Chapter 7

Ontology Engineering

The Representation of Knowledge

- knowledge **has many meanings**.
- *Data*, *facts*, *information* are often used to indicate *knowledge*.

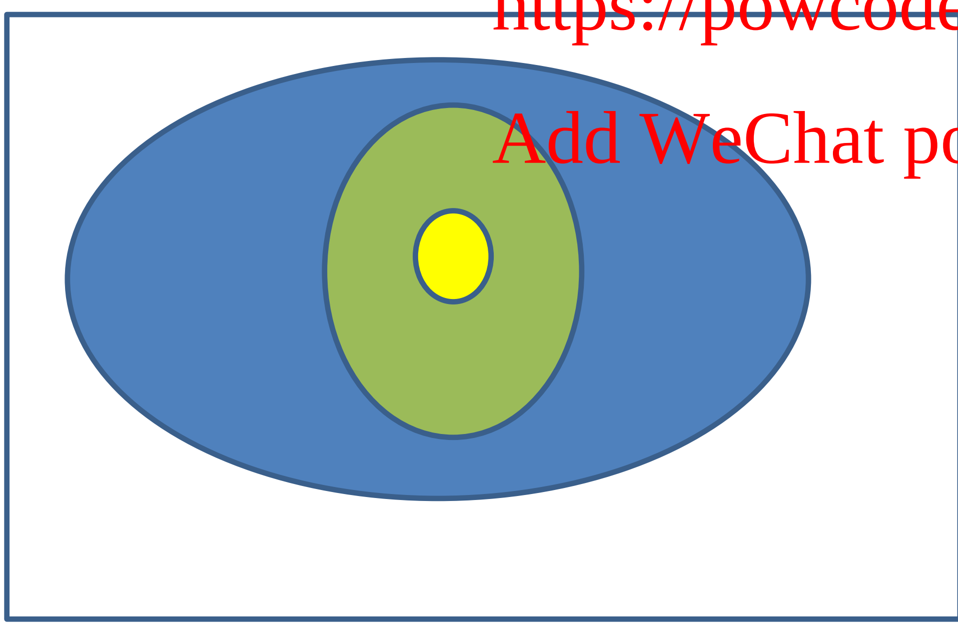
linked documents vs linked data

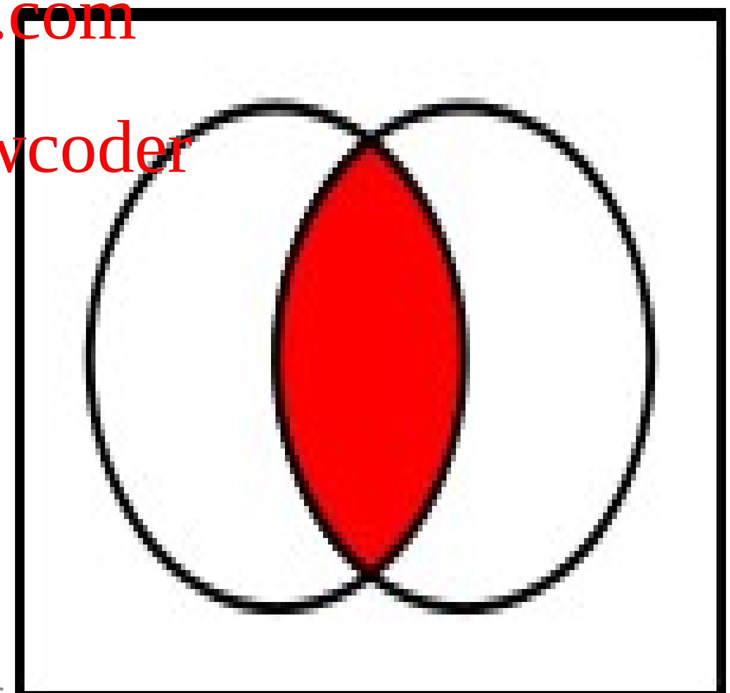
- **Web 1 was about linked documents, Web 2 is about social interactions and Web 3 will be about linked data!**
- **In the process of linked data, performing effective logic and knowledge processing with computers is gaining prime importance.**
- **Noise, data, information, and knowledge can be considered as a hierarchy, on which data sits on the top of noise, and information sits on the top of data, and knowledge on the top of information.**

Lecture Outline

1. **Introduction**
2. Constructing Ontologies Manually
3. Reusing Existing Ontologies
4. Semiautomatic Ontology Acquisition
5. Ontology Mapping
6. Architecture

Categories are the base of semantic Web and are called **(i)** domains (in databases), **(ii)** types (in Artificial Intelligence), **(iii)** classes (in object oriented programming), and **(iv)** concepts (in logic). **Sets** can show **Categories**. For instance, subclasses can be shown with:





Methodological Questions

  - Which languages and tools should be used in which circumstances, and in which order?
  - What about issues of quality control and resource management?

