OntoWiki: Community-driven Ontology Engineering and Ontology Usage based on Wikis

Martin Hepp
University of Innsbruck
Digital Enterprise Research Institute
Innsbruck, Austria
martin.hepp@deri.org

Daniel Bachlechner
University of Innsbruck
Digital Enterprise Research Institute
Innsbruck, Austria
daniel.bachlechner@deri.org

Katharina Siorpaes
University of Innsbruck
Digital Enterprise Research Institute
Innsbruck, Austria
katharina.siorpaes@deri.org

ABSTRACT
Ontologies are consensual representations of a domain of discourse and the backbone of the future Semantic Web. Currently, however, only a fraction of Web users can take part in the process of building ontologies. In this paper, we show that standard Wiki technology can be easily used as an ontology development environment, reducing entry barriers for the participation of users in the creation and maintenance of ontologies, and describe our first OntoWiki prototype.

Categories and Subject Descriptors
H.5.3 **[Information Interfaces and Presentation]:** Group and Organization Interfaces – *collaborative computing, computer-supported cooperative work, Web-based interaction;* I.2.4 **[Artificial Intelligence]:** Knowledge Representation Formalisms and Methods – *Representation languages, Representations, Semantic networks;* E.2 **[Data]:** Data Storage Representations.

General Terms
Management, Documentation, Human Factors, Languages

Keywords
Ontologies, Wiki, Semantic Web, Community-driven Ontology Engineering, RDF, RDF-S, OWL

1. INTRODUCTION
Ontologies are explicit conceptualizations of a domain of discourse [6, 7]. In short, they are unambiguous representations of concepts, relationships between concepts (for example, but not limited to, a hierarchy), instances, and axioms. Unambiguous in this sense means two things: First, the representation should allow humans to precisely grasp the meaning of any element, so that humans have a well-defined vocabulary at hand when annotating data, expressing queries, or drawing conclusions. Second, the representation should have a formal semantics, so that it supports machine reasoning. For a comprehensive overview, see [5]. However, it is important to note that ontologies are not just formal representations of a domain, but much more *community contracts* about such formal representations. Since a discourse is a dynamic social process, during which previous propositions are often modified, especially refined, or discarded, and new topics need to be added, such a community contract cannot be static, but must be able to reflect the community consensus at any point in time.

Currently, both ontology tools and ontology languages impose high entrance barriers for potential users. This likely contributes to the fact that the most popular approach of creating ontologies is engineering-oriented, i.e., a small number of individuals carefully constructs the representation of the domain of discourse, and releases the results at some point in time to a wider community of users. However, (1) the sequential paradigm of this approach and (2) the fact that a small group constructs the ontology for a bigger group has several weaknesses:

First, the *ontology evolution is not under the full control of the ontology user community*. For example, missing entries cannot be added by any user who reveals the need for a new concept, but has to be added by the small group of creators. In natural language, in comparison, the evolution of the vocabulary is under the control of the user community. Anybody can invent and define a new word or concept in the course of communications.

Second, ontology users cannot easily grasp the intension of a concept. Somebody using an ontology e.g. for annotating instances or expressing queries has little help in determining whether a given concept is suitable for his or her needs, since the formal part of the ontology only *constrains* the interpretation of a concept, but does, with the exception of very expressive ontologies, not actually *define* the meaning of this concept. This leaves the ontology user with sparse natural-language descriptions, e.g. in the form of the Dublin Core field `dc:description`. Such is hardly sufficient to check whether the ontology creators read the concept in the same manner as the potential ontology user does.

We propose to directly use the infrastructure and culture of Wikis as an ontology engineering workbench, in the sense that anybody can add a new element to the ontology, and refine or modify existing ones. We especially suggest the use of multimedia elements to improve the richness and disambiguity of informal concept definitions in an ontology. Also, we regard it as beneficial if the definition of a concept is not separated from the discussion that lead to shaping the intension of this concept, since the history of a conceptualization is a valuable part of the respective definition.

2. THE ONTOWIKI IDEA
Ontologies can have a varying degree of expressivity, ranging from flat collections of consensual concepts to abundantly axiomatized models. Many ontologies have a subsumption hierarchy that allows to infer implicit class membership, but this is not a mandatory property. In its least expressive form, an ontology is a collection of named concepts with a natural language definition of their meaning, i.e. a controlled vocabulary.

Though more expressive ontologies support more sophisticated reasoning, even such flat ontologies can be very useful. Already having unique identifiers (e.g. URIs) assigned to concepts described in natural language is beneficial. Now, we can

observe on one hand that there are very few real ontologies available; a large share of ontologies published on the Web are outdated, dead collections created in some academic research context. On the other hand, the English version of Wikipedia contains more than 726,000 entries, which means it holds unique identifiers for 726,000 concepts.

Our basic idea is to use a Wiki as a mechanism so that a community can create an URI for any needed concept, describe the concept using natural language, refine and modify the definition, and link this approach with the wealth of concepts already defined in Wikipedia. The main paradigm of our work is simplicity, i.e. we want to support only as much functionality as can be used productively by a large share of the community.

