Tracking Logical Difference in Industrial-Scale Ontologies*

Yizheng Zhao[1,2,3], Ghadah Alghamdi[3], Renate A. Schmidt[3], Hao Feng[4]
Giorgos Stoilos[5], Damir Juric[5], Mohammad Khodadadi[5]

[1] State Key Laboratory for Novel Software Technology, Nanjing Univeristy, China
[2] School of Artificial Intelligence, Nanjing University, China
[3] School of Computer Science, The University of Manchester, UK
[4] North China University of Science and Technology, China
[5] Babylon Health Limited, London, UK

This paper considers the problem of tracking the changes between different versions of an ontology, which can be done based on the notion of *logical difference*. The logical difference between two ontologies are the axioms in one that are not entailed by the other, reflecting information gained and information lost between the ontologies [5].

For ontology engineers, tracking the logical difference in industrial ontologies is a critical task for the purpose of maintaining them: to track what has changed in a new version of an ontology, to ensure that the extension is safe in the sense of being a conservative extension [9], and to identify information gained and information lost [4]. This provides a means of discovering issues in the ontologies and supports quality assurance reviews. Being able to compute the logical difference between two ontologies is also important when merging/aligning ontologies and integrating different ontologies [3,12].

Presently available ontology difference and alignment/merging tools compute the *structural difference* between ontologies [11,1], which is a syntactic notion, or *approximate* the logical difference between ontologies [5,3,7,12], which is a semantic notion. A particular challenge is the size of ontologies used in real-world applications. Our targets are the SNOMED CT and NCIt ontologies. Being the most comprehensive, multi-lingual medical ontology in the world, the core SNOMED CT ontology contains over 335K concepts definitions [13]. The NCIt ontology defines terminologies in the biomedical domain and contains more than 60K concept definitions [2].

This paper explores how the logical difference between two ontologies can be computed using a forgetting- or uniform interpolation-based approach, as proposed in [8]. Forgetting is an ontology re-engineering technique which preserves the semantics of definitions of concepts and relationships among them [6,10,14]. Current forgetting tools have difficulties in tracking the logical difference in very large ontologies. To address this shortcoming we introduce a new forgetting method designed for the task of tracking the logical difference in large-scale $\mathcal{ALC}$-ontologies. The method is sound and terminating, and can compute forgetting solutions for ontologies as large as SNOMED CT and NCIt. Our evaluation shows

* This is an extended abstract of a paper published in the proceedings of AAAI 2019.

the method can achieve considerably better success rates (>90%) over existing tools such as LETHE and FAME 1.0 on the restrictions to $\mathcal{ALC}$ of SNOMED CT and NCIt. Based on this forgetting method we have developed a new logical difference tool to analyze the differences between the Australia and Canada country extensions of the base SNOMED CT International ontology and recent versions of the NCIt ontologies. While the Canada extension was found to be a conservative extension of SNOMED CT International, there is significant diversion in the Australia extension, as is reflected in large logical difference sets.

References

1. R. S. Gonçalves, B. Parsia, and U. Sattler. Ecco: A hybrid diff tool for OWL2 ontologies. In *OWLED*, volume 849 of *CEUR Workshop Proc.*, 2012.
2. F. W. Hartel, S. de Coronado, R. Dionne, G. Fragoso, and J. Golbeck. Modeling a description logic vocabulary for cancer research. *J. Biomed. Inform.*, 38(2):114–129, 2005.
3. E. Jiménez-Ruiz, B. C. Grau, I. Horrocks, and R. B. Llavori. Supporting concurrent ontology development: Framework, algorithms and tool. *Data Knowl. Eng.*, 70(1):146–164, 2011.
4. M. C. A. Klein, D. Fensel, A. Kiryakov, and D. Ognyanov. Ontology versioning and change detection on the web. In *Proc. EKAW'02*, volume 2473 of *LNCS*, pages 197–212. Springer, 2002.
5. B. Konev, M. Ludwig, D. Walther, and F. Wolter. The logical difference for the lightweight description logic EL. *J. Artif. Intell. Res.*, 44:633–708, 2012.
6. B. Konev, D. Walther, and F. Wolter. Forgetting and Uniform Interpolation in Extensions of the Description Logic $\mathcal{EL}$. In *Proc. DL'09*, volume 477 of *CEUR Workshop Proc.*, 2009.
7. P. Kremen, M. Smid, and Z. Kouba. Owldiff: A practical tool for comparison and merge of OWL ontologies. In *DEXA Workshops*, pages 229–233. IEEE, 2011.
8. M. Ludwig and B. Konev. Practical Uniform Interpolation and Forgetting for $\mathcal{ALC}$ TBoxes with Applications to Logical Difference. In *Proc. KR'14*. AAAI Press, 2014.
9. C. Lutz, D. Walther, and F. Wolter. Conservative Extensions in Expressive Description Logics. In *Proc. IJCAI'07*, pages 453–458. AAAI/IJCAI Press, 2007.
10. C. Lutz and F. Wolter. Foundations for Uniform Interpolation and Forgetting in Expressive Description Logics. In *Proc. IJCAI'11*, pages 989–995. IJCAI/AAAI Press, 2011.
11. N. F. Noy and M. A. Musen. PROMPTDIFF: A fixed-point algorithm for comparing ontology versions. In *Proc. AAAI'02*, pages 744–750. AAAI Press / The MIT Press, 2002.
12. A. Solimando, E. Jiménez-Ruiz, and G. Guerrini. Minimizing conservativity violations in ontology alignments: algorithms and evaluation. *Knowl. Inf. Syst.*, 51(3):775–819, 2017.
13. K. A. Spackman. Snomed rt and snomed ct. promise of an international clinical ontology. *M.D. Computing 17*, 2000.
14. K. Wang, Z. Wang, R. W. Topor, J. Z. Pan, and G. Antoniou. Eliminating Concepts and Roles from Ontologies in Expressive Descriptive Logics. *Computational Intelligence*, 30(2):205–232, 2014.