Improved characterisation of clinical text through ontology-based vocabulary expansion

Luke T Slater[1,2*], William Bradlow[2,1], Simon Ball[2,1], Robert Hoehndorf[7], Georgios V Gkoutos[1,2,3,4,5,6],

1 Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham
2 Institute of Translational Medicine, University Hospitals Birmingham, NHS Foundation Trust
3 NIHR Experimental Cancer Medicine Centre
4 NIHR Surgical Reconstruction and Microbiology Research Centre
5 NIHR Biomedical Research Centre
6 MRC Health Data Research UK (HDR UK) Midlands
7 Computer, Electrical and Mathematical Sciences & Engineering Division, Computational Bioscience Research Center, King Abdullah University of Science and Technology

* l.slater.1@bham.ac.uk

Abstract

Background Biomedical ontologies contain a wealth of metadata that constitutes a fundamental infrastructural resource for text mining. For several reasons, redundancies exist in the ontology ecosystem, which lead to the same concepts being described by several terms in the same or similar contexts across several ontologies. While these terms describe the same concepts, they contain different sets of complementary metadata. Linking these definitions to make use of their combined metadata could lead to improved performance in ontology-based information retrieval, extraction, and analysis tasks.

Results We develop and present an algorithm that expands the set of labels associated with an ontology class using a combination of strict lexical matching and cross-ontology reasoner-enabled equivalency queries. Across all disease terms in the Disease Ontology, the approach found **51,362** additional labels, more than tripling the number defined by the ontology itself. Manual validation by a clinical expert on a random sampling of expanded synonyms over the Human Phenotype Ontology yielded a precision of **0.912**. Furthermore, we found that annotating patient visits in MIMIC-III with an extended set of Disease Ontology labels led to semantic similarity score derived from those labels being a significantly better predictor of matching first diagnosis, with a mean average precision of **0.88** for the unexpanded set of annotations, and **0.913** for the expanded set.

Conclusions Inter-ontology synonym expansion can lead to a vast increase in the scale of vocabulary available for text mining applications. While the accuracy of the extended vocabulary is not perfect, it nevertheless led to a significantly improved ontology-based characterisation of patients from text in one setting. Furthermore, where run-on error is not acceptable, the technique can be used to provide candidate synonyms which can be checked by a domain expert.

Background

Metadata are a fundamental feature of biomedical ontologies, describing a wealth of natural language information in the form of labels and descriptions [1]. OWL ontologies implement metadata in the form of annotation properties, and these can be used to describe multiple natural language labels for a single term. Open Biomedical Ontologies (OBO) [2] define a series of conventional annotation properties that can be used for the expression of labels and synonyms. These features are widely used; as of 2017 the Human Phenotype Ontology (HP) [3] contained 14,328 synonyms for 11,813 classes [4]. Because such labels are associated with ontology terms, ontologies constitute a controlled domain vocabulary.

The role of a controlled domain vocabulary makes ontologies a valuable resource for text mining, particularly in information retrieval and extraction tasks [5]. Furthermore, association of entities described in text with ontologies enables their integration with other datasets annotated by those ontologies, as well as caters the application of ontology-based analysis techniques such as semantic similarity, semantic rule-mining, or relational machine learning.

However, due to limitations of resources for expert curation of ontologies and the sheer scale of their contents, the labels obtainable from single ontologies are not exhaustive. Combined with the tendency for alternative presentation of semantically equivalent concepts in biomedical text [6], ontology labels are not always a good fit for text corpora that discuss the same concepts [7]. By expanding the set of synonyms in an ontology, particularly with synonyms that provide a better fit for text corpora, the performance of natural language processing tasks that depend on them can be improved.

This potential is reflected by previous work in the field. One approach that used analysis of existing synonyms across ontology hierarchy to determine new synonyms reported an increase in performance of a task retrieving articles from a literature repository [8]. Another rule-based synonym expansion approach to extending the Gene Ontology showed improved performance in named entity recognition (NER) tasks [9]. A combined machine-learning and rule-based approach to learning new HP synonyms from manually annotated PubMed abstracts improved performance of an annotation task over a gold standard text corpus [10].

Ontology-based annotation software such as OBO Annotator [11], ConceptMapper [12], and the NCBO Annotator [13] contain routines to consider rule-based morphological and positional transformations of terms to increase NER recall. Parameters that control the use of these features have a strong influence on annotation performance [14]. Previous work has also investigated synonym acquisition and derivation for the purposes of improving the performance of lexical ontology matching and alignment tasks [15]. Outside of automated synonym generation, organised efforts have been made to manually extend an ontology's synonyms for a particular purpose. For example, HP was expanded with layperson synonyms to enable its use in applications that interact directly with patients [16].

