Increasing the Accuracy of Wiki Searches Using Semantic Knowledge Engine and Semantic Archivist

Gretchen Lowerison
Coreidea Innovation Inc.
95 Salem Pl., London, ON N6K 1T8
+1.519.488.6298
glowerison@coreidea.com

Michael Lowerison
Coreidea Innovation Inc.
95 Salem Pl., London, ON N6K 1T8
+1.519.488.6298
mlowerison@coreidea.com

ABSTRACT

Current practices in industries such as aerospace attempt to aggregate information from a wide area as part of their decision making process. However, collecting knowledge that is critical to a project is often daunting and time consuming. This paper describes the conceptualization and early development of a framework consisting of a semantic knowledge engine, archivist tool, and knowledge-mapping tool using a wiki front-end as a means for users to enter knowledge using a familiar web-based interface.

Categories and Subject Descriptors

I.2.4 [**Knowledge Representation Formalisms and Methods**]: Relation systems; Semantic networks.

General Terms

Algorithms, Management, Design, Reliability, Theory.

Keywords

Wikis; knowledge building; semantic analysis; data aggregation; knowledge management; search and retrieval

1. INTRODUCTION

To be successful in today's knowledge economy requires that learners, scholars, researchers, and practitioners be able to effectively and efficiently, collect, store, evaluate, connect, and retrieve knowledge and best practices. This is typically achieved with the use of a knowledge management system that, though effective for most purposes, can result in fragmentation of information. As an example, Ash [1] reports on a concern at Boeing. He writes: "what's happened at Boeing, like many companies, is that the advent of personal computing has fractured the organizational database and the company's critical knowledge is now scattered among personal computers and in a plethora of formats" (¶ 3).

Current practices in industries such as aerospace, attempt to aggregate information from a wide area as part of their decision making process, e.g. when designing an aircraft component or selecting an appropriate product such as adhesive. Typically, multiple sources provide information, including industry contacts, former colleagues, product manufacturer data sheets, technical data sheets, material specification documents, process specification documents, etc. Some of this information is stored locally and coded based on specific use. Other information, such as information obtained from contacts, may be lost if not recorded. Retracing one's steps or determining if one can use a specific product for more than one purpose, e.g., can the same product be used as a surfacer and a structural adhesive, may require an excessive amount of time.

1.1 Wiki use

Recently, wiki use has increased in popularity proving to be an effective tool in encouraging users to consolidate knowledge and understanding thereby allowing that knowledge to become more visible. A good example of this is the Wikipedia [http://www.wikipedia.org/] where the global community is taking an active part in this consolidation process. Being part of a Web 2.0 environment such as a wiki collaborative group places users into a community of practice thereby expanding their opportunities for understanding [2].

By using the wiki "information is continually being generated, captured, referenced, and enhanced resulting in knowledge" (¶ 7) [3]. Final contributions are the result clarification and consensus making the development of knowledge an ongoing process rather than an end in of itself [4].

For learners, learning and understanding can occur at a deeper level; for business, the wiki offers a competitive advantage. The availability of information and connections to related ideas and best practices have the potential to benefit problem solving and decision-making. However, in order for this to be true it must be possible to locate relevant information with a high degree of accuracy.

It is believed that a semantic enabled wiki can help address these issues. This project describes the conceptualization and early development of the Coreidea System: a framework consisting of a semantic knowledge engine, archivist tool, and knowledge mapping tool (working name Chimera), using a wiki front-end as a means for users to enter knowledge.

2. COREIDEA FRAMEWORK

The componentized approach of the Coreidea framework divides the overall system into a series of separate, independently operating services. Working together, these components provide a coherent knowledge management system, allowing for both automatic higher-level semantic analysis of new content and easy retrieval of existing information, such as knowledge and best practices. The Wiki front-end presents a means for users to enter content using a familiar web-based interface. When the user edits or creates a new Wiki page, the Coreidea system channels their changes through a data aggregation and analysis pipeline. To simplify later processing, the first step involves data translation and enrichment. This it accomplishes using a translation gateway, which converts the data into a form that the rest of the system will understand, and feeds it onto the bus. On the bus, the data is now available for further interpretation. (see Fig. 1).

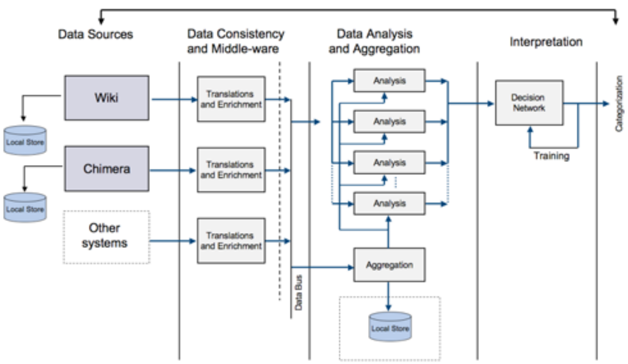

Figure 1. Semantic knowledge engine conceptualization model

The Archivist service reacts to the availability of new information and performs a multi-part analysis, while the aggregation service updates its word frequency tracking data matrices. The Archivist combines several document analysis techniques through a user trainable decision network under the belief that a user directed approach will mitigate the divergence from user expectation. One such technique involves performing a latent semantic analysis on the user content. Another involves performing a TF-IDF (term frequency – inverse document frequency) analysis. Both approaches yield an abstract 'orthogonality' from an existing corpus of documents, which, while often accurate, says nothing about specific context.

Both of these document analysis techniques require word statistics, which makes the role of the aggregation service clear. As an aside, from a technical point of view, synchronizing the communication between the bus and aggregator and Archivist and aggregator, while maintaining a strict decoupling between the two services, is one of the challenges.

Both the Archivist and aggregator needs to treat new data in a carefully sequenced manner. The analysis uses data that must be absent from the data matrices it is processing, requiring the use of event based synchronization techniques in the underlying messaging layer.

The system stores a decision network for each known topic. It feeds the product of each separate document analysis into the decision network for each topic yielding an overall probability that that topic is appropriate for the content. The Archivist then presents these choices to the user, who can accept or reject each topic. It uses the decisions that the user made to back-calculate through the decision network, allowing the network to learn. With adequate training, the choices made by this processing pipeline returns expected user results.

Chimera, another front-end application, presents the user with a data entry oriented interface. The user enters in relational data using dynamic fields, which the system will internally store and aggregate. A query language interface provides a means for the user to create expert system style queries against the aggregated store. This allows for the creation of reports and automated analyses that the student, professional, or manager can use to make well-informed decisions. Chimera also feeds this data onto the Coreidea bus, initially through a translation gateway to normalize it, and then following the same process as with the Wiki front-end.

A significant deficit of many wiki systems is the need to perform exact or morphologically similar searches. With the semantic archivist data store, searches using concepts, rather than specific words, become possible.

3. REFERENCES

[1] Ash, J. 2007. Boeing on knowledge management Retrieved November 13, 2008, from http://blog.jackvinson.com/archives/2007/10/29/boeing_on_knowledge_management.html

[2] Jarche, H. 2008. Skills 2.0: Web 2.0 gives learning professionals an aptitude adjustment. T+D, 62,4, 22-24.

[3] Albrycht, E. 2006. Knowledge is a means, not an end Retrieved October 13, 2008, from http://www.wikithat.com/wiki_that/2006/02/knowledge_is_a_.html (Feb 24, 2006).

[4] Lowerison, G. 2009. Using wikis to support learning in post-secondary economics. Unpublished doctoral dissertation, Dept. of Education, Concordia University, Montreal, Quebec, Canada.