ELSEVIER

International Conference in Computational Science (ICCS 2010)

Exploring ontology metrics in the biomedical domain

N. Manouselis[a], M.A. Sicilia[b], D. Rodríguez[b]♣

[a]*Agricultural University of Athens, 75 Iera Odos str., Botanikos, GR-118 55, Athens, Greece*
Dept of Computer Science, University of Alcalá, Ctra. Barcelona km. 33.6, 28871, Alcalá de Henares, Madrid, Spain

Abstract

Ontologies are gaining popularity in many domains as a way of representing, dealing and reasoning with large volumes of information, and they are starting to play a major role in computational science applications, e.g. in biomedical studies. The creation and maintenance of ontologies has become an engineering process that needs to be managed and measured. Therefore, as part of any ontology engineering processes, metrics can play a role helping in identifying possible problems or incorrect use of ontology elements. However, empirical studies that measure existing ontologies and set standard values are required for metrics to be effectively used. This paper reports a preliminary exploration of ontology metrics from a large set of ontologies extracted from the Open Biomedical Ontologies (OBO) repository. In this paper, we also present an ontology metrics tool, called Ontometrics, used to collect the set of metrics presented. Results show that such metrics can help identifying peculiarities of the ontologies that could be used as the basis for future ontology metric evaluation studies.

1. Introduction

Ontologies [4] are explicit representations of domain concepts and their relationships. More formally, an ontology defines the vocabulary of a problem domain and a set of constraints on how terms can be combined to model the domain, using a formal language. Like in the software engineering domain, generally the construction of ontologies follows an iterative and incremental approach and a number of metrics has been proposed to measure different quality criteria such as reliability, reusability, cohesion, etc. (see for example [5]). However, unlike in the software engineering and other domain, there is a lack of empirical studies that validate how the different metrics are capable of judging quality properties and their interpretation.

This lack of understanding of the proposed metrics calls for studies that explore and characterize the different measurable aspects of existing ontologies that are considered to have some degree of maturity. The results of these studies could then be used to propose interpretations for the metrics that can be subject to hypothesis contrast and

♣ * Corresponding author. Tel.: +34-918856534; fax: +34-918856646.
E-mail address: daniel.rodriguezg@uah.es

eventually lead to a practical applicability of these metrics that relies on a reliable base of empirical studies and contrasts.

This paper reports on an initial exploratory study on measuring a set of ontologies published in the Open Biomedical Ontologies (OBO) repository[a] [7] which are collected using a common topical criteria and maintained with a coherent set of tools. The study has been carried out by implementing and using an open source software framework for computing ontology metrics expressed in the Ontology Web Language (OWL) [3]. OWL has been chosen as the language is a W3C recommendation and it is a common and widespread language for expressing formal ontologies. The ontologies in OBO have been downloaded in OWL format and a limited number of existing ontology metrics proposed were computed. The overall figures found are reported, along with an exploratory study on potential categories of ontologies with diverging characteristics for which different metric interpretation or different quality criteria could be appropriate.

The rest of the paper is structured as follows. Section 2 covers the background on ontology metrics followed by a short description of the framework used in Section 3. Next, Section 4 describes the metrics analyses performed. Finally, Section 4 concludes the paper and outlines future work.

2. Background and overview

Although ontologies have been a resource in the field of artificial intelligence since mid 1970's, from the inception of the Semantic Web, in which ontologies are the principal recourse to integrate and deal with online information, a new set of standards to represent has been proposed. The Ontology Web Language (OWL) [3] is one of such standards that belongs to a family of knowledge representation languages created for the Semantic Web. The OWL has reached the status of W3C[b] (World Wide Web Consortium) recommendation. From a technical point of view OWL extends the RDF (Resource Description Framework) and RDF-S (RDF Schema) languages allowing us to integrate a variety of applications using XML as interchange syntax. There are three OWL flavours, OWL Lite, OWL DL (Description Logics) and OWL Full, being the OWL depending on the expressiveness and reasoning capabilities provided.

