BIOINFORMATICS

Vol. 00 no. 00 2008
Pages 1–4

BOWiki: An ontology-based wiki for annotation of data and integration of knowledge in biology

Robert Hoehndorf,[a,c,d] Joshua Bacher,[b,c] Michael Backhaus,[a,c,d] Sergio E. Gregorio, Jr.,[e] Frank Loebe,[a,d] Kay Prüfer,[c] Alexandr Uciteli,[c] Johann Visagie,[c] Heinrich Herre [a,d] and Janet Kelso [c]

[a]Department of Computer Science, Faculty of Mathematics and Computer Science, University of Leipzig, Johannisgasse 26, 04103 Leipzig, Germany; [b]Institute for Logics and Philosophy of Science, Faculty of Social Science and Philosophy, University of Leipzig, Beethovenstrasse 15, 04107 Leipzig, Germany; [c]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany; [d]Research Group Ontologies in Medicine (Onto-Med), Institute of Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Härtelstrasse 16–18, 04107 Leipzig, Germany; [e]Communications and Publications Services, International Rice Research Institute, College, 4030 Los Baños, Laguna, Philippines

ABSTRACT

Ontology development and the annotation of biological data using ontologies are time-consuming exercises that currently requires input from expert curators. Open, collaborative platforms for biological data annotation enable the wider scientific community to become involved in developing and maintaining such resources. However, this openness raises concerns regarding the quality and correctness of the information added to these knowledge bases. The combination of a collaborative web-based platform with logic-based approaches and Semantic Web technology can be used to address some of these challenges and concerns.

We have developed the BOWiki, a web-based system that includes a biological core ontology. The core ontology provides background knowledge about biological types and relations. Against this background, an automated reasoner assesses the consistency of new information added to the knowledge base. The system provides a platform for research communities to collaboratively integrate information and annotate data.

The BOWiki and supplementary material is available at `http://www.bowiki.net/`. The source code is available under the GNU GPL from `http://onto.eva.mpg.de/trac/BoWiki`.

Contact: bowiki-users@lists.informatik.uni-leipzig.de

1 INTRODUCTION

Biological ontologies have been developed for a number of domains, including cell structure, organisms, biological sequences, biological processes, functions and relationships. These ontologies are increasingly being applied to describe biological knowledge. Annotating biological data with ontological categories provides an explicit description of specific features of the data, which enables users to integrate, query and reuse the data in ways previously not possible, thereby significantly increasing the data's value.

Developing and maintaining these ontologies requires manual creation, deletion and correction of concepts and their definitions within the ontology, as well as annotating biological data to concepts from the ontology. In order to overcome the arising acquisition bottleneck, several authors suggest using community-based tools such as wikis for the description, discussion and annotation of the functions of genes and gene products [Wang, 2006, Hoehndorf et al., 2006, Giles, 2007].

However, an open approach like wikis frequently raises concerns regarding the quality of the information captured. The information represented in the wiki should adhere to particular quality criteria such as internal consistency (the wiki content does not contain contradictory information) and consistency with biological background knowledge (the wiki content should be semantically correct). To address some of these concerns, logic-based tools can be employed.

We have developed the BOWiki, a wiki system that uses a core ontology together with an automated reasoner to maintain a consistent knowledge base. It is specifically targeted at small- to medium-sized communities.

2 SYSTEM DESCRIPTION

The BOWiki is a semantic wiki based on the MediaWiki[1] software. In addition to the text-centered collaborative environment common to all wikis, a semantic wiki provides the user with an interface for entering structured data [Krötzsch et al., 2007]. This structured data can be used subsequently to query the data collection. For instance, *inline queries* [Krötzsch et al., 2007] can be added to the source code of a wikipage, which will always produce an up-to-date list of results on a wikipage.

The BOWiki significantly extends the MediaWiki's capabilities. It allows users to characterize the entities specified by wikipages as *instances* of ontological categories, to *define new relations* within the wiki, to *interrelate* wikipages, and to *query* for wikipages satisfying some criteria. In particular, the BOWiki provides features beyond those offered by common wiki systems (for details see the Implementation section and table 1): typing wikipages (table 1), *n*-ary semantic relations among wikipages (table 1), semantic

[1] `http://www.mediawiki.org`

search (special page, inline queries), reasoner support for content verification, adaptability to an application domain, import of bio-ontologies for local accessibility and simple reuse, graphical ontology browsing and OWL [McGuinness and van Harmelen, 2004] export of the wiki content.

