



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΑΓΡΟΝΟΜΩΝ ΤΟΠΟΓΡΑΦΩΝ ΜΗΧΑΝΙΚΩΝ - ΜΗΧΑΝΙΚΩΝ ΓΕΩΠΛΗΡΟΦΟΡΙΚΗΣ  
**ΔΠΜΣ ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ**  
**ΧΩΡΙΚΕΣ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ**

# Διαλειτουργικότητα - Ολοκλήρωση Βάσεων Χωρικών Δεδομένων

**Μαργαρίτα Κόκλα**  
Επικ. Καθηγήτρια  
Εργαστήριο Χαρτογραφίας  
ΣΑΤΜ - ΜΓ ΕΜΠ

# ΔΙΑΛΕΙΤΟΥΡΓΙΚΟΤΗΤΑ

# Παράδειγμα

- Σκοπός: δημιουργία εφαρμογής για τον εντοπισμό και έλεγχο της μόλυνσης κατά μήκος του ποταμού Δούναβη.

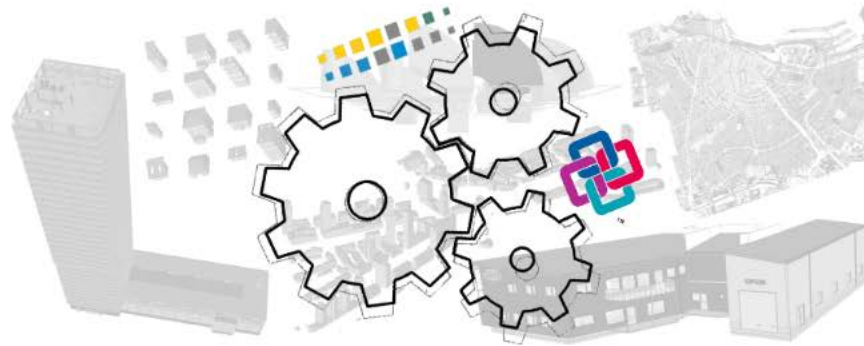


<http://www.ssqq.com/travel/images/danuberiver2016x007.jpg>

- Ο ποταμός περνάει ή αγγίζει τα σύνορα δέκα χωρών: Γερμανία, Αυστρία, Σλοβακία, Ουγγαρία, Κροατία, Σερβία, Βουλγαρία, Ρουμανία, Μολδαβία και Ουκρανία. Η λεκάνη απορροής του εκτείνεται σε εννέα ακόμη χώρες.
- Πρόβλημα: διαφορετικές βάσεις δεδομένων από διαφορετικές χώρες: διαφορές σε γλώσσα, προδιαγραφές, χρόνο ενημέρωσης, κλπ.



## GeoBIM benchmark 2019



### Recent news

Jan 11 The three final journal papers are published and open access online. See them in the 'presentations&publications' page.

Sep 02 Final results of the benchmark are now available in the 'presentations&publications' page. Thank you to all the people being involved in the initiative.

Dec 22 The benchmark is finished, thank you to all the involved people for the participation! The website is up-to-date with final participants and (almost) final list of software. In the 'event' page of this website, you can also find the materials and recordings of the GeoBIM benchmark workshop. Have a nice Christmas break everyone!!

[All news](#)

### Important dates and next steps

#### March 2019

- Complete materials available
- Start of declaration of interest from participants

#### July 8th, 2019

- GeoBIM benchmark meeting with participants and proponents

#### October 31st, 2019

- Deadline for data processing and benchmark answer submission

#### December 2nd-3rd, 2019

- EuroSDR-AMS GeoBIM benchmark workshop

### Team

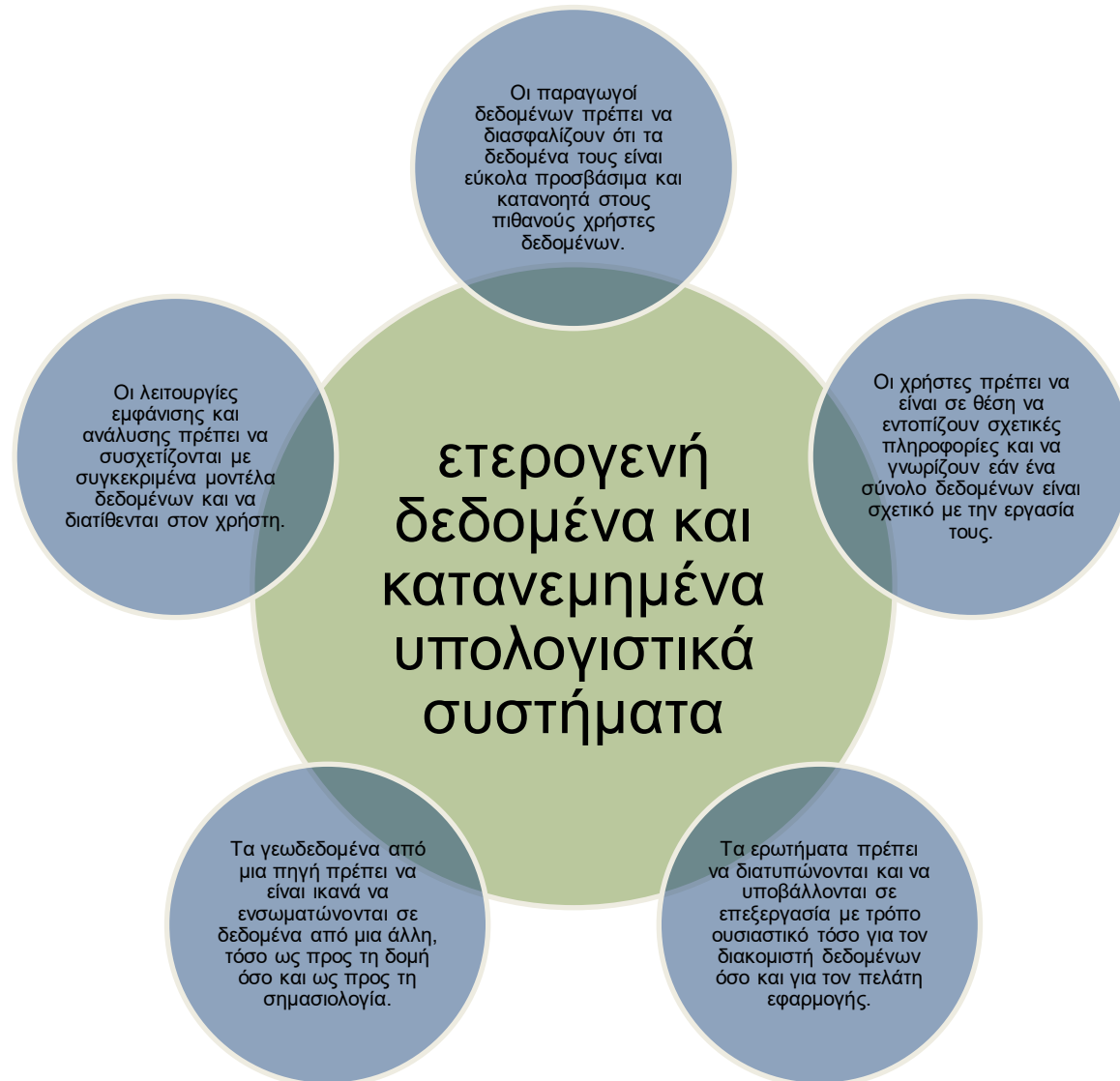
Francesca Noardo  
 Ken Arroyo Ohori  
 Jantien Stoter  
 Filip Biljecki  
 Claire Ellul  
 Lars Harrie  
 Thomas Krijnen  
 Giorgio Agugiaro  
 Margarita Kokla

# Ορισμοί διαλειτουργικότητας

- *ικανότητα δύο ή περισσότερων συστημάτων να ανταλλάσσουν πληροφορίες και να χρησιμοποιούν τις πληροφορίες που έχουν ανταλλάξει* (IEEE Standard Computer Dictionary)
- *ικανότητα επικοινωνίας, εκτέλεσης προγραμμάτων ή μεταφοράς δεδομένων μεταξύ διαφόρων λειτουργικών μονάδων με τρόπο που απαιτεί ο χρήστης να έχει λίγη ή καθόλου γνώση των μοναδικών χαρακτηριστικών αυτών των μονάδων* (International Organization for Standardization)

- Η διαλειτουργικότητα αναφέρεται συνήθως σε προσπάθειες από κάτω προς τα πάνω που δεν επιβάλλονται από μια κεντρική αρχή ούτε οδηγούνται από μία μόνο εφαρμογή.
- Τα **συστήματα και τα μοντέλα δεδομένων** των χρηστών είναι **ετερογενή**, έχουν αναπτυχθεί **ανεξάρτητα** το ένα από το άλλο.

# Στόχοι



Τα αυτόνομα συστήματα πρέπει να αντιμετωπίσουν δύο σημαντικές προκλήσεις:

Πρόκληση	Λύση
ανταλλαγή δεδομένων και επεξεργασία ερωτημάτων και άλλων αιτημάτων	κοινό σύνολο υπηρεσιών διαθέσιμο και προσβάσιμο μέσω επικοινωνιών δικτύου
κοινή κατανόηση των δεδομένων και των αιτημάτων	κοινή γλώσσα και μοντέλο αναπαράστασης



## Διαλειτουργικότητα σημαίνει:

- **«Ανοιχτότητα»** (openness) όσον αφορά τη βιομηχανία λογισμικού, δηλαδή την ελεύθερη δημοσίευση των εσωτερικών δομών των δεδομένων.
- **Ελεύθερη ανταλλαγή** (free exchange) δεδομένων, αφού κάθε σύστημα θα γνωρίζει τη μορφή των δεδομένων των άλλων συστημάτων - πρότυπα ανταλλαγής.
- **Κοινή αλληλεπίδραση των χρηστών** (commonality in user interaction), δηλαδή ο βαθμός εξοικείωσης των χρηστών με τα περιβάλλοντα αλληλεπίδρασης.

## Διαλειτουργικότητα σημαίνει:

- **Απλοποίηση** (simplification) στις μορφές των δεδομένων και στα πρότυπα, στην αλληλεπίδραση μεταξύ του χρήστη και του συστήματος, στη γνώση που απαιτείται από τον χρήστη.
- **Διαφάνεια** (transparency), δηλαδή αποδέσμευση του χρήστη από τις λεπτομέρειες υλοποίησης.
- **Ομοιότητα** (similarity), δηλαδή ο βαθμός στον οποίο τα σύνολα δεδομένων, τα συστήματα λογισμικού ή οι οργανισμοί χρησιμοποιούν το ίδιο λεξιλόγιο, ακολουθούν τις ίδιες συμβάσεις και επομένως μπορούν να διαλειτουργούν.

