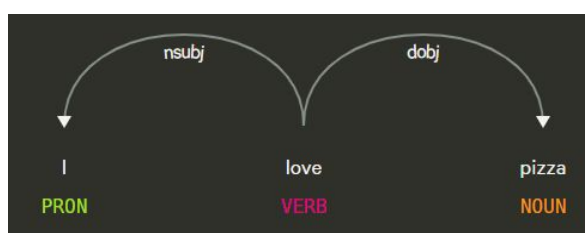


Semantic Technologies in the Cloud and HPC domains

A short introduction to Semantic Technologies, and SODALITE's approach for their application in the software engineering domains.

Defining the word *semantics* is considered as a great introductory point in order for the reader to gain a good grasp of the Semantic Technologies concepts. In all its forms, communication can be abstractly separated in two codependent notions, *syntax* and *semantics*. Syntax is how you structure the expression of something, while semantics is the actual meaning behind what you are trying to express.



Take, for instance, the phrase "*I love pizza*". Syntactically speaking, this phrase is structured based on strict linguistic rules, incorporating a pronoun as a subject, followed by a verb and a noun as an object. On the other hand, semantics are hidden in words. *Love* is broadly known as this heart-warming feeling and *pizza* as a

round, cheesy delicacy. While syntax can be described as a defined and commonly accepted set of structures, semantics tend to be harder to define formally. Plus, semantics are usually subjected to ambiguity, caused by factors such as time, locality, personal experiences, familiarity with the domain, and many more. For example, the notion of pizza might be differently perceived and described between an Italian chef and a lactose-intolerant consumer.

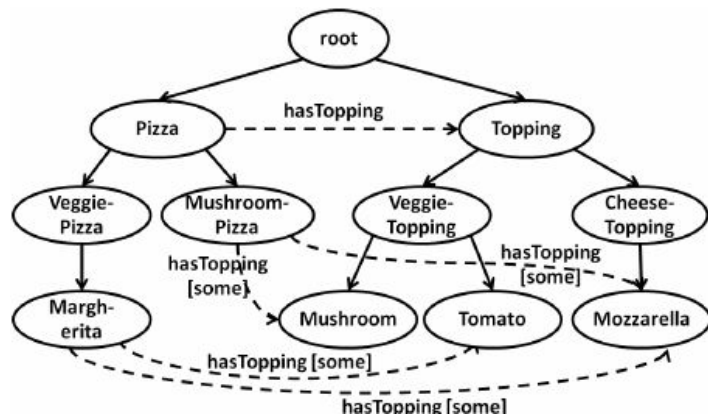
When it comes to technology and computers, knowledge exchange has been established on strong syntactical foundations. Communication protocols (e.g. HTTP), generic programming languages (e.g. Python, Java), and domain-specific languages (e.g. HTML, JSON, CSS) are only a tiny part of an expanding arsenal for the creation and sharing of content and functionalities. However, the emerging demand for more efficient exploitation of the huge volumes of collected knowledge dictates the need for exploring the meaning behind computer data. To this end, great efforts have been made in the field of knowledge representation into formal, well-defined structures, which enable computers to store, manipulate and exchange semantics, rather than plain data. The most iconic and highly-adopted accomplishment of these efforts is *Ontologies*.

Ontologies

An ontology constitutes a model for the representation of knowledge in a given domain. It incorporates the definition and hierarchical categorization of concepts in and around the domain of discourse, along with relationships between concepts and datatype properties (e.g. literal values, etc.) of entities. The definition of an ontology requires great familiarity with

the domain and is also subject to the aforementioned ambiguity. Thus, it is usually performed in close collaboration with domain experts. Nevertheless, a simplistic example of the pizza domain (which is actually quite common in ontology tutorials) is presented below as a graph.

