

15ο ΠΑΝΕΛΛΗΝΙΟ ΣΥΝΕΔΡΙΟ ΑΚΑΔΗΜΑΪΚΩΝ ΒΙΒΛΙΟΘΗΚΩΝ

***Διαχείριση οντολογιών: μελέτη και
εμβάθυνση στα βασικά προβλήματα που
την αφορούν και παρουσίαση
υπαρχουσών βιβλιοθηκών οντολογιών***

ΓΑΪΤΑΝΟΥ ΠΑΝΩΡΑΙΑ gaitanou@benaki.gr
ΓΕΡΓΑΤΣΟΥΛΗΣ ΜΑΝΟΛΗΣ manolis@ionio.gr

1 Νοεμβρίου 2006

Περιεχόμενα

- Από τον Παγκόσμιο Ιστό στο Σημασιολογικό Ιστό
- Ορισμοί και βασικά συστατικά οντολογιών
- Κατηγορίες και λόγοι δημιουργίας οντολογιών
- Συνδυασμός οντολογιών – Προβλήματα
- Υποστήριξη πολλαπλών εκδόσεων
- Εξέλιξη οντολογιών
- Βιβλιοθήκες οντολογιών
- Ερευνητικές κατευθύνσεις
- Παρατηρήσεις - Συμπεράσματα

Από τον Παγκόσμιο Ιστό στο Σημασιολογικό Ιστό (1/2)

- Παγκόσμιος Ιστός (Tim Berners-Lee) = σύνολο από ιστοσελίδες και διασυνδέσεις → τρόπος αναπαράστασης της πληροφορίας.
- Πληροφορία σήμανσης (markup information) = πληροφορίες που αφορούν στην παρουσίασή της (μέγεθος γραμματοσειράς, χρώμα κλπ.) → οι διασυνδέσεις γίνονται σε σχετικό περιεχόμενο.
- Υπολογιστές → δεν επεξεργάζονται την πληροφορία, είναι υπεύθυνοι μόνο για τον τρόπο παρουσίασης.
- Χρήστες → ερμηνεύουν την πληροφορία (σημασιολογικό περιεχόμενο).

Από τον Παγκόσμιο Ιστό στο Σημασιολογικό Ιστό (2/2)

- Ο τρέχων Ιστός → όχι μόνο μια απλή πλατφόρμα που επιτρέπει σε χρήστες και οργανισμούς να φιλοξενούν ιστοσελίδες.
- Δημιουργία ενός μηχανισμού που θα επιτρέπει στις μηχανές να συλλέγουν περιεχόμενο από διαφορετικές πηγές, να επεξεργάζονται την πληροφορία και να ανταλλάσσουν τα αποτελέσματα με άλλους πράκτορες και χρήστες.
- Οντολογίες: βασικό εργαλείο σημασιολογικής διαλειτουργικότητας:
 - λεξιλόγια για την περιγραφή εννοιών και των σχέσεών τους
 - εργαλεία αυτόματης "διερμηνείας" και συσχέτισης των μεταδεδομένων

Οντολογία - Ορισμοί

- Φιλοσοφικός ορισμός:

«Η επιστήμη που μελετά την ύπαρξη, τη φύση και τις σχέσεις των όντων (ιδιαίτερο σύστημα, σύμφωνα με το οποίο εξετάζονται τα προβλήματα της φύσης.)»

(Webster's Third New International Dictionary)

- Επιστημονικός ορισμός:

«Η οντολογία είναι ένας τυπικός, σαφής προσδιορισμός μιας διαμοιρασμένης εννοιολογικής αναπαράστασης».
(Gruber, 1993)

Ανάλυση ορισμού (Fensel, 2001)

- «**εννοιολογική αναπαράσταση**»: αφηρημένο μοντέλο ενός φαινομένου
- «**διαμοιρασμένη**»: ύπαρξη μιας συμφωνίας μεταξύ εκείνων που χρησιμοποιούν τις οντολογίες
- «**τυπικός**»: ακριβή, μαθηματικά, περιγραφή
- «**σαφής**»: ακρίβεια των εννοιών και των εμφανώς ορισμένων σχέσεών τους