## Επίπεδα διαλειτουργικότητας

Οργανωτική διαλειτουργικότητα

Σημασιολογική διαλειτουργικότητα

Συντακτική διαλειτουργικότητα

Τεχνική διαλειτουργικότητα

# Τεχνική διαλειτουργικότητα

- καλύπτει τα **τεχνικά ζητήματα της σύνδεσης συστημάτων και υπηρεσιών**. Περιλαμβάνει βασικές πτυχές, όπως ανοικτές διεπαφές, υπηρεσίες διασύνδεσης, ανταλλαγή, ενοποίηση και παρουσίαση δεδομένων, προσβασιμότητα και ασφάλεια των υπηρεσιών
- συνδέεται συνήθως με το **υλικό / λογισμικό, τα συστήματα και τις πλατφόρμες** που επιτρέπουν την **επικοινωνία μεταξύ συστημάτων**
- επικεντρώνεται σε **πρωτόκολλα επικοινωνίας και στην απαραίτητη υποδομή** για τη λειτουργία τους

# Συντακτική διαλειτουργικότητα

- συνδέεται συνήθως με **μορφές δεδομένων**
- τα μηνύματα που μεταφέρονται από τα πρωτόκολλα επικοινωνίας πρέπει να έχουν μια **καλά καθορισμένη σύνταξη και κωδικοποίηση**
- πολλά πρωτόκολλα μεταφέρουν δεδομένα ή περιεχόμενο, και αυτό μπορεί να αναπαρασταθεί με τη χρήση γλωσσών σήμανσης, όπως HTML και XML

# Σημασιολογική διαλειτουργικότητα

- ασχολείται με τη διασφάλιση ότι το ακριβές νόημα των ανταλλασσόμενων πληροφοριών γίνεται κατανοητό από οποιαδήποτε άλλη εφαρμογή που δεν είχε αρχικά αναπτυχθεί για τον σκοπό αυτό.
- παρέχει τη δυνατότητα στα συστήματα να **συνδυάσουν** ληφθείσες πληροφορίες με άλλες πληροφορίες και να τις επεξεργαστούν με τρόπο ουσιαστικό.

# Οργανωτική διαλειτουργικότητα

- ασχολείται με τον καθορισμό των στόχων των οργανισμών ή επιχειρήσεων, τη μοντελοποίηση διαδικασιών και τη συνεργασία των οργανισμών ή επιχειρήσεων που επιθυμούν μεν να ανταλλάσσουν πληροφορίες αλλά έχουν διαφορετικές εσωτερικές δομές και διαδικασίες
- στοχεύει στην αντιμετώπιση των απαιτήσεων της κοινότητας των χρηστών, καθιστώντας τις υπηρεσίες διαθέσιμες, εύκολα αναγνωρίσιμες, προσιτές και προσανατολισμένες προς τον χρήστη.

# Επίπεδα διαλειτουργικότητας

Επίπεδο	Στόχος	Αντικείμενα	Λύσεις	Επίπεδο επίτευξης
Τεχνική	Ασφαλής ανταλλαγή δεδομένων	Σήματα	Πρωτόκολλα μεταφοράς δεδομένων	Πλήρως υλοποιημένη
Συντακτική	Επεξεργασία λαμβανόμενων δεδομένων	Δεδομένα	Πρότυπα ανταλλαγής δεδομένων, XML	Πλήρως υλοποιημένη
Σημασιολογική	Επεξεργασία και ερμηνεία λαμβανόμενων πληροφοριών	Πληροφορία	Οντολογίες	Θεωρητικά τεκμηριωμένη, αλλά με πρακτικά προβλήματα υλοποίησης
Οργανωτική	Αυτόματη σύνδεση διαδικασιών μεταξύ διαφορετικών συστημάτων	Διαδικασίες (ροή εργασιών)	Αρχιτεκτονικά μοντέλα, τυποποιημένες διαδικασίες	Απουσία εννοιολογικής σαφήνειας, αόριστες έννοιες με μεγάλο εύρος ερμηνείας



# Open GeoSpatial Consortium

- ένας διεθνής, μη κερδοσκοπικός οργανισμός με αντικείμενο την **ανάπτυξη και εφαρμογή ανοικτών προτύπων για γεωχωρικές πληροφορίες και υπηρεσίες προκειμένου να υποστηρίζεται η διαλειτουργικότητα μεταξύ διαφορετικών πηγών, υπηρεσιών και εφαρμογών γεωγραφικών δεδομένων.**
- Αυτά τα πρότυπα αναπτύσσονται μέσω μιας διαδικασίας συναίνεσης και είναι ελεύθερα διαθέσιμα για χρήση.

# OGC standards

<http://www.ogc.org/docs/is/>

- [Geography Markup Language Encoding Standard \(GML\)](#)
- [Observations and Measurements conceptual model](#)
- [OGC GeoSPARQL - A Geographic Query Language for RDF Data](#)
- [Web Feature Service \(WFS\) Implementation Specification](#)
- [Web Coverage Processing Service \(WCPS\)](#)
- [CityGML](#)
- .....

# ISO/TC 211 geographic information standards <https://committee.iso.org/home/tc211>

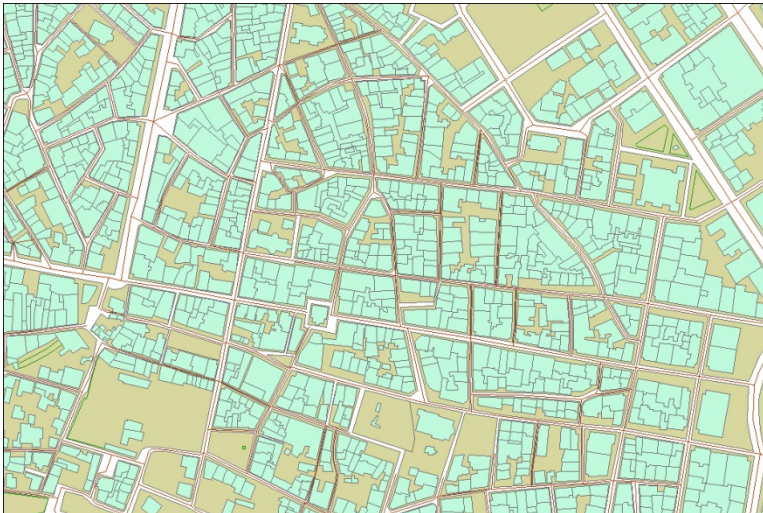
- Reference model for Open distributed processing (RM-ODP)
- Geometry
- Topology
- Temporal concepts
- Coverage geometry
- Tracking and navigation
- Multimodal routing and navigation
- Moving features
- Spatial referencing by coordinates
- Data quality
- Metadata
- Rules for application schema
- Methodology for feature cataloguing

# Είδη ετερογενειών

- Συντακτική Ετερογένεια
- Σχηματική Ετερογένεια
- Σημασιολογική Ετερογένεια

# Συντακτική Ετερογένεια

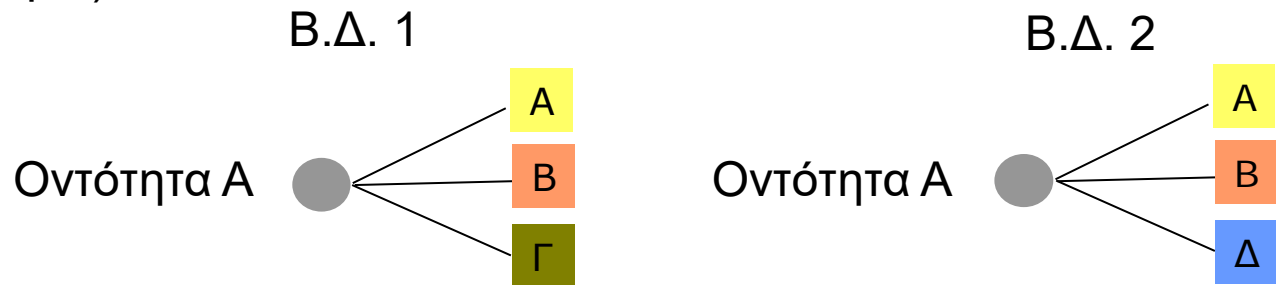
- ΣΔΒΔ που χρησιμοποιούν **διαφορετικά λογικά μοντέλα δεδομένων** (π.χ., σχεσιακά και αντικειμενοστραφή).
- **διαφορετικές γεωμετρικές αναπαραστάσεις** των γεωγραφικών αντικειμένων, π.χ., vector και raster.



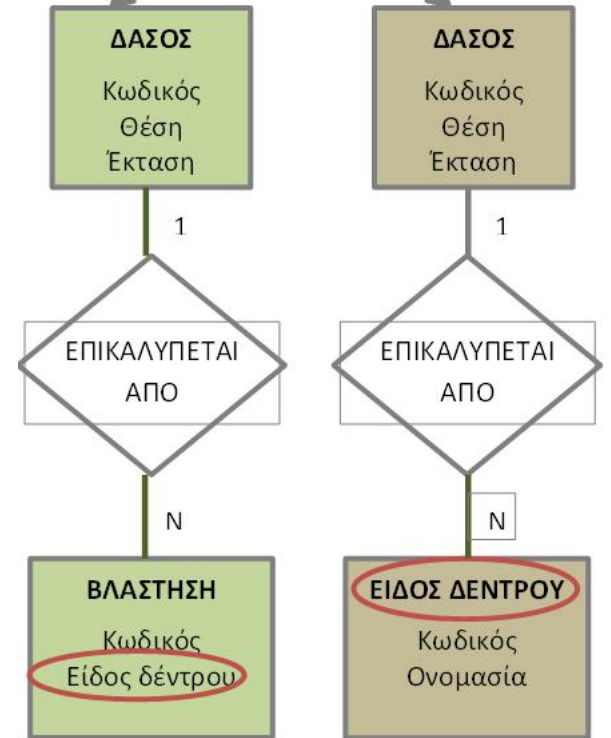
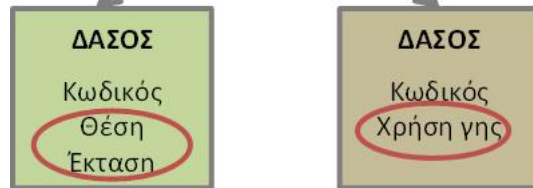
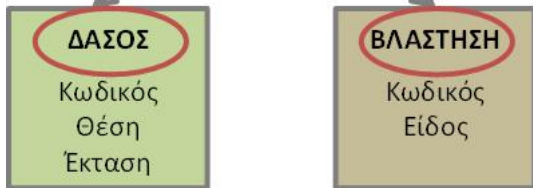
# Σχηματική Ετερογένεια

διαφορετικά εννοιολογικά μοντέλα δεδομένων λόγω διαφορετικών τρόπων ταξινόμησης των πραγματικών οντοτήτων:

- Οντότητα - Οντότητα (σχέσεις 1:M και N:M, διαφορετικά χαρακτηριστικά, ή περιορισμοί).



