1. GENERAL INFORMATION

Project title: Pragmatic Semantics for the Web of Data

Acronym: PraSem

Principal Investigator: Dr. Stefan Schlobach

2. SUMMARY

2A. SCIENTIFIC SUMMARY

What is the problem? The success of the Web of Data (WOD) is based on the thorough understanding of, and agreement upon, the semantics of data and ontologies. But the Web of Data as a whole is complex, and inherently messy, contextualised, opinionated, in short: it is **a market-place of ideas, rather than a database.** Existing paradigms are inappropriate for dealing with this new type of knowledge structures.

Why is it important? The urgency of dealing with the non-standard characteristics of the Web of Data has been recognised, and separate initiatives try to tackle its individual manifestations, e.g. inconsistencies, contexts, vagueness, provenance, etc. **Tomorrow's Web of Data requires novel semantics** with efficient (generic) implementations to ensure semantic clarity, reuse and interoperability.

How will we solve it? We recently introduced pragmatic semantics [29] 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.

What will be the main results?

1) Well-understood formalisms for pragmatically interpreting semantic data on the Web.
2) Algorithms to calculate the related inferences, such as entailment and truth, and guidelines for using the new paradigm in application domains.
3) A system integrating various implementations of generic and specific multi-optimisation methods, implementations of a variety of truth-orderings, and representation standards for semantic meta-data.
4) A thorough analytic and empirical evaluation of the proposed semantic paradigm.

2B. LAYMEN'S ABSTRACT

Het huidige Web of Data (WOD) maakt dat computerapplicaties automatisch feiten kunnen vinden, verwerken en op basis hiervan nieuwe feiten kunnen afleiden (bv. dat Rutte minister-president van Nederland is). Dit is mogelijk omdat deze informatie gecodeerd staat in termen van openbare ontologieën op basis van formalismen zoals RDF en OWL (zoals bijvoorbeeld DBPedia dat doet). In deze talen kan men bijna alles over objecten en klassen van objecten zeggen. Naar de eigenschappen van RDF en OWL is veel onderzoek gedaan en ze vormen als wereldwijde standaarden de basis van het huidige Web of Data.

Niettemin is de interpretatie binnen de applicaties vaak ingewikkeld: online informatie over Rutte is verspreid over een veeltal bronnen, met verschillende niveaus van detail en geloofwaardigheid. De informatie verandert over tijd, is onzeker en subjectief en is sterk afhankelijk van de context van het gebruik. Iedereen kan alles over alles en iedereen zeggen. Het Web of Data, dat tot nu toe als een database beschouwd werd, is in het echt een marktplaats van ideeën. Het blijkt dat de standaarden en talen waarover we nu beschikken niet voor een dergelijke marktplaats zijn gemaakt, waardoor we in de praktijk veel van de echte betekenis (de semantiek) van de informatie kwijtraken.

Door het onderzoeken van Pragmatic Semantics voor het Web of Data zal PraSem een belangrijke bijdrage leveren aan een nieuw, overkoepelend raamwerk. Dit gebeurt door criteria voor waarheid stap voor stap op nieuw te definiëren, en in een proces van multi-objective optimalisatie te integreren. Op deze manier kunnen gebruikers de bestaande formalismen en talen blijven gebruiken en wordt recht gedaan aan het complexe karakter van het WOD door de methoden van interpretatie hierop te laten aansluiten.

2C. KEYWORDS

Web of Data, Semantic Web, Knowledge Representation, Semantics, Optimisation, Nature-Inspired Reasoning, Ontology Languages

3. CLASSIFICATION

NWO classification: 2.7. Redeneersystemen

NOAG theme: Intelligente systemen

ICT discipline: Artificial Intelligence

4. RESEARCH TEAM

The core research team will consist of the PhD student to be funded by NWO, the principal investigator and promoter. The principle investigator will also act as co-promoter.

Name	Role in the project	Expertise	Affiliation
	PhD student	Pragmatic Semantics	VUA
Dr. Stefan Schlobach	Principal Investigator (co-promoter)	Knowledge Representation and Reasoning, Web of Data	VUA
Prof. Frank van Harmelen	Promoter	Knowledge Representation and Reasoning, Web of Data	VUA

The project will be embedded in the research of the Department of Computer Science, which has a unique mix of expertise for the proposed research, including Complex and Distributed Systems, High Performance Computing and Computational Intelligence. The Knowledge Representation and Reasoning group has been collaborating with all three groups in the past. There has also been a longstanding collaboration with the research unit on Scientometrics at the Group of Organisational Science, which will be paramount in the evaluation w.r.t the application scenario in Scientometrics.

