Development of a wiki-based, expert community-driven nanosystem vocabulary

Laura M. Bartolo
Kent State University
Tel: +1 330 672 1691
Fax:+ 1 330 672 3816
lbartolo@kent.edu

Cathy S. Lowe
Kent State University
Tel: +1 330 672 0021
Fax:+ 1 330 672 3816
clowe@kent.edu

Sharon C. Glotzer
University of Michigan
Tel: +1 734 615 6296
Fax:+ 1 734 764 7453
sglotzer@umich.edu

Christopher Iacovella
University of Michigan
Tel: +1 734 615 6296
Fax:+ 1 734 764 7453
cri@umich.edu

Abstract:
The NSDL Materials Digital Library Pathway (MatDL) is utilizing a wiki to facilitate the expert community-driven development of a controlled vocabulary on assembly of nanosystems. Vocabulary development is being initiated using preferred subject terms gathered from Dublin Core metadata captured during the execution of nanoscale computer simulation codes of a research group whose members are defining terms and relationships between terms.

Keywords:
Materials science, vocabulary, nanoscience, metadata, wiki.

1. Introduction

Information scientists and materials scientists are working together through the NSDL Materials Digital Library Pathway (MatDL) to capture, in Dublin Core XML format, optimal description of nanoscale computer simulation output as research codes are executed. The long term goal of the work is to enable users, such as research groups and students, to efficiently and effectively manage their results for internal use, for exchange with outside collaborators, for use in educational settings, and for submissions to digital libraries. Part of this goal includes using subject terms to lay the foundation for the development of an expert community-built, wiki-based, web-accessible controlled vocabulary which will serve as a reference resource on assembly of nanosystems for upper level undergraduates as well as for beginning graduate research lab assistants.

A controlled vocabulary serves to reconcile various words used to express a concept and to differentiate among meanings associated with a given word. The use of controlled

vocabulary facilitates more precise information search and retrieval (1). Controlled vocabulary development is often accomplished very successfully with a top-down approach in which a small group carefully develops and subsequently releases the vocabulary to the community. In an emerging field such as nanoscience, it can be difficult for such an approach to keep pace with vocabulary growth and variation. A vocabulary developed in this manner could be out-of-date upon its release. A bottom-up, community-driven solution may offer a better strategy under these circumstances, providing a useful service to the community from the outset which becomes increasingly valuable as the project grows and matures.

Computer-mediated communication technologies involving the Internet such as digital libraries, email, blogs, and wikis, have made it possible to overcome geographical barriers providing new avenues and opportunities for scientific discourse (2) as well as to more effectively harness member expertise to support development of the community knowledgebase and vocabulary (3,4,5).

As defined in Wikipedia (http://en.wikipedia.org), the popular, free, on-line encyclopedia, a wiki is a kind of website that allows users to add, remove, or edit content, very quickly and easily. Wikis are considered to be a low-barrier, easily adoptable method of facilitating communication, collaboration, and content sharing (6, 7). They have been successfully used by scientific research groups to serve a variety of purposes such as creating and maintaining up-to-date, easily accessible documentation and manuals (8) and providing curation functionality to augment a gene regulation data-mining tool (9). Wiki-based technology is also being used as the basis for the development of an ontology engineering tool (10). Some experts within the scientific community have instituted peer-review of wiki content in order to ensure that users have access to trustworthy, authoritative scientific information (11).

This pilot project focuses on using a wiki as a tool to facilitate development of a nanoscale system controlled vocabulary starting with the preferred subject vocabulary gathered from metadata generated by the research codes of one computational nanoscience and soft matter simulation research group. While the wiki is publicly viewable, development of wiki content is restricted to nanoscience experts. Though other, more specialized tools are available for vocabulary development, they tend to require specific technical expertise that can discourage community participation. The primary objective of this project was to gather the vocabulary, definitions, and relationships from domain experts using a method that presented the fewest possible technical barriers for contributors.

2. Method

2.1 Participants

University of Michigan (U-M) nanoscience research group members are piloting and testing this effort as part of their normal workflow. Two U-M classes (i.e., an upper level undergraduate class in molecular engineering and a graduate class in thermodynamics) will also participate in conjuction with their coursework once the wiki is established.

2.2 Software

Free, open source Mediawiki software (http://www.mediawiki.org/wiki/MediaWiki) was chosen as the wiki engine for this project because of factors including, ease of installation, stability, large user and contributor base, as well as functionality such as multiple layers of access control and the ability to render and display LaTex mathematical expressions.

2.3 Procedure

Input file parameters and values for the U-M research group's master simulation code and all of its modules are currently captured as DC metadata in XML format upon execution of the code. The DC metadata transparently describes all associated parameters and values necessary to identify, describe, or recreate the simulation. The metadata includes DC elements such as: title, creator, and date as well as subject terms which are drawn from parameters, e.g., simulation type, code module, and integrator scheme. DC description

includes additional parameter names and associated values such as: number of particles; system number density; system temperature as a function of time; time step for integration of equations of motion; and number of dimensions. Subject terms also provide a pool of preferred terms to begin the development of a community-driven, refereed, web-accessible dictionary, glossary, and thesaurus.

New subject terms are added to the wiki as they appear in the metadata generated along with the research output. Currently, only research group members are permitted to add definitions and links to related terms. Once the wiki becomes established, students as well as external collaborators from other research groups will be incorporated into the workflow. Terms identified in the wiki are augmented with links to concrete example resources available in MatDL. Administrative responsibility and site maintenance resides with nanoscience research group members and information scientists at MatDL.

3. Results & Discussion

Currently, approximately 40 preferred terms representing basic simulation types (e.g., molecular dynamics); interaction types (e.g., Lennard Jones); and analysis routines (e.g., radial distribution function); as well as other system classifications are represented in the wiki along with initial definitions and links to related terms. These definitions will be refined by group members while additional terms are also incorporated into the wiki.

Once established, this vocabulary for assembly of nanosystems should prove to be a valuable resource for graduate lab assistants joining the research group as well as for upper level undergraduates using selected research codes in their coursework. The corpus of preferred terms used by the research group may also serve as starting point for controlled vocabulary development within the broader nanoscience community. It is hoped that as the project gathers momentum, research groups collaborating with the pilot group as well as other groups within the community will also contribute their preferred vocabulary. Variations among preferred vocabularies used in different groups are expected. Wider vocabulary variations are expected among more dissimilar groups which may still wish to collaborate on areas where their work overlaps. For example, researchers taking theoretical and experimental approaches to a similar problem may use very different vocabulary to express the same concepts.

An expert community-driven wiki can collect vocabulary representing differing perspectives on nanoscale system assembly. Information scientists and materials scientists working together can establish associations between vocabulary from various approaches leading to a consensus-based, bottom-up controlled vocabulary. This approach would provide an authoritative resource along with a mechanism to retain vocabulary currency in a new and rapidly evolving field. On a broader scale, the outcomes of this effort would offer ongoing benefits to nanoscience education and research.

4. References:

1. A. Taylor. Verbal Subject Analysis. In The Organization of Information. Englewood, CO: Libraries Unlimited, Inc., 1999, pp. 131-171.
2. P. Genoni, et al. The use of the Internet to activate latent ties in scholarly communities. First Monday, 2005, vol. 10, n. 12.
3. C. Wiederkehr, et al. GermOnline, a cross-species community knowledgebase on germ cell differentiation. Nucleic Acids Research, 2004, vol. 32, Database issue, pp. D560-D567.
4. G. Winsor, et al. Pseudomonas aeruginosa Genome Database and PseudoCAP: facilitating community-based, continually updated, genome annotation. Nucleic Acids Research, 2005, vol. 33, Database issue, pp. D338–D343.
5. C. Friedman, et al. A vocabulary development and visualization tool based on natural language processing and the mining of textual patient reports. Journal of Biomedical Informatics, 2003, vol. 36, pp. 189–201.
6. J. Frumkin. The wiki and the digital library. OCLC Systems & Services: International Digital Library Perspectives, 2005, vol. 21, n. 1, pp.18-22.

7. D. Fichter. Intranets, wikis, blikis, and collaborative working. Online, vol. 29, n. 5, pp. 47-50.
8. I. Sauer, et al. Blogs and wikis are valuable software tools for communication within research groups. Artifical Organs, 2005, vol. 29, n. 1, pp. 82-89.
9. H. Maier, et al. LitMiner and WikiGene: Identifying problem-related key players of gene regulation using publication abstracts. Nucleic Acids Research, 2005, vol. 33, Web Server issue, pp. W779-W782.
10. M. Hepp, et al. OntoWiki: Community-driven ontology engineering and ontology usage based on wikis. Proceedings of International Symposium on Wikis (WikiSym 2005), October 16-18, 2005, San Diego, California, USA.
11. D. Butler. Experts plan to reclaim the web for pop science. Nature, 2006, vol. 439, n. 2, pp. 516-517.