HHS Public Access

Author manuscript

Artif Intell Med. Author manuscript; available in PMC 2018 June 01.

Published in final edited form as:

Artif Intell Med. 2017 June ; 79: 9–14. doi:10.1016/j.artmed.2017.05.002.

From SNOMED CT to Uberon: Transferability of Evaluation Methodology Between Similarly Structured Ontologies

Gai Elhanan[1], **Christopher Ochs**[1], **Jose L. V. Mejino Jr**[2], **Hao Liu**[1], **Christopher J. Mungall**[3], and **Yehoshua Perl**[1]

[1]Computer Science Department, New Jersey Institute of Technology, Newark, NJ, USA

[2]Department of Biological Structure (Structural Informatics Group), University of Washington, Seattle, WA, USA

[3]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA

Abstract

Objective—To examine whether disjoint partial-area taxonomy, a semantically-based evaluation methodology that has been successfully tested in SNOMED CT, will perform with similar effectiveness on Uberon, an anatomical ontology that belongs to a structurally similar family of ontologies as SNOMED CT.

Method—A disjoint partial-area taxonomy was generated for Uberon. One hundred randomly selected test concepts that overlap between partial-areas were matched to a same size control sample of non-overlapping concepts. The samples were blindly inspected for non-critical issues and presumptive errors first by a general domain expert whose results were then confirmed or rejected by a highly experienced anatomical ontology domain expert. Reported issues were subsequently reviewed by Uberon's curators.

Results—Overlapping concepts in Uberon's disjoint partial-area taxonomy exhibited a significantly higher rate of all issues. Clear-cut presumptive errors trended similarly but did not reach statistical significance. A sub-analysis of overlapping concepts with three or more relationship types indicated a much higher rate of issues.

Conclusions—Overlapping concepts from Uberon's disjoint abstraction network are quite likely (up to 28.9%) to exhibit issues. The results suggest that the methodology can transfer well between same family ontologies. Although Uberon exhibited relatively few overlapping concepts, the methodology can be combined with other semantic indicators to expand the process to other concepts within the ontology that will generate high yields of discovered issues.

Corresponding Author: Gai Elhanan, gelhanan@gmail.com, Computer Science Department, New Jersey Institute of Technology, Newark, NJ 07102, USA.

Keywords

anatomy ontology; quality assurance; semantic complexity; disjoint abstraction network; overlapping concepts

1. Introduction

Biomedical ontologies, as included in BioPortal [1], are computable representations of biomedical knowledge and, as such, are vital for effective delivery of clinical, research and administrative healthcare. Voluminous research has demonstrated that, while mostly correct, biomedical ontologies may harbor incorrect representations of knowledge and even critical errors [2–9].

Biomedical ontology curation efforts focus mostly on prospective aspects of structure and content creation, while limited resources are directed at retrospective content representation analysis. Considering the size of some of the currently dominant biomedical ontologies, the time for systematic, concept-by-concept, retrospective analysis is long past. Therefore, it is essential to maximize the effectiveness of scarce human resources by developing algorithmic approaches that enable efficient retrospective analysis of biomedical ontologies.

Abstraction networks, whether area taxonomies or their derivative partial-area taxonomies [10–12], offer an alternative view of biomedical ontologies based on their semantic structure. This alternative representation highlights elements that are structurally dissimilar and have been shown [8, 13] to possibly contain a disproportionately higher number of potential erroneous concepts, thus enabling a better view into potential problems and more efficient, higher yield retrospective reviews.

However, not all biomedical ontologies are alike. Aside from the obvious differences in content coverage, biomedical ontologies often differ in structural design. Therefore, it is not reasonable to expect that a specific structurally derived methodology will work across the vast majority of biomedical ontologies or continue to exhibit similar efficiency. Previous research [14, 15] has demonstrated that the biomedical ontologies in NCBO's BioPortal [1, 16] can be divided into 81 structural families with the expectation that structurally derived methodologies will work with similar efficiency within each such family.

Large biomedical ontologies such as SNOMED CT [17], Uberon [18, 19], GO [20], ChEBI [21] and, for practical reasons, NCIt [22] belong to the same family [15] that is based on the existence of semantic relationships (e.g., attribute relationships in SNOMED CT or restriction on classes in native OWL biomedical ontologies) and the existence of concepts with multiple parents (See family 3–20 in Table 5 in [15]). The disjoint partial-area taxonomy [23], a specialized method for taxonomy analysis, effectively highlights SNOMED CT concepts that overlap across partial-areas. Such overlapping concepts are more prone for modeling issues [6, 7]. To investigate the transferability of the methodology between members of same family of ontologies, we applied it to Uberon, a complex and relatively large, cross-species anatomical ontology.