We consider both adaptability to the application domain and content verification as the BOWiki's two most outstanding novel features. Adaptability means that during setup, the software reads an OWL ontology selected by the user that provides a type system for the wikipages and the relations that are available to connect them. New relations can be introduced using specific wiki syntax, while the types remain fixed after setup.

While semantic wikis allow for the structured representation of information, they often provide little or no quality control, and do not verify the consistency of captured knowledge. Using the imported ontology as a type system in the BOWiki enforces the use of a common conceptualization and provides additional background knowledge about the selected domain. This background knowledge is used to check user-entered, semantic content by means of an OWL reasoner. For example, the ontology can prevent typing an instance of p45 both with *Protein* and *DNA molecule* at the same time. Currently, the performance of automated reasoners remains a limiting factor. Nevertheless, the reasoner delivers a form of quality control for the BOWiki content that should be adopted wherever possible.

The BOWiki was primarily designed to describe biological data using ontologies. In conjunction with a biological core ontology [Valente and Breuker, 1996] like GFO-Bio [Hoehndorf et al., 2007] or BioTop [Schulz et al., 2006], the BOWiki can be used to describe biological data. For this purpose, we developed a module that allows OBO flatfiles[2] to be imported into the BOWiki. By default, these ontologies are only accessible for reading; they are neither editable nor considered in the BOWiki's reasoning. Users can then create wikipages containing information about biological entities, and describe the entities both in natural language text and in a formally structured way. For the latter, they can relate the described entities to categories from the OBO ontologies, and these categories are then made available for use by the BOWiki reasoning.

In contrast to annotating data with ontological categories, i.e., asserting an undefined association relation between a biological datum and an ontological category, it is possible in the BOWiki to define precisely the relation between a biological entity (e.g. a class of proteins) and another category: a protein may not only be *annotated* to *transcription factor activity*, *nucleus*, *sugar transport* and *glucose*. In the BOWiki, it may stand in the *has_function* relation to transcription factor activity; it can be *located_at* a nucleus; it can *participate_in* a *sugar transport* process; it can *bind* glucose. The ability to make these relations explicit renders annotations in a semantic wiki both exceptionally powerful and precise.

The BOWiki can be used to describe not only data, but also biological categories, or to create relations between biological categories. As such, the BOWiki could be used to create so-called cross-products [Smith et al., 2007] between different ontologies.

[2] `http://www.cs.man.ac.uk/~horrocks/obo/`

3 IMPLEMENTATION

Within our MediaWiki extension, users can specify the type of entity described by a wikipage (see table 1). One of the central ideas of the BOWiki is to provide a pre-defined set of types and relations (and corresponding restrictions among them). We deliver the BOWiki with the biological core ontology GFO-Bio, but any *consistent* OWL [McGuinness and van Harmelen, 2004] file can be imported as the type system. Types are modelled as OWL classes and binary relations as OWL properties. Relations of higher arity are modeled according to use case 3 in [Noy and Rector, 2006], i.e., as classes whose individuals model relation instances. Wikipages as (descriptions of) instances of types give rise to OWL individuals, which may be members of OWL classes (their types).

An OWL ontology can provide background knowledge about a domain in the form of axioms that restrict the basic types and relations within the domain. This allows for automatic verification of parts of the semantic content created in the BOWiki: users may introduce a new page in the wiki and describe some entity; they may then add type information about the described entity; and this added type information is then automatically verified. The verification checks the logical consistency of the BOWiki's content – as OWL individuals and relations among them – with the restrictions of the OWL ontology's types and relations, like those in GFO-Bio. The BOWiki uses a description logic [Baader et al., 2003] reasoner to perform these consistency checks. We implemented the BOWikiServer, a stand-alone server that provides access to a description logic reasoner using the Jena 2 Semantic Web Framework [Carroll et al., 2003] and a custom-developed protocol. A schema of the BOWiki's architecture is illustrated in figure 1.

