Noname manuscript No.
(will be inserted by the editor)

SWEET Ontology Coverage for Earth System Sciences

Nicholas DiGiuseppe
University of California, Irvine
Irvine, California 92617-3440
nicholas.digiuseppe@uci.edu
· Line C. Pouchard
Oak Ridge National Laboratory
1 Bethel Valley Road
Oak Ridge, TN 37831
pouchardlc@ornl.gov
· Natalya F. Noy
Stanford Center for Biomedical Informatics Research
Stanford University
Stanford, CA, U.S.A. 94070
noy@stanford.edu

the date of receipt and acceptance should be inserted later

Abstract Scientists in the Earth and Environmental Sciences (EES) domain increasingly use ontologies to analyze and integrate their data. For example, the NASA's SWEET ontologies (Semantic Web for Earth and Environmental Terminology) have become the *de facto* standard ontologies to represent the EES domain formally [22]. Now we must develop principled ways both to evaluate existing ontologies and to ascertain their quality in a quantitative manner. Existing literature describes many potential quality metrics for ontologies. Among these metrics is the coverage metric, which approximates the relevancy of an ontology to a corpus [28]. This paper has three primary contributions to the EES domain: (1) we present an investigation of the applicability of existing coverage techniques for the EES domain; (2) we present a novel expansion of existing techniques that uses thesauri to generate equivalence and subclass axioms automatically; and (3) we present an experiment to establish an upper-bound coverage expectation for the SWEET ontologies against real-world EES corpora from DataONE [19], and a corpus designed from research articles to specifically match the topics covered by the SWEET ontologies. This initial evaluation suggests that the SWEET ontology can accurately represent real corpora within the EES domain.

1 The SWEET Ontologies and Domain Relevancy

Scientists in the Earth and Environmental Sciences (EES) domain increasingly use ontologies to analyze and integrate their data. Ontologies define the terms in

Address(es) of author(s) should be given

a domain of discourse (shared metadata terms), provide constraints on the values and define formal semantics that enable automated reasoning. Furthermore, the World-Wide Web Consortium (W3C) has defined OWL, a formal language for representing and sharing ontologies on the Web, enabling scientists to publish and integrate metadata using standard Web protocols. One can think of an ontology as a taxonomy of terms with added rules and relationships that can be used by computer algorithms. Ontologies and semantic descriptions of the scientific data and processes provide the necessary objects supporting the production of new knowledge by allowing interoperability of the processes, shared annotations and integration of the data.

Concurrent with the increasing capacity of ontologies to approximate meaningfully the semantic understanding of domain knowledge, there is an increase of ontology use within the domain of Earth and Environmental Sciences (EES) [11]. Indeed, there is work that supports different activities with ontologies ranging from ontology-expanded search and discovery [21,23] to matching tools and datasets using ontology rules [16].

Unfortunately, there is no single accepted standard for quantitatively measuring the quality of a given ontology, or sufficient experimental evidence to support a single ontology (or set of ontologies) as accurately representing the EES domain [9, 28]. Yet what is fairly common within the EES community, is the use of the SWEET ontology set (arguably the paramount EES ontology set) notwithstanding a lack of formal studies that examine or evaluate the quality of it.

Researchers have proposed a host of potential metrics to ascertain the quality of an ontology including *coverage* metrics [3, 4, 6, 8, 17, 18, 26, 28]. Coverage metrics approximate the relevancy of an ontology to a given corpus by determining the number of ideas and relationships that are *covered* by the ontology (i.e., they exist within the ontology). However, there is a dearth of studies that examine existing coverage metrics within the EES domain using EES ontologies. Thus, it remains unclear whether existing coverage techniques translate well when used within the EES domain.

This work presents a study on the applicability of existing coverage techniques for use within the EES domain as well as experiments to evaluate empirically the relevancy of the SWEET ontologies to the EES domain.

One insight from the literature is that while ontologies represent structured descriptions of the domain knowledge, ontology definitions contain a variety of natural-language terms (e.g., class names). Thus, the ability to calculate their coverage score using existing Natural-Language-Processing (NLP) methods can provide insights on how close an ontology corresponds to a particular text corpus.

Yao and colleagues have developed a methodology for assessing ontology coverage with respect to a corpus in a biomedical domain [28]. Their methodology leverages NLP techniques such as stemming, stop-word removal, normalization and part-of-speech tagging in order to calculate coverage for all classes and equivalence axioms within an ontology. Thus, our first contribution is the application of this methodology to a very different domain and a set of ontologies: the domain is the Earth and Environmental sciences and the ontologies are the SWEET ontologies.

Yao and colleagues calculated coverage only for the terms and did not calculate coverage for subclass axioms. In order to evaluate coverage more precisely, we expand Yao's methodology to enable subclass-axiom coverage using our novel approach, Synonym Synergy.

Synonym Synergy is a technique that leverages thesauri to approximate subclass relationships using synonyms relationships. It approximates these relationships by determining which word combinations are only partial synonyms. For example, if a term A is a synonym of term B, but term B is not a synonym of term A, then we call this a partial synonym (as the synonym relationship is not bidirectional). We apply this expanded methodology to the EES domain to measure its applicability.

In addition to this generalization of existing techniques, we perform an experiment to calculate a upper-bound coverage for the SWEET ontology set. Our experiment utilizes two corpora: The first corpus, from DataONE [19], is a real-world corpus taken from a variety of stakeholders within the EES domain and includes more than 46,000 documents. The second corpus is a semi-randomly selected corpus (we refer to this corpus as the *tailored* corpus) using existing scientific articles that topically are synonymous with the SWEET ontologies. The tailored corpus represents a possible upper bound using the expanded methodology, because the documents chosen specifically deal with the topics within SWEET. Thus, by comparing the SWEET ontologies' coverage score from the DataONE corpus against the tailored corpus, we can empirically evaluate the relevancy of the SWEET ontologies for real-world EES corpora.

Our results suggest that the expanded methodology is applicable and that the SWEET ontologies accurately represent the EES domain.

This work makes the following contributions:

1. We present an empirical investigation of the applicability of existing ontology-coverage techniques from the biomedical field for use within the EES domain. Our study investigates the assumptions and differences between these fields' uses of ontologies.
2. We present a novel approach to generate automatically subclass axioms from an existing corpus using Synonym Synergy. Synonym Synergy utilizes multiple thesauri to determine whether the synonym relationship between word pairs is bidirectional or one-sided. This approach generalizes existing techniques to allow for increased applicability within the EES domain.
3. We present an experiment to evaluate the quantitative relevancy of the SWEET ontology set to the EES domain using our expanded coverage methodology. We leverage a real-world corpus containing over 46,000 documents, and a tailor-made corpus using semi-randomly selected research documents that are topically synonymous with the SWEET ontologies. This experiment establishes a likely *best-case* or *baseline* coverage that can be expected from the SWEET ontologies.

The remainder of the paper is outlined as follows: first we discuss the related work for this field, then we present motivations for this work, next we discuss the coverage methodology used in this work, then we present our experiment and results, next we discuss the implications of these results, and finally we present a conclusion and the future directions for this work.

2 Related Work

The increasing relevance of ontologies over the past years, researchers developed many different approaches for measuring and evaluating the quality of ontologies.

In 2005 for example, Brank and colleagues identified four different types of techniques for ontology evaluation which have been elaborated by other researchers [2]: (i) defining and comparing an ontology against a previously defined "gold standard" by using some measures of semantic similarity; (ii) evaluation of the ontology through an application based approach by defining the fitness of a given ontology to satisfy a given task [9]; (iii) extracting evaluation information from related data to evaluate the similarity with a related text corpus [28]; and (iv) manual evaluation, which typically involves human subjects comparing and measuring ontologies against a predefined set of requirements or measures.

In this paper, we focus on (iii), extracting evaluation information from related data to evaluate the similarity with a related text corpus.

Previous research has explored the application of ontologies in earth sciences and the evaluation of ontologies in this context. For instance, Wiegand and Garcia [27] proposed a task-based method of evaluating geospatial ontologies. They developed a task ontology, to accompany an ontology of geospatial terms, and used the explicit description of tasks to evaluate the appropriateness of the ontology. Tripathi and Babaie [25] evaluated how amenable the SWEET ontologies are to extending them to other domains, specifically the hydrogeology domain. They demonstrated that SWEET ontologies can be extended successfully while still maintaining their modular structure and the existing links. These methods thus addressed a complementary aspect of ontology evaluation to the one we are addressing.

