

An Automatic Approach to Extending the Consumer Health Vocabulary

Michal Monselise[†], Jane Greenberg, Ou Stella Liang, Sonia Pascua, Heejun Kim, Mat Kelly, Joan P. Boone, Christopher C. Yang

Drexel University, Philadelphia, PA, USA

Abstract

Purpose: Given the ubiquitous presence of the internet in our lives, many individuals turn to the web for medical information. A challenge here is that many laypersons (as “consumers”) do not use professional terms found in the medical nomenclature when describing their conditions and searching the internet. The Consumer Health Vocabulary (CHV) ontology, initially developed in 2007, aimed to bridge this gap, although updates have been limited over the last decade. The purpose of this research is to implement a means of automatically creating a hierarchical consumer health vocabulary. This overall purpose is improving consumers’ ability to search for medical conditions and symptoms with an enhanced CHV and improving the search capabilities of our searching and indexing tool HIVE (Helping Interdisciplinary Vocabulary Engineering).

Design/methodology/approach: The research design uses ontological fusion, an approach for automatically extracting and integrating the Medical Subject Headings (MeSH) ontology into CHV, and further convert CHV from a flat mapping to a hierarchical ontology. The additional relationships and parent terms from MeSH allow us to uncover relationships between existing terms in the CHV ontology as well. The research design also included improving the search capabilities of HIVE identifying alternate relationships and consolidating them to a single entry.

Findings: The key findings are an improved CHV with a hierarchical structure that enables consumers to search through the ontology and uncover more relationships.

Research limitations: There are some cases where the improved search results in HIVE return terms that are related but not completely synonymous. We present an example and discuss the implications of this result.

Practical implications: This research makes available an updated and richer CHV ontology using the HIVE tool. Consumers may use this tool to search consumer terminology for medical conditions and symptoms. The HIVE tool will return results about the medical term linked with the consumer term as well as the hierarchy of other medical terms connected to the term.

Citation: Monselise, Michal, Jane Greenberg, Ou Stella Liang, Sonia Pascua, Heejun Kim, Mat Kelly, Joan P. Boone, and Christopher C. Yang. “An automatic approach to extending the consumer health vocabulary.” *Journal of Data and Information Science* (2020). <https://doi.org/10.2478/jdis-2021-0003>

Received: Mar. 11, 2020

Revised: Jul. 16, 2020;

Aug. 26, 2020

Accepted: Aug. 27, 2020



[†] Corresponding author: Michal Monselise (E-mail: michal.monselise@gmail.com).

Originality/value: This is a first attempt in over a decade to improve and enhance the CHV ontology with current terminology and the first research effort to convert CHV's original flat ontology structure to a hierarchical structure. This research also enhances the HIVE infrastructure and provides consumers with a simple, efficient mechanism for searching the CHV ontology and providing meaningful data to consumers.

Keywords Consumer Health Vocabulary; Ontological fusion; Medical ontologies

1 Introduction

Vocabulary systems in the form of controlled vocabularies, thesauri, and ontologies are used to index resources stored in information systems and aid in information retrieval, discovery, and access. These types of vocabulary systems generally structure and encode the relationship among terms. Structuring processes frequently adhere to standards, such as the ANSI/NISO Z39.19 *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* (NISO, 2010). Structured vocabularies following this standard encode hierarchical, associative, and equivalent relationships among terms. Over nearly the last two decades, the World Wide Web Consortium (W3C) has introduced encoding standards, such as Simple Knowledge Organization Systems (SKOS) (Miles et al., 2005) and the Web Ontology Language (OWL) (Antoniou & van Harmelen, 2004), providing new opportunities for structured vocabularies (NISO, 2010) across the Web. These standards build on eXtensible Mark-up Language (XML) and integrate the Resource Description Framework (RDF) model, making vocabularies machine processable and enabling support for computational approaches. These innovations support linked data (Heath & Bizer, 2011) and the goals of the Semantic Web (Shadbolt, Berners-Lee, & Hall, 2006) by providing the necessary infrastructure for connecting information resources across the Web. Linked data capabilities have excelled in selected areas of medicine and biology; however, progress specific to consumer health seems limited, which is surprising, as millions of individuals turn to digital media for health information on a daily basis looking for genuine answers and assistance.

