

Knowledge management using Wikipedia

Dongeun Sun, Seongbin Park*, and Hyosook Jung

Department of Computer Science Education
Korea University, Seoul, Korea
{sunde41, psb, est0718}@comedu.korea.ac.kr

Abstract. In this paper, we present an ontology-based system that helps users manage knowledge using Wikipedia. The system analyzes ontologies and uses the structural information about the ontologies to re-structure contents of Wikipedia for better browsing. Using the system, users can acquire knowledge easily from Wikipedia. We show how the system can be used for life science applications.

1 Introduction

Wikipedia is a freely available online encyclopedia developed by a community of users as free content encyclopedia to which anyone can contribute [1]. In this paper, we present an ontology-based system that allows users to manage knowledge using Wikipedia. The system combines contents of Wikipedia and structural information obtained from ontologies so that users can acquire knowledge from Wikipedia easily. More specifically, the system analyzes ontologies represented in OWL [2] and contents of Wikipedia are re-structured in such a way that users can browse the contents easily. Users can also edit ontologies files so that personalized ontologies can be combined with the contents of Wikipedia. Figure 1 shows the features of the proposed system.

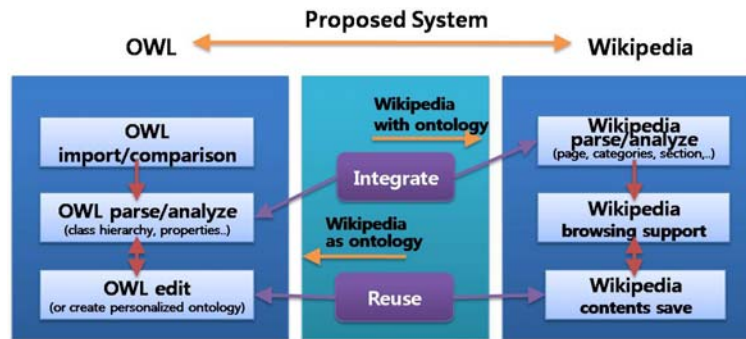


Fig. 1. Features of the proposed system

* To whom correspondence should be addressed.

This paper is structured as follows. Section 2 describes related works. Section 3 explains the proposed system and describes illustrative examples that show how the system can be used. Section 4 concludes the paper.

2 Related Works

There are two areas that are related to our work.

First, there are approaches to leverage Wikipedia by using ontology, which generally begin with converting machine-processable Wikipedia to more machine-understandable one with Semantic Web technologies. Semantic Wikipedia [4, 5] attempts to add Wikipedia to semantic relations by typed links between articles and other formalized elements, and tries to extend template to one that can help to write semantically enhanced articles easily. Likewise, WikiOnt [6] is built to describe and exchange Wikipedia articles into semantic one. WikiOnt provides the schema or vocabulary of Wikipedia articles, and Wikipedia articles that are described using the WikiOnt can be automatically converted to RDF instances by PHP scripts. Ontologies can be also used with wikis instead of Wikipedia. For example, Platypus Wiki [7] introduces RDF based approaches, OntoWiki [8] deploys a modified wiki installation as an ontology engineering platform, and SweetWiki [9] provides easy access to Ontology with WYSIWYG editor, etc.

Second, there are approaches to build new ontologies from Wikipedia contents, or to use Wikipedia as ontology itself, or to use Wikipedia to analyze and evaluate existing ontology. [10] constructs a broad general-purpose ontology from Wikipedia and [11] extracts a specific biological ontology from Wikipedia's biological text. [12] uses Wikipedia entries as controlled vocabulary or ontology. [13] uses Wikipedia category graph to compare the class hierarchy in ontology. As for relationships other than hierarchy, WikiRelate! [14] and ESA [15] indicate semantic relatedness between Wikipedia articles as relationship in ontology.

3 The system

In order to use the system, a user needs to load an ontology represented in OWL. There are three options for importing ontologies. The system imports ontologies from users local system, or url or Swoogle which is a search engine for Semantic Web documents [3]. Using crawler, Swoogle can discover RDF documents and embedded RDF content in HTML document.

After importing an ontology, the system parses the ontology into classes, individuals, and their hierarchy relationship. The parsed classes, individuals are represented as tree structure, in class hierarchy view and class detail view that the system supports. A user can edit the ontology while navigating information related to it. When a user selects a class, it shows subclasses or properties of the class. When the user selects a class as a query and requests information about the class, it provides information related to the document. For example, it shows categories in Wikipedia associated with the query, links in Wikipedia that point to the document, and the content structure of the document. In browsing

Wikipedia, a user can save the text data, image, PDF files into project, which is the data collection including ontology, text, image, PDF in our system.