2. Material and Methods

2.1. Source terminology

Unlike other anatomical ontologies, Uberon [18, 19] is a comparative, cross-animal species anatomical ontology. Focused on vertebrates, it represents body parts, organs and tissues and is designed to integrate with other ontologies. Currently, Uberon holds more than 14,000 classes with a rich network of relationships among the classes using constructs from the OWL2-DL [19, 24] (maximum depth of 36, 148 property types, 25,143 restrictions, 5,185 equivalent class axioms). For comparison, SNOMED CT holds more than 320,000 classes with 152 property types and a maximum depth of 28. Uberon is provided in both OBO [25] and OWL formats and can be explored through a variety of platforms, including BioPortal. For this study the OWL format of the January 2016 release of Uberon was used and Uberon was viewed (inferred version) using our locally sourced Ontology Abstraction Framework (OAF) tool [14, 26, 27] for OWL ontologies. All available information from the OWL format of Uberon was accessible.

2.2. Overlapping concepts and disjoint partial-area taxonomies

The reasoning and methodology for generation of partial-area taxonomies has been described extensively [11, 27]. For brevity, we will briefly describe the methodology behind disjoint partial-area taxonomies while a comprehensive description can be found here [7, 23].

Areas, in the abstraction networks discussed here, are created by the set of their defining relationships. Figure 1(a) depicts part of an area with a subhierarchy of 12 concepts (out of 460) from Uberon. All the concepts within the area have the same set of five relationships (listed at the top of Figure 1(a). A root of an area is a concept that has no ancestors in the same area. An area may have more than one root. In the example in Figure 1(a), **meso-epithelium**, **endo-epithelium** and, **ecto-epithelium** are the roots (marked by the bold dashed boxes around them). Each root is labeled with its total number of descendants in the area (though many are omitted from the figure). Each root and its descendants are surrounded by colored, dashed bubbles (e.g., **ecto-epithelium** and its 52 descendants are enclosed in a green dashed bubble). As some concepts may have more than one parent, they may be associated with more than one root. For example, **epithelium of vomeronasal organ** has **endo-epithelium** and **ecto-epithelium** as roots and therefore resides within the green and the blue bubbles.

Partial-areas are a collection of nodes within an area. A partial-area is a set comprising one root and all its descendants (within the area) that summarizes semantically similar concepts. Since an area may have more than one root, it may have more than one partial-area. Partial-area taxonomy is a network of partial-area nodes, connected by hierarchical relationships [10]. Figure 1(b) shows the partial-area taxonomy based on the concepts in Figure 1(a). The partial-areas are named after their root concept and labeled with the total number of their concepts. The partial-areas are shaded according to the corresponding bubble in Figure 1(a). For example, the partial-area **ecto-epithelium** (53 concepts) in Figure 1(b) summarizes the root concept **ecto-epithelium** and its 52 descendants in the green bubble. A given concept

may reside in more than one partial-area. Such concepts are considered overlapping concepts. Figure 1 shows that the three partial-areas in (b) have six overlapping concepts between them (see Figure 1(a)). Four concepts are shared between the **meso-epithelium** (red) and the **endo-epithelium** (blue) partial-areas (**prostate epithelium**, **cloacal epithelium**, **embryonic cloacal epithelium** and, **urogenital sinus epithelim**), one concept is shared between the **meso-epithelium** (red) and the **ecto-epithelium** (green) partial-areas (**brain endothelium**), and one concept is shared between the **endo-epithelium** (blue) and the **ecto-epithelium** (green) partial-areas (**epithelium of vomeronasal organ**).

