Measuring and Avoiding Information Loss During Concept Import from a Source to a Target Ontology

James Geller[1][a], Shmuel T. Klein[2][b] and Vipina Kuttichi Keloth[1][c]
[1]*Dept. of Computer Science, New Jersey Institute of Technology, U.S.A.*
[2]*Dept. of Computer Science, Bar Ilan University, Ramat Gan 52900, Israel*

Keywords: Biomedical Ontologies, Concept Import, Information Content, Information Loss.

Abstract: Comparing pairs of ontologies in the same biomedical content domain often uncovers surprising differences. In many cases these differences can be characterized as "density differences," where one ontology describes the content domain with more concepts in a more detailed manner. Using the Unified Medical Language System across pairs of ontologies contained in it, these differences can be precisely observed and used as the basis for importing concepts from the ontology of higher density into the ontology of lower density. However, such an import can lead to an intuitive loss of information that is hard to formalize. This paper proposes an approach based on information theory that mathematically distinguishes between different methods of concept import and measures the associated avoidance of information loss.

1 INTRODUCTION

The field of Medical Informatics has developed a rich ecosystem for research, development, and applications of biomedical terminologies and ontologies. The NCBO BioPortal (NCBO, 2019) provides access to over 772 such resources, containing, as of May 20, 2019, over 9.4 million classes (which would be called "concepts" in other repositories). BioPortal keeps a rich set of statistics about the upload and use of ontologies. These statistics allow the analysis of the quality of ontology maintenance by the curators of individual BioPortal entries (Geller et al., 2018).

BioPortal takes a "big tent" inclusive approach toward the question of *What qualifies as a biomedical ontology?* This is expressed both in the content and structure of some of the resources accessible through bioportal. Thus, (Stato, 2019) is a general purpose statistics ontology that is not specific to medicine. MeSH, the Medical Subject Headings (MeSH, 2019) is contained in BioPortal, although it is widely acknowledged that it is structurally not an ontology at all.

Another major resource for biomedical ontologies is the Unified Medical Language System (UMLS) (UMLS, 2019), developed by the National Library of Medicine (NLM), an institute under the US Government National Institutes of Health (NIH). The most important component of the UMLS is the Metathesaurus (Meta, 2019).

A new version of it is released twice a year and over a long period of time ("decades"), every new release has expanded on the previous version. According to the most recent release notes (Metanotes, 2019), the UMLS contains 3,848,696 concepts and 12,362,080 concept names from 210 distinct terminology sources. The staff of the NLM integrates the different terminologies such that each group of terms with identical meaning is tied together as a single concept and assigned a Concept Unique Identifier (CUI). However, individual terms are maintained with their source information.

1.1 Concept Import

The unique richness of the UMLS makes it possible to compare its subterminologies on a concept basis. Researchers have observed (He et al., 2014) that paths between pairs of concepts that are identical by their CUIs may be different in two different terminologies. Specifically, if a pair of concepts (A, B) exists in both terminologies T_1 and T_2, such that there is a path from A to B consisting of one or more IS-A links (similar to subclass links), then the following situations can arise.

[a] https://orcid.org/0000-0002-9120-525X
[b] https://orcid.org/0000-0002-9478-3303
[c] https://orcid.org/0000-0001-6919-1122

Geller, J., Klein, S. and Keloth, V.
Measuring and Avoiding Information Loss During Concept Import from a Source to a Target Ontology.
DOI: 10.5220/0008354904420449
In *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2019)*, pages 442-449
ISBN: 978-989-758-382-7

- There may be direct IS-A links from A to B in T_1 and T_2, with no intervening concepts. This expresses that A is a more specific concept than B.
- There may be paths of IS-A links and intervening concepts in T_1 and T_2 that are identical. Thus there may be a concept Z between A and B in both T_1 and T_2, such that $(A \text{ IS-A } Z)$ and $(Z \text{ IS-A } B)$. This would be a path of two IS-A links.
- There may be paths of IS-A links and intervening concepts such that the intervening concepts in T_1 are different from those in T_2; furthermore the paths may be of different lengths, including a length of one in either T_1 or T_2.

