Ontologies and Semantic Technologies

Bergmann Gábor

Budapesti Műszaki és Gazdaságtudományi Egyetem
Méréstechnika és Információs Rendszerek Tanszék

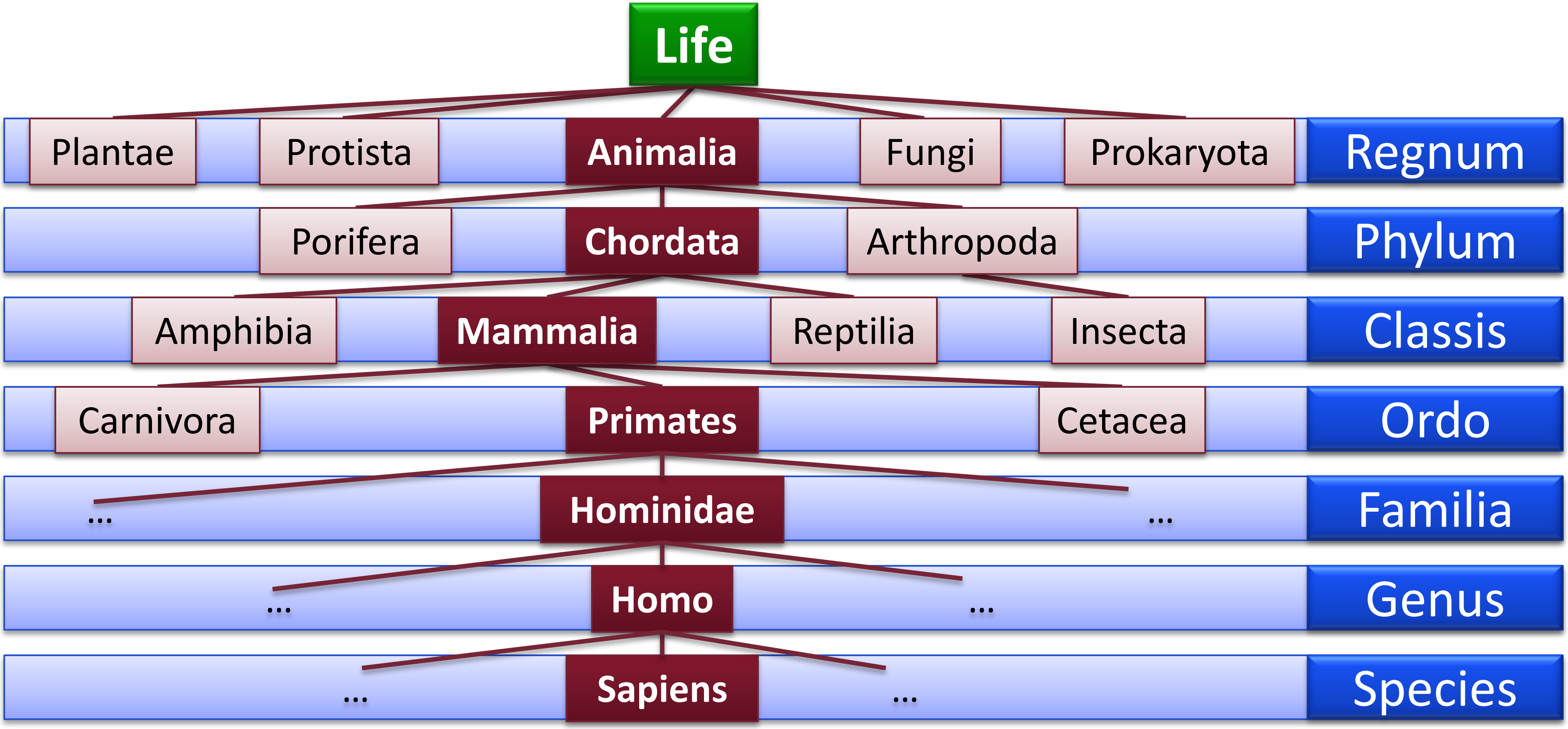
Agenda

- Ontologies
- RDF and Semantic Web
- Semantic Technologies and Resources
- Semantic Integration

ONTOLOGIES

Taxonomy

- **Taxonomy** = hierarchy of domain concepts



- Ontology ≅ taxonomy + relationships + definitions

Ontology

- **Ontology** = „the study of existence”
- Computer representation of **domain knowledge**
  - Identifying **concepts** to categorize **individuals**
  - **Relationships** that can hold between individuals
  - **Axioms** on concepts and their relationships
    - Including **taxonomy** of domain concepts (supertypes)
- Created by
  - **domain experts**
  - **knowledge engineers**
- Intended for automatic processing, reasoning

Domain Ontologies

- Domain Ontologies for **expert systems**
  - Reasoning based on axioms and formal logic
  - Ontology-based search
- Sample Ontologies
  - Open Biological and Biomedical Ontologies (OBO)
    - Chemical information
    - Cells, cell types, proteins, etc.
    - Anatomy (Upper/Human/…)
    - General Medical Science, disease ontologies
    - Medical software, imaging methods, spectrometry etc.
  - Food Ontologies, Wines, Air Traffic, Legal, etc.