Overlapping concepts suggest elaboration of semantics from multiple roots and convergence of knowledge. Such concepts merit a focused view [7]. The disjoint partial-area taxonomy advances the partial-area taxonomy by creating more uniform roots [23]. It partitions partial-areas with overlapping concepts into disjoint partial-areas whose concepts do not overlap with other disjoint partial-areas (see (c) in Figure 1). Concepts in any given disjoint partial-area have a single root. Each disjoint partial-area may summarize one or more hierarchically related concepts. Once a disjoint partial-area taxonomy is created, the three original partial-areas result in eight disjoint partial-areas that include the three original ones (albeit with fewer concepts) and five new disjoint partial-areas (See (c) in Figure 1). Two overlapping concepts (**embryonic cloacal epithelium** and **urogenital sinus epithelium**) are folded into a single disjoint partial-area of two concepts). Each concept from (a) is now summarized in (c) by exactly one disjoint partial-area. Disjoint partial-areas are named and labeled in a similar manner to regular partial-areas but their shading indicates partial-areas that are the source of the overlap. In Figure 1(c) it is easy to identify the newly created disjoint partial-areas based on their multiple shading. For example, the now disjoint partial-area **epithelium of vomeronasal organ** is shaded half blue and half green based on its overlapping role in the **endo-epithelium** and **ecto-epithelium** partial-areas. The total number of concepts in all eight disjoint partial-areas is smaller than the total in Figure 1(b), representing the elimination of overlap.

As indicated earlier, overlapping concepts may be a target-rich environment for retrospective analysis [7]. Figure 1 anecdotally supports this premise as the overlapping concept **brain endothelium** mapping to **ecto-epithelium** may require closer examination.

2.3. Methodology

Based on the OWL format version (January 2016) of Uberon and using our OAF-based algorithmic approach [26, 27], a disjoint partial-area taxonomy was generated and two sets of data were created. A test set of 100 randomly selected overlapping concepts was created first. Subsequently, a control set of 100, randomly selected, non-overlapping concepts was created, roughly matching the test set for semantic complexity based on a pre-determined cut-off point of up to two relationship types versus three or more relationship types. Both data sets were randomly merged and blindly presented for evaluation.

2.4. Evaluation of application of methodology to Uberon

A two-step evaluation was performed. First, the randomly combined data set of 200 concepts was given to a general medicine and ontologies domain expert (GE) for initial

comprehensive analysis. Utilizing an inferred version of Uberon, the domain expert reviewed each concept from the data set using the OAF tool. At their discretion, the reviewers could also reference external sources. Each reviewed concept was categorized according the following three categories: No issues, non-critical issues, and presumptive error. A concept may have more than one noncritical issue and/or presumptive error. The reasoning for each issue or presumptive error was documented by the domain expert.

To clarify, non-critical issues include more subjective, “nice to have”, observations that were not deemed significant enough to be presumptive errors. For example, the concept **lower jaw molar epithelium** has **dental epithelium** as a parent which, by itself, is correct. However this is considered redundant since **lower jaw molar epithelium** has, amongst others, **molar epithelium** as a parent which, in turn, has **dental epithelium** as a parent. Thus, **dental epithelium** is considered a redundant parent, which is a non-critical issue. Another example for a non-critical issue is refinement of parents or relationship targets. The concept **muscularis mucosa** has a *part_of* relationship with **digestive tract** and **mucosa**. These targets can be replaced and refined with **gastrointestinal system mucosa**. Presumptive errors are clearer problems that, with varying degree of urgency, should be addressed. For example, **minor vestibular gland** has **paraurethral gland** as a parent. However, the term **paraurethral gland** is rather a synonym than a parent. Another example for a presumptive error is **postganglionic sympathetic fiber** which has the relationship *extends_fibers_into* **sympathetic ganglion** but by virtue of it is being post ganglionic it is rather from and not into.

In the second round of the evaluation, the results from the first round were reviewed by a highly experienced anatomical ontology domain expert (JLVM). Discovered issues (including presumptive errors) from the first round were submitted to be confirmed or rejected in the second round. No resolution discussions were conducted between experts. Both experts were blinded as to the methodology source of the concepts.

3. Results

The abstraction network generated for Uberon (January 2016) identified 5,437 partial-areas. Sixty of these partial areas contained concepts that overlapped (1.1% of partial-areas). Overall, the partial-areas with overlapping concepts included 1,162 Uberon concepts (8.1% of Uberon) of which 302 were overlapping concepts. The 60 partial-areas with overlapping concepts were converted into 204 disjoint partial-areas. As can be seen from Table 1, 45.7% of the overlapping concepts reside in areas with three or more relationships (138 concepts). Due to our randomized study sample selection and control sample matching methodology, the distribution of semantically complex (level 3 or above, 45 concepts) and less semantically complex concepts (relationship levels 1 and 2, 55 concepts) between both control (non-overlapping) and the test (overlapping) samples was similar (see Table 3).