Prof. Maarten van Steen	Complex Systems, Networks, Large Scale Distributed Systems	Computer Science
Prof. Peter van de Besselar	Application: Scientometrics	Organisational Science
Prof. Henry Bal	High Performance Computing.	Computer Science
Prof. Guszti Eiben	Computational Intelligence	Computer Science

5. RESEARCH SCHOOL

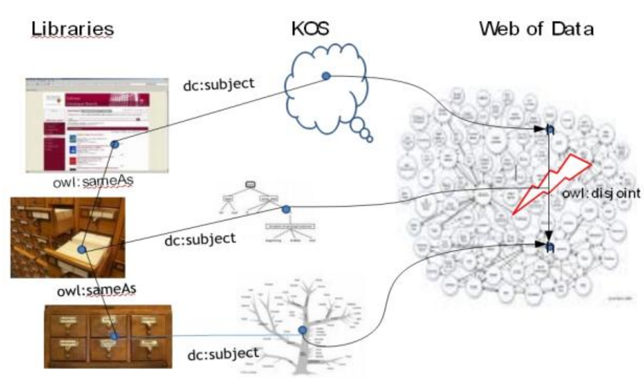
SIKS

6. DESCRIPTION OF THE PROPOSED RESEARCH

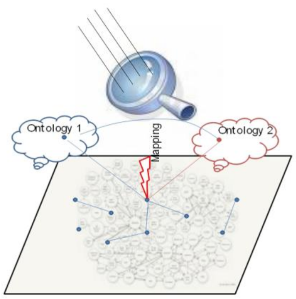
6A. SCIENTIFIC ASPECTS

WHAT IS THE PROBLEM?

The Web of Data (WOD) connects data in a similar way as the WWW connects documents. Atomic data-units called resources are connected via typed links with arbitrary resources anywhere on the Web, and together these RDF triples form a gigantic graph of linked data. The meaning of the types can be fixed using standardised schema and ontology languages such as RDFS and OWL. The semantics of these languages are based on logical paradigms that were designed for small and hand-made knowledge bases, and come with a classical model-theory assigning truth to formulae, and entailment based on this truth. In a highly complex, dynamic, context-dependent, opinionated, contradictory and multi-dimensional semantic network as the WOD, these Semantics are insufficient, as they are one-dimensional, often prone to logical fallacies, and usually intractable.



Two simple scenarios will illustrate some of the high-level problems. **Scenario 1 is about heterogeneous publishing of data.** Many libraries describe their books with controlled vocabularies. Linking collections and those vocabularies to the Linked-Open Data cloud, a collection of hundreds of interconnected data-graphs on the WOD, has huge benefits for libraries as search becomes more powerful, and metadata of documents is automatically enriched. Suppose a library in China annotates a book about Amsterdam with a concept ch:SmallTown. The Dutch National Library, on the other hand, annotates the same book with subject nl:BigCity. What happens now when the two libraries add their vocabularies and data to the WOD? What should be the desired answer to a query for big cities? Linking the libraries' vocabularies to the Linked-Open Data cloud will lead to conflicts and hamper access to the document in question rather than support it.



Scenario 2 is about opinionated interpretation of data, and is taken from Scientometrics, the Science of measuring and predicting Science. Scientometric researcher often use the Web as a proxy for studying science itself. Scientists leave online traces while doing research and a lot of this data is structured and part of the Web of Data. In some way, the WOD becomes a magnifying glass to measure activity in Science. The problem, however, is that multiple views are omnipresent: research blogs are biased, there are networks of publications of different impact levels, social networks that overshadow reliable analysis, which all comes on top of the usual technical problems of instance- unification, homonymy and synonymy. Modelling this highly complex Science Web with standard ontology languages is impossible as long as standard semantics are enforced.

