CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Introducing Hypertension FACTS: Vital Sign Ontology Annotations in the Florida Annotated Corpus for Translational Science
Authors	Hicks, Amanda;Hogan, William;Pepine, Carl;Boire, Nathan;Herring, Chloe;Seppälä, Selja
Publication date	2018
Original Citation	Hicks, A., Hogan, W. R., Pepine, C., Boire, N., Herring, C. and Seppälä, S. (2018) 'Introducing Hypertension FACT: Vital Sign Ontology Annotations in the Florida Annotated Corpus for Translational Science', Proceedings of the Thirty-First International Florida Artificial Intelligence Research Society Conference (FLAIRS 2018), Florida, USA, May 21-23, 2018, pp. 493-499.
Type of publication	Conference item
Link to publisher's version	https://aaai.org/proceeding/01-all-papers-flairs-2018/
Rights	© 2018, Association for the Advancement of Artificial Intelligence. All rights reserved.
Download date	2026-11-06 20:23:04
Item downloaded from	https://hdl.handle.net/10468/17184

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

The Thirty-First International Florida
Artificial Intelligence Research Society Conference (FLAIRS-31)

Introducing Hypertension FACTS: Vital Sign Ontology Annotations in the Florida Annotated Corpus for Translational Science

Amanda Hicks,[1] William Hogan,[1] Carl Pepine,[1]
Nathan Boire,[1] Chloe Herring,[1] Selja Seppälä[2]
[1]The University of Florida, [2]University College Cork
aehicks@ufl.edu, hoganwr@ufl.edu, Carl.Pepine@medicine.ufl.edu,
n.boire@ufl.edu, chloeeherring@ufl.edu, selja.seppala@ucc.ie

Abstract

We introduce the Florida Annotated Corpus for Translational Science (FACTS), which currently consists of 20 case reports about hypertension that we annotated with Vital Sign Ontology (VSO) classes. We describe the annotation method, the annotation results, interannotator agreement measure, and the availability of the corpus and supporting tools for annotating corpora with OWL ontologies. We also discuss issues and limitations of VSO for annotating vital sign data in case reports.

1 Introduction

Manually annotated corpora are a key component for constructing sophisticated automatic natural language processing (NLP) and natural language understanding (NLU) systems. The sophistication of the available manually annotated gold standard corpora determines the sophistication achievable by such systems (Palmer and Xue 2010). Most semantically annotated corpora contain annotations of all mentions of a concept in the corpus to facilitate named entity recognition. For example, the CRAFT corpus annotates mentions of concepts with an ontology as a first step toward using NLU to construct a formal representation of biomedical literature (Bada, Eckert et al. 2012). A key distinction in ontological knowledge representation is the distinction between instance level facts, e.g., facts about a patient, and general level facts, e.g., facts about the general causes of a disease. Most semantically annotated corpora do not distinguish between these levels.

We are developing the Florida Annotated Corpus for Translational Science (FACTS) to support the automatic construction of formal representations of patient data based on text. Furthermore, we plan to annotate the corpus with multiple and, wherever possible, competing ontologies to comparatively evaluate the usefulness of different ontologies for extracting and representing patient-level information in text. As such, FACTS will achieve a necessary first step toward creating an annotated corpus that can support comparative ontology evaluation. The focus on case reports in particular is a novel approach to extract patient-level data from text and to evaluate the suitability of ontologies for extracting and representing biomedical knowledge of varying degrees of generality. Utilizing PubMed case reports results in a gold standard that is freely available to the scientific community without violating patient confidentiality.

