A Non-traditional Inference Paradigm for Learned Ontologies*

Vít Nováček

Digital Enterprise Research Institute
National University of Ireland, Galway
E-mail: vit.novacek@deri.org

1 Main Thesis Focus

The purpose of this document is to give an overview of author's prospective doctoral thesis in terms of goals, plans, adopted methodology and current achievements. The thesis' general focus is the Semantic Web, AI, automatic ontology acquisition and reasoning.

When considering ontologies as general knowledge repositories, ideally reflecting substantial amount of information present on the web, it is obvious that developing them purely manually is infeasible task not only due to the extensive size of data, but also due to the highly dynamic nature of the environment. Therefore the need for automated methods of ontology creation and maintenance is well acknowledged in the community. However, there has been no explicit support for automatically learned ontologies in the main branches of research concerning inference in the Semantic Web.

We believe that efforts leading to bridging these two rather disparate lines of research are more than worthwhile and will prove beneficial for both automated ontology development and reasoning, considering the noisy, context-dependent and inconsistent character of mainly unstructured web data we *have to* deal with when making the Semantic Web real. The nature of this knowledge is hard to be captured by traditional (logical) reasoning paradigms that usually require quite extensively (and expensively) specified descriptions in order to allow any usable reasoning. We plan to develop an alternative formal semantics of the Semantic Web data and implement respective reasoning tool prototype that would be able to deal with this situation better in the context of ontology learning. This is reflected in the tentative thesis' title *A Non-traditional Inference Paradigm for Learned Ontologies*.

2 Motivations, Addressed Tasks and Proposed Solutions

Within implementation of the thesis topic prototype, we adhere to these required features:

- the ability to refine the learned knowledge on the fly by incorporation of a specific reasoning paradigm and respective tools;
- a query-processing mechanism able to infer valuable and useful knowledge from learned ontologies by tools basing on the same reasoning paradigm;

* This work has been supported by Academy of Sciences of the Czech Republic, 'Information Society' national research program, the grant AV 1ET100300419 and by the EU IST 6th framework's Network of Excellence 'Knowledge Web' (FP6-507482).

– a query-transformation layer that would allow to interface the system with the Semantic Web standard tools and languages (for evaluation and inter-operation purposes);
– a knowledge-transformation layer that would allow to export the knowledge in the Semantic Web standards (again, for evaluation and inter-operation purposes).

In the following overview of the respective tasks and solution sketches, we base on our ANUIC (*Adaptive Net of Universally Interrelated Concepts*) framework for representation of learned fuzzy ontologies [14, 15].

2.1 Task SW-1 (*reasoning support for ontology acquisition*)

To the best of our knowledge, there has been little effort dedicated to the development of methods that could refine a learned ontology dynamically on the fly by means of specifically tailored reasoning procedures. If a basic foundational and precise ontology for the given domain has been developed, it can be used as a top-level "seed" model for our ANUIC framework. The assertions (with weights initially set to 1.0) in this general seed will help refining the more specific dynamic insertions within ontology learning process (e.g. by decreasing weights of learned assertions that are inconsistent according to the seed ontology). The documents processed by ontology learning can contribute to the refinement of the weights by themselves – if there are certain more trusted or domain-relevant documents, the weights of the assertions learned from them should be favoured.

This will be accompanied by a mechanism of propagation of the weight changes in the vicinity of the influenced nodes in the semantic network induced by the ontology. Note that there will be no restriction on the propagation – even the seed ontology can be eventually changed if the empirical character of the field is different. The application of inherent rules (the idea introduced in the next section) will play as significant role as the seed model in the direct inference support of the acquisition process.

Evaluation of this task is quite straightforward – we can compare the ontologies learned with the inference support with ontologies learned by the same methods without the inference. Appropriate evaluation measures can be adapted according to [9, 3]. One possible option is to identify the differences and present them to potential users of the ontology and/or to an evaluation committee, elicitating the reasonability and usability of extensions/retractions caused by the reasoning process when compared to the "purely learned" ontology.