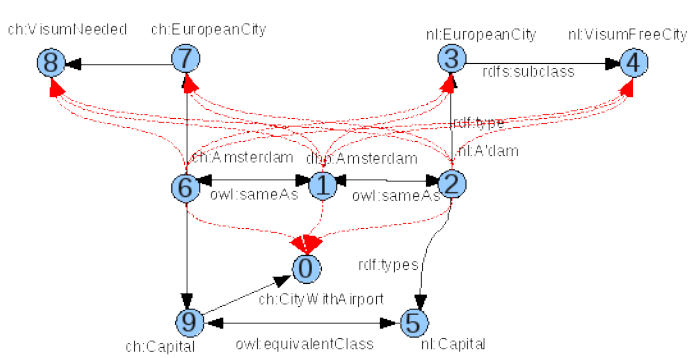
On the Web in general, and the Web of Data in particular, almost every bit of information is context-dependent, biased towards a particular viewpoint, opinionated, dated, uncertain or vague. **The WOD is a market-place of ideas, not a database, and has to be dealt with accordingly.** As making the representational languages more complex is not an option, we have to adapt the formal semantics of existing formalisms to the new requirements.

> *The goal of PraSem is to introduce, and critically analyse, a new semantic paradigm that is appropriate for the complex, and inherently inconsistent, noisy and multi-dimensional nature of the Web of Data.*

SCIENTIFIC TOPICS AND RESEARCH QUESTIONS

PraSem will introduce and assess the potential of novel semantics that can help overcome the weakness of traditional semantics when dealing with a messy, multi-dimensional, contextualised and complex knowledge structure such as the Web of Data.

Consider a prototypical example: a Dutch dataset describes European cities, among them Amsterdam, which is a capital and does not require a visa for travel. In good practice the resources are linked to existing sources, e.g. DBpedia, by an owl:sameAs predicate. Similar data is published in China (using the namespace ch:), but now for European cities a visa is required. Both pieces of information are locally correct and the linking



follows the correct principles. Still, considering 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. To address this issue, we need to deal with truth at different contexts.

Recent approaches to extend current methods, e.g., with quantitative information about vagueness and uncertainty, or towards multi-dimensionality are highly useful for specific applications. But they necessarily fall short of representing the full rich of the Web of Data. They also fail on a second, critical, requirement for our new semantics: as current ontology languages are already perceived as being too complex for practical use, the burden cannot be put on the modeller. **We need to adapt the semantics, not the languages**.

More concretely, we consider ontologies to be sets of RDF(S)/OWL triples. Without loss of generality an interpretation consists of a domain and interpretation function, assiging individuals to objects, concepts and classes to subsets of this domain, and properties and roles to binary relations, and which extended over the operators of the underlying representation language. Models are interpretations satisfying all the axioms. Axioms or triples are then *classically entailed* by the ontology if they are *satisfied in all its models*. Unfortunately, even in our simple example things go wrong as there cannot be a model where the instance dbp:Amsterdam is both visa-free and a city where a visa is required. By definition, everything is entailed: the semantics becomes useless.

PRAGMATIC SEMANTICS

Pragmatic semantics integrate different world-views instead of defining meaning with respect to a single one. The idea is to make as much information in the data explicit, and turn it into first-class semantics citizens. First, it allows integrating classic model-theoretic notions of truth with explicit knowledge about the structure of the knowledge base. But also semantic meta-data, such as popularity, scarcity, abnormality, etc., and even background knowledge from other sources, can be integrated.

Most of this additional knowledge induces some kind of ordering on the formulas, which we will call truth orderings. A simple example of such a truth ordering is the one induced by the size of the minimal subontology classically entailing a formula. Other examples are orderings derived as the ratio of sub-models (models for parts of an ontology) in which a formula is satisfied versus the total number of sub-models, or the ratio between sub-ontologies of O in which a formula holds holds versus the number of all sub-ontologies are interesting candidates. Those are **orderings based on subset of the ontology,** a well-known class often used when dealing with inconsistent ontologies [44].

Another relevant class are **orderings induced 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 (diameter of the induced sub-graphs). Such a notion is a proxy for confidence of derivation. Other graph-based measures, e.g. based on random-walk distance or edge-weights, induce orderings that are clustering-aware, with sub-ontologies entailing a formula have more cohesion than others. Finally, taking node properties such as PageRank into account, orderings can be used as proxies for popularity.

While those two classes of orderings make structural properties of the ontology explicit and use them to implicitly contextualise meaning, others are **based on external information** outside the ontology itself. Examples for this are the Google count and similarity [30] based on frequency of labels of resources on the WWW.