However, no work to our knowledge has considered linking different ontology terms for the purposes of vocabulary expansion. Many biomedical concepts are described by the same terms in equivalent or similar contexts across several ontologies. For example, terms describing hypertension exist in many ontologies and medical terminologies. The *hypertension* (HP:0000822) term describes the condition in the context of a phenotype, while *hypertension* (DOID:10763) from the Disease Ontology (DO) [17] describes it in the context of a disease. Specific-disease or application ontologies also extend upon definitions provided by general domain ontologies. For example, the Hypertension Ontology (HTN) [18] extends the HP and DO hypertension classes, adding additional information including labels. Furthermore, the subtle distinctions between concepts that biomedical ontologies capture, including phenotype versus disease, do not necessarily influence many of the commonly applied text mining tasks, because these contexts share the same labels.

We hypothesise that because ontologies are constructed with different loci, ontologies that define terms describing the same concepts will contain different, but valid, synonyms for a particular context. These loci consist in contexts, domain experts, and source material. By considering all of these terms, we can construct extended vocabularies that may improve the power of ontology-based

text mining tasks. In this paper, we describe and implement a synonym expansion approach that combines lexical matching and semantic equivalency to obtain new synonyms for biomedical concepts. We use the approach to extend several ontology vocabularies, and evaluate them both manually, and in an ontology-based patient characterisation task.

Results

The synonym expansion algorithm is available as part of the Komenti text mining framework, which is available under an open source licence at `https://github.com/reality/komenti`, while the files used for validation are available at `https://github.com/reality/synonym_expansion_validation`.

Algorithm

We developed a synonym expansion algorithm that derives additional synonyms for a class by matching it with classes from other ontologies, making use of the AberOWL ontology reasoning framework [19]. The algorithm performs the following process, for each class provided as input (in this context, 'every ontology' is any of the ontologies that are included in AberOWL):

1. Extract the labels and synonyms of any classes in any ontology with a label or synonym that exactly matches the first label of the input class.

2. Run an equivalency query against every ontology using the Internationalised Resource Identifier (IRI) of the input class, extracting labels and synonyms for any classes returned.

3. Of the candidate synonyms produced by the first two steps, discard any that were:

   - Defined in ontologies that were found to produce incorrect synonyms.
   - Have the form of a term identifier.
   - Contain the input class label as a substring.

The algorithm uses two different methods for identifying matching classes. Strict lexical matching is used to identify otherwise unlinked terms that contain a label which is the same as the first label of the input class. Mapping terms across ontologies via shared labels or metadata is a well established technique used in ontology alignment [20]. Cross-ontology equivalency queries are used to obtain additional classes in the case that ontologies describe semantically equivalent classes, but do not share the same label. This can occur due to inferred semantic equivalency, ontologies becoming out of sync, or omission of annotation properties in a referencing class. Only the first label for the input class is used, because additional labels and synonyms were found to less uniquely identify the class in question, and lead to more incorrect candidate synonyms.

After the main matching stage, the set of labels is pruned down to remove incorrect values. Some ontologies include term identifiers as labels which cannot be exploited by text-mining applications. Therefore, candidate synonyms that contained a colon or underscore were removed. The algorithm also removes labels sourced from GO-PLUS [21], MONDO [22], CCONT [23], and phenX [24], because we found these ontologies consistently produced incorrect synonyms. We also removed labels that include the input label as a substring, as these add no value to NER systems (as the smaller string would match, making the longer string redundant).

Ontology Expansion

We applied the vocabulary expansion algorithm to all 9,908 subclasses of *disease* (DOID:4) in the Disease Ontology (DO). DO itself asserts 24,878 labels and synonyms for these classes. The expanded DO vocabulary contained 76,240 labels and synonyms. We also applied the algorithm to

the 14,406 non-obsolete subclasses of *Phenotypic abnormality* (HP:0000118) in HP. HP itself asserts 95
29,805 labels and synonyms. The number of labels and synonyms following expansion was 54,765. 96
Therefore, the algorithm found 24,960 additional synonyms across HP. 97

For the DO term *hypertension* (DOID:10763), 28 labels and synonyms were found. 3 of these 98
were from DO itself. The algorithm found 70 synonyms not including the word 'hypertension'. Of 99
these, 56 were obtained via lexical matching, and 14 by equivalency query. The sources of these 100
synonyms is summarised in Table 1. After making the list unique, there were 28 labels and 101
synonyms. Therefore, the algorithm found 25 new synonyms. 102

Table 1. Source of the 70 non-unique synonyms found for the term *hypertension* (DOID:10763) per-ontology. Of these synonyms, 28 were unique. Bracketed numbers, where given, are the labels found by the equivalency method only (in these cases, lexical matches were made for multiple classes in the ontology).