- Χαρακτηριστικό - Χαρακτηριστικό (π.χ., διαφορετικά πεδία τιμών).
- Οντότητα - Χαρακτηριστικό, όταν τα αντικείμενα μιας Β.Δ. θεωρούνται ιδιότητες σε μια άλλη Β.Δ.
- Διαφορετικές αναπαραστάσεις των ίδιων δεδομένων, π.χ., διαφορετικές μονάδες, διαφορετική χωρική ανάλυση.



# Σημασιολογική Ετερογένεια

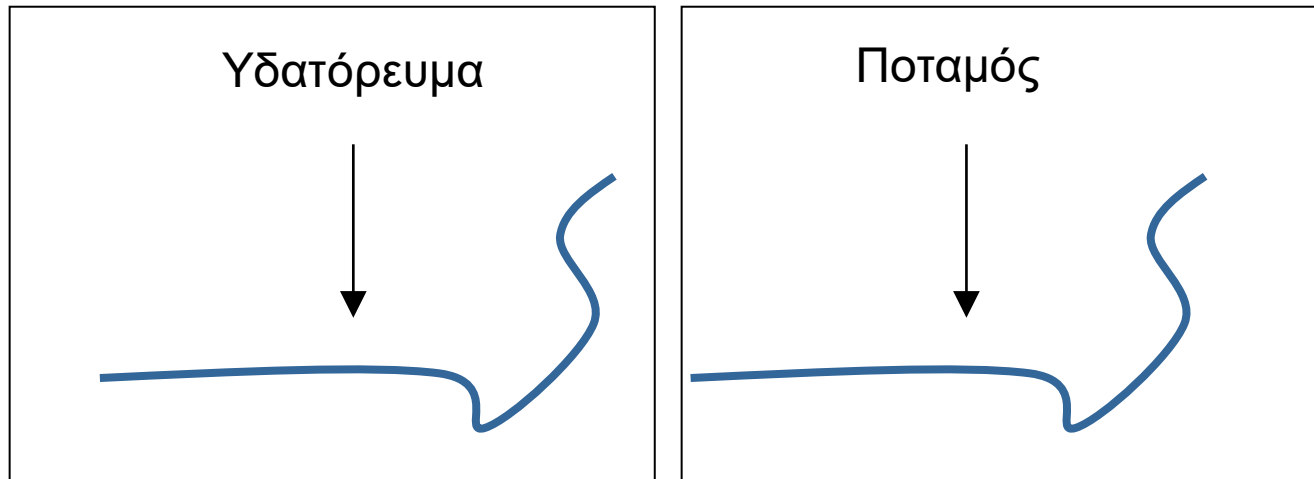
διαφορές στο **νόημα**, την **ερμηνεία** ή την **προτεινόμενη χρήση** των ίδιων ή σχετιζόμενων δεδομένων

- **ονομαστική ετερογένεια**
- **γνωσιακή ετερογένεια**



# Ονομαστική Ετερογένεια

Συνώνυμα: η ίδια έννοια περιγράφεται με διαφορετικούς όρους



# Ονομαστική Ετερογένεια

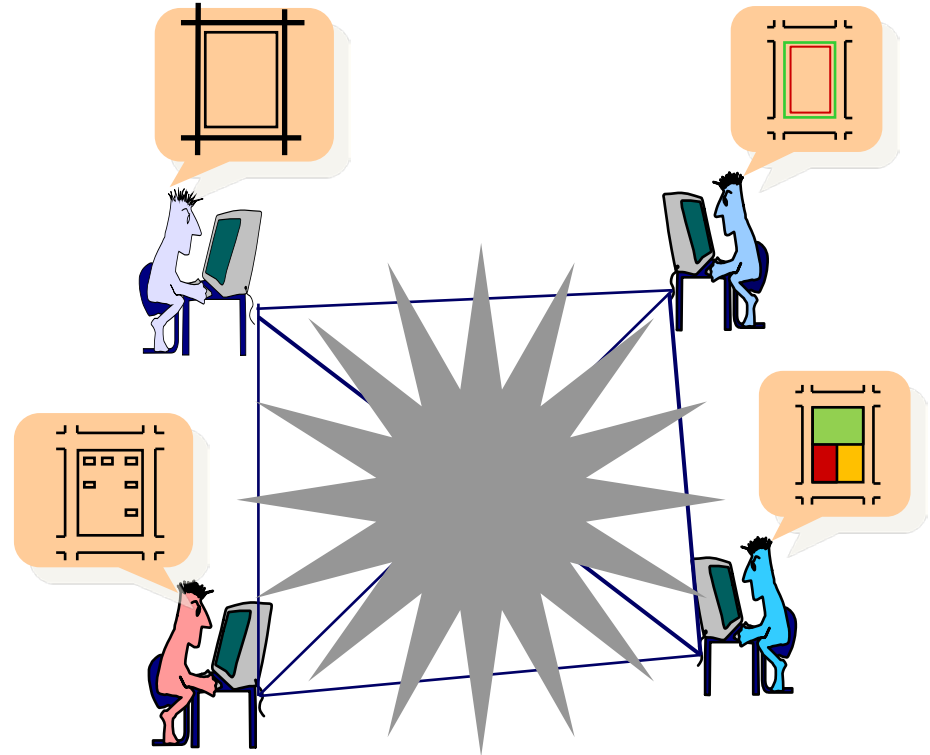
Ομώνυμα: οι ίδιοι όροι αποδίδονται σε διαφορετικές έννοιες

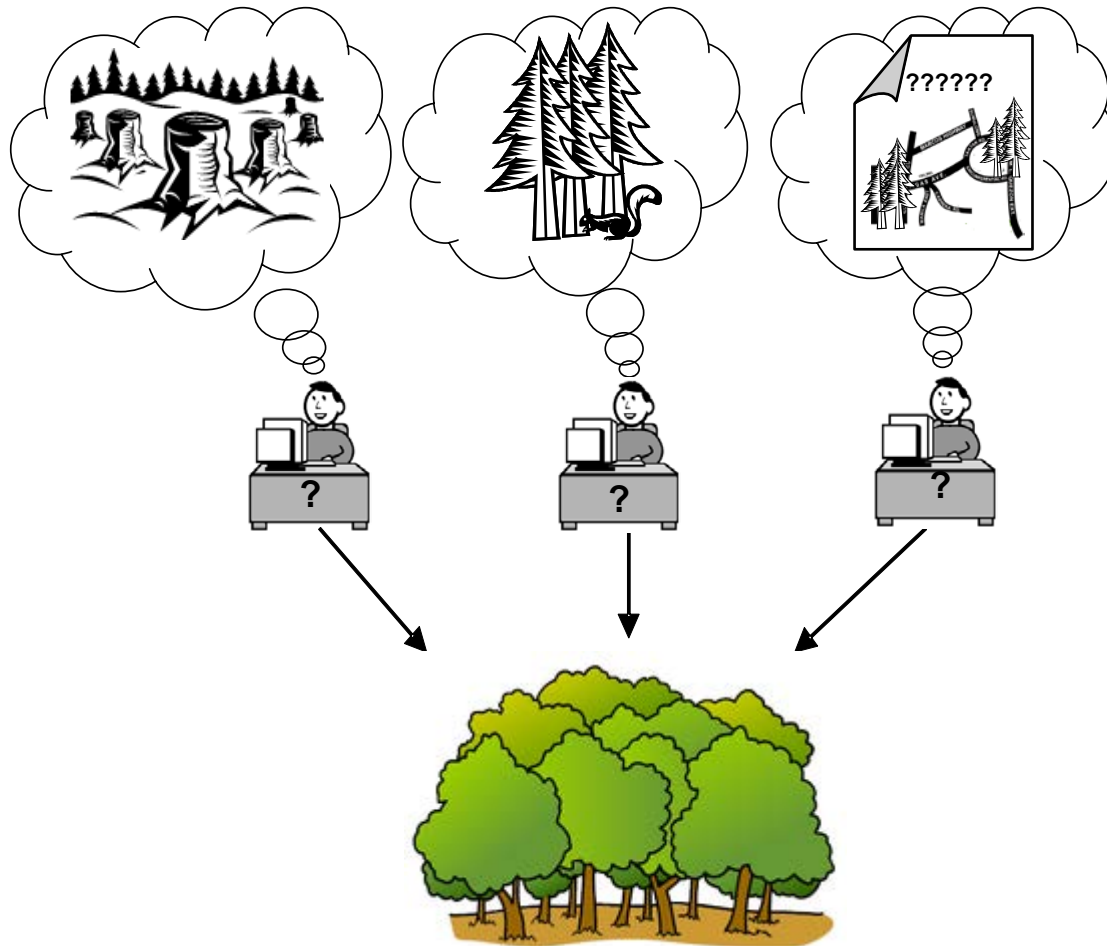
**Ρεύμα** (<http://www.greek-language.gr/>):

1. κίνηση μάζας υγρού ή αερίου προς ορισμένη κατεύθυνση
2. (ηλεκτρικό) ~, η ροή ηλεκτρονίων σε αγωγό.
3. η ροή οχημάτων που κινούνται σε δρόμο. || (επέκτ.) το τμήμα του δρόμου όπου επιτρέπεται η κίνηση αυτοκινήτων προς την ίδια κατεύθυνση.

# Γνωσιακή Ετερογένεια

διαφορετικές **αντιλήψεις** του χώρου π.χ. διαφορετικοί **ορισμοί κατηγοριών**, διαφορετική **γεωμετρική περιγραφή**.





# Γνωσιακή Ετερογένεια

- Δάσος: περιοχή εντός των διοικητικών ορίων μιας δασικής υπηρεσίας, με ή χωρίς δέντρα (νομικός ή διοικητικός ορισμός)
- Δάσος: μεγάλη περιοχή με πυκνή κάλυψη από δέντρα ή άλλα φυτά (ορισμός κάλυψης γης)
- Δάσος: περιοχή με δασώδη βλάστηση υπό διαχείριση για την παραγωγή ξυλείας και άλλων προϊόντων ή για έμμεσα οφέλη όπως αναψυχή ή προστασία λεκανών απορροής υδάτων (ορισμός χρήσης γης)
- Δάσος: οικολογικό σύστημα που αποτελείται από βιοτικά και αβιοτικά στοιχεία που αλληλεπιδρούν και στο οποίο επικρατούν τα δέντρα με κάλυψη μεγαλύτερη από το 20% της περιοχής (οικολογικός ορισμός)

# Οντολογία

- Σαφής περιγραφή των **βασικών οντοτήτων**, αλλά και των μεταξύ τους πιθανών **σχέσεων** που αναφέρονται σε ένα συγκεκριμένο **τομέα εφαρμογής**
- **Ουδέτερη**, αλλά και εύκολα **υλοποιήσιμη** περιγραφή ή θεωρία ενός συγκεκριμένου τομέα, η οποία μπορεί να γίνει **αποδεκτή** και να **επαναχρησιμοποιηθεί** από όλους
- Μια οντολογία πρακτικά παρέχει ένα κοινό λεξιλόγιο ώστε άνθρωποι ή υπολογιστικά συστήματα να μπορούν να μοιραστούν, επαναχρησιμοποιήσουν, ολοκληρώσουν, κλπ. πληροφορίες για ένα πεδίο.