The different orderings cover different aspects of the "true" semantics of the Web of Data. To combine those aspects *pragmatic entailment* is defined through multi-objective optimisation. A *pragmatic closure* C for an ontology O and orderings f1 to fn is then a set of formulas that is Pareto-optimal [45] w.r.t. the optimisation problem max[f1 (C),…,fn (C)].

Interoperability is then achieved by enriching an ontology with meta-information about semantic orderings, as well as agreement on the weighting of orderings. As there are possibly several pragmatic closures (different solutions on the Pareto-front) also agreement on the weighting of features is required. We will refer to the entailment induced by a given set of orderings as an **instantiation of the family of pragmatic semantics**.

CALCULI FOR PRAGMATIC SEMANTICS

Another way of looking at it is that the Web of Data is a Complex System [28], 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 a true Complex System [35]. 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). Traditional semantics deal with this problem by an intrinsic reduction of the complexity: only one world-view, one perspective is considered at the time, the Web of Data is seen as a database. With pragmatic semantics, this advantage gets lost, and the computational price has to be paid, which applies that classical top-down reasoning becomes impossible.

It is often claimed that such systems have to evolve according to biological evolution rules [36], and web-scale semantics and reasoning should emerge from controlled interactions between autonomous components. In [37] 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. 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. Not all of the conceivable calculi for pragmatic semantics have to be inspired by Computational Intelligence approaches, but PraSem will focus on this family of algorithms.

RESEARCH GOAL AND QUESTIONS

The idea of Pragmatic Semantics is to allow users to integrate different levels of knowledge and abstraction directly and explicitly into the semantics of their ontology. This has implications that are not yet well understood, both from theoretical and an application perspective. The main goal of the research is thus threefold:

1. To investigate the theoretical properties of pragmatic semantics in general, and a number of application-specific instantiations, including formal (such as monotonicity, completeness, decidability etc) and computational properties. **(Properties of the Semantics)**

2. To assess whether the proposed formalisms are effective solutions to the semantic interoperability problem on the WOD for a selected number of practical scenarios **(Quality of the Semantics)**.

3. To investigate whether we can build efficient (approximate) calculi for a number of instantiations of pragmatic semantics **(Practical applicability of the Semantics)**.

Scientific question: To understand, define and evaluate pragmatic semantics as a proposed semantic formalism for the Web of Data?

EXPECTED SCIENTIFIC OUTCOME

The principle result of PraSem will be a **comprehensive theory of pragmatic semantics** for the Web of Data. Based on this theory, we will provide **generic methods** for applying and extending those semantics. The methods will be realized through software that provides reasoning and interoperability for large-scale semantic data.

1. Methodologically, we will develop
    a. a theory, i.e. formalisms to pragmatically interpret semantic data on the WOD

   b. algorithms to calculate the related inferences, such as entailment and truth, and
   c. guidelines for using the new paradigm in application domains.
2. We will build a prototype integrating various
   a. implementations of generic and specific multi-optimisation methods,
   b. implementations of a variety of truth-orderings.
3. We will develop standards for representing semantic meta-data.

Finally there will be a thorough analytic and empirical evaluation of the proposed approach.

The research results will be disseminated in the usual way in the respective high-quality literature, such as the proceeding of conferences such as IJCAI, AAAI, ISWC, ESWC or GECO, and journals such as JWS, SWJ, AI Journal, CIJ etc.

RESEARCH METHODS

Starting from the initial definition of pragmatic semantics provided in [29] we will conduct the necessary research to answer our scientific question in a practical, i.e. application driven, iterative way. We will start with the two scenarios described above, for which data is available from previous research projects NWO's STITCH and KNAW's SMS.

For both scenarios a thorough requirement analysis will result in a number of candidate in the three classes of subontology, graph-based and external knowledge-based orderings. Specific algorithms need to be developed, others can be reused (such as graph algorithms). We will implement/reuse a generic optimization algorithm, but specific algorithms for relevant combination of truth-orderings will probably be required for efficiency reasons.

The proposed semantics will be validated in two ways: first their quality, and secondly whether they can efficiently calculate. For the first,:

1. we will run controlled experiments on ontologies constructed by merging of data-sources, e.g. having the semantics identify the local consequences over global consequences.
2. we will run end-to-end experiments, measuring to what extent derived pragmatic consequences are valid in a particular application context.

Both methods will be applied to our two case scenarios.