The results of the first and second rounds of evaluation are displayed in Table 2. Overall, between the two samples (200 concepts), 39 and 34 concepts were found to harbor issues in the first and the second rounds respectively, including 25 and 20 concepts (respectively) with presumptive errors. In both rounds of evaluation, concepts in the overlapping study sample

exhibited statistically significant (Fisher's exact test, 2-tail) higher rate of discovered issues (noncritical and presumptive errors) than in control. Overall, the second round approved 70.6 percent and 100 percent of the first round reported issues (overlapping and non-overlapping respectively, without statistical significance). The anatomy domain expert (LVM) agreed with somewhat fewer findings, especially presumptive errors in overlapping concepts.

Table 3 illustrates the distribution of all discovered issues (both first and second rounds) based on the semantic complexity of the concepts and the sample type (control vs. test). As can be seen in Table 3, there is no demonstrable significance for overlapping concepts for levels of complexity below Level 3. However, higher levels (>=3) overlapping concepts were much more likely to exhibit issues than lower semantic level overlapping concepts.

All discovered issues (first round) were reported to Uberon's editorial team. Of all reported issues, 96 percent were commented by Uberon's curators. Eleven non-critical reported issues (all related to redundant parents) were identified as non-relevant, resulting from display issues, and were removed from the results. Of the rest, Uberon's curators concurred (i.e. identified as an already known issue or newly discovered issue) with 70.2% of all issues and 79.2% of all presumptive errors. We do not consider the Uberon curator's review as a third round of our evaluation but rather as the reference standard against which our results should be compared. Accordingly, this represents a positive predictive value (PPV) per concept of 89.2%. Due to the nature of our evaluation and the curators' review specificity, sensitivity and negative predictive value (NPV) cannot be measured.

4. Discussion

While desirable, retrospective analysis of existing ontological content is usually not a high priority for ontology curators. Limited resources usually mandate focus on creation and expansion. While there are no reports, to date, that ontological errors have ever harmed patients, Rector et. al. [2] eloquently described how deficiencies in SNOMED CT's content prevented them from using it in practical applications. As possible recognition of such issues, large-scale medical ontologies, such as SNOMED CT, are generally not in direct use in clinical applications. End-user exposure to such ontologies is usually through limited, third-party, well-curated lists and applications. Such degree of separation alleviates the immediate need for retrospective analysis of the original content. However, using ontological content in such a manner, while simplifying current use, may cause long term issues in respect to re-use and data integration. To enable direct integration of biomedical ontologies, for their valued hierarchical and semantic meta-data, it is essential to quantify their deficiencies and, if not marginal, to establish efficient methods for correction. This is not a trivial task since prominent biomedical terminologies, such as SNOMED CT and even Uberon, are very large and complex structures, making systematic concept-by-concept review impractical.

Our research was conducted in the latter context. We strive to develop and evaluate methodologies that will make evaluation of existing and new content of biomedical ontologies more efficient by identifying parameters that indicate concepts with higher

probability for harboring problems that merit review, thus enabling more efficient and focused review. Overlapping concepts and disjoint partial-areas are one such approach.

However, non-transferable methodologies that work for just one biomedical ontology are not as desirable as methodologies that can work across such ontologies. The validity of disjoint partial-areas as hotspots for findings in biomedical ontologies has been demonstrated for various hierarchies of SNOMED CT [6, 7]. SNOMED belongs to a larger (>60) family of biomedical ontologies [14, 15] based on their structural characteristics (existence of object property restrictions and the existence of concepts with multiple parents). Thus, demonstrating that disjoint partial-area-taxonomies are effective on other biomedical ontologies with similar characteristics (i.e., same family of ontologies) is a step towards demonstrating a broader transferability. Previously, we calculated that 6/6 or 8/9 [15] successful demonstrations of the methodology will be required to make the assumption that the methodology will work for the majority of biomedical ontologies within the family. While not all the biomedical ontologies within this specific family are of the same scale (i.e. SNOMED vs. most other in terms of size and expressivity), more than a handful are rich enough for the methodology to function properly. For example, the ChEBI is roughly 40 percent the size of SNOMED CT with only 9 properties compared to 152 in SNOMED, yet the partial-area taxonomy operates effectively on it.

