

# Εργαστήριο Σημασιολογικού Ιστού

## Ενότητα 1: Σημασιολογία και Μεταδεδομένα

Μ.Στεφανιδάκης

10-2-2017

# Η αρχή: Το όραμα του Σημασιολογικού Ιστού

Tim Berners-Lee, James Hendler and Ora Lassila, “The Semantic Web”, Scientific American, May 2001, p. 29-37.

*...The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users...*

*...the effectiveness of such software agents will increase exponentially as more machine-readable Web content and automated services become available...*

*...this structure will open up the knowledge and workings of humankind to meaningful analysis by software agents, providing a new class of tools by which we can live, work and learn together.*

# Σημασιολογία – “semantics”

- ▶ Τι **σημαίνουν** τα δεδομένα;
  - ▶ Ποια η **έννοιά** τους;
- ▶ **Μετάδοση** έννοιας και ενσωμάτωση στη **γνώση**..
  - ▶ ..έχοντας ως αποτέλεσμα την κατάλληλη **δράση**!
- ▶ Πώς θα γίνει αντιληπτή από τις “**μηχανές**”;
  - ▶ Ός άνθρωποι, χρησιμοποιούμε εύκολα την **έμμεση** σημασιολογία
  - ▶ Για τις εφαρμογές όμως απαιτείται **ρητή** δήλωση της σημασιολογίας

# Έμμεση Σημασιολογία

**Παράδειγμα:** ονόματα συναρτήσεων ή μεταβλητών

```
quicksort :: (Ord a) => [a] -> [a]
quicksort [] = []
quicksort (x:xs) =
    let smallerOrEqual = [t|t<=xs,t<=x]
        greater = [t|t<=xs,t>x]
    in quicksort smallerOrEqual ++ [x] ++ quicksort greater
```

- ▶ Τι καταλαβαίνετε;
  - ▶ Ανεξάρτητα από το αν γνωρίζετε τη γλώσσα!
  - ▶ Τα σχόλια θα βοηθούσαν περαιτέρω στη μεταφορά σημασιολογίας..
- ▶ Τι μπορεί όμως να καταλάβει η “μηχανή”;

# Ιστοσελίδα και μηχανή

Τι “αντιλαμβάνεται” η μηχανή:

```
<!doctype html>
<html>
  <head>
    <meta charset="utf-8">
    <title>????????????????</title>
  </head>

  <body>
    <h1>????????????????</h1>
    ?????????????????????
    ?????????????????????
  </body>
</html>
```

- ▶ Προσοχή όμως: αν και δεν “αντιλαμβάνεται” το νόημα του περιεχομένου του τίτλου, η μηχανή ξέρει την έννοια “τίτλος”!
  - ▶ π.χ. έχει ιδιαίτερο βάρος στις μηχανές αναζήτησης

# Ποια πληροφορία βλέπετε εσείς;



- ▶ Τι θα μπορούσε να καταλάβει η “μηχανή” αν είχε τη δυνατότητα για:
  - ▶ OCR (αναγνώριση κειμένου)
  - ▶ Αναγνώριση εικονιζόμενων
  - ▶ Αναγνώριση σχεδίων κειμένου (τηλέφωνα, ονόματα..)
  - ▶ Αναγνώριση λεξιλογίου (“μηνών”, “εμβολιασμέν-”, “μέσα σπίτι”..)

# Κλασική χρήση δεδομένων

- ▶ Στο κλασικό σενάριο χρήσης, τα δεδομένα είναι “εσωτερική υπόθεση” του παραγωγού τους
  - ▶ Αποθηκευμένα σε βάσεις δεδομένων απροσπέλαστες από τρίτους (“data silos”)
  - ▶ Χρησιμοποιούνται από εφαρμογές που έχει γράψει ο ίδιος ο παραγωγός των δεδομένων
  - ▶ Για να εμφανιστούν στο web μετασχηματίζονται από τον ίδιο τον παραγωγό σε ιστοσελίδες
  - ▶ Δεν χρειάζεται μετάδοση της σημασιολογίας των δεδομένων γιατί είναι γνωστή εκ των προτέρων!

# Οι αλλαγές που έφερε το Web

- ▶ Σήμερα όλο και περισσότερες εφαρμογές βρίσκονται καταναμεμημένες στον παγκόσμιο ιστό
- ▶ Τα δεδομένα **αποσυνδέονται** από τις εφαρμογές
  - ▶ Ο παραγωγός και ο καταναλωτής των δεδομένων **δεν είναι κατ' ανάγκη ο ίδιος!**
- ▶ Μπορούμε να συνεχίσουμε με το κλασικό μοντέλο χρήσης των δεδομένων;
  - ▶ Μπορούν τα δεδομένα να χρησιμοποιηθούν από τρίτους;
  - ▶ Σε εφαρμογές που δεν γνωρίζει ο παραγωγός των δεδομένων;
  - ▶ Από εφαρμογές που δεν σχεδιάστηκαν ειδικά για τα δεδομένα αυτά;
- ▶ Αυτός είναι ο στόχος των Σημασιολογικών Δεδομένων



