A Controlled Natural Language Interface for Semantic Media Wiki Using the Rabbit Language

Jie Bao[1], Paul R. Smart[2], Dave Braines[3] and Nigel R. Shadbolt[2]

[1] Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180, USA. baojie@cs.rpi.edu
[2] School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, United Kingdom. ps02v@ecs.soton.ac.uk
[3] Emerging Technology Services, IBM United Kingdom Ltd, Hursley Park, Winchester, Hampshire, SO21 2JN, United Kingdom. dave_braines@uk.ibm.com

1 Introduction

Semantic wikis extend the idea of collaborative content editing (made popular by systems such as Wikipedia) to the realm of semantically-enriched representations and formal knowledge models. While a conventional wiki includes structured text and untyped hyperlinks, a semantic wiki is based on the representation of metadata elements. Semantic MediaWiki (SMW) [1] is probably the most popular and mature semantic wiki. It relies on the same wiki engine as Wikipedia and uses constructs from Resource Description Framework (RDF) and Web Ontology Language (OWL).

Despite their potential value as collaborative knowledge editing systems, semantic wikis often present a number of usability issues for human end users. This is due, at least in part, to the complexity of the knowledge representation formalisms associated with languages such as OWL. How can we enable users to create and edit structured knowledge content (in the form of RDF models and OWL ontologies) without reneging on the kind of simple user interaction mechanisms that makes conventional wiki systems, such as Wikipedia, so popular? One answer to this question is to capitalize on the availability of controlled natural languages (CNLs) that provide support for ontology model development. CNLs such as Rabbit [2], Sydney OWL Syntax (SOS) [3] and Attempto Controlled English (ACE) [4], all support the creation of semantically-enriched knowledge models, while preserving the production and comprehension benefits of natural language. As such, CNL interfaces for semantic wiki systems may provide a potent mechanism for encouraging the large-scale participation of user communities in the creation of semantically-enriched online content. Consistent with this view, Kuhn [5] presents a prototype application, called AceWiki, which co-opts the principles of a semantic wiki with the usability features of the ACE CNL.

In this paper, we present our efforts to develop a CNL interface for semantic wikis using the Rabbit CNL. Our main contributions are: 1) An extensible platform for CNL interface for semantic wikis with potential to support multiple CNLs; 2) An OWL meta model for knowledge representation on semantic wiki; 3) A prototype implementation for form-based ontology authoring and controlled natural language generation in Rabbit English.

Due to space limitation, complete features of the system are given in a technical report [6].

2 System Architecture

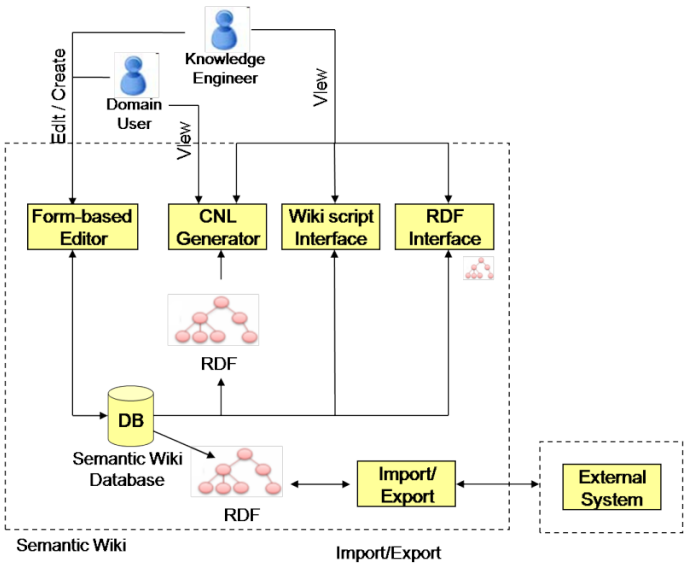

Fig. 1. Architectural overview of a system providing a CNL interface to SMW

Figure 1 illustrates the architecture of a prototype system that provides a CNL interface to a semantic wiki system, namely SMW. The architecture includes the following components:

- **Form-based Editor:** This is an editing interface that allows users to create and edit knowledge statements using the target CNL (Rabbit in the prototype system).
- **CNLG Module:** This is a Controlled Natural Language Generation (CNLG) module that can generate CNL sentences based on the knowledge statements (RDF triples) associated with a wiki page. The CNLG Module accepts a set of knowledge statements and renders these as embedded CNL statements on semantic wiki pages.
- **Wiki Script Interface:** This interface supports the generation of semantic wiki scripts using the usual wiki editing interface. These wiki scripts can describe knowledge in either informal way (e.g., plain text) or in a formal way (i.e., using a semantic annotation syntax).
- **RDF Interface:** This is an interface that supports an RDF-based view of a wiki page. It is intended for users that are familiar with RDF models.
- **Database (DB):** This is the wiki database that stores both semantic (i.e. RDF triples) and non-semantic (i.e. the free text contents of wiki pages) data.
- **Import/Export Modules:** A number of import/export modules handle the communication with external tools and knowledge technology components.

