SNOMED CT Expo 2019
Kuala Lumpur | Oct 31-Nov 1
Leading health innovation through data

201901 Modularity Meets Forgetting: A case with the SNOMED CT - Ontology

Jieying Chen,The University of Manchester (United Kingdom)

Co-authors

1. Renate Schmidt
2. Yongcheng Gao
3. Ghadah Alghamdi
4. Dirk Walther

Summary

We present a methodology and results of an ongoing project with IHTSDO for obtaining semantics preserving abstractions of SNOMED CT. Use cases include creating views for a medical specialty and hiding unwanted information. Abstractions can be used for model design, quality assurance and analytics.

Audience

Research/academic, Technical

Learning Objectives

1. SNOMED Abstraction,
2. Quality Assurance,
3. Model Development,
4. Modularity,
5. Forgetting.

Abstract

(As the paragraphs cannot be displayed properly in the system, please check the full abstract details in the link: http://bit.ly/2GN52LL.)

SNOMED CT is a comprehensive, precise and widely-used medical ontology covering ample clinical specialities and requirements. The latest release from January 2019 contains more than 340 000 axioms, with the number of axioms increasing by about 10% compared to the version of 2016. The number of concept names and size of the ontology will still keep on increasing.

Maintaining and developing such large-scale ontologies poses significant challenges for semantic web applications. Ontology modularization and forgetting provide very useful automated support tools to ontology developers for extracting smaller sets of relevant axioms from an ontology. In this presentation we discuss the relationship of three existing module extraction techniques and investigate their benefits when combined with forgetting, also known as uniform interpolation.

It is widely recognized that the creation of ontology extracts is a useful operation in the reuse, creation, curation, decomposition, integration and general use of ontologies. For example, for reviewing and analyzing the information relating to the concept “kidney disease (disorder)” which has more than 1200 sub-concepts in SNOMED CT, developers would benefit from being able to work with an extract that succinctly summarizes all information in the ontology relating to kidney diseases. Another concrete scenario is a doctor wishing to find diseases that have an inflammatory morphology and a finding site of kidney structure based on morphologies and/or finding sites. Instead of querying SNOMED CT as a whole, it would be more efficient to simply query a smaller extract of the ontology containing sufficiently many axioms to compute the same answer as if it was computed on the entire ontology.

A method commonly used for creating extracts of an ontology is modularization. In general, a module of an ontology is a subset of the ontology that in a specific context can function in the same way as the original ontology. Different notions of modules have been proposed. Among them are notions such as the so-called plain, self-contained and depleting modules [1]. The MEX system extracts minimal depleting and self-contained modules from ontologies formulated as acyclic EL-terminologies. Recently introduced minimal subsumption modules [2,3] are subsets of an ontology that preserve EL-subsumption queries. The evaluation in [3] shows that minimal subsumption modules are generally much smaller than MEX-modules. However, deciding the preservation of subsumption queries can be expensive. So, the algorithm for computing minimal subsumption modules from EL-terminologies runs in exponential time.

Approximate modules, such as modules based on syntactic locality [4], can be computed efficiently. For instance, the algorithm of extracting locality-based modules runs in polynomial time in the size of the ontology. Empirical investigations by [1,5] in application-close scenarios involving SNOMED CT have found that, while graph-based approaches to modularization have reasonable coverage, the obtained extracts are large. While relatively small extracts can be obtained with locality-based modules, a down-side is lower precision due to the presence of a large number of symbols in the module outside the desired signature specified as input.

An alternative method for creating a compact representation of a subset of the information in an ontology is uniform interpolation/forgetting. The returned set of axioms after uniform interpolation is called a uniform interpolant, which can be regarded as a view of an ontology. Forgetting allows the creation of abstractions of an ontology by hiding specified class or property names (the forgetting signature), without loosing the underlying logical definitions of the remaining class and property names (the interpolation signature). Because forgetting preserves all information in the specified interpolation signature, it has high precision, but the axioms in a forgetting solution (or uniform interpolant) are in general not axioms belonging to the original ontology. For example, the ontology O consists of the following axioms:

O = { Drug_interaction_with_drug (finding) $\sqsubseteq$ Drug_interaction (finding),

Drug_or_medicament (substance) $\sqsubseteq$ Substance (substance),

Adverse_drug_interaction_with_drug (disorder) $\sqsubseteq$ Adverse_drug_interaction (disorder) $\sqcap$

Drug_interaction_with_drug (finding) $\sqcap$ $\exists$ Associated_with (attribute). Drug_or_medicament (substance) }.