The main concepts (classes) in this sample ontology are *Pizza* and *Topping*, followed by further specifications of sub-concepts (subclasses) like *Veggie-Pizza* and *Cheese-Topping*. The arrows indicate potential relationships between classes, such as *hasTopping* between a *Pizza* and a *Topping* entity. Moreover, this schema contains *axioms* that are able to automatically infer further classifications. For instance, a



Pizza entity that has a *Mushroom* topping can be safely considered as a *Mushroom-Pizza*. The generation of such inferences is the baseline of what is called *Semantic Reasoning*.

Semantic Reasoning

Semantic Reasoning is the process of inferring hidden or underlying knowledge from existing knowledge, based on predefined rules and axioms. While our case above is oversimplified, more complex reasoning rules can be applied by elaborate software tools named *Semantic Reasoners*. Needless to say, by allowing the manipulation of existing knowledge in meaningful ways and the derivation of new, Semantic Reasoning constitutes one of the epitomes of Semantic Technologies. But what are the rest?

The Semantic Web

As we have stated earlier, the design of a purposeful ontology schema is a delicate process that requires domain expertise and resources. Thankfully, ontology creation allows the importing and reuse of previously created ontologies. As a result, by exploiting existing work, not every aspect of a domain needs to be thoroughly modelled from scratch, significantly reducing the required effort and time. To this direction, there are several established ontologies available for reuse, like the [Good Ontologies](#). Indicatively:

- [Friend Of A Friend \(FOAF\)](#) describes people and social relationships on the Web.
- [Dublin Core](#) models generic metadata of artifacts.
- [The Music Ontology](#) contains specifications for knowledge related to the music industry.

The most important outcome of this feature is the ability to adopt and extend established international standards, as defined by domain experts, and use such schemas to represent

own knowledge in a globally recognizable structure. For example, think of an institution that generates, preserves and shares biomedical data conforming to the [Gene Ontology](#). Then imagine another institution that retrieves and aggregates these structured semantics with local knowledge, and applies semantic reasoning for advanced research purposes. This example can be extended to a distributed network of computer systems and knowledge repositories worldwide, exchanging and manipulating interconnected semantics (not just data!) on a global scale. In a nutshell, this is the principal idea of what is called *the Semantic Web* (or *Web 3.0*).

The SODALITE approach

The SODALITE project incorporates Semantic Technologies in order to address the complexity and diversity of the software engineering domain. More specifically, four interlinked subdomains have been identified and are being modelled by the consortium's multidisciplinary members: a) software applications, b) Cloud/HPC infrastructures, c) performance optimization, and d) deployment and lifecycle of applications. The outcome of this process will be an elaborate ontology to act as the core of SODALITE's Semantic Knowledge Base (KB). The KB will fuse and host domain knowledge - both at an abstract and concrete level - generated by the system's users, application ops experts, resource experts, deployment mechanisms, optimization algorithms and more.

On top of the ontology, a sophisticated Semantic Reasoner will be performing a series of meaningful tasks to support the system's functionality. Such tasks include:

- The generation of recommendations in deployment design time, deriving from resource patterns, infrastructure availability, etc.
- The detection of invalid deployment plans, based on known problems, incompatibilities and antipatterns.
- The decision support for performance optimizations, based on performance optimization patterns and approaches.
- The preservation of mappings between abstract deployment models, designed by the users, and concrete deployment models, generated by SODALITE, in order to facilitate the monitoring of the deployment lifecycle on an abstract level.

Conclusion

To wrap up, Semantic Technologies are an evolving part of Computer Science. They address the challenges of modelling and manipulating the semantics underlying data, and offer the possibility of sharing such knowledge on a global scale via a network of interconnected and associated knowledge repositories (Semantic Web). With its focus on Semantic Technologies, SODALITE aspires to exploit the team's substantial expertise in computer engineering to significantly contribute to the Semantic Web initiative, both by providing solid and detailed semantic models for the related domains, and by highlighting the broad capabilities of Semantic Reasoning mechanisms.