We will start out with the following step-by-step procedure:

1. Requirement analysis for a use-case at hand
2. Identification and implementation of truth orderings
3. Implementation of optimisation algorithm
4. Evaluation (controlled and end-to-end)

Because the use-case scenarios are built on previous multi-disciplinary projects we have data and application problems to start work on day 1. The first iteration of the theoretical skeleton of the approach is clearly worked out, which will allow the PhD student to conduct his research in an iterative way.

RELATED RESEARCH

The existing Semantic Web knowledge representation formalisms have been originally developed for describing crisp and static knowledge about a domain of application, and as such are essentially incapable of dealing with various contextual aspects of knowledge on the Web of Data, nor with a number of phenomena such as uncertainty, vagueness, ambiguity, which are a commonplace. On recognizing these limitations, much of the research effort of the Semantic Web community has been devoted to finding adequate ways of handling the newly identified tasks, resulting in a rich and heterogeneous body of work, e.g. on:

- Reasoning with multiple ontologies, such as Integration and modularization, contextualisation and temporalisation. For the first *e-Connections* [7], [8] and *Distributed Description Logics* [9] are typical examples. The extensive work on *contextualization* mostly provides extensions to DL and OWL-DL languages for representing contexts and context-dependent knowledge explicitly [10-12] as does [13] with temporalization.
- Reasoning with imperfect knowledge: non-standard extensions to the DL languages for representing uncertainty and vagueness, e.g. [14-17].
- Reasoning with multiple RDF graphs: extensions to the RDF framework based on formal aspects of multi-dimensionality, such as *named graphs* [18], *networked graphs* [19] and *multidimensional graphs* [20]. Also related: Ontology matching for semantic interoperability ([21-23] for some overviews, and more specifically for context dependency of matching [24-26]).

Common to those approaches is their attempt to represent the complexity of the information explicitly, and as such put the burden on the shoulders of the user. All this work however indicates that devising semantics for this more complex and contextualised information is extremely hard, even when various aspects of the complexity are treated independently.

The most relevant attempt to introduce a new semantic paradigm has been Emergent Semantics [27], 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. PraSem is orthogonal to emergent semantics: it can be seen as an attempt to explicitly capture as much semantic information as possible. For this, the semantic properties need to be captured in the truth functions, and the complexity be dealt within the optimisation process. For developing truth functions we will study the extensive literature on Complex Systems and graphs [31,32,34].

Very significant work on collective intelligence has proven to be effective for tasks such as network routing [38] and data clustering [39], and using nature-inspired methods has become standard for optimization problems [40,41]. They have also been investigate in the context of the Web of Data recently, such as for storage and querying [42,43], but to the best of our knowledge PraSem is novel in its attempt to use such methods as calculi for explicit semantic systems.

EMBEDDING IN THE RESEARCH ENVIRONMENT

The Knowledge Representation and Reasoning Group at the Department of Computer Science is one of the foremost groups in the adaptation of formal knowledge representation to the Semantic Web. Many types of reasoning (classical [1], approximate [2], diagnostic [3,33], parallelised [5], abductive [4], multi-dimensional [6], goal-directed [4], etc) have been developed and most often been studied in the context of the Web of Data, both in theoretical ways, in standardisation efforts, and in practical applications and implementations.

Our group has a track-record of several high-profile research projects (EU/NWO) on the cross-line of Knowledge Representation and Semantic technology (Open-knowledge, SEKT, Knowledge-Web, LarKC, STITCH, BEST, OntoWeb, WonderWeb, I-Catcher, On-to-knowledge,SWAP). We are in a unique position to combine an in-depth understanding of the Web of Data in many facets through our year-long co-operation in practical applications, and expertise in the development of logical formalisms, and abstractions. A particularly relevant research project has recently been the SOKS project with studied methods from computational intelligence in Semantic Web research.

Renowned experts in our department complement our expertise with necessary in depth knowledge in Collective Intelligence (Eiben), Complex and Distributed Systems theory (van Steen), and Distributed and High performance reasoning (Bal), with access to high performance compute clusters. This will allow us to test scalability of our proposed semantics. The cooperation with our colleagues from scientometrics (van Besselar) will be the continuation of a previous successful joint project.

6B. APPLICATION PERSPECTIVE

The Future Web will be ubiquitous and information on it abundant. Multi-dimensional, context dependent, dynamic and personalised implicit information is continuously derived and updated, where appropriate security arrangements over locally hosted data will ensure full privacy.

