

Symptom ONtology - Error Check

J. Allen Baron

2022-04-07

Contents

Background	1
Goal	2
Error Check	2
Extract Terms	2
Check for Requirements Missing	2
IDs	2
Labels	3
Namespace	3
Check for Error-producing Duplicates	3
URIs	3
IDs	3
Labels	3
Synonymous Labels as Separate Terms	3
Incorrect critical information	4
Namespace incorrect	4
URI != ID	4

Background

While writing definitions I became curious about symptoms and wondered what was in the Symptom ontology. I wrote a quick SPARQL query to look at the list of symptom terms and, in doing so, came across an error. This prompted me to wonder about the possibility of more egregious errors in the ontology because it lacks the Github CI checks that the Human Disease Ontology has.

Goal

To check the Symptom ontology for the most egregious errors.

The errors I will check for include:

1. Missing critical information
 - ID
 - label
 - namespace
2. Duplicate critical information.
 - URI
 - ID
 - label
 - synonymous labels*
3. Incorrect critical information
 - namespace
 - ID not matching URI

* Checking for synonymous labels will require some degree of manual review.

Any and all errors of these types should be corrected.

Error Check

Extract Terms

A SPARQL query should *optionally* extract ID, label, and definition information to allow identification of classes with missing info. This is the query used:

```
SELECT ?class ?id ?label ?def
WHERE {
    ?class a owl:Class .

    OPTIONAL { ?class oboInOwl:id ?id . }
    OPTIONAL { ?class rdfs:label ?label . }
    OPTIONAL { ?class obo:IAO_0000115 ?def . }
}
```

Check for Requirements Missing

IDs

class	id
http://purl.obolibrary.org/obo/SYMP_0019182	NA
http://purl.obolibrary.org/obo/SYMP_0019183	NA
http://purl.obolibrary.org/obo/SYMP_0019184	NA

class	id
http://purl.obolibrary.org/obo/SYMP_0019185	NA
http://purl.obolibrary.org/obo/SYMP_0019186	NA
http://purl.obolibrary.org/obo/SYMP_0019187	NA

Labels

None missing.

Namespace

class	ns
http://purl.obolibrary.org/obo/SYMP_0019182	NA
http://purl.obolibrary.org/obo/SYMP_0019183	NA
http://purl.obolibrary.org/obo/SYMP_0019184	NA
http://purl.obolibrary.org/obo/SYMP_0019185	NA
http://purl.obolibrary.org/obo/SYMP_0019186	NA
http://purl.obolibrary.org/obo/SYMP_0019187	NA

Check for Error-producing Duplicates

URIs

No duplicates.

IDs

class	id	label
http://purl.obolibrary.org/obo/SYMP_0000101	SYMP:0000101	partial paralysis
http://purl.obolibrary.org/obo/SYMP_0000624	SYMP:0000101	obsolete appendicitis

Labels

No duplicates.

Synonymous Labels as Separate Terms

To identify synonymous labels that are separate terms requires manual review. To make review of the 950 terms easier, review will be based on the assumption that synonyms will be very similar to one another with only a few characters changed. The approximate string matching distance is calculated and only those with less than four character changes are reviewed.

Only the synonyms identified are listed here:

id1	symp1	id2	symp2
SYMP:0000683	tachypnoea	SYMP:0000603	tachypnea
SYMP:0000879	low-grade fever	SYMP:0000081	low grade fever

Incorrect critical information

Namespace incorrect

The Symptom ontology namespace should be “symptoms”.

```
## All non-missing namespaces are correct, equal to 'symptoms'.
```

URI != ID

This issue once cropped in the DO. The ID number should match the class URI number. For these it does not.

class	id	label
http://purl.obolibrary.org/obo/SYMP_0000624	SYMP:0000101	obsolete appendicitis
http://purl.obolibrary.org/obo/SYMP_2000000	SYMP:0020000	airway stenosis
http://purl.obolibrary.org/obo/SYMP_2000001	SYMP:0020001	apthous ulcer
http://purl.obolibrary.org/obo/SYMP_2000002	SYMP:0020002	anal abscess
http://purl.obolibrary.org/obo/SYMP_2000003	SYMP:0020003	acholic stool
http://purl.obolibrary.org/obo/SYMP_2000004	SYMP:0020004	ageusia
http://purl.obolibrary.org/obo/SYMP_2000005	SYMP:0020005	acquired color vision deficiency
http://purl.obolibrary.org/obo/SYMP_2000006	SYMP:0020006	agitation
http://purl.obolibrary.org/obo/SYMP_2000007	SYMP:0020007	anterior scleral inflammation