# Χρήσεις των οντολογιών

1. εννοιολογικός σχεδιασμός Β.Δ.
2. ανάκτηση και εξαγωγή πληροφοριών
3. μετάφραση φυσικής γλώσσας
4. αναπαράσταση, διαχείριση, ανταλλαγή, ενοποίηση πληροφοριών
5. «έξυπνη» αναζήτηση πληροφοριών, semantic web

# Συστατικά των οντολογιών

- Έννοιες (concepts) και η περιγραφή του νοήματός τους συνήθως μέσω ορισμών
- Ιδιότητες (properties) και επιτρεπόμενες τιμές
- Σχέσεις (relations) μεταξύ των εννοιών
- Αξιώματα (axioms), περιορισμοί (constraints), κανόνες (rules)
- Χαρακτηριστικές περιπτώσεις (instances), παραδείγματα

Μια οντολογία με ένα σύνολο περιπτώσεων συνιστά μια βάση γνώσης



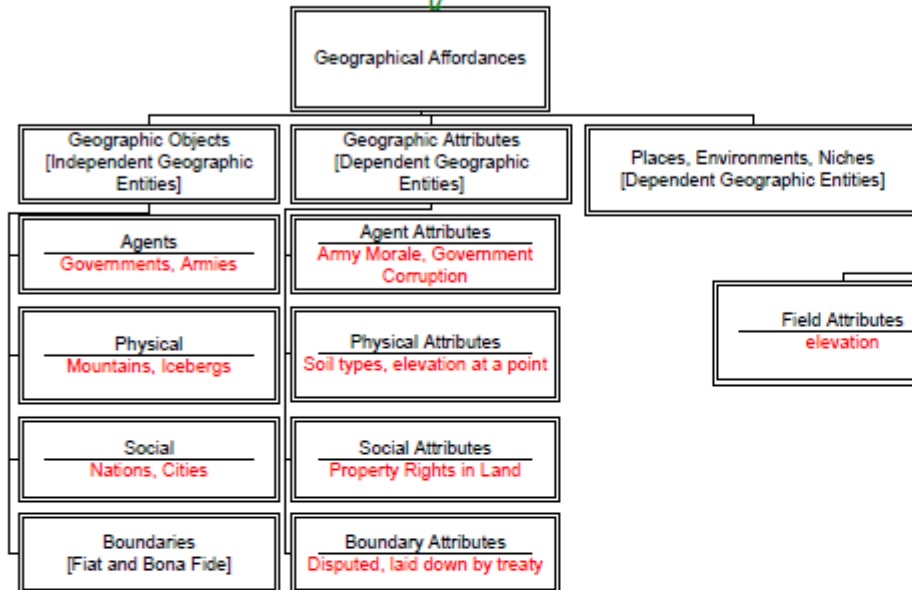
# OWL (Web Ontology Language)

- γλώσσα σημασιολογικού ιστού που αναπτύχθηκε από το World Wide Web Consortium (W3C)  
(<https://www.w3.org/OWL/>)
- παρέχει τυποποίηση για την περιγραφή πολύπλοκης γνώσης για έννοιες και συσχετίσεις μεταξύ τους
- βασίζεται στο Resource Description Framework (RDF)

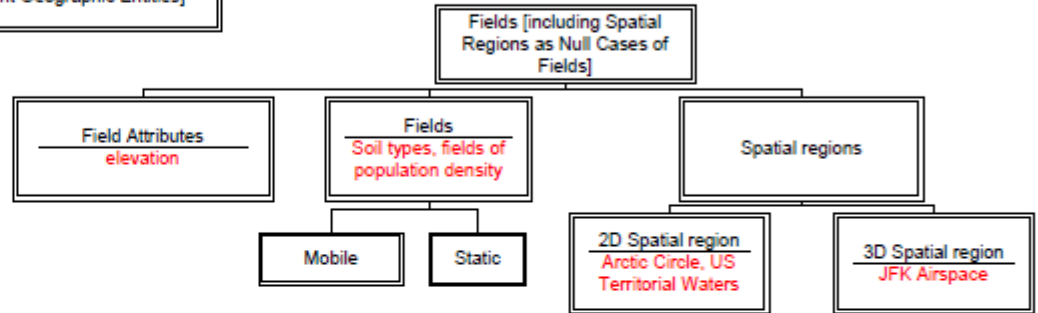
# Geographic Ontology

6 August 2002

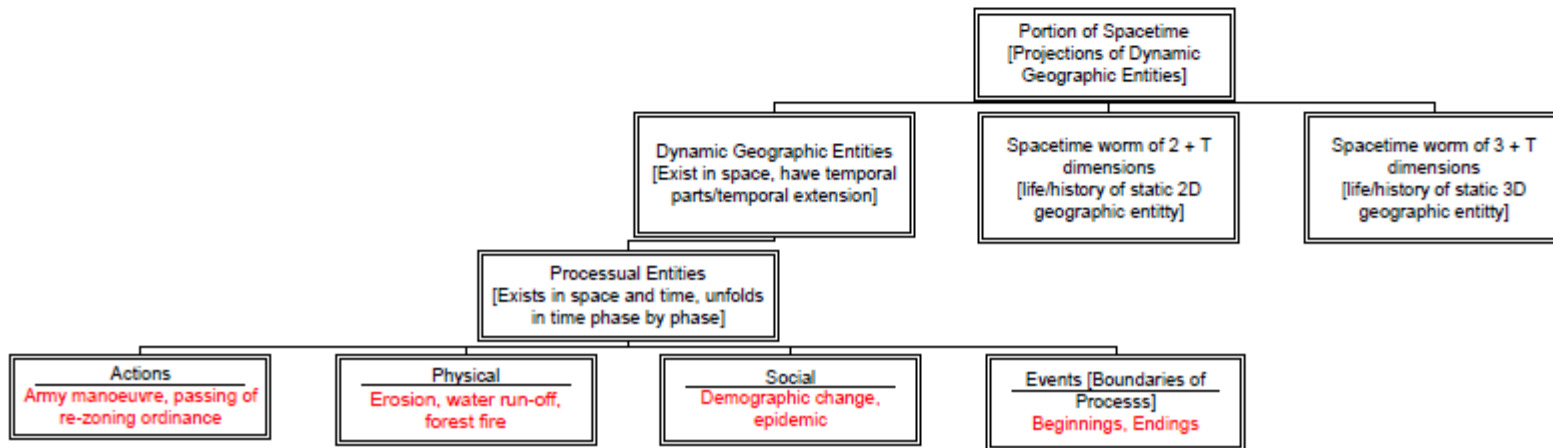
## SNAP<sub>tr</sub> view



## FIELD view

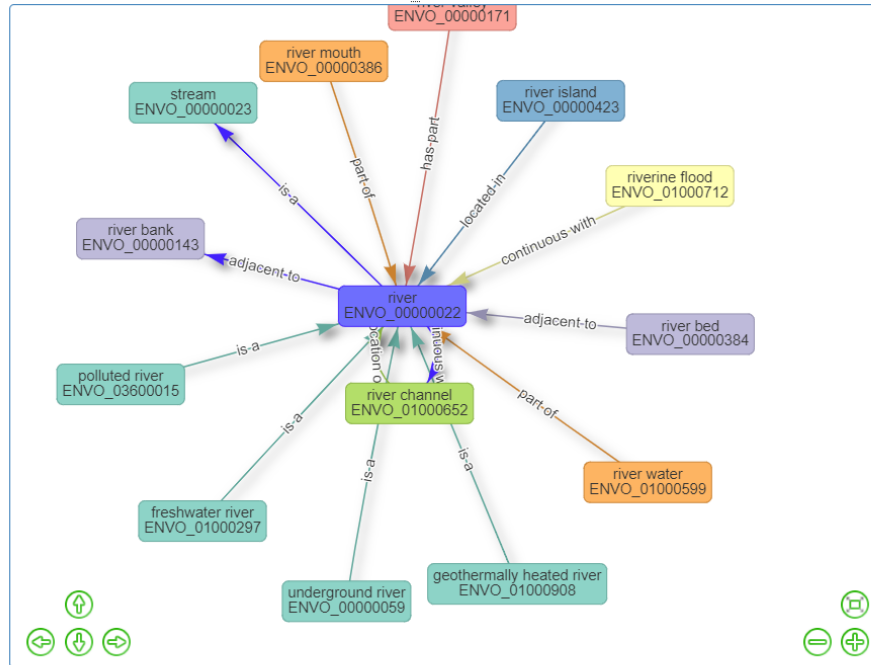


## SPAN view



Basic Formal Ontology (<https://basic-formal-ontology.org/>)

Visualized term: river ([http://purl.obolibrary.org/obo/ENVO\\_0000022](http://purl.obolibrary.org/obo/ENVO_0000022))



Legend

Relationship	Color	Visibility
Extended nodes (*)	Red	-
is a	Teal	<input checked="" type="checkbox"/>
continuous with	Yellow	<input checked="" type="checkbox"/>
adjacent to	Purple	<input checked="" type="checkbox"/>
has part	Red	<input checked="" type="checkbox"/>
located in	Blue	<input checked="" type="checkbox"/>
part of	Orange	<input checked="" type="checkbox"/>
location of	Green	<input checked="" type="checkbox"/>
Select/Deselect all		<input type="checkbox"/>

List of extended nodes (\*):

- river (ENVO\_0000022)

Create clusters | Open all clusters | Auto rearrange on | Hierarchical layout | Search nod | Search Node

[https://www.ebi.ac.uk/ols/ontologies/envo/terms/graph?iri=http://purl.obolibrary.org/obo/ENVO\\_0000022](https://www.ebi.ac.uk/ols/ontologies/envo/terms/graph?iri=http://purl.obolibrary.org/obo/ENVO_0000022)

river English Translate into...

+ See more

DEFINITIONS

EXAMPLES

RELATIONS

SOURCES

English > More languages...

