



HHS Public Access

Author manuscript

Proceedings (IEEE Int Conf Bioinformatics Biomed). Author manuscript; available in PMC
2023 December 20.

Published in final edited form as:

Proceedings (IEEE Int Conf Bioinformatics Biomed). 2019 November ; 2019: 1177–1178. doi:10.1109/
bihm47256.2019.8983220

Ontology of Consumer Health Vocabulary: providing a formal and interoperable semantic resource for linking lay language and medical terminology

Muhammad Amith,

School of Biomedical Informatics, University of Texas Health Science Center at Houston Houston, TX, United States

Kirk Roberts,

School of Biomedical Informatics University of Texas Health Science Center at Houston Houston, TX, United States

Cui Tao,

School of Biomedical Informatics University of Texas Health Science Center at Houston Houston, TX, United States

Licong Cui,

School of Biomedical Informatics University of Texas Health Science Center at Houston Houston, TX, United States

Hua Xu

School of Biomedical Informatics University of Texas Health Science Center at Houston Houston, TX, United States

Abstract

The Consumer Health Vocabulary has been an important contribution to the health informatics field since its introduction in 2006. Many studies have utilized the vocabulary for various scientific research to bridge the gap between consumers and health experts. Given the flat file format of the Consumer Health Vocabulary dataset, we developed a SKOS-based ontology of the dataset. As an ontology, this dataset can be semantically linked to other resources to provide consumer-level meaning. In addition with this artifact, we plan to further expand the terminology.

Keywords

Taxonomy; Terminology; Biomedical Informatics; Semantic Web

I. INTRODUCTION

Consumer health informatics is defined as “the field devoted to informatics from multiple consumer or patient views” [1]. One important area of research from this domain is the communication gap resulting from misunderstanding of expert level terminology that is communicated to patients. One solution is to link synonymous consumer terms (or “lay” terms) with medical terminology in various modalities and tasks, such as health care applications, medical records, and information retrieval endeavors [2].

In 2006, Zeng and Tse generated the Consumer Health Vocabulary (CHV) that assigns UMLS with similar consumer language [2]. The vocabulary was previously hosted on the University of Utah website, <http://consumerhealthvocab.chpc.utah.edu/CHVwiki/>, and was available in a flat file format. From its inception, it has been an important resource for many previous studies. Currently, the site is no longer available.

Ontologies are expressive terminological artifacts that link domain concepts to data using languages like Web Ontology Language (OWL2) [3] and Resource Description Format (RDF) [4]. Simple Knowledge Organization System (SKOS) is a categorical scheme for controlled vocabularies using RDF to design a standard vocabulary [5]. Essentially, *Concepts* in SKOS represent a thought or idea. The *Concept* class is annotated with alternative labels and has links to other *Concepts* indicating their semantic relationship.

To sustain this work and further enhance the CHV for interoperability, we propose to align the CHV data to SKOS and transform the data to an ontology artifact for reuse and distribution, called the Ontology of Consumer Health Vocabulary (OCHV).

II. METHOD

To create the Ontology of Consumer Health Vocabulary, we used the last updated Consumer Health Vocabulary dataset, dated from 2011. This last version has 158,508 UMLS terms that were either noted as formal medical terms or consumer language, and each term has an associated UMLS CUI identifier.

We developed Java software code to extract the data from CHV and organize it using the SKOS scheme. The software code utilizes a combination of OWL API (v.5.1) [6], Super CSV (2.4) [7], and Google Guava (v.28) [8] libraries to perform this task. We utilized Super CSV to read and extract the data from the flat files. With OWL API, we organized the terms and linked them according to the SKOS categorical scheme, and we also added some sub-classes for the core *Concept* of SKOS.

Figure 1 shows the abstract model to organize the OCHV. Mentioned earlier, we created two subclasses for *Concept* - *CHV Concept* and *UMLS Concept*. Each of these sub-concepts indicate whether the concepts in CHV was a formal UMLS term and/or lay consumer term, and each had a preferred label assigned to each. In addition, we also annotated the alternative labels for CHV Concepts that have supplemental terms associated. Between the *CHV Concepts* and *UMLS Concepts*, there are several concepts that are linked using SKOS' *has related match* property to identify their semantic relatedness.

After the processing procedures described, the ontology is outputted in an OWL2 file format. The Ontology of Consumer Health Vocabulary contains 115,645 classes and 18 properties. This draft version of the Ontology of Consumer Health Vocabulary [9], is available at the National Center for Biomedical Ontology BioPortal.

With the UMLS code and their associated lay term, we can potentially extend various ontologies and terminologies with consumer language, and lead towards resources and tools

that can be consumer-friendly. We intend to continue developing the OCHV, which includes expanding the terminology with new lay terms that may have emerged since 2011.

III. CONCLUSION

We present a representation of the Consumer Health Vocabulary as an ontology called the Ontology of Consumer Health Vocabulary (OCHV). This ontology encodes data from the original CHV dataset that was previously available on <http://consumerhealthvocab.chpc.utah.edu>. Within OCHV, each UMLS term is linked with the consumer language equivalent. With the ever-evolving consumer language, our next goal is to mine consumer corpora for additional consumer terms that can populate OCHV. OCHV is currently hosted on the BioPortal website, <https://bioportal.bioontology.org/ontologies/OCHV>.