Consumers seeking health information face the full range of consumer-oriented to scientific information, across the Web, as well as commercial information that has not necessarily been vetted. The realization of a consumer health vocabulary has historical roots in public libraries and early digital libraries (Rubenstein, 2012); and, in 2007, researchers at the Brigham and Women's Hospital and National Library of Medicine joined together developing the first-generation *Consumer Health Vocabulary* (CHV) (Zeng, 2007). Sponsored by the National Institutes of Health-National Library of Medicine (NIH-NLM), the initial CHV vocabulary was



developed as a flat file (lacking hierarchies) and maintained over close to a four-year period. In 2004, drawing on Semantic Web (Heath & Bizer, 2011) principles and infrastructure, an interconnected group of researchers developed the Open Consumer Health Vocabulary (OCHV), which was recently (2016) registered in National Center for Biological Ontologies' Bioportal in (Salvadores et al., 2013). These two initiatives stand as important contributions to the consumer health ontology space, although, unfortunately, they have discontinued, chiefly due to funding limitations and change in research priorities.

Today, given the promise of the Big Data rush and AI, it seems more pressing than ever to advance the original CHV ontology. The research presented in this paper takes a step toward addressing this need and developing a best practice for extending CHV. The research is motivated by the importance of consumer health to society and value of ontologies for both users and computational activities. The following sections cover background and data sources and a review of relevant prior research in this area as well as our goals of objectives of the research discussed in this study. The paper also reviews the research methods and procedures, including the incorporation of the Helping Interdisciplinary Vocabulary Engineering (HIVE) (Greenberg et al., 2011) technology; and presents the results and contextual discussion. Finally, the conclusion restates key outcomes and considers next steps.

2 Background and data sources

2.1 Consumer health vocabulary

The consumer health vocabulary (CHV) is a controlled vocabulary containing a mapping between consumer terms and UMLS medical terms. CHV was initially proposed in an expert panel at MEDINFO in 2001 (19). In the early 1990s, with the advancement of the Web, the imbalance of domain knowledge between professionals and laypersons was gradually tilted as consumers across many societal sectors increasingly gained access to information online. A 2008 study based on the popular OHC site, PatientsLikeMe, found that 43% of the patient-described symptoms have a match in the Unified Medical Language System Metathesaurus (Smith & Wicks, 2008). To help consumers with identifying all symptoms by matching them with UMLS terms, there was a need to bridge the gap between consumer and medical terminology. CHV was proposed as a solution to bridge this gap.

Over the years since the initial development of CHV, there have been a few efforts to expand the vocabulary. One such effort by Zeng et al. (2007) used human review as well as logistic regression and other formulas to expand the vocabulary. Another study by Doing-Harris and Zeng-Treitler (2011) proposes an expansion of CHV using social network mining.



Consumer terms come from a variety of sources including social media, discussion boards and patient diaries. Given this heterogeneity of consumer terms, one can anticipate that multiple consumer terms map to the same preferred UMLS term. As part of background exploratory work for the research reported on below, we found there are about 58,000 UMLS terms that are mapped to over 150,000 consumer terms. This means for every UMLS term, we have an average of three consumer terms associated with the UMLS term. The Medical Subject Heading (MeSH) thesaurus is another ontology that is a part of UMLS and offers possibilities for enhancing CHV. The next section discusses MeSH.

2.2 MeSH

The Medical Subject Heading (MeSH) thesaurus is another ontology system, often referred to as a controlled vocabulary, which published by the National Library of Medicine. The main goal of MeSH is to organize all articles and journals published in the life sciences. Particularly, all MEDLINE/PubMed articles are indexed using the MeSH controlled vocabulary. MeSH organizes all medical terminology into a tree structure. There are 16 initial categories which are then divided further into subcategories. The thesaurus has up to 13 levels of subcategories. The most compelling case for the research presented here is that, in addition to organizing terms in categories, both MeSH and CHV are a part of the collection of UMLS vocabularies. As such, they share a unique identifier called CUI (Concept Unique Identifiers).