IS A

stream • film

HAS PART

channel • water • River mouth • Backwater • streamflow +7 relations

PART OF

water system • drainage system • The Loire Valley between Sully-sur-Loire and Chalennes

HAS KIND

tributary • channel • distributary • tidal river • wadi +44 relations

HAS INSTANCE

Aar • Acheron • Adige • River Aire • Alabama +351K relations

AWARD RECEIVED

biotope of the year

COUNTRY OF ORIGIN

India

DESCRIBED BY SOURCE

Brockhaus and Efron Encyclopedic Dictionary • Explanatory Dictionary of the Living Great Russian Language • Gujin Tushu Jicheng

DIFFERENT FROM

stream

MODEL ITEM

Amazon • Yangtze • Congo • Mississippi River • River Thames

ON FOCUS LIST OF WIK

ru:Река

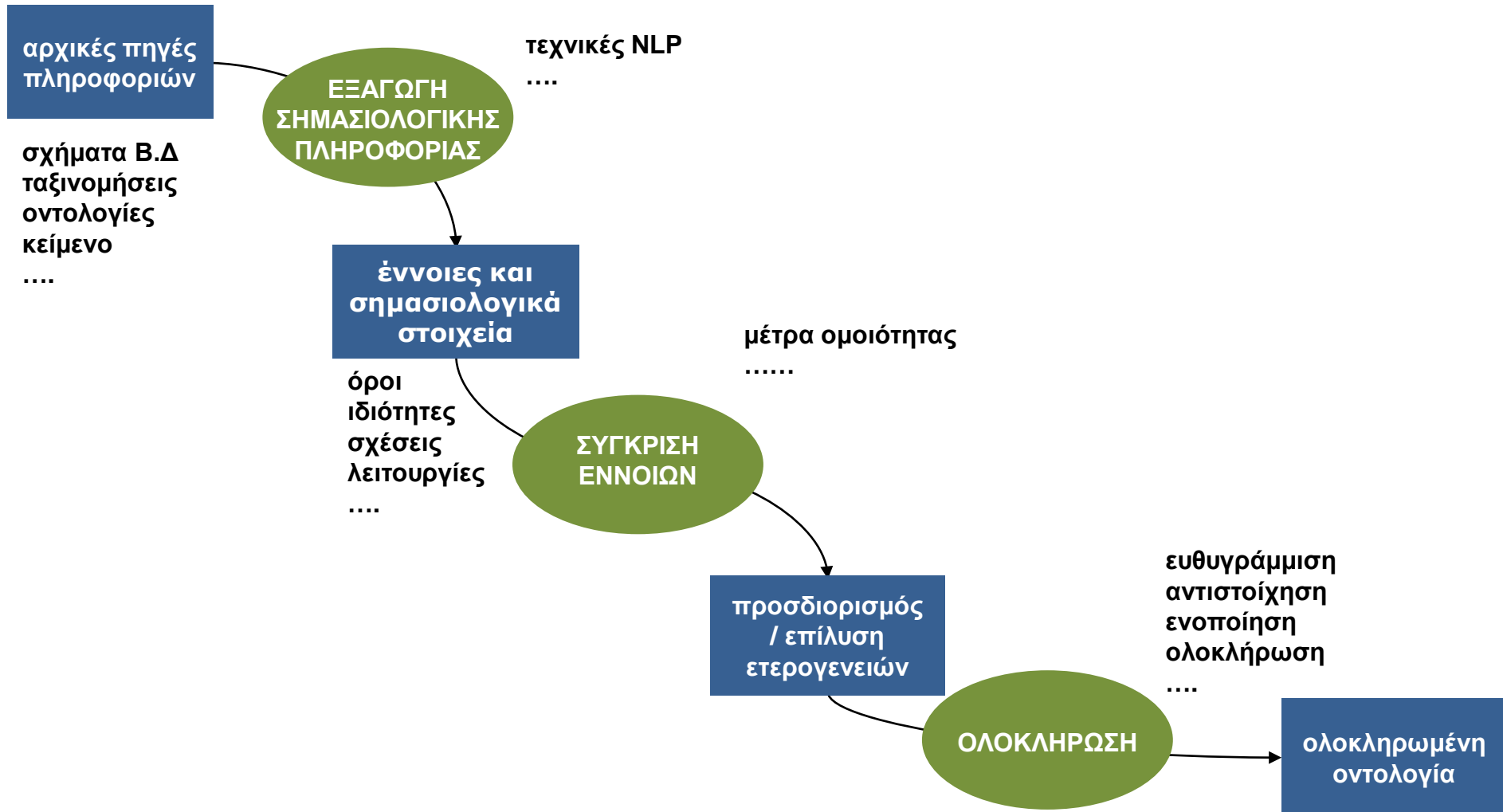
https://babelnet.org/synset?id=bn%3A00067948n&orig=river&lang=EN

# **ΣΗΜΑΣΙΟΛΟΓΙΚΗ ΟΛΟΚΛΗΡΩΣΗ (SEMANTIC INTEGRATION)**

# Σημασιολογική Ολοκλήρωση (Semantic Integration)

- **Συσχέτιση πληροφοριών από διαφορετικές πηγές:** τα συστήματα και τα μοντέλα δεδομένων των χρηστών είναι ετερογενή, έχουν αναπτυχθεί ανεξάρτητα το ένα από το άλλο.
- **Τρία θεμελιώδη ζητήματα:**
  1. Είδη ετερογενειών – τρόπος εντοπισμού και επίλυσης
  2. Επιμέρους διαδικασίες σημασιολογικής ολοκλήρωσης
  3. Είδη και επίπεδα ολοκλήρωσης

# Διαδικασίες Σημασιολογικής Ολοκλήρωσης



# Διαδικασία 1: Εξαγωγή Σημασιολογικής Πληροφορίας

αρχικές πηγές  
πληροφοριών

σχήματα Β.Δ  
ταξινομήσεις  
οντολογίες  
κείμενο  
....

ΕΞΑΓΩΓΗ  
ΣΗΜΑΣΙΟΛΟΓΙΚΗΣ  
ΠΛΗΡΟΦΟΡΙΑΣ

ΤΕΧΝΙΚΕΣ NLP

έννοιες και  
σημασιολογικά  
στοιχεία

όροι  
ιδιότητες  
σχέσεις  
λειτουργίες  
....

**ΣΤΟΧΟΣ:**  
εξαγωγή και  
τυποποίηση  
σημασιολογικής  
πληροφορίας





**Pre-processing**

**Entity extraction and linking**

**Concept extraction and linking**

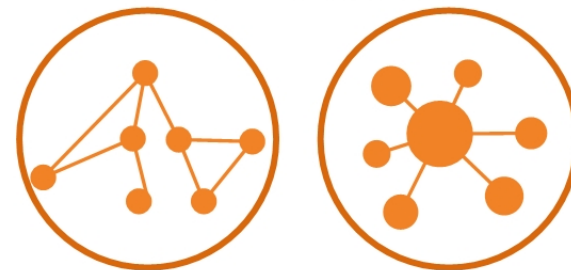
**Relation extraction and linking**



**Annotated content**



**ontologies and  
knowledge bases**



# Entity extraction / linking

The **2004 Indian Ocean earthquake** occurred in the coast of **Sumatra, Indonesia** on **December 26, 2004**.

The earthquake generated a tsunami which caused severe damage and casualties in the populated coastal areas of **Indonesia, Sri Lanka, India, and Thailand**.

Entity extraction

**2004 Indian Ocean earthquake**

IS A earthquake • tsunami  
COUNTRY Indonesia

**Sumatra**

IS A Island  
COUNTRY Indonesia  
PART OF Indonesia • Greater Sunda Islands  
CONTINENT Asia  
POPULATION 50,365,538  
LATITUDE N 0° 0' 0"  
LONGITUDE E 102° 0' 0"

**Indonesia**

IS A country  
PART OF Austronesia • Southeast Asia  
CAPITAL Jakarta  
CONTINENT Asia  
POPULATION 267,663,435  
LATITUDE S 5° 0' 0"  
LONGITUDE E 120° 0' 0"



Entity linking

# Concept extraction / linking

The **earthquake** occurred in the coast of **Sumatra, Indonesia** on **December 26, 2004**.

The **earthquake** generated a **tsunami** which caused severe damage and casualties in the populated **coastal areas** of **Indonesia, Sri Lanka, India, and Thailand**.

Concept extraction

## earthquake

SYNONYM TO	quake · seism · temblor · seismic activity
DEFINITION	shaking and vibration at the surface of the earth resulting from underground movement along a fault plane or from volcanic activity
IS A	geological phenomenon · phenomenon
HAS INSTANCE	2002 Hindu Kush earthquakes · 2002 Molise earthquake · 2007 Noto earthquake
HAS EFFECT	disaster · landslide · tsunami
HAS QUALITY	epicenter · Hypocenter
MEASUREMENT SCALE	moment magnitude scale · Richter scale



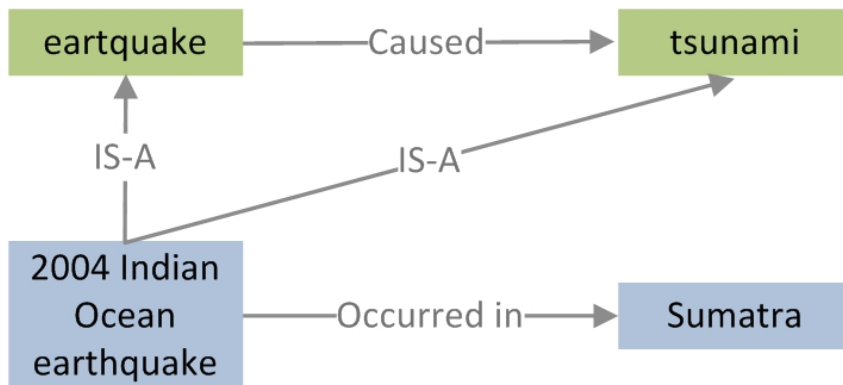
Concept linking

# Relation extraction / linking

The 2004 Indian Ocean earthquake occurred in the coast of Sumatra, Indonesia on December 26, 2004.

The earthquake generated a tsunami which caused severe damage and casualties in the populated coastal areas of Indonesia, Sri Lanka, India, and Thailand.

Relation extraction



Relation linking

Type the content to annotate:

The 2004 Indian Ocean earthquake occurred in the coast of Sumatra, Indonesia on December 26, 2004.  
The earthquake generated a tsunami which caused severe damage and casualties in the populated coastal areas of Indonesia, Sri Lanka, India, and Thailand.

Or select a text file:

Choose File No file chosen

Output type:

JSON

Document format:

plain text

Restore defaults

Term  Sentiment  Sentence  Person  Location  Organization  Hashtag  URL  UserID

Test Pipeline

download

Annotation types:  Location  Sentence  Sentiment  Term

The 2004  Indian Ocean earthquake occurred in the coast of Sumatra,  Indonesia on December 26, 2004. The earthquake generated a tsunami which caused severe damage and casualties in the populated coastal areas of  Indonesia,  Sri Lanka,  India, and  Thailand.

