A Quality Assurance Methodology for ChEBI Ontology Focusing on Uncommonly Modeled Concepts

Hao Liu[1], Ling Chen[2], Ling Zheng[3], Yehoshua Perl[1], James Geller[1]

[1]Department of Computer Science
New Jersey Institute of Technology,
Newark, NJ USA
{hl395, perl, geller}@njit.edu

[2]Borough of Manhattan Community
College CUNY
New York, NY USA
lchen@bmcc.cuny.edu

[3]CSSE Department
Monmouth University,
West Long Branch, NJ USA
zdzhengling@gmail.com

***Abstract*—The Chemical Entities of Biological Interest (ChEBI) ontology is an important knowledge source of chemical entities in a biological context. ChEBI is large and complex, making it almost impossible to be error-free, given the scarce resources for quality assurance (QA). We present a methodology to locate concepts in ChEBI with a high probability of being erroneous. An Abstraction Network, which provides a compact summarization of an ontology, supports the methodology. By investigating a sample of ChEBI concepts, we show that uncommonly modeled concepts residing in small units of the Abstraction Network of ChEBI are statistically significantly more likely to have errors than other concepts. The finding may guide ChEBI ontology curators to focus their limited QA resources on such concepts to achieve a better QA yield. Furthermore, this study, combined with previous work, contributes to progress in showing that this methodology can be applied to a whole family of similar ontologies.**

Keywords—ChEBI; chemical ontology; chemical concept; quality assurance; modeling error;

I. INTRODUCTION

The Chemical Entities of Biological Interest (ChEBI) ontology [1] is a large important knowledge source that facilitates reference to chemical entities within the biological field. It annotates small distinguishable entities such as atoms, ions, and polymers and their relationships to each other. ChEBI has been used to support chemical analysis. For example, a method for determining optimal semantic similarity and particularity thresholds was applied to the ChEBI ontology [2].

Quality assurance (QA) is an essential part of the ontology lifecycle to make sure that there is no modeling error in an ontology [3]. Errors in an ontology can propagate to its applications. However, due to limited available QA resources, it is impossible to perform thorough quality assurance on a large ontology like ChEBI or on a large hierarchy of an ontology, e.g., ChEBI's *Chemical Entity* hierarchy with 106,707 concepts (in July 2017), without automatic/semi-automatic techniques.

One practical ontology QA approach is to identify sets of concepts, which are more likely to have errors than other concepts. Focusing audits on such sets of concepts can achieve a high QA yield in terms of the ratio of the number of concepts with modeling errors to the number of reviewed concepts. Thus, the question is how can we find such sets of concepts?

The SABOC team has previously demonstrated that Abstraction Networks are an effective tool to identify sets of concepts in ontologies that are more likely to have errors [4]. An Abstraction Network (AbN) is a compact summary of an ontology's content and structure, which is automatically derived from the ontology. There are different kinds of AbNs depending on the ontologies' structure. The *Partial-area taxonomy* (introduced in the Background) is an AbN that was derived for several ontologies [3, 5], e.g., NCI Thesaurus (NCIt) [6], and SNOMED CT [7].

A partial-area (explained in the Background section) represents a group of concepts with the same structure and semantics, that is, all have the same set of relationships and are descendants of the same concept. Concepts in partial-areas that summarize few concepts ("small partial-areas") expose their uncommon modeling, which is of special interest to us.

A modeling error in an ontology can be either an omission error or a commission error. Examples of omissions are missing a parent concept or missing a lateral relationship. Examples of commission errors are an incorrect parent or an incorrect lateral relationship. Omission errors represent knowledge in an ontology that is not complete, while commission errors mean that the modeling is wrong. Some ontologies, e.g. NCIt, are intentionally modeled with incomplete knowledge. Hence, commission errors are more severe and attract more interest from ontology curators. A previous study [8] on a random sample of 400 concepts in the February 2016 release of ChEBI, found that 41.8% of the concepts exhibited errors. Thus, we focused in this study on commission errors with the goal to find a semi-automatic QA technique, which can identify ChEBI concepts with a higher probability of commission errors.

The purpose of this study is to test whether the partial-area taxonomy-based QA methodology, focusing on **concepts in small partial-areas**, improves the error yield for the large *Chemical Entity* hierarchy of ChEBI.