The technical and societal impact of the Web of Data is potentially huge as long as this distributed information can be combined in reasonably correct and well-understood ways. The complexity of the Web of Data is growing so dramatically that the current formalisms can become a hinder, with people deserting the well-understood modelling languages. New semantics for existing representation languages are needed so that users can employ (existing) semantic technology in an adaptive, emerging and dynamic, but well-understood way.

Rsearch in PraSem is intrinsically fundamental; the provided results are mostly theoretic and not immediately applicable in practice, and can even provide negative results. But the impact can be significant if Pragmatic Semantics will eventually be adapted as a formalism for interoperability of applications on the Web of Data.

7. DESCRIPTION OF THE PROPOSED PLAN OF WORK

PLANNING

The research will be conducted in three phases:

Phase 1: (12 months. Literature study, state-of-the-art and prototyping). The first 12 months of the research are fully specified in this proposal. After familiarizing herself with the state of the art in the most relevant fields of Knowledge Representation on the Semantic Web, multi-objective optimisation and nature-inspired calculation, and the application scenarios, the PhD student will start developing instantiations of pragmatic semantics based on the proposal in [27]. This will lead to the development of a prototypical implementation of the core reasoning functionality for a variety of specific instantiations of the formalism.

Phase 2: (24 months. Novel methods, theory development and evaluation). Based on the prototype the evaluation methodology must be further developed and exhaustively applied on the initial usage scenario, which will lead to a better understanding of the specific choices made in year 1. This will help developing a generic theory, as well as extending the initial set of methods and implementations. The second and third year will see extensive work on generalisation, consolidation and evaluation.

Phase 3: (12 months: Finishing thesis). The final year is reserved for finishing the PhD thesis, and deployment of the ideas in the usual way.

MILESTONES

M1(month 3): Literature overview, state of the art
M2(month 6): Requirement analysis for 2 use-case scenarios
M3(month 9) Efficient implementation of a generic multi-objective optimisation algorithm
M4(month 12. Pilot): A prototypical implementation of an integrated system providing semantic interoperability based on pragmatic semantics for chosen use-case scenarios.
M5(month 15) A thorough evaluation of proposed formalism and current implementation
M6(month 18) A collection of truth functions (designed and implemented)
M7(month 24) Novel algorithms for calculating pragmatic closure for chosen instantiations.
M8(month 30) Comprehensive theory of pragmatic semantics
M9(month 36) Collection of software (open source)
M10(month 48) PhD thesis on Pragmatic Semantics.

8.USE OF INSTRUMENTATION

We will use standard desktop equipment provided by the hosting institution. For larger-scale experiments we have access to the VU's DAS3 cluster.

9.LITERATURE

THE FIVE MOST IMPORTANT PUBLICATIONS OF THE TEAM

Christophe Guéret, Shenghui Wang, Paul T. Groth, Stefan Schlobach: Multi-Scale Analysis of the Web of Data: a Challenge to the Complex System's Community. Advances in Complex Systems 14(4): 587-609 (2011)

Shenghui Wang, Stefan Schlobach, Michel C. A. Klein: Concept drift and how to identify it. J. Web Sem. 9(3): 247-265 (2011)

Christophe Guéret, Paul T. Groth, Frank van Harmelen, Stefan Schlobach: Finding the Achilles Heel of the Web of Data: Using Network Analysis for Link-Recommendation. International Semantic Web Conference (1) 2010: 289-304

Ian Horrocks, Peeter Patel Schneider, Frank van Harmelen (2003). From SHIQ and RDF to OWL: The Making of a Web Ontology Language. Journal of Web Semantics Vol. 1, No. 1., pp. 7-26.

Stefan Schlobach, Ronald Cornet: Non-Standard Reasoning Services for the Debugging of Description Logic Terminologies. IJCAI 2003: 355-362

REFERENCES

[1] I. Horrocks, Patel P. Schneider, and F. van Harmelen. From shiq and RDF to OWL: The making of a web ontology language. Journal of Web Semantics, 1(1):7–26, 2003.

[2] P. Groot, A. ten Teije, and F. van Harmelen. Towards a structured analysis of approximate problem solving: a case study in classification. In Proceedings of the Ninth International Conference on Principles of Knowledge Representation and Reasoning (KR'04), Whistler, Colorado, June 2004.

[3] Schlobach, S. & Cornet, R.(2003): Non-Standard Reasoning Services for the Debugging of Description Logic Terminologies. In Proceedings of IJCAI'03. Morgan Kaufmann.

