



Towards Higher-order OWL

Martin Homola¹ · Ján Klůka¹ · Petra Hozzová² · Vojtěch Svátek³ · Miroslav Vacura³

Received: 29 February 2020 / Accepted: 28 May 2020 / Published online: 25 June 2020
© Gesellschaft für Informatik e.V. and Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

We summarize our ongoing endeavour towards proposing a suitable higher-order description logic that could serve as the semantic foundation for higher-order OWL, similarly to *SR_{OTQ}* serving as the semantic foundation of regular OWL.

Keywords Description logics · OWL · higher-order logic

1 Introduction and motivation

At the beginning there was a formally funded mobility project *LAAOS: Logical Aspects of Adaptable Ontological Schemas*, that our groups at Comenius University in Bratislava and University of Economics, Prague won for two years (2012–2013). A key problem addressed was rooted in the observation that different ontologies may be used to define schemas for data sets in the same domain; and the selection of the actual ontology used with a given data

set may be based on different representational or reasoning complexity requirements, or by other factors. However, sometimes it is desired to translate data from one schema to another, or to integrate data sources pertaining to different schemas. Our focus in the project was on documenting the differentiating factors and heterogeneities between ontologies (which later lead to the concept of *background modeling* [21]), and applying ontology transformation patterns to translate and integrate data in such cases. For more details on LAAOS refer to the published overview of its immediate results [6].

This work was supported from the Slovak–Czech bilateral project LAAOS supported by APVV under no. SK-CZ-0208-11 on the Slovak side and MŠMT under no. 7AMB12SK020 on the Czech side. The work was consecutively further supported by Slovak National VEGA projects no. 1/1333/12 and 1/0778/18 and on the Czech side by the EU ICT FP7 project no. 257943 (LOD2) and by the CSF project no. 18-23964S.

During the LAAOS project we realized that many times the developers of LOD vocabularies [1] apparently intend to express higher-order classes and relations. For example, they may want to classify existing classes into (meta) classes, like saying that the tiger (which is a class, *Panthera tigris*) belongs to the meta class *Endangered_Species*; or allowing classes to be used as subjects or as objects of a property, like saying that Yo-Yo Ma's *primary_instrument* is the *Cello*. Such modelling practice is often referred to as *metamodelling*.

✉ Martin Homola
homola@fmph.uniba.sk

Ján Klůka
kluka@fmph.uniba.sk

Petra Hozzová
petra.hozzova@tuwien.ac.at

Vojtěch Svátek
svatek@vse.cz

Miroslav Vacura
vacuram@vse.cz

However, OWL [5] does not really allow for this (notably, RDFS does), if we disregard punning which has some problems as we discuss below. For this reason the authors of LOD vocabularies often resort to approximate modelling and stay within the first-order bounds. For instance, the latter example above is real; it comes from Music Ontology [20], where the property *primary_instrument* takes values from the MusicBrainz Instrument Taxonomy in which the types of instruments are represented as individuals.

¹ Comenius University in Bratislava, Mlynská dolina, 842 48 Bratislava, Slovakia

² Vienna University of Technology, Favoritenstraße 9–11, Vienna 1040, Austria

³ University of Economics, Prague, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic

Apart from Music Ontology, we have documented cases of purported classes modelled as individuals, e.g., in Military Ontology¹, GoodRelations² (now also integrated into Schema.org³ including some of these individuals), and other LOD vocabularies.

Even in wider ontology practice apart from LOD vocabularies, Gu et al. [10] surveyed the actual use of metamodelling in OpenCyc, SUMO, FMA, CHEBI, GO, NCI, and OGG ontologies, and have documented that there are thousands of cases when the same names are used both as individuals and as classes and hundreds of cases when the same names are used both as individuals and as properties.

