

Εισαγωγή στις Οντολογίες και το Σημασιολογικό Ιστό

Μανόλης Γεργατσούλης

Χρήστος Παπαθεοδώρου

Ομάδα Βάσεων Δεδομένων και Πληροφοριακών
Συστημάτων, Τμήμα Αρχιονομίας – Βιβλιοθηκονομίας
Ιόνιο Πανεπιστήμιο



Ο Παγκόσμιος Ιστός

- Ο παγκόσμιος ιστός είναι μια μεγάλη και επιτυχημένη προσπάθεια αν αναλογιστεί κανείς
 - την ποσότητα της διαθέσιμης πληροφορίας, και
 - το ρυθμό αύξησης των ανθρώπων που τον χρησιμοποιούν.
- Έχει αρχίσει να διαπερνά τις περισσότερες πλευρές της καθημερινής μας ζωής αλλά και της ζωής των επιχειρήσεων και οργανισμών.
- Η επιτυχία του βασίζεται στην απλότητα του.
- Η απλότητα του HTTP και της HTML που στοχεύουν στη μεταφορά και προβολή υπερκειμένων, επέτρεψε στους κατασκευαστές λογισμικού, στους φορείς παροχής πληροφοριών και τους τελικούς χρήστες, την εύκολη προσπέλαση στο νέο μέσο και συνέβαλε στο να αναπτύξουν σε πολύ μικρό χρονικό διάστημα μια σημαντική κρίσιμη μάζα.



Ο Παγκόσμιος Ιστός (συνέχεια)

- Αυτή η απλότητα όμως ενδέχεται να αποβεί εμπόδιο στην παραπέρα ανάπτυξη του παγκόσμιου ιστού.
- Αυτό το οποίο έχουμε σήμερα αποτελεί την πολύ αρχική έκδοση του παγκόσμιου ιστού. Η επόμενη έκδοση θα είναι ακόμα μεγαλύτερη και πολύ πιο ισχυρότερη από τη σημερινή.

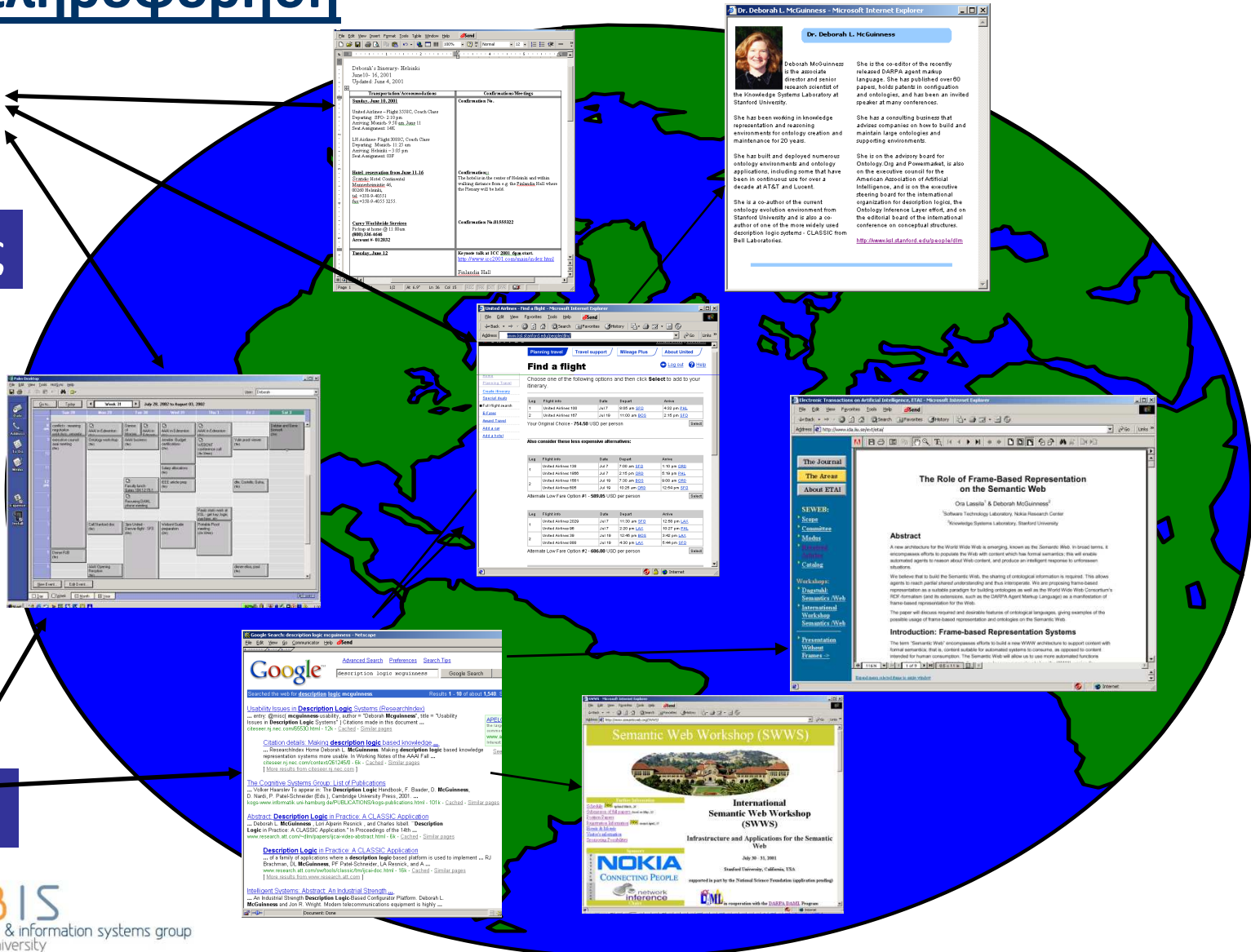
Παγκόσμιος Ιστός: Πλούσια Πηγή Πληροφοριών για Διαχείριση και Ερμηνεία από τον Άνθρωπο - Υπερπληροφόρηση



