DCPAPERS
Proc. Int'l Conf. on Dublin Core and Metadata Applications 2015

Advancing Materials Science Semantic Metadata via HIVE

Yue Zhang
Metadata Research Center
Drexel University, USA
yue.zhang@drexel.edu

Jane Greenberg
Metadata Research Center
Drexel University, USA
janeg@drexel.com

Adrian Ogletree
Metadata Research Center
Drexel University, USA
aogletree @drexel.com

Garritt Tucker
Materials Science and
Engineering Department
Drexel University, USA
gtucker@coe.drexel.edu

Keywords: metadata; materials science; metal; SKOS; HIVE; automatic indexing

This poster reports on the process and the initial results of an ontology for metals developed in Simple Knowledge Organization System (SKOS), a World Wide Web Consortium (W3C) standard, and integrated into the HIVE technology.

1. Materials Science: A Need for Semantic Ontologies

Metadata challenges in the materials science community have surfaced due to national and international data sharing policies and the Material Genome Initiative (NSTC, 2014). Among the more pressing challenges is the need to develop semantic ontologies, given their capacity to support information retrieval and discovery, interoperability, and linking of related resources. Researchers engaged in the Materials Science Metadata Infrastructure Initiative (M^2I^2) are addressing this need by working with the Helping Interdisciplinary Vocabulary Engineering (HIVE) technology in the area of metals.

Materials science is an interdisciplinary field that is advancing the discovery of new materials and enhancing existing materials. Like many interdisciplinary fields, there is a challenge in developing a single semantic ontology due to the breadth of topics that comprise the field. Materials science spans chemistry, engineering, mathematics, and physics, among other disciplines and sub-disciplines. A more productive way to pursue this challenge is to integrate domain-specific vocabularies dynamically when indexing resources. We are exploring this approach at the Metadata Research Center, Drexel University, as part of the M^2I^2 project, working, initially, in the sub-domain of metals.

2. HIVE—Helping Interdisciplinary Vocabulary Engineering

Helping Interdisciplinary Vocabulary Engineering (HIVE) is an automatic Linked Open Data (LOD) technology that integrates interdisciplinary semantic ontologies encoded with the Simple Knowledge Organization System (SKOS), a World Wide Web Consortium (W3C) standard. The integration is dynamic and takes place during an indexing operation. An overview of HIVE's architecture is found on the HIVE Wiki (https://cci.drexel.edu/hivewiki/index.php/Main_Page) and the Code is in GitHub (https://github.com/MetadataResearchCenter/hive-mrc).

3. The Metals Ontology: Methodology

The original corpus of terms for the metals ontology was extracted from a set of Wikipedia pages addressing the topic of metal as a material. We identified and defined terms from this set of pages and established their conceptual relationships by their hyperlinks and categories in Wikipedia. The vocabulary was automatically transformed to the SKOS standard by script,

https://doi.org/10.23106/dcmi.952137227

written in C++. The metals ontology includes 44 concepts. The ontology can be described as a high-level or general controlled vocabulary, and is useful for indexing or topical representation in a digital library or repository.

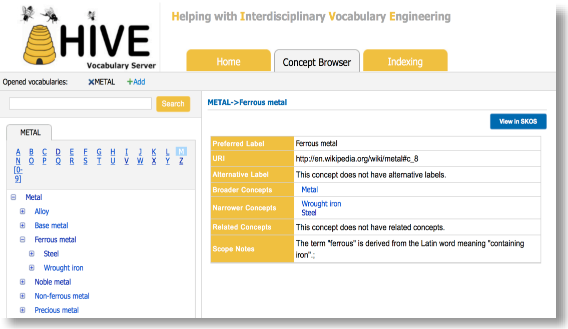

FIG. 1: Ontologies of metals (partial)

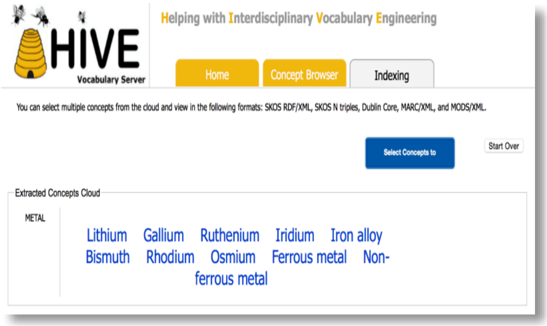

FIG. 2: Sample indexing in materials generated by HIVE

The metals ontology was uploaded to the HIVE demonstration site (http://hive.cci.drexel.edu:8080/home.html). The concept browser provides access to the ontology terms via search, and allows access via an alphabetical list (Figure 1). Each concept has narrower, broader, and related terms. For example, the term base metal has a broader term metal, and has narrower terms including Lead, Iron, Nickel, Cooper and Zinc. The indexing operation allows a user to invoke an automatic sequence by uploading a document (e.g. txt, docx, pdf) or entering a URI, for a digital resource, and then selecting the metal ontologies to add. The text is parsed for meaning and matched against the metals ontology. Figure 2 presents the output from running an automatic indexing of the Wikipedia page on metals (https://en.wikipedia.org/wiki/Metal) through HIVE. A user can then select appropriate ontology terms from the output for indexing the resource.

4. Status and Next Steps

The initial focus has been to develop a basic and high-level metals ontology for HIVE. A current focus is to enhance HIVE's indexing with the metals ontology, via machine learning algorithms such as KEA++ and MAUI (Frank et al., 1999; Witten et al., 1999). We are working with a group of selected articles and keywords assigned by domain experts (the gold standard) to train HIVE in materials science. We are also working more specifically in the area of naoncrystalline metadata to develop an approach for engaging the scientists in ontology development (Greenberg et al, 2015). Our goal is to further develop semantic ontologies from other sub-domains of Materials Science as we grow HIVE; and as part of this work we will continue to investigate users' preferences for ontologies and functionalities of HIVE.

Acknowledgements

We would like to acknowledge the U.S. National Science Foundation, #OCI-0830944.

References

Frank, E., Paynter, G. W., Witten, I. H., Gutwin, C., & Nevill-Manning, C. G. (1999). Domain-specific keyphrase extraction.

Greenberg, J., Zhang, Y., Ogletree, A., Tucker, G. (2015. In press) Threshold Determination and Engaging Materials Scientists in Ontology Design. 9th Metadata and Semantics Research Conference. Manchester, UK, September 9-11, 2015.

https://doi.org/10.23106/dcmi.952137227

National Science and Technology Council (NSTC). (2014). Materials Genome Initiative: Strategic Plan. Retrieved from https://www.whitehouse.gov/sites/default/files/microsites/ostp/NSTC/mgi_strategic_plan_-_dec_2014.pdf

Witten, I. H., Paynter, G. W., Frank, E., Gutwin, C., & Nevill-Manning, C. G. (1999, August). KEA: Practical automatic keyphrase extraction. In Proceedings of the fourth ACM conference on Digital libraries (pp. 254-255). ACM.

https://doi.org/10.23106/dcmi.952137227