The two concepts A and B are called anchor concepts. Following the literature (Rector et al., 2006), the difference between paths of different lengths expresses a density difference. As these paths appear in IS-A paths that are conventionally drawn akin to the vertical direction, these have been called *vertical density differences.*

The above observations raised the following intriguing question. If one designates T_1 as a source terminology and T_2 as a target ontology, and the path between the anchor concepts in T_1 is longer than that in T_2, does this mean that the intervening concepts from T_1 that are missing in T_2 can or even should be imported into T_2? In consultation with ontology curators and medical experts, it was determined that the results of an algorithm comparing paths in two ontologies may not be used for automated import from a source ontology into a target terminology. However, these results could be presented to target ontology curator(s) for a decision whether an import would be useful for improving it. It was furthermore observed that even with the help of such an algorithm, the actual work of the target ontology curator remains formidable (He and Geller, 2016). The reasons given by ontology curators for not importing valid concepts include that *they do not want to clutter up their ontology* with concepts for which *no use case exists* or which *no user has ever requested* (Curators, 2019).

In contrast to research on vertical density differences, this paper reports on work on *horizontal density differences* (Keloth et al., 2018; Keloth et al., 2019). Importing concepts from a source into a target ontology, based on horizontal density differences, would lead to a loss of information, and in this paper we analyze how to quantify and avoid this loss.

1.2 Relationship to Data and Ontology Integration

A rich literature exists on ontology alignment, matching and integration. The extensive work of Shvaiko, Euzenat, et al. (Shvaiko et al., 2018) may provide an excellent entry point into this field. Synonym substitution is one tool that can be used for the purpose of integration; this was proposed by (Huang et al., 2009), (Huang et al., 2007) using WordNet (WordNet, 2019) as additional resource besides the UMLS.

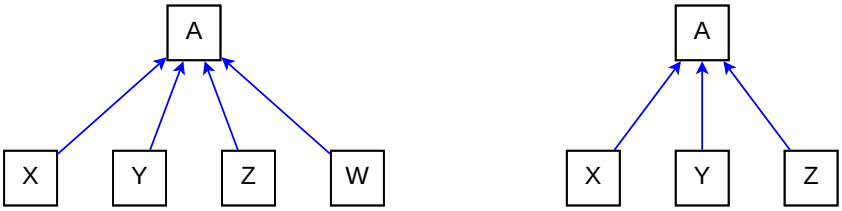
Our goals in this paper are more limited in that we are not attempting automated integration and are also limiting ourselves to a form of local "point wise" import. On the other hand, we are addressing the question of what loss of information occurs and how to avoid it, if the human curator agrees to an import. Ontologies can function as tools in (database) schema integration (Geller et al., 1992; Rahm, 2016).

1.3 Horizontal Density Differences

Figure 1 shows a bare bones example of a horizontal density difference. Terminology 1 (the source) contains a concept A that also exists in Terminology 2 (the target). Furthermore, by using both Terminology 1 and Terminology 2 in the version provided by the UMLS, because A has the same CUI in Terminology 1 and Terminology 2, we may assert that it is the same concept (unless the team of the NLM made a mistake during their integration). Furthermore, we observe that $(X \text{ IS-A } A)$ (i.e., X is a subclass or subconcept of A) in both the source and the target ontology, and again the identity of X is assured by having the same CUI. The same applies to Y and Z. However, there is a density difference. Terminology 1 has an additional concept W that does not exist in Terminology 2. We also assume that W does not exist *anywhere* in Terminology 2.

After importing W into Terminology 2, A has the same children in both terminologies. However, at this point, the information that X, Y and Z were originally in Terminology 2 and that W is "a recent addition," is completely lost. We note that the situation described in Figure 1 is not "theoretical."