Whenever a user edits a wikipage in the BOWiki, the consistency of the changes with respect to the core ontology is verified using the BOWikiServer. Only consistent changes are permitted. In the event of an inconsistency, an explanation for the inconsistency is given, and no change is made until the user resolves the inconsistency.

In addition to verifying the consistency of newly added knowledge, the BOWikiServer can perform complex queries over the data contained within the wiki. Queries are performed as retrieval operations for description logic concepts [Baader et al., 2003], i.e., as queries for all individuals that satisfy a description logic concept description.

A performance evaluation of our implementation using the Pellet description logic reasoner [Sirin and Parsia, 2004] for ontology classification showed, that presently, only small- to medium-sized wiki installations can be supported. The time needed for consistency checks increases as the number of wiki pages increases[3].

4 DISCUSSION

Using different reasoners

The BOWikiServer provides a layer of abstraction between the description logic reasoner and the BOWiki. Depending on the description logic reasoner used, different features can be supported. Currently, the BOWikiServer uses the Pellet reasoner [Sirin and Parsia, 2004]. Pellet supports the explanation of inconsistencies,

[3] The results of our performance tests can be found on the wiki at `http://bowiki.net`.

BOWiki syntax	OWL abstract syntax
Generic	
1 `[[OType:C]]`	Individual(**page** type(C))
2 `[[R::page2]]`	Individual(**page** value(R page2))
3 `[[R::role1=page1;...;roleN=pageN]]`	Individual(R-id type(R)) Individual(R-id value(subject **page**)) Individual(R-id value(R-role1 page1)) ... Individual(R-id value(R-roleN pageN))
4 `[[has-argument::` `name=roleName;type=OType:C]]`	SubClassOf(**page** gfo:Relator) ObjectProperty(R-roleName domain(**page**) range(C))
Examples	
1 on page Apoptosis: `[[OType:Category]]`	Individual(Apoptosis, type(Category))
2 on page Apoptosis: `[[CC-isa::Biological_process]]`	Individual(Apoptosis value(CC-isa Biological_process))

3 on page HvSUT2:
`[[Realizes::`
`function=Sugar_transporter_activity;`
`process=Glucose_transport]]`
	Individual(Realizes-0 type(Realizes))
	Individual(Realizes-0 value(Realizes-subject HvSUT2))
	Individual(Realizes-0 value(Realizes-function Sugar_transporter_activity))
	Individual(Realizes-0 value(Realizes-process Glucose_transport))

4 on page Realizes:
`[[has-argument::`
`name=function;`
`type=OType:Function_category]]`
	SubClassOf(Realizes gfo:Relator))
	ObjectProperty(Realizes-function domain(Function_category))

Table 1. Syntax and semantics of the BOWiki extensions. The table shows the syntax constructs used in the BOWiki for semantic markup. The second column provides a translation into OWL. (**page** refers to the wikipage in which the statement appears; "R-id" is a name for an individual whose "id" part is unique and generated automatically for the occurrence of the statement). Because OWL has a model-theoretic semantics, this translation yields a semantics for the BOWiki syntax. In the lower half of the table we illustrate each construct with an example and present its particular translation to OWL.

which can be shown to users to help them in correcting inconsistent statements submitted to the BOWiki. It also supports the nonmonotonic description logic ALCK with the auto-epistemic **K** operator [Donini et al., 1997]. This permits both open- and closed-world reasoning [Reiter, 1980] to be combined, which has several practical applications in the Semantic Web [Grimm and Motik, 2005] and the integration of ontologies in biology [Hoehndorf et al., 2007]. On the other hand, reasoning in the OWL description logic fragment [McGuinness and van Harmelen, 2004] is highly complex. It is possible to use reasoners for weaker logics to overcome the performance limitations encountered with Pellet.