Ontology	Source	Number of Synonyms
GWAS_EFO_SKOS	Lexical	16
MESH	Lexical	4
HTN [18]	Lexical, Equivalency	3 (2)
CRISP	Lexical	1
CCTOO [25]	Lexical	6
ONTONEO	Lexical, Equivalency	4 (3)
NCIT [26]	Lexical	6
COSTART	Lexical	7
BAO	Lexical, Equivalency	2
CSSO	Lexical	2
ODAE	Lexical, Equivalency	2
DO	–	3
DTO	Equivalency	2

In this example, there were no synonyms uniquely found via equivalency, however, if we use 103
bradycardia as the input class, we can identify two new synonyms from PhenomeNET [27], 104
bradyrhythmia and *reduced heart rate*, which were not otherwise obtained via lexical matching. This 105
is because PhenomeNET establishes a semantic equivalency between *decreased heart rate* 106
(MP:0005333), which does not share its first label with the HP class. 107

Manual Validation 108

To evaluate the correctness of synonyms in the expansion of HP, a clinical expert manually validated 109
the correctness of 866 synonyms found for 500 randomly selected terms. Table 2 summarises the 110
results, which show a precision of 0.912. 195 terms were marked as ambiguous, in the case that the 111
synonyms were in a foreign language or the clinician did not have enough expertise of the term to 112
determine whether the synonym was correct. 113

Table 2. Metrics for clinical expert validation of 866 generated synonyms for 500 terms. Synonyms already included in HP were not included in the validation. Synonyms were marked ambiguous if not English, or if the validator did not have enough expertise to confidently judge it.

Terms	Total Synonyms	TP	FP	Ambiguous	Precision
500	866	613	59	195	0.912

Information Retrieval 114

To evaluate whether the extended vocabularies could lead to more to greater performance in 115
information retrieval tasks, we compared the number of MIMIC-III annotations and MEDLINE 116

results returned for all non-obsolete subclasses of *Abnormality of the cardiovascular system* [117] (HP:0011025). HP asserts 2,205 labels and synonyms for these classes, while the expanded set of [118] labels numbers 5,336. The results are summarised in Table 3. [119]

Table 3. Amount of labels for *Abnormality of the cardiovascular system* (HP:0011025) before and after synonym expansion, and results of the two text mining tasks using them as vocabularies. MEDLINE results are the sum of the number of results returned by each query.

Vocabulary	Labels	MIMIC-III Annotations	MEDLINE Results
HP Labels	2,205	1,104	8,191,564
Expanded HP Labels	5,336	1,447	13,513,342

Patient Characterisation [120]

To evaluate the value of the DO expansion, and to identify whether the additional annotations were [121] informative for the purposes of characterising entities described by biomedical text, we investigated [122] its application within the MIMIC-III dataset [28]. We annotated a sample of 1,000 patient visits [123] using classes from the Disease Ontology (DO) that contained cross-references to ICD-9, both before [124] and after label expansion using the presented algorithm. We then used those annotations to [125] calculate a measure of semantic similarity between the patient visits, and evaluated the rankings [126] with respect to the ICD-9 codes they were annotated with. We used the mean reciprocal rank and [127] the mean average precision to measure how well rankings predicted matching first diagnoses. The [128] results of the ranking task are shown in Table 4, with the expanded vocabulary leading to an [129] increased performance in both cases. To determine whether the result was significantly different, we [130] used the Wilcoxon rank-sum test to compare the ranks of patients with matching first diagnoses, [131] yielding a p-value of 0.0007063. [132]

Table 4. Comparison of the annotations of texts for 1,000 randomly sampled MIMIC-III patient visits before and after expansion, and their associated performance with respect to how predictive semantic similarity scores calculated from the annotations were of shared first diagnosis.

Investigation	Annotations	MAP	MRR
Unexpanded	1,380,216	0.88	0.947
Expanded	2,088,765	0.913	0.986

Discussion [133]

The results clearly demonstrate that for two biomedical ontologies, our approach vastly increases the [134] amount of labels and synonyms available for their terms. Using hypertension as an example, we [135] demonstrated that a range of different ontologies contribute additional synonyms, leading to 25 new [136] unique labels for the term. By leveraging these we can effectively enrich vocabularies for terms. [137]

While we only manually validated a small subset of terms from HP, this indicated a fairly high [138] precision for candidate terms. Through analysis of the false positives, we found that many of them [139] were caused by errors in the ontologies that the synonyms were sourced from. For example, several [140] synonyms for *motor aphasia* (HP:0002427) were marked as incorrect since they refer to dysphasia, [141] including "Broca Dysphasia." Aphasia and dysphasia are different conditions. The first refers to a [142] partial loss of language, and the latter to a full loss of language. All of these incorrect synonyms [143] were sourced from *Aphasia, Broca* (MESH:D001039) in MESH. [144]

Though this is not reflected in the results, we also found during the development of the algorithm [145] that certain ontologies produced consistently incorrect synonyms. Several of these ontologies are [146] meta-ontologies, automatically constructed from several ontologies using alignment and integration [147] methods, and it's possible that errors in that process were the cause of the incorrect synonyms. [148]

Certain annotation properties were also incorrectly detailed by the AberOWL API as being labels, such as *europe pmc* and *kegg compound*. Candidate synonyms defined by problematic ontologies or matching the list of annotation properties are automatically removed. Expansion of the list of ontologies discluded from the sources for labels might further improve the precision of the algorithm, but may potentially come at the cost of correct synonyms.