The system inherits the inherent collaborative nature of wikis, and can co-opt with many existing wiki features for improving collaboration. For example, in our approach it is easy to track change history and provenance information of ontology editing, and revert an erroneous edit. The browser-based editing environment offers

high portability and accessability that many desktop-based tools lack. In particular, the semantic wiki-based approach allows both formal and informal descriptions of knowledge, thus may cover more life cycle phases of knowledge engineering. Finally, as the key components of our system (CNL verbalizers, the OWL meta model, and form-based editor) are themselves stored as wiki pages, the system is highly transparent and easier to extend.

We have implemented a prototype semantic wiki system based on the system architecture presented in Figure 1 (see http://tw.rpi.edu/proj/cnl/). The system enables users to create and edit knowledge models using a form-based interface and it renders semantic content as embedded CNL statements using the Rabbit CNL.

3 Semantic Wiki Extensions for CNL Editing

In order to accommodate CNLs, such as Rabbit, within a semantic wiki system, we need to address a number of expressivity constraints associated with conventional semantic wiki systems, such as SMW. SMW does not provide full support for OWL modeling formalisms, and this introduces a mismatch between the kind of knowledge statements that can be represented in Rabbit and the kind of knowledge statements that can be created in SMW.

In order to address this limitation, we developed a meta-model extension to SMW, called SMW-mOWL. SMW-mOWL represents an OWL ontology using a set of 'semantic templates[1]', each of which encodes information about a particular type of ontology element (i.e. classes, properties and individuals). The `Template:Class` template thus encodes information about OWL classes, while the `Template:Property` and `Template:Individual` templates encode information about OWL properties and OWL individuals, respectively. Each of the various axioms and expressions associated with an ontology element (such as a class) is also represented as an instance of a specific semantic template. Thus, the template `Template:Some` represents the `owl:somesValuesFrom` restriction, while the template `Template:ClassRelation` can be used to represent the `rdfs:subclassOf` axiom. The motivation behind this meta-modeling approach is based on a number of design considerations. These include the following:

- **Correspondence to the OWL Abstract Syntax (OWL-AS):** SMW-mOWL is intended to have a direct correspondence to OWL-AS. Ontology elements, axioms and class expressions in SMW-mOWL are all represented as corresponding wiki templates. For example, the class:

  ```
  Class(Rabbit partial Animal restriction(eats
  someValuesFrom(FreshVegetable)))
  ```

  is represented in the semantic wiki as a page using the `Template:Class`, `Template:ClassRelation` and `Template:Some` templates. This results in the following meta-model representation of the aforementioned class:

  ```
  {{Class|label=Rabbit|plural=Rabbits}}
  ```

[1] A semantic template extends the usual wiki template with the ability to turn wiki script into semantic annotations with a predefined skeleton, or to control look-and-feel of pages that use the template.

```
{{ClassRelation|type=subClassOf|class=Animal}}
{{Some|on property=eats|on class=FreshVegetable}}
```

This correspondence between OWL-AS and SMW-mOWL has a couple of advantages. Firstly, OWL-AS can be used as an intermediate syntax for knowledge exchange between SMW and other tools. Secondly, it provides an extensible framework for supporting multiple CNLs within the SMW environment.

- **UI Correspondences:** The design of SMW-mOWL is intended to have a direct correspondence to the SMW knowledge editing system. By utilizing what are called 'semantic forms' within SMW, each template can be edited using a form-based interface. Having an OWL meta-model thus provides us with a low-cost solution to the problem of implementing a full OWL ontology editor within the SMW environment.
- **Query Convenience:** The use of a template-based mechanism for SMW-mOWL allows us to store the OWL ontology in the SMW database and to use SMW-QL (the query language for SMW) to retrieve information from the model.