3 Need for Coverage

To help clarify the requirements and limitations of existing evaluation metrics based on coverage we present three use cases to scope our problem space, which are representative of a larger domain.

Use Case 1: Scientist Susie has multiple ontologies and needs to know which best matches her domain.
Use Case 2: Scientist Robert has multiple corpora and needs to know which is best matched by his lab's ontology.
Use Case 3: Scientist Jane is trying to optimize her ontology and needs to know the importance of its entities, such as classes and axioms (e.g., which terms are used frequently and which terms are used seldom, if ever).

Although there are a variety of ontology-evaluation metrics, when considering these representative use-cases, only a few are applicable: manual investigation [9], use of a gold standard [17], use of a fitness function [9], and coverage-based techniques [28]. Each of these metrics requires a different input and provides a different type of evaluation. Thus, we briefly describe the contributions and limitations of each technique.

Manual investigation approaches use experts to examine the ontology by hand in order to provide a score that represents the quality of ontologies or expected success for a given task. This approach has multiple high cost requirements including: (1)

a group of humans with a sufficient skill set to determine the ontologies quality, (2) sufficient time to investigate manually each idea and relationship represented within the ontology, (3) sufficient expertise to overcome the cognitive complexity of ontology representations [7], and (4) a process for settling disagreements between human scores. Additionally, this approach must be repeated whenever there are changes to the task, the ontology, or the domain. These concerns make using a manual investigation problematic. However, if skilled humans with sufficient time can agree upon a score for a given ontology for a specific task, this approach likely represents the most meaningful score of an ontology's quality and relevance. For our use cases, this type of approach is less than ideal because each scientist would have to gather a group of skilled people with sufficient time to understand their ontologies and their corpora and then have them agree upon a score for each task; it is not very efficient or practical.

Gold standard evaluations provide a quality score that describes the degree to which the ontology matches the classes or relationships that exist within the gold standard. While this approach sounds initially appealing, a preliminary difficulty lies in generating a quality gold standard, which is both time consuming and mostly manual resulting in a somewhat error prone process. A second difficulty is that the gold standard is only applicable to the set of circumstances for which it is constructed. Indeed if the domain evolves, or if an ontology describes a different set of topics than those built into the gold standard, this evaluation method provides less than ideal results. However, if a gold standard exists, and the ontology and standard describe the same domain, this evaluation can provide a meaningful approximation of the relevancy of the ontology. For our use cases, this type of approach is problematic because the answers to the scientists' questions are required to generate the gold standard, which would then be used to evaluate the answers. In other words, the same information that would be used to generate the gold standard would be used to deduce the answer to each individual question, such that the work involved in creating a gold standard would need to be repeated for each specific use case; a very time consuming process.

Fitness function evaluations typically provide a quality score that approximate the usefulness of an ontology for a given task. This approach requires that a heuristic be developed and tested to ensure that it accurately approximates a score for a given task. The development and testing of this heuristic requires multiple input ontologies so as to create a varied input set (i.e., to ensure the heuristic is a functional approximation requires multiple inputs to test and refine its properties), someone of sufficient expertise in the given domain to develop a heuristic for each use case, and the time to do this. While the time to create this function and test it is likely less than the time to create a gold standard (and much less than a manual approach), it is not trivial; and neither is the domain expertise to create it. For our use cases this approach is problematic because our scientists might not have more ontologies to be able to test a fitness function (if they must develop one), and would have to develop a different function for each use case, which is likely quite time consuming.

Coverage approach evaluations generate a quality score that represents how well a given dataset is "covered" (is relevant to) an ontology. This approach requires a dataset that reasonably approximates a given domain or problem along with an ontology. Ensuring that a dataset represents a domain can be somewhat problematic, as some domains are large, or are constantly evolving. However, if a dataset

Table 1: Comparison between biomedical ontologies used by Yao and colleagues [28], and the SWEET ontologies.

	SNOMED-CT	ICD9-CM	MeSH	SWEET (as a whole)	SWEET (per ontology)
Num of Classes	395,036	22,400	229,698	4,527	20.86
Num of Properties	41	10	32	358	1.64
Max Depth	32	6	0	10	0.04
Max Siblings	20,010	21	0	51	0.23
Avg Siblings	1	1	0	19	0.08

exists or can be found, this approach is fast, requires no manual building of a "correct answer" (as opposed to the gold standard and fitness function methods) and is entirely automatic (as opposed to the manual approach). For our use cases, this approach is the most reasonable as the scientists have corpora (their domain relevant dataset), and while these might need to be extended to ensure sufficient domain representativeness, this process is likely the fastest and requires no additional building of tools or techniques. Further, unlike the other approaches which repeat the time-consuming aspect of their approach for each use case (e.g., are not generalizable across different use cases) this approach works for each use case with minimal additional cost.

4 Applicability of Existing Techniques

Due to their specific scope, existing coverage-techniques may be ineffective when used outside of their expected domain. One cause of this limitation is that ontologies from different domains contain different underlying assumptions (e.g., large, multi-topic ontologies versus small, single-topic ontologies), and understanding the differences between domains can enable greater success in the generalization of techniques across domains. Given the potential of coverage techniques to address problem domain represented by our use cases for the EES domain, we now consider the applicability of existing techniques.

This work focuses on techniques from the biomedical domain described by Yao and colleagues [28]. The technique presented in this work utilizes a thesaurus to evaluate the coverage of classes and equivalence relationships for a given ontology. In other words, this method evaluates only coverage for ideas, and a single relationship.

While there a large diversity of ontologies within this domain that have different features and emphasize different relationships, we consider those used within the experiments described by Yao: International Classification of Diseases, Clinical Modification (ICD9-CM), Systematized Nomenclature of Medicine, Clinical Terms (SNOMED-CT) and Medical Subject Headings 2009 (MeSH).

For simplicity, we present Table 1 that depicts a juxtaposition between the different types of relationships and components used within the biomedical ontologies and the SWEET ontologies (the paramount ontologies within the EES domain). This table clearly demonstrates that these types of ontologies are largely different, particularly in their size. While both ontology sets contain the same components including classes, properties, siblings, and subclasses; the quantity within each is quite different. Consider that any given SWEET ontology is likely to have around twenty classes, whereas the smallest of our biomedical ontologies contains over 20,000—a difference of three orders of magnitude. SWEET ontologies as a whole contain a total of 4.5K classes, it is still two orders of magnitude smaller than most of these biomedical ontologies.

Initially these differences in size might appear to be trivial because both ontologies are organized according to the same properties. The methodology for generating coverage performs a type precision analysis, meaning a larger ontology-under-test requires a larger corpus to cover every topic and relationship represented in the ontology. In other words, the interpretation of results and necessary size of the corpus to ensure a *representative* topic-set for the domain are affected largely by the size and scope of the ontology under test. This means that directly translating a technique designed to evaluate biomedical-ontology-coverage to the EES domain will likely generate results with a significantly different representation of *quality*.

However, the basic premise of coverage evaluations is still valid for EES ontologies as they share a similar underlying structure. As will be discussed in the following section, our work expands the biomedical techniques to include a coverage approach for the subclass axiom, enabling a greater applicability for the EES domain and discusses the interpretation of coverage for the EES domain.

5 Synonym Symmetry Methodology

As previously mentioned, the coverage methodology used in this work was introduced by Yao and colleagues [28] for the biomedical domain. Their approach allows for the quantitative assessment of relevancy for an ontology to a given corpus. To expand the applicability of their approach for the EES domain, we add a step to perform Synonym Synergy, which enables the automatic classification of subclass axioms. Also, it should be noted that for this process we used OWL API[1].

The remainder of this section is laid out as follows: next is the formal definitions for coverage, then an explanation and examples of our novel expansion to the previously introduced methodology, Synonym Synergy and lastly the eight-step process to generate coverage for an ontology.

5.1 Coverage Definitions

For disambiguation of discussion, we present the four formal definitions of *coverage* that are classified upon completion of this eight step process. Let A be the set of classes $a_0, a_1...a_n$ within ontology X and let B be the set of classes $b_0, b_1...b_m$ within ontology Y. Then define the *Class Coverage* of ontology Y on ontology X with the following equation:

$$CC = \frac{\sum_{i=0}^{i=m} P_i}{|B|}, \text{ where } P_i = \begin{cases} 1 \text{ if } b_i \in A \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

[1] http://owlapi.sourceforge.net/

This measure captures the fraction of terms from ontology B that are also present in A.