2 Background

Ontologies are formal representations of any of the following: a) things in the world, b) knowledge and concepts relevant to a particular domain, or c) the meanings of terms and data. They contain definitions that aid both human understanding and computational parsing of meaning and have a wide range of applications including data integration through semantic interoperability (Martín, Anguita et al. 2008), decision support, hypothesis generation, NLP (Pyysalo, Ginter et al. 2007, Meystre, Savova et al. 2008, Noy, Shah et al. 2009), and annotation and indexing of biomedical datasets (Shah, Jonquet et al. 2009). As machine-readable representations of knowledge, biomedical ontologies render biomedical knowledge computable, thereby supporting computer-aided discovery of knowledge that would otherwise lie dormant in databases and scientific text (Rubin, Shah et al. 2008, Musen, Noy et al. 2012, Wittkop, TerAvest et al. 2013, Henschel, Anwar et al. 2015).

Copyright © 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

A corpus is a collection of documents that have been prepared for use in some investigation. Corpora are frequently used in linguistics to investigate various aspects of language use. They are also widely used in informatics to extract information from texts. An annotated corpus contains additional information about tokens in the text. Semantic annotations of biomedical text can be used for a variety of tasks. (Jovanović and Bagheri 2017) reviews the benefits of semantic annotations for semantic searches to retrieve documents and other biomedical resources. Semantic annotations can also be used for formally representing the knowledge of a domain, thereby aiding knowledge representation, integration, and discovery (Bada, Eckert et al. 2012).

Corpora can be annotated manually, automatically, or with a combination of manual and automatic methods (semi-automatically). While manual annotations are required for the production and evaluation of automatic NLP and NLU systems, they are labor-intensive and therefore time consuming and expensive to develop. However, automatically processing text relies on the ability to annotate that text (Palmer and Xue 2010). This has led informaticians to develop automatic and semi-automatic systems for annotating corpora (Savova, Masanz et al. 2010, Tseytlin, Mitchell et al. 2016), but these systems still ought to be evaluated against gold standard annotations, which are created manually.

(Bada, Eckert et al. 2012) reviews biomedical NLP tasks that have been improved by the use of gold standard corpora. The manual creation of gold standard annotations represents one of the major bottlenecks for automatically processing and formally representing unstructured patient data.

Semantic annotations of biomedical text can be used for a variety of tasks. (Jovanović and Bagheri 2017) reviews the benefits of semantic annotations for semantic search for retrieving documents and other biomedical resources. Semantic annotations can also be used for formally representing the knowledge of a domain, thereby aiding knowledge representation, integration, and discovery (Bada, Eckert et al. 2012).

Gold standard semantic annotations can be used to train machine learning NLP and NLU systems, as test sets for evaluating NLP and NLU systems, to generate robust lexica and thesauri, and to bootstrap rule-based annotation systems. Semantically annotated gold standard copora have also proven to be useful for comparatively evaluating semantic annotation systems (Jovanović and Bagheri 2017). The manual creation of gold standard annotations represents one of the major bottlenecks for automatically processing and formally representing unstructured patient data.

The Florida Annotated Corpus for Translational Science (FACTS) was created to alleviate the semantic annotation gold standard bottleneck. The first collection of documents in this corpus – Hypertension FACTS – consists of a group of 20 open-access, full-text PubMed case reports about hypertension that we manually annotated using Vital Sign Ontology (VSO) (Goldfain, Smith et al. 2011). We have created a gold standard corpus with VSO annotations as a first step to help solve the problem of extracting and formally representing patient data from unstructured text. The gold standard corpus will facilitate accurate automatic and semi-automatic processing of unstructured patient data and provide a tool for comparatively evaluating the suitability of semantic resources such as (semi-)automatic annotation systems and ontologies for extracting and formally representing patient data from text.

FACTS is primarily developed as a tool for comparatively evaluating the suitability of ontologies to extract and represent patient-level information as it is expressed in text rather than structured databases. If ontologies truly are to aid in integrating and formalizing data and knowledge from heterogeneous sources, the suitability of various ontologies for representing data and knowledge in different formats ought to be evaluated.