The current study indicates that, in Uberon, 24% of overlapping concepts exhibit issues regarding their modeling and content as opposed to 11% of non-overlapping concepts (see Table 2 and Table 4). Even though significantly more issues were detected in overlapping concepts, we were not able to demonstrate statistical significance in regards to presumptive errors. However, the distinction between non-critical issues and presumptive errors was done for practical reasons and, in fact, Uberon's curators, our "reference standard", concurred with the majority of both. This is the first successful demonstration of the transferability of the methodology between biomedical ontologies within the same family. At this time, we have also successfully tested the methodology on the gene and neoplasm subhierarchies of NCIt (to be published). Altogether, these are major steps towards validation of the methodology within a family of similarly structured biomedical ontologies.

A general search for Uberon on PubMed returns only seven search results and a Google Scholar title search returns five results, none of which deal with retrospective review. However, Uberon encourages its user community to submit identified issues to the Uberon Issue Tracker [29] on GitHub. Currently the site lists more than 400 open issues and is an indication of an active user community with an ongoing interest and effort to address deficiencies. In that context, the results of this study represent a first general attempt to assess Uberon's content and quantify possible problems. Based on the second round results, we identified 11% of our control sample concepts (non-overlapping) and 24% of the test sample concepts (Table 2) as concepts with issues (8% and 12% respectively for presumptive errors). Since the vast majority of Uberon's concepts are non-overlapping we can only speculate that the general rate of issues and presumptive errors is only somewhat higher than our findings for the control sample.

For automated medical transcription services an error rate of about three percent is considered acceptable. However, there is no such recognized threshold for biomedical ontologies even though such errors may hinder the practical use of biomedical ontologies [2]. Assuming that presumptive errors represent true errors, the presumptive error rate of 8% found in the second round of this study is rather low for a relatively large anatomical ontology such as Uberon. For comparison, our evaluation of ChEBI indicated nearly 40% of concepts with issues [to be published] and 15.5% of concepts with severe errors of commission [28].

Uberon's content quality is also important since Uberon concepts play a role in other biomedical ontologies. In a recent study [to be published] we found that 15 ontologies hosted in the NCBO BioPortal import portions of Uberon. For example, the Cell Ontology imports 910 of Uberon's classes and the Experimental Factor Ontology (EFO) imports 528. Five of these ontologies also utilize Uberon classes to model their own ontology's content. For example, 42 Uberon classes are used in the modeling of EFO classes (e.g., as a superclass or in a class restriction). The reuse of Uberon's content by other ontologies implies that errors in Uberon can propagate to other biomedical ontologies. Thus, efforts that improve Uberon's content will propagate such improvements to dependent biomedical ontologies.

Some of the presumptive errors we found were potential indicators for recurring problems within Uberon as they repeated within samples. For example, the concept **anterior spinocerebellar tract** has a *part_of* relationship with the **medulla oblongata**. However, only a segment of the anterior spinocerebellar tract courses through the medulla. Without explicitly specifying whether the term represents only a portion of the tract or the entirety of it (as it currently implies), the *part_of* relationship is too non-specific and misleading. Similar presumptive errors related to ambiguity regarding the entirety or portion of anatomical elements recurred. In another example, **decussation of trochlear nerve** demonstrated yet another recurring theme with our presumptive errors. The concept has the relationship *extends_fibers_into* **trochlear nerve**. However the decussation, by definition, is just an indication of crossed tracts and does not extend fibers. We found other problems with the use of *extends_fibers_into*. Of the two examples reported above the first is stated whereas the second is inferred. We did not follow reported issues to their roots as part of the scope of this study.

Initially we also reported several recurring non-critical issues related to redundant parents. For example, **lumbar spinal cord white matter** appears to have two parents, **white matter of spinal cord** and **white matter**. However, **white matter** is the parent of **white matter of spinal cord**, thus redundant in the context of **lumbar spinal cord white matter**. Upon review of the feedback from Uberon's curators it was revealed that such redundant parents resulted from display issues and we omitted them from our findings. However, it is important to note that similar display issues exist not only in our OAF tool, but also when viewing Uberon in environments such as BioPortal and Protégé [30] (and several other tools) that include super-classes from equivalence axioms. Other tools like Ontobee [31] do not include these and only show the parents from superclass axioms.

Semantic complexity in general, as represented by concepts with three or more types of relationships, was also shown to be an overall indicator for potential problems (Table 3). Semantic complexity was not an indicator for potential problems within the control or the test samples (Table 4). However, these secondary analyses demonstrate that the combination of measurements of semantic complexity and overlapping concepts is more effective at identifying concepts with issues than either alone (as demonstrated by the dark gray box in the bottom right corner of Table 4).