Ταξινόμια # Οντολογία



Οντολογία - Βασικά συστατικά

- Κλάσεις: αναπαριστώνται με έννοιες
 - α) *πρωταρχικές έννοιες* (primitive concepts)
 - β) *έννοιες εξ' ορισμού* (defined concepts)
- Σχέσεις: εκφράζουν ένα είδος αλληλεπίδρασης μεταξύ των εννοιών ενός πεδίου.
- Συναρτήσεις: εκπροσωπούν μια ειδική περίπτωση σχέσης.
- Αξιώματα: χρησιμοποιούνται προκειμένου να αναπαριστούν προτάσεις που είναι πάντοτε αληθείς.
- Στιγμιότυπα: εκφράζουν συγκεκριμένα στοιχεία.

Κατηγορίες οντολογιών (1/2)

Σύμφωνα με το βαθμό τυπικότητας της αναπαράστασής της (Uschhold και Gruninger, 1996):

- Άτυπη (highly informal), εκφρασμένη σε φυσική γλώσσα.
- Ημι-άτυπη (semi-informal), π.χ. διατυπωμένη σε ένα περιορισμένο και δομημένο υποσύνολο κάποιας φυσικής γλώσσας.
- Ημι-τυπική (semi-formal), διατυπωμένη σε μια τεχνητή και αυστηρά ορισμένη γλώσσα.
- Αυστηρά τυπική (rigorously formal): ορισμοί όρων με αυστηρή σημασιολογία, θεωρήματα και αποδείξεις ιδιοτήτων, όπως η *ορθότητα* (soundness) και η *πληρότητα* (completeness).

Κατηγορίες οντολογιών (2/2)

- Οντολογίες αναπαράστασης γνώσης → οντότητες αναπαράστασης χωρίς να προσδιορίζουν τι συγκεκριμένο αναπαριστούν.
- Γενικές ή κοινές οντολογίες → γενική γνώση γύρω από τον κόσμο, βασικές έννοιες (χρόνος, χώρος, συμβάντα κλπ.)
- Οντολογίες ανώτερου επιπέδου → γενικές έννοιες, κάτω από τις οποίες συσχετίζονται όλοι οι όροι σε ήδη υπάρχουσες οντολογίες.
- Οντολογίες μεταδεδομένων → λεξιλόγιο για την περιγραφή του περιεχομένου πληροφορίας, η οποία είναι ηλεκτρονικά διαθέσιμη.
- Οντολογίες πεδίου ορισμού → γνώση γύρω από ένα συγκεκριμένο πεδίο, π.χ. ιατρική κλπ.
- Οντολογίες μεθοδολογίας ή εργασιών → όροι που αναφέρονται σε συγκεκριμένες εργασίες, π.χ. διάγνωση κλπ.

Λόγοι δημιουργίας οντολογιών

- Διαμοιρασμός της κοινής αντίληψης της δομής της πληροφορίας ανάμεσα σε ανθρώπους ή υπολογιστικούς πράκτορες.
- Επαναχρησιμοποίηση της πληροφορίας του τομέα.
- Δημιουργία ρητών υποθέσεων για κάθε γνωστικό τομέα.
- Διαχωρισμός της γνώσης του κάθε τομέα από τη λειτουργική γνώση.
- Ανάλυση της γνώσης του κάθε τομέα.

Διαχείριση οντολογιών: Συνδυασμός και συσχέτιση

- Επαναχρησιμοποίηση διαφορετικών οντολογιών → συνδυασμός → ενοποίηση οντολογιών (συνένωση σε 1 οντολογία ή χωριστά).
- Διαδικασία ενοποίησης οντολογιών:
 - Εύρεση των σημείων εκείνων στις οντολογίες που έχουν κοινά χαρακτηριστικά.
 - Συσχέτιση των εννοιών εκείνων που είναι σημασιολογικά κοντά, μέσω της ισοδυναμίας και των σχέσεων υπαγωγής.
 - Έλεγχος της συνοχής και της συνάφειας του αποτελέσματος.
- Ευθυγράμμιση εννοιών: δύσκολη διαδικασία → αλλαγή σε μία τουλάχιστον από τις οντολογίες → νέα έκδοση της οντολογίας.
- Αν οι οντολογίες δεν αναπαριστώνται στην ίδια γλώσσα → αναγκαία η μετάφρασή τους.

