KB_Bio_101: A Repository of Graph-Structured Knowledge

Vinay K. Chaudhri, Michael Wessel, and Stijn Heymans

Artificial Intelligence Center, SRI International, Menlo Park, CA, 94025

1 Introduction

The goal of Project Halo is to develop a "Digital Aristotle" — a reasoning system capable of answering novel questions and solving advanced problems in a broad range of scientific disciplines and related human affairs [3]. As part of this effort, SRI has created a system called Automated User-Centered Reasoning and Acquisition System (AURA) [12], which enables educators to encode knowledge from science textbooks in a way that it can be used for answering questions by reasoning.

A team of biologists is currently using AURA to encode a popular biology textbook that is used in advanced high school and introductory college courses in the United States [15]. The knowledge base called `KB_Bio_101` is an outcome of this effort and contains concept taxonomy for the whole textbook and detailed rules for 20 chapters of the textbook. The current focus in the project is to expand the `KB_Bio_101` to cover all the 56 chapters of the book by December 2013. In the longer-term, `KB_Bio_101` will be expanded both in expressiveness and coverage. In terms of expressiveness, the Project Halo team is investigating the use of defaults, exceptions, negations, disjunctions and a process language. In terms of scope, the KB will likely be expanded to cover multiple textbooks potentially spanning a full undergraduate curriculum.

AURA uses a knowledge representation and reasoning system called Knowledge Machine (KM) [8]. KM supports a variety of representation features that include a facility to define classes and organize them into a hierarchy and define concept partitions (disjointness and covering axioms), ability to define relations (also known as slots) and organize them into a relation hierarchy, support for nominals, a facility to define horn rules, a procedure language, a situation mechanism, and a STRIPS representation for actions. KM performs reasoning by using inheritance, description-logic style classification of individuals, backward chaining over rules, and a heuristic unification. In addition, KM can use its situation mechanism and STRIPS representation of actions to simulate their execution. While the AURA team has experimented with the use of all of these features, the current core of AURA leverages only a small subset. The Project Halo team has invested significant effort to identify these core features and to specify them in a declarative manner. One example of such an effort is the work to specify the heuristic unification in KM using an answer set programming framework [7]. The net result of these efforts is that the team is now able to export the `KB_Bio_101` in a variety of standard declarative languages, for example, first order logic with equality [9], SILK [11], description logics (DLs) [5] and answer set programming [10].

The `KB_Bio_101` is a central component of an electronic textbook application called Inquire Biology [2] aimed at students studying from it. SRI has worked with teachers and students to collect a large number of questions that are of practical interest for this application. Working from those questions, the team has formulated logical reasoning tasks that must be performed by a reasoner.

The `KB_Bio_101` presents a unique opportunity for us to test our reasoners and to motivate further development. Recognizing that logical reasoning is only one component of the overall task of answering questions, the team at SRI is in the process of formulating similar challenges for knowledge representation [1] and natural language generation [6] which are also centered on `KB_Bio_101`. Taken collectively, these multiple challenges position us to make major leaps in AI in general, and knowledge-based question answering in particular.

2 Representation of Graphs in a Standard DL Syntax

There are two problems that need to be addressed to provide a representation of graphs in the DLs: defining a syntax for describing graphs and defining a family of graph expressiveness layers. We explain this in more detail next.

In principle, role value maps would be needed in order to truthfully represent the content of the `KB_Bio_101`. Role-value maps are a standard-way of expressing graph-structured descriptions in DL syntax. Unfortunately, unrestricted role value maps quickly lead to undecidability. There are decidable variants of role value maps, e.g. the restricted role-value-maps in a description logic with existential restrictions and terminological cycles ($\mathcal{EL}$ with cyclical TBoxes) of Baader [4], and we will check the applicability of this work to `KB_Bio_101`.

In recent work on description graphs [14] and description graph logic programs [13], a DL knowledge base is extended using a graph structure. While this proposal allows representation of graphs, it does not extend the conventional DL syntax in a graceful manner in that the conventional syntax can be completely abandoned in favor of this new syntax. The OWL export of `KB_Bio_101` extends the conventional syntax of OWL to encode graph structures.