Figure 2 shows functionalities provided by the proposed system.

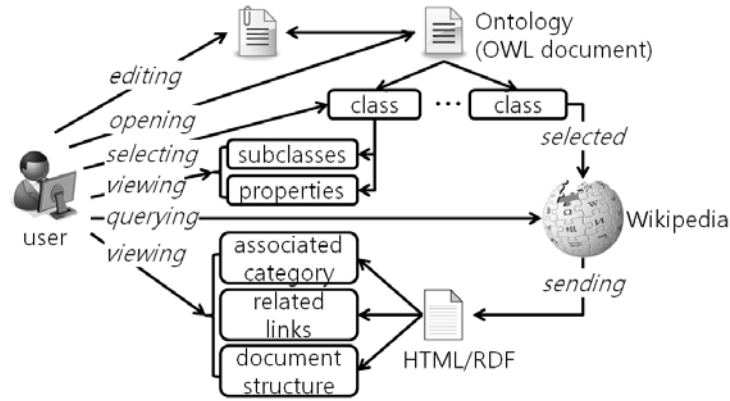


Fig. 2. Functionalities of the proposed system

4 Illustrative examples

4.1 Example 1

Assume that a student wants to know about Gene but does not know where to start. To browse Wikipedia with an ontology, the student needs to create a project that consists of an ontology and Wikipedia's data that the ontology is related with. Next, the student needs to import an ontology file. Assume that the student selects Swoogle option. In Swoogle option (Fig 3), the student can search the word ("gene") included in ontology. The system shows the student the ontology file name and included terms and class(instance) number.

After the importing step is done, the system shows the user the selected ontology hierarchy and properties and switch its view into Wikipedia Browser. The system coordinates the terms, properties and relationship in ontology imported and coordinates the categories and page sections in Wikipedia. In other words, the system gives an overall information concerned with ontology imported. (Fig 4) From them, the user can start studying Gene. When the user selects a root class, in this time, Gene, the system shows the user related Wikipedia page, named Gene as well as another Wikipedia pages list with the same class definition and related Categories. Recursively, the user scans root classes like Allele, Marker, Chromosome etc, resulting in getting a knowledge of Gene. In a related Wikipedia pages, if necessary, the user can store a Gene section contents and a necessary text, image and reference PDF in project. And at any time, the user

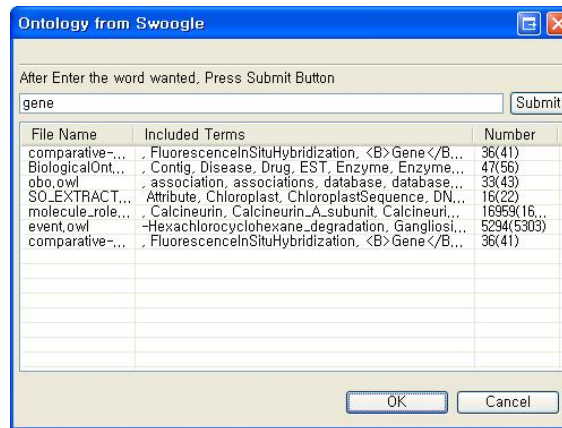


Fig. 3. Importing ontologies using Swoogle

can view these text, image PDF without browsing Wikipedia pages, and can edit text in the system itself. The user repeats this process with sub class and gets to know Gene, extensively into Life Science. And through analyzed ontology, the user relates and refines a knowledge of Gene. During these processes, the user can re-define ontology, that is to say, can create personalized Gene ontology. Additionally, user can print out the summary of a knowledge of Gene in overview viewer.

4.2 Example 2

Assume that sixth graders in an elementary school learn about animal classification which is to group species of animals according to their shared physical characteristics in a science class. In the sixth-grade curriculum, animals are divided into vertebrates and invertebrates according to whether they have vertebrae. Vertebrates are divided into mammalia, bird, reptilia, amphibia and fish. Invertebrates are divided into arthropoda, annelida, mollusca, coelenterata echinodermata, and platyhelminthes. The students learn physical characteristics of each animal group, and classify animals according to the characteristics. However, it is difficult for novices to classify animals because they should understand the structures and mechanisms specific to those groups well.

