



ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

# ΤΕΧΝΟΛΟΓΙΕΣ ΔΙΑΧΕΙΡΙΣΗΣ ΠΟΛΙΤΙΣΜΙΚΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

## Το όραμα του Σημασιολογικού Ιστού

*Γιώργος Καρυδάκης  
Τμήμα Πολιτισμικής Τεχνολογίας και  
Επικοινωνίας*



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

# Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



# Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αιγαίου**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



# Ενότητα 2: Το όραμα του Σημασιολογικού Ιστού

Τεχνολογίες Διαχείρισης Πολιτισμικής  
Πληροφορίας  
Γ. Καρυδάκης

# Το διαδίκτυο σήμερα

- Κατάλληλο για κατανάλωση από ανθρώπινους χρήστες
- Ακόμα και η δομημένη πληροφορία χάνεται κατά την παρουσίαση
- Η αναζήτηση βασίζεται ακόμα σε λέξεις κλειδιά και κειμενική ανάλυση
- Ακόμα και αυτή η φτωχή σε εκφραστικότητα ανάλυση αρκεί και καθιστά το www επιτυχές
- Πρόσβαση στο περιεχόμενο από τις μηχανές

# Προβλήματα

- Αναζήτηση πληροφοριών
  - Λέξεις κλειδιά
- Εξαγωγή πληροφοριών
  - Αναγκαία ανθρώπινη παρέμβαση
  - Ομοίως για περιήγηση, ανάκτηση, ερμηνεία
- Συντήρηση πληροφορίας
  - Ασυνέπειες στην ορολογία
  - Μη επικαιροποίηση
- Προβολή πληροφοριών
  - Όψεις

# Προβλήματα της αναζήτησης με λέξεις κλειδιά

- Υψηλή ανάκληση (recall) και χαμηλή ακρίβεια (precision)
- Ή χαμηλή ανάκληση
- Τα αποτελέσματα είναι ευαίσθητα στο λεξικό
- Απαιτείται το φιλτράρισμα των αποτελεσμάτων από τον χρήστη
- Φανταστείτε τις παρακάτω φράσεις:  
Θα βρέξει αύριο

# Στόχοι του ΣΙ

- Αναπαράσταση του περιεχομένου με τρόπο κατανοητό από τις μηχανές
- Εκμετάλλευση της αναπαράστασης αυτής για την ολοκλήρωση της γνώσης
- Η γνώση να οργανωθεί σε εννοιολογικό χώρο σύμφωνα με το νόημά της
- Αυτοματοποιημένα εργαλεία για ... γνώσης
  - συντήρηση
  - ανακάλυψη
- Απάντηση ερωτήσεων



# Περίπτωση χρήσης

- Τρέχον ηλεκτρονικό μέτα-κατάστημα
  - Βασίζονται σε ελεγχόμενες αρχές
  - Απαιτεί εκτεταμένο και προσαρμοσμένο προγραμματισμό
  - Η συντήρηση είναι εξαιρετικά ακριβή
  - Βασίζονται στην ανάλυση του κειμένου
    - Σφάλματα
    - Φτωχή περιγραφή και περιεχόμενο

# Περίπτωση χρήσης

- Σημασιολογικό ηλεκτρονικό μέτα-κατάστημα
  - πράκτορες λογισμικού να ερμηνεύουν
    - πληροφορίες
    - Υπηρεσίες
  - Δεν απαιτεί προγραμματισμό αφού αποτελείται από ερωτήματα
  - Η υποδομή βασίζεται σε πρότυπα
  - Σημασιολογική αναζήτηση

# Τεχνολογίες

- Ρητά Μεταδεδομένα
- Οντολογίες
- Λογική
- Συμπερασμός
- Πράκτορες

# Οντολογία

- Προέρχεται από τη φιλοσοφία
  - Η μελέτη της φύσης της ύπαρξης
- Στην επιστήμη των υπολογιστών
  - Ρητή και τυπική προδιαγραφή μιας έννοιας

# Συστατικά οντολογίας

- Όροι
  - σημαντικές έννοιες (κλάσεις) της γνωστικής περιοχής
  - π.χ. καθηγητές, το προσωπικό, τους φοιτητές, τα μαθήματα, τμήματα
- Σχέσεις
  - μεταξύ των όρων ως ιεραρχίες κλάσεων
  - π.χ. οι καθηγητές είναι μέλη του προσωπικού
- Ιδιότητες:
  - π.χ. ώρες μαθήματος
- Περιορισμοί
  - π.χ. μόνο τα μέλη ΔΕΠ μπορούν να διδάξουν μαθήματα
- Δηλώσεις μη συσχέτισης (disjoint)
  - π.χ. σχολή και γενικά το προσωπικό είναι ξένα μεταξύ τους
- Λογικές σχέσεις
  - π.χ. κάθε τμήμα πρέπει να περιλαμβάνει τουλάχιστον 10 σχολή