II. BACKGROUND

A. ChEBI

ChEBI [9] is maintained by the European Molecular Biology Laboratory–European Bioinformatics Institute (EMBL-EBI). Various applications have been developed based

on the ChEBI ontology. A prediction method [10] was proposed to utilize information from the ChEBI ontology for identifying drugs' target groups. Hill et al. [11] integrated the ChEBI structural hierarchy into the Gene Ontology to enable data integration across the biology and chemistry domains.

ChEBI is provided in the W3C standard Web Ontology Language (OWL) and OBO formats. In this study, using the OWL format, object properties (relationships) are used only in restrictions. There is a stated version and an inferred version of OWL files. ChEBI provides the stated version. The reasoner HermiT [12] was used to get the inferred version of ChEBI.

The ChEBI ontology consists of three hierarchies. In this study we used concepts from the largest, the *Chemical Entity* hierarchy, (98.7% of ChEBI concepts). It annotates chemical compounds within molecular entities. Besides the IS-A relationships, ChEBI's *Chemical Entity* hierarchy has nine lateral semantic relationship types, such as *has part*, *has role*, and *is conjugate base of*. The other two hierarchies are the *Subatomic Particle* and the *Role* hierarchy. The former hierarchy is mainly used to categorize particles smaller than atoms, while the *Role* hierarchy defines the roles of molecular entities in different contexts. The ChEBI ontology employs a star rating annotation: "3-star" indicates that a concept was manually annotated by the ChEBI curator team; "2-star" means that the concept is annotated by a third party; and "1-star" usually represents a concept marked as deleted or obsolete. Thus, "1-star" concepts are excluded from our research.

QA requests by users of ChEBI are easily made via ChEBI's GitHub issue tracking system (https://github.com/ebi-chebi/ChEBI/issues). ChEBI's curators review and verify these requests. Approved changes are made available in subsequent releases. For example, a user may report a wrong reference to a certain compound, or a wrong definition or relationship for a compound, by creating an issue report on GitHub. These reports are reviewed on a weekly or monthly basis. If the curators agree with a request, they make the corresponding changes to ChEBI. The volume of requests and the response time reflect the limited QA resources of ChEBI's curator team. At the time of this writing, there are 289 accumulated open issues and 3175 closed requests.

B. Partial-area taxonomy

In our previous research, Abstraction Networks (AbN) have been proven successful for summarizing and visualizing ontologies [4], and for supporting quality assurance of ontologies [13]. Different types of AbNs have been developed for various ontologies. In this study, we use the partial-area taxonomy AbN. Fig. 1 illustrates the derivation of the partial-area taxonomy for an excerpt of concepts from ChEBI's *Chemical Entity* hierarchy.

Fig. 1(a) depicts a subhierarchy of 17 concepts (drawn as ellipses, labeled with their names). The arrows denote IS-A links. Concepts with the same set of lateral relationships are grouped together in a dashed bubble, labeled by the common set of relationships. For example, *Atom*, *Nonmetal atom*, *S-block element atom*, *Polymer*, *Ionic Polymer*, and *Polyanionic polymer* are grouped in the green bubble because they all have only the *has part* relationship.

An area taxonomy is an AbN which consists of nodes called areas and child-of links connecting areas. An area, depicted as a color-coded box, represents the group of concepts with the same set of relationships, i.e. in the same bubble in Fig. 1(a). Fig. 1(b) shows the area taxonomy for Fig. 1(a). An area is labeled by its set of relationship types. The concepts in the green bubble are represented compactly by the green area {*has part*}. The concept *Chemical Entity* and its descendants in the grey bubble are now represented by the area {*Ø*}. (The symbol *Ø* represents the empty set.) Similarly, the concept *Hydrogen atom* and its descendants are represented by the area {*has part, has role*}. Areas with the same number of relationship types have the same color and are aligned at the same level. For example, the areas {*has part, has role*} and {*has part, is conjugate base of*} appear in the third level, colored in blue. Child-of links (arrows in Fig. 1(b)) are derived based on the underlying IS-A relationships in ontologies. See [3] for further details.