Let C be the set containing the equivalence relations $c_0, c_1...c_o$ within ontology X and let D be the set containing the equivalence relations $d_0, d_1...d_q$ within ontology Y. Then we define *Equivalence Coverage* (EC) of ontology Y on ontology X with the following equation:

$$EC = \frac{\sum_{i=0}^{i=q} P_i}{|D|}, \text{ where } P_i = \begin{cases} 1 \text{ if } d_i \in C \\ 0 \text{ otherwise} \end{cases} \tag{2}$$

Let E be the set containing the subclass relations $e_0, e_1...e_r$ within ontology X and let D be the set containing the subclass relations $f_0, f_1...f_s$ within ontology Y. Then we define *Subclass Coverage* (SC) of ontology Y on ontology X with the following equation:

$$SC = \frac{\sum_{i=0}^{i=s} P_i}{|F|}, \text{ where } P_i = \begin{cases} 1 \text{ if } f_i \in E \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

Lastly, let CC_x, EC_x, SC_x be the *Class Coverage*, *Equivalence Coverage*, and the *Subclass Coverage* respectively for ontology X. We then define the *Breadth Coverage* (BC) of ontology X with the following equation:

$$BC = g_1 CC + g_2 EC + g_3 SC, \text{ where } BC <= 1 \tag{4}$$

where g_1, g_2, and g_3 are user defined constants. These constants can be used to weight a particular type of coverage more than another. For example, consider a use case where normal classes are of little importance, subclass axioms are of primary importance, and equivalence axioms are completely irrelevant; in this case g_1 might be 0.1, g_2 might be 0, and g_3 might be 0.9. Note that each type of coverage will always have a score that is less than or equal to one.

5.2 Synonym Synergy

Due to the large quantity of subclass axioms within ontologies representing the EES domain, we present a novel expansion to the standard methodology presented by Yao and associates entitled SYNONYM SYNERGY. This approach requires a list of synonym relationships between word pairs. More simply, for a pair of words that contain a synonym relationship between them, our heuristic calculates whether it is bi-directional or uni-directional. In this case, we say that a pair of words has a *bi-directional* relationship if both words are synonyms of each other. In contrast, we say that a pair of words has a *one-sided* or *uni-directional* relationship if only one word is the synonym of the other. Visually, this could be represented as a graph where each word is a node and a directional edge represents a synonym relationship from one word to another. Figure 1 presents some examples of this type of analysis.

Consider the following example with the words *bovine* and *cattle*. The synonyms for the word *bovine* includes *cattle*, whereas the the synonyms for the word *cattle*

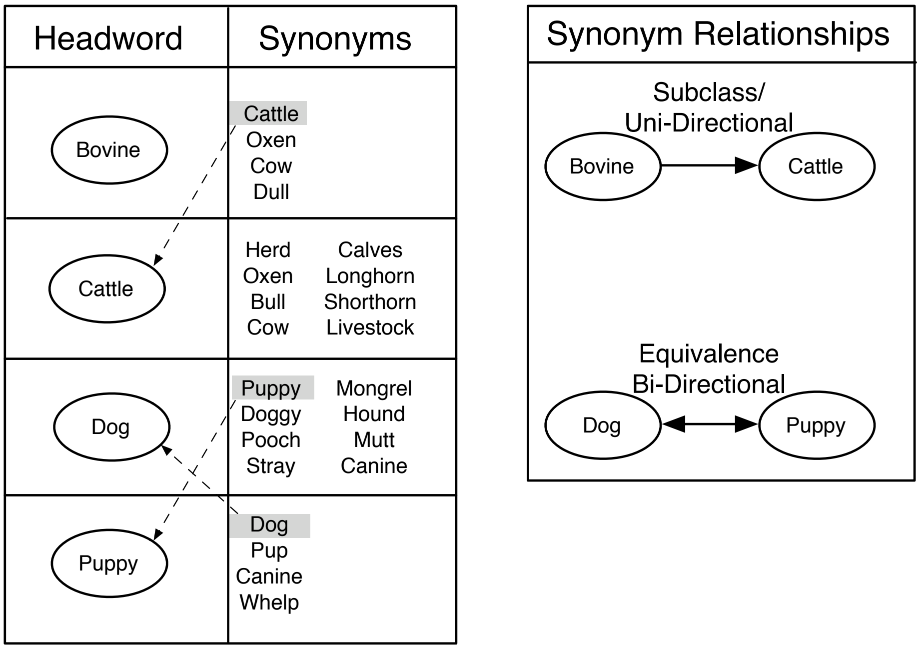

Fig. 1: An example of using synonym sets to determine equivalence and subclass relationships.

does not include *bovine*. In this case, the relationship is *one-sided* and can be interpreted in the following way: all bovines are cattle, but not all cattle are bovines. In other words, *bovine* is a subclass of *cattle*.

Synonym Synergy relies on the quality of the thesauri used as an incorrect listing of synonyms will greatly hamper this process. We limit this threat by leveraging seven separate thesauri (see Section 7.4 for a more thorough treatment of our thesauri).

In addition, even though this process is likely to insert erroneous subclass relationships into the corpus ontology, these erroneous relationships should have little to no impact upon the results. Consider an example of the word *cow*. While the most common use of the word *cow* refers to a four-legged animal used on farms, there is another use of the term that refers to intimidating someone. This means that it is likely that *cow* will become a subclass of *intimidate*. However, because the methodology performs a type of precision analysis (see Section 5.1), this false subclass relationship will likely be ignored.

In other words, when we calculate coverage, the heuristic attempts to match each relationship found within the ontology-under-test to the corpus ontology; meaning that coverage is only impacted if the ontology-under-test has that exact same spurious relationship as the corpus ontology using Synonym Synergy. Indeed, it is exceedingly unlikely that a given EES ontology, which by design topically only represents EES relationships, will have the exact same spurious relationship as generated by this approach. Therefore, these spurious relationships will likely have little or no impact upon the coverage of the ontology-under-test.

Actually computing the uni-directional or bi-directional relationships for each word is fairly straight forward. Using a standard *dictionary data structure* (i.e., a structure utilizing key-value pairs), we insert each key as a head word, and the values as a list of synonym words. Consider an abstract example.Imagine that when creating a corpus ontology, the system finds the word `X`. First, the system queries the thesauri to gather a list of synonym words for `X`. Then the system inserts `X` as a key, with the synonyms as its value. Next, for each word in `X`'s' value list, the system queries the thesauri to determine whether it contains `X`. If it does, the system determines this is a bi-directional relationship (as both words are synonyms of each other), and if not then it determines this is a uni-direction relationship, or a subclass (as only one word was a synonym of the other). More concretely, consider Figure 1. If the word bovine was read, it would be inserted into the dictionary as a *key*, and the words: cattle, oxen, cow, and dull would be inserted into a list that comprises the *value*. Then for each word (cattle, oxen, cow and dull) the system would check whether they contain the word bovine as a synonym (by querying the thesauri). It finds the case of cattle, it does not contain the synonym bovine, meaning that this relationship is uni-directional (or a subclass relationship). This process is repeated for each word encountered, and while we envision that there may be many other ways to compute this directionality, and do so in an optimized fashion, this method is presented because of its simplicity to understand and replicate.

In short, SYNONYM SYNERGY allows for the use of the natural-language definitions to inform our corpus ontology of potentially meaningful subclass relationships as defined by the synonym relationship between word pairs.

5.3 Coverage Methodology

Table 2: The statistics for each of thesauri used in this work.