In a recent paper (Keloth et al., 2019), we showed that there are many instances of horizontal density differences. This study was based on two popular medical terminologies, MEDCIN and the National Cancer Institute thesaurus (NCIt). It was shown that identical concepts with different sets of children in NCIt and MEDCIN appear 1966 times. More interestingly, 1049 of these concepts *do not have any children in common* in NCIt and MEDCIN. Table 1 shows an ex-



Terminology 1 Terminology 2

Figure 1: Horizontal Density Difference.

ample parent concept that appears in NCIt and MEDCIN and has four common children. The right column shows an example additional child concept in MEDCIN. Table 2 shows seven more examples, listing only the number of common children instead of showing them, in the column with the header C#.

In (Keloth et al., 2019) the authors examined different approaches of how to use this insight for importing concepts from MEDCIN into NCIt. However, they did not deal with the issue of potential loss of information during an import.

Figure 2(a) shows the original situation. Figure 2(b) shows naive import of D and E from the source ontology into the target ontology. In this case a user of the target ontology cannot tell that there is a "historical" difference between the concepts X, Y, and Z versus the concepts D and E. This is a form of information loss.

In Figure 2(c) we attempt to avoid this loss of information by creating an artificial intermediate node Inter1 which maintains a memory of the fact that D and E were imported. However, this leads to an imbalance of the structure that is not logical, because in the source ontology X, Y, Z, D, and E are all at the same level. As level is commonly used to imply generality (in tree-structured ontologies), placing two groups of concepts that were originally at the same level in the source ontology into two different levels in the target ontology corresponds to mutating the structure of the target ontology in an undesirable way. Figure 2(d) shows an alternative solution with two new intermediate nodes. Now the concepts X, Y, Z, D, and E are back to being at the same hierarchical level while still maintaining full information of the provenance of the imported concepts and the original concepts. However, in solution 2(d) we pay the price of having to introduce two artificial nodes.

The idea of introducing intermediate structuring nodes that have little meaning in the ontology might be objected too. However, it is not totally unprecedented. In the NDF-RT (National Drug File - Reference Terminology) (NDF-RT, 2019) groups of drug concepts are combined together by similar intermediate concepts (that, however, have chemical justifications).

2 MEASURING INFORMATION

We shall try to quantify the rather fuzzy concepts exposed above. The big difficulty is that even if a single concept is imported into an ontology of thousands of concepts, "every concept is now suspect." In other words, it is impossible to tell by looking at a concept whether it was originally in the ontology, or whether it was imported. Thus, we will use a "backwards approach," focusing on the gain of information achieved by making the structural changes during import that avoid the original "global" loss.

Measuring information is the main objective of Information Theory and is quite well understood since Shannon's pioneering work in 1948 (Shannon, 1948). The typical scenario is that of a discrete random variable X, taking on a finite number of possible values $x_1, \ldots, x_n$. One also generally assumes that there is a given probability distribution, assigning the probability p_i to the event $X = x_i$, for $1 \leq i \leq n$. The average amount of information conveyed by the random variable X, called its *entropy*, is then defined as $H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i$, and is measured in bits. Resnik (Resnik, 1995) has used entropy to measure *semantic similarity* between concepts within one single ontology, which is a different problem than the one posted here.

The first obstacle to overcome when trying to extend the notion of entropy to the concepts of an ontology as those in Figures 1 and 2 is that there is no underlying probability distribution. One way to avoid it, is to assume uniform probabilities, that is, $p_i = \frac{1}{n}$ for all i, in which case $H(X) = \log_2 n$. Indeed, the information amount given in the left hand side ontologies shown in Figure 2 is $\log_2 5 = 2.32$ bits, which is the number of bits necessary to encode a possible choice among the five alternatives X, Y, Z, D and E.

For the purposes of this analysis we will ignore any connections between concepts that do not imple-

Table 1: Example of Common Children and Extra Child.