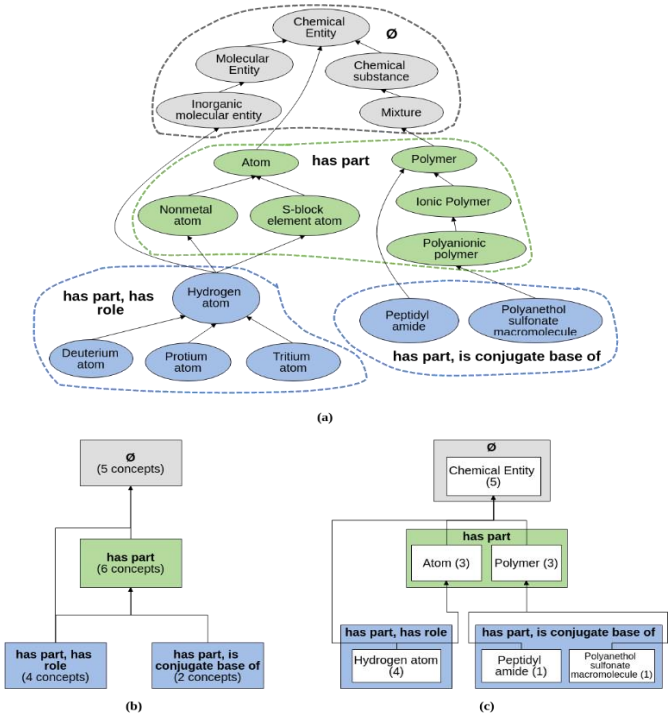

Fig. 1. (a) Excerpt of 17 concepts from ChEBI's *Chemical Entity* hierarchy. (b) Area taxonomy for (a). (c) Partial-area taxonomy for (a).

Areas summarize concepts with the same structure (relationships). An area may have multiple roots, which are concepts such that their parents are not in the area. For example, *Atom* and *Polymer* in the green area {*has part*} are roots, each imposing its semantics on its descendants in the area. For example, the descendants of *Atom* are kinds of atoms and the descendants of *Polymer* are kinds of polymers. The partial-area taxonomy is a refinement of the area taxonomy. A partial-area is composed of an area root concept and all its descendant concepts in the same area. The size of a partial-area is the

number of concepts in it. Each partial-area is labeled by the name of its root concept, expressing its semantics, with its size in parentheses. Partial-areas are shown as white boxes in Fig. 1(c). For example, *Atom (3)* is a partial-area in the green area summarizing three concepts, *Atom* and its two children, in the green area. A partial-area taxonomy is an AbN composed of nodes called partial-areas and hierarchical child-of links (arrows in Fig. 1(c)) connecting them. The compact summarization and visual simplification provided by the partial-area taxonomy, make it easier to identify anomalies in modeling the ontology. An area taxonomy and a partial-area taxonomy can be created automatically by a software tool called Ontology Abstraction Framework (OAF) [14] available in the NCBO BioPortal.

III. METHODS

As described above, there are two kinds of modeling errors in an ontology: omission errors and commission errors. Table I shows one omission error example and five commission error examples. For example, *3-buten-1-amine* in row 1 is missing the relationship *is conjugate base of* (an omission error). In row 4, *(S)-3-hydroxybutyric acid* has an incorrect hierarchical relationship (a commission error). In this study the term "error" will be reserved for commission errors, unless otherwise noted, following the preference of ontology curators described above.

TABLE I. EXAMPLE CONCEPTS WITH COMMISSION AND OMISSION ERRORS

Error Type	Concept Name	Error Description
Omission	3-buten-1-amine	Missing the *is conjugate base of* relationship
Commission: Incorrect relationship target	N-acetyl-D-glucosaminyldiphos -phodolichol	The charge for the target of its relationship *is conjugate acid of* should be 1- not 2-.
Same as above	γ -Glu-Glu	Same as above
Commission: Incorrect hierarchical relationship	(S)-3-hydroxybutyric acid	Its grandparent *(ω−1)-hydroxy fatty acid* is incorrect, should be *(ω−2)-hydroxy fatty acid*.
Same as above	zorbamycin	*zorbamycin* is a secondary amide, not a primary amide.
Same as above	cefaclor	Same as above

Given the fact that concepts within the same partial-area share the same structure and semantics, a partial-area that accommodates just a few concepts stands out as an "outlier," which often needs more QA attention. We consider concepts as "uncommonly modeled" if they appear as outliers through the lens of the partial-area taxonomy. The motivation for auditing concepts in outlier sets is that if a concept is in a small partial-area while related concepts reside in large partial-areas, this raises suspicions about the correctness of the modeling of the concepts in the outlier set. For example, 421 polymer concepts with the same modeling in one partial-area appear to be correctly modeled. However, a partial-area with only two concepts (out of thousands of concepts in the ontology) may indicate error(s). It is, of course, possible that these two uncommonly modeled concepts are correct, but there is a higher possibility that they have errors. Once any errors are corrected, these concepts may become part of another (larger) partial-area. We formulate the following hypothesis.

