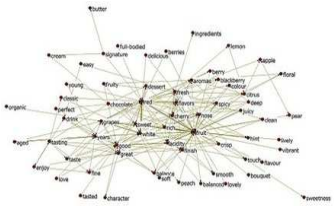
Establishing a framework for scholarly editing and publishing in the 21st century

Beirut, 9-10 March 2015

# SEMANTICS & DIGITAL DATA PROCESSING

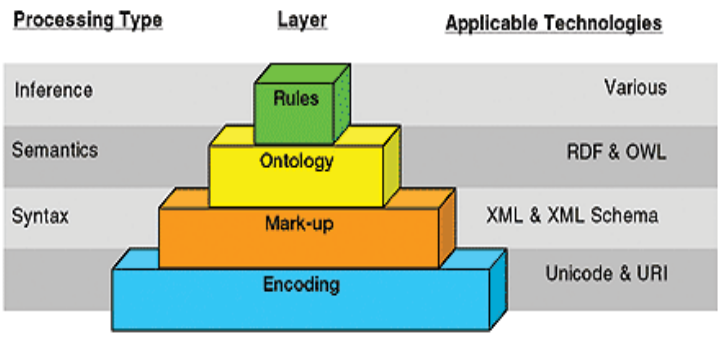
Mokhtar BEN HENDA





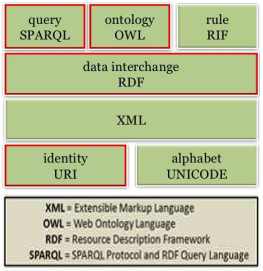
## Semantics: Better computational description of science

- Information is given explicit meaning so that machines can process it more intelligently;
- Instead of just creating standard terms for concepts as is done in XML, the Semantic Web also allows users to provide formal definitions for the standard terms they create so that machines can use inference algorithms to reason about the terms;
- A crucial component to the Semantic Web is the definition and use of ontologies,



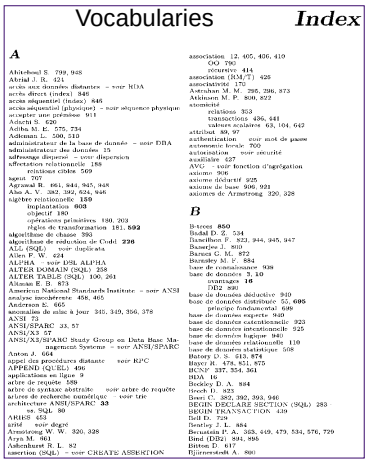
# Basics of semantic Web

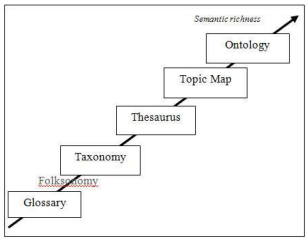
- XML
  - XML is a language for transmitting structured information. If the goal of the web is to enable not only communication between people, but also between machines, then XML seems a good basis not only for documents to be read by people, but for data to be read by machines.
- RDF
  - RDF(Resource Description Framework) is a directed, labeled graph data format for representing information in the Web. It is just a data model that does not have any significant semantics. RDF Schema is used to define a vocabulary for use in RDF models. In particular, it allows to define the classes used to type resources and to define the properties that resources can have.
- OWL
  - OWL was designed to provide a common way to process the content of web information instead of displaying it. It is primarily concerned with defining terminology that can be used in RDF documents. Syntactically, an OWL ontology is a valid RDF document and as such also a well-formed XML document.
- SPARQL
  - Is a set of specifications that provide languages and protocols to query and manipulate RDF graph content on the Web or in an RDF store