Furthermore, the manual validation revealed that many of the returned synonyms were in non-English languages. While OWL ontologies do allow for parameters that distinguish which language the property is in, AberOWL does not index them. Therefore, it is not currently possible to distinguish between English and non-English synonyms. These items were marked as ambiguous, and not counted in the overall precision. This could also be controlled partially by discluding additional ontologies from results. For example, WHOFRE is a non-ontology mapping of French vocabulary to UMLS.

For any uses, where a reduced vocabulary accuracy is not acceptable, the algorithm should be used as a candidate label generator, to be checked by a domain expert before further use.

We also demonstrated that that our expansion of the HP vocabulary increases the amount of data returned by two information retrieval tasks, for a subset of cardiovascular terms. We have not, however, shown in the information retrieval experiment whether the extra information returned is useful or relevant. We can assume, based on our manual validation results that some of the additional data returned are incorrect, though most should be correct.

Finally, our experimental validation showed a clear and significant increase for a patient stratification task over MIMIC-III. This indicates that for certain tasks, our approach can increase the quality of entity characterisations gained by information extraction, and in turn the power of ontology-based analyses, even without manual validation of the produced labels.

Limitations and Future Work

The most important potential limitation of the algorithm itself is that it violates the notion that the IRI of a concept uniquely identifies it, rather than its name. This is due to the fact that OWL ontologies do not follow the unique name assumption. False positives, in theory, could be generated by a lexical match on a homonym, which then has different synonyms itself. We believe, however, that this effect should be limited in the case of a highly specific biomedical language. Furthermore, any such error would be most likely be mitigated by the dataset context limitation. For example, synonyms derived from different contexts, incorrectly associated with a medical concept, are unlikely to be present within clinical letters. We further limit this effect by only using the first label, reasoning that the most unique synonym will be included as a synonym.

False synonyms could also be removed on the basis of a corpus search. For example, if a candidate synonym never, or, at least, rarely, appears in the same document as another label, used for this term across a literature corpus, it's possible that it refers to a different concept from a disjoint context. This could also be performed by analysing the metadata of text corpora. For example, if two terms are never, or, at least, rarely, associated with literature from the same journals, the same field, or the same content tags, it's possible they have different meanings. In a further study, we would investigate whether synonymy can be identified using word embeddings.

While equivalency returns fewer synonyms, and not necessarily many that are unique from the ones gained by lexical matching, they can also be treated with a higher level of confidence. For this reason, using only this method could be considered as a parameter in the case that a higher accuracy is required.

Conclusions

We have demonstrated that an inter-ontology approach to vocabulary expansion is a powerful method for adding informative labels and synonyms to terms used in text mining. These synonyms are found with a fairly high precision, and lead to a greater rate of document retrieval in clinical and

literature settings. Most importantly, we have shown that the approach improves the power of an   197
ontology-based characterisation and analysis of patients via clinical text.   198

Methods   199

All files described in the validation (excluding the MIMIC-III data files), along with the commands   200
necessary to repeat the experiments are available at   201

    `https://github.com/reality/synonym_expansion_validation/`.   202

Algorithm   203

We implemented the algorithm as a module in the Komenti semantic text mining framework using   204
the Groovy programming language [29]. It makes use of the AberOWL API [19] for label matching   205
and semantic queries, documented at `http://www.aber-owl.net/docs/`.   206

    OWL ontologies use a number of conventional annotation properties to define labels and   207
synonyms. These span a range of confidence and degree of synonymy. In this paper, we consider   208
frequently used annotation properties, summarised in Table 5. These are the annotation properties   209
consolidated into the 'synonym' property by the AberOWL API. Another oboInOwl synonym,   210
hasRelatedSynonym is excluded, because the labels provided by these synonyms are too imprecise.   211

Table 5. Summary of conventionally used annotation properties considered in this experiment. Definitions come from the description of the annotation properties in their respective top-level ontologies.

Annotation Property	Identifier	Definition
label	rdfs:label	"a human-readable version of a resource's name [30]."
altLabel	skos:core#altLabel	"An alternative lexical label for a resource [31]."
has_exact_synonym	hasExactSynonym	"An alias in which the alias exhibits true synonymy [32]."
has_narrow_synonym	hasNarrowSynonym	"An alias in which the alias is narrower than the primary class name. Example: pyrimidine-dimer repair by photolyase is a narrow synonym of photoreactive repair [32]."
has_broad_synonym	hasBroadSynonym	"An alias in which the alias is broader than the primary class name. Example: cell division is a broad synonym of cytokinesis [32]."
alternative term	IAO_0000118	"An alternative name for a class or property which means the same thing as the preferred name (semantically equivalent) [33]."