# Ο ρόλος των οντολογιών στον Παγκόσμιο Ιστό

- Οι οντολογίες παρέχουν μια κοινή κατανόηση ενός πεδίου
- Σημασιολογική διαλειτουργικότητα
- Αντιμετωπίζουν αποτελεσματικά θέματα ορολογίας

# Τεχνολογίες οντολογιών RDF

- RDF είναι ένα μοντέλο δεδομένων για
  - αντικείμενα
  - σχέσεις μεταξύ αντικειμένων
- RDF Schema
  - γλώσσα περιγραφής λεξιλογίου
- Περιγράφει
  - ιδιότητες
  - κατηγορίες RDF

# Τεχνολογίες οντολογιών OWL

- Πλουσιότερη γλώσσα οντολογίας
  - σχέσεις μεταξύ κλάσεων
    - π.χ., συσχέτιση
  - πληθικότητα
    - π.χ. "ακριβώς ένα"
  - ...



# Λογική

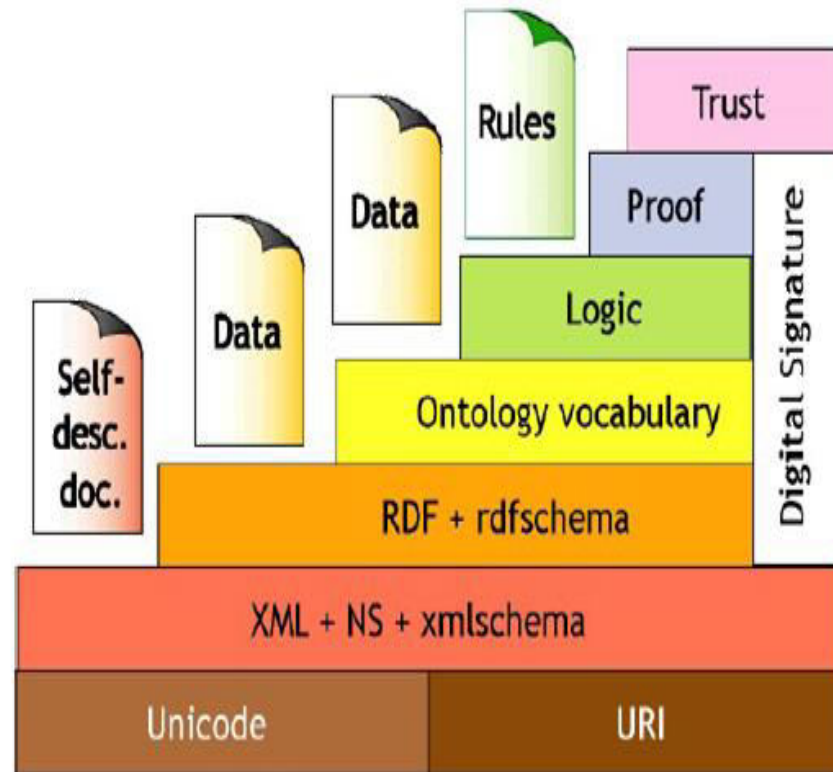
- Μελετά τις αρχές του συλλογισμού
- Έκφραση τυπικής γνώσης
- Συμπερασμός (reasoning)
  - Αυτοματοποιημένη εξαγωγή συμπερασμάτων από την δεδομένη γνώση
  - π.χ. ο messi είναι αθλητής αφού κάθε ποδοσφαιριστής είναι αθλητής

# Λογική έναντι Οντολογίες

- Λογική μπορεί να αποκαλύψει οντολογική γνώση που έχει δεν έχει δηλωθεί ρητά
- Η λογική μπορεί να εξάγει γνώση πέρα από την ρητή δήλωση της μέσω οντολογιών
- Μπορεί επίσης να βοηθήσει στην αποκάλυψη απρόσμενες σχέσεις και ασυνέπειες
- Η λογική είναι πιο γενική από οντολογίες

• Μπορεί επίσης να ανακαλύψει ασυνέπειες

# Επίπεδα ΣΙ



# Επίπεδα ΣΙ

- XML
  - Συντακτική βάση
- RDF
  - βασικό μοντέλο δεδομένων
  - Schema απλή γλώσσα οντολογίας
- Οντολογία
  - Πιο εκφραστική γλώσσα από RDF Schema
  - OWL

# Επίπεδα ΣΙ

- Λογική
  - ενισχύει περαιτέρω τις γλώσσες οντολογίας μέσω κανόνων
- Απόδειξη
  - παραγωγή, ανταλλαγή, επικύρωση
- Εμπιστοσύνη
  - υπογραφές, συστάσεις, αξιολόγηση