Example Ontology

- Concepts/classes: Animal, Person, Male, Female, Man, Woman, SingleChild, etc.
- Attributes: name, weight, etc.
- Relationships: mate, child, parent, etc.

No strict distinction

- Axioms and definitions
  - $child = parent^{-1}$, $mate = mate^{-1}$
  - $Person \subseteq Animal$
  - $Male \cap Female = \emptyset$, $Male \cup Female = Animal$
  - $Woman = Female \cap Person$
  - $SingleChild = Animal \cap \forall parent.\exists_{=1} child$

Formal Background

- Description Logics (**DL**)
  - Individuals, concepts, roles (properties)
  - Axioms ≈ set operations and role tree navigation
    - A-Box: axioms about individuals
    - T-Box: axioms about concepts and roles
  - Reasoning with tableau calculi
- Dialects: $\mathcal{SHOIQ}(\mathcal{D})$, $\mathcal{SHIQ}$, $\mathcal{SHIN}$, $\mathcal{ALCN}$, etc.
  - Varying expressive power
  - Usually weaker than FOL (first-order logic)
  - Expressivity vs. reasoning complexity

Open World Semantics

- Can we enumerate all diseases?
  - Traditional databases have Closed World Semantics
    - E.g. if not explicitly listed as a disease, then not a disease
- Most ontologies: **Open World** Semantics
  - Not proven true/false → not treated as false or true
  - Why? Ontologies can never be complete
  - Examples
    - E.g. if not listed / implied as a viral disease, still can be one
    - Patient 42 has lepers. Does Patient 42 have a flu? Unknown!
    - Patient 2501 died of lepers. Did she die of flu? No!
      (by multiplicity 1 of cause of death)

Unique Name Assumption

- Can two identifiers correspond to the same thing?
  - Patient 42 carries a skin disease.
  - A disease of patient 42 was found to be of viral nature.
  - Are they two different diseases?
- Usually NO **Unique Name** Assumption
- Two things can be the same, unless contradicted
  - disjoint classes (Patient 2501 has a hereditary disease)
  - explicit control: owl:sameAs, owl:differentFrom
- Why? Distributed knowledge gathering

OWL

- W3C: Web Ontology Language (**OWL**)
  - **owl:Class** (≈ concept) = set of individuals
  - **rdf:Property** (≈ role) = link to data or other individuals
  - Individual (≈ A-box), nominal ($\mathcal{O}$ in $\mathcal{SHOIQ}$), enum
  - Datatype, **axioms**, etc.
  - Uses URIs, builds on RDF+XML syntax (see later)
- Subsets for scalability
  - DL-compatible subset: OWL DL = $\mathcal{SHOIN}(\mathcal{D})$
  - OWL Full is stronger, has multi-level metamodeling
  - OWL 2 has further subsets (profiles)

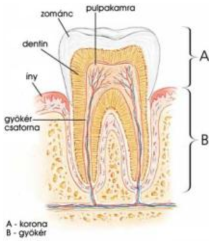
RDF AND SEMANTIC WEB

Metadata

- **Metadata**: description of data,
  - For people
  - For machines
- Example: image metadata
  - Generated partly automatically
  - „on this picture: John Doe, Jean-Baptiste Grenouille”
- Example: text document metadata
  - Author, literary category, year of publishing, etc.
- Metadata-based search

Syntactic Interpretation

- Can machines understand what we mean?
  - Textual / syntactic services can not
- Example: show me pictures depicting „fog”!



- Example: show me poems by female authors!
- Semantic solution
  - Machines should process the meaning, not the form
  - Use standardized concepts „fog”, „female”, „author”…
    - Refer to it in metadata and queries

Resource Description Framework

- W3C: Resource Description Framework (**RDF**)
  - **rdf:Resource** → something we talk about
    - a document (e.g. this photo)
    - a standardized meaning (e.g. tooth, Hungary)
    - identified by a URI
  - **rdf:Property** → relation type between resources
    - e.g depicts, taken_in etc.
    - also identified by a URI
  - Triplets → statements about properties of resources
- Open world, no unique names
- RDFS: RDF Schema

RDF Statements

- RDF statement = triplet
  - (resource, property, value)
  - resource, property are URIs
  - value: URI of other resource or raw data
- Example triplets
  - (this_photo, taken_in, Hungary)
  - (this_photo, file_name, „DSC00001.JPG”)
  - (this_photo, depicts, John Doe)
  - (this_photo, rdf:type, Photo)
  - (rdf:type, rdf:type, rdf:Property)