Manual Validation   212

To evaluate the performance of the algorithm, we randomly selected 500 classes from the expanded   213
version of HP for manual validation. Synonyms already asserted by HP were removed from the set,   214
because they were already assumed to be correct, and would not contribute to measuring the   215
performance of the synonym expansion algorithm. A clinical expert (WB) marked each synonym as   216
correct, incorrect, or ambiguous. The expert was asked to answer correctly or incorrectly on the   217

basis: "if a patient has *synonym*, would it also be true that they have *original label*?" Entries were ₂₁₈
marked as ambiguous if the synonym was in a different language, or the validator otherwise did not ₂₁₉
have enough knowledge of the phenotype to determine whether or not the synonym was correct. ₂₂₀

Information Retrieval ₂₂₁

We used the Komenti semantic text mining framework, which implements Stanford CoreNLP's ₂₂₂
RegexNER [34] to annotate 1,000 randomly sampled entries from the NOTEEVENTS table in ₂₂₃
MIMIC-III (MIMIC) [35]. MIMIC is a freely available healthcare database, containing a variety of ₂₂₄
structured and unstructured information concerning around 60,000 admissions to critical care ₂₂₅
services [28]. We annotated the sample with all subclasses of *Abnormality of the cardiovascular* ₂₂₆
system (HP:0011025), comparing the number of annotations before and after synonym expansion. ₂₂₇
This investigation was performed on 17/01/2020. ₂₂₈

Using the same set of subclasses of *Abnormality of the cardiovascular system* (HP:0011025), we ₂₂₉
compared the sum of article counts returned for a disjunctive query of all labels and synonyms for ₂₃₀
each term, before and after synonym expansion. MEDLINE is a searchable database of literature ₂₃₁
metadata in the life sciences, containing more than 25 million article references [36]. MEDLINE was ₂₃₂
queried on the 27/01/2020. ₂₃₃

Patient Characterisation ₂₃₄

We sampled 1,000 patient visits from the MIMIC-III (distinct from those used in the information ₂₃₅
retrieval experiment). We then concatenated all text records for each patient visit from the ₂₃₆
NOTEEVENTS table into one text file, and pre-processed the text to remove newlines, improve ₂₃₇
sentence delineation, and lemmatise words. We also retained the primary diagnosis, which was the ₂₃₈
first listed ICD-9 code in the DIAGNOSES_ICD table. These codes are produced by clinical coding ₂₃₉
specialists, by examining the texts associated with the visit. ₂₄₀

We limited the classes considered for our annotation vocabulary to those which DO contained a ₂₄₁
database cross-reference to ICD-9, of which there were 2,118. This was to reduce noise from terms ₂₄₂
not represented in ICD-9. We obtained the unexpanded and expanded synonyms for these terms on ₂₄₃
08/07/2020. Both sets of labels were also lemmatised (both lemmatised and unlemmatised forms ₂₄₄
were used for annotation). ₂₄₅

The Komenti semantic text-mining framework was used to annotate the text associated with ₂₄₆
each patient visit. As before, this made use of the CoreNLP RegexNER annotator [34]. Negated ₂₄₇
annotations were excluded using the komenti-negation algorithm [37]. We then used the set of terms ₂₄₈
associated with it to produce a semantic similarity matrix for patient visits, using the Resnik ₂₄₉
measure of pairwise similarity for each annotated term [38], normalised into a groupwise measure ₂₅₀
using the best match average method [39]. Information content was calculated using the probability ₂₅₁
of the term appearing as an annotation in the totality of the set of annotations [38]. The similarity ₂₅₂
matrix was computed using the Semantic Measures Library [40]. ₂₅₃

We evaluated the similarity matrix using mean reciprocal rank and mean average precision to ₂₅₄
measure performance in predicting shared primary patient diagnosis. A true case was considered to ₂₅₅
be whether a pair of patient visits had the same primary diagnosis (as per the MIMIC-III database). ₂₅₆
For mean average precision, we considered only the 10 most similar patients for each patient. The ₂₅₇
p-value was calculated using the built-in *wilcoxon.test* function of R version 3.4.4 [41]. ₂₅₈

Competing interests ₂₅₉

The authors declare that they have no competing interests. ₂₆₀

Author's contributions   261

Acknowledgements   267

References

1. Hoehndorf, R., Dumontier, M., Gkoutos, G.V.: Evaluation of research in biomedical ontologies. Briefings in Bioinformatics **14**(6), 696–712 (2013). doi:10.1093/bib/bbs053

2. The OBI Consortium, Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R.H., Shah, N., Whetzel, P.L., Lewis, S.: The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology **25**(11), 1251–1255 (2007). doi:10.1038/nbt1346

