Collaborative climate change research on the semantic web

Christian Battista
Jobe Microsystems/
University of Western Ontario
1201 Richmond Street #504
London, Ontario, N6A 3L6
1-226-820-1136
christian@jobemicrosystems.com

Benjamin Coe
Jobe Microsystems/
University of Guelph
921 Midland Avenue #305
Scarborough, Ontario, M1K 4G2
1-416-265-7860
ben@jobemicrosystems.com

ABSTRACT

WikiEarth (http://www.wikiearth.net) is a website designed for encouraging collaboration between researchers across the academic spectrum, and also serves as a test case to determine the limitations and benefits of using an ontological data structure to manage the input of natural science based data from around the world.

Drawing upon Wikipedia's model of massive user collaboration, WikiEarth's motivation is to extend beyond this by formalizing the relationships between the data being entered. A semantic ontology is a natural candidate for data representation for three reasons: first, the hierarchical class structure of an OWL-Ontology helps avoid redundancy when developing simulations, as an operation can be applied to a class and all its subclasses; secondly, a framework like Jena helps eliminate human error and reduce the amount of data entry that needs to be performed; and finally, important restrictions regarding data entry are imposed by the ontological structure and Jena, as opposed to by a proprietary system developed for one specific application.

Utilizing this infrastructure, a WikiEarth Climate Demonstration was successfully conceptualized, constructed, deployed and subsequently unveiled at the 2009 World Student Environmental Summit. The success of this application demonstrates that ontologies could be effectively purposed for a high-traffic production system.

Categories and Subject Descriptors

B.2 ARITHMETIC AND LOGIC STRUCTURES
I.2.4 Semantic networks
H.3.5 Online Information Service

General Terms

Algorithms, Design, Human Factors

Keywords

Climate change, sustainability, online collaboration

1. INTRODUCTION

In this document we will describe the planning, design, challenges and findings of the WikiEarth Climate Demonstration. It is important to emphasize the difference between the eventual WikiEarth website (which does not currently exist), and the WikiEarth Climate Demonstration (which does exist), which we are describing in this document. We will begin with a discussion of the design of the system, followed by findings of the current application of WikiEarth Climate Demonstration at the WSES, from which future recommendations can be constructed.

2. DESIGN

Ontology development is reliant on having a well defined problem. We cannot begin to design until we know what we're making. Because of the amount of flexibility, it was tempting to start too high (very abstract) or low (very specific empirical properties) levels of complexity. Instead we started at a middle ground in the problem, but providing ourselves room to grow. We began by describing properties of academic institutions and dietary items. We were later asked to also include state information about the diet items (whether they were solid or liquid), and we were able to do this without disrupting the system. There already exists an ontology (the Food Ontology) that serves this purpose; however, for this small demonstration, the WikiEarth society and WSES co-ordinators deemed that a smaller, simpler ontology should be used. As WikiEarth grows, existing ontologies such as the Food and Geo Ontology will be integrated into the system.

3. IMPLEMENTATION

Despite the steep learning curve, Jena was successfully used for this project. We used a portion of its available functionality to load the ontology into Java and put it in a state we can call on. Jena automated certain steps such as the creation of inverse relationships. We made extensive use of this feature, thus making it easy to traverse the ontology. The diet keeps track of the diet items, and the diet items keep track of the diet. No matter where you are, it's easy to refer yourself to the appropriate location.

We also used Jena to act as a reasoner for the impositions on class properties and relational properties. This allows for a high level abstraction—a relationship can be based on an abstract concept like 'a diet has food' and we can then add in as much as food as we want, but never have to change the relationship after that, no matter how much food we enter. This gives us a basis so we can have relationships involving just the top level properties (an object has a distribution chain), and is good for abstracting relationships. Where we to employ a relational database, we

would have to build to suit that, whereas with ontologies we can focus on abstract relationships as opposed to the actual data itself.

Given the processing power required and the latency introduced by running both the Jena reasoner and the XML-RPC calls, caching was used to ensure that classes only needed to be reloaded into PHP when changes were made to the underlying ontology. Furthermore, we could cache the results of the simulations that were run, such as the carbon footprint calculation, unless changes were made to the ontology.

4. CRITICISMS

While the site was met with a great deal of excitement and enthusiasm from the delegates of the World Student Environmental Summit, some criticisms do exist. The primary concern is that an ontology is only as good as the data that has been entered into it. Thus, students who did not put in the required effort to track their food items did not improve the knowledge base we were trying to build. However, this was a rarity, and we were able to collect diet information and make a variety of interesting conclusions based on the data.