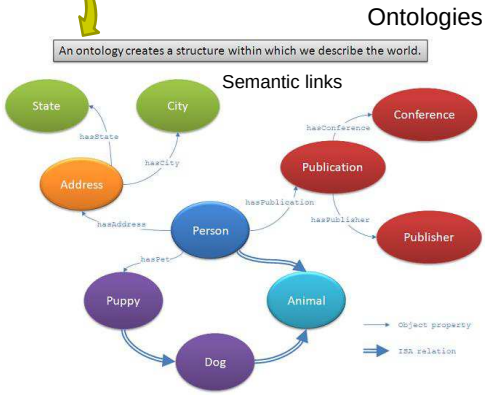



# Ontologies

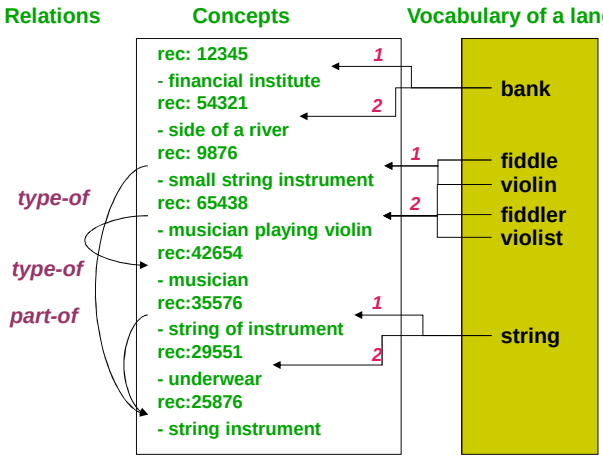
- Principle:
  - From characters string to word meaning
  - From words listing to words relations
  - From indexes to ontologies



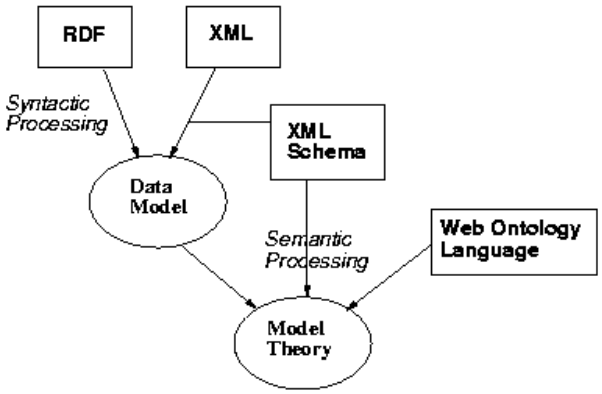



# Data Models

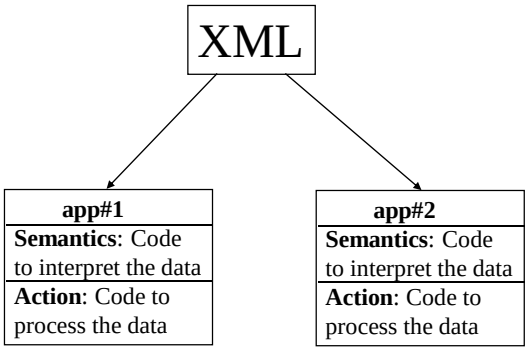





# Capturing Semantics in XML Documents



## Meaning (semantics) applied on a per-XML-application basis

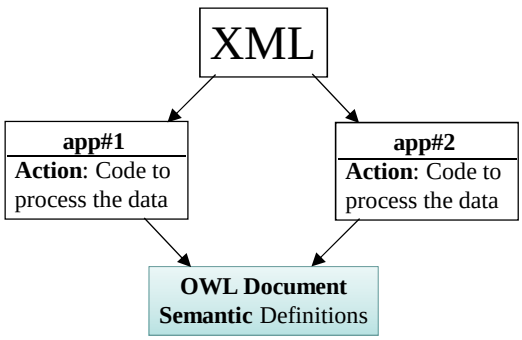







## OWL (Ontology Web Language)

- Handles separate concept definitions (semantics) from application
- Express concept definitions using a standard vocabulary



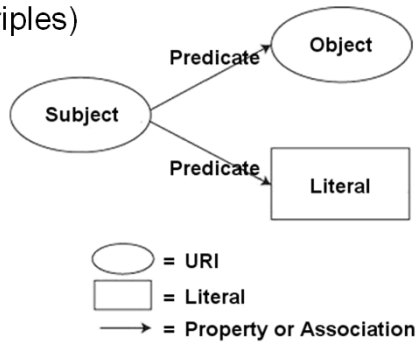
## OWL and logics

- OWL relies on Description Logics
- Logics provide automatic
  - Check of consistency of concept definitions
  - Completion of concept definitions
  - Classification of new instances and concepts
  - Extraction of implicit knowledge in the documents
- OWL greatly expands the vocabulary for multiple possible constructs
- XML Schema provides some of those properties to some extent