The shared CUI provides an important system features that could be leveraged pursuing ontological fusion. Specific to the goal of enhancing CHV, the fact that the two ontologies share a unique identifier provides a pathway for automatically enhancing the consumer health vocabulary using MeSH. Pursuing this step will allow researchers to capture health consumers' ever-evolving preferences in a machine-readable format, which could further be utilized by algorithms to improve the quality of information delivered to consumers' health-related queries.

3 Related work

3.1 Interoperability of ontologies

As we aim to extend and enhance CHV using another ontology, we look to other research on achieving and improving the interoperability among ontologies. There are a number of projects aimed at linking multiple ontologies to improve our ability to index and retrieve data. Zeng (2019) presents an insightful account detailing key approaches for merging two or more ontologies; and this work serves as a guide for



achieving interoperability. Research by Isaac et al. (2009) also details merging of ontologies using different techniques and provides insight to various metrics that help determine whether terms can be matched focusing on library data. Francesconi et al. (2008) also discuss a method for merging ontologies; however, their research focuses on combining five ontologies that are of importance in the EU. Chan and Zheng (2002) discuss matching ontologies using different algorithms and evaluation criteria. Background research also points to a study by Slater, Gkoutos, and Hoehndorf (2020), and their method for combining biomedical ontologies into a single meta-ontology. This piece addresses a range of challenges drawing, given the different semantic levels and relationships found in ontologies. The research reviewed here helped us to gain insight into various approaches and challenges, and provided grounding for how we might enhance the CHV.

3.2 Expanding CHV

As part of our review, we also examined research projects aimed at expanding CHV. It seems that previous projects mainly focused on expanding the ontology using unstructured text. One example is the study by Jiang and Yang (2015) that enhanced CHV with new terms by mining online health communities and determined whether terms should be added to the ontology using co-occurrence metrics. Another example is a study by He et al. (2017) exploring mining a social Q&A site to find candidate terms for expanding CHV. This stands as important work, although the results of this have not been incorporated into the publically accessible CHV. Overall, reporting on expanding the CHV and evaluating the outcomes are limited. Our aim is to expand the CHV, and, as part of our approach we seek to improve the interoperability of CHV using structured data from a well maintained ontology.

4 Research goals and objectives

The overriding goal of this research is to extend and enhance the original CHV. Many new medical conditions have emerged since the last CHV update. Examples such as illnesses associated with opioid addiction and vaping, have consumer terms associated with them; however, neither the conditions nor their associated consumer terms are currently captured in CHV. Another chief limitation to address is that the current CHV ontology is a flat terminology. Relationships among terms is central to ontology construction, and their ability to support computational activity and machine learning. To address this limitation, we took steps to enhance CHV with parent terms and additional associations and capture relationships useful for to consumers in learning about an illness as well as future data driven activity



supporting machine learning. These enhancements will help us overcome the challenges associated with CHV, particularly having multiple term variants at the same level. Additionally, adding parent terms, as part of a hierarchy, will help us identify where to insert new terms more easily. Our key objectives in this research are to convert CHV to a hierarchical ontology using terms from MeSH. This research will outline steps to perform this process automatically. By adding a hierarchical structure to CHV, we will enable consumers to produce more relevant results when searching for terms using this ontology. Our plan is to eliminate all redundant and irrelevant results as well as improve the discovery of terms using their relationships.

5 Methodology

5.1 Methods

We present our approach for automatically extending and enhancing CHV using identifier mapping and ontological fusion. To achieve these goals, we performed the following steps:

- *Implementation of a hierarchical structure.* We enhanced and improved CHV by converting the controlled vocabulary from a flat list to a hierarchical structure. This goal required converting the data to a hierarchical structure. The CHV ontology is a mapping from consumer terms to MeSH terms integrated within the UMLS. We selected MeSH because MeSH and CHV share an identifier, the language is understandable by the general consumer, and because it offers a hierarchical structure and performed a join using the UMLS terms that both ontologies share.
- *HIVE (Helping Interdisciplinary Vocabulary Engineering) integration/development.* We worked with the HIVE application to view and study the enhanced CHV, with the longer-term goal of enabling automatic assignment of subject metadata with CHV using HIVE. HIVE is both a model and a system that supports automatic metadata generation by drawing descriptors from multiple Simple Knowledge Organization System (SKOS)-encoded controlled vocabularies (Greenberg et al., 2011). This goal was accomplished by creating new unique identifiers that mapped to multiple alternate terms. In developing identifiers, our convention was to use the URI associated with the institution of the authors and assign a sequential integer identifier to each member of the vocabulary. The result is a significant reduction of the number of terms returned in HIVE's search results by removing all alternate terms from the search results.