To use the system, a student loads an animal ontology represented by OWL. The ontology defines the categories of animals and their relations according to their similarities and differences. If a class in the ontology is selected, the student can see its subclasses and properties. Figure 5 is a screenshot shown when a student loads an ontology and clicks arthropod. The student can see its subclasses such as myriapoda, crustacean, hexapoda, and chelicerata, and its properties such as has segmented bodies and jointed limbs, and lays eggs.

If the student wants to obtain information about each animal group in detail, the student can browse related Wikipedia pages. After clicking a selected class,

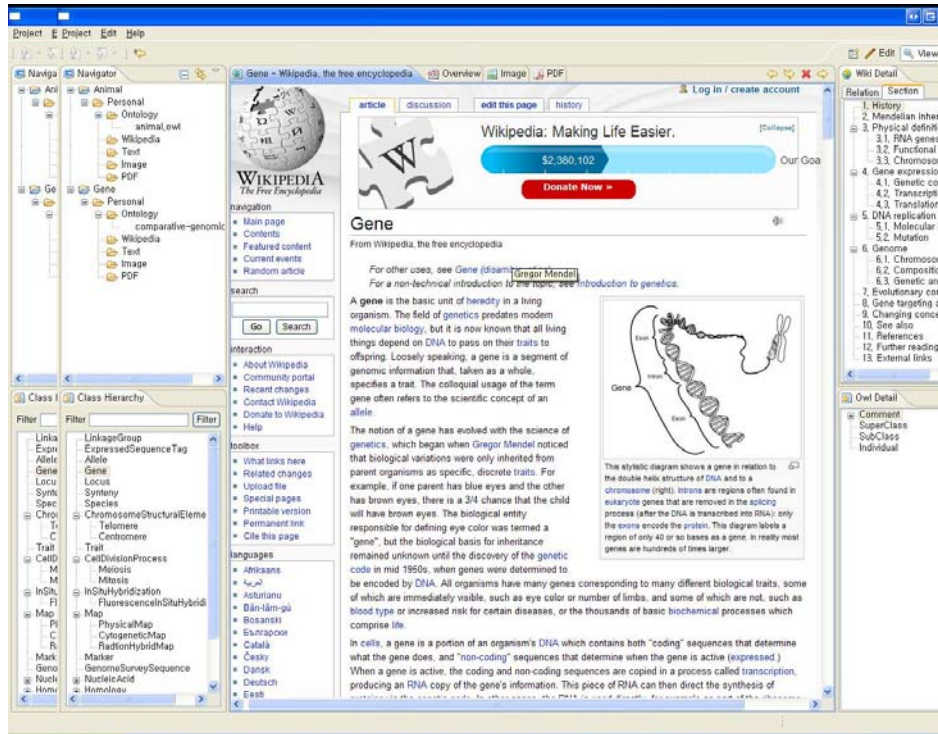


Fig. 4. Screenshot of the re-structured contents of Wikipedia

the student can see a Wikipedia page related to the class. It provides useful information to understand the class such as its classification, evolution, diversity, senses, etc. It is possible that the student can save a fragment of the page such as text or image. Figure 6 shows a part of a Wikipedia page which the student clips in our system and figure 7 shows an image in a Wikipedia page which the student saves in our system.

If the student finds additional information about the animal classification while browsing the Wikipedia pages such as subclasses or properties of a class which are not in the loaded ontology, the student can edit the ontology in order to add the subclasses or properties. For example, the student is reading a Wikipedia page about primate and finds that it has subclasses such as tarsier, monkey and ape, but there is just human as a subclass in the loaded ontology. The student adds tarsier, monkey and ape as its subclasses into the loaded ontology.

5 Conclusions

In this paper, we presented an ontology-based system that can help users acquire knowledge from Wikipedia. We showed a simple example that illustrates how the