Συνδυασμός οντολογιών – Προβλήματα (1/4)

Αναντιστοιχίες → 2 επίπεδα

- επίπεδο της γλώσσας ή το επίπεδο μετα-μοντέλου (language or meta-model level) → χρησιμοποιούνται τα πρωταρχικά στοιχεία μιας γλώσσας προκειμένου να οριστεί και να προσδιοριστεί μια οντολογία (αναντιστοιχίες ανάμεσα στους διαφορετικούς μηχανισμούς που ορίζουν τις κλάσεις, τις σχέσεις κλπ.)
- επίπεδο οντολογίας ή το επίπεδο μοντέλου (ontology or model level) → ενυπάρχει η πραγματική οντολογία ενός τομέα (η διαφορά στον τρόπο με τον οποίο μοντελοποιείται το εκάστοτε πεδίο ορισμού).

Συνδυασμός οντολογιών – Προβλήματα (2/4)

- Αναντιστοιχίες στο επίπεδο της γλώσσας
 - Συντακτικό
 - Λογική αναπαράσταση
 - Σημασιολογία πρωταρχικών στοιχείων
 - Εκφραστικότητα γλώσσας

Συνδυασμός οντολογιών – Προβλήματα (3/4)

- Αναντιστοιχίες στο επίπεδο οντολογίας
 - Εμβέλεια
 - Κάλυψη μοντέλου και διαβάθμιση
 - Παράδειγμα
 - Περιγραφή έννοιας

Συνδυασμός οντολογιών – Προβλήματα (4/4)

- Αναντιστοιχίες ορολογίας

→ Αναπαράσταση συνώνυμων και ομώνυμων όρων σε διαφορετικά πεδία

- Κωδικοποίηση πεδίων (π.χ. τρόπος αναγραφής ημερομηνίας)

Υποστήριξη πολλαπλών εκδόσεων (Versioning) (1/3)

- Ύπαρξη πολλαπλών διαφορετικών παραλλαγών μιας οντολογίας σε κάποιο περιβάλλον.

Παραλλαγές : α) τροποποιήσεις σε μια ήδη υπάρχουσα οντολογία.

β) διαφορετικές εκδόσεις οντολογιών του ίδιου τομέα ανεξάρτητα ανεπτυγμένες.

- Ορισμός: η δυνατότητα να χειριζόμαστε αλλαγές στις οντολογίες με τη δημιουργία και διαχείριση διαφορετικών παραλλαγών αυτών

- → δημιουργία μιας μεθοδολογίας που θα διακρίνει και θα αναγνωρίζει εκδόσεις, διαδικασίες για ενημερώσεις και αλλαγές στις οντολογίες, ώστε να παρακολουθούνται οι σχέσεις ανάμεσα στις διάφορες εκδόσεις.

Υποστήριξη πολλαπλών εκδόσεων (Versioning) (2/3)

Αιτίες αλλαγών στις οντολογίες:

- αλλαγές στον τομέα
- αλλαγές στη διαμοιρασμένη εννοιολογική αναπαράσταση
- αλλαγές στις προδιαγραφές

Υποστήριξη πολλαπλών εκδόσεων (Versioning) (3/3)

Επιπτώσεις αλλαγών

Ασυμβατότητες:

- A) ως προς τα **δεδομένα** (π.χ. ιστοσελίδες). Όταν μια οντολογία μεταβάλλεται, τα δεδομένα αυτά μπορεί να λάβουν διαφορετική ερμηνεία ή να χρησιμοποιηθούν άγνωστοι όροι.
- B) ως προς **άλλες οντολογίες**, οι οποίες χρησιμοποιούν την τροποποιημένη οντολογία. Αλλαγές στην αρχική οντολογία θα μπορούσαν να επηρεάσουν και την τελική.
- Γ) ως προς ορισμένες **εφαρμογές** που χρησιμοποιούν μια οντολογία → δυσλειτουργία.