Another example when higher-order constructs may be needed is if one wished to integrate two data sources pertaining to two ontologies which are first-order, but same entities are treated as individuals in one of them and as classes in the other. For example, the instances of the `Instrument` class of Music Ontology⁴ are in fact intended to be subsumed under a taxonomy of instruments, such as the MusicBrainz instruments taxonomy⁵ which is expressed in SKOS⁶. In this representation, each instrument is an individual, associated with others using the `skos:broader` and `skos:narrower` properties.

This apparent demand for higher-order constructs lead us to investigate a suitable description logic that could provide the semantic foundation for higher-order OWL, as much as *SRQLQ* does this job for regular OWL [5]. This is an ongoing work, which was partly covered by two national projects with broader but related scope on the Slovak side. We summarize our main results in this report.

2 Foundations and semantics

A basic requirement on a logic suitable for our intended application is the syntactic possibility to use classes and properties *as individuals*, i.e., in the usual position of individuals in expressions such as axioms, assertions, and queries. This requirement is already fulfilled by the *punning* feature of OWL 2, which allows any name to be used as an individual, a class, and a property in an ontology.

However, the uses of a name in the different roles are not semantically related in OWL 2, since the name has three interpretations that do not influence one another. Punning thus does not satisfy the property we came to call *intensional*

regularity: If two names A and B in a KB \mathcal{K} are equal as individuals, they are equal as classes, i.e.,

$$\mathcal{K} \models A = B \implies \mathcal{K} \models A \equiv B.$$

This property ensures that when we are, for instance, integrating the international and the Slovak zoological taxonomy, we can assert that the international binomial name of the giraffe species equals its Slovak counterpart (`Giraffa_camelopardalis`=`Žirafa_štíhla`; the former is abbreviated as `G._camelopardalis` below) and obtain the expected result that the respective sets of instances of the two species are also equal. It would thus be inconsistent to assert that, e.g., a particular specimen of the species denoted by the international name (`zarafa`: `G._camelopardalis`) is not a specimen of the species denoted by the Slovak name (`zarafa`: `¬Žirafa_štíhla`).

On the other hand, it seems inappropriate to reduce classes to sets of their instances and require that a higher-order description logic should possess the *extensionality* property of set theory:

$$\mathcal{K} \models A \equiv B \implies \mathcal{K} \models A = B.$$

For instance, `G._camelopardalis` is the only extant species of the genus `Giraffa`. If we are concerned with extant specimens only, we might want to assert that the sets of instances of these two classes of animals are equal (`G._camelopardalis` \equiv `Giraffa`), but they are still different concepts. In particular, `G._camelopardalis` is an instance of `Species`, while `Giraffa` is an instance of `Genus`, and these two second-order classes are expected to be disjoint (`Genus` \sqcap `Species` $\equiv \perp$).

Finally, the study of LOD vocabularies indicated that *typing* of classes and properties based on the nature of their instances is desirable, since mixed-type classes and properties often conflate several different entities that should be modelled separately.

In summary, the desired DL should support using classes and properties as individuals, intensional regularity, non-extensionality, and typing.

Perhaps the first decidable DL supporting all the requirements except for typing was studied by Motik. In his 2007 paper [16], he proved the OWL Full semantics [19] undecidable, and proposed two alternative semantics for a relaxed DL syntax allowing all names in a KB's vocabulary to be used in all roles. The minimal π -semantics was adopted in OWL 2 as punning. The stronger ν -semantics was intentionally regular and non-extensional. It was a DL adaptation of the semantics of HiLog, a higher-order Prolog proposed by Chen, Kifer, and Warren [4], which was essentially also used as the semantics of RDF(S) [11]. Its core idea is to construct a model of a KB by taking a non-empty domain Δ , interpret each name as a domain element called the name's *intension*,

¹ <http://rdf.muninn-project.org/ontologies/military.html>

² <http://www.heppnetz.de/ontologies/goodrelations/v1.html>

³ <https://schema.org/>

⁴ <http://purl.org/ontology/mo/>

⁵ <http://purl.org/ontology/mo/mit#>

⁶ <http://www.w3.org/2004/02/skos/>

and map each domain element to both a *class extension* (a subset of Δ) and a *property extension* (a subset of Δ^2). When used as a class or property, a name denotes the respective extension of its intension. Motik showed that standard reasoning tasks are decidable under the v -semantics for logics up to *ALCHOIC*, provided the *unique name assumption* is adopted for all names used as properties in the knowledge base.

