Evaluating OWL 2 Reasoners in the context of Clinical Decision Support in Lung Cancer Treatment Selection

M. Berkan Sesen[1], Ernesto Jiménez-Ruiz[2],
René Bañares-Alcántara[1], Sir Michael Brady[3]

[1] Department of Engineering Science, University of Oxford, UK
[2] Department of Computer Science, University of Oxford, UK
[3] Department of Oncology, University of Oxford, UK

Abstract. This paper evaluates the performances of the OWL 2 reasoners HermiT, FaCT++ and Pellet in the context of an ontological clinical decision support system in lung cancer care. In the first set of experiments, we compare how the classification and realisation times of the LUCADA and LUCADA-SNOMED CT ontologies vary as we expand their TBoxes with additional guideline rule knowledge. In the second set of experiments, we investigate the effect of increasing the ABox of the LUCADA ontology on the realisation times.

1 Introduction

Lung cancer is the most common and deadliest type of cancer, and is responsible for 21% of all cancer-related deaths globally. In England, care decisions for lung cancer patients are made by multidisciplinary teams (MDTs) that are comprised of clinical staff from diverse backgrounds. These teams meet weekly in cancer centres across the country in order to come to treatment decisions for each patient in their care. Usually, MDTs make use of their combined experience and knowledge of published clinical guidelines to decide upon the next stage of treatment for a patient [1]. The National Lung Cancer Audit (NLCA) data reveals that one of the major problems in the management of lung cancer care in England is the substantial level of unjustified variation in treatment decisions between different cancer centres [14, 13].

In order to reduce variability in clinical practice, clinical guidelines provide well defined sets of directions and evidence based standards to assist clinicians on decisions about appropriate clinical procedures [6]. However, as unstructured and free-text documents, clinical guidelines are usually not readily accessible at the point of decision making in the MDT meetings. Fortunately, clinical decision support (CDS) systems that computerise and automate the daily management of guidelines can facilitate access to guideline information in these meetings.

The computerisation of guideline rules can be achieved by structured logical languages which can express guideline rule eligibility and decision criteria. To date, many proprietary expression languages [4, 9, 11, 19, 20] have been proposed in order to encode and interpret guideline rules that are in a machine readable format. The interpretation of computerised guideline rules are carried out by execution engines that can match the encoded guideline rule criteria against existing patient records in order to infer rule applicability for different patient records.

In [16], we proposed OWL 2 [2] as a suitable candidate for encoding guideline rule criteria in the context of a CDS system for lung cancer care and we outlined a purely ontological guideline rule inference framework. In this paper, we focus on performance evaluations of off-the-shelf OWL 2 reasoners for inferring patient rule applicability based on the guideline rule inference framework presented in [16].

2 LUCADA ontology

Since 2004, the NLCA has collected all lung cancer patient data in England within the English Lung Cancer Dataset (LUCADA) [13] in order to gain a better understanding of the care delivered during referral, diagnosis and treatment of lung cancer patients. We have manually built a domain specific OWL 2 lung cancer ontology based on the LUCADA data model.[4] The LUCADA ontology provides the semantic layer of the Lung Cancer Assistant [16], an ontology-based system that is capable of providing guideline rule-based decision support during lung cancer MDT meetings.

SNOMED CT [15] is the reference ontology of choice across the information systems within the National Health Service (NHS). Thus, to facilitate interoperability with other NHS applications, we integrated LUCADA with a lung cancer-specific module of SNOMED CT. To this end, we have *(i)* identified the classes in SNOMED CT related to those in LUCADA and established correspondences (i.e. mappings) between them; and *(ii)* extracted a small fragment of SNOMED CT that captures the meaning of such relevant classes (i.e., a domain-specific module). SNOMED CT, however, is a complex ontology describing more than 300,000 classes; as a result, computing mappings with LUCADA is infeasible without suitable tool support. Thus, to perform task (i) we used the interactive-mode of the ontology matching system LogMap [7, 8]. Additionally, in order to perform task (ii), we used the ontology modularization technique described in [3]. Table 1 provides a side by side comparison of LUCADA and the integrated ontology LUCADA-SNOMED CT in terms of number of entities, axioms and expressivity.