Thesauri	Headwords	Synonym pairs	Synonyms Per Headword
The Synonym Finder [14]	20,249	758,611	37.46
Webster's New World Roget's A-Z Thesaurus [13]	29,925	329,669	11.01
21st Century Synonym and Antonym Finder [12]	7,507	146,806	19.55
The Oxford Dictionary of Synonyms and Antonyms [24]	8,487	105,902	12.47
A Dictionary of Synonyms and Antonyms [5]	3,771	57,366	15.21
Scholastic Dictionary of Synonyms, Antonyms and Homonyms [10]	2,147	19,759	9.20
WordNet [20]	115,201	306,472	2.66

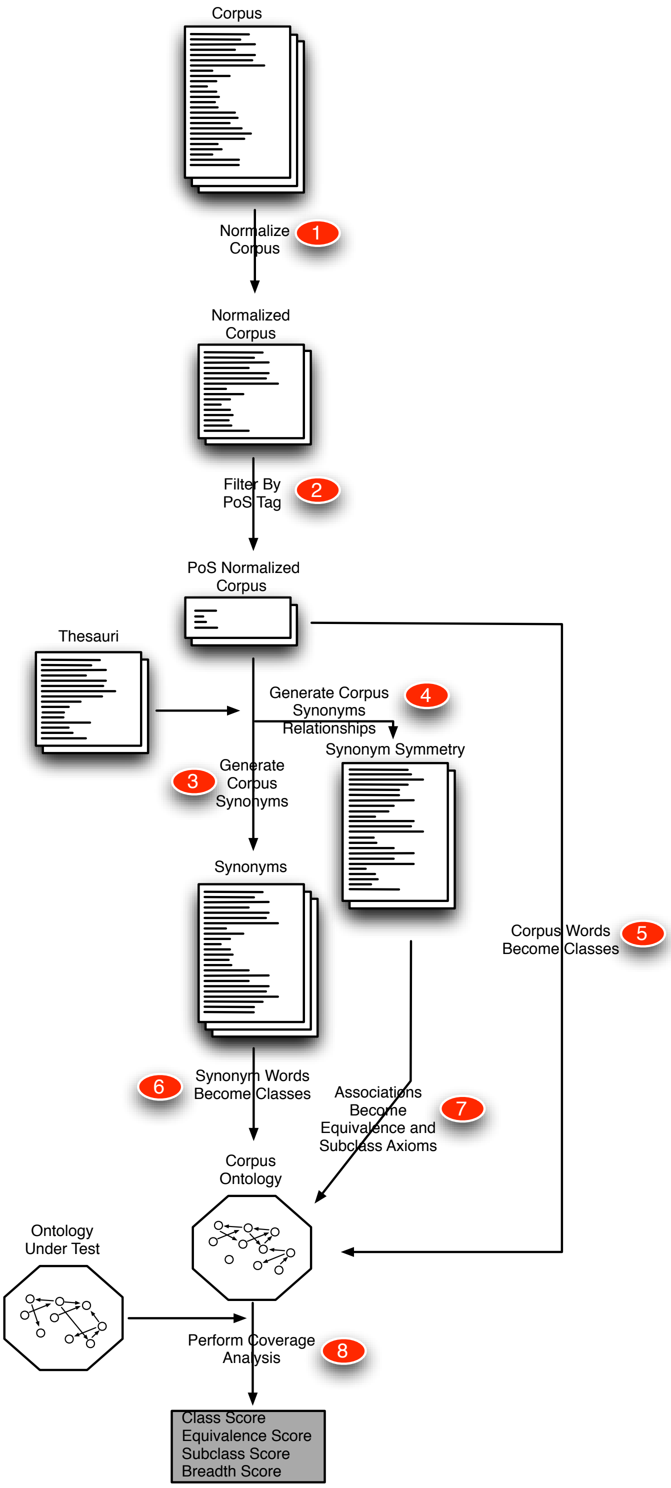

Fig. 2: A process model for coverage methodology with the step numbers in red circles. At a high level, the corpus is turned into an ontology and compared against the ontology under test.

The coverage methodology presented in this section can be completed in eight steps as displayed in Figure 2. Each of these steps was introduced by Yao and colleagues for the biomedical field, and save the introduction of Synonym Synergy, are only altered slightly to better match the EES domain [28]. The goal is to calculate coverage for an existing ontology, referred to as the *ontology under test*. At a high level of abstraction, this process transforms an existing corpus into an ontology, and then calculates a type of *precision* score for the elements within the ontology under test.

Step 1: Normalize corpus. This step takes as input an existing corpus and produces a normalized corpus. While normalization can take many forms depending on the needs of the user, we describe the natural-language processing techniques used in this paper. First, stem all words within the corpus to reduce words to their base, thus normalizing tense and plurality. In our case, we used the Porter stemmer[2], which is a rule-based stemmer. Next remove all stop words. Stop words are short, function words which provide no semantic meaning, but are necessary for proper speech (e.g., is, the, who, are, and on). Then remove all grammar and words containing only numeric characters. This filter is important because most ontology class names contain some alphabetic characters. Lastly change all words to be lowercase to prevent issues with case sensitivity.

Step 2: Filter Part of Speech (PoS) Tag. This step takes as input a normalized corpus and produces a corpus only containing words from specific parts of speech. Because our goal is to transform a corpus full of words into classes and axioms, we want to filter out those words which are likely not going to be classes. In this case, we used the Natural-Language-Toolkit part of speech tagger (PoS) to remove all words which were not nouns, verbs, adverbs, or adjectives [1].

Step 3: Generate Corpus Synonyms. This step takes as input all words from the normalized, PoS filtered corpus and multiple thesauri to produce a list of words that are all synonyms of at least one corpus word. In our case, we utilize seven thesauri (as shown in Table 2) so as to capture any possible synonym of a word found in our corpus [5, 10, 12–14, 20, 24]. This step ensures that the corpus ontology we generate will have a large variety of words to describe any particular topic discussed in the corpus. This helps to prevent an artificially low coverage score because of word choice.

Step 4: Generate Corpus Synonym Relationships. Step four is unique in that it encompasses our novel expansion Synonym Synergy. This step takes as input all the words from the normalized, PoS filtered corpus and all the *synonym relationships* from our seven thesauri and produces all equivalence and subclass axioms. By *synonym relationships* we refer to the potential uni-directionality or bi-directionality of a word pair experiencing a synonym. We say that two words are equivalent (and thus contain this axiom) if they both are synonyms of each other (i.e., their synonymy is bi-directional). However, if for a pair of words only a single word is a synonym of the other, we classify them as a subclass (i.e., their synonymy is uni-directional). The rational and limitations of classifying subclasses this way was be discussed in the previous section.

Step 5: Corpus Words Become Classes. This step takes as input the words from the normalized, PoS filtered corpus and inserts them as classes into a blank

[2] http://tartarus.org/martin/PorterStemmer/

ontology (heretofore referred to as the corpus ontology). This step is fairly straightforward.
Step 6: Synonym Words Become Classes. This step takes as input the words from generated synonym list and inserts them as classes into a corpus ontology. This step is fairly straightforward.
Step 7: Associations Become Equivalence and Subclass Axioms. This step takes as input the results from calculating the synonym symmetry, and inserts the equivalence and subclass axioms for the corpus ontology. This step is fairly straightforward.
Step 8: Perform Coverage Analysis. This step takes as input the corpus ontology and the ontology under test and produces the *class, equivalence, subclass,* and *breadth* coverage scores. Following the formulas presented in Section 5.1 we can generate all these scores. For simplicity, we present an example of how the first three scores are generated; note that the breadth score is simply a linear combination of the first three scores that equate to less than or equal to one. Consider a corpus ontology that contained classes

$$A, B, C, D, E, F$$

and the ontology under test contained classes

$$A, B, C, X, Y, Z$$

We would iterate through each element in the ontology under test set and determine whether it existed in the corpus ontology set. We would find that while A, B, and C exist in the corpus ontology set while X, Y, and Z do not. This would result in a score of 3/6 or 0.5 because three elements matched, and the total cardinality of the ontology under test is six.

This methodology provides the ability to quantitatively and automatically calculate the relevancy of an ontology for a specific domain.

5.4 Scalability

To better understand the scalability of Synonym Synergy, there are two factors that need to be considered: creating an ontology from a corpus, and evaluating coverage for an ontology.