Comparison with other approaches

WikiProteins [Giles, 2007] is a software project based also on the MediaWiki software, focused on annotating Swissprot [Boeckmann et al., 2003]. Similar to the BOWiki, it utilizes ontologies like the Gene Ontology [Ashburner et al., 2000] and the Unified Medical Language System [Humphreys et al., 1998] as a foundation for the annotation. It is generally more targeted at creating and collecting definitions for terms than on formalizing knowledge in a logic-based and ontologically founded framework. As a result, it contains a mashup of lexical, terminological and ontological information. In addition, WikiProteins neither supports *n*-ary relations nor provides a description logic reasoner to retrieve or verify information. It therefore lacks the quality control and retrieval features that are central to the BOWiki. On the other hand, because of the different use-cases that WikiProteins supports, it is designed to handle much larger quantities of data than the BOWiki, and it is better suited for creating and managing terminological data.

The Semantic Mediawiki [Krötzsch et al., 2007] is another semantic wiki based on the Mediawiki software. It is designed to be applicable within the online encyclopedia Wikipedia. Because of the large number of Wikipedia users, performance and scalability requirements are much more important for the Semantic Mediawiki than for the BOWiki. Therefore, it also provides neither a description logic reasoner nor ontologies for content verification.

The IkeWiki [Schaffert et al., 2006], like the BOWiki, includes the Pellet description logic reasoner for classification and verification of consistency. In contrast to the BOWiki, parts of the IkeWiki's functionality require users to be experts in either Semantic Web technology or knowledge engineering. As a consequence, the BOWiki lacks some of the functionality that the IkeWiki provides (such as creating and modifying OWL classes) as it targets biologist users, most of whom are not trained in knowledge engineering.

Conclusion

We developed the BOWiki as a semantic wiki specifically designed to capture knowledge within the biological and medical domains. It has several features that distinguish it from other semantic wikis and from similarly targeted projects in biomedicine, most notably its ability to verify its semantic content for consistency with respect to background knowledge and its ability to access external OBO ontologies.

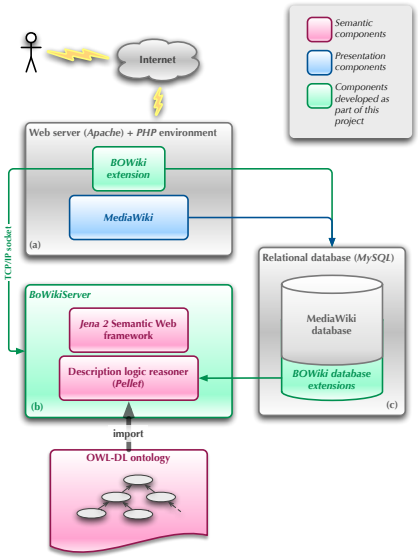

Fig. 1: BOWiki Architecture. (a) The BOWiki extension to the MediaWiki software processes the semantic data added to wiki pages. The semantic data is subsequently transferred to the BOWikiServer using a TCP/IP connection. (b) To evaluate newly entered data or semantic queries, the BOWikiServer requires an ontology in OWL-DL format (provided during installation of the BOWiki). Consistent semantic data will be stored. If an inconsistency is detected, the edited page is rejected with an explanation of the inconsistency. The BOWikiServer currently uses the Jena 2 Semantic Web framework together with the Pellet reasoner. (c) After successful verification the semantic data is stored in a separate part of the SQL database.

The BOWiki allows a scientific community to annotate biological data rapidly. This annotation can be performed using biomedical ontologies. In addition to data annotation, the specific type of relations between entities can be made explicit. It is also possible to integrate different biological knowledge bases by creating partial definitions for the relations and categories used in the knowledge bases.

The BOWiki employs a type system to verify the consistency of the knowledge represented in the wiki. The type system is provided in the form of an OWL knowledge base. If the type system is a core ontology for a domain (i.e., it provides background knowledge and restrictions about the categories and relations for the domain), its use contributes to maintaining the ontological adequacy of the BOWiki's content, and thereby the content's quality.

ACKNOWLEDGEMENTS

We thank Christine Green for her help in preparing the English manuscript.

REFERENCES

M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, and *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 25(1):25–29, 2000.

F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, Cambridge, UK, 2003.

B. Boeckmann, A. Bairoch, R. Apweiler, M.-C. Blatter, A. Estreicher, E. Gasteiger, M. J. Martin, K. Michoud, C. O'Donovan, I. Phan, and *et al.*. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*, 31 (1):365–370, January 2003.