- Many of these questions for the Semantic Web have been studied in other contexts
  - E.g. **(i)** software engineering, **(ii)** object-oriented design, and **(iii)** knowledge engineering

Lecture Outline

1. Introduction
2. **Constructing Ontologies Manually**
3. Reusing Existing Ontologies
4. Semiautomatic Ontology Acquisition
5. Ontology Mapping
6. On-To-Knowledge SW Architecture

8 main Stages in Ontology Development:

1. Determine scope
2. Consider reuse
3. Enumerate terms
4. Define taxonomy
5. Define properties
6. Define facets (cardinality, symmetry, transitivity,………..)
7. Define instances
8. Check for anomalies

Not a linear process!

Lecture Outline

1. Introduction
2. Constructing Ontologies Manually
3. **Reusing Existing Ontologies**
4. Semiautomatic Ontology Acquisition
5. Ontology Mapping
6. On-To-Knowledge SW Architecture

Existing **Domain-Specific** Ontologies

DBPedia is a great source of Knowledge with all people around the world contributing to improving its status

There are many domains for ontology, for instance:

- **Medical domain**: Cancer ontology from the National Cancer Institute in the United States

- **Cultural domain**:

  - Art and Architecture Thesaurus (AAT) with 125,000 terms in the cultural domain
  - Union List of Artist Names (ULAN), with 220,000 entries on artists
  - Iconclass vocabulary of 28,000 terms for describing cultural images
- **Geographical domain**: Getty Thesaurus of Geographic Names (TGN), containing over 1 million entries

Integrated Vocabularies

- **Merge** independently developed vocabularies into a single large resource
- E.g. Unified Medical Language System integrating 100 biomedical vocabularies
  - The **UMLS metathesaurus contains 750,000 concepts**, with **over 10 million links between them**
- The semantics of a resource that integrates many independently developed vocabularies is rather low
  - But very useful in many applications as starting point

Upper-Level Ontologies

- Some attempts have been made to define **generally applicable ontologies**
  - Not domain-specific
- Cyc, **with 60,000 assertions on 6,000 concepts**
- Standard Upperlevel Ontology (SUO)

Topic Hierarchies

- Some “ontologies” do not deserve this name:
  - simply sets of terms, loosely organized in a hierarchy
- This hierarchy is typically not a strict taxonomy but rather mixes different specialization relations (e.g. **is-a**, **part-of**, **contained-in**)
- Such resources are often very useful as starting point

Linguistic Resources

- Some resources were originally built not as abstractions of a particular domain, but rather as linguistic resources
- These have been shown to be useful as starting places for ontology development
  - E.g. **WordNet**, **with over 90,000 word senses**

Ontology Libraries

- Attempts are currently underway to construct highly sophisticated online libraries of valuable online ontologies

  1) Rarely existing ontologies can be reused without changes

  2) Existing concepts and properties must be refined using:

  rdfs:subClassOf and **rdfs:subPropertyOf**

  – Alternative names must be introduced which are better suited to the particular domain using **owl:equivalentClass** and **owl:equivalentProperty**

Ontology Repositories

- https://www.w3.org/wiki/Ontology_repositories

Design your own ontology

- https://www.youtube.com/watch?v=sK2rFFf53uU&ab_channel=OpenHPITutorials

- https://protege.stanford.edu/publications/ontology_development/ontology101.pdf

Lecture Outline

1. Introduction
2. Constructing Ontologies Manually
3. Reusing Existing Ontologies
4. **Semi-automatic Ontology Acquisition**
5. Ontology Mapping
6. On-To-Knowledge SW Architecture

The Knowledge Acquisition Bottleneck

- Manual ontology acquisition remains a (i) time-consuming, (ii) expensive, (iii) highly skilled, and sometimes (iv) cumbersome task.