Χρήστης



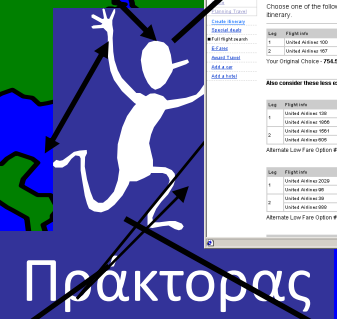
Χρήστης



Η τεχνολογία του Σημασιολογικού Παγκόσμιου Ιστού

- Το όραμα του Tim Berners-Lee για τον Σημασιολογικό Παγκόσμιο Ιστό προβλέπει:
 - σημασιολογία της πληροφορίας η οποία είναι εύκολο να γίνει κατανοητή μηχανικά, και
 - εκατομμύρια από μικρές εξειδικευμένες υπηρεσίες συλλογισμού (reasoning services) οι οποίες παρέχουν υποστήριξη στην επιτυχή επίτευξη εργασιών βασιζόμενες στην προσπελάσιμη πληροφορία.

Ο Μελλοντικός Παγκόσμιος Ιστός: Πλούσια Πηγή Πληροφοριών για Διαχείριση και Ερμηνεία από τον Η/Υ

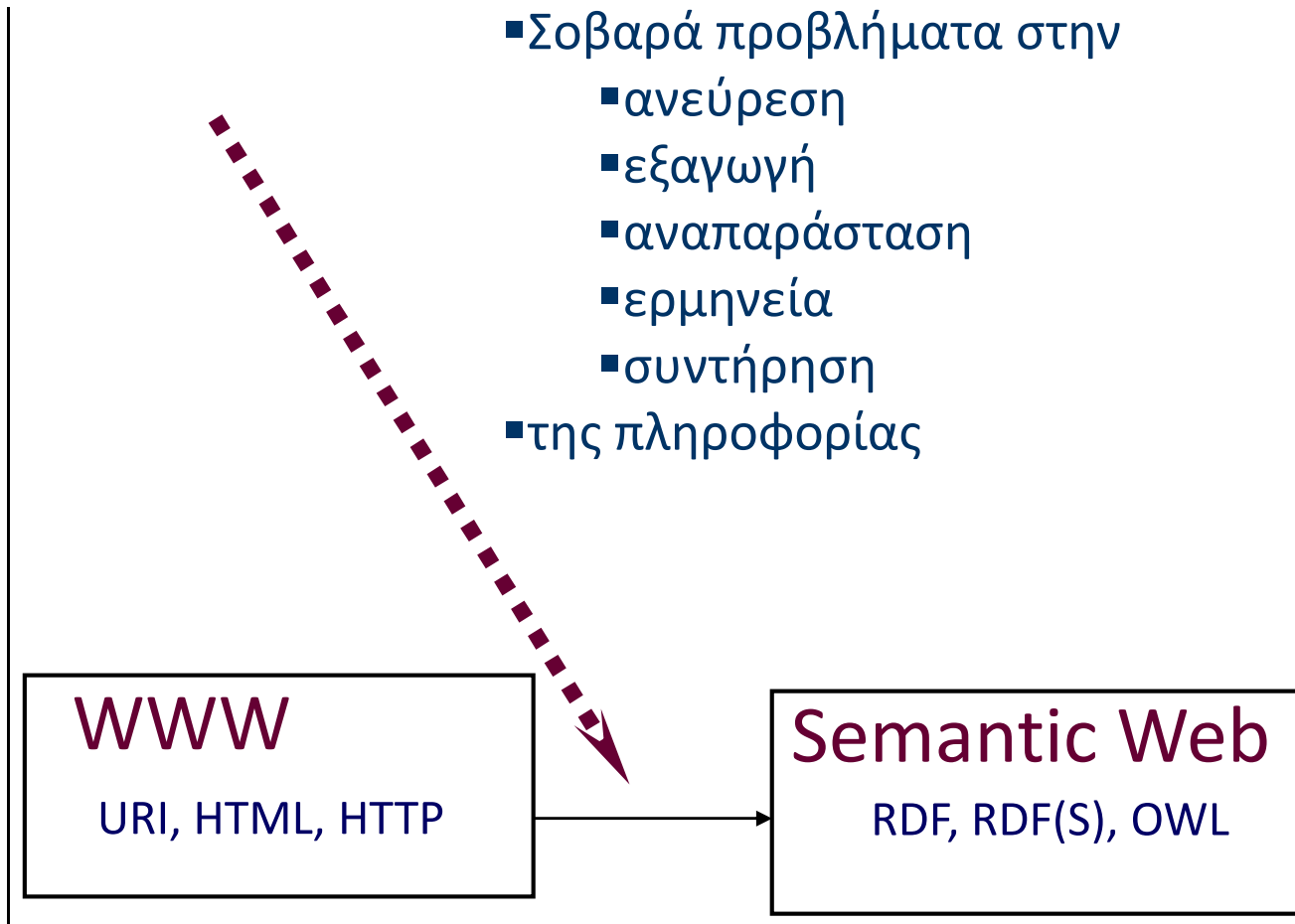


The image shows a globe with several overlapping screenshots of web pages and software interfaces. Arrows point from the icons to the relevant screenshots:

- Χρήστης (User):** Points to a screenshot of a web page titled "Dr. Deborah L. McGuinness" with a profile picture and biographical information.
- Πράκτορας (Actor):** Points to a screenshot of a flight search engine showing flight options from London to Athens.
- Πράκτορας (Actor):** Points to a screenshot of a journal article titled "The Role of Frame-Based Representation on the Semantic Web" by Ora Lassila & Deborah McGuinness.
- Πράκτορας (Actor):** Points to a screenshot of a Semantic Web Workshop (SWWS) advertisement for the International Semantic Web Workshop (SWWS) in July 2003.
- Πράκτορας (Actor):** Points to a screenshot of a Google search results page for "description logic McGuinness".