Restrictions in description graphs prohibit the use of certain forms of cycles are too severe for `KB_Bio_101` which needs cyclicity in addition to the ability to express graphs. While the work on description graphs acknowledges the need for more expressive formalisms that go beyond tree structures, the nature of `KB_Bio_101` is sufficiently different from the setting in description graphs that it requires further research and could prove to be a data set that drives research beyond the current state.

3 Reasoning with Graph-Structured Descriptions

Similarity reasoning and relationship reasoning are two tasks that are of great practical interest to our application. In a similarity reasoning task, we are given two graph structured descriptions A and B, and the task is to compute new descriptions that correspond to their intersection and difference.

In the relationship reasoning task, we first create an ABOX by instantiating each concept in the TBOX, and then given two individuals A and B, we wish to compute all possible paths of a certain length between those individuals.

4 Summary

An initial version of the `KB_Bio_101` in OWL is now available. We are interested in identifying collaborators interested in exploiting this KB in the context of their tool set. We will work with them to first define an acceptable translation, and then participate in an experimental evaluation of the results of the reasoning tasks suggested above.

5 Acknowledgments

This work has been funded by Vulcan Inc.

References

1. Deep knowledge representation and reasoning challenge. https://sites.google.com/site/dkrckcap2011/ and https://sites.google.com/site/2nddeepkrchallenge/.
2. Inquire: An Intelligent Textbook. http://aivideo.org/2012/.
3. Project Halo. http://www.projecthalo.com.
4. Franz Baader. Restricted role-value-maps in a description logic with existential restrictions and terminological cycles. In *Proceedings of the International Workshop on Description Logics (DL 2003)*, 2003.
5. Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2nd edition, 2007.
6. Eva Banik, Claire Gardent, Donia Scott, Nikhil Dinesh, and Fennie Liang. KBGen Text Generation from Knowledge Bases as a New Shared Task. In *International conference on Natural Language Generation*, 2012.
7. Vinay K. Chaudhri and Tran C. Son. Specifying and Reasoning with Underspecified Knowledge Bases Using Answer Set Programming. In *Proc. of International Conference on Knowledge Representation and Reasoning (KR)*, 2012.
8. Peter E. Clark and Bruce Porter. Knowledge machine userss guide. Technical report, University of Texas at Austin.
9. Melvin Fitting. *First-Order Logic and Automated Theorem Proving*. Springer, 1996.

10. M. Gelfond and V. Lifschitz. Logic programs with classical negation. In D.Warren and Peter Szeredi, editors, *Logic Programming: Proceedings of the Seventh International Conference*, pages 579–597, 1990.
11. Benjamin N. Grosof. SILK: Higher Level Rules with Defaults and Semantic Scalability. In Axel Polleres and Terrance Swift, editors, *Web Reasoning and Rule Systems, Third International Conference (RR 2009)*, volume 5837 of *Lecture Notes in Computer Science*, pages 24–25. Springer, 2009.
12. David Gunning, Vinay K. Chaudhri, Peter Clark, Ken Barker, Shaw-Yi Chaw, Mark Greaves, Benjamin Grosof, Alice Leung, David McDonald, Sunil Mishra, John Pacheco, Bruce Porter, Aaron Spaulding, Dan Tecuci, and Jing Tien. Project Halo Update – Progress Toward Digital Aristotle. *AI Magazine*, Fall 2010.
13. Despoina Magka, Boris Motik, and Ian Horrocks. Modeling Structured Domains using Description Graphs and Logic Programming. In *Proceedings of the International Workshop on Description Logics (DL 2012)*, 2012.
14. Boris Motik, Bernardo Cuenca Grau, Ian Horrocks, and Ulrike Sattler. Representing Ontologies using Description Logics, Description Graphs and Rules. *Artificial Intelligence*, 173:1275–1309, 2009.
15. Jane B. Reece, Lisa A. Urry, Michael L. Cain, Steven A. Wasserman, Peter V. Minorsky, and Robert B. Jackson. *Campbell Biology.* Benjamin Cummings, 9th edition, 2011.