Parent
C0149516 : Chronic sinusitis

Common Children	Example Extra Child in MEDCIN
C0008712 : Chronic sphenoidal sinusitis C0008683 : Chronic frontal sinusitis C0008698 : Chronic maxillary sinusitis C0008681 : Chronic ethmoidal sinusitis	C0155827 : Chronic pansinusitis

Table 2: Examples of Extra Children.

Parent	C #	Example Extra Child in MEDCIN
Anti-Arrhythmia Agents	13	pilsicainide hydrochloride
Testosterone	4	testosterone methyl
Loop Diuretics	5	Xipamide
Cranial Nerve Neoplasms	15	overlapping neoplasm of cranial nerve
Glycogen Storage Disease	7	GLYCOGEN STORAGE DISEASE Ic
Mastectomy	6	Bilateral mastectomy
Retinoids	4	aliretinoin

ment the concept taxonomy, i.e., we will not consider any lateral/semantic relationships such as "location." Nevertheless, a second complication arises from the fact that the structures of many important real life ontologies may be more generally directed acyclic graphs (DAGs), such as SNOMED CT (SNOMED CT, 2019), considered the most important clinical ontology, and the NCIt (NCIt, 2019) mentioned earlier, rather than the over-simplified approximation as a tree-like hierarchy dealt with in the example figures above. We shall, however, restrict this preliminary investigation only to trees, and tackle the general problem in future work.

In previous work on structural families of ontologies in BioPortal (Ochs et al., 2016) over 140 tree-shaped biomedical ontologies were observed. Examples include the Healthcare Common Procedure Coding System (HCPCS) and the Drug Ontology (DRON).

To deal with the general case, we shall derive our suggested measure inductively. Assume a tree structure with $r+1$ levels indexed 0 (for the level of the root) to r, and with n_i nodes on level i, for $0 \le i \le r$. Denote the number of nodes on level i that are further sub-partitioned as m_i, so that $n_i - m_i$ is the number of leaves at level i. These m_i nodes have, respectively, $n_{i+1,1}, n_{i+1,2}, \ldots, n_{i+1,m_i}$ children on level $i+1$, so that

$$\sum_{j=1}^{m_i} n_{i+1,j} = n_{i+1} \qquad \text{for } 0 \le i < r.$$

For the example tree displayed in Figure 3, $r = 3$, $(n_0, \ldots, n_3) = (1,4,8,7)$, $(m_0, \ldots, m_3) = (1,3,2,0)$ and $(n_{1,1};\, n_{2,1}, n_{2,2}, n_{2,3};\, n_{3,1}, n_{3,2}) = (4;\, 3,2,3;\, 4,3)$.

We define the information content I of the tree $\mathcal{T}$ by summing, within each level and over all the levels, the logarithm of the branching multiplicity of the nodes, taking a weighted average for the nodes within each level. Formally

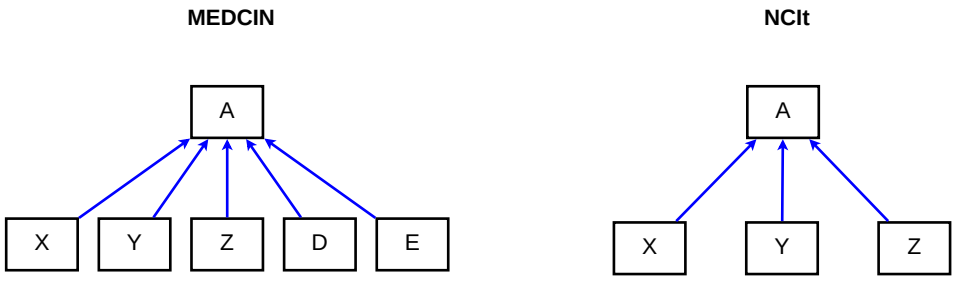
$$I(\mathcal{T}) = \sum_{i=1}^{r} \sum_{j=1}^{m_{i-1}} \frac{n_{i,j}}{n_i} \log_2(n_{i,j}). \qquad (1)$$