In a nutshell, OWL ontologies are composed of (i) *classes* that can be nested as sets of individuals, (ii) *individuals* as instances of classes, i.e., objects of the domain and (iii) *properties* as binary relations between individuals. It is also possible to specify property domains, cardinality ranges and reasoning on ontologies. From these basic elements a number of authors have proposed metrics to measure the quality of ontologies.

For example, Yao et al [1] proposed a set on metrics to measure the cohesion in ontologies such as the number of root classes, number of leaf classes and average depth of inheritance tree of all leaf nodes. Authors validated the metrics both theoretically and empirically. However, their empirical validation consisted of comparing statistically the results of the metrics and subjective evaluation of eighteen evaluators.

Tatir et al. [2] propose three orthogonal dimensions to evaluate the quality of an ontology. The first dimension is the quality of the ontology schema representing the real world. The second dimension evaluates the content of the ontology, i.e, how well populated is the ontology and if it reflects the real world. Finally, the third dimension evaluates if the set of instances and relations agree with the schema. Schema metrics proposed by the authors include relationship, attribute and inheritance richness. Instance metrics are divided into (i) knowledge base metrics such as class richness, average population and cohesion and (ii) class metrics which include importance, fullness, inheritance richness, relationship richness connectivity and readability. Authors analysed two general purpose ontologies SWETO and TAP as well as the GlycO (Glycomics *Ontology*) ontology.

From the set of metrics proposed in the literature, a subset has been selected as a first target. In particular in the OntoMetrics framework reported in this work, the following metrics have been considered for the development of the present exploratory study:

─────────

[a] http://www.obofoundry.org/
[b] http://www.w3c.org/

- No. of Classes (*noc*) is a simple count of the number of classes contained in the ontology.
- No. of Instances (*noi*) is a simple count of the number of instances contained in the ontology.
- No. of Properties (*nop*) is a simple count of the number of properties contained in the ontology.
- *Number of Root Classes* metric (*norc*) [1] corresponds to the number of root classes (those without superclasses) explicitly defined. The range of this metric is from 1 to ∞. There is at least one root and the larger the number of root classes, the more diverse the ontology will be.
- *Number of Leaf Classes* metric (*nolc*) is the sum of all leaf classes, i.e., those without subclasses, in an ontology. As a root class without inheritance can also be considered a leaf class, and the range of this metric is from 1 to ∞.
- *Average Population* metric (*ap*) [2] measures the average distribution of instances across all classes. Formally, it is defined as:

$$AP = \frac{|I|}{|C|}$$

  where *I* represent the instances and *C* represent the classes. According the authors, this metric is supposed to be used in conjunction with the *Class Richness* metric as an indication if there is enough information in the ontology.
- *Class Richness* metric (*cr*) is the ratio between the number of classes that have instances divided by the total number of classes. Formally, CR is defined as:

$$CR = \frac{|C'|}{|C|}$$

  where *C´* is the number of classes with instances (used in the ontology) and *C* is the total number of classes defined in the ontology. This metric provides an indication of how many instances are really related to classes defined in the schema.
- *Explicit Depth of Subsumption Hierarchy* (*dosh*)
- *Relationship Richness* metric (*RR*) [2] is defined as the ratio of the number of relationships defined in the schema divided by the sum of the number of subclasses.

$$RR = \frac{|p|}{|sc| + |p|}$$

  where *p* is the number of relationships and *sc* is the number of subclasses, which is equivalent to the number of inheritance relationships plus the number of relationships) [2]. This metric reflects the diversity of relations and placement of relations in the ontology. An ontology that contains many relations other than class-subclass relations (values close to 1) is richer than a taxonomy with only class-subclass relationships (values close to zero).
- *Inheritance Richness* metric (*ir*) [2] defined as the average number of subclasses per class. Formally, *IRM* is defined as:

$$IR = \frac{\sum \left| H^c \left(C_1, C_i \right) \right|}{|C|}$$

  where the dividend is the sum of the number of subclasses (C_1 – one level) for a class C_i which is defined as $H^c(C_1, C_i)$. The divisor (*C*) is the total number of classes in the ontology. According to the authors, this metric represents the distribution of information across different levels of the ontology's inheritance tree and serves as an indicator of how well knowledge is grouped into different categories and subcategories in the ontology. Values close to zero indicate flat or horizontal ontologies representing perhaps more general knowledge while large values represent vertical ontologies describing detailed knowledge of a domain.