J. J. Carroll, I. Dickinson, C. Dollin, D. Reynolds, A. Seaborne, and K. Wilkinson. Jena: Implementing the Semantic Web recommendations. Technical Report HPL-2003-146, Hewlett Packard, Bristol, UK, 2003.

F. M. Donini, D. Nardi, and R. Rosati. Autoepistemic description logics. In M. E. Pollack, editor, *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence, IJCAI 1997, Nagoya, Japan, Aug 23-29*, volume 1, pages 136–141, San Francisco, 1997. Morgan Kaufmann.

J. Giles. Key biology databases go wiki. *Nature*, 445(7129):691, 2007.

S. Grimm and B. Motik. Closed world reasoning in the Semantic Web through epistemic operators. In B. Cuenca Grau, I. Horrocks, B. Parsia, and P. Patel-Schneider, editors, *Proceedings of the OWLED'05 Workshop on OWL: Experiences and Directions, Galway, Ireland, Nov 11-12*, volume 188 of *CEUR Workshop Proceedings*, Aachen, Germany, 2005. CEUR-WS.org.

R. Hoehndorf, K. Prüfer, M. Backhaus, H. Herre, J. Kelso, F. Loebe, and J. Visagie. A proposal for a gene functions wiki. In R. Meersman, Z. Tari, and P. Herrero, editors, *Proceedings of OTM 2006 Workshops, Montpellier, France, Oct 29 - Nov 3, Part I, Workshop Knowledge Systems in Bioinformatics, KSinBIT 2006*, volume 4277 of *Lecture Notes in Computer Science*, pages 669–678, Berlin, 2006. Springer.

R. Hoehndorf, F. Loebe, J. Kelso, and H. Herre. Representing default knowledge in biomedical ontologies: Application to the integration of anatomy and phenotype ontologies. *BMC Bioinformatics*, 8(1):377, 2007.

B. L. Humphreys, D. A. B. Lindberg, H. M. Schoolman, and G. O. Barnett. The Unified Medical Language System: an informatics research collaboration. *J Am Med Inform Assoc*, 5(1):1–11, 1998.

M. Krötzsch, D. Vrandečić, M. Völkel, H. Haller, and R. Studer. Semantic wikipedia. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(4):251–261, 2007.

D. L. McGuinness and F. van Harmelen. OWL Web Ontology Language overview. W3C recommendation, World Wide Web Consortium (W3C), 2004.

N. Noy and A. Rector. Defining N-ary relations on the Semantic Web. W3C working group note, World Wide Web Consortium (W3C), 2006.

R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13(1-2):81–132, 1980.

S. Schaffert, R. Westenthaler, and A. Gruber. IkeWiki: A user-friendly semantic wiki. In H. Wache, editor, *Demos and Posters of the 3rd European Semantic Web Conference, ESWC 2006, Budva, Montenegro, Jun 11-14*, 2006.

S. Schulz, E. Beisswanger, J. Wermter, and U. Hahn. Towards an upper-level ontology for molecular biology. *AMIA Annu Symp Proc*, 2006:694–698, 2006.

E. Sirin and B. Parsia. Pellet: An OWL DL reasoner. In V. Haarslev and R. Möller, editors, *Proceedings of the 2004 International Workshop on Description Logics, DL2004, Whistler, British Columbia, Canada, Jun 6-8*, volume 104 of *CEUR Workshop Proceedings*, pages 212–213, Aachen, Germany, 2004. CEUR-WS.org.

B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S.-A. Sansone, R. H. Scheuermann, N. Shah, P. L. Whetzel, and S. Lewis. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotech*, 25(11):1251–1255, 2007.

A. Valente and J. Breuker. Towards principled core ontologies. In B. R. Gaines and M. A. Musen, editors, *Proceedings of the 10th Knowledge Acquisition Workshop, KAW'96, Banff, Alberta, Canada, Nov 9-14*, pages 301–320, 1996.

K. Wang. Gene-function wiki would let biologists pool worldwide resources. *Nature*, 439(7076):534, 2006.