

Ontology-based approach to creating semantic wikis

Ortiz Troncoso, Alvaro

*Museum für Naturkunde - Leibniz Institute for Research on Evolution and Biodiversity,
Invalidenstrasse 43, 10115 Berlin, Germany alvaro.OrtizTroncoso@mfn.berlin*

Received (Day Month Year)

Revised (Day Month Year)

Abstract

Purpose: Maintaining a semantic wiki is challenging. Coping with increasingly complex wikis led to the development of a methodical approach for simplifying the creation and maintenance of semantic wikis -

Approach: The methodical approach used involves modeling the semantic relationships underlying the information in the wiki as an ontology, and then programmatically creating the wiki from the ontology constructs -

Findings: The approach was implemented as open source software that maps ontology constructs to the wiki artifacts, implementation details are discussed. -

Practical implications: Semantic wikis support research at the Museum of Natural History Berlin by providing services for developing and documenting collection management and imaging procedures or annotating archive material. Additionally to these internally available wikis, some wikis are available on the Web, as publication medium for conference presentations or providing additional material for exhibitions. A total of 16 semantic wikis are currently online on the museum's servers. -

Value: A methodical approach for creating and maintaining wikis greatly simplifies these tasks. Furthermore, reusing vocabularies and taxonomies throughout several projects becomes manageable. -

Keywords: semantic web, ontology engineering, collaborative authoring

Paper type: Case study

1. Introduction

Wikis are web platforms allowing for collaborative authoring of information. With certainty, the best known wiki is Wikipedia, but wikis are also being used to setup corporate intranets, private collaboration environments or knowledge bases. A semantic wiki is a wiki whose functionality has been extended to encode information in a structured and machine-readable way: information encoded semantically forms a "consistent logical web of data" [Berners-Lee, T. *et al.* (1998)], that can be used for enriching information with metadata, checking the conformance of information against an agreed upon schema, or accessing and combining data from different

sources [Oren, E. *et al.* (2006)]. Semantic wikis therefore combine the metadata and conformance aspects required by knowledge management, with the possibility to access and build-upon a corpus of collaboratively collected knowledge.

Semantic wikis have been used at the Museum für Naturkunde Berlin (Museum of Natural History Berlin) since 2015. A total of 16 semantic wikis are currently online on the museum's server. Most of these wikis are part of the museum's intranet, and support research by providing services for developing and documenting collection management and imaging procedures or annotating archive material. On the other hand, some wikis are available on the Web, as publication medium for conference presentations or providing additional material for exhibitions. An example of a publicly accessible semantic wiki running at the Museum of Natural History Berlin is the Wiki supplementing the Panda exhibition (http://biowikifarm.net/v-mfn/panda_en/index.php).

The construction and provisioning of wikis has been the subject of two research projects at the Museum [Patzschke, E., Ortiz-Troncoso, A., and Abele, A. (2016)]. The experience gained demonstrates that creating and maintaining a semantic wiki is challenging, a major difficulty being that the complexity of the information stored in the wiki tends to increase as the user base grows, and new ideas are integrated into the wiki's semantic structure [Kiniti, S. and Standing, C. (2013)]. This increase in complexity can take several forms: the number of document types increases, the number of attributes per document increases, external vocabularies are added to the wiki, category trees become more complex. On the other hand, complex information offers more possibilities for gaining new insights, as categories and properties can be combined through semantic queries into new data. For example, querying the biographies of collectors in the wiki of the museum's historical division can be visualized as a map showing where the museum has collected specimens in the past (see [Ortiz Troncoso, A. (2016)] for this and other examples of semantic queries).

2. Methodical approach

Coping with this increasing complexity led to the development of a methodical approach for simplifying the creation and maintenance of semantic wikis. One way to create a semantic wiki is to model the semantic relationships underlying the information in the wiki as an ontology, and then to customize a wiki to reflect the constructs described by the ontology [Di Iorio, A. *et al.* (2006)].

At this point it might be necessary to clarify the concept of ontology. A basic definition of what an ontology is in the realm of knowledge engineering was given by Gruber and Borst (cited in [Corcho, O., Fernández-Lpez, M., and Gómez-Pérez, A. (2003)]): an ontology is made of clear-cut and formal definitions of the concepts in a domain, and these definitions are accepted by the community of practitioners of that domain. Another way to put Gruber and Borst's definition: an ontology is a collaboratively constructed, formalized mind map. Another definition of ontology can be gained from information theory. Pieterse, V. and Kourie, D. G. differentiate

between five types of fundamental information structures: linear lists, hierarchical taxonomies, poly-hierarchical lattices, thesauri combining hierarchies with various kinds of relations between concepts, and ontologies, which extend thesauri with a set of rules for inferring new knowledge. This definition emphasizes the role of logical inference rules, and paves the way for the role ontologies play in machine learning (see Zhou, L. for a road map).

Ontology engineering is a vast topic by itself, yet a simple methodology could follow these basic steps [Noy, N. and McGuinness, D. (2001)]: obtaining domain information by consulting experts and gathering specific vocabularies (e.g. in the case of Museum of Natural History Berlin, zoological taxonomy, stratigraphical vocabulary and others); subsequently, modeling the ontology in an external tool such as Protégé [Musen, M. A. (2015)], and storing the result as an RDF/XML file [Musen, M. A. (2015)]. Software implemented at the Museum automates the conversion of the ontology file into web pages and forms in the wiki. This software is distributed under an Open Source License (<https://github.com/MfN-Berlin/RDF-to-SemanticWiki>), in the hope that it can be useful for computer scientists researching computer-mediated collaboration and ontology engineering.

3. Implementation details

The process by which the conversion of the ontology into a wiki is automated is implemented as follows: the RDF/XML file representing the ontology is parsed into an in-memory semantic model. Then, for each object in the model, a data access object (DAO) is created using a factory class. A connection is established with the wiki's web service (API). Finally, each DAO object saves itself to the wiki, using the wiki's API. With an ontology as input, the software knows two commands: "import" creates the necessary wiki pages, "delete" removes the corresponding wiki pages. The layout of the wiki is customizable through templating: templates can be used to specify the order in which elements should appear, as well as to filter out or hide some elements. The software is written in Python, templates are in XSLT. The resulting wiki pages are compatible with the semantic extension to the popular MediaWiki software (<https://www.semantic-mediawiki.org>).

Ontology constructs are mapped to the wiki artifacts according to these rules: ontology classes are mapped to forms, templates and page categories; data attributes are mapped to property pages of the corresponding data type; relations between objects are rendered as links between pages. The cardinality of attributes is enforced through appropriate form input fields. Class inheritance is supported. Furthermore, navigation elements are added to the wiki, to enable creating, listing and editing content following the structure defined by the ontology. The information stored in the wiki using these forms and templates is thus structured and consistent, and can be accessed using semantic queries (<https://www.semantic-mediawiki.org>).

4. Conclusions and future work

The method presented here and the associated software support research at the Museum of Natural History Berlin by providing a way to create collaboration environments for working on complex scientific projects, while enforcing structure and consistency. Conceivably, the method and software could be applied to other research institutions and even to areas of industry that face similar knowledge management challenges. Furthermore, as ontologies can be reused and extended [Noy, N. and McGuinness, D. (2001)], the methodical approach facilitates knowledge transfer between projects, as taxonomies and vocabularies found useful in one project can be used to construct a wiki for a second project. As discussed above, the implemented software supports the conversion of semantic classes, attributes and relations, class inheritance and cardinality. This is a subset of the constructs defined by the RDF/XML specification [Musen, M. A. (2015)]. More work is needed to extend the possibilities of the software so as to pave the way for more expressive collaboration environments.

5. Acknowledgements

Software development at Museum für Naturkunde Berlin was sustained in part by the project “Knowledge Transfer Concept for Research Contents, Methods and Competences in Research Museums”, funded by the German Federal Ministry of Education and Research (BMBF), grant no. 01IO1632.

References

References

- Berners-Lee, T. et al. (1998). Semantic web road map.
- Corcho, O., Fernández-Lpez, M., and Gmez-Prez, A. (2003). Methodologies, tools and languages for building ontologies. where is their meeting point? *Data & knowledge engineering*, 46(1):4164.
- Di Iorio, A., Fabbri, M., Presutti, V., and Vitali, F. (2006). Automatic deployment of semantic wikis: a prototype. In *SemWiki*.
- Gandon, F. and Schreiber, G. (2014). Rdf 1.1 xml syntax. [Online; accessed 1-July-2019].
- Kiniti, S. and Standing, C. (2013). Wikis as knowledge management systems: issues and challenges. *Journal of Systems and Information Technology*, 15(2):189201.
- Musen, M. A. (2015). The protg project: a look back and a look forward. *AI Matters*, 1(4):412.
- Noy, N. and McGuinness, D. (2001). Ontology development 101: A guide to creating your first ontology.
- Oren, E., Miller, K., Scerri, S., Handschuh, S., and Sintek, M. (2006). What are semantic annotations. *Relatorio tecnico*. DERI Galway, 9:62.
- Ortiz-Troncoso, A. (2016). Erfahrungen aus dem Projekt Wiki-Ansatz und kollaboratives Arbeiten im Forschungsmuseum, *Citizen Science-Wiki-Workshop 08.04.16 Ludwig-Maximilians-Universität München*, DOI: 10.13140/RG.2.1.1597.2247
- Patzschke, E., Ortiz-Troncoso, A., and Abele, A. (2016). Wiki-ansatz und kollaboratives arbeiten im forschungsmuseum: Schlussbericht.

- Pieterse, V. and Kourie, D. G. (2014). Lists, taxonomies, lattices, thesauri and ontologies: paving a pathway through a terminological jungle. *KO Knowledge Organization*, 41(3):217229.
- Zhou, L. (2007). Ontology learning: state of the art and open issues. *Information Technology and Management*, 8(3):241252.

Corresponding author

Ortiz Troncoso, Alvaro can be contacted at: alvaro.OrtizTroncoso@mfh.berlin.