“Versioning” # “Evolution”

- “evolution” → δυνατότητα εκτέλεσης αλλαγών, χωρίς να βλάπτεται η ακεραιότητα της οντολογίας.
- “versioning” → δυνατότητα που έχει ένα σύστημα διαχείρισης οντολογιών να διαχειρίζεται τις διάφορες αλλαγές μέσα από τη δημιουργία διαφορετικών «*διεπαφών εκδόσεων*».

Πρακτικά

“evolution” → αλλαγές που αφορούν στην παρούσα στιγμή
→ δεν είναι δυνατόν να ανακαλέσουμε την οντολογία όπως ήταν προτού εφαρμόσουμε την ή τις αλλαγές
→ κανένα ιστορικό αλλαγών

“versioning” → διατηρεί εκδόσεις με το πώς ήταν η οντολογία πριν την ή τις αλλαγές (προσπελάσιμες εκδόσεις)
→ πλήρες ιστορικό των αλλαγών

Εξέλιξη οντολογιών (Ontology evolution)

Ορισμός → έγκαιρη προσαρμογή στις μεταβαλλόμενες εργασιακές απαιτήσεις, στις τάσεις στα στιγμιότυπα, στον τρόπο χρήσης εφαρμογών που βασίζονται σε οντολογίες, καθώς και τη συνεχή διαχείριση των αλλαγών αυτών.

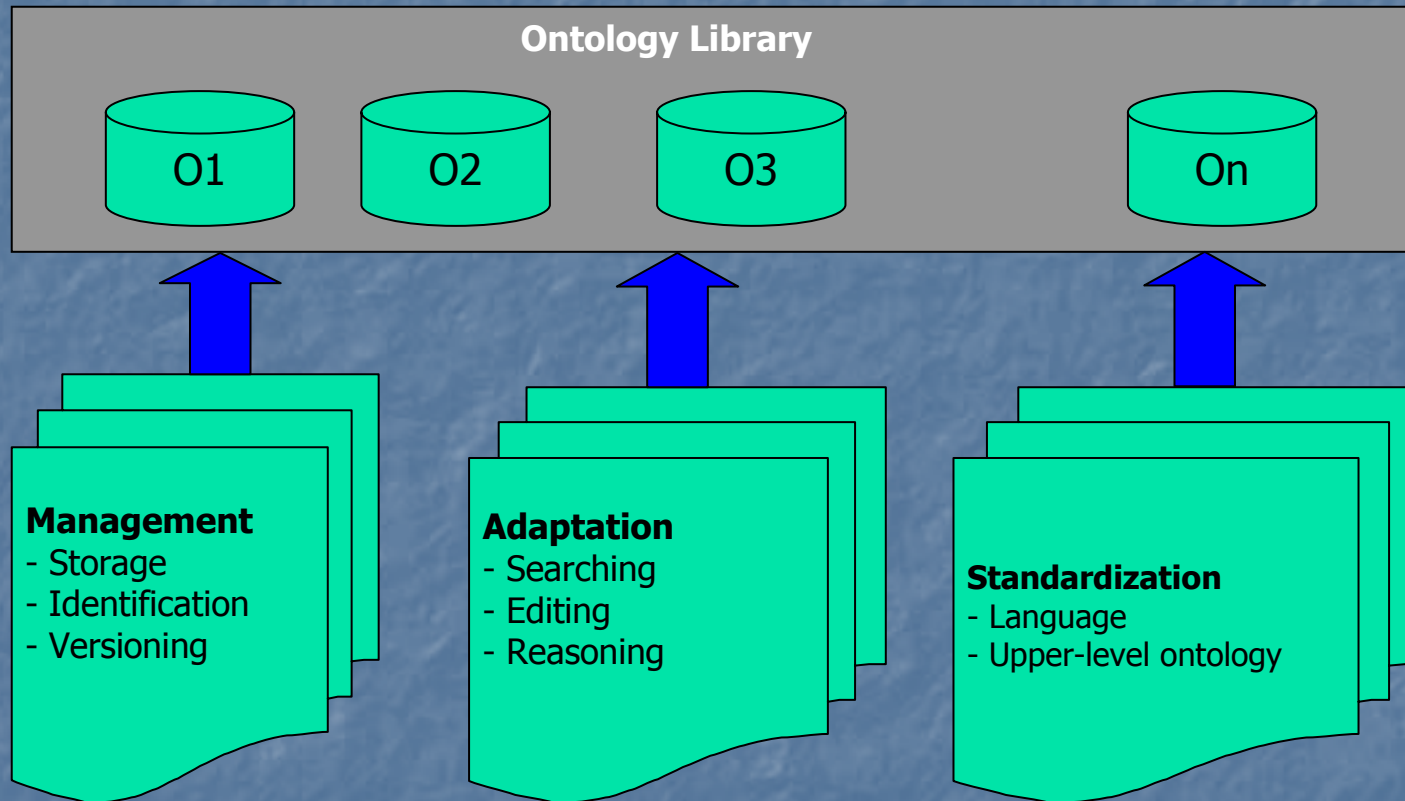
- Σκοπός (σε κατανεμημένα περιβάλλοντα) → η αύξηση της γνώσης σε κάθε κόμβο ενός ανοιχτού κατανεμημένου συστήματος, με την απόκτηση περιγραφών πηγών από τις οντολογίες των άλλων κόμβων.
- Περιλαμβάνει 6 στάδια:

(1) εντοπισμός αλλαγών	(2) αναπαράσταση αλλαγών
(3) σημασιολογία αλλαγών	(4) διάδοση αλλαγών
(5) υλοποίηση των αλλαγών	(6) επικύρωση αλλαγών

Ontoview

- Εργαλείο υποστήριξης, για την εξέλιξη οντολογιών που είναι διαθέσιμες on-line στο διαδίκτυο.
- Κύριες λειτουργίες
 - διαχείριση των εκάστοτε αλλαγών που παρουσιάζονται στις οντολογίες
 - επίτευξη διαλειτουργικότητας μεταξύ των διαφόρων εκδόσεων μιας οντολογίας (συγκρίνει τις διάφορες εκδόσεις και τονίζει τις διαφορές μεταξύ τους)
 - προσδιορισμός της εννοιολογικής σχέσης ανάμεσα στις διαφορετικές εκδόσεις των εννοιών
- Αποθηκευτικό μέσο, παρέχοντας μια διαφανή διεπαφή για την επικοινωνία μεταξύ τυχαίων εκδόσεων των οντολογιών.

Βιβλιοθήκες οντολογιών



WebOnto (1/2)

- Αναπτύχθηκε από το Knowledge Media Institute του Open University (UK).
- Σχεδιάστηκε για να υποστηρίξει τη συνεργατική δημιουργία, ανάγνωση και επεξεργασία των οντολογιών.
- Παρέχει:
 - α) μια διεπαφή άμεσου χειρισμού, η οποία εκθέτει οντολογικές εκφράσεις
 - β) ένα εργαλείο "συνομιλίας" με οντολογίες (ontology discussion tool), το Tadzebao → υποστηρίζει ταυτόχρονα και σύγχρονες και ασύγχρονες "συνομιλίες" με οντολογίες.

WebOnto (2/2)

- Αρχιτεκτονική πελάτη-εξυπηρετητή
- Γραφικό περιβάλλον
- Έλλειψη συστηματικής ταξινόμησης, προσδιορισμός οντολογιών σύμφωνα με το όνομά τους.
- Δεν υποστηρίζει πολλαπλές εκδόσεις.
- Γλώσσα αναπαράστασης: OCML
- Μη διαθέσιμη πρότυπη οντολογία ανώτερου επιπέδου (μία πιο εκλεπτυσμένη δομή).