### Annotations at this location

#### Sentiment

emotion	sadness
polarity	negative
sarcasm	no

<https://cloud.gate.ac.uk/shopfront/displayItem/environmental-annotator>

# Διαδικασία 2: Σύγκριση Κατηγοριών



# Εκτίμηση ομοιότητας

- Η εκτίμηση ομοιότητας είναι η βάση για την αντιστοίχιση μεταξύ των κατηγοριών
- Μπορεί να είναι είτε **ποσοτική**, είτε **ποιοτική**
- **Ποσοτική εκτίμηση ομοιότητας**: υπολογίζεται με μέτρα ομοιότητας και εκφράζεται στο διάστημα  $[0, 1]$ , όπου το 0 σημαίνει ότι οι έννοιες είναι διαφορετικές και 1 σημαίνει ότι οι έννοιες είναι ισοδύναμες.
- **Ποιοτική εκτίμηση ομοιότητας**: υπολογίζεται με τη μορφή κατηγοριών που καθορίζουν τη σχέση μεταξύ δύο εννοιών, όπως ισοδυναμία, συμπερίληψη, επικάλυψη κ.λπ.

# Διαδικασία 2:

## Σύγκριση Κατηγοριών

- **Στόχοι:**

1. προσδιορισμός ομοιοτήτων και ετερογενειών
2. επίλυση ετερογενειών

- **Μεθοδολογίες:**

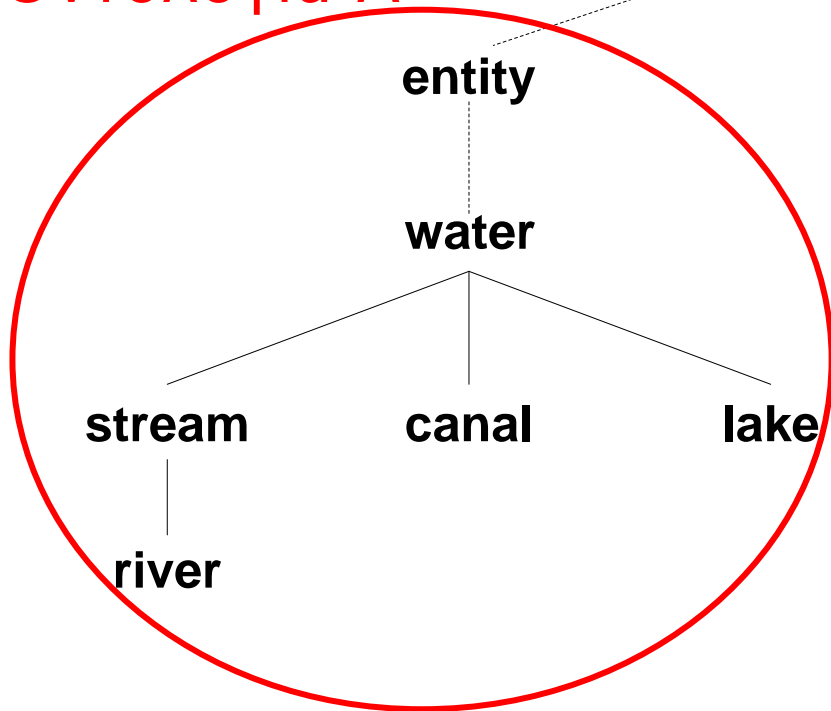
1. για τον προσδιορισμό ομοιοτήτων και ετερογενειών
  - ποσοτική εκτίμηση ομοιότητας
    - τεχνικές σύγκρισης όρων
    - τεχνικές σύγκρισης χαρακτηριστικών στοιχείων
    - υπολογισμός σημασιολογικής απόστασης
2. για την επίλυση ετερογενειών
  - ποιοτική εκτίμηση ομοιότητας - συνήθως βασίζεται στη λήψη αποφάσεων



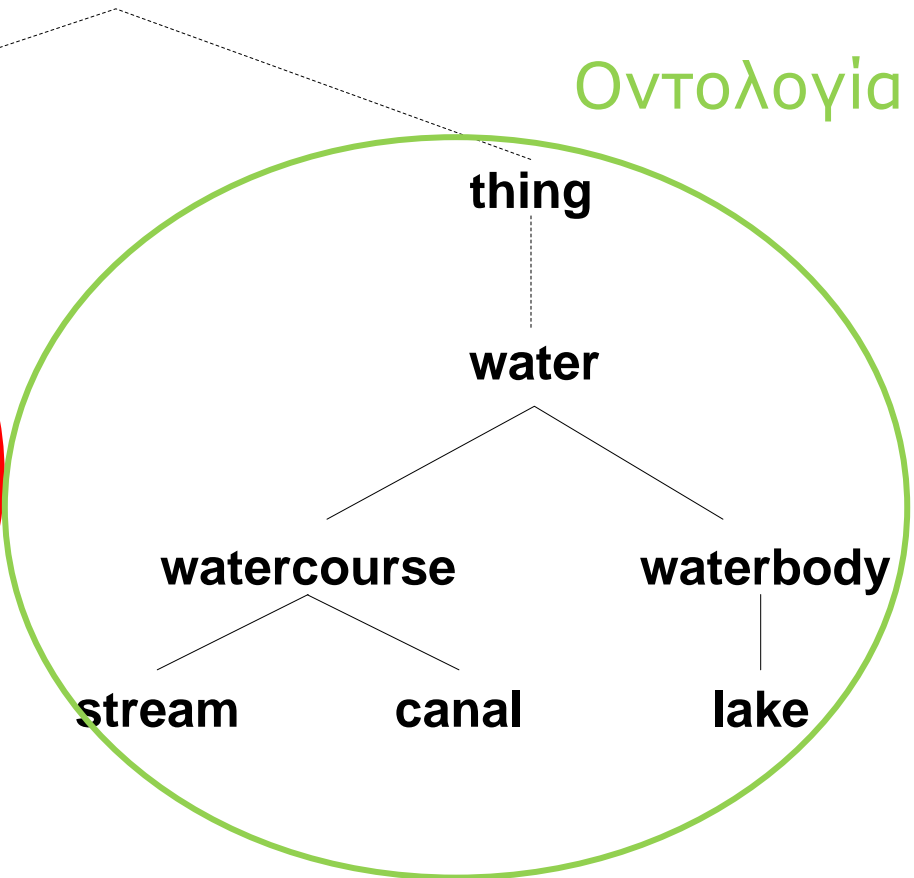
Ontology A		
Water:	Part-of: the Earth's surface	Attributes: {cover}
Stream:	Has-Parts: {ford, meander, midstream}	Attributes: {{cover}, {nature}, {flow}}
River:	Has-Parts: {estuary, rapid, waterfall}	Attributes: {{cover}, {nature}, {size}, {flow}}
Lake:	Has-Parts: {inlet}	Attributes: {{cover}, {surroundness}}
Canal:	Has-Parts: {lock, lockage}	Attributes: {{cover}, {nature}, {purpose}}
Ontology B		
Water:	Part-of: the Earth's surface	Attributes: {cover}
Watercourse:		Attributes: {cover}
Stream:	Has-Parts: {midstream, riverbank}	Attributes: {{cover}, {nature}, {flow}}
Canal:	Has-Parts: {lock}	Attributes: {{cover}, {nature}, {purpose}}
Waterbody:		Attributes: {cover}
Lake:	Has-Parts: {inlet}	Attributes: {{cover}, {surroundness}}

# Σημασιολογική Γειτονιά

Οντολογία A



Οντολογία B



# Rodriguez και Egenhofer (2003, 2004)

$$S(a^p, b^q) = w_w S_w(a^p, b^q) + w_u S_u(a^p, b^q) + w_n S_n(a^p, b^q)$$

για  $w_w, w_u,$  and  $w_n \geq 0$ , όπου

w: synonym words

u: distinguishing features

n: semantic relations (semantic neighborhood)

$$S(a, b) = \frac{|A \cap B|}{|A \cap B| + \alpha(a, b)|A/B| + (1 - \alpha(a, b))|B/A|} \quad 0 \leq a \leq 1$$

$$\alpha(a^p, b^q) = \left\{ \begin{array}{l} \frac{\text{depth}(a^p)}{\text{depth}(a^p) + \text{depth}(b^q)} \text{depth}(a^p) \leq \text{depth}(b^q) \\ 1 - \frac{\text{depth}(a^p)}{\text{depth}(a^p) + \text{depth}(b^q)} \text{depth}(a^p) > \text{depth}(b^q) \end{array} \right\}$$

## Σύγκριση κατηγοριών - παράδειγμα

$$S(a^p, b^q) = w_w S_w(a^p, b^q) + w_u S_u(a^p, b^q) + w_n S_n(a^p, b^q)$$

$S(\text{stream}^A, \text{stream}^B) = 0.615$  (stream<sup>A</sup> is 61.5% similar to stream<sup>B</sup>)

$S(\text{canal}^A, \text{canal}^B) = 0.693$  (canal<sup>A</sup> is 69.3% similar to canal<sup>B</sup>)

$S(\text{lake}^A, \text{lake}^B) = 0.77$  (lake<sup>A</sup> is 77% similar to lake<sup>B</sup>)

$S(\text{water}^A, \text{water}^B) = 0.85$  (water<sup>A</sup> is 85% similar to water<sup>B</sup>)

# Εντοπισμός Σηματολογικών Ετερογενειών – Σύγκριση Κατηγοριών

Ισοδυναμία



**ΣΓΠ Α**

*Ποτάμια: φυσικές ροές υδάτων*

**ΣΓΠ Β**

*Ποτάμια: φυσικές ροές υδάτων*

Διαφορά



**ΣΓΠ Α**

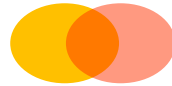
*Λίμνες: φυσικές ζώνες υδάτων που περιβάλλονται από ξηρά*