## RDF (Resource Description Framework)

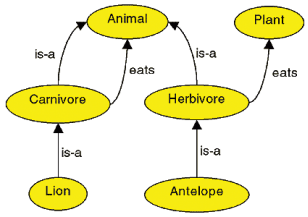
- Provide basic syntax for OWL
- Use of URI for unique identification of concepts, instances and relations
- Expression of relations between objects and concepts (RDF triples)

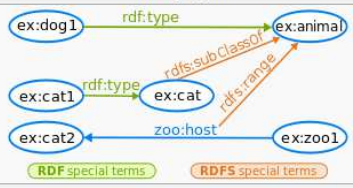



- Problem: no structure

# RDF Schema

- Add basic structure to RDF
  - Class/Subclass declaration
  - Instances
  - Properties (relations)
  - Multiple inheritance



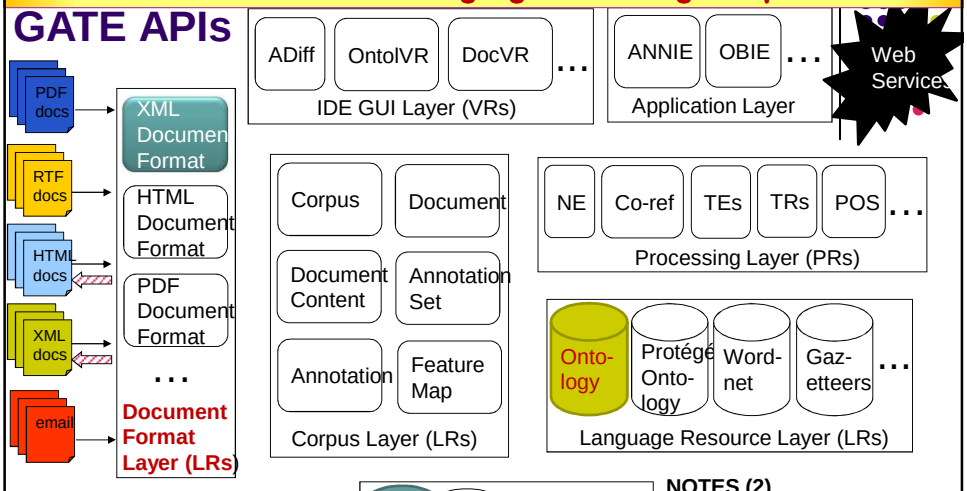

| In English | The graph |
|---|---|
| • Dog1 is an animal<br>• Cat1 is a cat<br>• Cats are animals<br>• Zoos host animals<br>• Zoo1 hosts the Cat2 |  |





## Sheffield Natural Language Processing Group

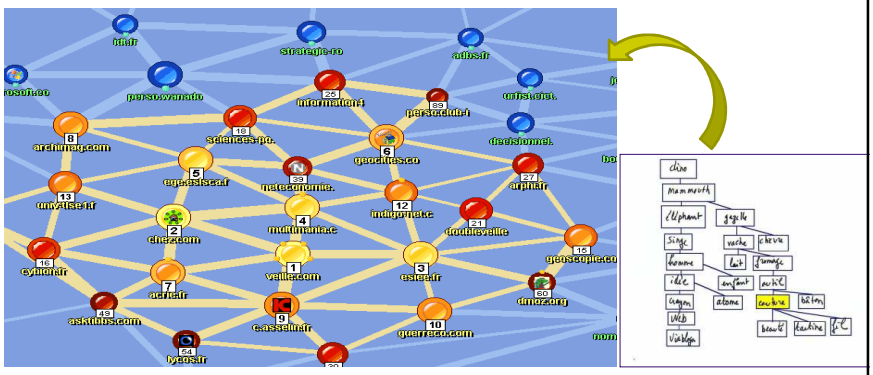
# GATE APIs




**NOTES**
- **everything is a replaceable bean**
- **all communication via fixed APIs**
- **low coupling, high modularity, high extensibility**

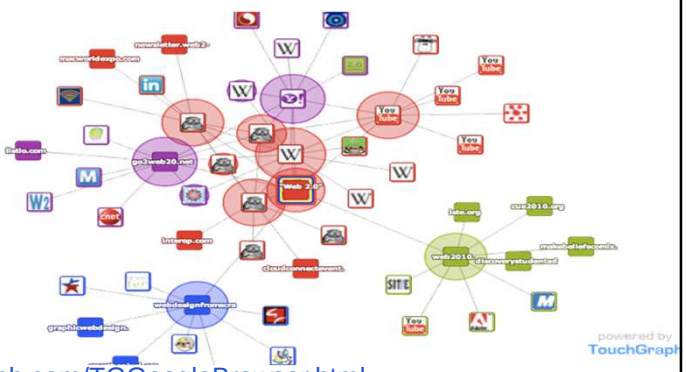
**NOTES (2)**
- **eg: Protégé LR & VR both wrapped in Res. (bean) API**
- **ontology repositories and inference are the same: KAON + Sesame + Orenge + ?**