Hypothesis 1: There exists a threshold value Θ differentiating small and large partial-areas, such that concepts in small partial-areas within the partial-area taxonomy for an ontology have a statistically significantly higher error rate than concepts in large partial-areas.

If Hypothesis 1 is confirmed even for one threshold value, it can be the basis for a QA methodology to guide the ChEBI ontology curators to focus on concepts in the small partial-areas whenever the QA resources are limited. To test the above Hypothesis 1, 500 concepts (0.5%) were randomly selected from the *Chemical Entity* hierarchy of ChEBI's July 2017 release. Concepts were presented in random order to our domain expert LC for review. LC is a chemistry professor with substantial experience in chemical ontology auditing. We analyzed the sizes of partial-areas of these 500 concepts. The evaluation of Hypothesis 1 depends on the threshold value differentiating small partial-areas from large ones.

The threshold Θ may be different for different ontologies. We consider two ways to obtain a threshold value. One is that the threshold is predefined, based on prior experience from other ontologies. Then we can conduct a study to test whether we achieve statistical significance for the error rate difference between small and large partial-areas. For example, in the study by Zheng et al. [15] of NCIt's *Neoplasm* subhierarchy, small partial-areas were predefined as partial-areas with up to 10 concepts and large partial-areas were predefined as those with at least 20 concepts. Partial-areas from 11 to 19 concepts were considered medium-sized. Zheng et al. demonstrated that NCIt *Neoplasm* concepts in small partial-areas have a statistically significantly higher error rate than such concepts in large partial-areas.

The other way is to choose the threshold value that maximizes the error rate difference between small partial-areas and large ones. Hence, the threshold is determined by the study results. In this study, we introduce this second method. Considering the variations in terms of the numbers of concepts for different partial-area sizes, we use the weighted average error rate instead of the average error rate to determine the desired threshold. We call such a threshold an *optimizing threshold*, since it optimizes the difference between the weighted error rates of the two ranges of partial-area sizes. The weighted average error rate is calculated using formula (1), where w_i is the total number of concepts in the partial-areas with the size *i*. E_i is the commission error rate of the reviewed concepts in partial-areas with the size = *i*. $\bar{E}$ is the weighted average commission error rate of all reviewed concepts of the partial-areas with sizes ranging over all existing sizes of small (large) partial-areas, respectively. Thus, the contribution of the concepts in the partial-areas of size *i*, to the weighted average error rate is $E_i * \frac{w_i}{\sum_i w_i}$.

$$\bar{E} = \frac{\sum_i w_i * E_i}{\sum_i w_i} \quad (1)$$

We calculated the weighted average error rate for all possible threshold values and picked the maximizing threshold as value for Θ. We calculated the two-tailed p-value of Fisher's

exact test [16] to evaluate the statistical significance for the optimizing threshold.

IV. Result

The *Chemical Entity* hierarchy had 106,707 concepts in the July 2017 release of ChEBI. There are 27,498 partial-areas in its partial-area taxonomy, from which an excerpt of 164 partial-areas summarizing 92,685 concepts (86.86%) is shown in Fig. 2 created by the OAF tool [14]. Fig. 2 summarizes the content and the structure of most of the *Chemical Entity* hierarchy. It captures the "big picture" of what this hierarchy is about by displaying most of the very large partial-areas. For example, the hierarchy has 35,141 *polyatomic entit(y)(ies)* and 4550 *carbohydrate derivative(s)*. Out of the 500 randomly selected concepts, only 476 concepts were reviewed, excluding 24 concepts with "1-star" that were marked either deleted or obsolete. There were 164 concepts exhibiting commission errors (164/476=34.45%). They were posted on the GitHub site of ChEBI for review by curators.

Table II shows the distribution of concepts and errors in terms of partial-area sizes in the partial-area taxonomy. For each partial-area size *i*, the columns include the number of concepts w_i, the number of audited concepts, the number of concepts with commission errors, and the corresponding error rate E_i. For example, there are 25,798 partial-areas of size = 1, in which 236 concepts were reviewed by LC of which 98 concepts (41.53%) were found to have commission errors. Similarly, there are nine concepts with commission errors out of all 20 audited concepts (45.00%) in partial-areas with size 2.