**ΣΓΠ Β**

*Θάλασσα: μεγάλη επιφάνεια αλμυρού νερού που περιβάλλεται μερικώς από ξηρά*

# Εντοπισμός Σηματολογικών Ετερογενειών – Σύγκριση Κατηγοριών

Επικάλυψη



**ΣΓΠ Α**

**Κανάλια:** *τεχνητές* ροές υδάτων που χρησιμοποιούνται για *μεταφορές και άρδευση*

**ΣΓΠ Β**

**Κανάλια:** *τεχνητές ή βελτιωμένες φυσικές* ροές υδάτων που χρησιμοποιούνται για *άρδευση*

Συμπερίληψη



**ΣΓΠ Α**

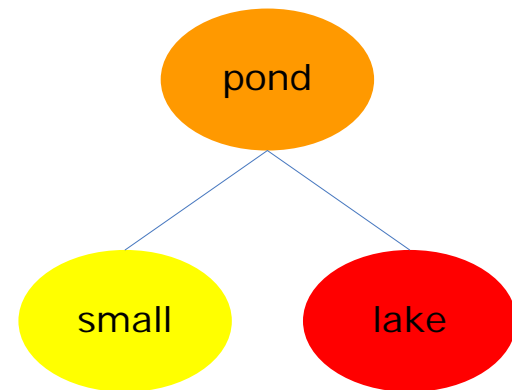
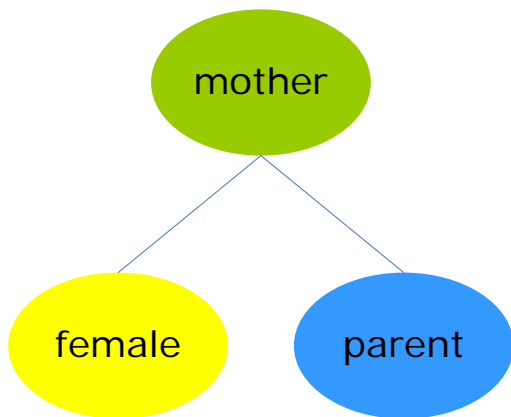
**Λίμνες:** *φυσικές ή τεχνητές* ζώνες υδάτων που περιβάλλονται από ξηρά

**ΣΓΠ Β**

**Λίμνες:** *φυσικές* ζώνες υδάτων που περιβάλλονται από ξηρά

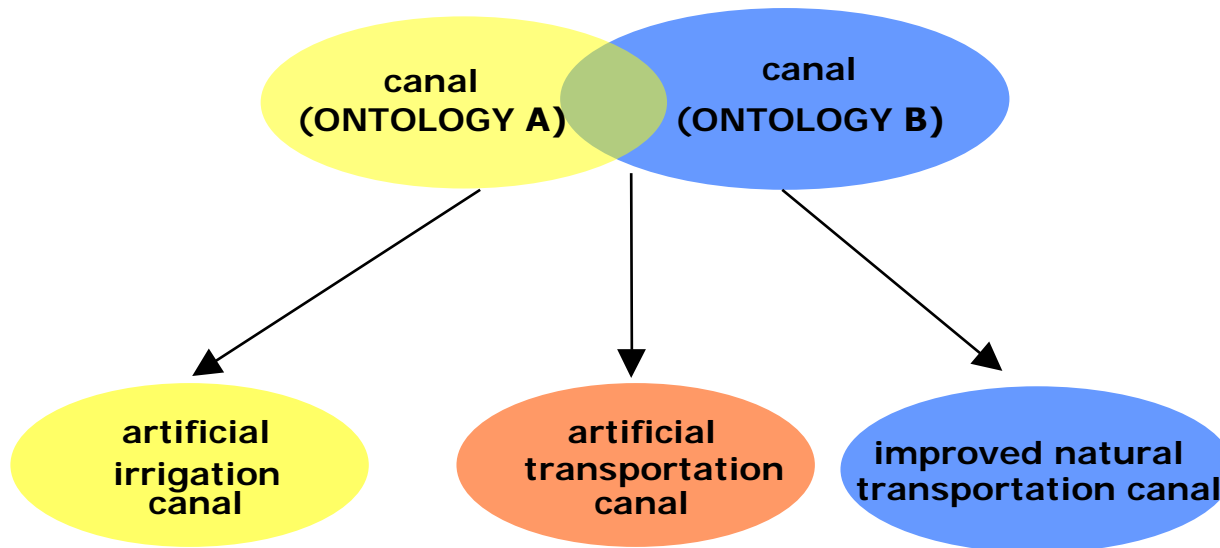
# Σημασιολογική Παραγοντοποίηση (Semantic Factoring)

- διαδικασία εννοιολογικής ανάλυσης
- αποσυνθέτει μια πολύπλοκη έννοια στις συνιστώσες έννοιες από τις οποίες ορίζεται (στοιχειώδεις έννοιες ή στοιχεία), που ονομάζονται **σημασιολογικοί παράγοντες**.



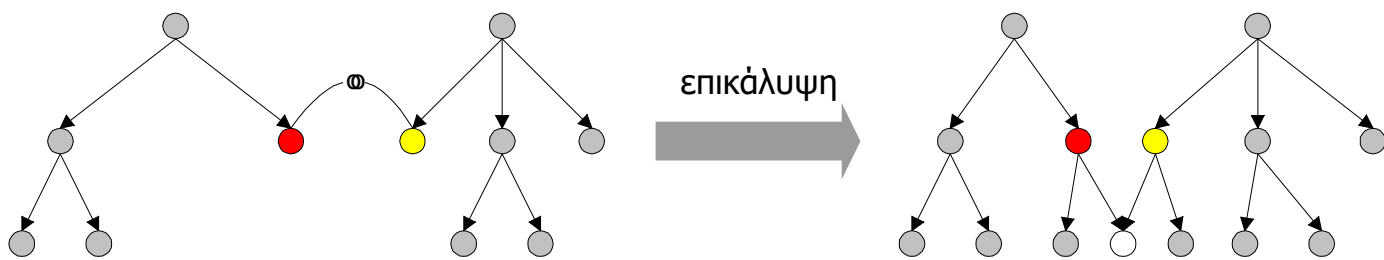
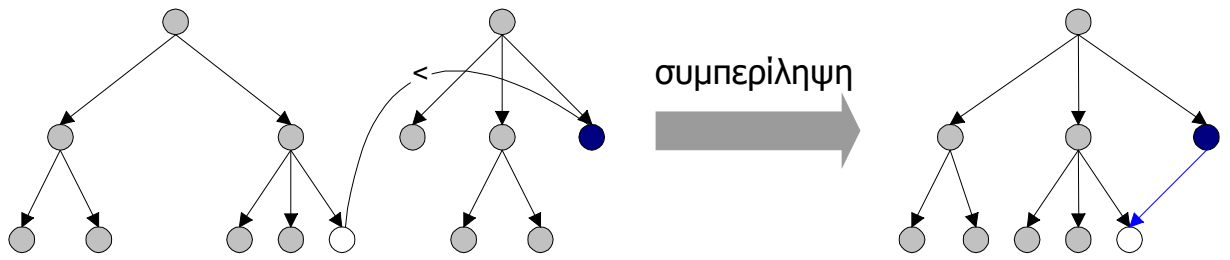
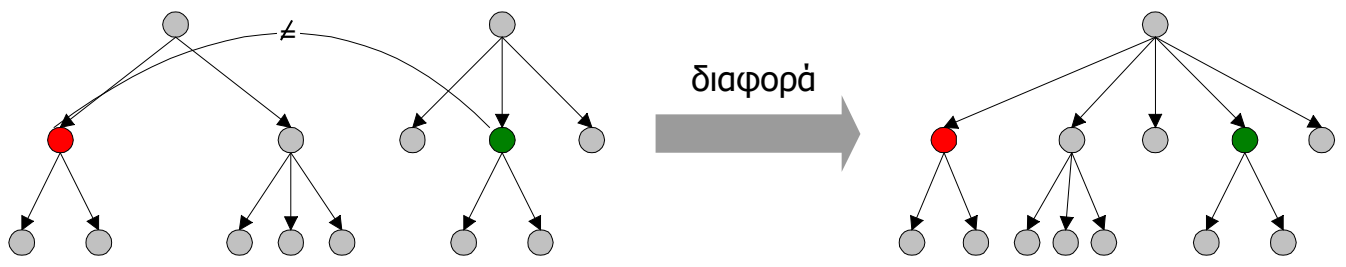
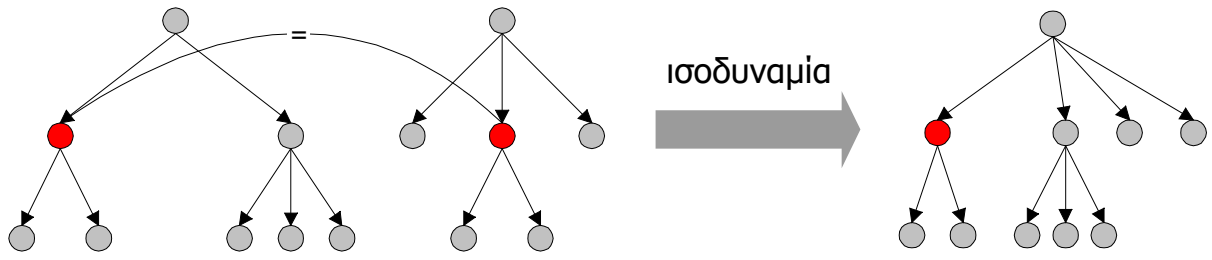
# Παράδειγμα: Σημασιολογική Παραγοντοποίηση Επικαλυπτόμενων Κατηγοριών

TERM	IS-A	COVER	PURPOSE	NATURE
<b>canal</b> (Ontology A)	way	water	transportation or irrigation	artificial
<b>canal</b> (Ontology B)	way	water	transportation	artificial or improved natural





# Εντοπισμός και Επίλυση Ετερογενειών



# Διαδικασία 3: Ολοκλήρωση



# Διαδικασία 3: Μορφές Ολοκλήρωσης (1) (Sowa, 2000)

- **Ευθυγράμμιση (alignment):** διαδικασία απλής αντιστοίχισης κατηγοριών διαφορετικών οντολογιών, συνήθως με χρήση εργαλείων μετάφρασης / μετατροπής. Δεν προκαλεί αλλοίωση.
- **Μερική συμβατότητα (partial compatibility):** διαδικασία ενοποίησης των όμοιων τμημάτων των αρχικών οντολογιών. Τα υπόλοιπα τμήματα των οντολογιών διατηρούνται ως έχουν. Προκαλεί αλλοίωση στα κοινά τμήματα.