2.2 Task SW-2 (*reasoning with learned ontologies*)

The ontology reasoning research in the Semantic Web has been focused mainly on the development of rigorous knowledge representation models and related formalised procedures of logical inference. However, the models in question (namely OWL [1] ontologies) require an indispensable amount of expert human intervention to be built and maintained. This makes the knowledge management based on this kind of explicit representation very expensive, especially in dynamic and data-intensive domains (e.g. medicine), or even infeasible, if the experts are not always available (e.g. semantic desktop).

The scalable ontology learning methods can overcome the problem of large domains. Moreover, automatic bottom-up knowledge acquisition prevents the possible bias in hand-crafted ontologies. The price we have to pay is that we must be able to deal with the less complex, noisy, possibly imprecise and very probably inconsistent knowledge then. Nonetheless, there could be implicit knowledge worth to infer even in the learned ontologies if there is a substantial amount of data in them. A possible way to an alternative approach to reasoning with learned ontologies rests with the development of a new kind of "loose", yet formal semantics. This semantics will support both refinement of ontology learning results (Section 2.1) and full-fledged reasoning with and querying of the learned ontologies themselves.

The semantics has been worked out in three levels that are jointly contributing to the process of formal interpretation of the learned content[1]:

1. **Declarative** semantics reflects direct meaning of learned knowledge *declared* in the ANUIC network of fuzzy modelling primitives. Interpretation of a node at this level is based on fuzzy intersection of sets induced by ranges of its properties (this interpretation is crucial for establishment of fuzzy analogical mappings, among other things). We further plan to design a natural extension of the ANUIC model by simple IF-THEN rules treated exactly in the same dynamic manner as the relations between ANUIC concepts.
2. **Procedural** semantics comprises the formal aspects of *procedures* of rule execution and analogy retrieval, mapping and transfer in the underlying model. We plan to incorporate the AI methods of heuristic reasoning [16, 10] into the engine based on the improved fuzzy ANUIC model. Very valuable concept in this respect is the notion of analogical reasoning [12] and its fuzzy extension [2]. The latter can be further developed in the scope of our work with different notions of fuzzy similarity [22, 11]. For the implemented inference engine, we have to provide a respective query-transformation layer in order to interface our system with other Semantic Web frameworks and standards.
3. **Interlocutive** semantics allows to further specify and/or refine meaning of stored knowledge in dynamic interaction with users (human or artificial agents – e.g. other ANUIC-based reasoners fed with different data in similar or otherwise relevant domains).

The evaluation of this task remains more or less open problem for now. However, besides measuring the computational efficiency of the inference, we could formalise a measure of "usefulness" of answers to certain types of queries and compare our system to the similar ones in an application-oriented assessment trial.

3 Current Achievements

At this time, an automated ontology acquisition platform *OLE* (*Ontology LEarning*) has been developed before and within the work on the thesis topic itself. *OLE* processes natural language English documents (in plain text, HTML, PDF or PostScript) and extracts an ontology from them. It makes use of NLP and machine learning techniques. An ANUIC (*Adaptive Net of Universally Interrelated Concepts*) model has been proposed and initially implemented for the fuzzy representation of learned ontologies in

[1] Only very brief description is given here, partially also due to space restrictions. The topic of the three-level formal semantics is currently under thorough development within a conference submission.

OLE. The progress of this work has been documented in several refereed papers[2] and presented by the author of this document at the respective events.

A technique of so called conceptual refinement improving the results of initial ontology extraction methods has been proposed and implemented for the task of taxonomy acquisition. Under a certain interpretation, it boosts the precision of taxonomy acquisition methods by more than 150%. The preliminary results of this work form the major recently published or accepted achievements [14, 15] and were presented by the author of this document at the ESWC 2006 conference (an ICEIS 2007 presentation to come in June, 2007). This initial proposal and implementation of the natural and intuitive mechanism coping with autonomous assignment of fuzzy relevance measures to the general learned relations (which has been considered as an open problem in this respect [19]) forms the most tangible and strongly related basic groundwork of the thesis, aimed at reasoning in the proposed ANUIC model. Current progress is continually documented at the project's webpage[3].

