Pragmatic Semantics for the Web of Data

Stefan Schlobach, Christophe Guéret, and Kathrin Dentler

Vrije Universiteit Amsterdam

Abstract. The Web of Data (WOD) is complex, and inherently messy, contextualised, opinionated, in short: it is a **market-place of ideas, rather than a database.** Logic-based semantics were not developed, and are not suitable, for such complex systems. A radically new approach to semantics is required. We introduce **pragmatic semantics** as a new semantic paradigm integrating elements from market-theory and classical semantics into a framework of optimisation over truth-orderings, each representing a particular world-view. We propose nature-based algorithms to implement those semantics.

1 Semantics for a Complex System

On the Web in general, and the Web of Data (WoD) in particular, almost every piece of information is strongly context-dependent and described from multiple viewpoints. The Web of Data is nowadays more and more recognised to be a Complex System [4], with interlinked information at different scales of abstraction. A well-argued claim in the Complex Systems literature suggests that it is impossible to construct logical systems that capture the full meaning of any Complex System. Thus, approaches to extend current methods, for example, with quantitative information about vagueness and uncertainty, or towards multi-dimensionality (for modelling contexts), although highly useful and ingenious for specific applications, necessarily fall short of representing the full richness of semantics of the Web of Data.

This implies that the classical semantics of data and ontology languages such as RDF(S) and OWL, which is based on truth with respect to **all** world-views (or models) is often inappropriate, as those languages are not designed to deal with inconsistent, multi-dimensional and contextualised knowledge. The most promising recent attempt to address this problem is the work on Emergent Semantics [2], defining semantics as the result of collective processes and interactions between nodes in a network - a collective agreement. Although this formalism can capture some of the emerging structure, the price is that meaning and truth are defined as results of processes or calculi, and the well-understood declarative, model-theoretic semantics of traditional formalisms are lost.

Nevertheless, the Semantic Web requires semantics to allow for interoperability and seamless interchange of data. In order to solve this apparent CATCH-22, we propose a new semantic paradigm based on multi-objective optimisation of a variety of truth criteria, which we will call **Pragmatic Semantics**. We will finally argue that Pragmatic Semantics require nature-inspired calculi, such as swarm or evolutionary algorithms.

2 Pragmatic Semantics

Let us introduce the problem with a simple example. Figure 1 shows some typical knowledge on the current Web of Data. A Dutch dataset (using a namespace `nl:`) describes some European cities, among them Paris and Amsterdam, which are capitals and do not require a visum for travel. In good practice for Linked-Open data the new resources are linked to some existing sources, e.g. DBpedia[1], by an `owl:sameAs` predicate. Similar data is published in China (using the namespace `ch:`), but now for European cities a visum is required. Both pieces of information are locally correct and the linking follows the right principles. Still, as soon as we consider the two classes `ch:VisumNeeded` and `nl:VisumFreeCity` to be disjoint, classical semantics collapse, and even useful derivable information, such as the fact that Amsterdam is a city with an airport as it is a capital city is lost.

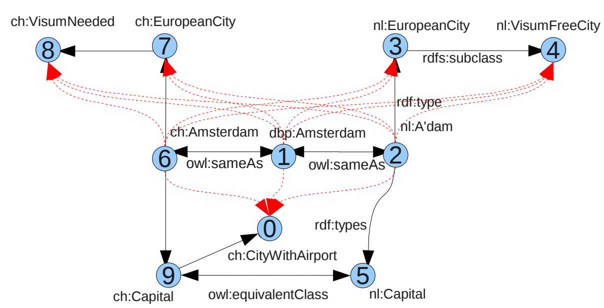

Fig. 1. Example: Three combined ontologies that are together inconsistent.

With Pragmatic Semantics we propose a formalism to integrate classic model-theoretic notions of truth with explicit knowledge about the structure of the knowledge base, and with the (possibly implicit) information-need often associated with data.

Let us recall some formal notions: let $\mathcal{O}$ be an ontology, for example a set of RDF(S)/OWL triples, or a set of axioms in a logical language such as a Description Logic. Without loss of generality we consider in classical semantics an interpretation of $\mathcal{O}$ to be a pair $(\mathcal{U}, \cdot^{\mathcal{I}})$ where $\mathcal{U}$ is a set of objects called the domain and $\cdot^{\mathcal{I}}$ an interpretation function assigning individuals to objects, concepts and classes to subsets of the domain, and properties and roles to binary relations. This interpretation function is usually extended over the operators of the underlying representation language. *Models* are those interpretations satisfying all the axioms in $\mathcal{O}$. An axiom or triple α is then classically entailed by $\mathcal{O}$ if it is satisfied in all the models of $\mathcal{O}$. Unfortunately, even in our simple example things go wrong as there cannot be a model where the instance `dbp:Amsterdam` is both visum free, and a city where a visum is required. By definition, everything is entailed, which makes the semantics useless.