Our methodology is described in Figure 1. We enhance and improve CHV and rename it the Combined Consumer Health Vocabulary (CCHV).

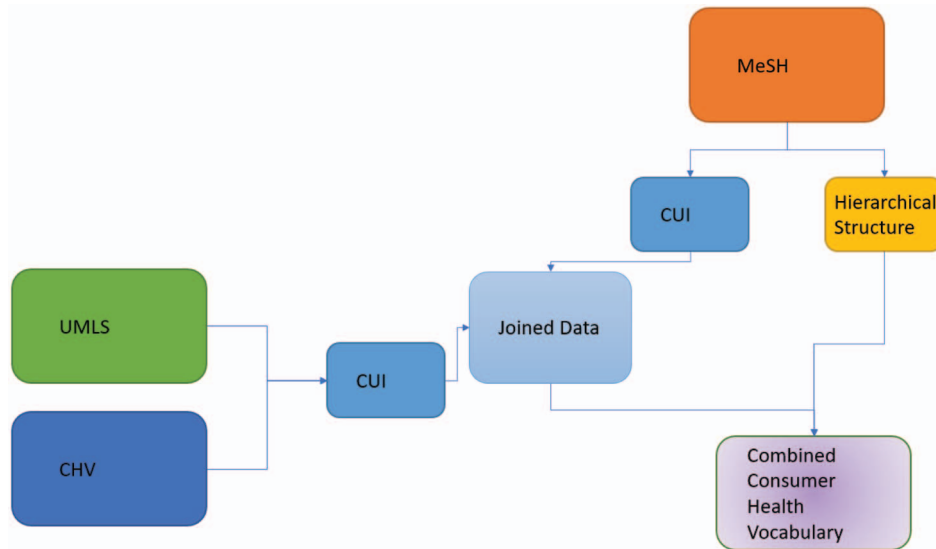


Figure 1. Diagram of methodology for producing the Combined Consumer Health Vocabulary.

5.2 Procedures

The original CHV ontology contains 158,519 terms and consists of two types of terms – UMLS terms that have a CUI associated with them and consumer terms that do not have a CUI directly associated with them. These terms are mapped to a UMLS term and are associated with a CUI indirectly. There are 57,819 unique UMLS terms in the CHV ontology. As Figure 1 shows, we start by joining our CHV and UMLS data to the MeSH terms using the common CUI field. While the consumer terms are mapped to the UMLS terms, they do not have a unique identifier. Therefore, the next step in processing this vocabulary was to assign a unique identifier to each consumer term in the ontology. This will differentiate the consumer terms from the UMLS terms as they will be represented as independent terms with their own identifier. We connect the consumer terms and the UMLS terms in the ontology with the alternate relationship. This is because in CHV, the terms are connected with both an alternate and a related relationship. Additionally, terms are connected with an alternate relationship in MeSH (they are defined as synonyms) so using an alternate relationship contributes to maintaining consistency in the new ontology. The unique identifiers were generated by assigning a URI to the terms



that ended in an eight-digit unique identifier. Using the UMLS terms in CHV, we identified all terms in MeSH that were parents of the UMLS terms in CHV and added these terms and their entire hierarchy. These parent terms were connected as parents to both UMLS terms and consumer terms in CHV as parents. Additionally, relationships between existing terms were retrieved from MeSH. This step is described in Figure 1 as the extraction of the hierarchical structure and led to additional alternate relationships in the ontology. We add these terms and relationships to CHV and transform the mapping from a flat mapping to a hierarchy with multiple connections. The new ontology is named CCHV. The final step in this procedure is to enhance HIVE. HIVE was enhanced to store all alternate terms as concepts. This means that we could use these relationships to discover more information in our data. Furthermore, HIVE has been modified to return only preferred terms when performing a search. This means a more streamlined result when performing a search in HIVE. Rather than returning all terms that contain the search query, we now return only the preferred terms for any term returned by the search query.