As a tool for ontology evaluation, a gold standard corpus of case reports has a two-fold advantage over other corpora. First, unlike clinical notes, a corpus of published case reports can be made freely available under an open license without violating patient confidentiality, thereby facilitating data sharing and replicating experiments. Second, case reports typically contain statements at various levels of generality and various types of modality. They frequently describe the typical progression or causes of a disease or disorder and thereby contain statements at a general level and of scientifically accepted knowledge. They describe an individual patient's clinical phenotype, thereby containing patient-level assertions. They may contain hypotheses concerning mechanisms of action, descriptions of unknown components of mechanisms of action, and suggestions for research and clinical practice. As such, case reports provide a rich test for the suitability of ontologies to support knowledge formalization at a variety of levels.

2.1 Related work

We review other semantically annotated gold standard corpora, grouping them in three categories of relevance to Hypertension FACTS: (1) corpora annotated with OBO Foundry ontologies, (2) corpora consisting of case reports, and (3) corpora about heart disease.

OBO Foundry ontology annotated corpora

A number of gold standard corpora have been semantically annotated with widely used ontologies from the OBO Foundry Ontologies (Smith, Ashburner et al. 2007), and terminologies such as MESH, SNOMED CT (Donnelly 2006) and UMLS (Bodenreider 2004). To date, FACTS has been annotated only with OBO Foundry ontologies, so here we focus on corpora that also use the OBO Foundry ontologies as an annotation schema.

The Colorado Richly Annotated Full Text Corpus CRAFT (Bada, Eckert et al. 2012) consists of 67 full-text biomedical journal articles that are semantically annotated with CHEBI, the Cell Ontology, Entrez Gene, Gene Ontology, NCBI Taxonomy, Protein Ontology, and the Sequence Ontology. The CRAFT gold standard corpus consists of scientific literature largely from microbiology. The annotation task during the construction of the gold standard was to annotate all mentions of all concepts in the relevant ontologies. This is the most common semantic annotation task for gold standards that use ontologies. In contrast, FACTS does not annotate all mentions of a concept in the text, but only those mentions that are at the instance level.

The Human Phenotype Ontology Gold Standard Corpus (Groza, Köhler et al. 2015) consists of 228 abstracts cited by the Online Mendelian Inheritance in Man database. The abstracts are semantically annotated with classes from the Human Phenotype Ontology (HPO) (Köhler, Doelken et al. 2014). Like the CRAFT corpus, the annotation task during the construction of the gold standard was to annotate all mentions of concepts in the HPO regardless of the level of generality, with the caveat that the mention should appear in the text in canonical form. For example, 'short and broad fingers' should be annotated with classes for short fingers and broad fingers, but 'fingers are short and broad' should not be annotated at all.

Case report annotated corpora

The Adverse Drug Events (ADE) corpus (Gurulingappa, Rajput et al. 2012) consists of 3000 MEDLINE case reports and is semantically annotated for all mentions of drugs, adverse effects, dosages and includes relation annotations. However, the annotation schema used in this corpus is not an ontology or terminology, but has been generated specifically for this annotation task and lacks formal semantic representations. It is therefore not suitable for evaluating ontologies or integrating information extracted from the case reports with other knowledge resources.

Semantics in a Clinical Scenario Corpus (Zhang and Patrick 2007) contains 75 case reports manually annotated by a single author for all mentions of the concepts in the annotation schema. Only five clinical scenarios have been manually annotated with an interannotator agreement (IAA) score computed with variable results for different tags (from .02-.91). Like the ADE corpus, this corpus has been annotated with a task-specific annotation schema.

Corpora on hypertension or heart disease

To the best of our knowledge, there are currently no publicly available, semantically annotated corpora that focus specifically on hypertension. *The PhenoCHF corpus* (Alnazzawi, Thompson et al. 2015), however, is the most closely related to the domain of Hypertension FACTS and is annotated with the UMLS Metathesaurus. While the UMLS is not an OBO Foundry ontology, it does provide an annotation schema with a formal semantic structure, so is noteworthy in this context. The PhenoCHF corpus consists of 10 full-text articles and 300 discharge summaries that are semantically annotated for all mentions of phenotypic information related to congestive heart failure on two levels: (1) a semantic type from the UMLS and (2) a more specific term from the UMLS Metathesaurus. The annotations in level two were produced semi-automatically with MetaMap. The corpus also contains relation annotations.