Taken altogether, overlapping concepts, semantic complexity and recurring issues present an attractive methodology for an effective, focused retrospective review with high yields. Even though the number of partial-areas with overlapping concepts and the number of overlapping concepts were relatively small, an initial focus on the intersection of semantically complex overlapping concepts may yield numerous findings. Next, recurring problems will be identified and semantically complex concepts with these parameters will be identified across the ontology and revisited. Subsequent steps may depend on the overlapping/non-overlapping ratio with a threshold to be determined. A high ratio may suggest that recurring problems should be visited in the remaining non-overlapping concepts with these indicators. With a low ratio it might be preferable to visit first the less semantically complex overlapping concepts. If needed, this process can be parceled for specific sub-domains of an ontology and replace systematic, concept-by-concept exhaustive review.

Arguably, the evaluation could have been more expansive. However, our research reflects the reality of everyday life where domain experts and their time are a limited resource, especially for such tasks. Review of Table 2 indicates an inter-round approval of 70.6% or higher (up to 100%) by the second round reviewer. Considering that the second reviewer is a highly experienced domain-specific expert, we believe that the two rounds of evaluation offer sufficient robustness. This is supported by the high concurrence rate (up to 79.2%) after review of reported issues by Uberon's curators. The most likely deficiency of our evaluation process may be the possibility of missed additional issues as the second round reviewer was not asked to conduct comprehensive review but rather confirm first round findings. Similarly, the curators' review cannot be used as a measure true positives or true negatives and therefore, sensitivity, specificity and NPV cannot be calculated. Nevertheless, the curators' review represents a PPV of 89.2% per concept.

5. Conclusion

The study validates that overlapping concepts, exposed through disjoint partial-area taxonomies, constitute a subset that is more likely to display modeling and content problems. Our success with Uberon demonstrates that the methodology is indeed transferrable between biomedical ontologies that share structural features. Combining this and other methodologies, such as semantic complexity, can offer an effective and efficient alternative to resource intensive systematic retrospective review of ontological content.

Acknowledgments

Research reported in this publication was 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 views of the National Institutes of Health.

References

1. BioPortal. http://bioportal.bioontology.org/
2. Rector AL, Brandt S, Schneider T. Getting the foot out of the pelvis: modeling problems affecting use of SNOMED CT hierarchies in practical applications. J Am Med Inform Assoc. 2011 Jul-Aug; 18(4):432–40. [PubMed: 21515545]
3. Ceusters W, Smith B, Kumar A, et al. Mistakes in medical ontologies: where do they come from and how can they be detected? Stud Health Technol Inform. 2004; 102:145–64. [PubMed: 15853269]
4. Zhu X, Fan JW, et al. A review of auditing methods applied to the content of controlled biomedical terminologies. J Biomed Inform. 2009 Jun; 42(3):413–25. [PubMed: 19285571]
5. Rector A, Iannone L. Lexically suggest, logically define: quality assurance of the use of qualifiers and expected results of post-coordination in SNOMED CT. J Biomed Inform. 2012 Apr; 45(2):199–209. [PubMed: 22024315]
6. Ochs C, Geller J, et al. Scalable quality assurance for large SNOMED CT hierarchies using subject-based subtaxonomies. J Am Med Inform Assoc. 2015 May; 22(3):507–18. [PubMed: 25336594]
7. Wang Y, Halper M, et al. Auditing complex concepts of SNOMED using a refined hierarchical abstraction network. J Biomed Inform. 2012 Feb; 45(1):1–14. [PubMed: 21907827]
8. Min H, Cohen B, et al. Detecting role errors in the gene hierarchy of the NCI Thesaurus. Cancer Inform. 2008; 6:293–313. [PubMed: 19221606]
9. Ochs C, Perl Y, et al. Quality assurance of the gene ontology using abstraction networks. J Bioinform Comput Biol. 2016 Jun.14(3):1642001. [PubMed: 27301779]
10. Wang Y, Halper M, et al. Structural methodologies for auditing SNOMED. J Biomed Inform. 2007 Oct; 40(5):561–81. [PubMed: 17276736]
11. Halper M, Gu H, et al. Abstraction networks for terminologies: Supporting management of “big knowledge”. Artif Intell Med. 2015 May; 64(1):1–16. [PubMed: 25890687]
12. Geller J, Ochs C, et al. New abstraction networks and a new visualization tool in support of auditing the SNOMED CT content. AMIA Annu Symp Proc. 2012; 2012:237–46. [PubMed: 23304293]
13. Halper M, Wang Y, et al. Analysis of error concentrations in SNOMED. AMIA Annu Symp Proc. 2007 Oct.11:314–8.
14. He Z, Ochs C, et al. A family-based framework for supporting quality assurance of biomedical ontologies in BioPortal. AMIA Annu Symp Proc. 2013 Nov 16.:581–90. [PubMed: 24551360]
15. Ochs C, He Z, et al. Utilizing a structural meta-ontology for family-based quality assurance of the BioPortal ontologies. J Biomed Inform. 2016 Jun.61:63–76. [PubMed: 26988001]
16. Noy NF, Shah NH, et al. BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Res. 2009 Jul; 37(Web Server issue):W170–3. [PubMed: 19483092]
17. SNOMED CT. http://www.ihtsdo.org/snomed-ct
18. Uberon. http://uberon.github.io/
19. Mungall CJ, Torniai C, et al. Uberon, an integrative multi-species anatomy ontology. Genome Biol. 2012 Jan 31.13(1):R5. [PubMed: 22293552]
20. Gene Ontology Consurtium. http://geneontology.org/
21. ChEBI. http://www.ebi.ac.uk/chebi/
22. NCIt. https://ncit.nci.nih.gov/ncitbrowser/pages/home.jsf?version=16.07d
23. Wang Y, Halper M, et al. Abstraction of complex concepts with a refined partial-area taxonomy of SNOMED. J Biomed Inform. 2012 Feb; 45(1):15–29. [PubMed: 21878396]
24. OWL 2 Web Ontology Language. https://www.w3.org/TR/owl2-overview/
25. The OBO Foundry. http://www.obofoundry.org/