# Τι δεν είναι ο ΣΙ

- Δεν προσπαθεί να επιβάλει
  - Έννοιες
  - Νόημα
  - Ερμηνεία
- Δεν (χρειάζεται) να είναι
  - Τεχνικά
  - Επισημειωμένο χειροκίνητα

# Επιβολή νοήματος

- Επιβολή νοήματος στους χρήστες μέσω OWL πρότυπα και RDFS
- Δεν επιβάλλουν έννοιες
- Προτείνει γλώσσα στην οποία οι χρήστες μπορούν να εκφράσουν τις δικές τους έννοιες
- Ελεύθερη επιλογή πεδίου και λεξιλογίου
- Παράδειγμα από τον σημερινό ιστό (HTML)
  - δεν επιβάλλει το περιεχόμενο
  - γλώσσα περιγραφής περιεχομένου
  - σύμβαση τυποποιημένης γλώσσας

# Επιβολή ερμηνείας

- Όλοι πρέπει να υιοθετήσουν προκαθορισμένη ερμηνεία των όρων που χρησιμοποιούν
- Το σύνθημα του Σημασιολογικού Ιστού δεν είναι η επιβολή μιας ενιαίας οντολογίας, αλλά μάλλον "ας ανθίσουν χίλιες οντολογίες"
- Η αντιστοίχισης γνώσης είναι κεντρικό θέμα στον ΣΙ



# Τεχνικά πολύπλοκο

- Οι χρήστες πρέπει να κατανοήσουν τις περίπλοκες λεπτομέρειες των τεχνολογιών αναπαράσταση γνώσης
- Αλήθειες
  - Οι τεχνολογίες είναι πολύπλοκες
  - Η σημασιολογία και η αναπαράσταση γνώσης είναι περίπλοκες διεργασίες
  - Η ανάπτυξη οντολογιών είναι ένας εξειδικευμένο τομέας της Μηχανικής Γνώσης
- Αλλά

# Χειροκίνητη επισημείωση

- Ο ΣΙ απαιτεί την χειροκίνητη σήμανση όλων των υφιστάμενων ιστοσελίδων
- Δεν αναπτύσσονται ξεχωριστές εκδόσεις σελίδων
- Μικρή έκταση πληροφορίας μπορεί να οδηγήσει σε παραγόμενη γνώση

# Τρέχουσα Κατάσταση

- Από πού προέρχονται τα μεταδεδομένα;
- Από πού προέρχονται οι οντολογίες;
- Πως αντιμετωπίζονται οι πολλές οντολογίες;
- Παραμένει ο ιστός στον ΣΙ;

# Από πού προέρχονται τα μεταδεδομένα;

- Μεγάλο μέρος των μεταδεδομένων προέρχονται από την επεξεργασία φυσικής γλώσσας και την μηχανική μάθηση
- Είναι δυνατόν με την υπάρχουσα τεχνολογία να παραχθεί σημασιολογική σήμανση για μεγάλο όγκο ιστοσελίδων με όρους από εκτενής και πλούσιες οντολογίες
- Το Web2.0 συνεισφέρει και στο Web3.0

# Από πού προέρχονται οι οντολογίες;

- Μεγάλο φάσμα σημασιολογικών δομών
- Μεγάλο εύρος έκτασης και ανάλυσης
- Πηγές
  - Από την κοινότητα του ΣΙ
  - εταιρείες με αντίστοιχη οργάνωση πληροφορίας
  - οργανισμοί έχουν εσωτερική γλωσσάρια
  - επιστημονικές κοινότητες με σχήματα μεταδεδομένων

# Πως αντιμετωπίζονται οι πολλές οντολογίες;

- Μα αυτή είναι η ιδέα!!!
- Ο ΣΙ στηρίζεται σε μεγάλο βαθμό στην ύπαρξη και την δυνατότητα ενσωμάτωσης πολλαπλών οντολογιών
- Θέματα οντολογιών
  - ευθυγράμμισης
  - αντιστοίχισης
  - ολοκλήρωσης
- Επιλύονται με μεθόδους
  - Επεξεργασίας φυσικής γλώσσας
  - Μηχανικής μάθησης
  - Θεωρίας γράφων
  - Στατιστικής
- Δεν είναι πλήρως λυμένο...

# Παραμένει ο ιστός στον ΣΙ;

- Ίσχυε παλαιότερα όταν η αναπαράσταση γινόταν για στενότερα πεδία
- Πρόσφατα παρατηρείται στροφή προς την ενίσχυση του δικτυακού στοιχείου
- Το όραμα είναι ένας καλύτερος ιστός