From a technological perspective, it is quite reasonable to wonder why a relational database was not used for this small demonstration. While this would certainly have been possible, it is important to remember that the end goal of the WikiEarth project is to have a massive collection of scientific data upon which simulations can be run. As we move towards more complex set of data, we hope to take advantage of description logics and object reasoners to perform time series simulations and more complex data manipulations, and it is through the use of Jena that this is made possible.

Further, scientific data is most often organized in the form of a taxonomy, so an ontology is a data structure that experts from various fields should be able to easily conceptualize and work with. As the success or failure of WikiEarth hinges on the quality of the data being entered, ease of use is of critical importance.

5. SUMMIT FINDINGS

The most interesting result surrounded the impact of imported beers and wines. Most people don't immediately think of their alcohol consumption as part of their 'diet proper', and so don't consider the environmental impacts of drinking imported alcohols. However, a student in British Columbia, Canada, with an entirely local diet substantially increased their CO_2 emissions by drinking a lot of German beer. Another interesting finding was that vegetarian diets tend to require consumers to import food from great distances, probably due to the restrictions on the growing seasons. Meats are available locally year-round, and this should be factored into statements that are made about whether or not being a vegetarian is 'greener'. The most obvious finding was that regions with very short growing seasons, like Sweden, or with limited area to grow food, like Japan, have high carbon costs associated with importing food.

6. FUTURE WORK

While the WikiEarth Climate Demonstration was deemed to be a resounding success, more work must be put into developing a stable high-volume production system, with the two main concerns being speed and stability. Speed can be improved by improving the server hardware, but also be improving the caching mechanisms. Effective methods for invalidating the cache must be developed, such that classes are reloaded—which is rather processor intensive—only when necessary. In terms of stability, synchronization issues were addressed, such that concurrent modification by a multitude of users is possible. Some issues, either in our systems or in our system's interaction with Jena, were not threadsafe. It is known that Jena has some issues when multiple threads are present in the JVM. Our system spawns a thread every time a new XML-RPC call is made, so this is an issue that must be resolved. However, this can be surmounted by working more with Jena's locking functionality.

Future steps also include the addition of temporal version control, which will allow the system to perform simulations that can span multiple versions of the data stored in the ontology.

Most importantly, we aim to extend the WikiEarth site such that it can be applied to new and globally significant problems, such desertification, deforestation, and high toxin levels in urban areas. This will be attempted through the uploading of data related to the sourcing of the raw materials and byproducts involved in the manufacture of the goods that fill the spaces around us. With this data available to the public, the life cycles of products and goods that people consume can be analyzed, and citizens of WikiEarth will gain a novel understanding of the impacts of a globalized society. Using this knowledge, it is hoped that they may shift their buying and spending patterns away from products that are tied to negative environmental effects.

7. ACKNOWLEDGMENTS

Our thanks to the WikiEarth Society and the World Student Environmental Summit Co-ordinators for making this project into a tangible reality. We would also like to thank the World Student Environment Summit Student Delegates for acting as test users during the initial deployment of this system.

8. REFERENCES

[1] Kuhn, T. AceWiki: A Natural and Expressive Semantic Wiki. Semantic Web User Interaction Workshop, CHI 2008.

[2] R. Tazzoli, P. Castagna, and S. E. Campanini. Towards a Semantic WikiWikiWeb. In 3rd International Semantic Web Conference, ISWC 2004.

[3] Castells P. and Macías, J. A. An Adaptive Hypermedia Presentation Modeling System for Custom Knowledge Representations. In Proceedings of the World Conference on the WWW and Internet, WebNet 2001, pp. 148-153.

[4] Rauschmeyer, A. Semantic-Web-Backed GUI Applications. Proceedings of the ISWC 2005 Workshop on End User Semantic Web Interaction, CEUR 2006.

[5] Schwabe, D., Szundy, G., Moura, S. S. D., and Lima, F. Design and Implementation of Semantic Web Applications. In WWW Workshop on Application Design, Development and Implementation Issues in the Semantic Web, WE-SW 2004, volume 105 of CEUR Workshop Proceedings, pp. 275–284.

[6] Noy, N.F.;McGuinness, D. "Ontology Development 101: A Guide to creating your first Ontology" Stanford KSL Technical Report KSL-01-05, 2000.