- Πράκτορας (Actor):** Points to a screenshot of a software interface showing a complex diagram or flowchart.
- Πράκτορας (Actor):** Points to a screenshot of a web page titled "Description Logic Systems (Message 00000)".
- Πράκτορας (Actor):** Points to a screenshot of a web page titled "The Cooperative Systems Group: List of Publications".
- Πράκτορας (Actor):** Points to a screenshot of a web page titled "Description Logic in Practice: A CLASSIC Application".
- Πράκτορας (Actor):** Points to a screenshot of a web page titled "Description Logic in Practice: A CLASSIC Application".
- Πράκτορας (Actor):** Points to a screenshot of a web page titled "Intelligent Systems: An Industrial Showcase".

Σημασιολογικός Παγκόσμιος Ιστός (Semantic Web)

- Σοβαρά προβλήματα στην
 - ανεύρεση
 - εξαγωγή
 - αναπαράσταση
 - ερμηνεία
 - συντήρηση
- της πληροφορίας



Η υλοποίηση του Σημασιολογικού Ιστού

- Η υλοποίηση του Σημασιολογικού Παγκόσμιου Ιστού είναι ένα εξαιρετικά φιλόδοξο εγχείρημα.
- Ακόμη και η 'μερική' υλοποίηση του θα είναι σημαντικό επίτευγμα.
- Απαιτείται η συμβολή πολλών επιστημόνων από διαφορετικά επιστημονικά πεδία, όπως:
 - Τεχνολογίες βάσεων δεδομένων
 - Ευφυείς διαμεσολαβητές (agents)
 - Ψηφιακές βιβλιοθήκες
 - Μεταδεδομένα
 - Αναπαράσταση γνώσης και εξαγωγή συμπερασμάτων
 - Φυσικές γλώσσες
 - Οντολογίες
 - Πολυμέσα
 - Τεχνικές αναζήτησης και ερωταπαντήσεων
 - Οπτικοποίηση και μοντελοποίηση
 -