The creation of an ontology for a corpus comprises steps one through seven of our previously described process. The time taken to perform these steps is largely dependent upon two things, the number of unique words within the corpus documents, and the number of headwords and synonym pairs within the combined thesauri used. In the case of the corpus used in our experiments (which will be discussed more fully in Section 6), it had over 27,000 unique words (the stats for our thesauri were described in Table 2). To create our corpus ontology took roughly 48 minutes on a Intel 2 core 2.13 GHz processor with 4GB of ram. To create our Tailored corpus (which will be discussed more fully in Section 6), which had slightly more than 10,000 words took TODO FILL ME IN OMG FILL ME IN. In both cases, nearly the entire time taken was used to during steps four and and seven (those steps that compute the directionality of a synonym relationship for a given word). It should be noted however, that this process (the creation of

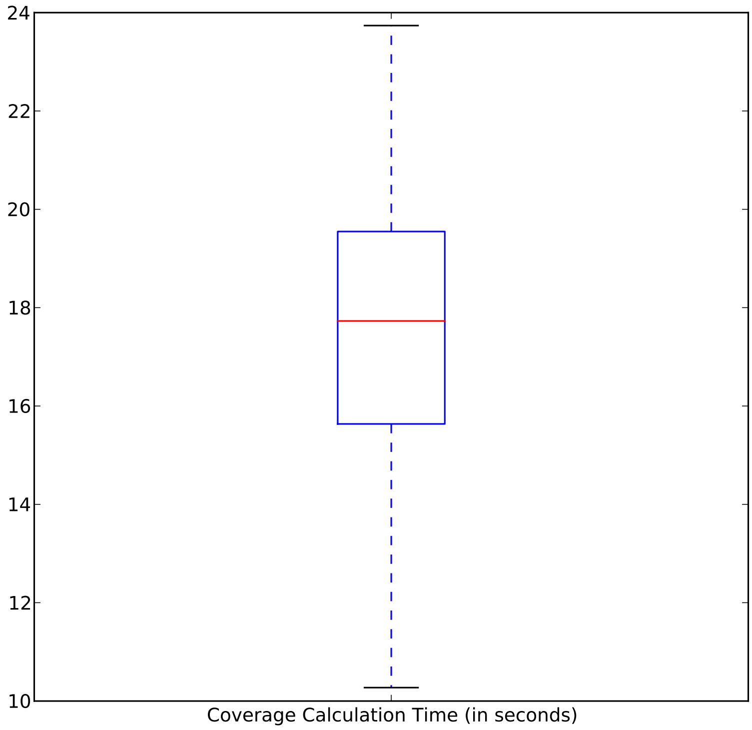

Fig. 3: A boxplot representing the time to calculate coverage for a single ontology in seconds.

a corpus ontology) is a one time expense. Indeed, as documents are added to the corpus, these can be added individually (or collectively) to the existing corpus ontology such that no work needs to be repeated. As such, we feel that this aspect appears to be very scalable in practice.

To better understand the scalability of coverage evaluation, we present Figure 3. This boxplot demonstrates (in seconds) how long it takes to generate coverage results (i.e., step eight of our process) for a single ontology. It should be noted that these results come from generating coverage for the SWEET ontologies, which tend to be small. However, in total we used 217 ontologies, and the median time is 18 seconds. This means that to generate coverage for all 217 ontologies takes roughly 65 minutes. While this expense is larger than the creation step, generating coverage for this large and diverse takes about an hour. Thus, we feel that this aspect also is very scalable in practice.

Table 3: Values of our Independent and Dependent variables.

Independent Variables	DataONE corpus, Tailored Corpus
Dependent Variables	Class, Subclass, Equivalence, and Breadth Coverage

6 Experiments and Results

One fundamental purpose for the methodology presented in the previous section is to provide a meaningful representation of the relevancy of an ontology or set of ontologies for a specific domain. This methodology elucidates understanding as scientist strive to refine existing ontologies to better approximate the community's semantic understanding. Within the Earth and Environmental Sciences domain, the SWEET ontologies have become the *de facto* standard. However, there are no studies that evaluate the relevancy of the SWEET ontologies for this EES domain. This dearth if studies poses a simple but important research question:

Research Question: In the best case, how well do the SWEET ontologies match the EES domain?

6.1 Experimental Setup

To answer our research question, we designed the following experiment. At a high level of abstraction, our experiment works as follows: we leverage multiple corpora, one of which is designed to specifically match the SWEET ontologies, and another which comes from an existing EES research lab to calculate the expected coverage scores for SWEET ontology. Then, by comparing these scores, we can determine an expected best-case relevancy score.

The outline of our experimental variables is given in Table 3.

Our experiment has two independent variables: the DataONE corpus [19], and the tailor-made corpus. The DataONE corpus comes from Data Observation Network for Earth foundation, an environmental-science distributed-framework with multiple data centers and organizations each representing a unique element of the EES domain. At the time of writing this, the DataONE corpus contains 46,428 documents, each of which is typically smaller than 2 MB. Our initial investigation leveraging dimensional reduction techniques revealed that this corpus contains nearly 1,000 unique topics and over 27,000 unique words. This corpus represents a real corpus used by EES researchers and practitioners.

The tailored corpus is a randomly selected set of documents that have topical harmony with some of the SWEET ontologies. We followed a straightforward and simple process to generate this corpus, which is comprised of four steps: (1) randomly select a SWEET ontology, (2) select three classes at random from said ontology, (3) search Google Scholar using the three classes as a single query, and (4) select the top two articles returned by Google Scholar and add them to the corpus. It is important to note that we repeated this process eight times, resulting in a corpus that was comprised of 16 documents from various EES research journals. For transparency, we briefly describe our rationale behind each of these steps.

Step one has us select a SWEET ontology at random. Because no one SWEET ontology is more important than any other, we wanted to ensure that the topics reflected in our tailored corpus were randomized with respect to specific SWEET ontologies. Step two has us select three class names from the selected ontology. SWEET ontologies on average have around 20 unique classes per ontology. We select class names from the ontology, to form a type of representative "topic" for the given ontology. The number three was used because on average, it was 15% of the total number of classes for any SWEET ontology, which we felt was sufficient to represent the ontology. Step three has us use Google Scholar to identify relevant scientific articles. We wanted to ensure that the articles selected were truly topically similar to the SWEET ontology selected in step one. Thus, we used all three class names from step two to perform the query with Google Scholar. Google Scholar was selected because anecdotally it has proven reliable and effective at discovering relevant articles within the EES domain. Step four has us add the top two articles returned by Google to the tailored corpus. We choose two articles per query because there are multiple ways to discuss a topic, and by selecting two articles there was an increased likelihood that, (1) the topic from the SWEET ontology was being discussed, and (2) the topic was being discussed in a unique way, ensuring a greater potential for coverage. This tailored corpus represents a *best case* scenario in which the documents that make up the corpus are exactly what the SWEET ontologies are trying to represent, for the eight selected SWEET ontologies.

Our process to generate an ontology and calculate coverage for the tailored corpus follows the same process as identified in Figure 2, but the "corpus" is replaced with the "tailored corpus", resulting in a "tailored ontology" after step 7.

For reference, the tailored corpus contains 16 documents, each of which is less than 100 KB. An initial investigation leveraging dimensional reduction techniques reveals that this corpus contains roughly 6-10 unique topics and over 10,000 unique words.

Our four dependent variables are: Class Coverage, Subclass Coverage, Equivalence Coverage, and Breadth Coverage. Each of these coverage types are explicitly described in Section 5.1. It is of note that while we include equivalence coverage, very few SWEET ontologies have any equivalence axioms. However, because a few ontologies possess it, we calculate it with the rest of the more relevant coverage scores.

The steps to perform our experiment are fairly straightforward. First, transform each corpus into an ontology using the coverage methodology. Second, generate coverage for each of the SWEET ontologies for each corpus. Third, compare results for each category. Note that the tailored corpus represents a likely *best* possible score given our coverage framework, and represents a potential *ideal* case. Thus, the difference between the coverage results from the DataONE corpus and the tailored corpus provides an empirical baseline as to the expected coverage given this methodology within the EES domain.

Table 4: Results from the paired t-tests between the two ontologies.

Dependent Variable	P-Value
Class Coverage	0.2723
Subclass Coverage	0.4048
Equivalence Coverage	NaN
Breadth Coverage	0.2525

6.2 Experimental Results

We present our results in two subsections, the first examines the difference between the DataONE corpus and the tailored corpus, and the second compares results between the SWEET ontologies used to create the tailored corpus, and those not used.

6.3 DataONE and the Tailored Corpus

Figures 4 and 5 present the coverage scores for the DataONE corpus and the Tailored corpus respectively. When considering the tailored corpus, it should be noted that out of 217 SWEET ontologies, 29 had zero for each dependent variable, 170 had positive scores only for class and breadth coverage, and 18 had a positive score on class, subclass and breadth coverage, but no ontology had a positive score for equivalence. However, when considering the DataONE corpus, it should be noted that out of 217 SWEET ontologies, 26 had zero for each dependent variable, 169 has positive scores only for class and breadth coverage, and 22 had a positive score on class, subclass and breadth coverage, but again no ontology had a positive score for equivalence. Further, the ontologies which received a score of zero on all dependent variables was different for each corpus.