RDF Concrete Syntax

- Concrete syntaxes: **RDF+XML**, RDFa, N3, etc.

```
<rdf:RDF xmlns:… >
 <foaf:Person rdf:ID="#me">
    <foaf:name>Bergmann Gábor</f
    <foaf:gender>male</foaf:gender>
    <foaf:img rdf:resource="http://…/…jpg"/>
    <foaf:holdsAccount> <foaf:Onlin
      <foaf:accountServiceHomepage
"http://www.facebook.com/"/>
      <foaf:accountName>…</foaf:
    </foaf:OnlineAccount> </foa
    <foaf:knows>
      <foaf:Person rdf:about="#662…">
        <foaf:name>Zoltán Szatmári</foaf:name>
        <foaf:holdsAccount>…</foaf:holdsAccount>
        <foaf:based_near><geo:Feature>
          <geo:name>Budapest, Hungary</geo:name>
        </geo:Feature></foaf:based_near>
      </foaf:Person>
    …
</rdf:RDF>
```

(«base»:#me, rdf:type, foaf:Person)

(«base»:#me, foaf:gender, „male”)

(«base»:#me, foaf:img, «URL»)

(«base»:#me, foaf:knows, «Zee»)
(«Zee», rdf:type, foaf:Person)

(«Zee», foaf:based_near, «nameless»)
(«nameless», rdf:type, geo:Feature)
(«nameless», geo:name, „Budapest…”)

RDF Application

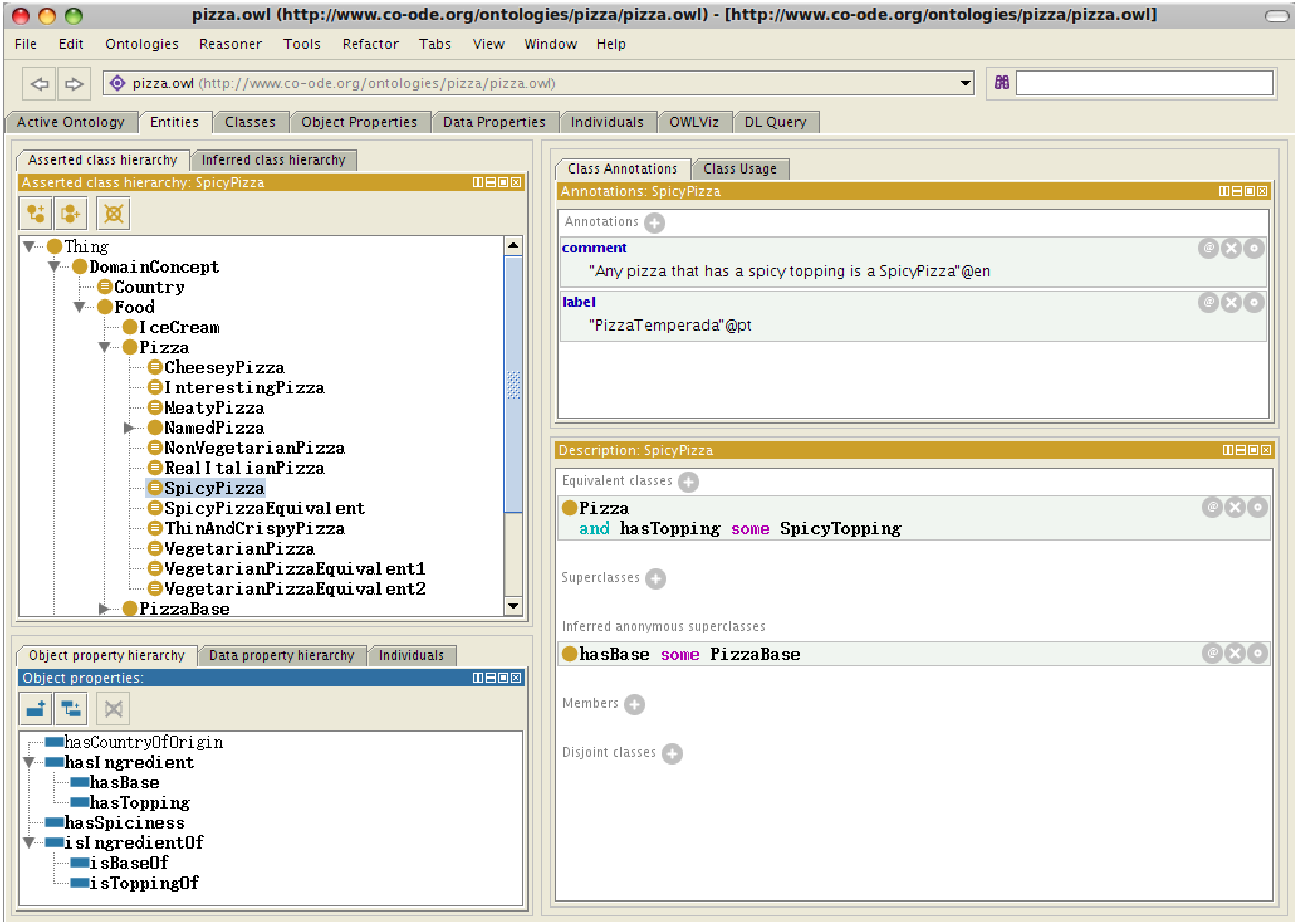
- RDF Site Summary (**RSS**)
  - Items with title, description, link, creator, date, …
  - RSS 2.0 abandons RDF, backronym
- OWL itself is an RDF document
  - Classes, properties identifiable by URIs
- **Semantic Web**
  - Is a photo of my Porsche a photo of a car?
  - Need standard URIs for RDF resource/property types
  - Use OWL ontologies to provide type URIs
  - Local metadata + ontologies = semantic web