6 Results

Transitioning to CCHV means that we can now query the vocabulary and retrieve results that contain more information about each term. Particularly, we can learn more about the parent, child, and related terms for each search result. Additionally, fewer irrelevant results are retrieved from a query since the search result only contains preferred terms. For example, while *bunion* and *bunions* both exist in the vocabulary as consumer terms, searching either one will return *bunion*, since *bunion* is the preferred term. Before enhancing CHV, we had no broader or narrower relationships. After transitioning to CCHV, a total of 8,269 MeSH parent terms were added, and some parent-child relationships were uncovered between terms that already existed in the hierarchy. Furthermore, the number of broader-narrower relationships between existing terms increased 0 to 55,571 relationships among terms. At the same time, there were 30,060 related relationships and 51,995 alternate relationships. The alternate relationships were retained while the related relationships were enhanced. By adding the related relationships from MeSH, we now have 324,545 related term relationships. These new relationships allow users to discover more information by clicking through the concepts and discovering more information. We also use the relationships in querying and indexing. The summary of the change in semantic relationships counts from the original CHV, to the enhance CHV, called CCHV, is presented in Table 1.



Table 1. Relationships between the original CHV and the new CCHV.

VOCABULARY	SEMANTIC RELATIONSHIP			
	ALTERNATE TERM	RELATED TERM	BROADER TERM	NARROWER TERM
Original CHV	51,995	30,060	0	0
CCHV	51,995	325,545	55,571	55,571

The change in the data is demonstrated in the examples in Figures 2 and 3. Figure 2 is a screenshot of the term *bunion* from CHV in HIVE before adding the additional information from MeSH. This term is assigned the CUI from UMLS. Figure 3 shows CCHV which contains the additional MeSH terms. In Figure 3, we see the same *bunion* term from the original CHV but enhanced with the broader and narrower terms. We combined all the relationships to *bunion* from all alternates of *bunion* to create one combined *bunion* term. Since we introduced new relationships, we also assigned a new ID to this term.

[List](#) [JSON-LD](#) [SKOS RDF/XML](#) [Dublin Core](#) [XML](#)

Preferred label *bunion*

URI <http://uth.tmc.edu/ontology/ochv#C0006386>

Alternate label Not provided

Notes label Not provided

Broader No broader concepts

Narrower No narrower concepts

Related No related concepts

Figure 2. Search results for the term *bunion* in the previous version of HIVE using CHV.

Adding terms and relationships from MeSH helps us discover relationships in our data and adds more context to the consumer terms. Since HIVE is a tool for indexing documents, uncovering more connections in the vocabulary used to index will mean more context on the terms in the document. In the example, we see that *Taylor's bunion* is a term that existed in CHV before the enhancement but now we know that it is a narrower term of *bunion*. However, the term *foot deformities* did not exist in the data and has been added from MeSH. Additionally, we have used the preferred label for each term so that searching for an alternate will return the preferred term. For example, in Figure 4, *hallux valgus* is an alternate term for *bunion*. Searching for *hallux valgus* returns the term *bunion* as a first result.



[List](#)
[JSON-LD](#)
[SKOS RDF/XML](#)
[Dublin Core](#)
[XML](#)

Preferred label [bunion](#)

URI <http://www.drexel.edu/ontology/cchv#E0001181>

Alternate label No alternate concepts

Notes label UML Concept

Broader

- Foot Deformities, Acquired
- foot deformities

Narrower

- taylor's bunion (disorder)
- bunionette

Related

- hallux valgus
- acquired hallux valgus
- congenital hallux valgus
- bunions

Figure 3. Search results for the term *bunion* in the improved version of HIVE using CCHV.

2 Enter search term [Search](#)

CCHV

[bunion](#)

[List](#)
[JSON-LD](#)
[SKOS RDF/XML](#)
[Dublin Core](#)
[XML](#)

Preferred label [bunion](#)

URI <http://www.drexel.edu/ontology/cchv#E0001181>

Alternate label No alternate concepts

Notes label UML Concept

Broader

- Foot Deformities, Acquired
- foot deformities

Narrower

- taylor's bunion (disorder)
- bunionette

Related

- hallux valgus
- acquired hallux valgus
- congenital hallux valgus
- bunions

Figure 4. Search results for the term *hallux valgus* in the improved version of HIVE using CCHV.