3. Köhler, S., Doelken, S.C., Mungall, C.J., Bauer, S., Firth, H.V., Bailleul-Forestier, I., Black, G.C.M., Brown, D.L., Brudno, M., Campbell, J., FitzPatrick, D.R., Eppig, J.T., Jackson, A.P., Freson, K., Girdea, M., Helbig, I., Hurst, J.A., Jähn, J., Jackson, L.G., Kelly, A.M., Ledbetter, D.H., Mansour, S., Martin, C.L., Moss, C., Mumford, A., Ouwehand, W.H., Park, S.-M., Riggs, E.R., Scott, R.H., Sisodiya, S., Vooren, S.V., Wapner, R.J., Wilkie, A.O.M., Wright, C.F., Vulto-van Silfhout, A.T., de Leeuw, N., de Vries, B.B.A., Washingthon, N.L., Smith, C.L., Westerfield, M., Schofield, P., Ruef, B.J., Gkoutos, G.V., Haendel, M., Smedley, D., Lewis, S.E., Robinson, P.N.: The Human Phenotype Ontology project: Linking molecular biology and disease through phenotype data. Nucleic Acids Research **42**(Database issue), 966–974 (2014). doi:10.1093/nar/gkt1026

4. Köhler, S., Vasilevsky, N.A., Engelstad, M., Foster, E., McMurry, J., Aymé, S., Baynam, G., Bello, S.M., Boerkoel, C.F., Boycott, K.M., Brudno, M., Buske, O.J., Chinnery, P.F., Cipriani, V., Connell, L.E., Dawkins, H.J.S., DeMare, L.E., Devereau, A.D., de Vries, B.B.A., Firth, H.V., Freson, K., Greene, D., Hamosh, A., Helbig, I., Hum, C., Jähn, J.A., James, R., Krause, R., F. Laulederkind, S.J., Lochmüller, H., Lyon, G.J., Ogishima, S., Olry, A., Ouwehand, W.H., Pontikos, N., Rath, A., Schaefer, F., Scott, R.H., Segal, M., Sergouniotis, P.I., Sever, R., Smith, C.L., Straub, V., Thompson, R., Turner, C., Turro, E., Veltman, M.W.M., Vulliamy, T., Yu, J., von Ziegenweidt, J., Zankl, A., Züchner, S., Zemojtel, T., Jacobsen, J.O.B., Groza, T., Smedley, D., Mungall, C.J., Haendel, M., Robinson, P.N.: The Human Phenotype Ontology in 2017. Nucleic Acids Research **45**(D1), 865–876 (2017). doi:10.1093/nar/gkw1039

5. Spasic, I., Ananiadou, S., McNaught, J., Kumar, A.: Text mining and ontologies in biomedicine: Making sense of raw text. Briefings in Bioinformatics **6**(3), 239–251 (2005). doi:10.1093/bib/6.3.239

6. Cohen, K.B., Palmer, M., Hunter, L.: Nominalization and Alternations in Biomedical Language. PLOS ONE **3**(9), 3158 (2008). doi:10.1371/journal.pone.0003158

7. Brewster, C., Alani, H., Dasmahapatra, S., Wilks, Y.: Data Driven Ontology Evaluation. In: International Conference on Language Resources and Evaluation (30/05/04) (2004)

8. Taboada, M., Rodriguez, H., Gudivada, R.C., Martinez, D.: A new synonym-substitution method to enrich the human phenotype ontology. BMC Bioinformatics **18** (2017). doi:10.1186/s12859-017-1858-7

9. Funk, C.S., Cohen, K.B., Hunter, L.E., Verspoor, K.M.: Gene Ontology synonym generation rules lead to increased performance in biomedical concept recognition. Journal of Biomedical Semantics **7**(1), 52 (2016). doi:10.1186/s13326-016-0096-7

10. Lobo, M., Lamurias, A., Couto, F.M.: Identifying Human Phenotype Terms by Combining Machine Learning and Validation Rules. https://new.hindawi.com/journals/bmri/2017/8565739/ (2017). doi:10.1155/2017/8565739

11. Groza, T., Köhler, S., Doelken, S., Collier, N., Oellrich, A., Smedley, D., Couto, F.M., Baynam, G., Zankl, A., Robinson, P.N.: Automatic concept recognition using the Human Phenotype Ontology reference and test suite corpora. Database: The Journal of Biological Databases and Curation **2015** (2015). doi:10.1093/database/bav005

12. Tanenblatt, M., Coden, A., Sominsky, I.: The ConceptMapper Approach to Named Entity Recognition, 6

13. Jonquet, C.: NCBO Annotator: Semantic Annotation of Biomedical Data, 3

14. Funk, C., Baumgartner, W., Garcia, B., Roeder, C., Bada, M., Cohen, K.B., Hunter, L.E., Verspoor, K.: Large-scale biomedical concept recognition: An evaluation of current automatic annotators and their parameters. BMC Bioinformatics **15**, 59 (2014). doi:10.1186/1471-2105-15-59