The last three columns of Table II show the weighted error rate for the small partial-areas, for the large partial-areas, and the error rate difference between the two categories according to the corresponding threshold value equal to the partial-area size in the corresponding row. For example, in row 2, the partial-area size 2 is selected as the threshold to distinguish small partial-areas and large partial-areas. The weighted error rate for the small partial-area category according to formula (1) is 41.71%, while the weighted error rate for the large partial-area category is 26.31%. The error rate difference between them is 15.41%. From Table II we can see that the maximizing threshold value is 2. Thus, we choose $\Theta = 2$ for Hypothesis 1.

Table III is the 2x2 contingency table for the commission erroneous concepts of the small partial-areas and the large partial-areas where the threshold value is the maximizing threshold 2. The count for erroneous concepts and concepts without errors for partial-areas is calculated using data from Table II. According to Table II, 236 concepts in the sample are from partial-areas with size = 1 (row 1) and 20 concepts are from partial-areas with size = 2 (row 2), thus a total number of 256 (=236+20) concepts from small partial-areas were audited. There are 107 (=98+9) concepts with commission errors in the small partial-areas. There are 149 (=256-107) concepts without commission errors from the small partial-areas. Similarly, we can calculate the number of erroneous concepts and the number of concepts without errors for the large partial-areas with size > 2, i.e., 57 and 163.

The two-tailed p-value of Fisher's exact test [16] is 0.0003 (< 0.05) based on Table III. That is, the difference of the weighted average error rates of concepts in small and large partial-areas has statistical significance, and Hypothesis 1 is confirmed.

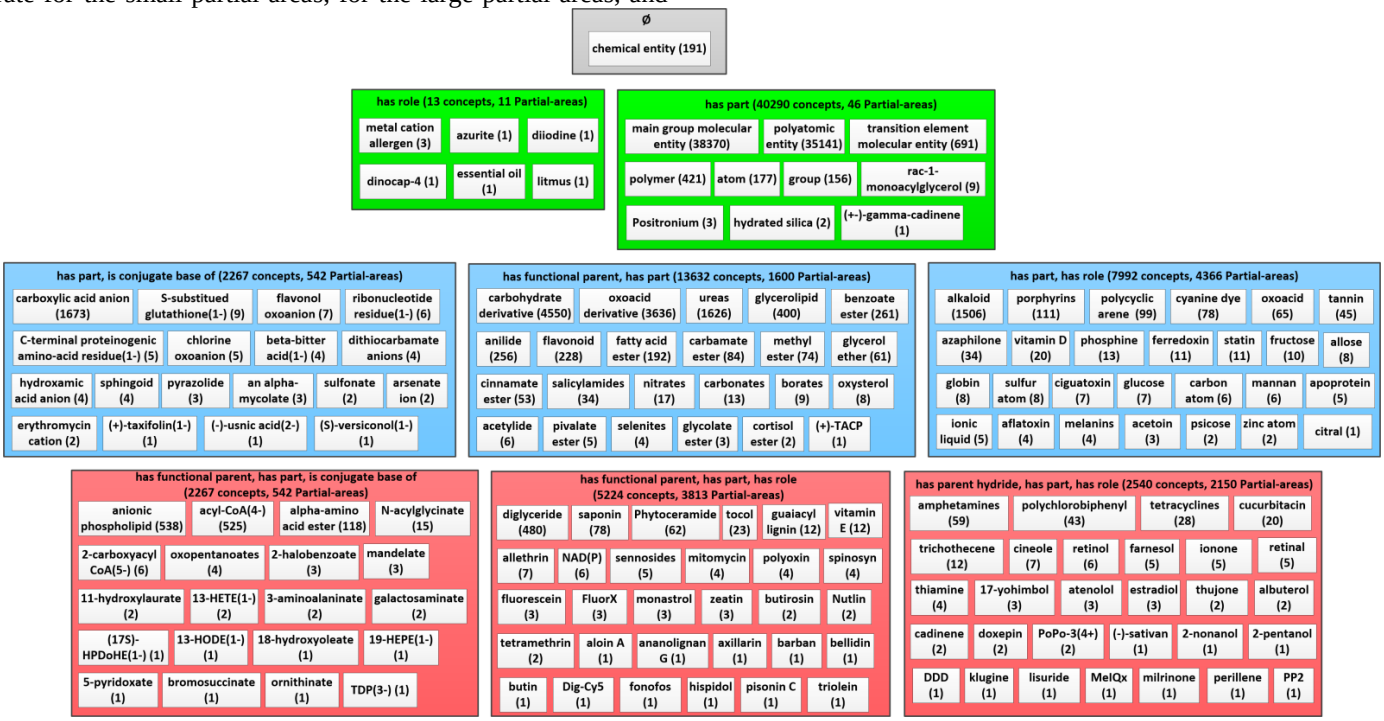

Fig. 2. An excerpt of 164 partial-areas summarizing 92,685 concepts (86.86%) from ChEBI's *Chemical Entity* partial-area taxonomy.