Οι μηχανές αναζήτησης στο Σημασιολογικό Παγκόσμιο Ιστό

- Τρέχουσα τεχνολογία μηχανών αναζήτησης = λέξεις κλειδιά:
 - υψηλός βαθμός επανάληψης ανάκτησης, μικρή ακρίβεια
 - ευαίσθητες στο λεξιλόγιο
 - δε λαμβάνουν υπ' όψιν το υπονοούμενο περιεχόμενο
- Μηχανές αναζήτησης στο Σημασιολογικό Παγκόσμιο Ιστό:
 - αναζήτηση με βάση έννοιες αντί λέξεις κλειδιά
 - σημασιολογική εστίαση/διεύρυνση των ερωτήσεων
 - ερωταπαντήσεις ως προς περισσότερα του ενός κείμενα
 - χρήση τελεστών μετασχηματισμού των κειμένων

Από τα μεταδεδομένα στη γνώση

- Ανάγκη: οργάνωση και ανάκτηση της πληροφορίας με βάση τη σημασία της, τις έννοιες και όχι τις λέξεις
- Συμβατικές βιβλιοθήκες: Ταξινομικά συστήματα, θησαυροί
- Ψηφιακή Τεχνολογία: Οντολογίες
 - Ορίζουν τις έννοιες που υπάρχουν στα μεταδεδομένα και τις μεταξύ τους σχέσεις → **ΠΑΡΑΓΩΓΗ ΓΝΩΣΗΣ!!**
 - Περιγράφουν ένα πεδίο γνώσης και διατυπώνουν μια κοινά αποδεκτή ορολογία → **ΠΑΡΑΓΩΓΗ ΓΝΩΣΗΣ!!**
 - **Αναπαριστούν** αυτή τη γνώση στον υπολογιστή
 - Επιτρέπουν τη σύνδεση του περιεχομένου (το οποίο είναι μηχαναγνώσιμο και επεξεργάζεται μηχανικά) με τις έννοιες
 - Επιτρέπουν την ανάπτυξη **μηχανισμών εξαγωγής συμπερασμάτων**

Ο ρόλος των Οντολογιών

- Οι Οντολογίες συνενώνουν δύο ουσιώδη συστατικά τα οποία συμβάλουν στην πλήρη ανάπτυξη του Παγκόσμιου Ιστού:
 - ορίζουν την τυπική σημασιολογία της πληροφορίας διευκολύνοντας την επεξεργασία της πληροφορίας από τον Η/Υ
 - ορίζουν σημασιολογία του πραγματικού κόσμου επιτρέποντας τη σύνδεση του περιεχομένου το οποίο επεξεργάζεται μηχανικά, με τη σημασία που του δίνουν οι άνθρωποι βασιζόμενοι σε κοινά αποδεκτή ορολογία.

Οντολογία (Ontology)

- Μακρά ιστορία που προέρχεται από την Φιλοσοφία - Αριστοτέλης
 - “Η μεταφυσική μελέτη της φύσης της ζωής και της ύπαρξης”
- Η έννοια της οντολογίας έχει υιοθετηθεί από την Τεχνητή Νοημοσύνη και σημαίνει
 - **“μια διαμοιρασμένη και κοινή κατανόηση κάποιου τομέα, η οποία μπορεί να ανταλλαγεί μεταξύ ανθρώπων και συστημάτων εφαρμογών” Gruber**

Το πρόβλημα

- Άνθρωποι, οργανισμοί και προγράμματα πρέπει να επικοινωνούν μεταξύ τους.
- Οι διαφορετικές ανάγκες και το διαφορετικό υπόβαθρο οδηγούν σε αποκλίνουσες οπτικές γωνίες και παραδοχές για πράγματα που στην ουσία είναι ίδια μεταξύ τους.
- Η έλλειψη κοινής αντίληψης οδηγεί σε:
 - προβλήματα στην επικοινωνία μεταξύ ανθρώπων και οργανισμών.
 - δυσκολίες στον προσδιορισμό των απαιτήσεων και κατά συνέπεια στην ανάπτυξη των προδιαγραφών των συστημάτων.
- Οι ανομοιόμορφες μέθοδοι μοντελοποίησης, οι γλώσσες και τα εργαλεία λογισμικού περιορίζουν σοβαρά:
 - τη δια-λειτουργικότητα
 - την επαναχρησιμοποίηση και το διαμοιρασμό εφαρμογών.
- Τελικά όλα τα παραπάνω οδηγούν στο να ξαναανακαλύπτουμε τον τροχό.