Returning to the example of Figure 3, we get

$$I(\mathcal{T}) = \log_2 4 + \left[\frac{3}{8}\log_2 3 + \frac{2}{8}\log_2 2 + \frac{3}{8}\log_2 3\right] + \left[\frac{4}{7}\log_2 4 + \frac{3}{7}\log_2 3\right] = 5.26 \text{ bits}.$$

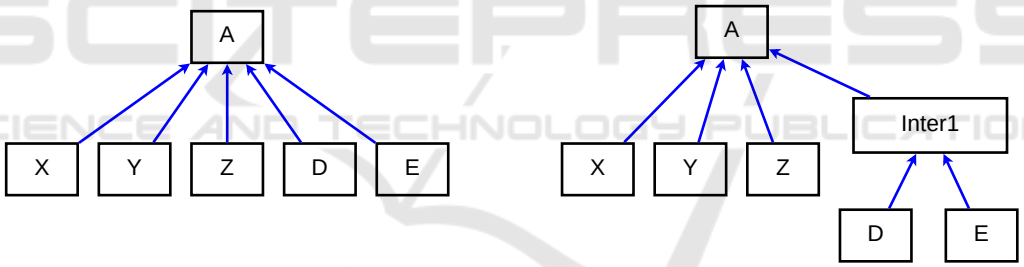
In particular, for a simple ontology with n concepts, represented by a tree of depth 1, that is, with just one level of n leaf nodes, the information content will be $\log_2 n$.

Suppose then that we are given an ontology, which is conveniently represented as a tree structure, and that we want to refine it by introducing an intermediate node R. Assume that this intermediate node is added between level $i-1$ and level i of the tree for some $i > 0$, that there are n_i nodes on level i and that k of them should now be connected to the new node R. The passage from the right hand tree in Figure 2(b) to that of Figure 2(c) is the special case for which $i = 1$, $n_1 = 5$ and $k = 2$. The general scenario is depicted in Figure 4. Note that for convenience, we assume that the k nodes of level i which are connected to R are siblings, in the sense that they were originally children of the same node on level $i-1$.

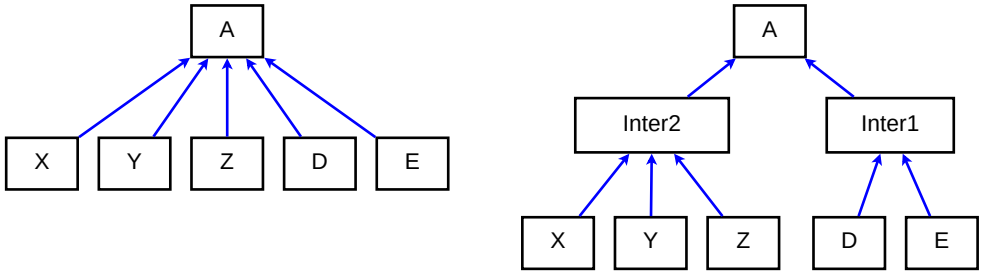


(a) Initial Situation

(b) Import of D and E with complete loss of information.



(c) Import of D and E maintaining the information that they have a common "provenance" from MEDCIN.



(d) Import of D and E with restructuring, making the provenance of all children of A explicit in the target ontology.

Figure 2: Different Approaches to Importing Concepts.