TABLE II. DISTRIBUTION OF CONCEPTS AND ERRORS IN THE PARTIAL-AREA TAXONOMY

Partial-area Size	Total # of Concepts	# of Concepts Audited	# of Concepts w/ Error	Error Rate	Weighted Error Rate for Small Partial-areas	Weighted Error Rate for Large Partial-areas	Error Rate Difference
1	25798	236	98	41.53%	41.53%	26.54%	14.99%
2	1464	20	9	45.00%	41.71%	26.31%	15.41%
3	983	14	1	7.14%	40.51%	26.47%	14.04%
4	475	6	1	16.67%	40.11%	26.51%	13.61%
5	405	5	1	20.00%	39.83%	26.53%	13.30%
6	384	4	1	25.00%	39.64%	26.54%	13.10%
7	308	2	0	0.00%	39.23%	26.61%	12.62%
8	272	1	0	0.00%	38.88%	26.67%	12.20%
9	189	2	2	100.00%	39.26%	26.55%	12.71%
10	130	1	0	0.00%	39.09%	26.58%	12.51%
11	154	0	0	0.00%	38.89%	26.62%	12.28%
12	216	2	0	0.00%	38.62%	26.67%	11.95%
13	91	0	0	0.00%	38.51%	26.69%	11.82%
14	84	1	1	100.00%	38.67%	26.64%	12.04%
15	150	3	1	33.33%	38.65%	26.63%	12.02%
16	112	2	2	100.00%	38.87%	26.55%	12.32%
>16	112057	177	47	26.55%			
Total	143027	476	164	34.45%			

TABLE III. THE 2*2 CONTINGENCY TABLE FOR SMALL AND LARGE PARTIAL-AREAS (P-VALUE = 0.0003)

	# of Concepts w/ Errors	# of Concepts w/o Errors
Partial-areas with size <=2	107	149
Partial-areas with size >2	57	163

Two types of commission errors were reported by our expert: 71 concepts (14.92%) had incorrect hierarchical relationships and 93 (19.54%) had incorrect relationship targets. Table I shows examples with their error descriptions. For example, in row 2, the incorrect target of the relationship *is conjugate acid of* for *N-acetyl-D-glucosaminyldiphosphodolichol* is a concept with 2- charge. It should have a 1- charge, since only one proton is removed. In row 5, *zorbamycin* is a secondary, not a primary amide, thus it has an incorrect hierarchical relationship error. The distribution of ChEBI concepts and errors between "2-star" and "3-star" concepts is in Table IV. It shows a much higher error rate for "3-star" concepts. Thus, we recommend to start by auditing "3-star" concepts in small partial-areas.

TABLE IV. THE DISTRIBUTION OF "2-STAR" AND "3-STAR" CONCEPTS WITH ERRORS

	# of Concepts in ChEBI	# of Concepts Audited	Error Rate
2-star concepts	51032	159	18.24%
3-star concepts	37803	317	42.59%

V. DISCUSSION

We utilized the partial-area taxonomy of ChEBI's *Chemical Entity* hierarchy to explore the QA methodology focusing on small partial-areas. The results show that a threshold value Θ = 2 maximizes the average error rate difference between small and large partial-areas. The weighted average error rate for concepts of small partial-areas of up to 2 concepts is 41.71%. Hence, if the total of 27,262 concepts of the small partial-areas would be reviewed, about 11,371 concepts are expected to require corrections. Thus, if the QA resources are too limited to review 27,262 concepts, then 41.71% of the concepts from small partial-areas reviewed are expected to be erroneous.

Ochs et al. [17] and He et al. [18] presented a family-based QA framework such that one methodology is applicable to a whole family of structurally similar ontologies. If the same QA methodology is successful on six out of six ontologies in the same family, then it will be successful for at least half of the ontologies in the family. To be considered successful, the error rate of study concepts should be statistically significantly higher than for control concepts. Ochs et al. classified 373 BioPortal ontologies into 81 structural families, according to structural features of those ontologies for which AbNs can be derived.