These figures both display an average class score of 36%, an average subclass score of around 4% and a breadth score around 12%. When calculating the breadth score, we weighted all categories equally. That being said, due to the zero scores in the equivalence coverage category, the breadth score appears artificially low. Indeed, if equivalence coverage (an axiom not very relevant to the SWEET ontologies) were removed from this calculation, the breadth score mean would be in the low thirties.

These figures also display a large potential range. While some ontologies score zero in all categories, other ontologies score nearly 90% in class coverage. It is significant that no ontology scores over 10% in subclass coverage.

Lastly, we performed a paired t-test for each dependent variable between the two corpora. The results of these t-tests are presented in Table 5. It should be noted that in this case, the equivalence coverage results were identical between the two corpora (all zeros) resulting in an incalculable score. In this case the graphs of these two results visually appear very similar, and our statistical test affirms that indeed there is no statistically significant difference between them.

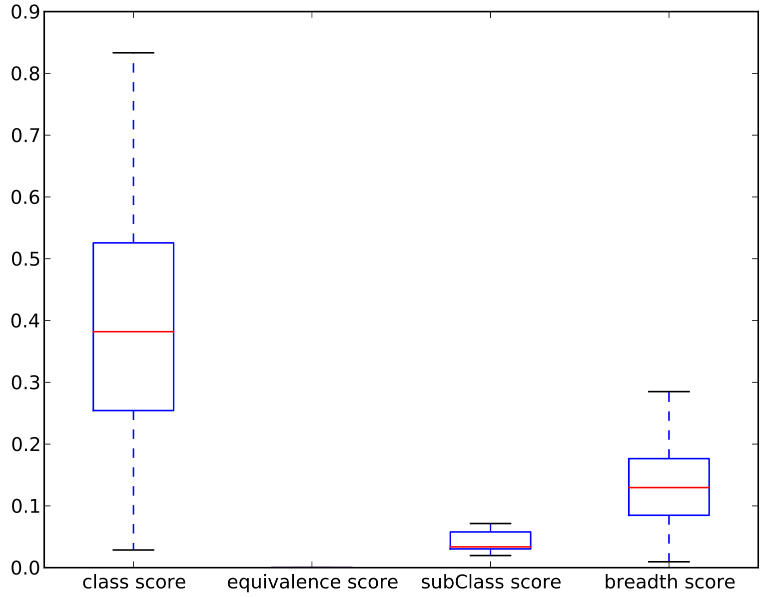

Fig. 4: A boxplot representing the coverage for each of our dependent variables on the DataONE corpus. The center line is the mean, the box represents the first and third quartiles and the dotted lines represent the min and max.

6.4 Tailored Corpus

Figures 6 and 6 represent two coverage analyses using the tailored corpus. Figure 6 presents coverage with only those SWEET ontologies which were *not used* to create the tailored corpus, and Figure 6 presents coverage of only the SWEET ontologies *used* to created the tailored corpus. Thus, Figure 6 can be considered the best possible score given out approach. For simplicity of discussion, we refer to the results utilizing *only* those ontologies that were used to create the tailored corpus as the *golden results*, and the others as the *non-golden results*.

When considering the golden results, it should be noted that out of 8 SWEET ontologies, none had zero for each dependent variable, and all had positive scores only for class, subclass and breadth coverage, but no ontology had a positive score for equivalence. However, when considering the non-golden results, it should be noted that out of 209 SWEET ontologies, 26 had zero for each dependent variable, 169 has positive scores only for class and breadth coverage, and 14 had a positive score on class, subclass and breadth coverage, but again no ontology had a positive score for equivalence. Further, the ontologies which received a score of zero on all dependent variables was different for each corpus.

Thus, of note first is that there are two primary differences between these results: (1) the golden results score positively in every variable (except equivalence),

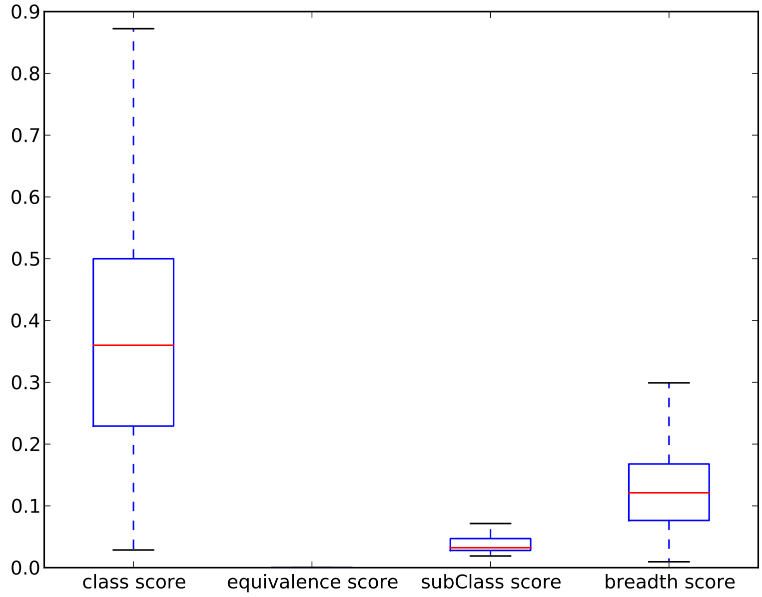

Fig. 5: A boxplot representing the coverage for each of our dependent variables on the tailored corpus. The center line is the mean, the box represents the first and third quartiles and the dotted lines represent the min and max.

whereas the non-golden results had only 14 ontologies gain a positive score in each variable; and (2) the golden results score significantly higher in class score and breadth than the non-golden results.

The golden results display a median class score of 59%, a median subclass score of around 6% and a breadth score around 21%. In contrast, the non-golden results display a median class score of 38%, a median subclass score of around 0% and a breadth score around 14%. When calculating the breadth score, we weighted all categories equally. That being said, due to the zero scores in the equivalence coverage category, the breadth score appears artificially low. Indeed, if equivalence coverage (an axiom not very relevant to the SWEET ontologies) were removed from this calculation, the breadth score mean would be in the low thirties.

These figures also display a difference in the large potential range of some variables. For example, when considering the non-golden results, some ontologies score zero in all categories, other ontologies score nearly 90% in class coverage. However, the golden results never score below 55% in terms of class score. It is significant that no ontology scores over 10% in subclass coverage.

Lastly, we performed a paired t-test for each dependent variable between the two results. The results of these t-tests are presented in Table 5. It should be noted that in this case, the equivalence coverage results were identical between the two corpora (all zeros) resulting in an incalculable score. In this case the

Table 5: Results from the paired t-tests between the two ontologies.

Dependent Variable	**P-Value**
Class Coverage	$p < 0.05$
Subclass Coverage	$p < 0.05$
Equivalence Coverage	NaN
Breadth Coverage	$p < 0.05$

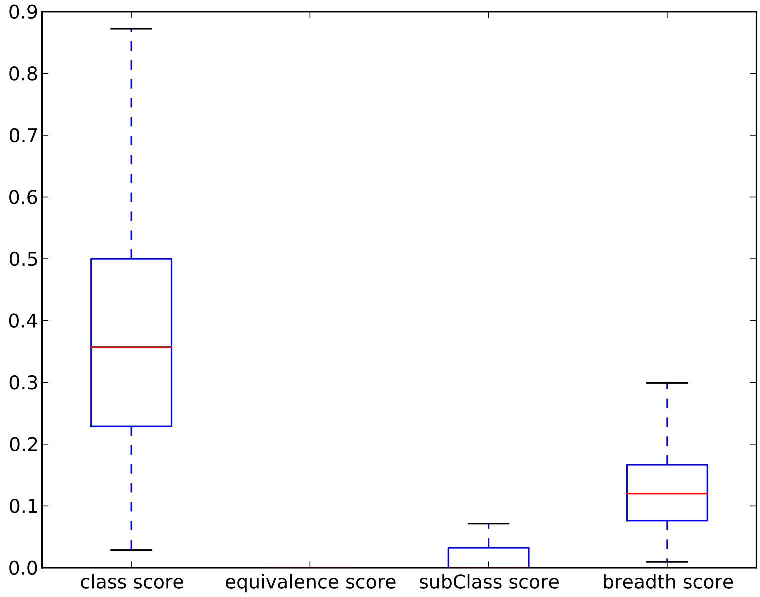

Fig. 6: A boxplot representing the coverage for each of our dependent variables on the tailored corpus when only considering those ontologies not used in creating the corpus. The center line is the mean, the box represents the first and third quartiles and the dotted lines represent the min and max.

graphs of these two results visually appear very dissimilar, and our statistical test affirms that indeed there is a statistically significant difference between them when considering class, subclass, and breadth scores.