[4] Szymon Klarman, Ulle Endriss and Stefan Schlobach. *ABox Abduction in the Description Logic ALC.* In *Journal of Automated Reasoning*, 46(1), pp. 43-80, 2011.

[5] E. Oren, S. Kotoulas, G. Anadiotis, R. Siebes, A. ten Teije, and F. van Harmelen. Marvin: A platform for large-scale analysis of semantic web data. In Proceedings of the International Web Science conference, 2009.

[6] Maarten Marx, Szabolcs Mikul´as, and Stefan Schlobach. Tableau calculus for local cubic modal logic and it's implementation. Logic Journal of the IGPL, 7(6):755–778, 1999.

[7] O. Kutz, C. Lutz, F. Wolter, and M. Zakharyaschev. E-connections of abstract description systems. Artificial Intelligence, 156(1):1–73, 2004.

[8] Bernardo Cuenca Grau, Bijan Parsia, and Evren Sirin. Combining OWL ontologies using E-Connections. Web Semantics: Science, Services and Agents on the World Wide Web, 4(1):40–59, 2006.

[9] Alex Borgida and Luciano Serafini. Distributed description logics: Assimilating information from peer sources. Journal of Data Semantics, 1:2003, 2003.

[10] Djamal Benslimane, Ahmed Arara, Gilles Falquet, Zakaria Maamar, Philippe Thiran,and Faiez Gargouri. Contextual ontologies: Motivations, challenges, and solutions. In Proceedings of the Advances in Information Systems Conference, Izmir, 2006.

[11] Paolo Bouquet, Fausto Giunchiglia, Frank van Harmelen, Luciano Serafini, and HeinerStuckenschmidt. C-OWL: Contextualizing ontologies. In Dieter Fensel, Katia P. Sycara, and John Mylopoulos, editors, International Semantic Web Conference, volume 2870 of Lecture Notes in Computer Science, pages 164–179. Springer, 2003.

[12] Krzysztof Goczyla, Wojciech Waloszek, and Aleksander Waloszek. Contextualization of a dl knowledge base. In Diego Calvanese, Enrico Franconi, Volker Haarslev, Domenico Lembo, Boris Motik, Anni-Yasmin Turhan, and Sergio Tessaris, editors, The Proceedings of the International Workshop on Description Logics (DL-2007), volume 250. CEUR-WS.org, 2007.

[13] Zhisheng Huang and Heiner Stuckenschmidt. Reasoning with multi-version ontologies: A temporal logic approach. In International Semantic Web Conference, pages 398–412, 2005.

[14] Thomas Lukasiewicz and Umberto Straccia. Managing uncertainty and vagueness in description logics for the semantic web. Web Semantics, 6(4):291–308, 2008.

[15] Guilin Qi, Jeff Z. Pan, and Qiu Ji. Extending description logics with uncertainty reasoning in possibilistic logic. In ECSQARU ’07: Proceedings of the 9th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, pages 828–839, Berlin, Heidelberg, 2007. Springer-Verlag.

[16] Umberto Straccia. A fuzzy description logic for the semantic web, 2005.

[17] Thomas Lukasiewicz and Umberto Straccia. Managing uncertainty and vagueness in description logics for the semantic web. Web Semantics, 6(4):291–308, 2008.

[18] Jeremy J. Carroll, Christian Bizer, Pat Hayes, and Patrick Stickler. Named graphs, provenance and trust. In WWW ’05: Proceedings of the 14th international conference on World Wide Web, pages 613–622, New York, NY, USA, 2005. ACM

[19] Simon Schenk and Steffen Staab. Networked graphs: a declarative mechanism for sparql rules, sparql views and rdf data integration on the web. In WWW ’08: Proceeding of the 17th international conference onWorldWideWeb, pages 585–594, New York, NY, USA, 2008. ACM.

[20] Manolis Gergatsoulis and Pantelis Lilis. Multidimensional RDF. In Robert Meersman and Zahir Tari, editors, On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE, pages 1188–1205, 2005.

[21] AnHai Doan and Alon Y. Halevy. Semantic integration research in the database community: A brief survey. In AI Magazine, 26(1), 2005.