In the previous study on NCIt's *Biological Process* hierarchy by Hua et al. [19], we formulated a similar hypothesis for small partial-areas. Although we reported the error rates for different partial-area sizes and the error rate of small partial-areas with sizes up to three was higher than that of large partial-areas, we did not calculate the statistical significance. According to the previously reported error rates, the two-tailed p-value of Fisher's exact test is 0.0011, based on Table V. Hence, Hypothesis 1 for the *Biological Process* hierarchy of NCIt is confirmed with statistical significance.

TABLE V. THE 2*2 CONTINGENCY TABLE FOR SMALL AND LARGE PARTIAL-AREAS FOR BIOLOGICAL PROCESS HIERARCHY (P-VALUE = 0.0011)

	# of Concepts w/ Errors	# of Concepts w/o Errors
Partial-areas with size <=3	27	195
Partial-areas with size >3	17	350

The concerns of SNOMED CT users about errors are documented in [20]. In the study of SNOMED CT's *Procedure* hierarchy by Ochs et al. [21], we obtained a similar result.

Concepts in small partial-areas with sizes up to three have more errors than large partial-areas. The two-tailed p-value of Fisher's exact test is $p < 0.019$. A study on the NCIt's *Neoplasm* subhierarchy by Zheng et al. [15] reported that concepts in small partial-areas with sizes ≤ 10 have a statistically significantly higher error rate than large partial-areas, with $p = 0.0113$.

According to Ochs et al. [17], NCIt's small *Biological Process* and *Neoplasm* subhierarchies and SNOMED CT's large *Procedure* hierarchy in the previous three studies, and ChEBI's *Chemical Entity* hierarchy in this study, belong to the same family of 76 ontologies. In summary, there have been four successful studies, out of the required six, of the QA technique showing that small partial-areas have statistically significantly more errors than large partial-areas. However, the threshold that defines small partial-areas varies. Hence, this study advances towards the goal of showing that small partial-area-based QA is applicable to the whole family. If two more such studies will be successful, then we can make a statement that the small partial-area-based methodology is applicable to this whole family as follows. For at least half of the remaining ontologies there exists a threshold Θ such that the error rate for concepts of small partial-areas is statistically significantly higher than for large partial-areas. A *substantially different* approach to QA using partial-areas is based on a refinement of the partial-area taxonomy into the **disjoint** partial-area taxonomy [22].

VI. Conclusion

Abstraction Networks of ontologies have been proven to define a framework for the identification of concept sets that are expected to have comparatively higher error rates. Small partial-areas in the partial-area taxonomy derived from an ontology likely reflect uncommonly modeled concepts in the ontology. In this paper we tested the QA methodology that concentrates on auditing the concepts in small partial-areas.

This study applied the small partial-area-based QA methodology to the ChEBI ontology. Our analysis revealed that small partial-areas have statistically significantly more errors than large partial-areas, with an optimal threshold of two. The results confirmed that in the ChEBI ontology small partial-areas with size up to two concepts harbor statistically significantly more commission errors compared to large partial-areas. Overall, this approach narrows down the places in the ChEBI ontology where limited QA efforts should be invested to obtain a higher QA yield. This study, in combination with three other previous studies, provides progress toward showing that the small partial-area-based methodology is successful in identifying likely errors for a whole family of 76 BioPortal ontologies.

Acknowledgment

Research reported in this publication was partially supported by the National Cancer Institute of the National Institutes of Health under Award Number R01CA190779. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] J. Hastings, G. Owen, A. Dekker, M. Ennis, N. Kale, V. Muthukrishnan, *et al.*, "ChEBI in 2016: Improved services and an expanding collection of metabolites," *Nucleic Acids Res*, vol. 44, pp. D1214-9, Jan 4 2016.

[2] C. Bettembourg, C. Diot, and O. Dameron, "Optimal Threshold Determination for Interpreting Semantic Similarity and Particularity: Application to the Comparison of Gene Sets and Metabolic Pathways Using GO and ChEBI," *PLoS One*, vol. 10, p. e0133579, 2015.

[3] H. Min, Y. Perl, Y. Chen, M. Halper, J. Geller, and Y. Wang, "Auditing as part of the terminology design life cycle," *J Am Med Inform Assoc*, vol. 13, pp. 676-90, Nov-Dec 2006.

[4] M. Halper, H. Gu, Y. Perl, and C. Ochs, "Abstraction networks for terminologies: Supporting management of "big knowledge"," *Artif Intell Med*, vol. 64, pp. 1-16, May 2015.