7 Discussion

While the previous section described the results of our experiment, we present our discussion here.

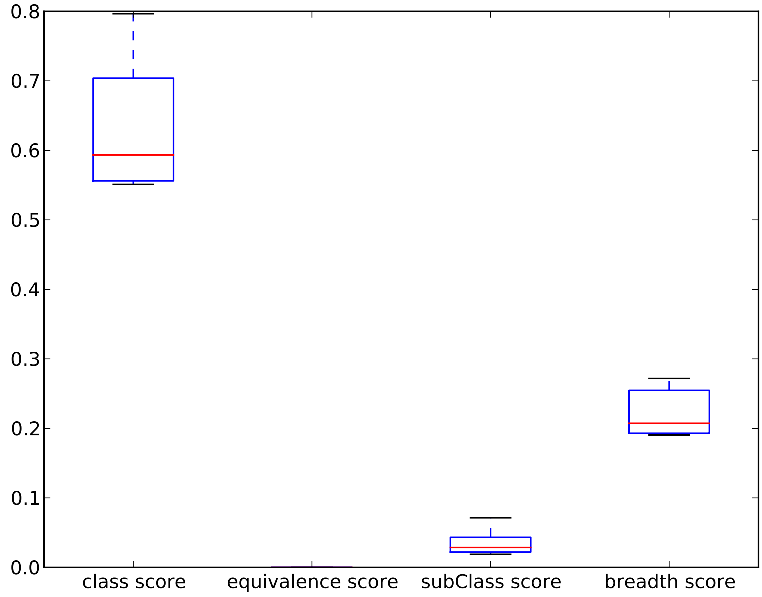

Fig. 7: A boxplot representing the coverage for each of our dependent variables on the tailored corpus when only considering those ontologies used in creating the corpus. The center line is the mean, the box represents the first and third quartiles and the dotted lines represent the min and max.

7.1 Relevancy of SWEET

When considering the results of the tailored corpus on all the SWEET ontologies, we were initially surprised to see scores so low. Indeed, finding that on average less than 40% of the classes of a given ontology were found within the corpus was somewhat shocking. However, it should be remembered that the SWEET ontology set has over two-hundred ontologies, and our process to generate the tailored corpus only drew topics from eight of them. Thus, when considering only those eight, having an median class score of 59%, signified a substantial improvement meaning that the scores would likely have been much higher if we had added articles to represent each ontology within the SWEET set. That being said, a likely interpretation of these results is that the SWEET ontologies cover a significant quantity of disjoint topics within the EES community, such that discussing one, does not entail even referencing the other; and simultaneously that some SWEET ontologies contain significant overlap (such that the topics found within eight of the set, can generate a score around 40% overall). This would explain the numbers with the tailored corpus within the class scores.

However, it is significant that overall, the SWEET ontologies received a statistically identical set of scores for the DataONE corpus. This result means that regardless of whether we used the tailored corpus, which was generated to specifically match the topics of SWEET, or a real-life corpus from the EES community, the SWEET ontologies are expected to get statistically identical relevancy scores. This finding is significant as it points to the quality and domain relevancy of the SWEET ontologies to the EES community.

It is also of significance that the ontologies which scored high (and low respectively) were different for each corpus. Indeed, some ontologies received all zeros with one corpus and received relatively high scores for the other corpus. This result demonstrates that while our corpora described topically different sets, the SWEET ontologies were still able to cover them.

> The answer to our research question is that the SWEET ontologies accurately represent the EES domain and that using this coverage methodology will likely provide a best-case score around 70% for classes, 7% for subclasses, and 40% for breath (if equivalence axioms are ignored).

7.2 Value of Synonym Synergy

As for the low subclass score, we attribute this to a deficiency within the Synonym Synergy approach, more than a demonstration of limitations within the SWEET ontologies. This is because not all subclass relationships can be identified using Synonym Synergy. For example, *hunger* is a subclass of *needs*, but they share no synonym relationship. In other words, while there is a relationship between these ideas conceptually, their alphabetic words share no association. However, we do not take this to mean that Synonym Synergy is not valid or useful. Indeed, our results demonstrate that while not all or even most relationships between ideas share an alphabetic tie, a significant portion do. Synonym Synergy allowed for the coverage of nearly 10% of all subclass relationships within the SWEET ontologies. Further, Synonym Synergy requires only a thesaurus as a new information source. As mentioned earlier, many existing techniques require specific hand-crafted heuristics or standards to evaluate quality, whereas this approach is lightweight and automatic in comparison while still producing meaningful scores. In other words, Synonym Synergy is valuable though not sufficient in isolation. While we believe this to be a good first step, and these results show promise, more approaches are needed to approximate the relationships between ideas within an ontology.

> Our results indicate that Synonym Synergy can successfully identify and enable the coverage of non-trivial subset of subclass axioms. However the use of directional synonymy is not sufficient to capture all subclass relationships.

7.3 Applicability of Coverage Methodology

One cause for the low scores in all categories was found in the way our methodology works, and the classes within the SWEET ontology. Remember that our methodology (as described in Section 5.3) parses words from natural text, meaning they are very unlikely to get camel-cased words (e.g., humanNeeds, pollutedLand, or bigBusiness) as they are typically not used within normal English text. In contrast, the SWEET ontology frequently uses camel-cased class names. Our methodology cannot match these with its existing implementation, meaning that these classes (or axioms containing these classes) are currently uncoverable. Consider that the SWEET ontologies we used contain 8,641 unique classes, and that 3,775 are camel cased; meaning 43.6% of the classes we observed were uncoverable using our methodology (i.e., at best, coverage could score 57% and the average class score was nearly 40%). This explains a large degree of the missed classes and subclasses.

Another limitation of this methodology is that it does not take into account misspellings or abbreviations. For example, if a document contained a typo (which is translated to a misspelled class), it would likely be ignored by our thesauri (as it is not a real word), and represent a hole where ontologies would be unable to find coverage, even if they are describing the same thing. This type of problem has been addressed in natural-language research with techniques like fuzzy-spelling, but none are implemented in our methodology. In addition, the limitation with abbreviations is that in many cases, these require a type of *usage dictionary* [15] to guide a heuristic in selecting the correct expansion of a word. These dictionaries are important as abbreviations have different meanings in different contexts. For example, consider `mem`. `Mem` is likely `memory` when found within computer programs, but is a letter of the alphabet when discussing in Hebrew and Arabic. Having a usage dictionary to expand these to the proper word for a given domain would also likely increase the quality of the coverage result.

> We found that this approach from the biomedical domain applicable in the EES domain, though it requires alterations to account for different axioms (i.e., subclass and restrictions) the use of camel-case, and how results are interpreted.

7.4 Multiple Thesauri

As mentioned in Section 5.3 and in Table 2, our approach leveraged seven thesauri including WordNet. However, many research endeavors using synonyms have relied exclusively upon WordNet as it is an easily available tool that provides synonym sets for any given word within its database. Indeed, looking back at Table 2 it is important that WordNet has more headwords than all of the other six thesauri combined. A reasonable question then might be, why not exclusively use WordNet, or in other words, why use seven thesauri?

One of the main efforts of the methodology is to create an ontology from the corpus that represents a domain. It does this by ensuring that issues such as word choice, are addressed. For example, without synonyms consider a corpus that used the word `puppy` but the ontology-under-test contained the class `dog`; this would result in a hole in the coverage notwithstanding the idea being present in the corpus. Thus it is of critical importance that *all* possible synonyms are captured when the thesaurus is checked in step 3.