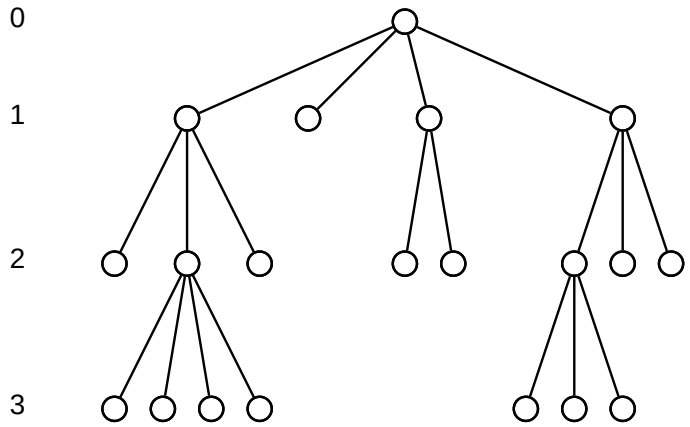


Figure 3: Example tree hierarchy.

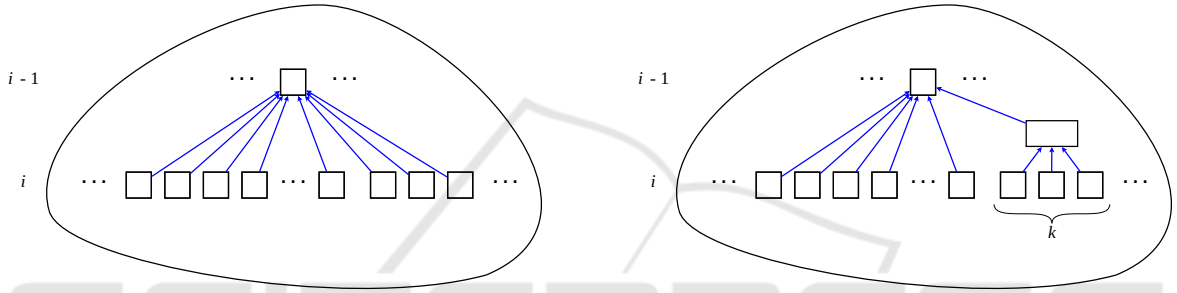


Figure 4: Schematic view of the inductive step in the definition of the information measure.

Since the modified tree structure has obviously added some information, we define the *additional* amount of information that has been added by inserting the intermediate node R as follows. A new choice among k elements has been adjoined, which should add another $\log_2 k$ bits, but only k of the n_i nodes are affected, so we define the added information amount as

$$\frac{k}{n_i}\log_2 k, \tag{2}$$

thereby extending the definition in eq. (1). Returning to the example of Figure 2(c), we get that the information at this stage is

$$\log 5 + \frac{2}{5}\log 2 = 2.72 \text{ bits}. \tag{3}$$

The addition of another intermediate node, as in the passage from the right hand tree in Figure 2(c) to that of Figure 2(d) is yet another example of the same generalization principle, so we get as information content of the structure with both intermediate nodes:

$$\log 5 + \frac{2}{5}\log 2 + \frac{3}{5}\log 3 = 3.67 \text{ bits}. \tag{4}$$

A technical problem arises from the fact that the definition of the additional information relies on having the levels of the tree well defined. However, the newly inserted intermediate nodes may disrupt the level numbering if one considers these nodes as equivalent to the original nodes in the tree. For example, the nodes D and E in the right hand side tree of Figure 2(c) would then be on level 2, while their former siblings X, Y and Z remain on level 1, as in the left hand side of the figure. As a result, the information added by the inter2 node would then be biased, because it would consider all the (remaining) nodes of level 1 to be affected, and not only 3 of the 5 nodes that were originally on level 1.

Since a model in which the level of a node is not influenced by the possible insertion of intermediate nodes seems more reasonable and closer to the real life scenario we wish to simulate, we shall ignore intermediate nodes for the calculation of the level of a node. This is consistent with the fact that these nodes have been artificial additions in the first place, that they are not content-bearing and are used only for technical convenience. Applying this new convention then yields, for our running example, the information amounts reported in eqs. (3) and (4).

One of the advantages of this approach of defining the information content as given in eq. (2) is that, by definition, this information measure is an increasing function of the complexity of the hierarchical struc-

ture: every newly introduced branching conveys additional information, and accordingly, adds a non-negative amount to the previously defined information estimate.