OntoLingua (1/4)

- Αναπτύχθηκε στο Knowledge Systems Laboratory του Stanford University (αρχές '90).
- Περιλαμβάνει:
 - Ένα εξυπηρετητή (αποθετήριο οντολογιών → δυνατότητα δημιουργίας, επεξεργασίας, τροποποίησης και βελτίωσης)
 - Μια γλώσσα αναπαράστασης
- Δυνατή η μετατροπή κάθε οντολογίας σε διαφορετικές διατάξεις

OntoLingua (2/4)

New Unnamed Ontology

Ontology name:

Vehicles-Tutorial

Ontology documentation (optional):

```
<UL><LI>This ontology will be a general ontology of vehicles  
which are typically bought and sold through the classified  
ads. This will include motorized as well as unmotorized  
vehicles.  
<LI>This ontology will need to be able to describe any
```

Ontologies included by this new ontology:

Physical-Quantities
Physicist-Query-Ontology
Product-Ontology
Pwrist2-Temp-Hpkb-Upper-Level-Kernel-Latest
Quantity-Spaces

Assert New Ontology

Cancel

Reset

Help

OntoLingua (3/4)

- Αρχιτεκτονική πελάτη-εξυπηρετητή
- Έλλειψη ταξινόμησης σύμφωνα με κάποιες πρότυπες κατηγορίες
- Θεωρία του πλέγματος (μία οντολογία συμπεριλαμβάνει τις οντολογίες εκείνες, κάτω από τις οποίες βρίσκεται αυτή τοποθετημένη με εσοχή)

Kif-Sets

Kif-Extensions

Frame-Ontology

Jat-Generic

Job-Assignment-Task

Basic-Matrix-Algebra

Tensor-Quantities

3d-Tensor-Quantities

Simple-Geometry

Mechanical-Components

Mace-Domain

Abstract-Algebra

Physical-Quantities

Standard-Dimensions

Vt-Design

Vt-Domain

Vt-Example

OntoLingua (4/4)

- Ταυτοποίηση οντολογίας από το όνομα
- Έλλειψη υποστήριξης πολλαπλών εκδόσεων
- Γραφική ανάγνωση οντολογιών (γρήγορες μεταβάσεις από τον όρο μιας οντολογίας σε έναν άλλο)
- Γλώσσα αναπαράστασης: KIF
- Οντολογία ανώτερου επιπέδου CYC, (HPKB-UPPER-LEVEL) = περίπου, 3000 έννοιες, αγγλικοί ορισμοί και κάποιες βασικές σχέσεις μεταξύ τους

DAML (1/2)

- Μέρος του DARPA Agent Markup Language (DAML) Program (Αύγουστος 2000)
- Στόχος → ανάπτυξη γλώσσας και εργαλείων για τη διευκόλυνση του Σημασιολογικού Ιστού
- Κατάλογος οντολογιών σε XML, HTML και DAML
- Δημόσια βιβλιοθήκη (υποβολή νέων οντολογιών)
- Αρχιτεκτονική πελάτη-εξυπηρετητή
- Ταξινόμηση σύμφωνα με το Open Directory Category
- Ταυτοποίηση: URI
- Μη διαθέσιμες λειτουργίες επεξεργασίας, συλλογιστικής υποστήριξης και υποστήριξης πολλαπλών εκδόσεων

DAML (2/2)

```
<ontology uri="http://www.davincinetbook.com:8080/daml/rdf/personal-info.daml" id="2">
  <description>DAML ontology for homework 1</description>
  <poc name="Mark Neighbors" organization="Booz-Allen & Hamilton" email="neighbors_mark@bah.com"/>
  <submitter name="Mark Neighbors" organization="Booz-Allen & Hamilton" email="neighbors_mark@bah.com"
    date="2000-10-31"/>
  <keyword>personal information</keyword>
  <dmoz>http://dmoz.org/dmoz5</dmoz>
  <dmoz>http://dmoz.org/dmoz6</dmoz>
  <funder>DARPA DAML Program</funder>
  <class>AnnotatedBulletList</class>
  <class>Bullet</class>
  <class>BulletList</class>
  <class>Company</class> .....
  .....
  <property>bulletList</property>
  <property>companies</property>
  <property>currentEmployerIDs</property>
  <property>currentProjectIDs</property>
  <property>description</property>
  .....
  <namespace>http://156.80.108.115/2000/10/daml-ont.daml</namespace>
  <namespace>http://www.w3.org/1999/02/22-rdf-syntax-ns</namespace>
  <namespace>http://www.w3.org/2000/01/rdf-schema</namespace>
</ontology>
```