2.2 Resources used

BRAT: The Brat Rapid Annotation Tool (BRAT) is an open-source software application for collaborative text annotation (Stenetorp, Pyysalo et al. 2012). It allows users to manually annotate text with entity labels and to relate annotated text spans to each other with relational annotations. The entities and relations can be freely defined for each annotation task. BRAT also supports associating annotations to external ontologies and databases. Its interface and usage are clear and intuitive, providing a user-friendly tool for the less-technical annotator. The annotation output is stored in standoff annotation files.

The Vital Sign Ontology: The Vital Sign Ontology (VSO) is a realist ontology that includes blood pressure, body temperature, heart rate, and respiration rate (Goldfain, Smith et al. 2011). These qualities are measured and monitored by healthcare professionals during every medical visit as indicators of an individual's current health status and to predict future health risks. In VSO, *blood pressure* is defined as "the pressure exerted by circulating blood on the walls of blood vessels," and is further differentiated into subtypes *systolic blood pressure* and *diastolic blood pressure* based on the phase of the cardiac cycle. *Body temperature* is defined as "the temperature of a part of the human body," and is differentiated by the anatomical location in which the measurement is taken. *Pulse rate* is defined as "the rate at which an artery pulses as blood passes through it." *Respiratory rate* is defined as "the rate at which an organism breathes" and is not further differentiated. VSO fills a gap in OBO ontology coverage of clinical signs, including cardiopulmonary physiology and medical device types. VSO can be paired with disease ontologies to help build associations between a certain vital sign profile and pathological processes in a disease course.

3 Methods

3.1 Corpus compilation

The FACTS corpus currently consists of a set of 20 open-access, full-text PubMed case reports about hypertension. A case report is a short (approximately two to six pages long) medical journal article that provides important clinical information about individual patients who present a notewor-

thy medical condition. Case reports include details about signs and symptoms, diagnosis and treatment, as well as test results.

The corpus was compiled by searching PubMed for case reports on hypertension within the previous five years (2011-2016) with the following query:

> (("hypertension"[MeSH Terms] OR "hypertension"[All Fields]) AND ("antihypertensive agents"[Pharmacological Action] OR "antihypertensive agents"[MeSH Terms] OR ("antihypertensive"[All Fields] AND "agents"[All Fields]) OR "antihypertensive agents"[All Fields] OR ("anti"[All Fields] AND "hypertensive"[All Fields]) OR "anti hypertensive"[All Fields])) AND ("loattrfree full text"[sb] AND "2011/11/01"[PDAT] : "2016/10 /31"[PDAT]) AND Case Reports[ptyp] AND (Case Reports[ptyp] AND "loattrfree full text"[sb] AND "humans"[MeSH Terms] AND English[lang])

At the time of the query, 139 case reports matched these criteria. We stratified these according to race, ethnicity, sex, and age (<18, ≥18). We then selected a stratified random sample of 20 case reports. For each case report, we collected the text of the article (in plain text format) and the corresponding PDF file, both of which were renamed with their eight-digit PMID number as such: PMID_XXXXXXXX. Each text file was duplicated to keep a copy of the raw text and preprocessed to create a new clean copy. The clean copies are used in the annotation tasks. They consist of three sections: (i) the title of the article, (ii) the abstract (when applicable), and (iii) the text of the case report, including the legends of any figures and tables. Sections are separated by a dashed line. All other information, such as authors and references, was removed. The cleaned texts were normalized so that new paragraphs are separated with a new line and then compared to the text in the PDF files to ensure their completeness. Since BRAT stores the text and the annotations in separate files, the final step of the corpus creation process consisted in creating an empty annotation file for each case report in the corpus. These files are named with the corresponding PMID and the '.ann' extension: PMID_XXXXXXXX.ann. BRAT writes the annotations made in the program to these '.ann' files on the fly. The same process was followed for cleaning the test corpus files, which were used for training the annotators, but are not included in the gold standard corpus.