3. RESEARCH CHALLENGES

When using Wiki technology for ontology engineering, we see the following research challenges.

Selection of a Proper Ontology Meta-Model: We have to define an ontology meta-model that is suitable for a large audience. In our current approach, we support (1) classes, (2) instances, and (3) relationships in the sense of RDF predicates. Relationships can be either data type properties, pointing to literal values using standard XSD data types, or object properties that point to an instance of any other class. In a sense, this model is very close to plain RDF. At this point in time, we completely leave out any kind of hierarchical order.

Handling of Conceptual Change: A standard Wiki already provides all functionality necessary to create a textual definition and a unique URI. For example, anybody could have added an entry for the Republic of Austria to Wikipedia, now available at

http://en.wikipedia.org/wiki/Austria.

We could immediately use this mechanism and propose to re-use this URI not only as the resource locator for retrieval of the description, but also as the identifier for the concept "Republic of Austria". Now the problem is that since everybody can alter the text, we never know whether the current version is a monotonic extension of any previous version. So anybody who used this URI for the annotation of instances or any other statement might find that his statement no longer holds with the modified version. We propose a very hands-on solution, based on a combination of the "history" functionality in the MediaWiki distribution, and a versioning scheme embedded in the URI for concepts, same as used by the W3C for W3C documents or the WSMO, WSMX, and WSML working groups [3]. The main idea is that the general URI (e.g. http://en.wikipedia.org/wiki/Austria) always refers to the latest version, while all intermediate versions have an additional URI of their own, e.g. by adding the date and time of creation (plus probably the IP address of the originator). In MediaWiki and thus Wikipedia, all intermediate versions already have unique identifiers. This allows referring either to the latest version or to any specific version. It also makes it possible to create statements *about* a specific version.

Helper Functionality for Collaborative Ontology Building: The success of Wikis would have been impossible without the existence of many small but effective scripts that help the community build and maintain the corpus of knowledge. We are currently evaluating various helper scripts for OntoWiki.

Partitioning: If two communities have conflicting views on the same subject, or if two communities use similar terms for different concepts, it would be helpful to partition the space of Wiki entries in subspaces. We are currently implementing this by allowing a user to propose a new ontology, which works like a subdirectory and is reflected in the URI of all concepts.

This is obviously a preliminary list of research challenges, and we are striving towards a more comprehensive understanding of the associated problems.

4. DISCUSSION AND CONCLUSION

We have proposed using standard Wiki technology as the platform for community-driven ontology building and maintenance. In the current stage, this provides only support for creating URIs plus human-readable definitions (using text and graphics) classes, instances, and relationships. This does not provide any reasoning support, but it offers the opportunity for a large user community to establish unique identifiers for needed concepts, and thus more current and more complete ontologies. Our future research will focus on how this skeleton can be extended towards a richer ontology meta-model without introducing new entry barriers for users. We are also evaluating how previous works can be included. There is already significant literature about collaborative ontology engineering in general, e.g. Tadzebao and WebOnto (see [4]). [1] describe collaborative ontology building in analogy to Wikis, but do not borrow more from the Wiki community than the pure name. Platypus Wiki [2] is a Wiki augmented by Semantic Web approaches, namely RDF, while we want to use Wikis for creating ontologies that can be used anywhere in the Semantic Web.

Acknowledgements: The work presented in this paper is partly funded by the European Commission under the project DIP (FP6-507483) and the Trans IT Entwicklungs- und Transfercenter at the University of Innsbruck.

5. REFERENCES

[1] Bao, J. and Honavar, V., Collaborative Ontology Building with Wiki@nt. A multi-agent based ontology building environment. In: Proceedings of the 3rd International Workshop on Evaluation of Ontology-based Tools (EON2004), (Hiroshima, Japan, 2004), 1-10.

[2] Campanini, S.E., Castagna, P. and Tazzoli, R., Platypus Wiki: a Semantic Wiki Wiki Web. In: Proceedings of the 1st Italian Semantic Web Workshop Semantic Web Applications and Perspectives (SWAP), (Ancona, Italy, 2004), 1-6.

[3] de Bruijn, J. and Kopecký, J. Persistent URIs for WSMO and WSML deliverables. WSMO Note 2 February 2005. http://www.wsmo.org/TR/NOTE-URIs/20050202/.

[4] Domingue, J., Tadzebao and WebOnto: Discussing, Browsing, and Editing Ontologies on the Web. In: Proceedings of the 11th Knowledge Acquisition for Knowledge-Based Systems Workshop, (Banff, Canada, 1998), 1-20.

[5] Fensel, D. Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce. Springer, Berlin etc., 2004.

[6] Gruber, T.R. Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies, 43, 907-928.

[7] Guarino, N., Formal Ontology and Information Systems. In: Proceedings of the FOIS '98, (Trento, Italy, 1998), 3-15.