It follows that we may define the *information loss* mentioned in the title of this work by taking the difference between the information contents of the hierarchical structure after and before the import of the concepts from the source to the target ontology. This is precisely the amount given in eq. (2), so it is defined as the number of bits lost by *not* including the additional intermediate node(s).

Another advantage is that the specific definition as a precise number of information bits to be derived from the structure can have a logical interpretation: the given number of bits is the minimal one, from the compression point of view, needed to communicate all the information displayed by the hierarchy, see any textbook dealing with coding, e.g., (Klein, 2016, Chapter 11).

As we have argued, the introduction of the intermediate nodes counteracts a hard to quantify global information loss. Thus, using one or two intermediate nodes avoids the issue of information loss by maintaining a record of the ontology from which the imported concepts are taken.

3 EXPERIMENTAL SETUP

The idea of trying to quantify a semantic concept by assigning it a measure that can be efficiently and precisely calculated is not new and has been applied in various fields. An example could be the *attraction factor* defined in (Choueka et al., 1983), allowing to sort the terms of an ontology according to the strength by which they "attract" the term(s) following them; thus *once upon a* has a high factor, being practically always followed by *time*, but *and* has a low factor, even though *and the* is very frequent, yet there are many other combinations starting with *and*. Another example would be (Geller et al., 2015), in which a measure is derived helping to identify term pairs with strong semantic correlation.

Devising a convincing experimental setup to evaluate the usefulness of a proposed measure does not seem to be a trivial task. The intuition of most readers will hardly differentiate between a structure that has been assigned, say, 4.8 bits, and one with only 3.6 bits; and it will be even harder to convince ourselves why the increase should be by precisely 33%.

A reasonable, yet very resource intensive, approach would be to make use of human informants. One could then prepare a large set of examples and ask the informants to classify them according to what they "feel" their information content should be. In a second stage, the results, averaged over all informants, could be compared with what would be obtained by classifying the examples according to the information measure proposed herein. A high correlation would then be supportive of the usefulness of our suggestions. The current paper, however, is only meant to present the basic ideas, and we leave their evaluation for future work.

4 CONCLUSIONS AND FUTURE WORK

The UMLS mapping of concepts from different ontologies makes it possible to observe potentially missing concepts by comparing pairs of ontologies. A domain expert can then decide whether such concepts should be imported or not. Many opportunities for such imports exist. However, when a concept is imported naively, the information that it was not originally in the target ontology is lost. Quantifying this loss is difficult, because it affects the whole target ontology. We have presented an approach to quantifying the loss of information by measuring the gain that is achieved by maintaining the source information during import, with the aid of "artificial" parent nodes.

In future work, we plan to extend the presented model from trees to Directed Acyclic Graphs (DAGs), which covers a much larger set of biomedical ontologies. We will also attempt to perform a user study with human informants. An algorithm for automatically generating intermediate nodes during import will also be provided.

REFERENCES

Choueka, Y., Klein, S. T., and Neuvitz, E. (1983). Automatic retrieval of frequent idiomatic and collocational expressions in a large corpus. *Journal Association Literary and Linguistic Computing*, 4:34–38.

Curators (2019). Personal communication with ncit and snomed curators.

Geller, J., Keloth, V. K., and Musen, M. A. (2018). How sustainable are biomedical ontologies? In *AMIA 2018, American Medical Informatics Association Annual Symposium, San Francisco, CA, November 3-7, 2018*.

Geller, J., Klein, S. T., and Polyakov, Y. (2015). Identifying pairs of terms with strong semantic connections in a textbook index. In *KEOD 2015 - Proceedings of the International Conference on Knowledge Engineering*

and Ontology Development, Volume 2, Lisbon, Portugal, November 12-14, 2015, pages 307–315.

Geller, J., Perl, Y., Neuhold, E., and Sheth, A. (1992). Structural schema integration with full and partial correspondence using the dual model. *Information Systems*, 17(6):443 – 464.