The basic idea of Pragmatic Semantics is that *weaker notions of truth* can be defined in a variety of ways, each of them covering one or more aspects of the original ontology, and most of them inducing some kind of ordering on the entailed formulas, which we will call *partial truth orderings*. Those could include,

[1] http://www.dbpedia.org

for example, explicitly modeled quantitative information about confidence or degree such as provided in probabilistic or fuzzy ontology languages. But even for classical Boolean ontology languages such as RDF(S)/OWL, weaker notions are abundant, and could be based, for example, on structural properties of the ontology, but also on external information extracted from background knowledge.

There are numerous other ways conceivable to define such partial truth orderings on formulas, e.g., **based on sub-ontologies**. A simple example is the ordering induced by the size of the minimal ontology which classically entails a formula φ. The ratio of sub-models (models for parts of an ontology) in which φ is satisfied versus the total number of sub-models, or the ratio between sub-ontologies of $\mathcal{O}$ in which φ holds versus the number of all sub-ontologies are other interesting candidates.

Another relevant class of orderings is given by the **graph properties** of the ontology, in case that the underlying data-model is graph-based. A `shortest path ordering` can be determined as the inverse of the longest shortest distance between all nodes in the ontology (ordering of the diameter of the induced sub-graphs). Such a notion can be used as a proxy for determining confidence of derivations. Other graph-based measures, e.g. based on random-walk distance or edge-weights induce orderings that are clustering-aware, where the sub-ontologies entailing a formula supposedly have more cohesion than others. Finally, taking node properties such as PageRank into account, orderings can be used as proxies for popularity.

The point of informally introducing all these truth-orderings is to motivate the following definition of pragmatic entailment. The different orderings cover different aspects of the true semantics of the Web of Data, and entailment is defined pragmatically as a multi-objective optimisation for those aspects.

Definition 1. *Let O be a not necessarily consistent ontology. Let furthermore $PO = \{\preceq_1, \ldots, \preceq_n\}$ be a set of partial orderings over sets of formulas induced by $\mathcal{O}$. A set of formulas Φ is then called a pragmatic closure of $\mathcal{O}$ w.r.t. PO, if and only if it is Pareto-optimal wrt. to the optimisation problem* $\max_{\Phi}[\preceq_1 (\Phi), \ldots, \preceq_n (\Phi)]$.

In order to achieve interoperability one needs to enrich an ontology with meta-information about choices for semantic orderings, and even given agreement on the orderings, there are possibly several pragmatic closures (different solutions on the Pareto-front), i.e. also agreement on the weighting of features is required. Note that these agreements are explicit, and correspond to agreement on particular aspects of the semantics of an ontology.

3 Nature-inspired calculi

Results from studying the Web of Data as a Complex System show that considering different scales and levels of interactions make it impossible to engineer a web-scale reasoner (whatever the semantics considered), as traditional, decomposition-based approaches, are doomed with bandwidth limitations between the coordinating components (i.e. the datasets). Changes in one dataset have potential implications for all the other datasets connected to it, and so on. Instead, such a system has to evolve according to biological evolution rules [1], and web-scale semantics and reasoning should emerge from controlled interactions between autonomous components.

In [3] we introduced such a calculus based on swarm intelligence where instead of indexing all triples and joining the results, swarms of lightweight agents (so-called boids) autonomously traverse the graph, each representing a reasoning rule, which might be (partially) instantiated. Whenever the conditions of a rule match the node a boid is on, it locally adds the new derived triple. This provides an index-free alternative for reasoning over large distributed dynamic networks of RDF(S) graphs.[2] It calculates the pragmatic closure under the condition of maximising popularity of nodes (as random walks of boids simulate PageRank calculation) and minimizing the length of sub-ontologies, two particular truth orderings.

4 Conclusion

This paper introduces a new paradigm of optimisation-based semantics. Given that partial truth orderings over formulas can be defined even for inconsistent, multi-dimensional and contextualised knowledge, pragmatic entailment can cover multiple aspects of the intended meaning of ontologies in an optimal way. These new semantics reflect the transformation of the Semantic Web from a (potentially large and distributed) database holding a single truth into a marketplace of ideas.

We briefly discussed a swarm-based calculus, which is obviously only a first minimal attempt at a better understanding of Pragmatic Semantics. We see this paper as a call for a research program with the goal of providing appropriate semantics for the Complex System we call Semantic Web.

References

1. Bar-Yam, Y.: About engineering complex systems: Multiscale analysis and evolutionary engineering. In: Brueckner, S. (ed.) Engineering Self-Organising Systems. LNCS, vol. 3464, pp. 16–31. Springer (2004)
2. Cudré-Mauroux, P.: Emergent Semantics – Interoperability in Large-scale Decentralized Information Systems. EPFL & CRC Press (2008)
3. Dentler, K., Schlobach, S., Guéret, C.: Semantic web reasoning by swarm intelligence. In: 5th International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS2009) (October 2009)
4. Guéret, C., Wang, S., Groth, P., Schlobach, S.: Mulit-scale analysis of the web of data: A challenge to the complex system's community. Advances in Complex Systems (ACS) 14(4), 587–609 (2011)

[2] See http://www.beast-reasoning.net/ for some experiments and available code.