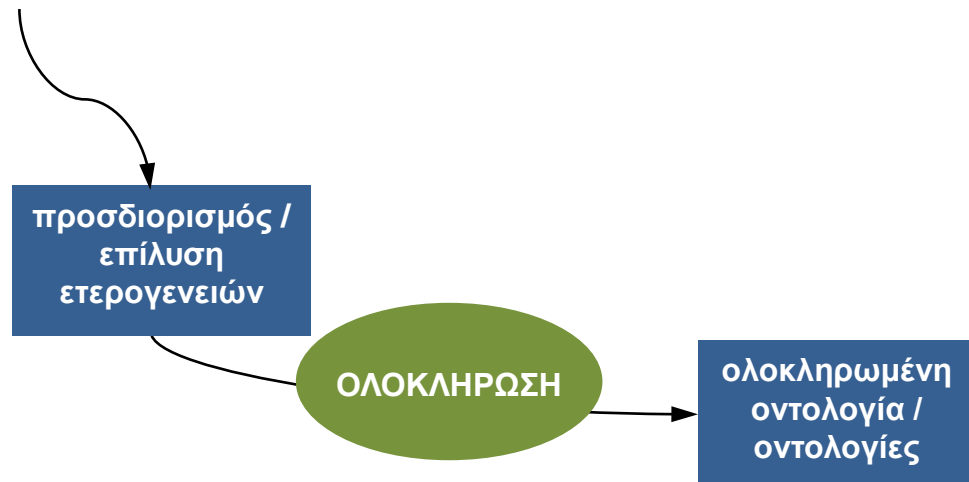
## Διαδικασία 3: Μορφές Ολοκλήρωσης (2) (Sowa, 2000)

- **Ενοποίηση (unification):** διαδικασία ενοποίησης όλων των κατηγοριών των αρχικών οντολογιών οι οποίες παραρποιοούνται ώστε να είναι μεταξύ τους συμβατές. Το αποτέλεσμα είναι μία τελική οντολογία. Προκαλεί αλλοίωση στις αρχικές οντολογίες.
- **Πραγματική ολοκλήρωση (true integration):** διαδικασία δημιουργίας μίας μοναδικής οντολογίας, η οποία αποτελείται από τις αρχικές κατηγορίες και κάποιες επιπρόσθετες απαραίτητες για τη συσχέτιση. Δεν προκαλεί αλλοίωση στις αρχικές οντολογίες.

# Διαδικασία 3η: Ολοκλήρωση

- INPUT: σημασιολογικοί παράγοντες, σημασιολογικά στοιχεία και σχέσεις μεταξύ των αρχικών κατηγοριών που κληρονομούνται στους σημασιολογικούς παράγοντες
- ΜΕΘΟΔΟΣ: Formal Concept Analysis (FCA) (Wille, 1992, Ganter and Wille, 1999)

- OUTPUT: μία τελική ολοκληρωμένη οντολογία

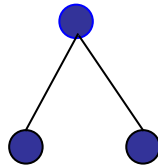


# Δομές για την Αναπαράσταση Γεωγραφικών Κατηγοριών

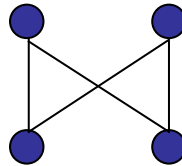
Hirtle (1995):

- δέντρα (trees)
- διατεταγμένα δέντρα (ordered trees)
- δικτυωτά (lattices)

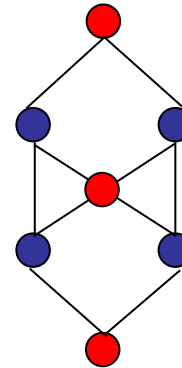
τα δικτυωτά είναι πιο ισχυρές και ευέλικτες δομές για την αναπαράσταση των πολλαπλών, επικαλυπτόμενων σχέσεων μεταξύ των γεωγραφικών εννοιών



**tree**

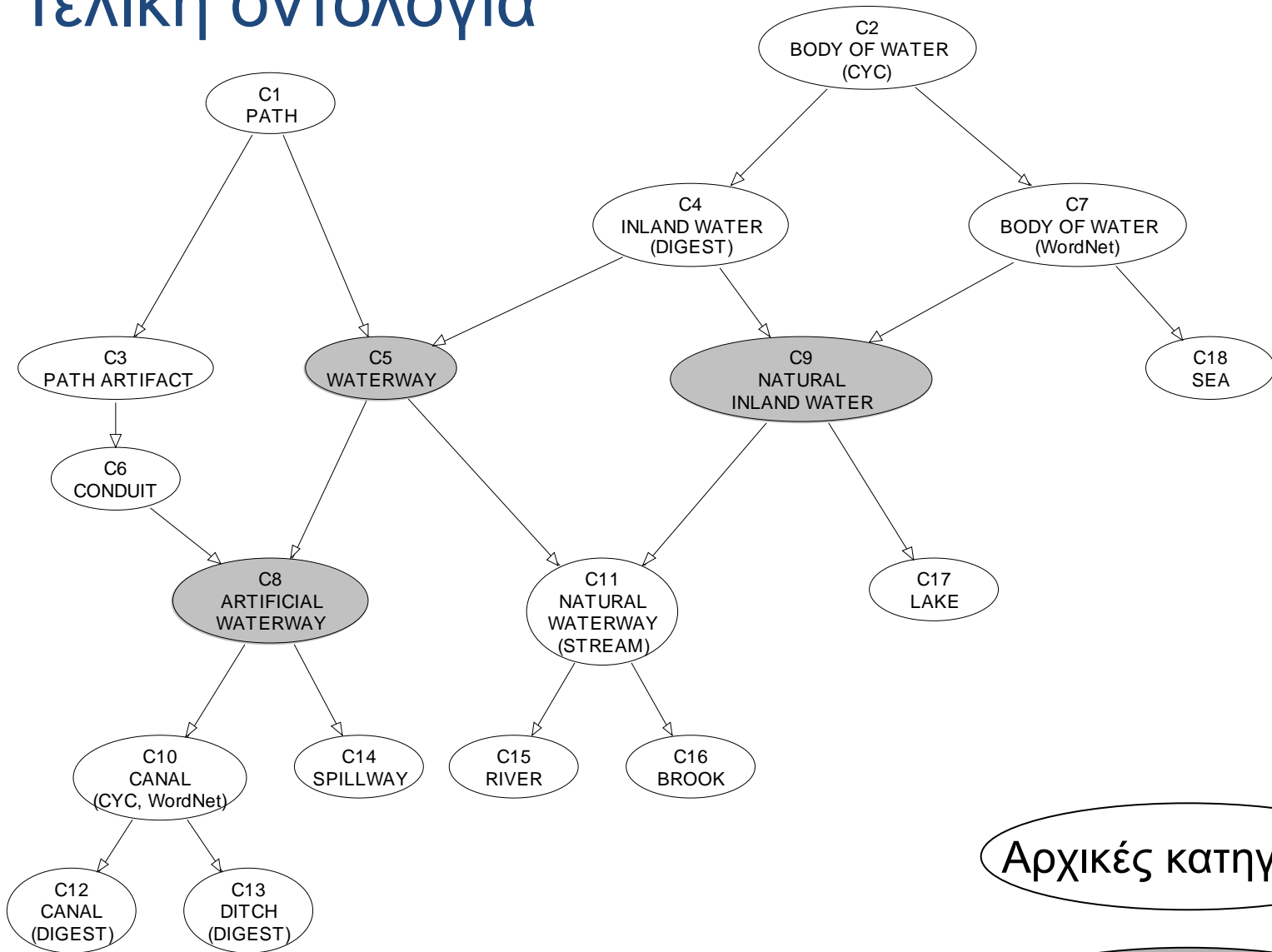


**Ordered tree**



**lattice**

# Τελική οντολογία



Αρχικές κατηγορίες

Νέες κατηγορίες

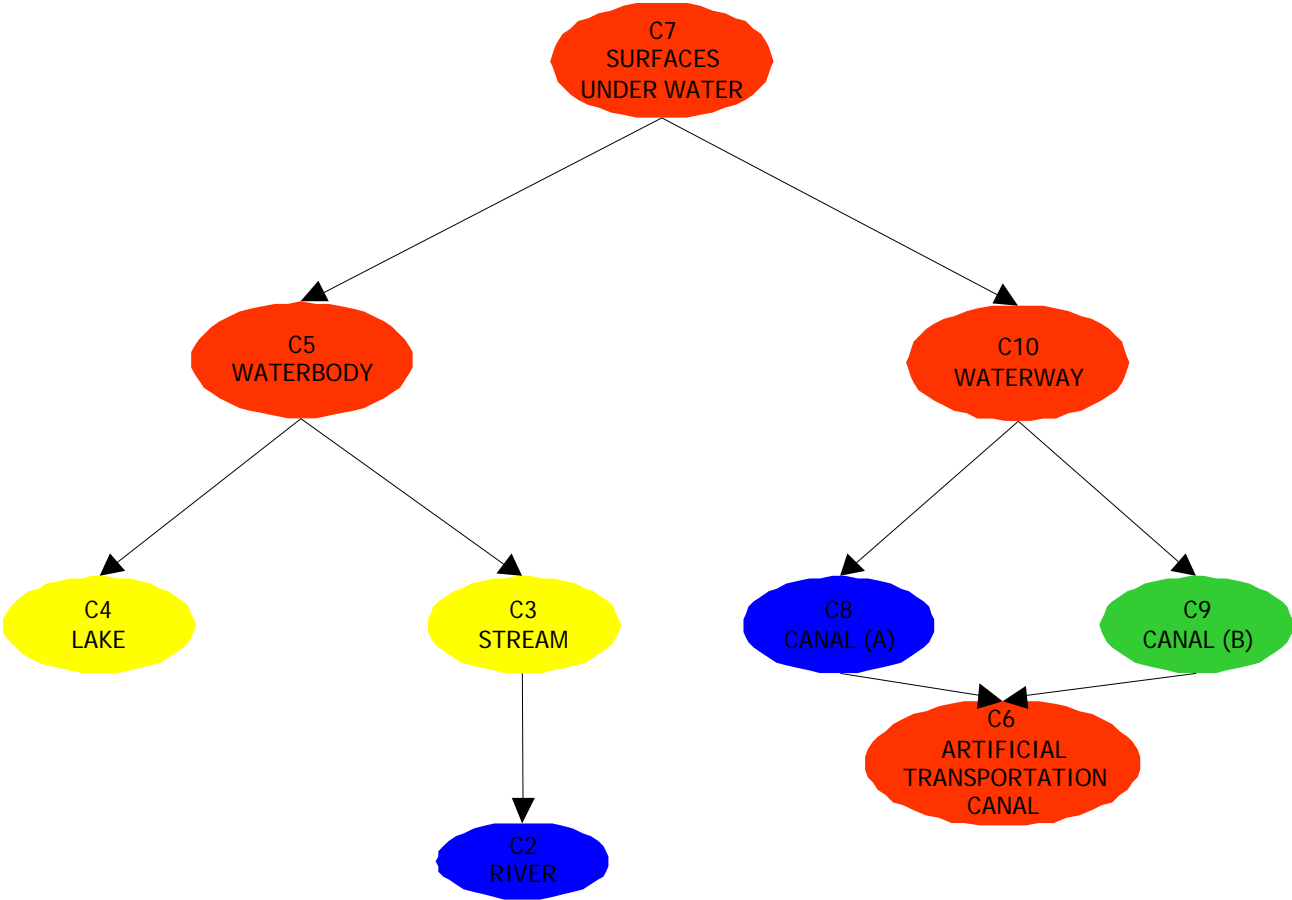
# Δημιουργία της Ολοκληρωμένης Οντολογίας

Ontology A

Ontology B

Κοινές κατηγορίες

Νέες κατηγορίες





# Ολοκλήρωση οντολογιών