4 Related Work

There are methods refining the ontology after the learning process, using external reference and pruning [5]. However, there are generally no suitable external resources for many practical domains, therefore our tool is more universal in this respect. Some approaches try to connect ontology learning and reasoning by transforming the learned knowledge into a shape acceptable by the "traditional" inference mechanisms. The *Text2Onto* tool removes inconsistent knowledge from the learned ontologies [7] in order to allow usual precise OWL reasoning. The approach in [8] translates ontologies acquired by application of Formal Concept Analysis into FOL formulas, which is even more simplistic. These approaches leave vast amount of the sense of the learned knowledge unrecognised (e. g. possible different contexts induced by consistent subsets, structural properties of the knowledge, implicit relations between concepts, etc.).

In [17], a fuzzy relational model of ontology is introduced. However, it is only very simple and IR-oriented one, with no proper semantics generally applicable in other domains. [6] focuses on mining knowledge from databases and uses for example fuzzy rules to refine the resulting ontologies. But the authors' concrete approach to this topic is rather unclear and the formal semantics is lacking again. There is an indirectly related research in fuzzy OWL [20] and fuzzy DL reasoning [21]. However, these approaches still exploit the "traditional" logics based knowledge representation, which we find inappropriate for reasoning with learned ontologies. AI methods of heuristic [16, 10] or analogical [12, 18] reasoning present alternative paradigms that have, however, not been connected to a mechanism of automatic real-world knowledge acquisition. This is a practical disadvantage our approach aims to tackle (among other things).

5 Selected Application Domains

Following the **medicine** use cases specified in [13, 4], the implementation of our framework for ontology learning and reasoning could massively help in the processing of the

[2] See `http://www.muni.cz/people/4049/publications` for the full list of author's publications to date.

[3] See `http://nlp.fi.muni.cz/projects/ole` – the top Google™ result of the "*ontology acquisition*" query on December 15, 2006; a web interface to the system libraries is present there as well.

dynamically changing medical knowledge. After initial definition of the seed model, ontologies learned by our tool from the natural language in medical records and even from the databases (after a preprocessing) can integrate the newly coming knowledge with the current facts on a single formal and technical basis. Moreover, the efficient and robust reasoning in our model can support the everyday decision process of medical experts in purely automatic way, utilising even data that have not been covered by formal medical manually developed ontologies.

The **semantic desktop** domain is related to new topics that have appeared recently within the major Semantic Web and AI research activities like *CALO* project[4] in USA and/or *NEPOMUK* project[5] in EU. The main aim of the projects is the development of an intelligent layer on the top of the current personal desktop systems. Possible application of our work in the scope of the semantic desktop research efforts is especially in the field of dynamic and automatic knowledge acquisition from the "raw" data. The model and reasoning paradigm we plan to develop could help in efficient semi-automatic discovery of implicit relations in the personal data and thus improve the process of their semantic re-organisation, meta-data annotation and querying.

6 Conclusion and Future Work

We have presented our current results and a vision of our doctoral thesis in the context of the Semantic Web and AI. Some of the missing links in the contemporary research have been identified. We have argued importance of the respective research questions and analysed the tasks that can fill in the gaps then. Possible solutions and evaluation methods have been roughly outlined. Examples of concrete application domains have been sketched, showing the practical relevance of the topic.

The work on the thesis was formally started in March, 2006. Supposed term of the thesis submission is the beginning of the year 2009. We plan to deliver the complete elaboration of the proposed ANUIC uncertain KR model and its semantics by the end of the year 2007, together with respective extension of the ontology learning framework. During the year 2008, we plan to devise and implement basic set of rule-based heuristic and analogical reasoning methods for the prototype and evaluate it, summing up the results in the thesis.

References

1. S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. A. Stein. *OWL Web Ontology Language Reference*, 2004. Available at (February 2006): `http://www.w3.org/TR/owl-ref/`.
2. B. Bouchon-Meunier and L. Valverde. A fuzzy approach to analogical reasoning. *Soft Computing*, 3:141–147, 1999.
3. C. Brewster, H. Alani, S. Dasmahapatra, and Y. Wilks. Data driven ontology evaluation. In *Proceedings of LREC 2004*, 2004.
4. Marco Eichelberg (edited by). Requirements analysis for the ride roadmap. Deliverable D2.1.1, RIDE, 2006.
5. A. Gangemi, R. Navigli, and P. Velardi. Corpus driven ontology learning: a method and its application to automated terminology translation. *IEEE Intelligent Systems*, pages 22–31, 2003.