26. Ochs C, Perl Y, Geller J. BLUSNO: A System for Orientation, Visualization, and Quality Assurance of SNOMED CT Using Abstraction Networks. ICBO. 2013:128–129.
27. Ochs C, Geller J, et al. A unified software framework for deriving, visualizing, and exploring abstraction networks for ontologies. J Biomed Inform. 2016 Aug.62:90–105. [PubMed: 27345947]
28. Yumak, H., Chen, L., et al. A Quality-Assurance Study of ChEBI. ICBO; 2016. http://icbo.cgrb.oregonstate.edu/node/320
29. Uberon Issue Tracker. https://github.com/obophenotype/uberon/issues
30. Protégé. http://protege.stanford.edu/
31. Ontobee. http://www.ontobee.org/

Highlights

- A generalization of a disjoint partial-area taxonomy methodology between similarly structured ontologies
- A two-step blinded evaluation demonstrates that the methodology effectively identifies areas in Uberon, an anatomical ontology, that produce higher yield for retrospective review
- First demonstration of the applicability of the methodology across ontologies with similar characteristics
- Overall, Uberon exhibits a relatively low rate of findings but disjoint partial-areas along with semantic complexity indicators can be used to reduce overall resource demand when conducting retrospective reviews of content

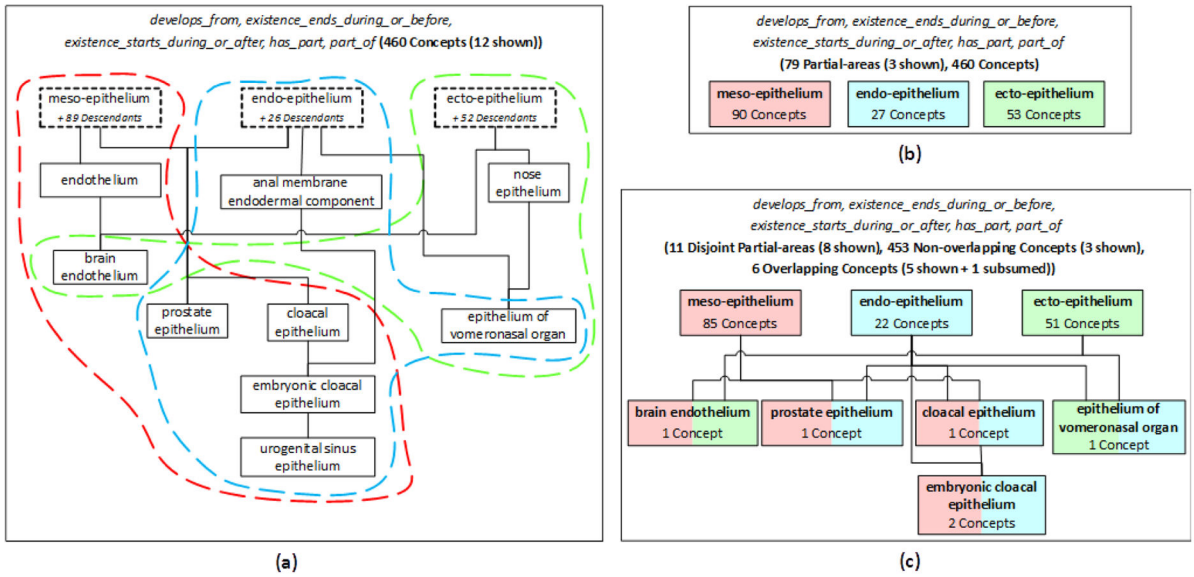

Figure 1.

(a) A subhierarchy of 12 concepts from Uberon that are modeled with *develops_from, existence_ends_during_or_before, existence_starts_during_or_after, has_part*, and *part_of* restrictions (i.e., they all belong to the same area, as indicated at the top of **(a)**). Concepts are represented as labeled boxes and parents are identified using upward direct arrows. **meso-epithelium**, **endo-epithelium**, and **ecto-epithelium** are the roots of this area, as indicated by a bold dashed outline. Each root is labeled with the total number of descendants in the area (though many are omitted). Each root and its descendant are surrounded by colored, dashed bubbles (e.g., **meso-epithelium** and all its descendants are enclosed in a red dashed bubble). Note that several concepts belong to multiple subhierarchies (for example, **epithelium of vomeronasal organ**). **(b)** The three partial-areas derived from the subhierarchy in (a). Partial-areas are shown as colored boxes (following the colored bubbles from (a)) and are labeled with the name of their