SEMANTIC TECHNOLOGIES AND RESOURCES

Semantic Web Technologies

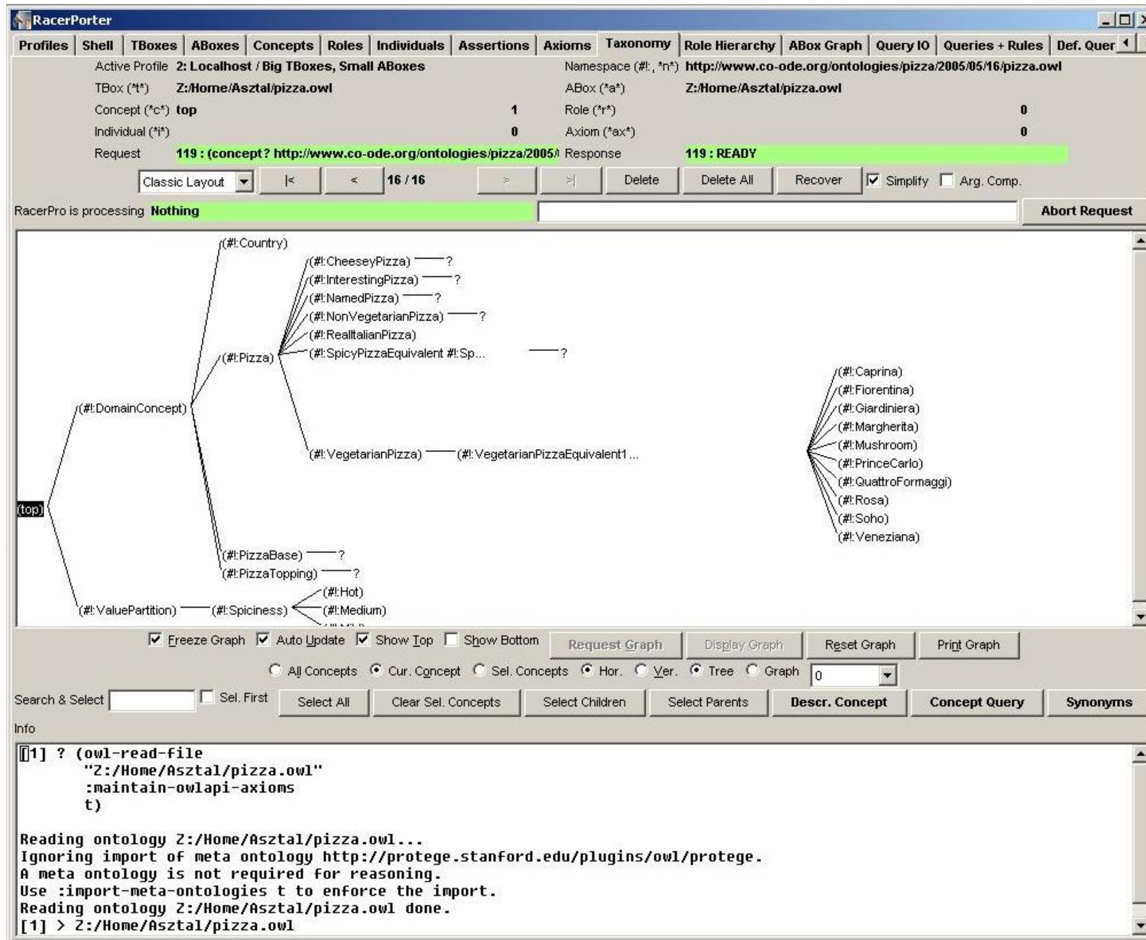
- Swoogle
  - Indexes RDF, OWL files
  - PageRank-like ranking
  - Descriptions collected from referencing documents
- SPARQL - RDF query language
  - SQL-like syntax
- Programming frameworks (e.g. Jena, Sesame)
  - APIs for RDF and OWL
  - In-memory and persistent storage + remote access
  - Manipulation, query and inference

