**KI - Künstliche Intelligenz manuscript No.**
(will be inserted by the editor)

# Interview with Prof. Dr. Ian Horrocks, Professor at the Department of Computer Science of the University of Oxford

**Birte Glimm**

Received: date / Accepted: date

Ian Horrocks is a full professor in the Oxford University Department of Computer Science and a visiting professor in the Department of Informatics at the University of Oslo. His research interests include logic-based knowledge representation and reasoning and semantic technologies, with a particular focus on ontology languages and applications. He was an author of the OIL, DAML+OIL, and OWL ontology language standards, chaired the W3C working group that standardised OWL 2, and developed many of the algorithms, optimisation techniques and reasoning systems that underpin OWL applications. He has participated in numerous national and international research projects, and is currently Scientific Director of the EU funded Optique project, which is deploying semantic technologies in the Oil & Gas and Power Generation industries. He is a Fellow of the Royal Society, a member of Academia Europaea, an ECCAI Fellow and a Fellow of the British Computer Society.

*KI: Hallo Ian, thanks a lot for taking the time to give us an interview. Ten years ago the first KI special issue on the semantic web was published. What are, in your opinion, the most important developments of the previous decade?*

Birte Glimm
University of Ulm
Institute of Artificial Intelligence
D-89081 Ulm
Tel.: +49 (0)731 50 24125
Fax: +49 (0)731 50 24119
E-mail: birte.glimm@uni-ulm.de

W3C standards were critical – this is what gave users the confidence to look at and even try out "semantic technologies"; without them (standards), we would not have got to first base. In my opinion (obviously) we got the standards about right (even if we can fight about some of the details), but to be honest, the existence of standards is more important than the details of what is standardised.

"W3C standards gave users the confidence to look at and try out semantic technologies."

Of course it was useful to be able to respond to the interest generated by standardisation with existing robust implementations, but in many ways the standards were themselves the drivers for the development of many of the robust implementations that we see today.

Profiles where also important in helping (potential) users to better understand performance/expressivity trade-offs, and to distinguish between principled language subsets and arbitrary incompleteness. New reasoning procedures/implementations for the profiles also represent an important advance, particularly in the increasing number of applications where the focus is on query answering over large datasets; I am thinking, e.g., of OBDA implementations for OWL QL (e.g., Ontop), triple store based implementations for OWL RL (e.g., Owlim, Oracle's RDF Semantic Graph, RDFox, etc.) and the new generation of EL implementations (e.g., ELK, Snorocket).

*KI: Almost all relevant standards have undergone a revision within the last decade. As co-chair of the W3C*

*OWL working group, what were the important developments for OWL?*

The addition of profiles (as discussed above) was the main substantive change. Another important addition was extensive support for datatypes – these are important in realistic applications, and were largely ignored in OWL 1.0. Also, the quality of the spec was greatly improved so that it could really be used as a specification for implementation. As an example, parsing and species validation of the RDF syntax in OWL 1.0 was specified in a way that was very difficult to understand and did not suggest an obvious implementation method; i.e., according to the OWL 1.0 spec., an RDF graph G is a valid OWL DL ontology if there exists an OWL DL ontology in abstract syntax whose serialisation in RDF equals G. This was a nightmare for implementers, and led to significant (in)compatibility issues. This and much more was cleaned up in OWL 2.0.

*KI: Logic plays an important role in the semantic web standards of the W3C, but is it really relevant in practice?*

I am not sure if logic is important per se. What is important in my opinion is precisely specifying the semantics of the standards so that the correct behaviour of implementations is well defined and verifiable. Logic has proven to be a good way to do that. Without such a specification we have chaos – implementations can do different things while all claiming to be correct. This really does matter to (most) users – without well defined and consistent behaviour, they will lose faith in and stop using "semantic technologies". E.g., in many applications, semantic technologies are being used to augment and/or replace databases (e.g., in the Optique project), and the developers/users of such applications expect to enjoy the benefits of semantic technology **in addition to** and not instead of the benefits of DBs, i.e., they expect precise formal semantics and a guarantee that query answers respect said semantics.

> "Without well defined and consistent [system] behaviour, users will lose faith in and stop using semantic technologies. "

*KI: OWL is often considered too complex for real use cases, at least when used without restrictions. Would a few often used features such as owl:sameAs not suffice?*

Clearly this is not the case in all applications. A good example is the SNOMED ontology, where they currently use only the EL profile. However, if we look at the modelling, we can easily see that some parts of the model are completely broken, and would lead to large numbers of inconsistencies if it were not for the fact that the model as a whole is massively under-constrained (the modelling of junction areas such as "groin" is a well known example). Fixing such errors seems to require (at least) the full expressive power of OWL. Even in applications that require less expressivity, it does not seem to be the case that the **same** few features are appropriate to a wide range of applications, and one of the things that we know from logic is that a surprisingly small number of features already provides equivalent expressive power to the whole language (see, e.g., the OWL Lite fiasco). This is where the profiles come in: if it is possible to identify a sufficient subset of features that is inside one of the profiles, then one can use more efficient reasoning techniques without giving up correctness – in other words, we can have the best of both worlds. Finally, it is worth pointing out that it is far from clear that selecting a few often used features, even if these features constitute a subset of one of the profiles, would admit a more efficient implementation – as far as I know this has not been demonstrated either theoretically or empirically.

