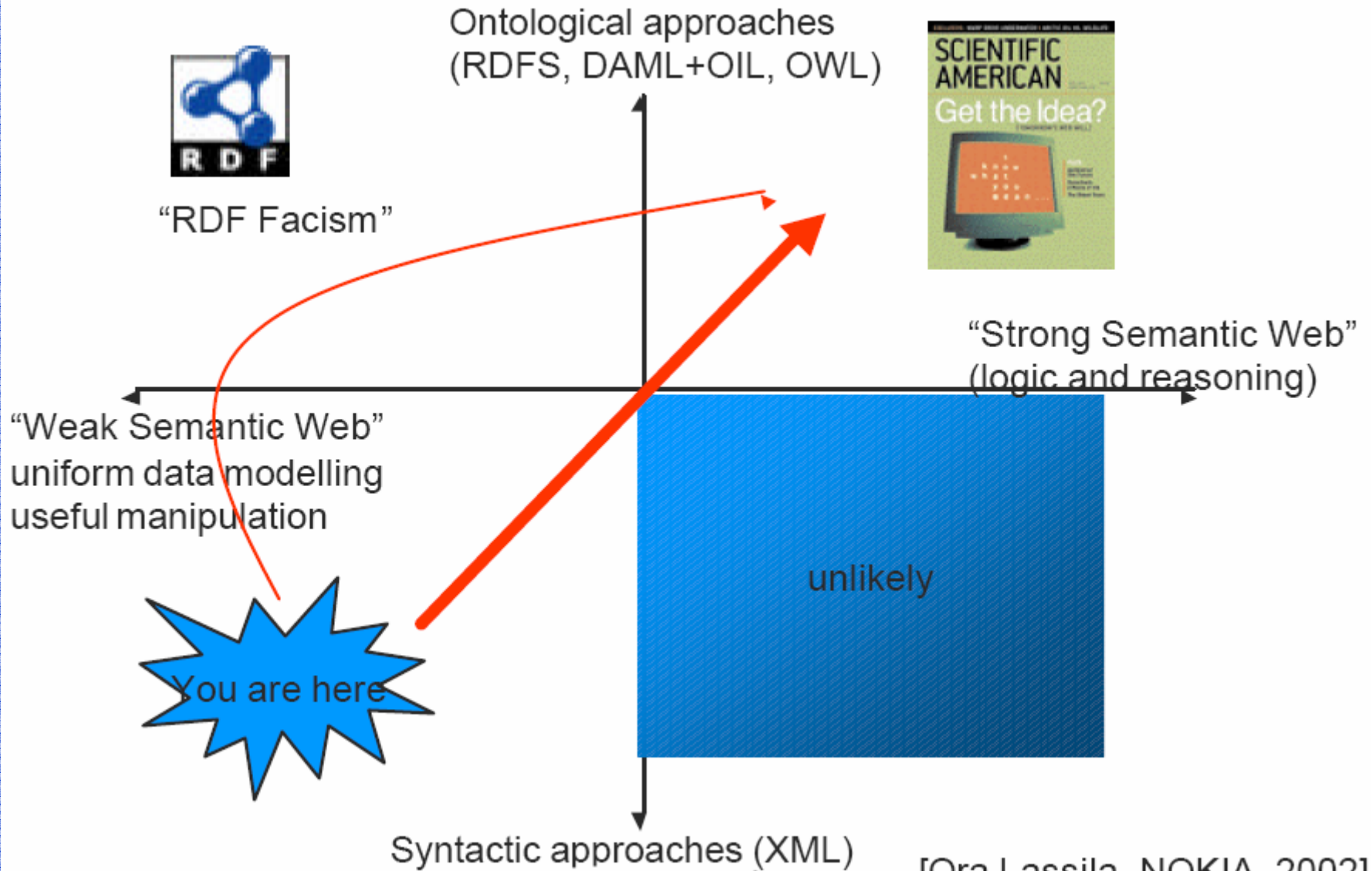
(Berners-Lee, 03)

# Αρχιτεκτονική Σημασιολογικού Ιστού

- Συντακτικό - XML
- Γραμματική – XML Schema
- Οντολογία *(Περισσότερα στο 2<sup>ο</sup> μέρος)*
- Γλώσσες στηριζόμενες στη λογική (1<sup>ο</sup> βαθμού)
- Ανάπτυξη του Λογικού επιπέδου



# Αναπαράσταση Δεδομένων: XML ή RDF



[Ora Lassila, NOKIA, 2002]

## Δήλωση

Da Vinci **ι** is the **Creator** of the Resource “<http://www.moyseio.org/DaVinci/MonaLisa>”

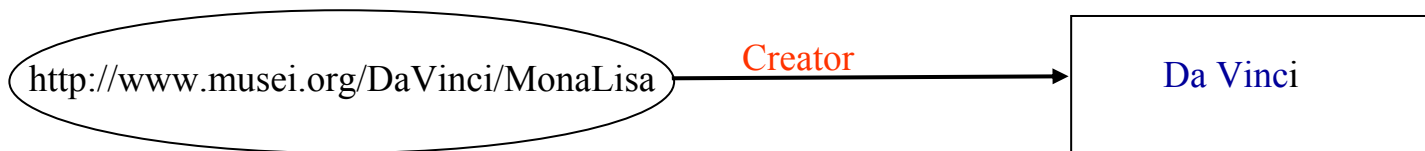
## Δομή

Πόρος (subject) <http://www.moyseio.org/DaVinci/MonaLisa>

Ιδιότητα (predicate) <http://www.moyseio.org/DaVinci/MonaLisa/#Creator>

Τιμή (object) “Da Vinci”

## Γράφος



- Εργαλεία
- Επισημείωση εγγράφων
  - χειρωνακτική (βαρετή, κουραστική, χρονοβόρα)
  - ημι-αυτόματη (μια υπολογίσιμη βοήθεια)
  - αυτόματη (ιδεατό αλλά ...)
- Επισημείωση δυναμικών σελίδων (Deep Web)
  - δυναμική επισημείωση (ποιος περιμένει...)
  - μετατροπή σχημάτων σε οντολογίες ή
  - χρήση μεσολαβητών



# Σημαιολογικός Ιστός: πλεονεκτήματα & μειονεκτήματα

- Ο ΣΙ δεν λύνει οριστικά το πρόβλημα
  - Πλεονεκτήματα:
    - Δομή το εννοιολογικό Ιστό
    - Εμπλουτίζει το περιεχόμενο με μεταδεδομένα,
    - Υιοθέτει οντολογίες για την ερμηνεία και επεξεργασία του περιεχομένου από τον Η/Υ.
    - Αυτόματη επεξεργασία περιεχομένου
  - Μειονέκτημα
    - Single-point-of-failure απειλή
    - Η C/S αρχιτεκτονική δεν ευνοεί μεγάλου κλίμακας εφαρμογές
    - Ποιος δημιουργεί μεταδεδομένα
    - Ποιος επισημειώνει περιεχόμενο
    - Σημαιολογική ανομοιογένεια



- Agents
- Databases Technologies
- Natural Language Processing
- Information Extraction
- Knowledge representation & reasoning
- Software engineering
- Web mining
- Trust & security
- Visualization & modeling
- User Interfaces
- ...

- Επισημείωση ιστοσελίδων (τουλάχιστον ημι-αυτόματη)
- Αυτόματη παραγωγή μεταδεδομένων
- Δημιουργία και συντήρηση οντολογιών
- Αναζήτηση πληροφορίας με τη χρήση οντολογιών. *(2<sup>ο</sup> μέρος)*