Η λύση

- Η εξάλειψη ή η μείωση της σύγχυσης σχετικά με τις έννοιες και τους όρους και τελικά η απόκτηση κοινής αντίληψης.
- Αυτή η κοινή αντίληψη μπορεί να αποτελέσει το ενοποιητικό πλαίσιο ανάμεσα στις διαφορετικές οπτικές γωνίες και να συμβάλει στην βελτίωση της επικοινωνίας, και της δια-λειτουργικότητας. Μπορεί να διευκολύνει την ανάπτυξη προδιαγραφών και τη βελτίωση της αξιοπιστίας συστημάτων και να παρέχει δυνατότητες επαναχρησιμοποίησης τους.

Τι είναι η Οντολογία;

- Μια **οντολογία** είναι μια τυπική (*formal*), κατηγορηματική (*explicit*) προδιαγραφή μιας διαμοιρασμένης (*shared*) εννοιολογικής αναπαράστασης (*conceptualization*) - **Gruber**
 - Ο όρος ‘εννοιολογική αναπαράσταση’ (*conceptualization*) αναφέρεται σε ένα αφηρημένο μοντέλο φαινομένων του κόσμου στο οποίο έχουν προσδιοριστεί οι έννοιες που σχετίζονται με τα φαινόμενα αυτά.
 - Ο όρος ‘κατηγορηματική’ (*explicit*) σημαίνει ότι το είδος των εννοιών που χρησιμοποιούνται, και οι περιορισμοί που αφορούν την χρήση αυτών των εννοιών είναι προσδιορισμένα με σαφήνεια.
 - Ο όρος ‘αυστηρή’ (*formal*) αναφέρεται στο ότι η οντολογία πρέπει να είναι μηχανικά αναγνώσιμη.
 - Ο όρος ‘διαμοιρασμένη’ (*shared*) αναφέρεται στο ότι η οντολογία πρέπει να αποτυπώνει γνώση κοινής αποδοχής στα πλαίσια μιας κοινότητας.

Η μορφή μιας οντολογίας

- Μια οντολογία μπορεί να πάρει διάφορες μορφές αλλά οπωσδήποτε θα περιλαμβάνει ένα λεξιλόγιο όρων και κάποιας μορφής προδιαγραφές για τη σημασία τους.
- Σχετικά με το βαθμό της τυπικότητας της αναπαράστασης μιας οντολογίας αυτή μπορεί να είναι:
 - **Άτυπη** (informal), εκφρασμένη σε μια φυσική γλώσσα.
 - **Ημι-άτυπη** (semi-informal): για παράδειγμα 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).

Τα βασικά συστατικά μιας Οντολογίας

- Πέντε κατηγορίες συστατικών:
 - Κλάσεις (classes):
 - Έννοιες που σχετίζονται με ένα πεδίο ή κάποιες εργασίες, οι οποίες είναι συνήθως οργανωμένες σε κάποιο ταξινομικό σύστημα.
 - Σε μια οντολογία που αφορά το πανεπιστήμιο: ο 'φοιτητής' και ο 'καθηγητής' αποτελούν δύο κλάσεις.
 - Σχέσεις (relations):
 - Ένας τύπος αλληλεπίδρασης μεταξύ εννοιών ενός πεδίου.
 - όπως: *subclass-of*, *is-a*

Τα βασικά συστατικά μιας Οντολογίας

(συνέχεια)

- **Συναρτήσεις (functions)**
 - Μια ειδική περίπτωση σχέσης στην οποία το n -οστό στοιχείο της σχέσης προσδιορίζεται μοναδικά από τα $n-1$ προηγούμενα στοιχεία.
 - Παράδειγμα: Η τιμή-μεταχειρισμένου-αυτοκινήτου μπορεί να προσδιορίζεται σαν συνάρτηση της αρχικής τιμής του καινούριου αυτοκινήτου, του μοντέλου του αυτοκινήτου, των χαρακτηριστικών του αυτοκινήτου και των χιλιομέτρων που έχει διανύσει.
- **Αξιώματα (axioms)**
 - Αναπαριστούν προτάσεις που είναι πάντα αληθείς
 - Παράδειγμα: αν ο Φ είναι δευτεροετής φοιτητής τότε μπορεί να εγγραφεί στο επιλεγόμενο μάθημα M .
- **Στιγμιότυπα (instances)**
 - Αναπαριστούν συγκεκριμένα στοιχεία
 - Παράδειγμα: ο φοιτητής με το όνομα 'Νίκος' είναι ένα στιγμιότυπο της κλάσης 'φοιτητής'