root concept and the total number of concepts summarized. **(c)** The eight disjoint partial-areas derived from (b). Each disjoint partial-area is shown as a box color-coded according to the partial-area(s) its concepts are summarized by and each is labeled with its overlapping root. Each concept from (a) is summarized by exactly one disjoint partial-area.

Table 1

General information regarding partial-areas with overlapping concepts (Uberon, January 2016).

	# partial-areas	# Concepts	# Overlapping Concepts	% Total Overlapping
# of Relationships				
1	126	529	164	54.3
3 & 4	440	1018	25	8.3
>4 (max. 8)	655	1805	113	37.4
All	1221	3352	302	

Table 2

Results of first and second round of evaluation for control (Non-overlapping) and test (Overlapping) samples. Statistical significance calculated using Fisher's Exact test, 2-tail (NS - No Significance).

		First Round		Second Round		
	Concepts with:	**Issues**	**No Issues**	**Issues**	**No Issues**	**% Approval Second**
All Findings	**Non-overlapping**	12	88	11	89	91.7
	Overlapping	27	73	24	77	82.8
	P-value	0.012		0.025		
Presumptive Errors Only	**Non-Overlapping**	8	92	8	92	100
	Overlapping	17	83	12	88	70.6
	P-value	NS		NS		

Table 3

Distribution of findings according to semantic complexity (<3, >=3 relationship types per concept) between control (Non-overlapping) and test (Overlapping) samples. Shaded cells at the lower part of the table indicate the groups that were compared to each other. Statistical significance calculated using Fisher's Exact test, 2-tail (NS - No Significance).

All Findings	Level #	<3				>=3				
	Sample	Non-overlapping		Overlapping		Non-overlapping		Overlapping		P-value
	# Concepts	55		55		45		45		
	Round	1st	2nd	1st	2nd	1st	2nd	1st	2nd	
	# Concepts with issues	9	8	15	11	3	3	13	13	
										0.025
										NS
										NS
										0.006
										0.006

Table 4

Matrix calculation of yields (% concepts with findings, second round) - Non-overlapping/Overlapping vs. semantic complexity (relationship types <3/>=3). Statistical significance calculated using Fisher's Exact test, 2-tail (NS - No Significance).

Sample / Level			Non-Overlapping	Overlapping	P-value
	# Concepts Non-Ov / Ov		100	100	
		%Findings 2nd Round	11%	24%	0.013
<3	55 / 55	17.3% (19/110)	14.5% (8/55)	20% (11/55)	NS
>=3	45 / 45	17.7% (16/90)	16.7% (3/45)	**28.9% (13/45)**	0.01
P-value		NS	NS	NS	