3.2 Creation of annotation guidelines

To facilitate IAA, we developed a set of annotation guidelines that describe how to determine whether a span ought to be annotated and how to determine with which class from the ontology to select for annotation. These guidelines were augmented and refined over time as primary annotators had questions or as sources of interannotator disagreement were identified. The annotation guidelines were made available to the annotation team and were consulted during primary annotations and the annotation harmonization process.

3.3 Creation of Bratify tools

Annotating with an ontology in BRAT requires creating BRAT-formatted configuration files for each ontology. We developed a Python script, Bratify, that takes OWL files as input and produces BRAT-formatted class hierarchy and labels to be added in the BRAT configuration files. Thus, the ontology hierachy and labels can be used for annotation.

Bratify runs a SPARQL query on an OWL file to extract all the child-parent pairs in the ontology. The query retrieves the ID and preferred label for each class. To use Bratify with VSO, we first merged all imports (vso-external.owl and vso-external-derived.owl) specified in the VSO source file (vso.owl, version 2012-04-25) to create a new single OWL file (vso_merged.owl). The program produces two distinct outputs to be copied in the ontology-specific configuration files: (i) one for the '[entities]' section of the annotation.conf file where the types of entities used for annotation are defined, and (ii) the other for the visual.conf file, which defines the terminological variants of each annotation entity to be displayed in the annotation interface.

The first output consists in the labels that will be used by the system for the annotations of entities, which need to be formatted with a restricted set of characters. The program replaces any non-compliant character with an underscore ('_'). For instance, the program gets the ID 'VSO:0000011' from the OWL file and outputs the ID 'VSO_0000011'. For this project, the program does it for all the VSO classes, which are displayed in the output file as a tab-indented hierarchy.

The second output is a list of all the annotation types used internally by the system and specified in the annotation.conf file. The annotation types are followed by any number of labels for display on the user interface, separated by a pipe ('|'). These labels are not subject to any character restrictions and can contain spaces and Unicode characters. For this project, the output copied in the visual.conf file is the BRAT-formatted class labels derived from the class IRIs or rdfs:label annotations followed by the preferred terms associated with each class (e.g., 'VSO_0000011 | cardiac cycle phase').

The program takes as parameters the paths to the OWL input file and the two output files, as well as a SPARQL query adapted to the architecture of the input ontology. The user can also select between two output styles: one in which the system types are terms and another in which they are ontology IDs.

Bratify is available under a GNU General Public License v3.0 at https://github.com/seljaseppala/bratify.

Class label	Class IRI	Number of annotations
blood pressure	VSO_0000004	92
blood pressure measurement datum	VSO_0000005	48
regulation of blood pressure	GO_0008217	21
vital sign	OGMS_0000029	10
systolic blood pressure	VSO_0000038	9
blood pressure measurement process	VSO_0000006	7
measurement datum	IAO_0000109	6
body temperature	VSO_0000007	3
diastolic blood pressure	VSO_0000015	3
pulse rate	VSO_0000029	3
pulse rate measurement datum	VSO_0000030	3
respiratory rate	VSO_0000034	2
respiratory rate measurement datum	VSO_0000035	2
body temperature measurement datum	VSO_0000008	1
noninvasive blood pressure measurement process	VSO_0000024	1

Table 1 Frequency of Vital Sign Ontology class annotations in the Hypertension FACTS corpus

3.4 Annotation process

The annotation procedure consisted of five phases: 1) creating an initial set of annotation guidelines, 2) creating BRAT-formatted configuration files from the VSO OWL file, 3) producing the primary sets of annotations by two primary annotators, 4) harmonizing the annotations, and 5) reviewing harmonized annotations and discussing possible misses.