Παράδειγμα Οντολογίας (σε OIL)

```
class-def animal
%plants are a class that is disjoint from animals
class-def plant subclass-of NOT animal
%it is necessary but not sufficient for a tree
% to be a plant:
class-def tree subclass-of plant
%branches are PART OF trees
class-def branch
  slot-constraint is-part-of has-value tree
%it is necessary and sufficient for a carnivore
% to be an animal:
class-def defined carnivore subclass-of animal
  slot-constraints eats value-type animal
%herbivores eat only plants OR part of plants
class-def defined herbivore subclass-of animal
  slot-constraint eats value-type plant OR
  (slot-constraint is-part-of has-value plant)
```

Κατηγορίες Οντολογιών

- Μερικές χαρακτηριστικές κατηγορίες οντολογιών είναι οι ακόλουθες:
 - **Οντολογίες πεδίου ορισμού** (domain ontologies): αναπαριστούν γνώση γύρω από ένα συγκεκριμένο πεδίο (π.χ. ιατρική, ηλεκτρονικά κ.λ.π.).
 - **Οντολογίες μεταδεδομένων** (metadata ontologies): παρέχουν ένα λεξιλόγιο για την περιγραφή του περιεχομένου ηλεκτρονικά διαθέσιμης πληροφορίας.
 - **Γενικές ή κοινές οντολογίες** (generic or common sense ontologies): στοχεύουν στο να αποτυπώσουν γενική γνώση γύρω από τον κόσμο, παρέχοντας βασικές έννοιες όπως ο χρόνος, ο χώρος, τα συμβάντα, κ.λ.π.
 - **Οντολογίες αναπαράστασης** (representational ontologies): παρέχουν οντότητες αναπαράστασης χωρίς να προσδιορίζουν τι συγκεκριμένο αναπαριστούν
 - π.χ. Frame Ontology (Gruber 1993): ορίζει έννοιες όπως frames, slots, slot constraints κ.λ.π.
 - **Οντολογίες μεθοδολογίας ή εργασιών** (method or task ontologies): παρέχουν όρους που αναφέρονται σε συγκεκριμένες εργασίες (π.χ. διάγνωση κ.λ.π.)