# Σημασιολογικά Δεδομένα

- ▶ **Ρητά δηλωμένα και προσπελάσιμη** και η σημασιολογία των δεδομένων
  - ▶ Μόνο τότε μπορούν να χρησιμοποιηθούν από εφαρμογές που **προσαρμόζουν τη συμπεριφορά** τους ανάλογα με το τι σημαίνουν τα δεδομένα
  - ▶ Εκτελώντας ενέργειες πάνω σε σετ δεδομένων για τα οποία **δεν σχεδιάστηκαν εξ' αρχής**
- ▶ Μεγαλεπήβολος στόχος;
  - ▶ “τεχνητή νοημοσύνη”;
- ▶ Ποια τα στοιχεία που απαιτούνται για την υλοποίησή του;

# Σημασιολογία και Μεταδεδομένα

- ▶ Μεταδεδομένα:
  - ▶ “δεδομένα για/σχετικά με/που αφορούν τα δεδομένα”
  - ▶ Είναι ακριβώς εκείνα που μεταδίδουν τη σημασιολογία των κανονικών δεδομένων
- ▶ Στις εφαρμογές του Σημασιολογικού Ιστού
  - ▶ Τα μεταδεδομένα μεταδίδονται ρητά μαζί με τα δεδομένα
  - ▶ Έχουν την ίδια αξία και είναι το ίδιο απαραίτητα με τα δεδομένα

# Διαθεσιμότητα στον παγκόσμιο ιστό

- ▶ Ο παγκόσμιος ιστός είναι η **de facto** πλατφόρμα ανταλλαγής δεδομένων
  - ▶ Τα δεδομένα πρέπει να είναι **απευθείας προσπελάσιμα** στο web
  - ▶ Σε μια **πρότυπη** μορφή **εύκολα επεξεργάσιμη** από τις μηχανές
  - ▶ **“Ωμά δεδομένα”** – δεν πρέπει να προσφέρονται “ερμηνευμένα” από τον πάροχό τους
    - ▶ Αυτό είναι πλέον αρμοδιότητα του καταναλωτή!

# Σημασιολογικό μοντέλο οργάνωσης

- ▶ Τα σημασιολογικά δεδομένα, όπως κάθε είδος δεδομένων, χρειάζονται ένα **μοντέλο οργάνωσης**
  - ▶ Data model
  - ▶ Το μοντέλο θα πρέπει να είναι απλό, ευέλικτο και να παρέχει τη μέγιστη δυνατότητα αλληλοδιασύνδεσης μεταξύ των δεδομένων
- ▶ Ο Σημασιολογικός Ιστός χρησιμοποιεί την οργάνωση σε **τριάδες** (triples)
  - ▶ Στη μορφή (**υποκείμενο, κατηγορημα, αντικείμενο**)

# Τα δεδομένα ως οντότητες

- ▶ Οι σημασιολογικές τριάδες έχουν την ισχύ μιας “δήλωσης” (assertion)
  - ▶ Ότι μια **οντότητα** (entity), για συγκεκριμένη **ιδιότητα** (attribute), έχει ορισμένη **τιμή** (value)
  - ▶ Τα δεδομένα ως οντότητες
- ▶ Απαιτείται μια μονοσήμαντη ονομασία των οντοτήτων
  - ▶ Σε τι “πράγμα” αναφερόμαστε;
  - ▶ Με παγκόσμια ισχύ – απαραίτητο για δεδομένα που ανταλλάσσονται στον παγκόσμιο ιστό!

# Παραγωγή νέας γνώσης

- ▶ Ο συνδυασμός δεδομένων και ρητά δηλωμένων μεταδεδομένων ισοδυναμεί με μια σειρά δηλώσεων
  - ▶ Ότι “κάτι ισχύει”
- ▶ Η γνώση που περιλαμβάνεται σε ένα τέτοιο σύνολο δηλώσεων είναι πεπερασμένη
- ▶ Μπορούμε με τη βοήθεια κανόνων να συμπεράνουμε νέα γνώση
  - ▶ Ισοδυναμεί με την παραγωγή νέων τριάδων από τις ήδη υπάρχουσες
- ▶ Χρειαζόμαστε πρότυπα και γλώσσες για την έκφραση οντολογικών κανόνων

# Ο Σημασιολογικός Ιστός και τα Συνδεδεμένα Δεδομένα

- ▶ Ο **Σημασιολογικός Ιστός** είναι το σύνολο των προτύπων που καθορίζουν το μοντέλο οργάνωσης και τις γλώσσες για την ανάπτυξη σημασιολογικών δεδομένων
  - ▶ Πρότυπα του οργανισμού W3C
- ▶ Τα **Συνδεδεμένα Δεδομένα** είναι ένα σύνολο καλών πρακτικών για την ανάπτυξη σημασιολογικών δεδομένων και εφαρμογών στον παγκόσμιο ιστό.
  - ▶ Η πρακτική πλευρά του Σημασιολογικού Ιστού
  - ▶ Δίνεται μεγάλη έμφαση στη **διασυνδεσιμότητα** και την **ανοικτότητα**
  - ▶ Έτσι συχνά μιλάμε για **Ανοικτά Συνδεδεμένα Δεδομένα** (Linked Open Data)