In fact, Machine Learning techniques may be used to alleviate

  – knowledge **acquisition** or extraction
  – knowledge **revision** or maintenance

Tasks Supported by Machine Learning :

I. Extraction of ontologies from existing data on the Web

II. Extraction of relational data and metadata from existing data on the Web

III. Merging and mapping ontologies by analysing extensions of concepts

IV. Maintaining ontologies by analysing instance data

Useful **Machine Learning Techniques** for Ontology Engineering:

- Clustering
- Incremental ontology updating
- Support for the knowledge engineering
- Improving large natural language ontologies
- Ontology learning

Lecture Outline

1. Introduction
2. Constructing Ontologies Manually
3. Reusing Existing Ontologies
4. Semi-automatic Ontology Acquisition
5. **Ontology Mapping**
6. Architecture

Ontology Learning

- https://www.youtube.com/watch?v=lXgex750Acs&ab_channel=OpenHPITutorials

Ontology Mapping

A single ontology will **rarely** fulfill the needs of a particular application; multiple ontologies will have to be combined

This raises the problem of **ontology integration** (also called **ontology mapping**)

Current major approaches in ontology mapping are:

- (i) linguistic,
- (ii) statistical,
- (iii) structural, and
- (iv) logical methods

(i) Linguistic methods

- The most basic methods try to exploit the linguistic labels attached to the concepts in source and target ontology in order to discover potential matches

(ii) Statistical Methods

- A significant statistical correlation between the instances of a source concept and a target concept, gives us reason to believe that these concepts are strongly related
- These approaches rely on the availability of a sufficiently large amount of instances that are classified in both the source and the target ontologies

(iii) Structural Methods

- Since ontologies have internal structure, it makes sense to exploit the graph structure of the source and the target ontologies and try to determine similarities, often in coordination with other methods
  - If a source target and a target concept have similar linguistic labels, then the dissimilarity of their graph neighborhoods could be used to detect homonym problems where purely linguistic methods would falsely declare a potential mapping

(iv) Logical Methods

- The most specific to mapping ontologies
- A serious limitation of this approach is that many practical ontologies don't carry much logical formalism with them
- In any case, if an ontology carries heavy logical formalism, logical methods can be effectively used for its mapping.

Ontology-Mapping Techniques Conclusion

- Although there is much potential, and indeed need, for these techniques to be deployed for Semantic Web engineering, this is far from a well-understood area
- For Ontology-Mapping, no off-the-shelf techniques are currently available, and it is not clear that this is likely to change in the near future.

- https://www.youtube.com/watch?v=VJHKcq_GuxY&ab_channel=OpenHPITutorials

Lecture Outline

1. Introduction
2. Constructing Ontologies Manually
3. Reusing Existing Ontologies
4. Semi-automatic Ontology Acquisition
5. Ontology Mapping
6. **Architecture**

Architecture

- Building the Semantic Web or in fact its architecture involves :

I. Knowledge Acquisition
II. Knowledge Storage
III. Query Languages, for processing the knowledge stored, and
IV. Knowledge Maintenance

Knowledge Acquisition

- Initially, tools must exist that use surface analysis techniques to obtain content from documents

  - Unstructured natural language documents: statistical techniques and shallow natural language technology

  - Structured and semi-structured documents: pattern recognition

Knowledge Storage

- The output of the analysis tools is sets of concepts, organized in a shallow concept hierarchy with at best very few cross-taxonomical relationships
- RDF/RDF Schema are sufficiently expressive to represent the extracted info
  - Store the knowledge produced by the extraction tools
  - Retrieve this knowledge, preferably using a structured query language

Query Languages

Without query languages, questions cannot be answered and since, **Semantic Web is involved with answering questions**, these query languages like **(i) SPARQL**, **(ii) DL**, and **(iii) SWQRL** play key roles in Web Semantic.

Knowledge Maintenance and Use

- **A practical Semantic Web repository must provide functionality for managing and maintaining the ontology**:
  - change management
  - access and ownership rights
  - transaction management
- There must be support for both
  - Lightweight ontologies that are automatically generated from unstructured and semi-structured data
  - Human engineering of much more knowledge-intensive ontologies

More ontology design methodologies

- https://www.youtube.com/watch?v=uH4BBeHy1NM&ab_channel=OpenHPITutorials

Ontology Evaluation

- https://www.youtube.com/watch?v=mol_BJklNH0&ab_channel=OpenHPITutorials