Table 1 summarizes the main metrics reported on the literature, classified by the kind of elements considered.

Table 1 Summary of ontology metrics used in this work

Metric	Acronym
No. of Classes	*noc*
No. of Instances	*noi*
No. of Properties	*nop*
No. of Root Classes	*norc*
Number of Leaf Classes	*nolc*
Average Population	*ap*
Class Richness	*cr*
Explicit Depth of Subsumption Hierarchy	*dosh*
Inheritance Richness	*ir*
Relationship Richness Metric	*rr*

3. Data collection using the OntoMetrics framework

In order to collect the metrics from the OWL files and the OBO repository, the OntoMetrics framework[c] has been developed. The OntoMetrics is an open source Java implementation that makes use the Java libraries of Protégé[d] an ontology editor that provides an application programmer's interface (API) for loading, saving and manipulating OWL and RDF files. The metrics computed were extracted from the text of the OWL files, with no associated reasoning process. This entails that the metrics do not consider elements that would be potentially inferred from reasoners, e.g. automatic classifications or subsumption relationships that are not explicit. While this may seem a major limitation, the fact that most of current ontologies in the biomedical domain do not contain axioms or defined classes (classes expressed through logical necessary and sufficient conditions) suggests that explicit elements are enough for the empirical assessment of metrics in the current state of development of ontologies.

The implementation of the metrics has in some cases be made through approximation or simplification of the original metric formulae. Only elements explicitly declared in the ontology were considered as ontology elements can also be created from reasoning or inference. The general case of OBO ontologies is that they are not equipped with these mechanisms so that aspect was discarded. Then, in the case of number of classes and number of root and leaf classes, named classes were gathered through the `OWLModel.getUserDefinedOWLNamedClasses()` method. Instances are counted through `OWLModel` method `getUserDefinedRDFIndividuals()` and properties through `getUserDefinedOWLProperties()`. Average population and class richness uses the previous metrics to relate number of classes to number of instances. Depth of subsumption hierarchy and inheritance richness is computed recursively via the `OWLClass.getSubclasses()` method, for the latter traversing all the list of declared classes. The same operation is combined with `OWLClass.getAssociatedProperties()` to compute relationship richness.

Appendix A provides the raw data obtained from the OWL files described in the OBO Foundry repository.

4. Analysis of ontologies in OBO Foundry repository

As stated previously, in this work we have used a limited set of ontologies from the OBO Foundry repository. The OBO Foundry is a collaborative project that aims to both establish principles for ontology development and to create ontologies in the biomedical domain. The site also provides relevant literature, mappings between different ontologies and specialised tools for such domain. Currently, the repository is composed of around one hundred ontologies mainly using the OBO format used by the OB-OEdit tool[e], but most ontologies are also translated into alternative formats including the OWL, which is also made available openly for download. The following three ontologies NIF_Cell.owl, NIF_Dysfunction.owl and relationship.owl were discarded as the ontology metrics software was not able to compute all the metric set for them.

---

[c] Available at http://sourceforge.net/projects/ontometrics/

[d] http://protege.stanford.edu/

[e] http://www.oboedit.org/

Table 2 provides basic descriptive statistics for the set of 75 ontologies studied. From the figures in the Table, we can conclude that biomedical ontologies are relatively large ones when considering their t-boxes, with an average of more than 3000 explicitly declared ones. However, there are some that are much smaller.