[5] Y. Wang, M. Halper, H. Min, Y. Perl, Y. Chen, and K. A. Spackman, "Structural methodologies for auditing SNOMED," *J Biomed Inform*, vol. 40, pp. 561-81, Oct 2007.

[6] S. de Coronado, M. W. Haber, N. Sioutos, M. S. Tuttle, and L. W. Wright, "NCI Thesaurus: using science-based terminology to integrate cancer research results," *Stud Health Technol Inform*, vol. 107, pp. 33-7, 2004.

[7] M. Q. Stearns, C. Price, K. A. Spackman, and A. Y. Wang, "SNOMED clinical terms: overview of the development process and project status," *Proc AMIA Symp*, pp. 662-6, 2001.

[8] H. Yumak, L. Chen, M. Halper, L. Zheng, Y. Perl, and G. Elhanan, "A Quality-Assurance Study of ChEBI," in *ICBO/BioCreative*, 2016.

[9] EMBL-EBI. *Chemical Entities of Biological Interest*. Available: https://www.ebi.ac.uk/chebi/

[10] Y. F. Gao, L. Chen, G. H. Huang, T. Zhang, K. Y. Feng, H. P. Li, *et al.*, "Prediction of drugs target groups based on ChEBI ontology," *Biomed Res Int*, vol. 2013, p. 132724, 2013.

[11] D. P. Hill, N. Adams, M. Bada, C. Batchelor, T. Z. Berardini, H. Dietze, *et al.*, "Dovetailing biology and chemistry: integrating the Gene Ontology with the ChEBI chemical ontology," *BMC Genomics*, vol. 14, p. 513, 2013.

[12] B. Glimm, I. Horrocks, B. Motik, G. Stoilos, and Z. Wang, "HermiT: An OWL 2 Reasoner," *J. Autom. Reason.*, vol. 53, pp. 245-269, 2014.

[13] M. Halper, Y. Perl, C. Ochs, and L. Zheng, "Taxonomy-Based Approaches to Quality Assurance of Ontologies," *J Healthc Eng*, vol. 2017, p. 3495723, 2017.

[14] C. Ochs, J. Geller, Y. Perl, and M. A. Musen, "A unified software framework for deriving, visualizing, and exploring abstraction networks for ontologies," *J Biomed Inform*, vol. 62, pp. 90-105, Aug 2016.

[15] L. Zheng, H. Min, Y. Chen, J. Xu, J. Geller, and Y. Perl, "Auditing National Cancer Institute thesaurus neoplasm concepts in groups of high error concentration," *Applied Ontology*, vol. 12, pp. 1-18, 2017.

[16] R. A. Fisher, "Statistical methods for research workers," in *Breakthroughs in Statistics*, ed: Springer, 1992, pp. 66-70.

[17] C. Ochs, Z. He, L. Zheng, J. Geller, Y. Perl, G. Hripcsak, *et al.*, "Utilizing a structural meta-ontology for family-based quality assurance of the BioPortal ontologies," *J Biomed Inform*, vol. 61, pp. 63-76, Jun 2016.

[18] Z. He, C. Ochs, A. Agrawal, Y. Perl, D. Zeginis, K. Tarabanis, *et al.*, "A family-based framework for supporting quality assurance of biomedical ontologies in BioPortal," *AMIA Annu Symp Proc*, vol. 2013, pp. 581-90, 2013.

[19] H. Min, Y. Perl, Y. Chen, M. Halper, J. Geller, and Y. Wang, "Auditing as Part of the Terminology Design Life Cycle," *J Am Med Inform Assoc*, vol. 13, pp. 676-690, Nov-Dec 2006.

[20] G. Elhanan, Y. Perl, and J. Geller, "A survey of SNOMED CT direct users, 2010: impressions and preferences regarding content and quality," *J Am Med Inform Assoc*, vol. 18 Suppl 1, pp. i36-44, Dec 2011.

[21] C. Ochs, Y. Perl, J. Geller, M. Halper, H. Gu, Y. Chen, *et al.*, "Scalability of abstraction-network-based quality assurance to large SNOMED hierarchies," *AMIA Annu Symp Proc*, vol. 2013, pp. 1071-80, 2013.

[22] Y. Wang, M. Halper, D. Wei, Y. Perl, and J. Geller, "Abstraction of complex concepts with a refined partial-area taxonomy of SNOMED," *J Biomed Inform*, vol. 45, pp. 15-29, Feb 2012.