3 Typed higher-order DL

Our initial proposal for a logic meeting all the requirements, called *TH(SROIQ)* [13], used a HiLog-based semantics and extended it up to *SROIQ*. Typing was supported on the syntactic level, similarly to Pan, Horrocks, and Schreiber's OWL FA [18] (which used extensional semantics): The set of names was partitioned to disjoint subsets of names of different types, where type 0 names denoted individuals, type $i+1$ names denoted classes of entities of type i , and type i, j names denoted properties with type $i+1$ domains and type $j+1$ ranges. Class expressions could only be formed if their subexpressions were of compatible types. Decidability and complexity of reasoning in *TH(SROIQ)* was proved using a translation to *SROIQ* based on ideas by Glimm et al. [9], who devised an approach to metamodel higher-order entities directly in *SROIQ*.

4 Meta modelling instantiation

The results of Glimm et al. [9] that we used in the work on *TH(SROIQ)* indicated that a more expressive and flexible higher-order DL was feasible. Namely, the relation of being an instance of a class could be meta modelled, i.e., made available as the *instanceOf* property in KB expressions. This property allows traversing the class-instance relations, in order to express, e.g., that *Species*, in terms of Carvalho and Almeida's MLT [3], *characterizes Organism*, i.e., all individuals of any species are organisms ($\exists \text{instanceOf}.\text{Species} \sqsubseteq \text{Organism}$), or that a species whose individuals are alive is not extinct ($\text{Species} \sqcap \exists \text{instanceOf}^{\neg}.\text{Alive} \sqsubseteq \neg \text{Extinct}$). Moreover, it can be used to traverse orders in both directions within one class description, e.g., to classify all individuals of the same species as *zarafa* without having to know what *zarafa*'s species is ($\exists \text{instanceOf}.\text{Species} \sqcap \exists \text{instanceOf}^{\neg}.\{\text{zarafa}\}$). Additionally, this expressive description logic, which we named *HIR(SROIQ)* [15], does not have to be typed: it can allow arbitrary names as instances. Nevertheless, we showed how a typed hierarchy in the spirit of *TH(SROIQ)* can be axiomatized within *HIR(SROIQ)*.

Intuitively, the syntax of *HIR(SROIQ)* is an extension of the syntax of *TH(SROIQ)*. The differences are that (i) the set of property names of *HIR(SROIQ)* always contains the name *instanceOf*, and that (ii) in *HIR(SROIQ)*, a type hierarchy is not imposed (each class or property can classify or connect any entities: any individual, class, or even property). The semantics of *HIR(SROIQ)* is also the same as for *TH(SROIQ)* except for *instanceOf*, which is always interpreted as the set of pairs $(i, c) \in \Delta^2$ such that the element i belongs to c 's extension. Decidability and complexity of reasoning in *HIR(SROIQ)* was, again, established based on ideas by Glimm et al. [9].

5 Metamodelling subsumption

We also explored the possibilities of extending our higher-order DL with the *subClassOf* property with fixed semantics. This property allows us to express, e.g., the MTL's *subordination* relationship between the pairs of consecutive taxonomic ranks of *Species*, *Genus*, and *Family*, i.e., the fact that each *Species* is a subclass of some *Genus* and each *Genus* is a subclass of some taxonomic *Family*, by axioms $\text{Species} \sqsubseteq \exists \text{subClassOf}.\text{Genus}$ and $\text{Genus} \sqsubseteq \exists \text{subClassOf}.\text{Family}$. Combined with assertions *zarafa*: *G._camelopardalis* and *G._camelopardalis*: *Species*, these axioms entail that *zarafa* is an instance of some *Family*, i.e., *zarafa*: $\exists \text{instanceOf}.\text{Family}$.