Table 2 Descriptive Statistics for the OBO Foundry repository

	ap	cr	dosh	Ir	noc	noi	nolc	nop	norc	rr
Count	75	75	75	75	75	75	75	75	75	75
Avg	2.64	0.01	10.12	0.93	3169.75	11318.8	2490.52	15.41	496.23	0.44
Variance	15.16	0.00	37.27	0.09	8.37E+07	2.87E+09	5.36E+07	947.27	3.65E+06	0.12
StdDev	3.89	0.05	6.10	0.30	9148.88	53541.6	7318.05	30.78	1910.5	0.35
Min	0	0	1	0	34	0	21	0	1	0
Max	31.76	0.41	41	1.62	75529	455734	60858	194	13737	1
Range	31.76	0.41	40	1.62	75495	455734	60837	194	13736	1
Std Sk	20.82	25.98	10.00	-4.93	24.66	28.26	25.18	13.08	20.12	0.01
Std Kr	76.05	104.35	20.86	5.41	96.18	117.62	99.54	29.22	61.96	-2.43

Table 3 Pearson's correlations

	ap	cr	dosh	ir	noc	noi	nolc	nop	norc	rr
ap		-0.027	0.141	0.24•	0.084	0.9••	0.051	-0.138	-0.089	-0.208
	—	0.82	0.227	0.038	0.473	0	0.663	0.238	0.446	0.074
cr			-0.094•	0.064	-0.094•	-0.059•	-0.093	0.356•	-0.074	0.036
		—	0.422	0.584	0.422	0.613	0.428	0.002	0.526	0.761
dosh				0.448•	0.224	0.24•	0.185	0.313•	-0.263•	-0.02
			—	0.000	0.054	0.038	0.112	0.006	0.023	0.868
ir					0.087	0.248•	0.024	0.14	-0.619•	-0.438•
				—	0.458	0.032	0.839	0.230	0	0
noc						0.192	0.995••	0.6•	0.153	0.174
					—	0.099	0	0	0.191	0.135
noi							0.151	-0.069	-0.024	-0.149
						—	0.196	0.554	0.840	0.202
nolc								0.603•	0.211	0.2
							—	0	0.07	0.085
nop									-0.1	0.474
								—	0.393	0
norc										0.338•
									—	0.003
rr										—

These variables do not follow the normal distribution as it is apparent from the skewness and kurtosis measures.

Next step in the exploratory study is contrasting correlations among variables. Table 3 shows the Spearman rank correlation coefficient between each pair of variables with a range between -1 and +1 as well as the p-value of its statistical significance. Those *p-values* below 0.05 are represented with a "•" and indicate statistically significant non-zero correlations at the 95% confidence level. More relevant, however, is that for the OBO repository there is a couple of pairs that are very positively correlated (marked with"••"): (i) *noi* (no. of instances) and *ap* (average population), and (ii) *noc* (number of classes) and *nolc* (number of leaf classes). This suggests that ontologies are in general quite flat and most classes contain single instances.
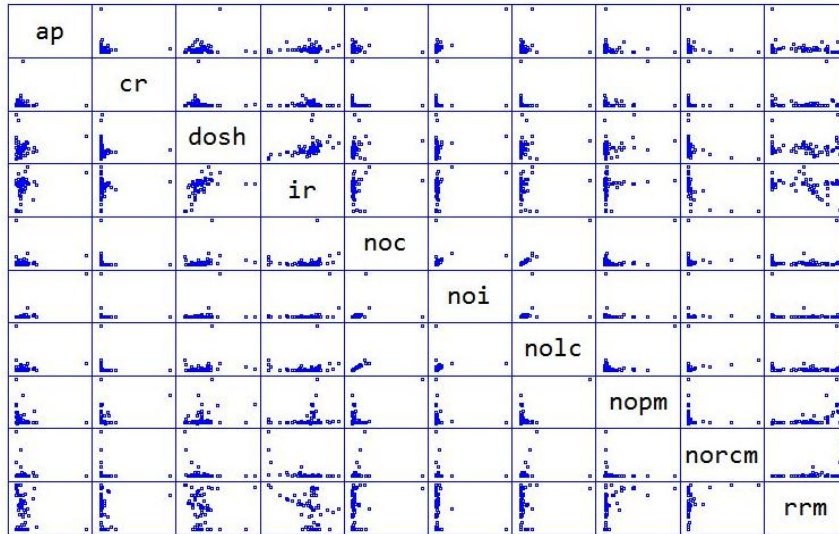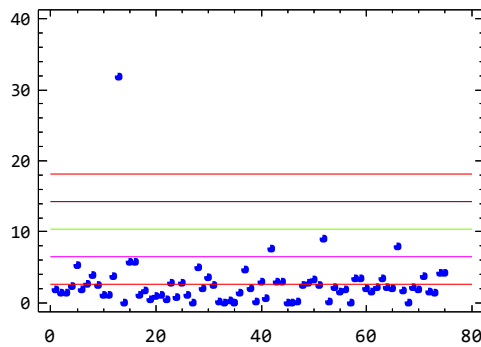
Figure 1 Correlations