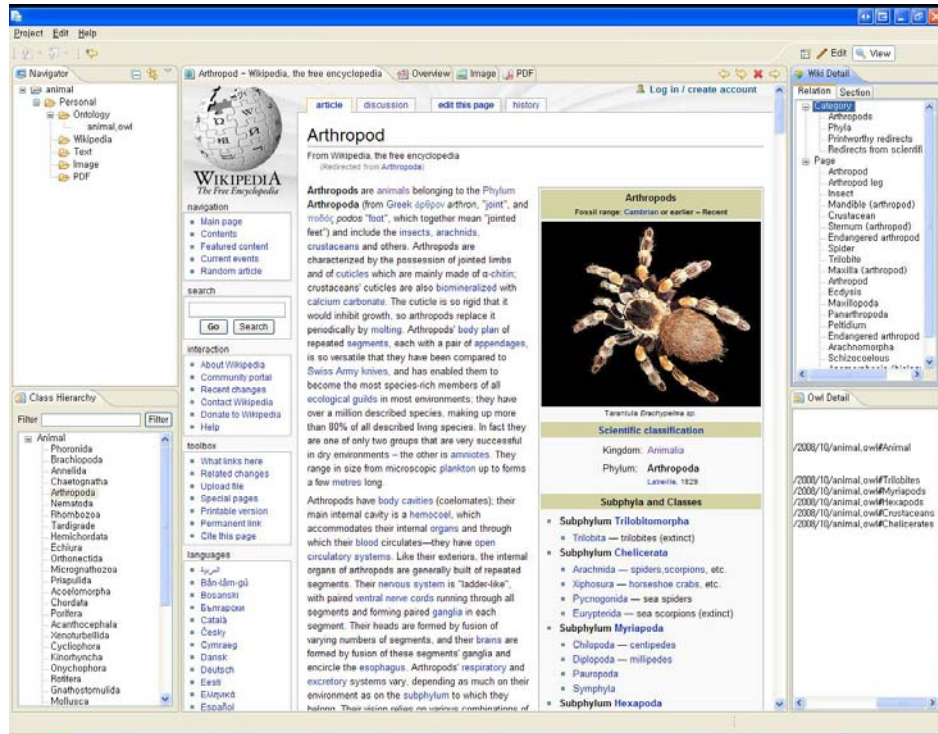


Fig. 5. Loading an ontology

proposed system can be used for life science applications. While ontologies are difficult concepts for novice users, we believe that our system allows non-expert users to design or re-design personalized ontologies without a steep learning curve.

References

1. <http://en.wikipedia.org/wiki/Wikipedia>
2. <http://www.w3.org/TR/owl-ref/>
3. <http://swoogle.umbc.edu>
4. M Krotzsch and D Vrandeic and M Volkel. Wikipedia and the Semantic Web-The Missing Links. Proceedings of Wikimania 2005
5. M Vlkel, M Krtzsch, D Vrandeic, H Haller. Semantic Wikipedia. Proceedings of the 15th international conference on World Wide Web 2006
6. Andreas Harth, Hannes Gassert, Ina O'Murchu, John Breslin, Stefan Decker. WikiOnt: An Ontology for Describing and Exchanging Wikipedia Articles. Proceedings of the 1st International Wikimedia Conference, 2005
7. S.E. Campanini, P. Castagna, and R. Tazzoli, Platypus Wiki: A Semantic Wiki Web, Proc. 1st Italian Semantic Web Workshop Semantic Web Applications and Perspectives, 2004