BRAT was used for all semantic-annotation tasks. There were two primary student annotators (NB & CH), a lead annotator with an ontology background (AH), a linguistics expert (SS), and two medical experts (WRH & CP). SS and WRH also have an ontology background. The first primary annotator was a master's student in the Department of Health Education and Behavior at the University of Florida. The second primary annotator was a current medical student at University of Florida, College of Medicine with a biochemistry background and master's degree in public health. The two primary annotators were each given initial training on the use of BRAT and key ontological concepts and practice annotating a test corpus. Once initial training was completed, the primary annotators worked separately to annotate the corpus. They submitted their annotations to the lead annotator who generated an automatic comparison of the annotations in the form of a diff file using the BRAT diff_and_mark.py script and reviewed the discrepancies in annotations. Weekly meetings were held with the two primary annotators, lead annotator, and linguistics expert to discuss annotations and achieve consensus about discrepancies. The annotation guidelines were iteratively refined in light of the discussions at the weekly meetings to promote consistency of annotations across time. When there were questions specific to the domain of hypertension, WRH and CP were consulted. After the group achieved consensus, the lead annotator then reviewed the annotations once more to ensure nothing was missed by both primary annotators. Two possible missed annotations were reviewed by the annotation team and a consensus achieved.

4. Results

The gold standard corpus consists of 20 case reports for hypertension that were annotated with a total of 211 annotations with classes from the VSO. In total, 15 of the 181 classes in VSO were used in annotations. Table 1 shows the frequency of each class in the gold standard corpus.

We measured IAA among the primary annotators using an f-measure. Cohen's kappa or Krippendorff's alpha typically provides a more accurate IAA measure because, unlike the f-measure, they take into account the probability that the annotators agreed by chance. However, these metrics can be skewed artificially low by rare cases, i.e., by annotations that do not occur frequently in the annotation set (Viera and Garrett 2005). Because our annotations contain a large number of rare cases (cf. Table 1), we use the f-measure to calculate IAA. This also renders our IAA scores comparable to those reported in the construction of the CRAFT corpus (Bada, Eckert et al. 2012). We calculated the IAA score for the corpus as a whole and also for the first and last ten case reports that were annotated, Hypertension 1 and Hypertension 2 respectively.

We consider annotations to be an exact match when they are on the exact same span and use the same class for annotation. We consider annotations to be a partial match when the spans have a tolerance of ± 1 character. This avoids pe-

nalizing annotation errors that may result from difficulty with the user interface, such as dropping the first letter of a word or including a white space at the end of a word. We also calculated the IAA score on classes for only those annotations that had an exact or partial span match. Table 2 shows the IAA scores.

	f-measure exact matches only	f-measure exact and partial matches	f-measure classes on matched spans only
Hypertension 1	0.50	0.54	0.93
Hypertension 2	0.60	0.69	0.87
Full Corpus	0.57	0.60	0.90

Table 2 Interannotator agreement score

The Hypertension FACTS annotations are licensed under the Creative Commons Attribution 4.0 license (CC BY) and available at www.github.com/ufbmi/FACTS_resources.

5. Discussion and future work

The IAA score for the corpus as a whole is low, .60 compared to the CRAFT corpus's score of ~.90. We have identified several reasons for this. First, because one annotator was recruited prior to the other, the annotators did not receive the same amount of practice annotating test corpora. Because the IAA score for Hypertension 1 is significantly lower than the agreement for Hypertension 2 (.54 and .69 respectively), this seems to be a likely cause of disagreement. In our current annotation efforts with the Human Disease Ontology, the gold standard annotations were not commenced until the primary annotators consistently achieved an IAA of ~.70 on the test corpus.