We studied four possible versions of the interpretation of *subClassOf* taking into account two factors: (a) whether the interpretation is set-theoretic, i.e., whether the pair (c, d) belongs to the interpretation iff the extension of c is a subset of the extension of d , or whether (c, d) belonging to the interpretation entails that the extension of c is a subset of the extension of d , but not necessarily vice versa; and (b) whether the condition from (a) holds only for intensions of named classes, or for all intensions in the domain, including those not referred to by any class name. The preferred combination of these options is not apparent. While the set-theoretic interpretation might seem intuitive, it leads to counterintuitive consequences, such as empty classes (e.g., *Unicorn*) being subclasses of all classes (e.g., $\text{Unicorn} \sqsubseteq \text{G._camelopardalis}$). Moreover, enforcing the (non-)set-theoretic property only for named classes or for all classes does not have clear advantages or disadvantages.

We denoted the resulting logics by *HIRS_{*}(SROIQ)*, with different subscripts for different combinations of the (a) and (b) options. We showed [14] that *subClassOf* for three of these interpretations (both non-set-theoretic ones and the set-theoretic one for named classes only) can be axiomatized within the logic itself, and that the approach of Glimm et al. [9] corresponds to the non-set-theoretic interpretation for named classes. The question of whether *HIRS_{*}(SROIQ)* with

set-theoretic interpretation of `subClassOf` for all classes is decidable remains open. We did not find a suitable axiomatization or extension of the tableaux algorithm for this logic, nor were we able to prove that such a logic is undecidable.

6 Related work

Other works looking into enabling metamodelling in DL include those of Motik [16], De Giacomo et al. [7] and Gu et al. [10], who also rely on HiLog-style semantics, similarly to our work. Giordano and Policriti [8] took a different non-extensional approach based on hereditarily finite sets in their work on the ALC^{Ω} DL with the power set constructor.

The second group of works rely on Henkin's general semantics [12], including Pan et al. [18] and Motz et al. [17]. This semantics is extensional, which may not be desired in some metamodelling use cases as we explained above.

The study of LOD vocabularies within the LAAOS project has also lead to a specification of a new ontology background modelling language called PURO [21] that could express the intended, or background, model of LOD vocabularies, including higher-order types and n -ary relations. One of its purposes has been to annotate each entity in a LOD vocabulary with a term of the PURO language denoting its intended ontological category (such as plain object, relation, type of types). Such annotations document the vocabulary's author's intentions. The PURO language does not directly correspond to any of the DLs described above, but these DLs aim to capture its semantics at least partly (notably the typed variants).

A similarly motivated axiomatic theory for conceptual modelling with higher-order types, called MLT, was proposed by Carvalho and Almeida [3]. MLT focuses on relationships of types of different orders, such as characterization and subordination described in previous sections. Brasileiro et al. [2] developed a partial implementation of MLT in OWL and SPARQL, to be used as a basis of multi-level ontologies. Similarly to PURO, the DLs discussed above do not correspond directly to MLT, though both $HTRS_*(SROIQ)$ and ALC^{Ω} capture some of its elements.

7 Future directions

PURO and MLT point out directions for expanding the expressivity of higher-order DLs in future research. The work on ALC^{Ω} might provide a promising basis for further investigation of the last unresolved semantics of the $HTRS_*(SROIQ)$ logics.