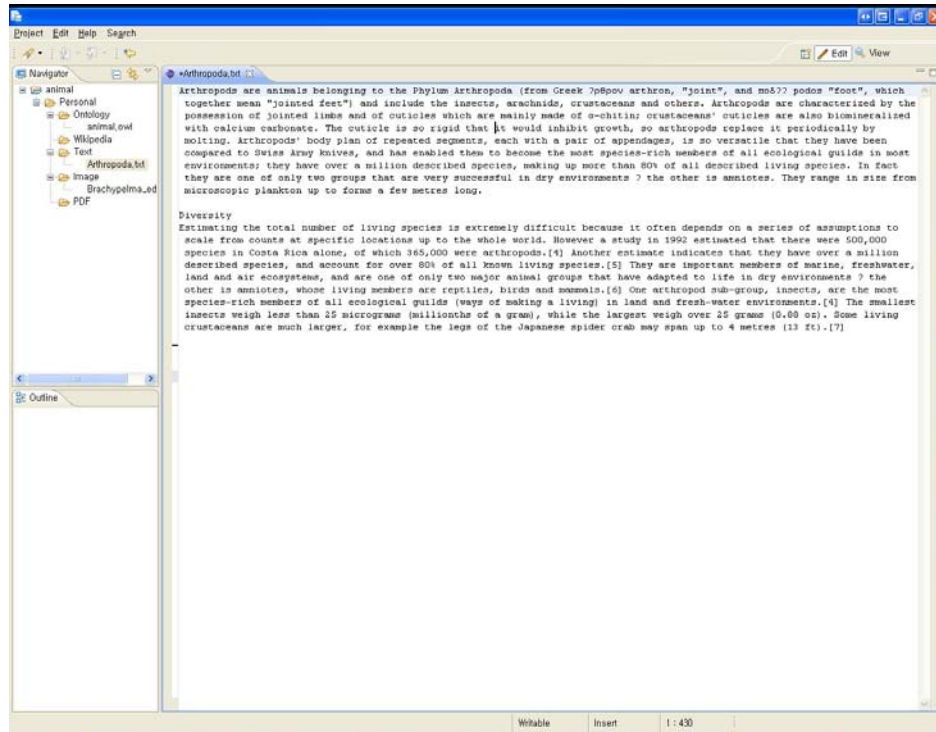


Fig. 6. A part of a Wikipedia page

8. M. Hepp, D. Bachlechner, and K. Siorpaes, OntoWiki: Community-Driven Ontology Engineering and Ontology Usage Based on Wikis, WikiSym 06, 2006
9. M Buffa, F Gandon. SweetWiki: semantic web enabled technologies in Wiki. Proceedings of the international symposium on Symposium 2006
10. F Wu, D Weld. Automatically refining the wikipedia infobox ontology, WWW08, 2008.
11. A. Herbelot and A. Copestake. Acquiring ontological relationships from wikipedia using rmrs. Proceedings of Workshop on Web content Mining with Human Language Technologies, ISWC06, 2006.
12. M Hepp, K Siorpaes, D Bachlechner. Harvesting Wiki Consensus: Using Wikipedia Entries as Vocabulary for Knowledge Management. EEE Internet Computing 2007
13. J Thom, J Pehcevski, A Vercoustre. Use of Wikipedia Categories in Entity Ranking. Arxiv preprint arXiv:0711.2917 2007
14. M Strube, S Ponzetto. WikiRelate! Computing semantic relatedness using Wikipedia. Proc. of AAAI 2006
15. E Gabrilovich, S Markovitch. Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis. Proceedings of the 20th International Joint Conference on 2007

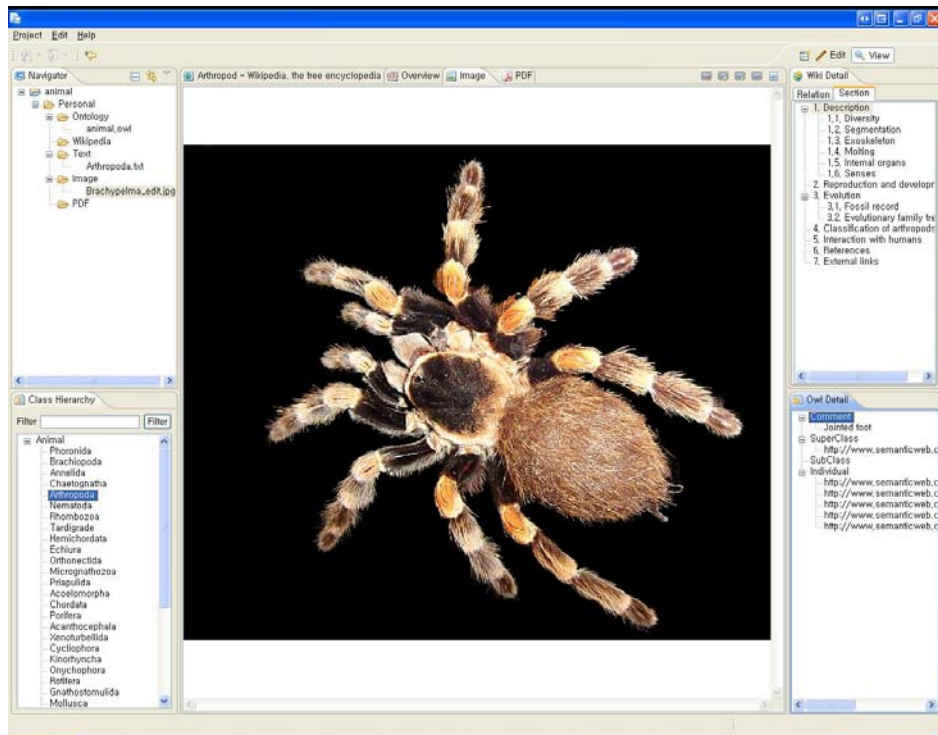


Fig. 7. An image in a Wikipedia page