# Semantic Web: cartographic searching



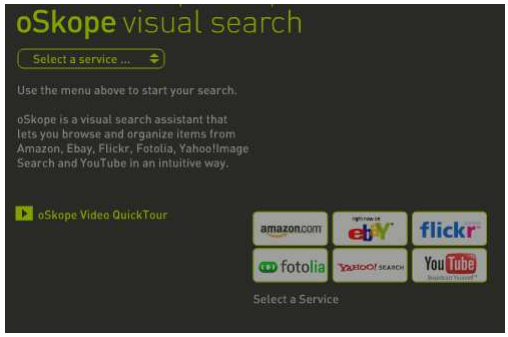
# Semantic Web: cartographic searching

- Conceptual charts
  - Reproduce semantic links like charts
  - One concept can have numerous semantic links



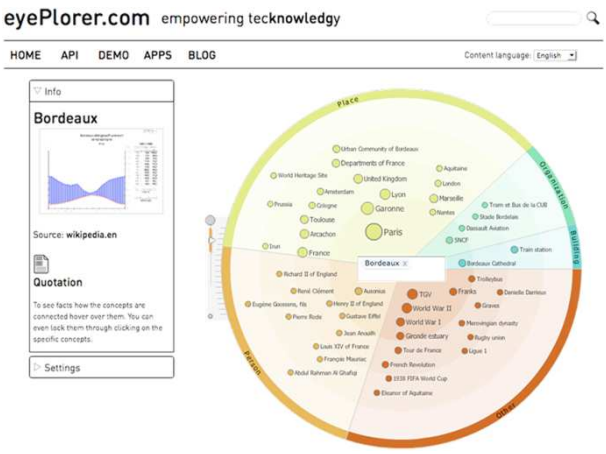
http://www.touchgraph.com/TGGoogleBrowser.html

# Semantic Web: cartographic searching



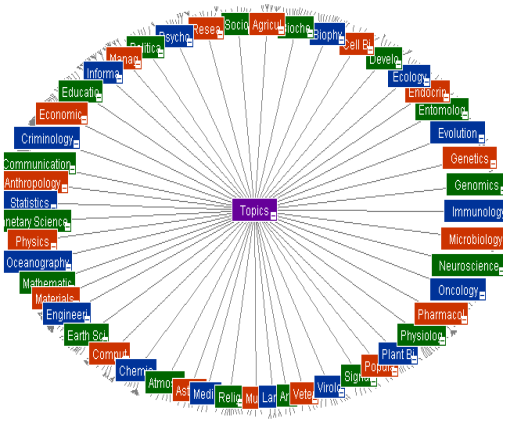

http://www.oskope.com

# Semantic Web: cartographic searching




- http://vionto.com/show/

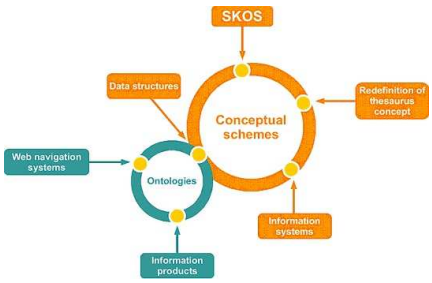
## Semantic Web: cartographic searching




Topic Maps : http://highwire.stanford.edu/lists/artbytopic.dtl

## SKOS: Simple Knowledge Org. Systems

- SKOS: specifications and standards to support within the framework of the Semantic Web, the use of knowledge organization systems (KOS) such as:
  - thesauri,
  - classification schemes,
  - subject heading lists
  - Taxonomies



http://www.ebusiness-unibw.org/tools/skos2owl/

# SKOS Play

- SKOS Play thesaurus is a visualization service of SKOS formatted taxonomies or vocabularies.
- More generally, it is used to view or print a knowledge organization system expressed in SKOS.


http://labs.sparna.fr/skos-play/upload?lang=en