In order to incorporate lung cancer guideline knowledge, we introduced the *patient scenario* class into both ontologies [16]. A guideline rule consists of an antecedent, i.e. rule body, which specifies the eligibility criteria for the rule and a consequent, i.e. rule head, which encapsulates the action(s) to take when the conditions in the antecedent are satisfied [5]. According to our guideline rule inference framework, we represent the guideline rule antecedents as defined *patient scenario* classes, whose equivalent class capture the semantics for rule eligibility criteria. As an example, the eligibility for the guideline rule[5] *"Consider radiotherapy for Stage I, II, III patients with good performance status"* is encoded as the following OWL 2 class equivalence axiom:

$$\begin{aligned}\mathsf{GR1} \equiv\ & \mathsf{GoodPerformancePatient} \sqcap \exists \mathsf{hasClinicalFinding}. \\ & (\mathsf{NeoplasticDisease} \sqcap \exists \mathsf{hasPreHistology.NonsmallCellCarcinoma} \sqcap \\ & \exists \mathsf{hasPreTNMStaging}.string \sqcap \forall \mathsf{hasPreTNMStaging}.\{I, II, III\})\end{aligned}$$

[4] Through a data sharing agreement between the University of Oxford and NLCA, we have been granted access to an anonymised version of LUCADA dataset.

[5] The guideline rules have been extracted from from National Institute for Clinical Excellence (NICE) document [12].

Table 1: Summary of the LUCADA and LUCADA-SNOMED CT ontology metrics

Metric \ Ontology	LUCADA-SNOMED CT	LUCADA
DL Expressivity	$\mathcal{ALCHIF}(\mathcal{D})$	$\mathcal{ALCHI}(\mathcal{D})$
# **Classes**	1553	376
# **Object properties**	63	37
# **Data Properties**	63	63
# **Equiv. class axioms**	1010	0
# **Subclass of axioms**	999	386
# **Prop. domain axioms**	97	97
# **Prop. range axioms**	30	30

Furthermore, we represent a *patient record* as a set of OWL 2 individual axioms with respect to the terminological knowledge captured within the LUCADA and the integrated LUCADA-SNOMED CT ontologies as exemplified in [16]. According to this, a patient record is characterised (on average) by 25 class and property assertion axioms. An OWL 2 reasoner can be used to determine whether a specific patient is a member of a particular patient scenario class, and therefore, subject to the recommendations or actions of the respective guideline rule.

3 Evaluation

We evaluated the scalability of our guideline rule inference framework with off-the-shelf OWL 2 reasoners: HermiT 1.3.7 [10], Pellet 2.3.0 [17] and FaCT++ 1.6.2 [18]. The tests have been performed on a Windows 7 64-bit desktop computer with 15 GiB of RAM and an Intel Xeon 2.27 GHz CPU. Overall, we report two sets of experimental results as given below. Note that all results reported here have been acquired as averages of at least 10 repetitions of the described experimental setup.

3.1 Increasing the TBox with patient scenarios

In the first set of experiments we compared how the classification and realisation times of LUCADA and LUCADA-SNOMED CT ontologies varied as we increased the guideline rule coverage (i.e. patient scenarios classes). To this end, we incrementally added to each ontology 40 patient scenarios, represented as equivalent class axioms (see Section 2), and recorded the times taken by each reasoner to perform classification (i.e. execution of precomputeInferences(CLASS_HIERARCHY) method) and realisation of only one patient individual (i.e. execution of the method getTypes()).