Even taking the discrepancy of practice annotations into account, the IAA score is lower than that achieved in the CRAFT corpus. Our annotation task differs from that used in CRAFT and is more complex. The task for CRAFT annotators is to annotate every mention of a class with the corresponding class in an ontology. In FACTS, the task is to annotate every instance-level mention with a corresponding class in the ontology. This is essentially two tasks: 1) identify spans that denote an individual and 2) annotate those spans with a corresponding class in the ontology. When we calculated the agreement of class annotations for those spans on which both primary annotators agreed, we found that IAA is comparable to that of CRAFT, suggesting that identifying spans that denote individuals is a difficult task. We are currently looking into strategies for improving IAA for task one, including having professional ontologists and/or philosophers with training in analytic metaphysics annotate the corpus for instance-level mentions and then having domain experts annotate the corpus with domain classes from an ontology.

During the annotation process we discovered gaps in the coverage of the VSO. For example, 'monitoring blood pressure' denotes a set, series, or collection of blood pressure measurement processes, but there is no class for a set, series, or collection of such processes. In general, we found that the lack of sets or collections in the VSO resulted in relevant spans of text not being annotated. Also, adding a class for *heart rate mesurement datum* would be useful for capturing the difference between pulse rate and heart rate measures. While (Goldfain, Smith et al. 2011) makes the conceptual distinction between heart rate and pulse rate and draws attention to the fact that this is also a clinical distinction, there is no class *heart rate* in VSO. Furthermore, we found that the phrase 'heart rate' was often used to describe pulse rate and that this is common in clinical practice. One annotator noted that while the pulse-oxometer measures pulse rate, it is often displayed on the monitor as heart rate. The annotations in this corpus are intended to capture actual language use rather than prescribed language use, so there are instances of the phrase 'heart rate' that are annotated with *pulse rate* and instances of 'heart rate' that are not annotated at all.

We are currently annotating Hypertension FACTS with the Human Disease Ontology and anticipate annotating it with more ontologies including the Symptom Ontology, Drug Ontology, Eagle-I Resource Ontology, NCBI Taxonomy, and the Foundational Model Anatomy Ontology. We also plan to annotate the corpus with relations, creating statement-level annotations. Finally, we will convert the annotations to a graph database and evaluate patient-level inferences and queries that leverage knowledge coded in the ontologies to comparatively evaluate the accuracy and coverage of various ontologies.

6. Conclusion

In this paper, we report the results of annotating publicly available hypertension case reports with the Vital Sign Ontology (VSO) to produce the first set of annotations in the Florida Annotated Corpus for Translational Science (FACTS). The same corpus is currently being annotated with classes from the Disease Ontology and will undergo new annotation rounds with classes from the Symptom Ontology, the Drug Ontology, the Eagle-I Resource Ontology, the NCBI Taxonomy, and the Foundational Model of Anatomy ontology. We found that the annotation task for this corpus presents some challenges due to the difficulty in identifying instance-level statements and coverage gaps in the ontology used for the annotations. We are currently developing strategies to improve IAA. We also discovered

areas of domain coverage in the VSO that can be augmented to improve the suitability for the VSO to annotate patient information in text. The gold standard annotations are available under a CC BY 4.0 license at www.github.com/ufbmi/facts_resources.

Acknowledgments

This work was supported in part by the NIH/NCATS Clinical and Translational Science Awards to the University of Florida UL1 TR000064. It was also supported in part by the OneFlorida Clinical Data Research Network, funded by the Patient-Centered Outcomes Research Institute #CDRN-1501-26692. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology, or the OneFlorida Clinical Research Consortium.

References

Aronson, AR. "Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program." *Proceedings of the AMIA Symposium*. 2001:17–21.

Alnazzawi, N., et al. 2015. "Using text mining techniques to extract phenotypic information from the PhenoCHF corpus." *BMC Medical Informatics and Decision Making* 15(Suppl 2): S3-S3.

