

The field of the research is *Ontology learning from texts*. My initial dissertation topic is *Learning the domain ontology: minimizing expert interruption*. By this, I mean two things:

First that the aim of the research is to develop a prototype of an automated system, which would be used for a specific domain ontology learning (extracting). Learning of the ontology requires linguistic resources, i.e. technical texts (articles, technical reports) of the domain, and combining NLP tools and computational linguistics algorithms.

Second, that minimizing the manual work done by the experts of the domain in every step of ontology building is an important task to achieve.

Learning the domain ontology: the idea.

Learning of the ontology is a technically and theoretically complex task. The present-day ontology learning researches more or less share the methodology, consisting of the steps depicted in Figure 1.:

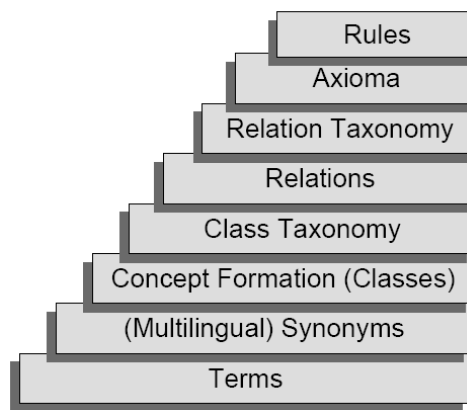


Figure 1. Ontology learning cake (Resource F. Cimiano's dissertation, 2007).

According to this methodology, learning of the ontology requires having a domain specific corpus, tools for natural language processing in order to extract terminology, refine it and define relationships between terms, i.e. hyperonymy - hyponymy relationships, synonymy relationships. As well as tools for classifying terms into facets in order to build a taxonomical representation. Inference rules and logical reasoning tools are needed in order to provide a logical representation of ontology. When there exists a part/parts of the ontology, then ontology mapping tools are necessary.

My approach to ontology learning is as follows:

- Building a domain specific corpus;
- Linguistically based terminology extraction;
- Building a domain specific thesaurus;
- Developing taxonomical structure of the thesaurus and enriching it with synonymy relationships;
- Manually restructuring the existing upper taxonomy;
- Creating a logical representation of the upper taxonomy;
- Mapping taxonomically enriched thesaurus to an upper domain ontology.