*KI: Database technology seems to play a key role now in the semantic web. Are databases the new reasoners or are specialised algorithms and implementations still needed?*

The attractive features of query rewriting techniques for OWL 2 QL might lead to this impression, but the QL profile is very restricted, and is almost invariably too weak to capture the desired ontological model. Complex mapping rules can provide some compensatory power – but are difficult to develop and maintain – and "approximation" of ontologies amounts to incomplete reasoning.

OWL 2 RL is more powerful, but is still much weaker than the full language. Moreover, the most efficient implementations of RL reasoners typically use specialised algorithms and datastructures and/or exploit specialised (graph) databases.

Exploiting databases becomes even more problematical when we go outside Horn fragments and need to consider multiple models. Recent work on the PAGOdA system has shown how it is possible to develop a hybrid system that exploits an RL reasoner to do much of the work when answering queries w.r.t. an unrestricted OWL 2 ontology, but for some combination of ontology, query and data it will always be necessary to use a specialised reasoner to ensure answer completeness.

*KI: You also have a lot of implementation experience. What are, in your opinion, the most important features for reasoning systems in the semantic web?*

A lot depends on the application. In most applications (perhaps modulo some "web" applications) correctness is of critical importance: if query answers cannot be trusted, then the system becomes useless. In some applications, the ability to explain answers may also be very important, as without this, users may start to doubt the correctness of inferences, particularly if they are unexpected – and these are likely to be the most important and useful inferences.

Of course performance is also important; in particular, it is important that the performance of the system is robust and consistent – users do not like poor performance, but (arguably) even worse is highly unpredictable performance, and small changes in inputs resulting in large changes in performance. This is where the profiles are of particular importance: when reasoning is of high worst-case complexity, it may still be possible to optimise systems so that they exhibit good performance in typical cases, but even small changes to the ontology and/or data can cause such optimisations to fail.

I would not have particularly emphasised usability, whatever that means for a reasoner; I would not have said that databases score very highly on the usability scale, but this does not seem to matter as they are typically invisibly embedded in other applications. I expect the same to apply to reasoning systems – most users will be unaware that the application that they are using is powered by a reasoning system.

The quality of infrastructure supporting ontology engineering and the like is also very important for the uptake of semantic technologies, but perhaps a bit beyond the scope of this "conversation".

*KI: To come back to a question from the interviews ten years ago: Is the semantic web still fiction or reality?*

I recently gave a keynote at ISWC in which I distinguished two perspectives on the semantic web: one in which the emphasis is on using semantic technologies to "improve" the web, and the other in which the emphasis is on the development and application of semantic technologies, with the web being just one potential application. From the former perspective, progress has been relatively modest: it is not very easy to identify examples where semantic technology has had a significant impact, although knowledge graphs and the like could be seen as a recognition of the need to exploit some kind of semantic technology in addition to statistical and machine learning techniques. From the latter perspective (the development and application of semantic technologies), there is a much more convincing success story – semantic technologies are supported in mainstream products (including offerings from the likes of IBM and Oracle), are deployed in large scale applications (e.g., in the Optique project), and their potential is now widely recognised. Having said that, it is clear that we are still some way short of a breakthrough in the application of semantic technologies – but I believe that the next few years could see such a breakthrough.

---

"Semantic technologies have enormous potential for advances and even breakthroughs, and for generating real impact."

---

*KI: What are, in your opinion, the current and coming research trends in the semantic web?*

If I knew stuff like that I would not be wasting my time as a university professor – I would be working towards my second billion. Seriously, the focus has clearly shifted dramatically towards query answering over (large) datasets. OBDA is already well established, but there are still many issues to be addressed, including, e.g., how to deal with equality (sameAs), which is critical for data integration. Effective query answering with more expressive ontologies is also a hot topic, and there is resurgent interest in Datalog/OWL-RL reasoning. Issues surrounding "big data" and data analytics are also of great interest, and many people are starting to think about ways of combining the desirable features of semantic and analytical systems, but exactly how this might work is unclear – this is still a very open area.

*KI: As a concluding question, would you advise young researchers to join the field?*

Of course! I think that it is a great area to work in: there is still enormous potential for advances and even breakthroughs, and for generating real impact. It also provides a very attractive mix of theoretical and applied research: there remain many challenging problems of both kinds, and a longstanding and mutually advantageous tradition of respect and collaboration between theoretical and applied researchers.

*KI: Ian, thank you very much for your time and for this interview.*