Figure 2 Outlier plot for the *ap* variable with std deviations

We next performed outlier detection analysis for each variable. To do so, the Studentized values measure how many standard deviations each value is from the mean and a few ontologies were consistently classified as outliers. These were the NCI Thesaurus (`ncithesaurus.owl`) and the `disease_ontology.owl`. Both are extremely large ontologies (140MB and 150Mbytes respectively). For example, using the *ap* (average population) variable, the most extreme value corresponds to the NCI thesaurus which is 7.48 standard deviations from the mean. A closer look into these ontologies reveals that these are special ontologies. The NCI thesaurus [8] is used to define a vocabulary of the cancer domain and related diseases, and not a formal ontology in an strict sense. The second one also defines a vocabulary of human diseases based on previous thesauri and terminologies. These can therefore be considered special ontologies that deserve separate examination.

5. The relationship between the rest of the metrics were contrasted using factor analysis.

Table **4** shows the resulting factor loadings using principal component analysis as an extraction method with 3 components were extracted.

Table 4 Factorial Analysis

	Component		
	1	*2*	*3*
cr	.068	.217	-.753
dosh	.666	.285	.230
ir	.893	-.122	-.088
noc	.210	.728	.371
noi	.394	-.007	.561
nop	.259	.891	-.272
norc	-.722	.238	.371
rr	-.433	.704	-.036

The first component is characterized by high values in *dosh* and *ir*, that roughly measure depth and breadth of the subsumption hierarchy. Not surprisingly, this component is negatively correlated with *norc*, i.e. these ontologies tend to have fewer roots of hierarchy trees. The second component is characterized by a high correlation with the number of classes and properties and also with relationship richness that relates both of them. The third component is correlated with the number of instances.

6. Conclusions and future work

Assessing the validity of ontology metrics is challenging as of today. This paper addressed such a problem presenting a ontology metrics framework, OntoMetrics, and a preliminary evaluation of a set of well established ontologies from the OBO foundry repository. Although it is difficult to establish concrete thresholds to define when an ontology needs to be improved, such metrics can help identifying problems in ontologies or uses of ontologies for special purposes (e.g., as a thesaurus). Therefore, metrics can be used in conjunction with domain knowledge as part of an ontology engineering process. The analysis done has pointed out to relationships between groups of metrics that can be used to devise new combined metrics characterizing different kinds of ontologies.

Future work will be directed to further develop our metrics framework with new metrics and further analysis of the metrics. We intend to perform comparisons between mature and not so mature ontologies to further validate how metrics can be used to evolve and improve ontologies. This can be done for example, using raking metrics such as those defined by Ho *et al.* [6] and inter-ontology metrics.

Acknowledgements

References

[1] Yao, H., Orme, A.M. and Etzkorn, L. (2005) Cohesion Metrics for Ontology Design and Application. *Journal of Computer Science* 1(1): 107-113.

[2] Tartir, S., Arpinar, I.B.,Moore, M., Sheth, A.P., Aleman-Meza, B.: OntoQA:Metric-based ontology quality analysis. In: Proc. of Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources (2006).

[3] M.K. Smith, C. Welty, and D.L. McGuinness, OWL Web Ontology Language Guide". W3C, 2004, `http://www.w3.org/TR/owl-guide/`

[4] T.R., Gruber, Toward Principles for the Design of Ontologies Used for Knowledge Sharing, *International Journal of Human-Computer Studies*, **43**(5-6) (1995) 907-928.