Bada, M., et al. 2012. "Concept annotation in the CRAFT corpus." *BMC Bioinformatics* 13(1): 1-20.

Bodenreider, O. 2004. "The Unified Medical Language System (UMLS): integrating biomedical terminology." *Nucleic Acids Res* 32(Database issue): D267-270.

Donnelly, K. 2006. "SNOMED-CT: The advanced terminology and coding system for eHealth." *Studies in health technology and informatics* 121: 279.

Goldfain, A., et al. 2011 "Vital Sign Ontology." *Bio-Ontologies 2011*, 71-74.

Groza, T., et al. 2015. "Automatic concept recognition using the Human Phenotype Ontology reference and test suite corpora." *Database (Oxford)*.

Gurulingappa, H., et al. 2012. "Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports." *Journal of Biomedical Informatics* 45(5): 885-892.

Henschel, A., et al. 2015. "Comprehensive meta-analysis of ontology annotated 16S rRNA profiles identifies beta diversity clusters of environmental bacterial communities." *PLoS Computational Biology* 11(10): e1004468.

Jovanović, J. and E. Bagheri 2017. "Semantic annotation in biomedicine: the current landscape." *Journal of Biomedical Semantics* 8(1): 44.

Köhler, S., et al. 2014. "The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data." *Nucleic Acids Res* 42(Database issue): D966-974.

Martín, L., et al. 2008. Ontology based integration of distributed and heterogeneous data sources in ACGT. *Proceedings of the First International Conference on Health Informatics - (Volume 1)*. L. Azevedo and A. Londral. Funchal, Madeira, Portugal: Science and Technology Publications, Lda: 301-306.

Meystre, S. M., et al. 2008. "Extracting information from textual documents in the electronic health record: a review of recent research." *Yearbook of Medical Informatics*: 128-144.

Musen, M. A., et al. 2012. "The National Center for Biomedical Ontology." *Journal of the American Medical Informatics Association : JAMIA* 19(2): 190-195.

Noy, N. F., et al. 2009. "BioPortal: ontologies and integrated data resources at the click of a mouse." *Nucleic Acids Research* 37: W170-173.

Palmer, M. and N. Xue 2010. Linguistic Annotation. *The Handbook of Computational Linguistics and Natural Language Processing*: Wiley-Blackwell: 238-270.

Pyysalo, S., et al. 2007. "BioInfer: a corpus for information extraction in the biomedical domain." *BMC Bioinformatics* 8(1): 50.

Rubin, D. L., et al. 2008. "Biomedical ontologies: a functional perspective." *Briefings in Bioinformatics* 9(1): 75-90.

Savova, G. K., et al. 2010. "Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications." *Journal of the American Medical Informatics Association* 17(5): 507-513.

Shah, N. H., et al. 2009. "Ontology-driven indexing of public datasets for translational bioinformatics." *BMC Bioinformatics* 10(2): S1.

Smith, B., et al. 2007. "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration." *Nature Biotechnology* 25(11): 1251.

Stenetorp, P., et al. 2012. BRAT: a web-based tool for NLP-assisted text annotation. *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*. Avignon, France: Association for Computational Linguistics: 102-107.

Tseytlin, E., et al. 2016. "NOBLE – Flexible concept recognition for large-scale biomedical natural language processing." *BMC Bioinformatics.* 17.

Viera, A. J. and J. M. Garrett 2005. "Understanding interobserver agreement: the kappa statistic." *Fam Med* 37(5): 360-363.

Wittkop, T., et al. 2013. "STOP using just GO: a multi-ontology hypothesis generation tool for high throughput experimentation." *BMC Bioinformatics* 14(1): 53.

Zhang, Y. and J. Patrick 2007. "Extracting semantics in a clinical scenario." Proceedings of the fifth Australasian symposium on ACSW frontiers-Volume 68: Australian Computer Society, Inc.