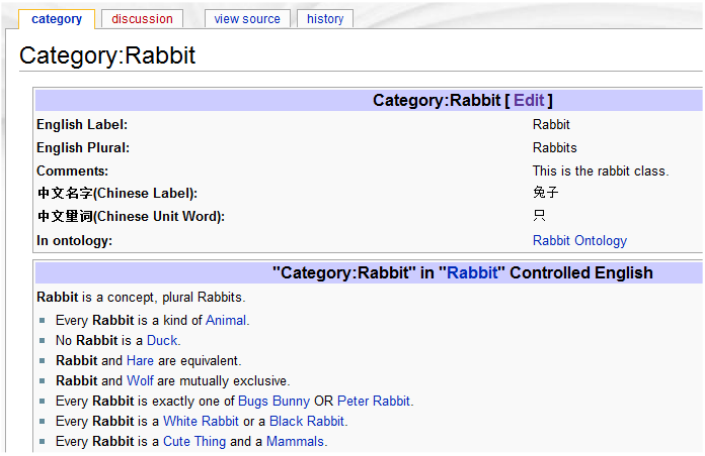

Fig. 2. Semantic wiki page containing Rabbit CNL sentences for an OWL class.

At the present time, the meta-modeling approach described above provides a significant fragment of representational support for OWL. The form-based interface, however, does impose limits on what kinds of expression a user can create within the system. The user cannot create class expressions of arbitrary complexity, for example, class expressions where the value of a quantifier restriction is itself a complex expression rather than a named class. The following class cannot therefore be created within the semantic wiki system using the form-based interface:

```
Class(Rabbit partial Animal restriction(eats
someValuesFrom(intersectionOf(Vegetable
restriction(isFresh value(true))))))
```

Our future work on the form-based interface, as well as the CNL interface (see section 2), will aim to address this shortcoming.

Figure 2 illustrates a semantic wiki page for an OWL class that has been created using the form-based interface associated with the semantic wiki system. The page contains Rabbit CNL sentences that correspond to the elements of the class description. These sentences are generated using the CNLG component described in Section 2.

4 Conclusions and Future Work

This paper summarizes our initial efforts to develop a generic CNL interface for semantic wiki systems. Thus far we have specified a system architecture to support the collaborative editing of community knowledge using semantic wikis. We have also developed a set of meta-model extensions to SMW in order to accommodate the expressivity features of OWL-compliant CNLs, like Rabbit. Finally, we have implemented a prototype system (see http://tw.rpi.edu/proj/cnl/) that combines a form-based editor with a CNL generation capability. This system enables users to create/edit ontologies within the SMW environment and to view the contents of those ontologies as embedded CNL (i.e. CNL statements embedded within a semantic wiki page).

Future work will focus on the development of enhanced editing and parsing capabilities. For example, we are currently working on a better user interface to support the direct entry of Rabbit sentences. We also aim to extend the semantic template system in order to accommodate multiple CNLs (e.g. Rabbit, SOS, ACE, etc.) within a single semantic wiki system. This integrative approach to CNL-mediated knowledge content creation will, we suggest, surmount some of the usability issues associated with semantic wiki systems and make large-scale collaborative knowledge modeling on the Semantic Web more accessible to a variety of user communities.

References

1. Krötzsch, M., Vrandecic, D., Völkel, M., Haller, H., Studer, R.: Semantic Wikipedia. Journal of Web Semantics **5** (2007) 251-261
2. Hart, G., Johnson, M., Dolbear, C.: Rabbit: Developing a Control Natural Language for Authoring Ontologies. 5th European Semantic Web Conference (ESWC'08), Tenerife, Spain (2008)
3. Cregan, A., Schwitter, R., Meyer, T.: Sydney OWL Syntax—towards a Controlled Natural Language Syntax for OWL 1.1. OWL Experiences and Directions Workshop, Innsbruck, Austria (2007)
4. Kaljurand, K., Fuchs, N.E.: Bidirectional mapping between OWL DL and Attempto Controlled English. 4th Workshop on Principles and Practice of Semantic Web Reasoning, Budva, Montenegro (2006)
5. Kuhn, T.: AceWiki: A Natural and Expressive Semantic Wiki. Semantic Web User Interaction Workshop at CHI 2008, Florence, Italy (2008)
6. Bao, J., Smart, P., Braines, D., Shadbolt, N. A Controlled Natural Language Interface for Semantic Media Wiki, In Tetherless World Constellation Technical Report, TW-2009-05, 2009 http://tw.rpi.edu/wiki/TW-2009-05