15. Pesquita, C., Faria, D., Stroe, C., Santos, E., Cruz, I.F., Couto, F.M.: What's in a 'nym'? Synonyms in Biomedical Ontology Matching. In: Alani, H., Kagal, L., Fokoue, A., Groth, P., Biemann, C., Parreira, J.X., Aroyo, L., Noy, N., Welty, C., Janowicz, K. (eds.) The Semantic Web – ISWC 2013. Lecture Notes in Computer Science, pp. 526–541. Springer, Berlin, Heidelberg (2013). doi:10.1007/978-3-642-41335-3_33

16. Köhler, S., Carmody, L., Vasilevsky, N., Jacobsen, J.O.B., Danis, D., Gourdine, J.-P., Gargano, M., Harris, N.L., Matentzoglu, N., McMurry, J.A., Osumi-Sutherland, D., Cipriani, V., Balhoff, J.P., Conlin, T., Blau, H., Baynam, G., Palmer, R., Gratian, D., Dawkins, H., Segal, M., Jansen, A.C., Muaz, A., Chang, W.H., Bergerson, J., Laulederkind, S.J.F., Yüksel, Z., Beltran, S., Freeman, A.F., Sergouniotis, P.I., Durkin, D., Storm, A.L., Hanauer, M., Brudno, M., Bello, S.M., Sincan, M., Rageth, K., Wheeler, M.T., Oegema, R., Lourghi, H., Della Rocca, M.G., Thompson, R., Castellanos, F., Priest, J., Cunningham-Rundles, C., Hegde, A., Lovering, R.C., Hajek, C., Olry, A., Notarangelo, L., Similuk, M., Zhang, X.A., Gómez-Andrés, D., Lochmüller, H., Dollfus, H., Rosenzweig, S., Marwaha, S., Rath, A., Sullivan, K., Smith, C., Milner, J.D., Leroux, D., Boerkoel, C.F., Klion, A., Carter, M.C., Groza, T., Smedley, D., Haendel, M.A., Mungall, C., Robinson, P.N.: Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Research **47**(D1), 1018–1027 (2019). doi:10.1093/nar/gky1105

17. Schriml, L.M., Arze, C., Nadendla, S., Chang, Y.-W.W., Mazaitis, M., Felix, V., Feng, G., Kibbe, W.A.: Disease Ontology: A backbone for disease semantic integration. Nucleic Acids Research **40**(D1), 940–946 (2012). doi:10.1093/nar/gkr972

18. Hicks, A., Miller, M.A., Stoeckert, C., Mowery, D.: The Hypertension Ontology (2019). doi:10.5281/zenodo.2605329

19. Hoehndorf, R., Slater, L., Schofield, P.N., Gkoutos, G.V.: Aber-OWL: A framework for ontology-based data access in biology. BMC Bioinformatics **16**(1), 26 (2015). doi:10.1186/s12859-015-0456-9

20. Kalfoglou, Y., Schorlemmer, M.: Ontology Mapping: The State of the Art. /core/journals/knowledge-engineering-review/article/ontology-mapping-the-state-of-the-art/9A273424C3873243A0DC50FD43C6AD4E (2003). doi:10.1017/S0269888903000651

21. Hill, D.P., Adams, N., Bada, M., Batchelor, C., Berardini, T.Z., Dietze, H., Drabkin, H.J., Ennis, M., Foulger, R.E., Harris, M.A., Hastings, J., Kale, N.S., de Matos, P., Mungall, C.J., Owen, G., Roncaglia, P., Steinbeck, C., Turner, S., Lomax, J.: Dovetailing biology and chemistry: Integrating the Gene Ontology with the ChEBI chemical ontology. BMC Genomics **14**(1), 513 (2013). doi:10.1186/1471-2164-14-513

22. Shefchek, K.A., Harris, N.L., Gargano, M., Matentzoglu, N., Unni, D., Brush, M., Keith, D., Conlin, T., Vasilevsky, N., Zhang, X.A., Balhoff, J.P., Babb, L., Bello, S.M., Blau, H., Bradford, Y., Carbon, S., Carmody, L., Chan, L.E., Cipriani, V., Cuzick, A., Rocca, M.D., Dunn, N., Essaid, S., Fey, P., Grove, C., Gourdine, J.-P., Hamosh, A., Harris, M., Helbig, I., Hoatlin, M., Joachimiak, M., Jupp, S., Lett, K.B., Lewis, S.E., McNamara, C., Pendlington, Z.M., Pilgrim, C., Putman, T., Ravanmehr, V., Reese, J., Riggs, E., Robb, S., Roncaglia, P., Seager, J., Segerdell, E., Similuk, M., Storm, A.L., Thaxon, C., Thessen, A., Jacobsen, J.O.B., McMurry, J.A., Groza, T., Köhler, S., Smedley, D., Robinson, P.N., Mungall, C.J., Haendel, M.A., Munoz-Torres, M.C., Osumi-Sutherland, D.: The Monarch Initiative in 2019: An integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic Acids Research **48**(D1), 704–715 (2020). doi:10.1093/nar/gkz997