References

- Bizer C, Heath T, Berners-Lee T (2009) Linked data-the story so far. *Int J Semant Web Inf Syst* 5(3):1–22
- Brasileiro F, Almeida JPA, Carvalho VA, Guizzardi G (2016) Expressive multi-level modeling for the semantic web. In: P. Groth, E. Simperl, A. Gray, M. Sabou, M. Krötzsch, F. Lecue, F. Flöck, Y. Gil (eds.) *The Semantic Web – ISWC 2016: 15th International Semantic Web Conference, Kobe, Japan, October 17–21, 2016, Proceedings, Part I*, pp. 53–69. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46523-4_4
- de Carvalho VA, Almeida JPA (2018) Toward a well-founded theory for multi-level conceptual modeling. *Softw Syst Model* 17(1):205–231. <https://doi.org/10.1007/s10270-016-0538-9>
- Chen W, Kifer M, Warren DS (1993) A foundation for higher-order logic programming. *J Log Program* 15(3):187–230
- Cuenca Grau B, Horrocks I, Motik B, Parsia B, Patel-Schneider P, Sattler U (2008) OWL 2: the next step for OWL. *J Web Semant* 6(4):309–322
- Čertický M, Homola M, Křůka J, Serra S, Svátek V, Šimko A, Vacura M, Zamazal O (2014) LAAOS: Logical aspects of adaptable ontological schemas (Project overview). In: *Znalosti 2014*. VŠE, Praha
- De Giacomo G, Lenzerini M, Rosati R (2011) Higher-order description logics for domain metamodeling. In: *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2011, San Francisco, California, USA, August 7–11, 2011*. AAAI
- Giordano L, Policriti A (2019) Adding the power-set to description logics. *Theor Comput Sci*. <https://doi.org/10.1016/j.tcs.2019.10.049>
- Glimm B, Rudolph S, Völker J (2010) Integrated metamodeling and diagnosis in OWL 2. In: *ISWC*
- Gu Z, Cao C, Zhang S (2019) An expressive sub-language of OWL 2 full for domain meta-modeling. In: M. Simkus, G.E. Weddell (eds.) *Proceedings of the 32nd International Workshop on Description Logics, Oslo, Norway, June 18–21, 2019, CEUR Workshop Proceedings*, vol. 2373. CEUR-WS.org
- Hayes PJ, Patel-Schneider PF eds. (2014) *RDF 1.1 Semantics*. W3C Recommendation
- Henkin L (1950) Completeness in the theory of types. *J Symb Log* 15:81–91
- Homola M, Křůka J, Svátek V, Vacura M (2014) Typed higher-order variant of *SROIQ* – why not? In: *DL*
- Kubincová P (2016) Higher-order description logics for metamodelling. Master's thesis, Comenius University in Bratislava
- Kubincová P, Křůka J, Homola M (2016) Expressive description logic with instantiation metamodelling. In: C. Baral, J.P. Delgrande, F. Wolter (eds.) *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016, Cape Town, South Africa, April 25–29, 2016.*, pp. 569–572. AAAI Press
- Motik B (2007) On the properties of metamodeling in OWL. *J Log Comput* 17(4):617–637
- Motz R, Rohrer E, Severi P (2015) The description logic SHIQ with a flexible meta-modelling hierarchy. *J Web Semant* 35:214–234. <https://doi.org/10.1016/j.websem.2015.05.002>
- Pan JZ, Horrocks I, Schreiber G (2005) OWL FA: A metamodelling extension of OWL DL. In: *OWLED*
- Patel-Schneider PF, Hayes P, Horrocks I (2002) OWL Web Ontology Language; Semantics and Abstract Syntax. Recommendation. W3C. <http://www.w3.org/TR/owl-semantics/>
- Raimond Y, Abdallah SA, Sandler MB, Giasson F (2007) The music ontology. In: S. Dixon, D. Bainbridge, R. Typke (eds.) *Proceedings of the 8th International Conference on Music Information Retrieval, ISMIR 2007, Vienna, Austria, September 23–27, 2007*, pp. 417–422. Austrian Computer Society
- Svátek V, Homola M, Křůka J, Vacura M (2013) Metamodelling-based coherence checking of OWL vocabulary background

models. In: Rodriguez-Muro M, Jupp S, Srinivas K (eds)

OWLED, CEUR Workshop Proceedings, vol 1080. <http://CEUR-WS.org>