Protégé Ontology Editor



RACER reasoner

- RacerPro: DL reasoner engine
- RacerPorter GUI



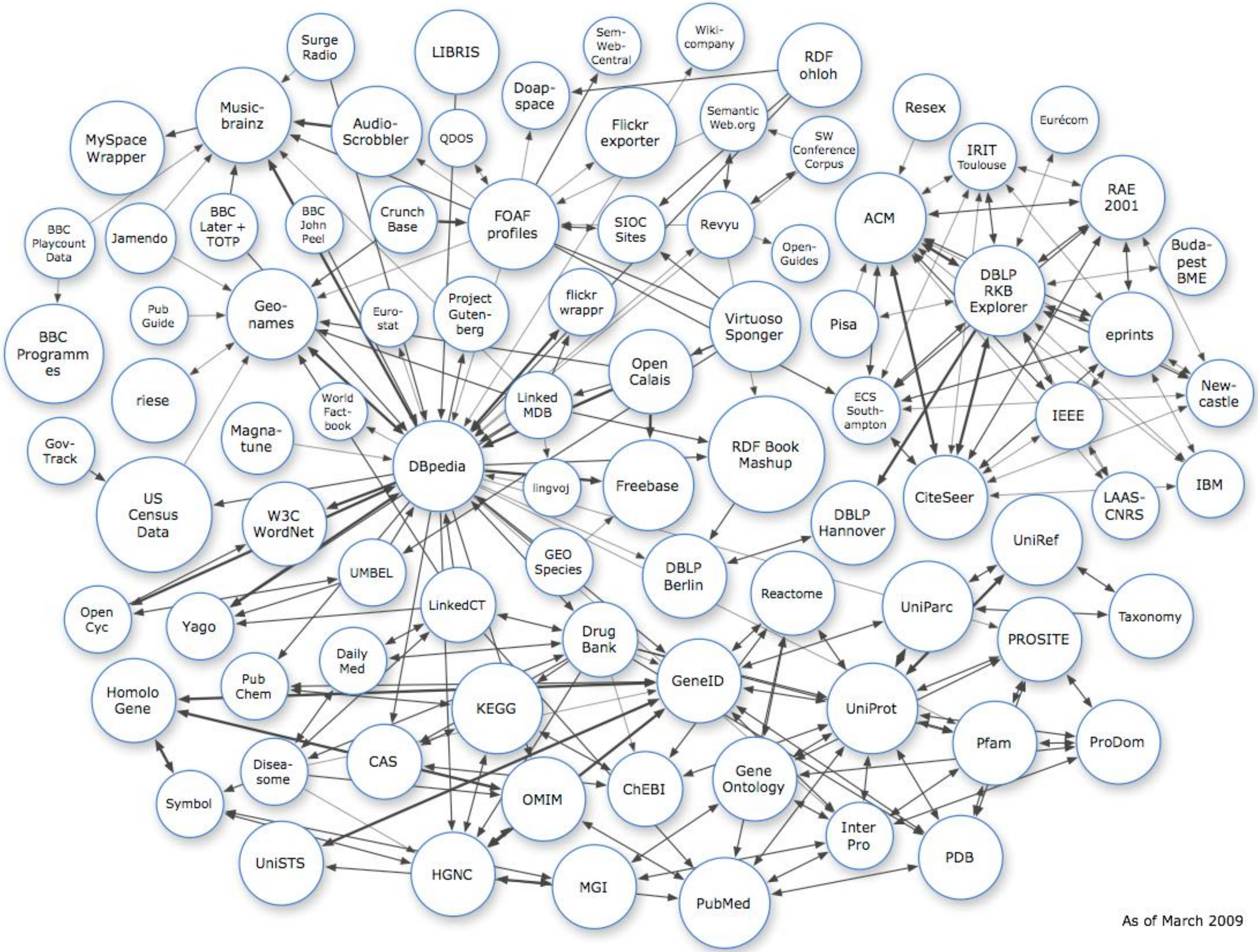
Semantic Web Ontologies

- Friend-of-a-Friend (FOAF) ontology
  - Classes: Person, Image, Document, OnlineAccount, etc.
  - Attributes: surname, birthday, title, etc.
  - Relationships: knows, made, depicts, weblog, topic, logo, openID, page, interest, etc.
  - Used / usable in Facebook, flickr, LauchPad…
- Dublin Core (DC) ontology
  - Librarian metadata for documents
  - Title, publisher, language, format, date, creator, etc.
  - Widespread usage (e.g. as an RSS Module)

Encyclopedic Knowledge Bases

- WordNet: lexical knowledge base of english words
  - Synonym sets (synsets)
  - Semantic relations, including
    - Antonym (opposite)
    - Hypernym, hyponym (super/subtype)
- DBpedia Knowledge Base
  - RDF information
  - Automatically extracted from Wikipedia
  - Manual annotation, links to WordNet etc.
  - SPARQL interface