Figures 1 and 2 summarise the reasoning times obtained for the LUCADA and LUCADA-SNOMED CT ontologies respectively. In both figures, we only report the total inference times (classification + realisation) for FaCT++ and Pellet since the individual realisation times for these two reasoners were negligible. However, for HermiT

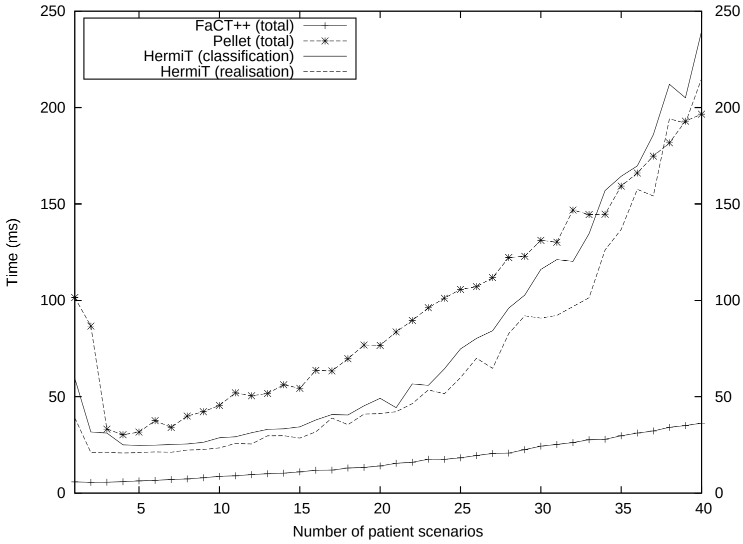

Fig. 1: Reasoning times for LUCADA containing 1 to 40 patient scenarios

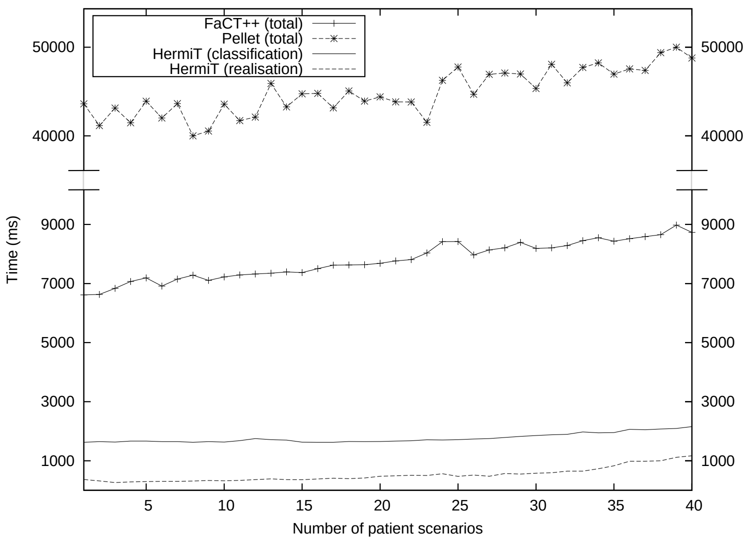

Fig. 2: Reasoning times for LUCADA-SNOMED CT containing 1 to 40 patient scenarios

we present classification and realisation times separately, since realisation takes up a significant portion of the total inference time (up to 0.2ms for LUCADA and 1s for LUCADA-SNOMED CT). We note that the classification times for all three reasoners are below one second for the LUCADA ontology, whereas they rise to 9 and 50 seconds re-