Examining other search queries and comparing only the number of search terms shows we typically find a reduction of results. In Table 2, we see an improvement in results for the query *liver failure*. In the previous version of HIVE, we get a result of the same term appearing twice while in the new HIVE it only appears once.

Table 2. A comparison of search results for the term “liver failure” between the previous and the improved versions of HIVE.

HIVE Results - Before Improvements	HIVE Results - After Improvements
Liver failure	Liver failure
Liver failure	

In Table 3, we present the result of the query *ovarian cysts*. The new version of HIVE returns fewer terms. Some of the terms are not directly related to the query but are returned in our search because of an alternate. In this case, the term *brca1 genes* is returned even though it isn’t an exact match with *ovarian cysts*.

Table 3. A comparison of search results for the term *ovarian cyst* between the previous and the improved versions of HIVE and CHV.

HIVE Results - Before Improvements	HIVE Results - After Improvements
cyst ovarian	brca1 genes
cysts ovarian	ovarian cysts
ovarian cyst	parovarian cyst
ovarian cystectomy	
ovarian cysts	
ovarian cysts	
parovarian cyst	
parovarian cyst	
parovarian cysts	

7 Discussion

The results presented above demonstrate the improvements made in both CHV and HIVE. The new CCHV ontology is a transformation from a flat mapping of consumer terms to UMLS terms to a hierarchical dataset that has a tree structure and contains both new terms from MeSH as well as new relationships between existing terms uncovered from MeSH. In many areas of research, a hierarchical ontology is considered an improvement on a flat mapping since it can convey more information through the relationships between the terms. One example of an area of research where such an improvement is observed is in protein function prediction using gene ontologies (Eisner et al., 2005). Another area of research that illustrates the benefits of a hierarchical ontology over a flat one is in the classification of web search results (Singh & Nakata, 2005). In addition to enhancing the structure of the ontology, our study has also added improvements to HIVE. HIVE has been enhanced



to produce more relevant search results. HIVE now returns one search result per preferred term rather than returning all alternate terms containing the search query. To demonstrate this, we tested a sample of 30 search queries and saw a reduction in the number of results with the new version of HIVE. The mean number of results returned in the previous version of HIVE was 140.4 and went down to 22.87 in the new version of HIVE. The *bunion* search query highlights the improvements in our results. Other research like the study by Kaisser, Hearst, and Lowe (2008) has shown that shorter search results are considered preferable. Therefore, a reduction in the number of terms returned in a query is a desirable result. Using the example of the search term *bunion*, in previous versions of HIVE and CHV, there were no broader, narrower, or alternate terms connected to the *bunion* entry in CHV. However, after enhancing both HIVE and CHV, we added the relationships to the alternates *hallux valgus*, *acquired hallux valgus*, *congenital hallux valgus*, and *bunions*. We also added the relationships to the narrower terms *Tailor's bunion* and *bunionette*. This has been illustrated in Figures 2–4. The related terms previously existed in our dataset but the relationships between them were not retrieved in the HIVE search. We also added the terms foot deformities and acquired foot deformities from MeSH. Having these terms in our dataset allows us to connect the term *bunion* with other types of foot deformities. The additional relationships between the terms in the updated ontology allow for additional data discovery for consumers.

While we have reduced the number of results per search, occasionally, we will get a result that is not directly related to the query. For example, when searching for *ovarian cysts*, our search query produces less duplication and streamlines the result by combining all alternates into one preferred term. However, our query also returned *brca1 genes*. While both terms are related to women's health, the result is not synonymous with the query. This is a definite area for improvement. The algorithm should be improved to display only the most relevant results for each query. One possible method is by using an algorithm for synonym detection like the one developed by Turney (2001).

8 Conclusion

This research has been performed with the aim of enhancing CHV and improving HIVE to produce improved search results. The new CHV contains more relationships between existing concepts. It contains more types of relationships since the previous version of CHV contained no broader or narrower terms and no related terms. CHV has transformed into a hierarchical structure by adding broader terms from MeSH. HIVE has been enhanced by improving query results when searching for medical terms. After enhancing CHV and HIVE, search queries now return fewer terms. The



terms returned are only preferred terms rather than returning multiple variations of the same term with slightly different spelling. We have consolidated all duplicate terms and very similar terms and produce a more streamlined result that is easier to sift through. We can still uncover the alternate terms by looking at the alternates of the query result. This research is a first step in enhancing CHV to include new terms that have been added to other medical ontologies since the last release of CHV.