Βιβλιοθήκη οντολογιών Shoe

- Λίστες οντολογιών
- Οντολογίες → αλφαβητική ευρετηρίαση και ταξινόμηση ως προς την εξάρτηση της οντολογίας (αυστηρή δομή δέντρου)
- Δεν υπάρχει περιβάλλον αναζήτησης και επεξεργασίας
- Υποστήριξη πολλαπλών εκδόσεων και συλλογισμού (το πρώτο σύστημα που εστίασε στο πρόβλημα της εξέλιξης και συνέχειας των οντολογιών)
- Γλώσσα αναπαράστασης: SHOΞ

Vub Ontology Server

- Αναπτύχθηκε από το Vrije Universiteit Brussels.
- Τεχνολογία οντολογιών + σημασιολογία βάσεων δεδομένων.
- Το σύστημα διαχείρισης βάσης δεδομένων (DBMS), εξοπλισμένο με ποικίλες συντακτικές δομές, δίνει τη δυνατότητα στα σχήματα βάσεων να παρουσιάζουν αντικείμενα, ταξινομίες, περιορισμούς ακεραιότητας και κανόνες παραγωγής.
- Έλλειψη υποστήριξης πολλαπλών εκδόσεων και συλλογισμού
- Γλώσσα αναπαράστασης: XML
- Μη υιοθέτηση πρότυπης οντολογίας.

Ερευνητικές κατευθύνσεις (1/2)

Δημιουργία ιδανικής βιβλιοθήκης οντολογιών:

- Αρχιτεκτονική πελάτη-εξυπηρετητή (συνεργατική επεξεργασία οντολογιών)
- Ταξινόμηση οντολογιών (διευκόλυνση της αναζήτησης, διαχείρισης και επαναχρησιμοποίησης)
- Προτυποποίηση σύμφωνα με το URL και το όνομα της οντολογίας
- Υποστήριξη πολλαπλών εκδόσεων

Ερευνητικές κατευθύνσεις (2/2)

- Σύγχρονο περιβάλλον φυλλομέτρησης
- Υποστήριξη συλλογισμού
- Χρήση πρότυπης γλώσσας
- Χρήση πρότυπης οντολογίας ανώτερου επιπέδου
- Λειτουργίες συντήρησης
- Σαφής τεκμηρίωση (σχετικά με τον τρόπο ανάπτυξης οντολογίας, την πολιτική ονοματοδοσίας κλπ.)

Παρατηρήσεις - Συμπεράσματα

- Ο τομέας διαχείρισης της γνώσης (συνεπώς και ο τομέας διαχείρισης οντολογιών) αναπτύσσεται με ταχύτατους ρυθμούς → εξελίξεις ραγδαίες → ανάπτυξη νέων μεθοδολογιών και τεχνικών.
- Οι οντολογίες, με τα ιδιαίτερα χαρακτηριστικά τους, συνεισφέρουν καθοριστικά στο όραμα της σωστής διαχείρισης της γνώσης, και της ολοκλήρωσης της δημιουργίας του Σημασιολογικού Ιστού.
- Διαρκώς εμφανίζονται περισσότερες ερευνητικές κατευθύνσεις, επιπρόσθετα χαρακτηριστικά, νέες ιδιότητες.
- Παρατηρούνται προβλήματα στη διαχείριση των οντολογιών, γίνονται προσπάθειες εξάλειψής τους.
- Ο χώρος των οντολογιών έχει σημαντικά ερευνητικά περιθώρια, με ιδιαίτερο ενδιαφέρον και ευοίωνο μέλλον.

Σας ευχαριστούμε για την προσοχή σας !!!