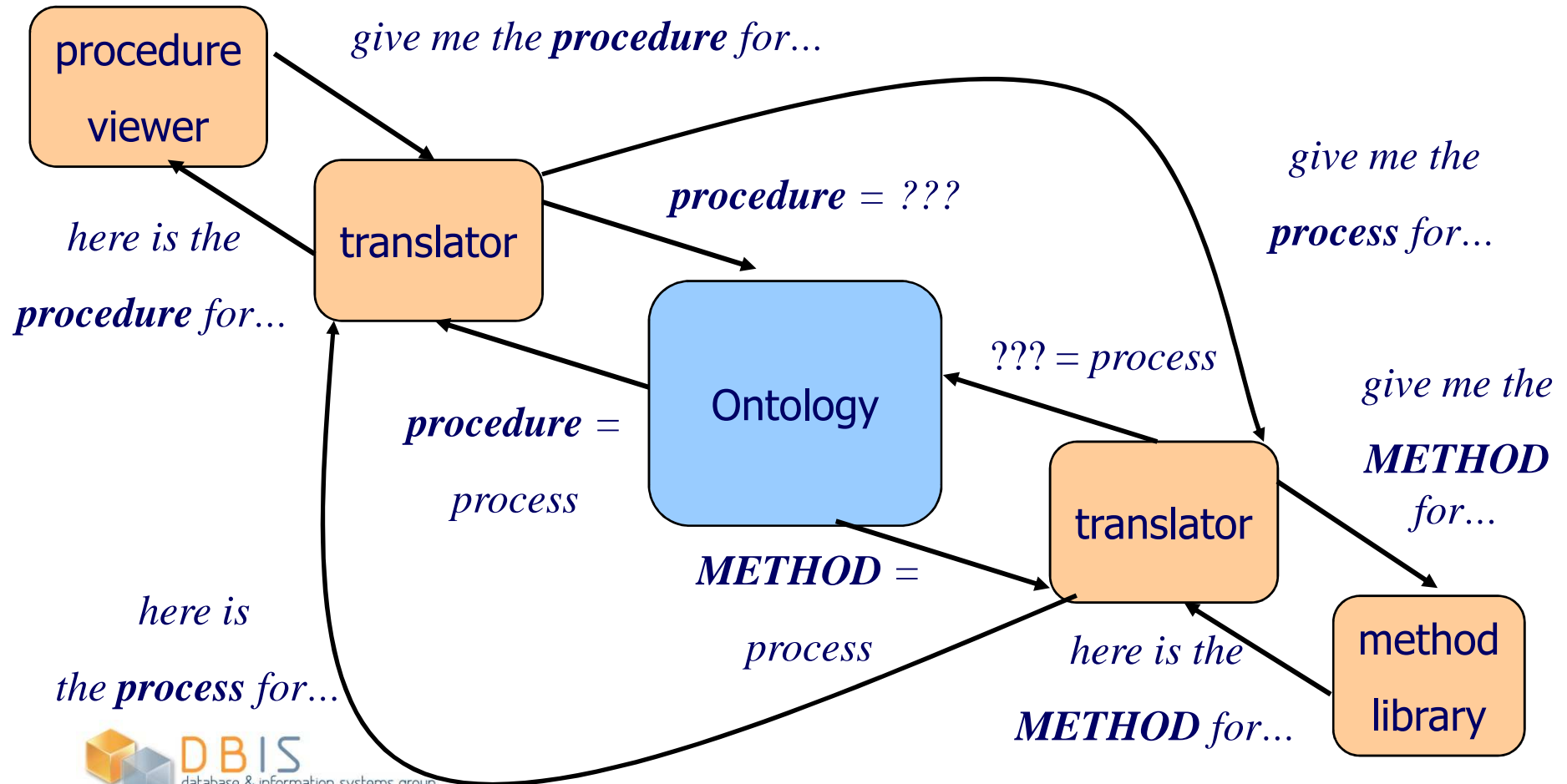
Μερικές Εφαρμογές των Οντολογιών

- Επικοινωνία μεταξύ ανθρώπων και οργανισμών.
 - Παρέχουν ολοκληρωμένο πλαίσιο εννοιών και ορολογίας μεταξύ ανθρώπων με διαφορετικές ανάγκες και οπτικές γωνίες στα πλαίσια ενός οργανισμού. Διευκολύνουν την επικοινωνία των ανθρώπων στα πλαίσια του οργανισμού.
- Δια-λειτουργικότητα (inter-operability) μεταξύ συστημάτων
 - Διάφοροι χρήστες χρειάζεται να ανταλλάσσουν δεδομένα ή χρησιμοποιούν διαφορετικά πακέτα λογισμικού.
 - Χρήση οντολογιών για την υποστήριξη μετάφρασης μεταξύ διαφορετικών γλωσσών και αναπαραστάσεων.
- Μηχανική Συστημάτων (systems engineering)
 - Προδιαγραφές.
 - Επαναχρησιμοποίηση τμημάτων.
 - Αξιοπιστία.

Παράδειγμα:

Δια- λειτουργικότητα

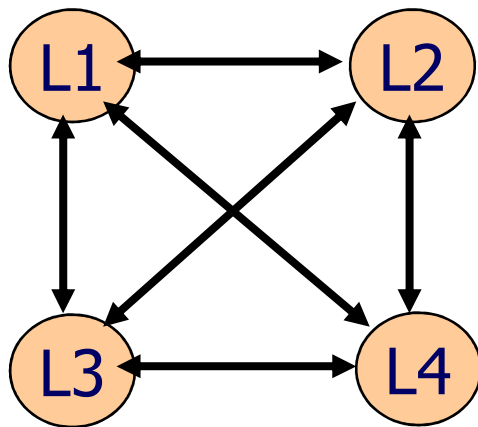
Ο όρος 'procedure' που χρησιμοποιείται από ένα εργαλείο μεταφράζεται στον όρο 'method' που χρησιμοποιείται από το άλλο με τη βοήθεια της οντολογίας, η οποία με τη σειρά της χρησιμοποιεί τον όρο 'process' για να εκφράσει την ίδια έννοια.