[5] M. Uschold, M.Grüninger,: Ontologies: Principles, Methods, and Applications, *Knowledge Engineering Review*, **11**(2) (1996) 93-113.

[6] Hu B., Dasmahapatra S., Lewis, P., Semantic metrics, *International Journal of Metadata, Semantics and Ontologies*, **2**(4), (2007) 242-258.

[7] Smith, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration, *Nature Biotechnology*, **25**, (2007) 1251-1255

[8] Ceusters W, Smith B. A Terminological and Ontological Analysis of the NCI Thesaurus. *Methods of Information in Medicine*, **44** (2005) 498-507.

Appendix

Table 5 Metrics collected from the OBO repository

Ontology	ap	cr	dosh	ir	noc	noi	nolc	nop	norc	rr
amphibian_anatomy.owl	1.83	0	5	0.26	706	1,290	656	3	525	0.78
ascomycete_phenotype.owl	1.4	0.01	7	0.97	324	455	253	1	10	0
bilateria_mrca.owl	1.38	0.02	7	0.77	120	166	91	6	28	0.34
brenda.owl	2.36	0	9	0.73	4,007	9,444	3,298	5	1,323	0.57
biological_process.owl	5.2	0	15	1.62	18,578	96,604	10,193	6	7	0.45
caro.owl	1.87	0.06	7	0.87	52	97	34	2	7	0.04
cell.owl	2.69	0	13	1.21	1,055	2,843	697	2	20	0.33
cellular_component.owl	3.81	0	11	1.49	2,762	10,522	2,159	6	7	0.39
cereal_anatomy.owl	2.51	0	10	0.82	874	2,198	623	3	206	0.53
dendritic_cell.owl	1	0.02	6	0.93	180	180	150	10	12	0.55
dendritic_cell_prerelease.owl	1	0.02	6	0.93	180	180	150	10	12	0.55
dictyostelium_discoideum_anatomy.owl	3.73	0.02	4	0.95	144	537	113	3	7	0.59
disease_ontology	31.76	0	21	1.42	14,348	455,734	9,829	2	7	0
emap.owl	0	0	1	0	13,737	0	13,737	2	13,737	1
envo.owl	5.71	0	10	0.99	1,239	7,075	916	5	75	0.31
envo_prerelease.owl	5.71	0	10	0.99	1,239	7,075	916	5	75	0.31
event.owl	1.12	0	14	1.03	2,787	3,135	2,011	2	280	0.43
evidence_code.owl	1.64	0.02	7	1.03	155	254	104	1	7	0
evoc.owl	0.49	0	8	0.98	1,006	488	811	1	16	0
fix.owl	0.88	0	16	1.37	1,169	1,028	605	6	27	0.44
fly_anatomy.owl	1.09	0	14	1.16	6,512	7,088	4,999	13	356	0.54
flybase_vocab.owl	0.49	0.01	8	1	719	355	597	2	7	0.04
fly_development.owl	2.84	0.02	4	0.95	138	392	132	4	7	0.64
fly_taxonomy.owl	0.81	0	36	1	6,605	5,350	5,030	1	7	0
fungal_anatomy.owl	2.72	0.04	8	0.95	79	215	45	4	15	0.27
gemina_symptom.owl	1.1	0	8	1	820	898	674	2	7	0
gene_regulation.owl	0.01	0	11	1.05	439	3	303	27	4	0.78
homology_ontology.owl	5	0.04	6	1.06	72	360	43	1	7	0
human-dev-anat-staged.owl	2	0	2	0	8,346	16,685	8,345	3	8,323	1
human_phenotype.owl	3.55	0	14	1.19	9,552	33,890	6,633	1	7	0
hymenoptera_anatomy.owl	2.4	0	10	1	1,104	2,649	954	2	8	0.48
image.owl	0.15	0	7	0.97	265	39	215	2	7	0