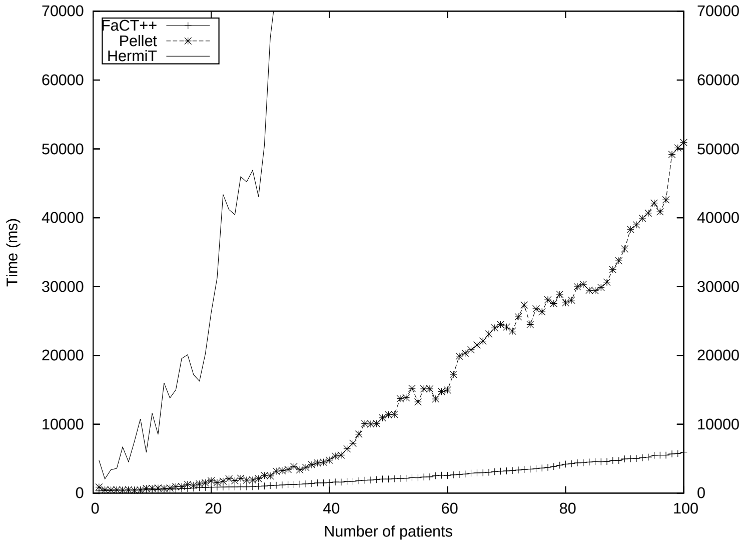

Fig. 3: Realisation times in LUCADA with 1 to 100 patient records

spectively with FaCT++ and Pellet for the integrated LUCADA-SNOMED CT ontology. Note that HermiT classifies the integrated ontology the fastest, with classification times ranging from 1.6s to 2.2s.

3.2 Increasing the ABox with patient records

In the second set of experiments, we incrementally added 100 patient records, represented as OWL 2 individuals axioms (see Section 2), to the LUCADA ontology which contained 40 patient scenarios. Figure 3 compares the realisation times (i.e. execution of the method getTypes() for each patient individual) obtained by all three reasoners. As expected, realisation times increase as more patients are added to the ontology. It is noticeable that FaCT++ and HermiT have very disparate behaviours. While the increase in realisation times with respect to the number of patient individual in the ontology is fairly gradual and linear for FaCT++, the realisation times for HermiT increase very quickly and clearly in a non-linear fashion. Although not as severe as the realisation times achieved by HermiT, Pellet realisation times are also considerably slower compared to FaCT++ and seem to increase non-linearly.

4 Conclusions

In this paper we evaluated empirically the classification and realisation performances of the three most commonly used OWL 2 reasoners within our guideline rule inference framework. We found that FaCT++ is the best choice for our application since it provides very fast inference times for both classification and realisation. We also found that HermiT provides the fastest TBox reasoning times for the integrated LUCADA-SNOMED CT ontology; but it performs poorly in ABox reasoning with both ontologies. Finally, we found that Pellet performs well in classifying the LUCADA ontology but struggles with the LUCADA-SNOMED CT ontology, which contains many axioms inherited from SNOMED CT.

Acknowledgements

The LCA project was funded by the CDT in Healthcare Innovation programme within the Institute of Biomedical Engineering, Oxford University. We would also like to acknowledge the clinical inputs from our collaborators Dr Michael Peake, Prof Fergus Gleeson and Dr Donald Tse during the elicitation of guideline rules from the literature. Jiménez-Ruiz was partially supported by the Seventh Framework Program (FP7) of the European Commission under Grant Agreement 318338, "Optique", and the EPSRC projects Score!, ExODA and MaSI[3].