Suppose that the interpolation signature

Σ = { Adverse_drug_interaction_with_drug (disorder), Drug_interaction (finding),

Substance (substance), Associated_with (attribute) }.

Then the uniform interpolant of O w.r.t. Σ is

{ Adverse_drug_interaction_with_drug (disorder) ⊑ Drug_interaction (finding) ⊓

∃ Associated_with (attribute). Substance (substance) },

while the module is the ontology itself. Compared with ontology modularization, a significant advantage of uniform interpolation is that the returned interpolants only use class and property names in the interpolation signature.

This presentation discusses ongoing work in collaboration with IHTSDO about abstraction on the content of SNOMED CT. Our interest is the computation of forgetting-based ontology extracts in real scenarios involving the SNOMED CT ontology. For SNOMED CT it requires nearly all of the names in the ontology to be forgotten, which poses a significant new challenge for uniform interpolation and forgetting tools. In particular, property names can be rather difficult to forget.

The approach introduced in this presentation uses a pipeline of three steps:

(i) extension and, if needed, partitioning of the interpolation signature,

(ii) ontology modularization—we use three different tools to extract different ontology modules: locality-based STAR-modules by the OWL API, modularization performed by the MEX system and minimal subsumption modularization, and

(iii) forgetting using the systems: Fame [6] and Lethe [7].

The presentation addresses the following issues posed by the creation of extracts of ontologies based on modularization and forgetting.

1. The interpolation signature is very small compared with the size of the signature of SNOMED CT and the forgetting signature is correspondingly very large. For instance, the average size of NHS Refsets is less than 0.8% of the signature in SNOMED CT and the ERA refset is only 0.02%.

2. Interpolation signatures defined by class names specified in a refset, which can be any subsets of SNOMED CT.

3. Extending the class set to include properties and classes picked up 'horizontally', thereby creating a neighbourhood of related class and property names for the given refset.

4. Since locality-based modules are often large (unless the seed signature is very carefully chosen) and contain foreign symbols, and for the other forms of modularization producing smaller results, the forgetting tools still had difficulties, a key issue was creating modules sufficiently small before the forgetting tool was applied and succeed in feasible time.

By computing three different ontology modules and applying forgetting method to the 2016 core version of SNOMED CT, we created ontology abstractions for a sample of reference lists used in the health care sector in the UK (NHS Refsets). We used 165 concepts in the ERA list as class names for primary renal disorders in our evaluation. The results show that precomputing MEX-modules and minimal subsumption modules significantly reduce the size of extracts and significantly improve the performance of our forgetting tool. The returned uniform interpolants only consist of class and property names in the interpolation signatures. Because forgetting preserves semantic equivalence high precision can be achieved.

For more details on technical aspects and the evaluation, please refer to the document at the following link: http://bit.ly/2GN6fmp.

Reference Documentation

[1] Konev, B., Lutz, C., Walther, D., Wolter, F.: Model-theoretic inseparability and modularity of description logic ontologies. Artif. Intell. 203, 66–103 (2013)

[2] Chen, J., Ludwig, M., Ma, Y., Walther, D.: Zooming in on ontologies: Minimal modules and best excerpts. In: Proc. ISWC'17. pp. 173–189 (2017)

[3] Chen, J., Ludwig, M., Walther, D.: Computing minimal subsumption modules of ontologies. In: Proc. GCAI'18. pp. 41–53 (2018)

[4] Grau, B.C., Horrocks, I., Kazakov, Y., Sattler, U.: Modular reuse of ontologies: Theory and practice. J. Artif. Intell. Res. 31(1), 273–318 (2008)

[5] Lopez-Garcia,P.,Boeker,M.,Illarramendi,A.,Schulz,S.: Usability-driven pruning of large ontologies: the case of SNOMED CT. Journal of the American Medical Informatics Association 19 (2012)

[6] Koopmann, P., Schmidt, R.A.: Forgetting Concept and Role Symbols in ALCH-Ontologies. In: Proc. LPAR'13. (2013).

[7] Zhao, Y., Alghamdi, G., Schmidt, R.A., Feng, H., Stoilos, G., Juric, D., Khodadadi, M.: Tracking logical difference in large-scale ontologies: A forgetting-based approach. In: Proc. AAAI'19 (2019)