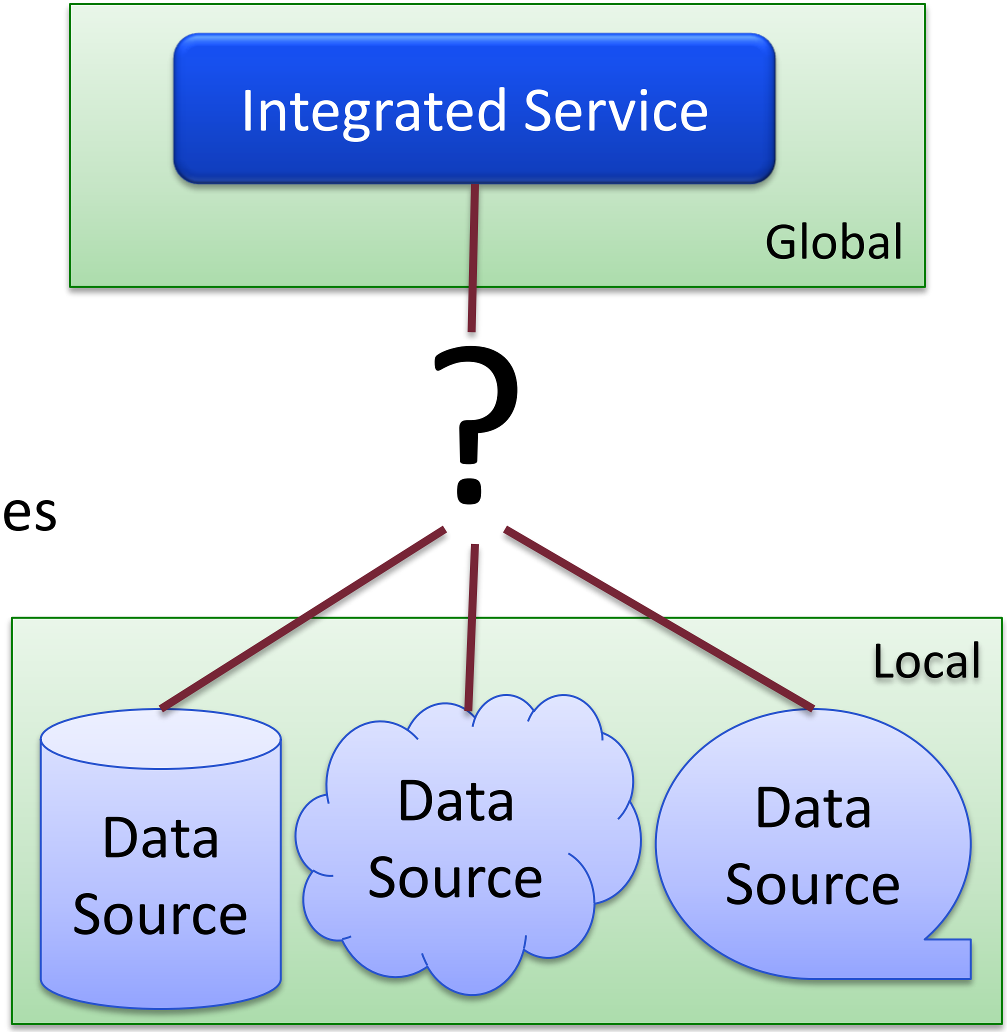
Linking Open Data (LOD)