Allowing users to search for consumer terms and obtain results of the preferred medical term can have great benefits to both consumers and medical practitioners (McCray et al., 1993). Using such a resource can help identify medical conditions or symptoms of a condition from discussions in messaging boards or improve results in search engines. By increasing the number of connections between terms, we allow for improved discovery of medical conditions and symptoms using consumer terms.

While this research significantly improves both CHV and HIVE, in performing the research we identify two main areas for further work. The first area is to devise an algorithm to identify alternates that are not synonymous with each other. We identified such an example with the relationship between *ovarian cysts* and *breast genes*. The two are not synonymous but related. However, when searching for one, the other was returned in our results. To further improve our results, we will devise a machine learning algorithm that will learn to identify the most appropriate relationship between terms and recommend whether the terms have a broader-narrower relationship, an alternate relationship, or a related relationship. This future research will further improve our search results. A second area of improvement is to perform the second stage of this analysis and enhance CHV with new terms that did not exist when CHV was last updated. There are many conditions that do not exist in the current version of CHV. One example is conditions related to vaping. There is currently no mention of these medical conditions in CHV and consumers and medical practitioners would both benefit from such enhancements to the vocabulary. The plan for CHV enhancement with new terms will also use a machine learning algorithm. Our plan is to scrape multiple message boards that contain questions about health and use natural language processing to identify relationships between consumer terms discussed in the message boards and UMLS terms.

Author contributions



References

- ANSI/NISO Z39.19-2005 (2010) Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies | NISO website. (n.d.). Retrieved from <https://www.niso.org/publications/ansiniso-z3919-2005-r2010>
- Antoniou, G., & van Harmelen, F. (2004). Web Ontology Language: OWL. In S. Staab & R. Studer (Eds.), *Handbook on Ontologies* (pp. 67–92). Springer. https://doi.org/10.1007/978-3-540-24750-0_4
- Chan, L.M., & Zeng, M.L. (2002). Ensuring interoperability among subject vocabularies and knowledge organization schemes: A methodological analysis. *IFLA Journal*, 28(5–6), 323–327. <https://doi.org/10.1177/034003520202800519>
- Doing-Harris, K.M., & Zeng-Treitler, Q. (2011). Computer-assisted update of a consumer health vocabulary through mining of social network data. *Journal of Medical Internet Research*, 13(2), e37. <https://doi.org/10.2196/jmir.1636>
- Eisner, R., Poulin, B., Szafron, D., Lu, P., & Greiner, R. (2005). Improving protein function prediction using the hierarchical structure of the gene ontology. 2005 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, 1–10. <https://doi.org/10.1109/CIBCB.2005.1594940>
- Francesconi, E., Faro, S., Marinai, E., & Peruginelli, G. (2008). A methodological framework for thesaurus semantic interoperability. In *Proceeding of the Fifth European Semantic Web Conference*, pp. 76–87. [https://scholar.googleusercontent.com/scholar?q=cache:5lsA9SO1Fx oJ:scholar.google.com/+Francesconi,+E.,+Faro,+S.,+Marinai,+E.,+%26+Perugi,+G.+ \(2008\).+A+methodological+framework+for+thesaurus+semantic+interoperability.+In+ Proceeding+of+the+Fifth+European+Semantic+Web+Conference+\(pp.+76-87\).&hl=en&as_sdt=0,39](https://scholar.googleusercontent.com/scholar?q=cache:5lsA9SO1Fx oJ:scholar.google.com/+Francesconi,+E.,+Faro,+S.,+Marinai,+E.,+%26+Perugi,+G.+ (2008).+A+methodological+framework+for+thesaurus+semantic+interoperability.+In+ Proceeding+of+the+Fifth+European+Semantic+Web+Conference+(pp.+76-87).&hl=en&as_sdt=0,39)
- Greenberg, J., Losee, R., Agüera, J.R.P., Scherle, R., White, H., & Willis, C. (2011). HIVE: Helping interdisciplinary vocabulary engineering. *Bulletin of the American Society for Information Science and Technology*, 37(4), 23–26. <https://doi.org/10.1002/bult.2011.1720370407>
- He, Z., Chen, Z.W., Oh, S., Hou, J.H., & Bian, J. (2017). Enriching consumer health vocabulary through mining a social Q&A site: A similarity-based approach. *Journal of Biomedical Informatics*, 69, 75–85. <https://doi.org/10.1016/j.jbi.2017.03.016>
- Heath, T., & Bizer, C. (2011). Linked data: Evolving the web into a global data space. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 1(1), 1–136. <https://doi.org/10.2200/S00334ED1V01Y201102WBE001>
- Isaac, A., Wang, S., Zinn, C., Mattheizing, H., van der Meij, L., & Schlobach, S. (2009). Evaluating thesaurus alignments for semantic interoperability in the library domain. *IEEE Intelligent Systems*, 24(2), 76–86. <https://doi.org/10.1109/MIS.2009.26>
- Jiang, L., & Yang, C.C. (2015). Expanding consumer health vocabularies by learning consumer health expressions from online health social media. In N. Agarwal, K. Xu, & N. Osgood (Eds.), *Social Computing, Behavioral-Cultural Modeling, and Prediction* (pp. 314–320). Springer International Publishing. https://doi.org/10.1007/978-3-319-16268-3_36
- Kaiser, M., Hearst, M.A., & Lowe, J.B. (2008). Improving search results quality by customizing summary lengths. In *Proceedings of ACL-08: HLT*, 701–709. <https://www.aclweb.org/anthology/P08-1080>