23. Ganzinger, M., He, S., Breuhahn, K., Knaup, P.: On the ontology based representation of cell lines. PloS One **7**(11), 48584 (2012). doi:10.1371/journal.pone.0048584

24. Hamilton, C.M., Strader, L.C., Pratt, J.G., Maiese, D., Hendershot, T., Kwok, R.K., Hammond, J.A., Huggins, W., Jackman, D., Pan, H., Nettles, D.S., Beaty, T.H., Farrer, L.A., Kraft, P., Marazita, M.L., Ordovas, J.M., Pato, C.N., Spitz, M.R., Wagener, D., Williams, M., Junkins, H.A., Harlan, W.R., Ramos, E.M., Haines, J.: The PhenX Toolkit: Get the Most From Your Measures. American Journal of Epidemiology **174**(3), 253–260 (2011). doi:10.1093/aje/kwr193

25. Lin, F.P.-Y., Groza, T., Kocbek, S., Antezana, E., Epstein, R.J.: The Cancer Care Treatment Outcomes Ontology (CCTO): A computable ontology for profiling treatment outcomes of patients with solid tumors. Journal of Clinical Oncology **35**(15_suppl), 18137–18137 (2017). doi:10.1200/JCO.2017.35.15$_s$uppl.e18137

26. Sioutos, N., de Coronado, S., Haber, M.W., Hartel, F.W., Shaiu, W.-L., Wright, L.W.: NCI Thesaurus: A semantic model integrating cancer-related clinical and molecular information. Journal of Biomedical Informatics **40**(1), 30–43 (2007). doi:10.1016/j.jbi.2006.02.013

27. Hoehndorf, R., Schofield, P.N., Gkoutos, G.V.: PhenomeNET: A whole-phenome approach to disease gene discovery. Nucleic Acids Research **39**(18), 119–119 (2011). doi:10.1093/nar/gkr538

28. Johnson, A.E.W., Pollard, T.J., Shen, L., Lehman, L.-w.H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L.A., Mark, R.G.: MIMIC-III, a freely accessible critical care database. Scientific Data **3**(1), 1–9 (2016). doi:10.1038/sdata.2016.35

29. The Apache Groovy Programming Language. http://groovy-lang.org/

30. RDF Schema 1.1. https://www.w3.org/TR/rdf-schema/#ch_label

31. SKOS Core Vocabulary Specification. https://www.w3.org/TR/swbp-skos-core-spec/#altLabel

32. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G.: Gene Ontology: Tool for the unification of biology. Nature Genetics **25**(1), 25–29 (2000). doi:10.1038/75556

33. Information-Artifact-Ontology/IAO. IAO (2019)

34. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J.R., Bethard, S., McClosky, D.: The Stanford CoreNLP Natural Language Processing Toolkit. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55–60 (2014). doi:10.3115/v1/P14-5010

35. Gkoutos, G.V., Schofield, P.N., Hoehndorf, R.: The neurobehavior ontology: An ontology for annotation and integration of behavior and behavioral phenotypes. In: International Review of Neurobiology vol. 103, pp. 69–87. Elsevier, ??? (2012)

36. MEDLINE®: Description of the Database. https://www.nlm.nih.gov/bsd/medline.html

37. Slater, L.T., Bradlow, W., Motti, D.F., Hoehndorf, R., Ball, S., Gkoutos, G.V.: A fast, accurate, and generalisable heuristic-based negation detection algorithm for clinical text. bioRxiv, 2020–0703187054 (2020). Chap. New Results. doi:10.1101/2020.07.03.187054

38. Resnik, P.: Using Information Content to Evaluate Semantic Similarity in a Taxonomy. arXiv:cmp-lg/9511007 (1995). cmp-lg/9511007

39. Wang, J.Z., Du, Z., Payattakool, R., Yu, P.S., Chen, C.-F.: A new method to measure the semantic similarity of GO terms. Bioinformatics **23**(10), 1274–1281 (2007). doi:10.1093/bioinformatics/btm087

40. Harispe, S., Ranwez, S., Janaqi, S., Montmain, J.: The semantic measures library and toolkit: Fast computation of semantic similarity and relatedness using biomedical ontologies. Bioinformatics **30**(5), 740–742 (2014). doi:10.1093/bioinformatics/btt581

41. Ihaka, R., Gentleman, R.: R: A language for data analysis and graphics. Journal of computational and graphical statistics **5**(3), 299–314 (1996)