[22] Yannis Kalfoglou and W. Marco Schorlemmer. *Ontology mapping: The state of the art.* In Yannis Kalfoglou, W. Marco Schorlemmer, Amit P. Sheth, Steffen Staab, and Michael Uschold, editors, *Semantic Interoperability and Integration,* volume 04391 of Dagstuhl Seminar Proceedings. IBFI, Schloss Dagstuhl, Germany, 2005.

[23] Jérôme Euzenat and Pavel Shvaiko. *Ontology Matching*. Springer-Verlag, Berlin Heidelberg (DE), 2007.

[24] G. Fenza, V. Loia, S. Senatore. *Local semantic context analysis for automatic ontology matching.* In Proceedings of IFSA-EUSFLAT, 2009.

[25] R. Albertoni, M. De Martino. Asymmetric and context-dependent semantic similarity among ontology instances. In Journal on Data Semantics, 2008.

[26] F. Duchateau, Z. Bellahsene, M. Roche. Context-based Measure for Discovering Approximate Semantic Matching between Schema Elements. In Proceedings of RCIS, 2007.

[27] Philippe Cudré-Mauroux: Emergent Semantics. Encyclopedia of Database Systems 2009: 982-985

[28] Christophe Guéret, Shenghui Wang, Paul T. Groth, Stefan Schlobach: Multi-Scale Analysis of the Web of Data: a Challenge to the Complex System's Community. Advances in Complex Systems 14(4): 587-609 (2011)

[29] Stefan Schlobach, Kathrin Denthler, Christophe Gueret. Pragmatic Semantics for the Web of Data. Submitted to ISWC Outrages ideas track. 2011. Available online: www.few.vu.nl/~schlobac/

[30] Rudi Cilibrasi and Paul Vitanyi (2004). The Google Similarity Distance, IEEE Trans. Knowledge and Data Engineering, 19:3(2007), 370–383

[31] M. Newman, A.-L. Barabási, D.J. Watts [eds.] *The Structure and Dynamics of Networks*. Princeton, N.J.: Princeton University Press. (2006)

[32] S. Wang and P. Groth (2010). Measuring the dynamic bi-directional influence between content and social networks. In Proceedings of the 9th International Semantic Web Conference (ISWC2010), Shanghai.

[33] Schlobach, S.: *Diagnosing Terminologies.* In: Veloso, M. M. & Kambhampati, S. (Hrsg.): *Proceedings, of* AAAI 2005, S. 670-675.

[34] Maarten van Steen. *Graph Theory and Complex Networks. An introduction.*2008

[35] Bar-Yam, Y., *Making Things Work: Solving Complex Problems in a Complex World* (Knowledge Press; 1st Edition, 2005).

[36] 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)

[37] Dentler, K., Schlobach, S., Gueret, C.: Semantic web reasoning by swarm intelligence. In: 5th International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS2009) (October 2009)

[38] Dorigo, M., and Caro, G. D. 1998. Antnet: Distributed stigmergetic control for communications networks. Journal of Artificial Intelligence Research 9:317–365.

[39] Deneubourg, J. L.; Goss, S.; Franks, N.; Sendova-Franks, A.; Detrain, C.; and Chr´etien, L. 1990. The dynamics of collective sorting robot-like ants and ant-like robots. In Proceedings of

the first international conference on simulation of adaptive behavior on From animals to animats, 356–363. Cambridge, MA, USA: MIT Press.

[40] Deb, K. *Multi-Objective Optimization using Evolutionary Algorithms* John Wiley & Sons, 2001.

[41] Coello Coello, C. A.; Lamont, G. B. & Van Veldhuizen, D. A. *Evolutionary Algorithms for Solving Multi-Objective Problems* Springer, 2007.

[42] Hannes Mühleisen, Tilman Walther, Robert Tolksdorf: *Multi-level indexing in a distributed self-organized storage system.* IEEE Congress on Evolutionary Computation 2011: 989-994

[43] Eyal Oren, Christophe Guéret, Stefan Schlobach: *Anytime Query Answering in RDF through Evolutionary Algorithms.* International Semantic Web Conference 2008: 98-113

[44] Zhisheng Huang, Frank van Harmelen, Annette ten Teije: *Reasoning with Inconsistent Ontologies.* IJCAI 2005: 454-459

[45] Vilfredo Pareto. *The New Theories of Economics.* Journal of Political Economy *5*: 485–502.

10.REQUESTED BUDGET

We request funding for a PhD student for 4 years, and the additional bench-fee. There are no additional funds required for travelling, apparatus or software.