- McCray, A.T., Aronson, A.R., Browne, A.C., Rindfleisch, T.C., Razi, A., & Srinivasan, S. (1993). UMLS knowledge for biomedical language processing. *Bulletin of the Medical Library Association*, 81(2), 184–194.
- Miles, A., Matthews, B., Wilson, M., & Brickley, D. (2005). SKOS Core: Simple Knowledge Organisation for the Web. In *Proceedings of the 2005 International Conference on Dublin Core and Metadata Applications: Vocabularies in Practice*. pp. 3–10.
- Rubenstein, E. (2012). From social hygiene to consumer health: Libraries, health information, and the American public from the late nineteenth century to the 1980s. *Library & Information History*, 28(3), 202–219. <https://doi.org/10.1179/1758348912Z.00000000016>
- Salvadores, M., Alexander, P.R., Musen, M.A., & Noy, N.F. (2013). BioPortal as a dataset of linked biomedical ontologies and terminologies in RDF. *Semantic Web*, 4(3), 277–284.
- Shadbolt, N., Berners-Lee, T., & Hall, W. (2006). The semantic web revisited. *IEEE Intelligent Systems*, 21(3), 96–101. <https://doi.org/10.1109/MIS.2006.62>
- Singh, A., & Nakata, K. (2005). Hierarchical classification of web search results using personalized ontologies. In *Proceedings of the 3rd International Conference On Universal Access in Human-Computer Interaction*, 10.
- Slater, L.T., Gkoutos, G.V., & Hoehndorf, R. (2020). Towards semantic interoperability: Finding and repairing hidden contradictions in biomedical ontologies. *bioRxiv* 2020.05.16.099309. <https://doi.org/10.1101/2020.05.16.099309>
- Smith, C.A., & Wicks, P.J. (2008). PatientsLikeMe: Consumer health vocabulary as a folksonomy. *AMIA Annual Symposium Proceedings*, 2008, 682–686.
- Turney, P.D. (2001). Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL. In L. De Raedt & P. Flach (Eds.), *Machine Learning: ECML 2001* (Vol. 2167, pp. 491–502). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-44795-4_42
- Zeng, M.L. (2019). Interoperability. *Knowledge Organization*, 46(2), 122–146. <https://doi.org/10.5771/0943-7444-2019-2-122>
- Zeng, Q., Tse, T., Divita, G., Keselman, A., Crowell, J., Browne, A., Goryachev, S., & Ngo, L. (2007). Term identification methods for consumer health vocabulary development. *Journal of Medical Internet Research*, 9(1), e4. <https://doi.org/10.2196/jmir.9.1.e4>



This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

