



KESA



Knowledge Extraction and Semantic Annotation

Maria Pia di Buono

Mario Monteleone

Annibale Elia

Department of Political, Social and Communication
Sciences

University of Salerno

April 19 - 24, 2015 - Barcelona, Spain

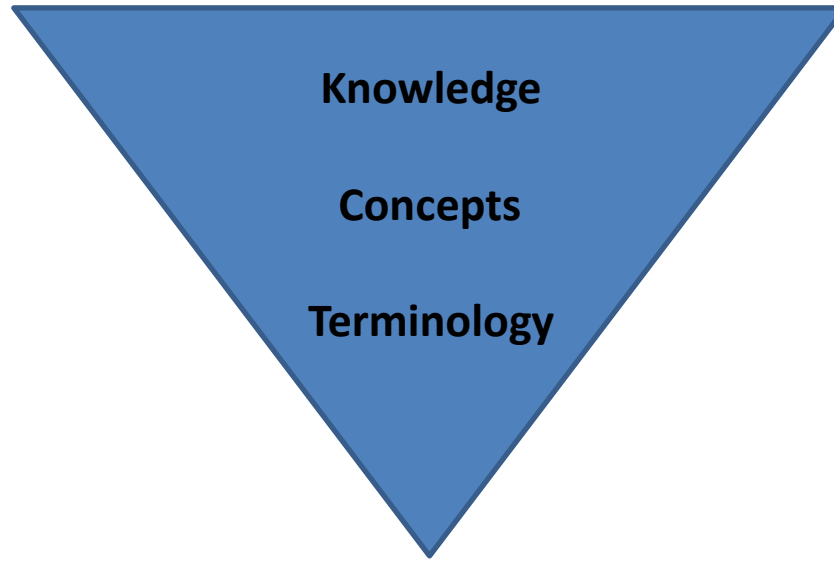
Human & Machine Semantics

Is there any possibility to make an efficient and effective matching between Human and Machine Semantics?

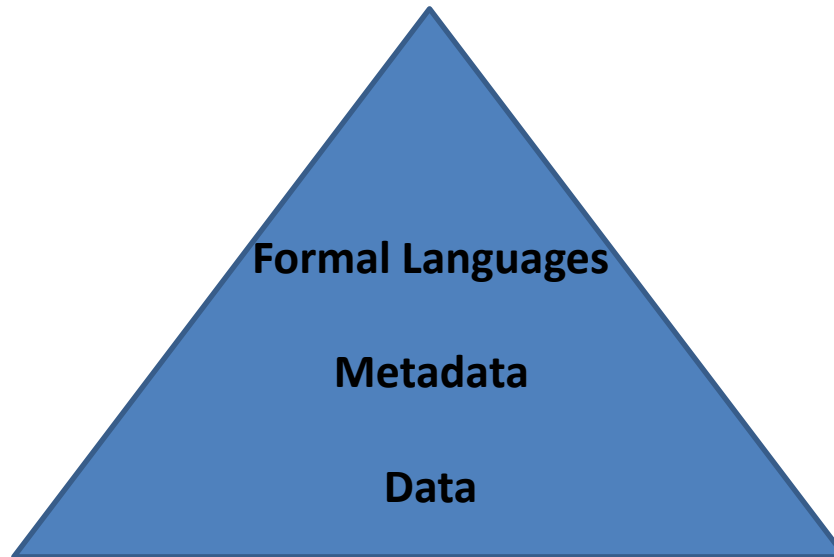
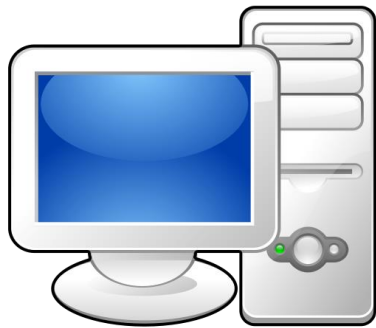
Semantics \neq Pragmatics



Abstraction Level



————— **Ontology** —————



Abstraction Level



Knowledge Representation

Knowledge Representation is heavily **based on** using **terminology**, due to the fact that many **terms** have precise **meanings** in a specific **domain** but not in others.

Therefore, a coherent and consistent **formal description language** is crucial and indispensable to achieve a **correct semantic representation** of a specific knowledge domain.

Semantic Annotation 1/2

A correct semantic representation is based on a semantic annotation process of resources.

We may differentiate three types of annotations:

1. **informal annotations** (they don't use a formal language)
2. **formal annotations**, that have formally defined constituents and are thus machine-readable, and
3. **ontological annotations**, that have formally defined constituents and use only ontological terms that are socially accepted and understood

(OREN, Eyal, et al. What are semantic annotations. Technical Report. *DERI Galway*, 2006.)

Semantic Annotation 2/2

Annotation task provides some data to some other data, creating a **relation** between annotating data and annotated data.

Several tools and paradigms exist to create annotations of Web resources, both manually, semi-automatically and fully automatically (i.e. OntoText, OntoGen, GoNTogle, NooJ, SemTag, TextToOnto, ...)

Knowledge Extraction 1/2

Knowledge Extraction is the retrieval, from structured and unstructured sources, of text elements and text data bringing and representing knowledge.

Extracting semantic content requires **concept-based systems**, which use sets of features you need to (pre)define in order to represent conceptualizations and their formalizations.

Knowledge Extraction 2/2

Extraction approaches require accurate recognition techniques for **semantic disambiguation** in order to cover several kinds of **descriptive data and metadata**.

The output must be in a **machine-readable** and **machine-interpretable** format and must represent the retrieved knowledge in order to facilitate (mainly) machine-based inferencing processes.

Ontology-Based Information Extraction 1/2

- **Ontology-based information extraction (OBIE)** is a Knowledge Extraction (KE) task
- OBIE uses ontologies and their specifications to "drive" the information extraction process.
- When tagging resources, terms and concepts in the source ontology form the basis for term matching.

Ontology-Based Information Extraction 1/2

OpenSemanticFramework Prescriptions for OBIE (2014):

- All ontology concepts have a definition
- All ontology concepts have alternative labels
- Where appropriate, ontology concepts have hidden labels to account for common misspellings
- Inference is used as appropriate during extraction processes.