SEMANTIC INTEGRATION

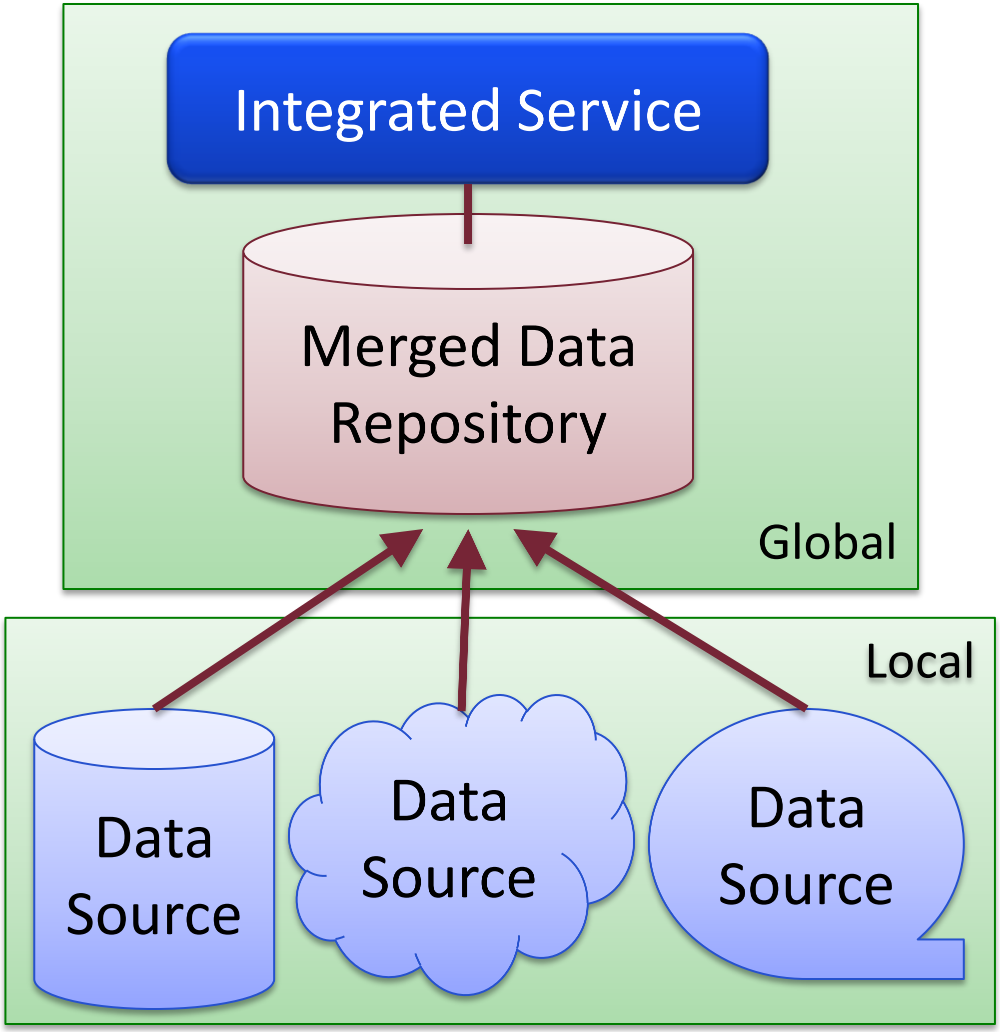
Data Integration

- Distributed, heterogenous sources
  - Relational DB
  - Web Service
  - Ad-hoc interfaces
- Unified global service
  - Usually query-only
  - Global query vs. local sources
- Terminology
  - Data Integration
  - Information Integration
  - Data Fusion



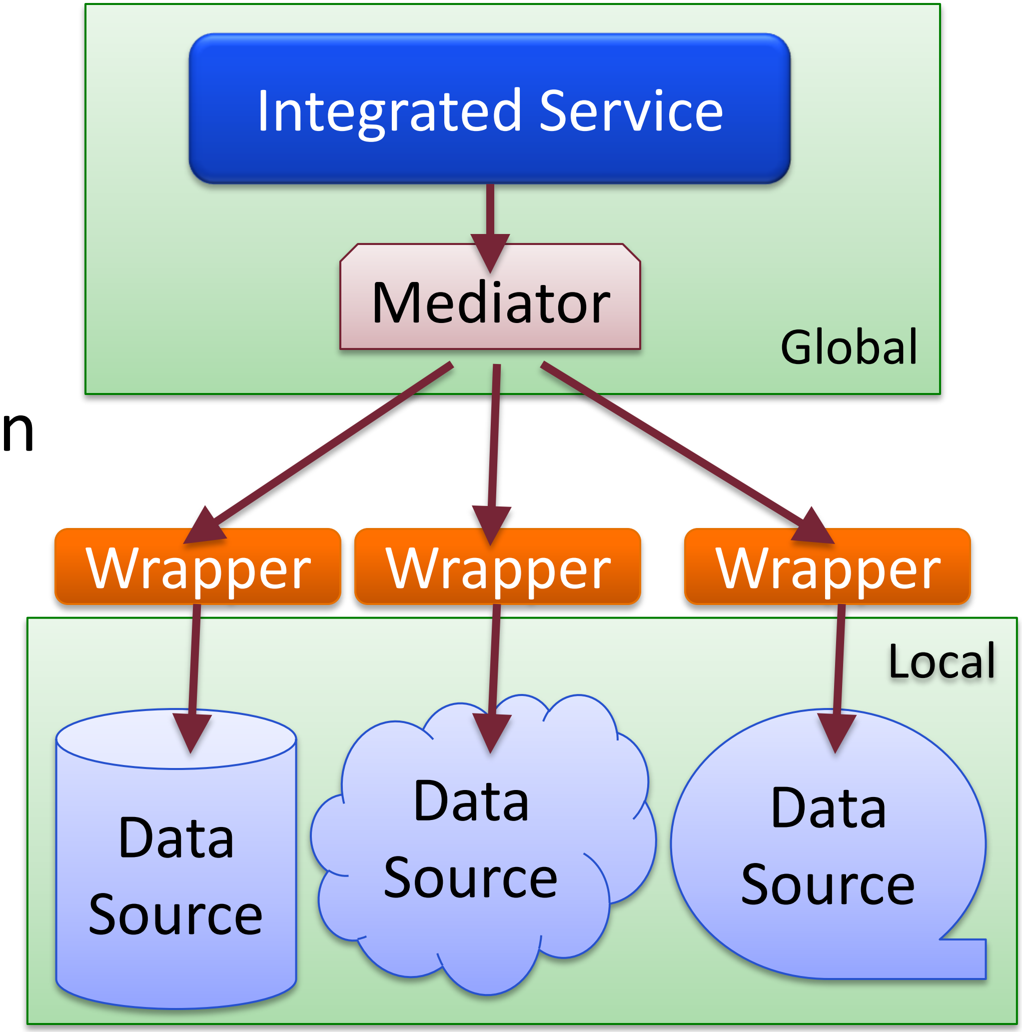
Merging (Data Warehousing)

- Solutions
  - **Merging (warehousing)**
  - Mediation (federated DB)
- Single central repository
  - Contains all merged data
- Straightforward
- Drawbacks
  - Merge cost
  - Outdated
  - Unless regularly refreshed
  - Maintainability issues



Mediation (Database Federation)

- Solutions
  - Merging (warehousing)
  - **Mediation (federated DB)**
- Query propagation
  - Result composition
- Problem: query translation
  - GaV
  - LaV
- Advantages
  - Always up-to-date
  - Lightweight



Mediation: Global as View

- Problem: query translation
  - **GaV (Global as View)**
  - LaV
- The global schema expressed as a view on locals
  - Transformations, projections, unions, joins
  - E.g. ‘ACME’$\times$($\pi_{Name,Price}$AcmeProducts)$\cup$…
- Query execution: simple view evaluation
- The same basic design as in the merging case
  - Unmaintainable with too many sources