One way to ensure that no synonyms are missed is to include more thesauri with different headwords and synonyms. If each thesaurus has only a small quantity of overlap, then we likely are going to ensure fewer missed synonyms. Indeed, Yao and colleges present a discussion on these seven exact thesauri and demonstrate that they have very little overlap (i.e., most headwords and synonyms within a thesaurus are unique when compared to the other six) [28].

Another concern with exclusively using WordNet is that while it has a lot of headwords, it has relatively few synonyms. We can easily calculate on average, how many synonyms exist per headword within a given thesaurus. Given our goal to not miss any synonym, ideally the perfect thesauri for this methodology would have high synonym-per-headword ratio and a high quantity of headwords. Table 2 displays these quantities. We find that while WordNet has the most headwords, it has the fewest synonyms-per-headword by a factor five to ten. This means that while there are many headwords in WordNet, it contains the fewest synonyms for any given headword (on average). In this case, we see that WordNet would not be sufficient, and in one way is the least ideal for our methodology.

> The coverage methodology utilized in our study requires a large quantity of headwords and synonyms-per-headword making WordNet in isolation not sufficient or ideal. A larger set of thesauri with non-overlapping words is more likely to generate meaningful results.

8 Conclusions and Future Directions

This work has presented an expansion to, and empirical investigation of the applicability of existing ontology coverage techniques for the Earth and Environmental Sciences (EES) domain as well as experiments to empirically evaluate the relevancy of the SWEET ontologies to the EES domain.

Due to the domain specificity of ontology organization (e.g., size or number of axioms), it was unclear whether the coverage techniques used in other fields would provide meaningful results for the EES domain. We investigated the assumptions and organizations of ontologies within the biomedical field and found them to be substantially different from the paramount ontologies within the EES domain (i.e., the SWEET ontologies). We found that while the underlying assumptions were similar enough to warrant the use of existing biomedical-ontology-coverage techniques within the EES domain, the results were likely to be significantly different.

In addition to the use of existing techniques, we provided a novel expansion called Synonym Synergy. Synonym Synergy uses thesauri to capture subclass relations and enable their coverage. Our study demonstrated that Synonym Synergy enabled the coverage of roughly 10% of all subclass relationships within the SWEET ontologies. These results reinforced that Synonym Synergy is valuable in that it captures those subclass relations where the alphabetic name of the classes also share a relationships, but does not capture those relationships where only the ideas share a relationship. In other words, Synonym Synergy is useful to capture some subclass relations, but not sufficient to capture a majority of subclass relations.

Lastly, we performed an experiment upon the SWEET ontologies. While prevalent within the EES domain, there were currently no studies to validate their use or relevancy. Our studies found that the SWEET ontologies had a statistically identical coverage score to a real-world EES corpus as they did to a hand-tailored ontology specifically chosen to represent the SWEET ontologies. This investigation suggested that the SWEET ontologies do sufficiently represent the EES domain, and are broad in their topic coverage.

In the future, this work can be improved upon in a few key areas. First, Synonym Synergy can be improved upon through additional heuristics that capture a greater percentage subclass relationships. Further, techniques which can capture other relationships (e.g., restrictions) would also be valuable. In addition, more experiments with varied corpora would provide greater confidence in generalizability of these results and increase understanding as to the domain-specificity of our findings. Lastly, simple additions that account for camel-cased class-names, and the matching of misspellings and abbreviations within the corpus will likely improve the results of the coverage.

9 Acknowledgments

This material is based upon work supported by the National Science Foundation, through Award CCF-1116943 and through Graduate Research Fellowship under Grant No. DGE-0808392. Michael Huhns was extremely helpful in directing and crystalizing this research. We would also like to thank Andrey Rzhetsky for providing the seven thesauri used in our experiment.

References

1. S. Bird, E. Loper, and E. Klein. Natural language processing with python. 2009.
2. J. Brank, M. Grobelnik, and D. Mladenić. A survey of ontology evaluation techniques. In *Conference on Data Mining and Data Warehouses (SiKDD 2005)*, 2005.
3. P. Cimiano, A. Hotho, and S. Staab. Learning concept hierarchies from text corpora using formal concept analysis. *J. Artif. Intell. Res.(JAIR)*, 24:305–339, 2005.
4. K. Dellschaft and S. Staab. On how to perform a gold standard based evaluation of ontology learning. In *The Semantic Web-ISWC 2006*, pages 228–241. Springer, 2006.
5. J. Devlin. *A dictionary of synonyms and antonyms.* 1961.
6. A. Doan, J. Madhavan, R. Dhamankar, P. Domingos, and A. Halevy. Learning to match ontologies on the semantic web. *The VLDB JournalThe International Journal on Very Large Data Bases*, 12(4):303–319, 2003.
7. A. Gibson, K. Wolstencroft, and R. Stevens. Promotion of ontological comprehension: Exposing terms and metadata with web 2.0. In *Workshop on Social and Collaborative Construction of Structured Knowledge at WWW 2007*, 2007.
8. U. Hahn and K. Schnattinger. Towards text knowledge engineering. *Hypothesis*, 1:2, 1998.
9. R. Hoehndorf, M. Dumontier, and G. V. Gkoutos. Evaluation of research in biomedical ontologies. *Briefings in Bioinformatics*, 2012.
10. S. I. *Scholastic dictionary of synonyms, antonyms, and homonyms.* 2001.
11. T. Kauppinen, L. Pouchard, and C. Kessler. *Proceedings of the First International Workshop on Linked Science (LISC 2011)*, volume CEUR Workshop Proceedings 783. 2011.
12. B. A. Kipfer. *21st century synonym and antonym finder.* Dell, 1993.
13. C. G. Laird. *Webster's New World Roget's A-Z Thesaurus.* SimonandSchuster. com, 2003.
14. N. LaRoche, J. J. I. Rodale, and L. Urdang. *The Synonym Finder.* Rodale, 1978.
15. D. Lawrie, D. Binkley, and C. Morrell. Normalizing source code vocabulary. In *Reverse Engineering (WCRE), 2010 17th Working Conference on*, pages 3–12. IEEE, 2010.

16. C. Lynnes. Toolmatch. *Proceedings of the ESIP Summer Meeting*, 2012.
17. A. Maedche and S. Staab. Measuring similarity between ontologies. In *Knowledge engineering and knowledge management: Ontologies and the semantic web*, pages 251–263. Springer, 2002.
18. D. Maynard, W. Peters, and Y. Li. Metrics for evaluation of ontology-based information extraction. In *International World Wide Web Conference*, 2006.
19. W. K. Michener, S. Allard, A. Budden, R. B. Cook, K. Douglass, M. Frame, S. Kelling, R. Koskela, C. Tenopir, and D. A. Vieglais. Participatory design of dataoneenabling cyberinfrastructure for the biological and environmental sciences. *Ecological Informatics*, 11:5–15, 2012.
20. G. A. Miller. Wordnet: A lexical database for english. *Communications of the ACM*, 38(11):2, 1995.
21. L. Pouchard, R. Cook, J. Green, G. Palanisamy, and N. Noy. Semantic technologies improving the recall and precision of the mercury metadata search engine. *AGU Fall Meeting Abstracts*, 1:1437, 2011.
22. R. G. Raskin. SWEET 2.1 Ontologies. *AGU Fall Meeting Abstracts*, pages B6+, Dec. 2010.
23. E. Rozell, P. Fox, J. Zheng, and J. Hendler. S2s architecture and faceted browsing applications. In *Proceedings of the 21st international conference companion on World Wide Web*, WWW '12 Companion, pages 413–416, New York, NY, USA, 2012. ACM.
24. A. Spooner. *The Oxford dictionary of synonyms and antonyms*. Oxford University Press, 2007.
25. A. Tripathi and H. A. Babaie. Developing a modular hydrogeology ontology by extending the sweet upper-level ontologies. *Computers & Geosciences*, 34(9):1022–1033, 2008.
26. K. Verspoor, J. Cohn, S. Mniszewski, and C. Joslyn. A categorization approach to automated ontological function annotation. *Protein Science*, 15(6):1544–1549, 2006.
27. N. Wiegand and C. Garca. A task-based ontology approach to automate geospatial data retrieval. *Transactions in GIS*, 11(3):355–376, 2007.
28. L. Yao, A. Divoli, I. Mayzus, J. A. Evans, and A. Rzhetsky. Benchmarking Ontologies: Bigger or Better? *PLoS Comput Biol*, 7(1):e1001055+, Jan. 2011.