References

1. Austin, M.: Information Integration and Decision Support for Multidisciplinary Team Meetings on Colorectal Cancer. Ph.D. thesis, University of Oxford (2008)
2. Cuenca Grau, B., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P.F., Sattler, U.: OWL 2: The next step for OWL. J. Web Sem. 6(4), 309–322 (2008)
3. Cuenca Grau, B., Horrocks, I., Kazakov, Y., Sattler, U.: Modular reuse of ontologies: Theory and practice. J. Artif. Intell. Res. 31, 273–318 (2008)
4. Fox, J., Johns, N., Lyons, C., Rahmanzadeh, A., Thomson, R., Wilson, P.: PROforma: a general technology for clinical decision support systems. Computer Methods and Programs in Biomedicine 54(12), 59 – 67 (1997)
5. Horrocks, I., Patel-Schneider, P.F., Boley, H., Tabet, S., Grosof, B., Dean, M.: SWRL: A Semantic Web Rule Language Combining OWL and RuleML. Tech. rep., World Wide Web Consortium (2004)
6. Isern, D., Sánchez, D., Moreno, A.: HeCaSe2: A Multi-agent Ontology-Driven Guideline Enactment Engine. In: 5th International Central and Eastern European Conference on Multi-Agent Systems. pp. 322–324 (2007)
7. Jiménez-Ruiz, E., Cuenca Grau, B.: LogMap: Logic-based and Scalable Ontology Matching. In: Int'l Sem. Web Conf. (ISWC). pp. 273–288 (2011)
8. Jiménez-Ruiz, E., Cuenca Grau, B., Zhou, Y., Horrocks, I.: Large-scale interactive ontology matching: Algorithms and implementation. In: European Conf. on Artif. Intell. (ECAI). pp. 444–449 (2012)
9. Miksch, S., Shahar, Y., Johnson, P.D.: Asbru: A Task-Specific, Intention-Based, and Time-Oriented Language for Representing Skeletal Plans. In: 7th Workshop on Knowledge Engineering: Methods & Languages (KEML-97) (1997)
10. Motik, B., Shearer, R., Horrocks, I.: Hypertableau reasoning for description logics. J. Artif. Intell. Res. 36, 165–228 (2009)
11. Musen, M.A., Tu, S.W., Das, A.K., Shahar, Y.: EON: A Component-Based Approach to Automation of Protocol-Directed Therapy. Journal of the American Medical Informatics Association 3(6) (1996)

12. NICE: The Diagnosis and Treatment of Lung Cancer (Update). National Collaborating Centre for Cancer (UK). NICE Clinical Guidelines, No. 121. (2011), available from: `http://www.ncbi.nlm.nih.gov/books/NBK99021/`
13. NLCA: The National Clinical Lung Cancer Audit (LUCADA) Data Manual (2010), available from: `http://www.hscic.gov.uk/lung`
14. Riaz, S.P., Lchtenborg, M., Jack, R.H., Coupland, V.H., Linklater, K.M., Peake, M.D., Mller, H.: Variation in surgical resection for lung cancer in relation to survival: Population-based study in england 20042006. European Journal of Cancer 48(1), 54 – 60 (2012)
15. Schulz, S., Cornet, R., Spackman, K.A.: Consolidating SNOMED CT's ontological commitment. Applied Ontology 6(1), 1–11 (2011)
16. Sesen, M.B., Bañares-Alcántara, R., Fox, J., Kadir, T., Brady, J.M.: Lung Cancer Assistant: An ontology-driven, online decision support prototype for lung cancer treatment selection. In: OWL: Experiences and Directions Workshop (OWLED) (2012)
17. Sirin, E., Parsia, B., Grau, B.C., Kalyanpur, A., Katz, Y.: Pellet: A practical OWL-DL reasoner. J. Web Sem. 5(2), 51–53 (2007)
18. Tsarkov, D., Horrocks, I.: FaCT++ Description Logic Reasoner: System Description. In: Third International Joint Conference on Automated Reasoning, IJCAR. pp. 292–297 (2006)
19. Tu, S.W., Campbell, J.R., Glasgow, J., Nyman, M.A., McClure, R., McClay, J., Parker, C., Hrabak, K.M., Berg, D., Weida, T., Mansfield, J.G., Musen, M.A., Abarbanel, R.M.: The SAGE Guideline Model: Achievements and Overview. Journal of the American Medical Informatics Association 14(5), 589 – 598 (2007)
20. Wang, D., Peleg, M., Tu, S.W., Boxwala, A.A., Ogunyemi, O., Zeng, Q.T., Greenes, R.A., Patel, V.L., Shortliffe, E.H.: Design and implementation of the GLIF3 guideline execution engine. Journal of Biomedical Informatics 37(5), 305–318 (2004)