infectious_disease_ontology.owl	0.04	0.01	11	1.01	493	19	404	1	4	0
information_artifact.owl	0.32	0.04	9	0.99	159	51	115	46	2	0.87
lipid.owl	0	0	15	1.01	715	0	514	46	1	0.95
loggerhead_nesting.owl	1.31	0.01	11	0.98	314	410	196	4	14	0.51
mammalian_phenotype.owl	4.65	0	16	1.17	7,428	34,506	4,779	2	7	0
mao.owl	1.98	0.02	4	0.55	170	336	151	5	77	0.64
medaka_anatomy_development.owl	0.07	0	5	0.04	4,364	315	4,353	4	4,195	0.96
mged.owl	2.94	0.41	9	1.1	233	685	184	121	1	0.77
minimal_anatomical_terminology.owl	0.6	0	2	0.01	467	279	464	20	463	0.99
mosquito_anatomy.owl	7.53	0	7	0.42	1,867	14,067	1,716	4	1,130	0.7
mosquito_insecticide_resistance.owl	2.86	0	10	0.99	4,322	12,340	4,008	13	27	0.01
mouse_pathology.owl	2.94	0	9	1.13	649	1,908	517	3	8	0
Ncithesaurus	0	0	17	1.12	75,529	0	60,858	194	20	0.92
obi.owl	0.05	0.01	41	1.03	2,638	141	1,452	83	7	0.94
OGMS.owl	0.15	0.03	8	0.98	115	17	72	0	2	0
pathway.owl	2.44	0	8	1.14	724	1,766	452	2	43	0.11
plant_environment.owl	2.82	0.01	9	1.02	506	1,428	380	2	7	0
plant_trait.owl	3.18	0	10	1.13	1,028	3,266	711	2	22	0.09
po_anatomy.owl	2.51	0	10	0.82	874	2,198	623	3	206	0.53
po_temporal.owl	8.95	0.01	8	0.95	286	2,559	238	2	14	0.04
provenance.owl	0.12	0.05	8	1.04	398	48	263	33	10	0.85
psi-ms.owl	2.1	0	6	1.11	1,445	3,037	1,251	4	101	0.48
quality.owl	1.5	0	11	1	2,121	3,174	1,766	22	7	0.78
rex.owl	1.78	0.01	13	1.16	558	991	316	7	34	0.48
sao.owl	0.02	0	11	1	763	18	503	81	2	0.92
sequence.owl	3.46	0	11	1.04	1,747	6,050	1,315	40	12	0.43
sequence_prerelease.owl	3.46	0	11	1.04	1,747	6,050	1,315	40	12	0.43
spatial.owl	2.02	0.02	3	0.89	135	273	120	34	15	0.6
spider_anatomy.owl	1.59	0.01	6	0.67	566	899	444	2	235	0.57
systems_biology.owl	2.15	0.01	10	1.09	503	1,079	326	2	7	0
teleost_anatomy.owl	3.4	0	12	0.72	2,666	9,070	2,303	5	851	0.63
tick_anatomy.owl	2.13	0	7	0.45	634	1,348	606	2	366	0.67
transmission.owl	2.03	0.09	7	0.79	34	69	21	1	7	0
uberon.owl	7.87	0	11	1.09	3,171	24,952	2,732	54	22	0.93
unit.owl	1.64	0	11	1	2,392	3,932	1,960	23	8	0.76
vaccine.owl	0.03	0	20	1.01	1,641	51	1,149	79	4	0.92
worm_anatomy.owl	2.15	0	9	0.77	7,063	15,211	6,533	6	2,778	0.57
worm_development.owl	1.79	0.03	7	0.91	78	140	53	2	7	0
worm_phenotype.owl	3.74	0	13	1.06	1,877	7,026	1,160	2	110	0.15
xenopus_anatomy.owl	1.56	0	7	0.61	823	1,282	748	6	322	0.78
yeast_phenotype.owl	1.36	0.01	7	0.97	301	408	232	1	10	0
zebrafish_anatomy.owl	4.16	0	12	0.84	2,437	10,142	2,085	5	452	0.79
zebrafish_anatomy_prerelease.owl	4.16	0	12	0.84	2,437	10,142	2,085	5	452	0.79