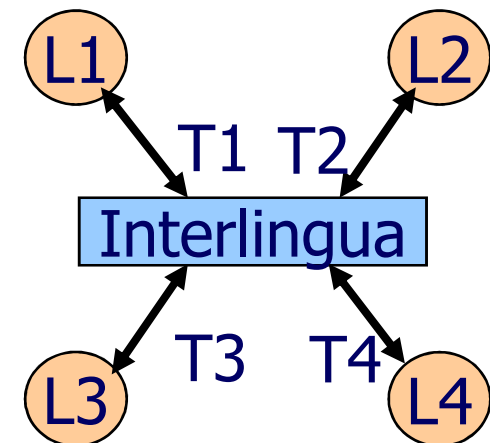
Οι Οντολογίες σαν Inter-Lingua

- Για την υποβοήθηση της δια-λειτουργικότητας οι οντολογίες μπορούν να χρησιμοποιηθούν για να υποστηρίξουν τη μετάφραση μεταξύ διαφορετικών γλωσσών και αναπαραστάσεων.

Μια προσέγγιση είναι να σχεδιάσουμε μοναδικούς μεταφραστές για κάθε ζευγάρι γλωσσών. Αυτό όμως θα απαιτούσε $O(n^2)$ μεταφραστές για n διαφορετικές οντολογίες.



Χρησιμοποιώντας οντολογίες σαν inter-lingua για την υποστήριξη της μετάφρασης, μπορούμε να μειώσουμε τον αριθμό των μεταφραστών σε $O(n)$ για n διαφορετικές οντολογίες, αφού θα απαιτούνται μόνο μεταφραστές από την μητρική οντολογία στην οντολογία ανταλλαγής



Μεθοδολογία Ανάπτυξης Οντολογιών

- Δεν υπάρχουν τυποποιημένες μεθοδολογίες για τη ανάπτυξη οντολογιών παρά μόνο εμπειρικοί κανόνες.
- Στην εργασία M. Uschold & M. Gruninger 1996) προτείνονται οι ακόλουθες φάσεις για την ανάπτυξη οντολογιών:
 - Προσδιορισμός σκοπιμότητας και πεδίου εφαρμογής.
 - Κατασκευή της οντολογίας:
 - Σύλληψη (capture).
 - Κωδικοποίηση (coding).
 - Ενοποίηση (integration) υπαρκτών οντολογιών.
 - Αξιολόγηση (evaluation).
 - Τεκμηρίωση (documentation).

Προσδιορισμός σκοπιμότητας και πεδίου εφαρμογής

- Σημαντικό να ξεκαθαριστεί το **γιατί** να κατασκευαστεί η οντολογία και **ποιες** είναι οι πιθανές χρήσεις της.
- Χρήσιμο είναι επίσης να προσδιοριστεί ποιοι πρόκειται να είναι οι **πιθανοί χρήστες** της οντολογίας.

Κατασκευή της οντολογίας

- **Σύλληψη**
 - Προσδιορισμός των βασικών εννοιών και των μεταξύ τους σχέσεων.
 - Παραγωγή σαφών προδιαγραφών των εννοιών και των μεταξύ τους σχέσεων σε μορφή κειμένου.
 - Συμφωνία για τους όρους με τους οποίους θα αναφερόμαστε στις έννοιες και σχέσεις.
- **Κωδικοποίηση**: ρητή αναπαράσταση της σύλληψης του προηγούμενου σταδίου σε μια τυπική γλώσσα.
- **Ενοποίηση υπάρχουσών οντολογιών**:
 - Θα πρέπει (και πως) να χρησιμοποιηθούν υπάρχουσες οντολογίες (ή τμήματα αυτών); Γενικά δύσκολο πρόβλημα.
- **Αξιολόγηση**: «έκφραση τεχνικών κρίσεων σχετικά με τις οντολογίες, το σχετιζόμενο με αυτές περιβάλλον λογισμικού, και την τεκμηρίωση σε σχέση με ένα πλαίσιο αναφοράς...» (Gomez-Perez, 1995).
- **Τεκμηρίωση**: όλες οι σημαντικές παραδοχές τόσο αναφορικά με τις βασικές έννοιες και με τα βασικά δομικά στοιχεία που χρησιμοποιήθηκαν για την έκφραση αυτών των εννοιών, πρέπει να τεκμηριωθούν.

Μερικές εφαρμογές των οντολογιών στο Σημασιολογικό Ιστό

- Διευκόλυνση της επικοινωνίας αγοραστών-πωλητών στο ηλεκτρονικό εμπόριο.
- Σημασιολογική αναζήτηση πληροφορίας
- Παροχή πληρέστερων περιγραφών υπηρεσιών οι οποίες θα ερμηνεύονται ευκολότερα από ευφυείς διαμεσολαβητές.