氏　　　　　名　　小出　誠二

学位（専攻分野）　博士（情報学）

学 位 記 番 号　　総研大甲第 1425 号

学位授与の日付　　平成２３年３月２４日

学位授与の要件　　複合科学研究科　情報学専攻
　　　　　　　　　学位規則第６条第１項該当

学 位 論 文 題 目　THEORY AND IMPLEMENTATION OF OBJECT
　　　　　　　　　ORIENTED SEMANTIC WEB LANGUAGE

論 文 審 査 委 員　　主　　査　　　　教授　　　武田　英明
　　　　　　　　　　　　　　　　　　教授　　　佐藤　健
　　　　　　　　　　　　　　　　　　准教授　　COLLIER, Nigel
　　　　　　　　　　　　　　　　　　准教授　　大向　一輝
　　　　　　　　　　　　　　　　　　教授　　　山口　高平　慶應義塾大学

RDF is an assertional language intended to be used to express propositions using precise formal vocabularies and its syntax is applicable to the World Wide Webs. RDF Schema (RDFS for short) is an semantic extension of RDF and it provides a minimum type system to describe web ontologies. OWL Web Ontology Language is a language for defining and instantiating Web ontologies. These three languages for Semantic Webs are intended to be integrated in Semantic Web Layered Architecture, namely OWL was planned to be realized on top of RDF and RDFS. However, this intention is not accomplished and it seems to be coming apart more and more. The objective of this doctoral study is recovering the language integration and provides a unified language system for Semantic Webs.

Firstly, the semantics of RDF, RDFS, and OWL are investigated in common semantics based on Tarskian denotational semantics. Unfortunately, the formal way of describing semantics in W3C recommendations is different between RDF(S) and OWL. RDF semantics is formalized based on Tarskian denotational model theory and RDFS is extended in the same framework, but OWL semantics is mainly described in the way called Direct Model-Theoretic Semantics that is appropriate for describing Description Logics and OWL DL. However, this different formalization has become to make it difficult to understand both languages in common views, and amounted to that RDF and OWL are apart from each other. In this dissertation, an overview of RDF semantics is given in the way described in the RDF documents of W3C Recommendations. Then, OWL semantics is also investigated and formalized based in the same way as RDF with referencing the description of OWL specifications in OWL Direct-Model Semantics in the documents of W3C Recommendations. Due to our semantic web language system is built on top of Common Lisp Object System, CLOS semantics and its computational model is also discussed. The semantic gap between OWL and object oriented languages are also pointed out.

Secondly, RDF semantics is realized on top of CLOS by straightforward mapping of RDF graph to CLOS objects such as a start node of graph to a CLOS object, an edge in graph to a slot-name, and an end node to a slot-value. RDFS class/instance relation is mapped to that in CLOS, and RDFS class/superclass relation is mapped to that in CLOS, because the semantics of RDFS is analogous to the semantics of CLOS. The problems arising from such straightforward mapping for RDF and RDFS are discussed and solved in the realization of SWCLOS. Then, all OWL features are implemented on top of RDF(S) by CLOS with preserving RDF(S) semantics. We distinguish substantial sorts and non-substantial sorts in ontology description, and procedural subsumption computation algorithm for OWL Full is developed. The system is named SWCLOS from the acronym of Semantic Web Common Lisp Object System.

Thirdly, the efficiency of SWCLOS implementation is tested by Lehigh University Benchmark, and the result showed the comparable performance to other OWL reasoners. SWCLOS returned correct answers for all LUBM queries, whereas it is reported in the benchmark report that two reasoners except one returned wrong answers for some queries. Some distinctive benchmark results are analyzed and improvements are achieved by several different engineering methods. The metamodeling capability of SWCLOS is also demonstrated in examples of SWCLOS metamodeling programming.

Through this study, we obtained deep understanding of semantics on RDF, RDFS, and OWL. The language integration was not so easy because of the semantic disparity between RDF(S) and OWL DL. For example, the subsumption in class hierarchy is weak in RDFS but strong in OWL DL. The semantics of OWL DL class is akin to set theory, but the semantics of RDFS class is based on but dissent from set theory, rather it is close to frame systems. The semantics of RDF(S) is basically categorized into higher order logic but OWL DL stays in first order logic. RDFS allows the membership loop and enables metamodeling of ontology but OWL DL cannot accept the membership loop and does not allow metamodeling. Entities in RDF universe stand in the Unique Name Assumption (UNA) for graph node but entities in OWL universe does not stand in UNA for objects in ontology. RDF semantics is not developed up to Open World but it is assumed in OWL semantics. These highly technical issues must be discussed and settled in order to integrate RDF(S) and OWL. The solution for membership loop, weak/strong subsumption, and non-Unique Name Assumption, Open World Assumption are addressed and implemented in SWCLOS.

In addition to these differences of semantic foundation of languages, what is worse, a misunderstanding is involved on the interpretation of RDF semantics in the documentation of OWL 1 for the RDF compatibility of OWL. An excessive RDF modeling is introduced into the RDF-compatible OWL theory with the name 'comprehension principle' and it is used as the pretext of criticizing RDF that 'comprehension principle' allows the paradox to invade upon systems. Such theoretical disorders in Semantic Web community are also discussed in order to rescue OWL Full theory from the theoretic disorder. It deserves to know that OWL 2 specifications eliminated the term 'comprehension principle' from the documentation of W3C, and elements of RDF are disappeared in the end.