He, Z. and Geller, J. (2016). Preliminary analysis of difficulty of importing pattern-based concepts into the national cancer institute thesaurus. In *Exploring Complexity in Health: An Interdisciplinary Systems Approach - Proceedings of MIE2016 at HEC2016, Munich, Germany, 28 August - 2 September 2016.*, pages 389–393.

He, Z., Geller, J., and Elhanan, G. (2014). Categorizing the relationships between structurally congruent concepts from pairs of terminologies for semantic harmonization. In *AMIA Joint Summits on Translational Science proceedings*, pages 48–53.

Huang, K., Geller, J., Halper, M., and Cimino, J. J. (2007). Piecewise synonyms for enhanced UMLS source terminology integration. In *AMIA 2007, American Medical Informatics Association Annual Symposium, Chicago, IL, USA, November 10-14, 2007*.

Huang, K., Geller, J., Halper, M., Perl, Y., and Xu, J. (2009). Using wordnet synonym substitution to enhance UMLS source integration. *Artificial Intelligence in Medicine*, 46(2):97–109.

Keloth, V. K., He, Z., Chen, Y., and Geller, J. (2018). Leveraging horizontal density differences between ontologies to identify missing child concepts: A proof of concept. In *AMIA 2018, American Medical Informatics Association Annual Symposium, San Francisco, CA, November 3-7, 2018*.

Keloth, V. K., He, Z., Elhanan, G., and Geller, J. (2019). Alternative classification of identical concepts in different terminologies: Different ways to view the world. *Journal of Biomedical Informatics, in press*.

Klein, S. T. (2016). *Basic Concepts in Data Structures*. Cambridge University Press.

MeSH (2019). Medical Subject Headings, https://bioportal.bioontology.org/ontologies/{mesh}.

Meta (2019). The UMLS Metathesaurus, https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/.

Metanotes (2019). Metathesaurus Release Notes, https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/relea\-se/notes.htm.

NCBO (2019). https://bioportal.bioontology.org/.

NCIt (2019). The National Cancer Institute thesaurus, https://ncithesaurus-stage.nci.nih.govncitbrowser/.

NDF-RT (2019). National Drug File Reference Terminology, https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/{ndfrt}/.

Ochs, C., He, Z., Zheng, L., Geller, J., Perl, Y., Hripcsak, G., and Musen, M. A. (2016). Utilizing a structural meta-ontology for family-based quality assurance of the bioportal ontologies. *Journal of Biomedical Informatics*, 61:63–76.

Rahm, E. (2016). The case for holistic data integration. In *Advances in Databases and Information Systems - 20th East European Conference, ADBIS 2016, Prague, Czech Republic, August 28-31, 2016, Proceedings*, pages 11–27.

Rector, A. L., Rogers, J., and Bittner, T. (2006). Granularity, scale and collectivity: When size does and does not matter. *Journal of Biomedical Informatics*, 39(3):333–349.

Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. In *IJCAI'95 Proceedings of the 14th international joint conference on Artificial intelligence - Volume 1, Montreal, Quebec, Canada, August 20 - 25, 1995*, pages 448–453. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(2):379–423.

Shvaiko, P., Euzenat, J., Jiménez-Ruiz, E., Cheatham, M., and Hassanzadeh, O., editors (2018). *Proceedings of the 13th International Workshop on Ontology Matching co-located with the 17th International Semantic Web Conference, OM@ISWC 2018, Monterey, CA, USA, October 8, 2018*, volume 2288 of *CEUR Workshop Proceedings*. CEUR-WS.org.

SNOMED CT (2019). https://www.snomed.org/

Stato (2019). Statistics Ontology, https://bioportal.\-bioontology.org/ontologies/{stato}.

UMLS (2019). The Unified Medical Language System, https://www.nlm.nih.gov/research/umls/.

WordNet (2019). https://wordnet.princeton.edu/.