Acknowledgment

Research was supported by the National Library of Medicine of the National Institutes of Health under Award Numbers R01LM011829 and R00LM012104, and the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number R01AI130460.

REFERENCES

- [1]. American Medical Informatics Association, “Consumer health informatics,” 2019. [Online]. Available: <https://www.amia.org/applications-informatics/consumer-health-informatics>
- [2]. Zeng QT and Tse T, “Exploring and developing consumer health vocabularies,” *Journal of the American Medical Informatics Association*, vol. 13, no. 1, pp. 24–29, 2006. [PubMed: 16221948]
- [3]. W3C OWL Working Group, “Owl 2 web ontology language document overview (second edition),” 2012. [Online]. Available: <https://www.w3.org/TR/owl2-overview/>
- [4]. Cyganiak R, Wood D, and Lanthaler M, “Rdf 1.1 concepts and abstract syntax,” 2014. [Online]. Available: <https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>
- [5]. Miles A, Matthews B, Wilson M, and Brickley D, “Skos core: simple knowledge organisation for the web,” in *International Conference on Dublin Core and Metadata Applications*, 2005, pp. 3–10.
- [6]. Horridge M and Bechhofer S, “The owl api: A java api for owl ontologies,” *Semantic Web*, vol. 2, no. 1, pp. 11–21, 2011.
- [7]. Graversen KB, “Super CSV,” 2015. [Online]. Available: <https://super-csv.github.io/super-csv/index.html>
- [8]. Bourrillion K and Levy J, “Guava: Google core libraries for java,” 2019. [Online]. Available: <https://github.com/google/guava>
- [9]. Amith M and Tao C, “Ontology of consumer health vocabulary,” 2019. [Online]. Available: <https://bioportal.bioontology.org/ontologies/OCHV>

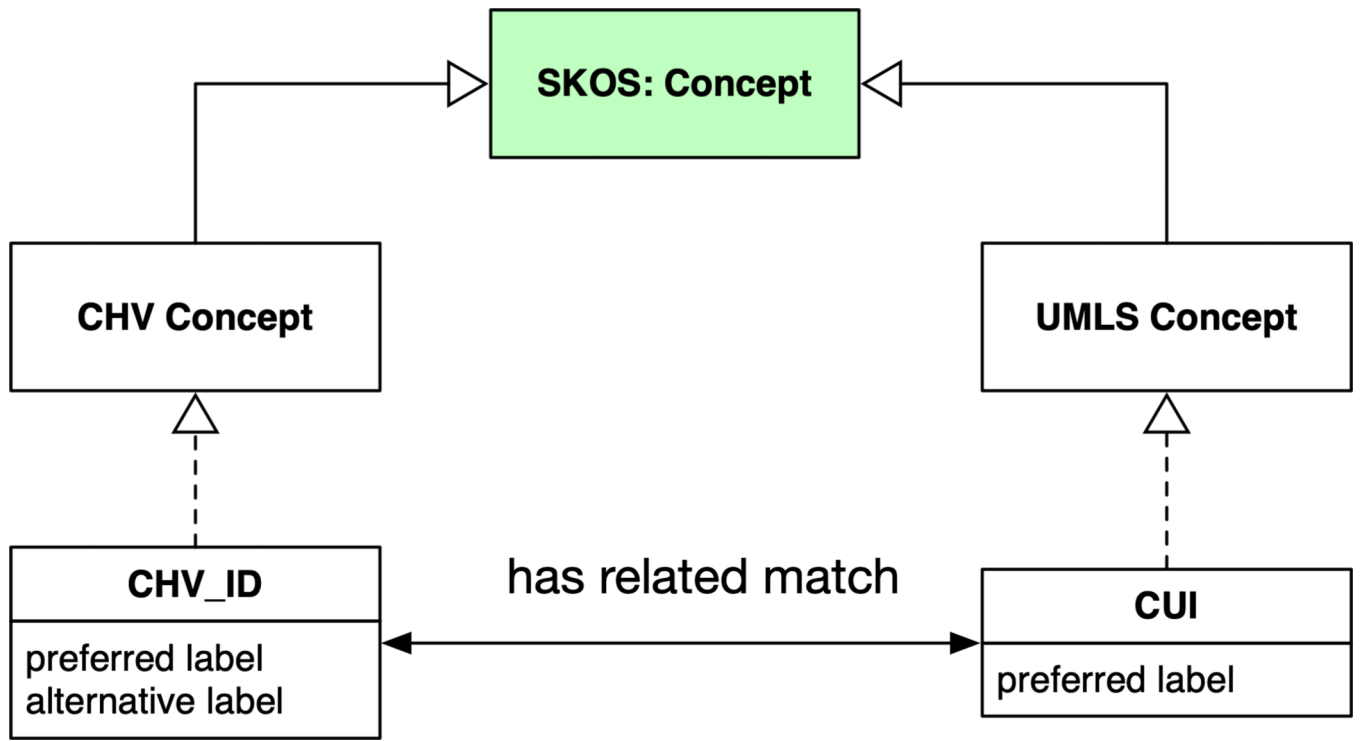


Fig. 1.
Class level structure of the Ontology of Consumer Health Vocabulary (OCHV).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript