Ontology-based approach to creating semantic wikis

Alvaro R. Ortiz Troncoso

Museum für Naturkunde - Leibniz Institute for Research on Evolution and Biodiversity,

alvaro.OrtizTroncoso@mfn.berlin

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. 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. A methodical approach for creating and maintaining wikis greatly simplifies the creation and maintenance of semantic wikis. Furthermore, reusing vocabularies and taxonomies throughout several projects becomes manageable. The approach was implemented as open source software that maps ontology constructs to the wiki artifacts, practical applications are discussed.

1 INTRODUCTION

Wikis are web platforms allowing for collaborative authoring of information. With certainty, the best-known wiki is Wikipedia, but wikis are also used to setup corporate intranets, collaborative authoring 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 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 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 servers. Most of these wikis are part of the museum's intranet, and support research by providing services for developing and documenting collection management 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[1].

The construction and provisioning of wikis has been the subject of two research projects at the Museum[2]. 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 and Standing, 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 and poly-hierarchic. 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 information. 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[3].

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 et al., 2006].

Gruber and Borst (cited in [Corcho et al., 2003]) give a basic definition of what an ontology is in the realm of knowledge engineering: an ontology is made of clear-cut and formal definitions of the concepts in a domain, and the community of practitioners of that domain accepts these definitions. Another definition of ontology can be gained from information theory: Pieterse and Kourie, [2014] 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 (table 1). This definition emphasizes the role of logical inference rules, and paves the way for the role ontologies play in machine learning.

Ontology engineering is a vast topic by itself, yet a methodical approach could be described as following these basic steps [Noy and McGuinness, 2001]: obtaining domain information by consulting experts and gathering specific vocabularies; subsequently, modeling the ontology in an external tool such as Protégé [Musen, 2015], and storing the result as an RDF/XML file [Gandon, Fabien and Schreiber, 2014].

---

1 http://biowikifarm.net/v-mfn/panda_en/index.php

2 https://doi.org/10.2314/GBV:872479447

3 https://doi.org/10.13140/RG.2.1.1597.2247

Table 1: Information systems, adapted from [Pieterse and Kourie, 2014]

Type	Definition	Example
List	A linearly organized collection of entries	Catalog, dictionary, index
Taxonomy	A hierarchically organized collection	Biological taxonomy
Lattice	A hierarchically organized, searchable collection. Poly-hierarchies are supported.	Recommendation systems
Thesaurus	A collection of entries on a certain domain. Entries are organized (poly-) hierarchically. Thesauri can represent relations of equivalence, association, opposition etc. (e.g. synonyms and antonyms)	Knowledge bases
Ontology	A thesaurus enhanced with inference rules, enabling automatic inferences.	Lexicons (including grammar)

Software Implementation details

The software described here automates the conversion of the ontology RDF/XML file into web pages and forms in the wiki. This software is distributed under an Open Source License, in the hope that it can be useful for computer scientists researching computer-mediated collaboration and ontology engineering[4].

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. Finally, the model is stored in 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 organize elements into tabs, menus and provide help texts. 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[5].

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 that follow the standard API provided by the wiki. An example of the generated wiki page can be seen in figure 1.

---

4 https://github.com/MfN-Berlin/RDF-to-SemanticWiki

5 https://www.semantic-mediawiki.org

Figure 1: A semantic MediaWiki form for describing a doctoral dissertation, rendered from a specification written in RDF/XML.

PRACTICAL APPLICATION AND PROSPECT

An ontology for describing research at the Museum of Natural History Berlin was developed in collaboration with the Human-Centered Computing lab at the Freie Universität Berlin [Benjamin et al., 2019]. The resulting ontology[6], based on the classification system of the German Research Foundation (DFG), was edited and enhanced with controlled vocabularies used by researchers at the museum. A wiki describing research projects, doctoral dissertations, museum collections and laboratories was created from this ontology and is currently in use as an internal resource on the Museum's intranet. The wiki currently holds data on 178 research projects, 11 doctoral dissertations, 9 laboratories and 32 collections. Entries are classified using 282 categories and 218 semantic attributes. Information is presented and organized using 36 semantic queries.

The method discussed here and the associated software support research at the Museum of Natural History Berlin by providing a way to create and manage a collaboration environment for working on complex scientific projects, while enforcing structure and consistency. Conceivably, the method and software could be applied by other research institutions and even by areas of industry that face similar knowledge management challenges. Furthermore, as ontologies can be reused and extended [Noy and McGuinness, 2001], the methodical approach facilitates knowledge transfer between projects, as taxonomies and vocabularies found useful in one project can be used to structure the work of further projects.

---

6 https://via.naturkundemuseum.berlin/ontologies/ikon.rdf

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, 2015]. More work is needed to extend the possibilities of the software to pave the way for more expressive collaboration environments.

ACKNOWLEDGMENTS

REFERENCES

Benjamin, J.J., Müller-Birn, C. and Kinkeldey, C., 2019. Understanding knowledge transfer activities at a research institution through semi-structured interviews. (Technical Report, available through http://refubium.fu-berlin.de)

Berners-Lee, T., 1998. Semantic web road map.

Corcho, O., Fernández-López, M. and Gómez-Pérez, A., 2003. Methodologies, tools and languages for building ontologies. Where is their meeting point?. *Data & knowledge engineering*, *46*(1), pp.41-64.

Di Iorio, A., Fabbri, M., Presutti, V. and Vitali, F., 2006. Automatic Deployment of Semantic Wikis: a Prototype. In *SemWiki*.

Gandon, F. and Schreiber, A.T., 2014. Rdf 1.1 xml syntax.

Kiniti, S. and Standing, C., 2013. Wikis as knowledge management systems: issues and challenges. *Journal of Systems and Information Technology*.

Musen, M.A., 2015. The protégé project: a look back and a look forward. *AI matters*, *1*(4), pp.4-12.

Noy, N.F. and McGuinness, D.L., 2001. Ontology development 101: A guide to creating your first ontology.

Oren, E., Möller, K., Scerri, S., Handschuh, S. and Sintek, M., 2006. What are semantic annotations. *Relatório técnico. DERI Galway*, *9*, p.62.

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), pp.217-229.