[4] See `http://caloproject.sri.com/`.
[5] See `http://nepomuk.semanticdesktop.org`.

6. Paulo Gottgtroy, Nikola Kasabov, and Stephen MacDonell. Evolving ontologies for intelligent decision support. In Elie Sanchez, editor, *Fuzzy Logic and the Semantic Web*, Capturing Intelligence, chapter 21, pages 415–440. Elsevier, 2006.
7. Peter Haase and Johanna Völker. Ontology learning and reasoning - dealing with uncertainty and inconsistency. In Paulo C. G. da Costa, Kathryn B. Laskey, Kenneth J. Laskey, and Michael Pool, editors, *Proceedings of the Workshop on Uncertainty Reasoning for the Semantic Web (URSW)*, pages 45–55, NOV 2005.
8. Hele-Mai Haav. An ontology learning and reasoning framework. In Yasushi Kiyoki, Jaak Henno, Hannu Jaakkola, and Hannu Kangassalo, editors, *Information Modelling and Knowledge Bases XVII*, volume 136 of *Frontiers in Artificial Intelligence and Applications*, pages 302–309. IOS Press, 2006.
9. J. Hartmann, P. Spyns, A. Giboin, D. Maynard, R. Cuel, M. C. Suarez-Figueroa, and Y. Sure. Methods for ontology evaluation (D1.2.3). Deliverable 123, Knowledge Web, 2005.
10. Jerry R. Hobbs and Andrew S. Gordon. Toward a large-scale formal theory of commonsense psychology for metacognition. In *Proceedings of AAAI Spring Symposium on Metacognition in Computation*, pages 49–54, Stanford, CA, 2005. ACM.
11. Zsolt Csaba Johanyák and Szilvester Kovács. Distance based similarity measures of fuzzy sets. In *Proceedings of SAMI 2005*, 2005.
12. Boicho Kokinov and Robert M. French. Computational models of analogy making. In L. Nadel, editor, *Encyclopedia of Conginitve Science*, volume 1, pages 113–118. Nature Publishing Group, London, 2003.
13. Lyndon Nixon and Malgorzata Mochol. Prototypical business use cases (D1.1.2). Deliverable 112, Knowledge Web, 2004.
14. V. Nováček and P. Smrž. Empirical merging of ontologies – a proposal of universal uncertainty representation framework. In *LNCS*, volume 4011, pages 65–79. Springer-Verlag Berlin Heidelberg, 2006.
15. Vít Nováček. Imprecise empirical ontology refinement. In *Proceedings of ICEIS 2007, vol. Artificial Intelligence and Decision Support Systems*. Kluwer Academic Publishing, 2007. In press.
16. Praveen K. Paritosh. The heuristic reasoning manifesto. In *Proceedings of the 20th International Workshop on Qualitative Reasoning*, 2006.
17. Rachel Pereira, Ivan Ricarte, and Fernando Gomide. Fuzzy relational ontological model in information search systems. In Elie Sanchez, editor, *Fuzzy Logic and the Semantic Web*, Capturing Intelligence, chapter 20, pages 395–412. Elsevier, 2006.
18. Christian D. Schunn and Kevin Dunbarr. Priming, analogy and awareness in complex reasoning. *Memory & Cognition*, 24(3):271–284, 1996.
19. Amit Sheth, Cartic Ramakrishnan, and Christopher Thomas. Semantics for the semantic web: The implicit, the formal and the powerful. *International Journal on Semantic Web & Information Systems*, 1(1):1–18, 2005.
20. G. Stoilos, G. Stamou, V. Tzouvaras, J.Z. Pan, and I. Horrocks. Fuzzy owl: Uncertainty and the semantic web. International Workshop of OWL: Experiences and Directions, Galway, 2005, 2005.
21. Umberto Straccia. A fuzzy description logic for the semantic web. In Elie Sanchez, editor, *Fuzzy Logic and the Semantic Web*, Capturing Intelligence, chapter 4, pages 73–90. Elsevier, 2006.
22. Wen-June Wang. New similarity measures on fuzzy sets and on elements. *Fuzzy Sets and Systems*, 85:305–309, 1997.