As its name implies, SWCLOS is not based on a logic system but based on Common Lisp Object System. It is semantically an amalgamation of CLOS and OWL on top of RDF; nevertheless it still conforms to object-oriented paradigm as programming language. It is the reason why we call it object-oriented semantic web language. The ground of enabling SWCLOS can be summarized as follows. First, the subsumption of CLOS is the same as the subsumption of RDF(S), and the structure of hierarchy and ordering of CLOS classes is the same as the structure of RDF(S). The dynamic property of CLOS and the Meta-Object Protocol of CLOS enabled to tailor the semantics of language within the realms of CLOS language. In fact, it was easy to realize RDF semantics on top of CLOS,

because the semantics is almost same except property-centric or object-centric. Then, OWL Full level capability is obtained by pursuing the compatibility to RDF and preserving it.

Thesis proposed the theory for OWL Full, a Semantic Web language, as the extension of RDF model theory and its implementation with the Object-oriented programming language which shows more correct inferences than other implementations of OWL Full still keeping comparative inference speed. The thesis is 159 pages long and consists of nine chapters and two appendixes which are summarized as bellow;

(1) Introduction. The background of the thesis is identified. While OWL Full is proposed as a semantic web language, its semantics is not clear yet. RDF semantics is defined by P. Heyes with set theory while OWL Full semantics is not defined yet. We need OWL Full semantics with non-unique name assumption and open world assumption. The other aspect is that there is good correspondence between OWL Full and Object-oriented programming paradigm. So in this thesis, OWL Full semantics is proposed based on the set theory and then realized it upon the Object-oriented programming language.

(2) Semantics of RDF, OWL and CLOS. In this chapter, the semantics of RDF, RDFS and OWL are investigated based on Tarkian denotational semantics. Firstly, RDF semantics are overviewed in the formal way of Tarkian denotational theory described in the RDF documents of W3C. Secondly, in order to develop OWL Full theory, OWL semantics is also studied based on Tarskian denotational semantics with referencing the description of OWL specifications in OWL Direct-Model Semantics. Thirdly, CLOS semantics and its computational model are investigated, and the semantic gap between OWL and object oriented languages are pointed out.

(3) Implementation of RDF, RDFS, and OWL on CLOS. In this chapter, SWCLOS is introduced which is RDF and RDFS implementation firstly on top of CLOS. It is realized by straightforward mapping of RDF graph such that a start node of graph to a CLOS object, an edge in graph to a slot-name, and an end node to a slot-value. RDFS classes/instance relation is mapped to CLOS classes/instance relation. The problems arising from such mapping are discussed and solved. Secondly, all of OWL features are realized on top of RDF(S). After substantial sorts and non-substantial sorts are distinguished, the procedural subsumption computation algorithm for OWL Full is developed. Several OWL specific features are explained with SWCLOS demonstrations.

(4) Benchmark Test by LUBM. The efficiency of implementation is tested by Lehigh University Benchmark (LUBM), and SWCLOS showed the comparable performance to other OWL reasoners reported in Guo et al. SWCLOS replied with correct answers to all LUBM queries, whereas no other reasoners but OWLJessKB replied correctly to all queries.

(5) Demonstration of OWL Full Metamodeling. SWCLOS is the first full-fledged

language as OWL Full processor, in which the capability of metamodeling objects is borrowed from the power of the dynamic and reflective features of Lisp and metamodeling capability of CLOS. We implemented many OWL axioms into CLOS using Meta-Object Protocol (MOP) of CLOS. Whereas the complete freedom of metamodeling certainly results in undecidability, most examples demonstrated as OWL Full undecidability are unreasonably extreme and make no sense from the view of engineering. In this chapter, several metamodeling examples of SWCLOS are shown within the understandable rationale of engineering from our practical experience, and a set of metamodeling criteria that enables SWCLOS to perform ontology metamodeling is addressed.

(6) OWL Full Theory. In this Chapter, OWL Full theory is developed with rearranging and rephrasing previously presented descriptions for RDF, RDFS, CLOS, and OWL in W3C documentations and Chapter 2. The discussion is based on Zermelo-Fraenkel set theory since the W3C recommendation of OWL semantics mentions comprehensive principle that is a foundation of naive set theory that was commenced by Gottlob Frege.

(7) Open World Assumption and Disjointness. Although OWL Full is said to use Open World Assumption (OWA), fully OWA is not desirable as an ontology building language. For this reason, the notion of auto-epistemic local closed world assumption is introduced in this chapter. An agent can introspectively check their knowledge with their extent of capabilities with it.

(8) Related Work. Some related papers are referred on frame-based and Object-oriented OWL systems and RDF and OWL theory.

(9) Conclusion.

The novel contribution of the thesis is summarized as; (a) the proposal of semantics for OWL Full with relation to Object-oriented programming paradigm, (b) the implementation of OWL Full based on this semantics, and (c) identification of metamodeling criteria. In particular, the last point is valuable because it contributes to design theoretically sound ontology.

Seiji Koide gave a 50 minute presentation in English to the exam committee which is open to the public. The major points in the presentation is as follows;

(1)   Goal, motivation and background

(2)   RDF semantics and OWL semantics

(3)   CLOS Semantics and Meta-object protocol

(4)   RDF(S) and OWL implementation

(5)   LUBM benchmark test results

(6)   Metamodeling demonstration

(7)   Advanced topics

There are questions for the presentation which are mostly clarification of the details in

the presentation;

(1)   The lack of data for the comparative system in LUBM loading time. The answer is that it is because of the original paper.

(2)   The over all evaluation of LUBM benchmark test in comparison to the comparative system. The answer is that the simple summarization is difficult because there are different reasons why SWCLOS is slower. The analysis and the engineering is shown in Chapter 4.

(3)   "Theory" is not proved in the thesis. The answer is that "theory" here is not something to prove because no formal semantics is given, so that the theory is proposed to give it to OWL Full.

The committee members all agreed that the thesis is enough worth giving PhD degree.