Reconstruction of Hierarchical Structures in a Thesaurus-based Ontology through Curation by Domain Experts

Tatsuya Kushida[a], Kouji Kozaki[b], Yasunori Yamamoto[c], Atsuko Yamaguchi[c], Takahiro Kawamura[d], Takako Nakagawa[a], Yuka Tateisi[a], Toshihisa Takagi[a,e]

[a] *National Bioscience Database Center, Japan Science and Technology Agency, Tokyo, Japan,*
[b] *Faculty of Information and Communication Engineering, Osaka Electro-Communication University, Neyagawa, Osaka, Japan,*
[c] *Database Center for Life Science, Research Organization of Information and Systems, Kashiwa, Chiba, Japan,*
[d] *National Agriculture and Food Research Organization, Tsukuba, Ibaraki, Japan,*
[e] *Toyama University of International Studies, Toyama, Toyama, Japan*

Abstract

We reconstructed hierarchical structures of Interlinking Ontology for Biological Concepts by life-science experts converting the skos:broader and dcterms:subject relations to rdfs:subClassOf relations. As a result, we could efficiently create more than 125,000 rdfs:subClassOf in the ontology, and we expect that the relations would be leveraged for inferring biological molecules' functions by the inheritance approach.

Keywords:

Expert curation, Life science, Ontology

Introduction

Interlinking Ontology for Biological Concepts (IOBC) is developed from Japan Scientific and Technological Thesaurus (JST thesaurus). It contains approximately 120,000 terms related to life science and chemical science. So far we sub-classified the skos:related relations within the IOBC to 35 relations such as SIO:000225 (has function) [1]. Consequently, we could infer biological and chemical functions to biological molecules using the sub-classified and sub-class-of relations of IOBC [2]. However the hierarchical structure has been described using skos:broader and dcterms:subject without distinguishing some ontological relationships such as sub-class-of, and is-part-of because it was developed from the thesaurus. In this study, we attempted to reconstruct the hierarchical relations for the purpose of more rigorously describing the relationships among biological concepts and inferring biological molecules' functions at scale.

Methods

In order to improve the hierarchical structures of IOBC, four life-sciences experts attempted the followings. (1) The experts considered separating each "skos:broader" relation into each of "rdfs:subClassOf", and "SIO:000068(is part of)" because they were represented by skos:broader in IOBC. In addition, (2) the experts considered converting each "dcterms:subject" relation to "rdfs:subClassOf," because some upper concepts were represented by dcterms:subject instead of rdfs:subClassOf in IOBC. Furthermore, they examined whether there was more appropriate term upper concept for each rdfs:subClassOf in IOBC. As the result, the experts defined 35 newly categorized terms as upper concepts. The newly defined terms include "biological phenomenon, process, and state," "substance", etc.

In the above reconstruction processes, each relation was decided by 3 experts' majority rule, and discussion, and the ontological consistency was finally confirmed by an expert.

Results

We separated 147,460 skos:broader into 110,582 rdfs:subClsssOf, and 135 SIO:000068 (is part of) (1-1 and 1-2 in Table.1). Meanwhile, we converted 118,355 dcterms:subject to 15,092 rdfs:subClassOf (2 in Table.1) . This ontology is available in BioPortal (http://purl.bioontology.org/ontology/IOBC).

Table.1 Reconstruction examples of the hierarchical structures

#	Before reconstruction	After reconstruction
1-1	erythrocyte aggregation (skos:broader) hemagglutination	erythrocyte aggregation (rdfs:subClassOf) hemagglutination
1-2	antigen recognition (skos:broader) antigen-antibody reaction	antigen recognition (SIO:000068 (is part of)) antigen-antibody reaction
2	cell aggregation (dcterms:subject) Immunology	cell aggregation (rdfs:subClassOf) biological phenomenon, process, and state

Conclusions

The expert curation greatly and exactly increasing the number of rdfs:subClassOf, and SIO:000068 (is part of) relations in the IOBC, we can more largely and accurately discover new functions, and roles of biological concepts such as chemical compounds and gene products using the is-a hierarchy and the part of structure, and combinating the IOBC, with other ontologies, such as ChEBI.

Address for correspondence

To whom correspondence should be addressed: kushida@biosciencedbc.jp

References

1. Kushida, T., et.al. (2017). Efficient construction of a new ontology for life sciences by sub-classifying related terms in the Japan Science and Technology Agency thesaurus. Proc ICBO2017. pp.1–6.
2. Kushida, T., et.al. (2018). Inference of Functions, Roles, and Applications of Chemicals Using Linked Open Data and Ontologies. In Semantic Technology: 8th JIST, LNCS 11341, pp.385–397.