Mediation: Local as View

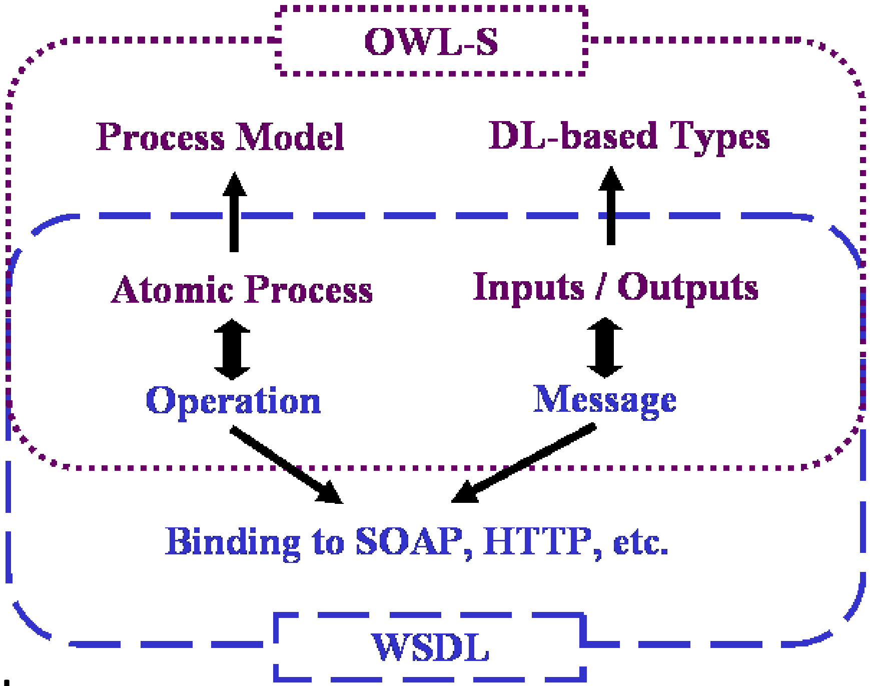
- Problem: query translation
  - GaV
  - **LaV (Local as View)**
- Each local schema as a view of the global
  - Projections, selections
  - E.g. only product data, only for ACME boxes
- Easier to add new sources
- Query execution
  - theory of federated databases
  - „answering queries using views”

Semantic Integration

- Heterogeneity
  - Non-relational sources
  - Different vocabulary and semantics
  - Different structure
  - Different representations
- Ontology-based Semantic Data Integration
  - Local scheme explained with linked ontologies
  - Semantic mapping between schemes
  - Query formulation based on ontology
  - Execution: automatic reasoning?

Semantic Service Integration

- Standards such as WSDL define the syntax
  - Currency? Unit of measurement?
  - The identifier of *what*?
  - Preconditions?
- Semantic Web Services
  - OWL-S
    - Process Model, DL types
  - WSMO
    - Goals of clients, mediators, etc.
- Dream: Semantic Service Discovery



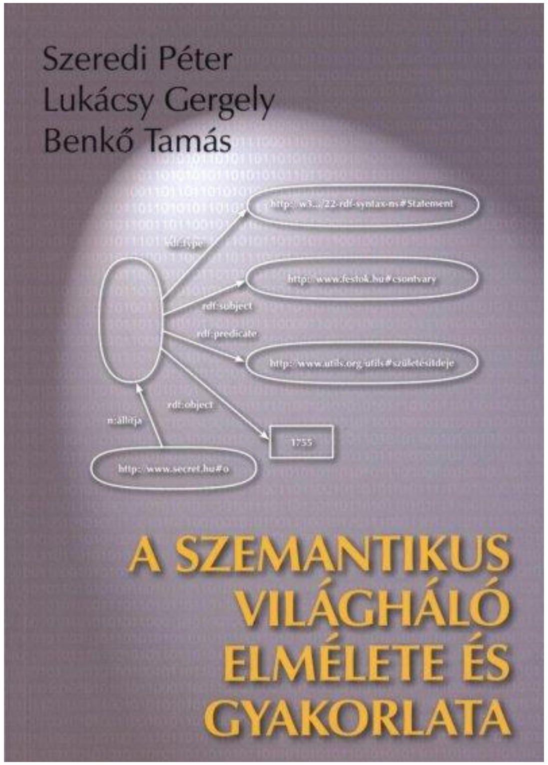
SUMMARY

Summary

- Semantic technologies
  - Metadata (RDF)
  - Ontologies (OWL)
  - Formal logic based reasoning
- Applications
  - Domain ontologies in expert systems
  - Semantic Web
  - Ontology-based semantic data integration
  - Semantic Service Discovery

Recommended reading

Benkő-Szeredi-Lukácsy:
A szemantikus világháló elmélete és gyakorlata.
Typotex, 2005.



BMEVIMIM222
Információ- és tudásintegrálás
(MSc intelligens rendszerek szakirány)