> Semantic Web Use Cases and Case Studies

Use Case: Semantic MDR and IR for National Archives

Tony Lee, Jin Woo Kim, and Bok Ju Lee, Saltlux, Kyu Hyup Kim and Yoon Jung Kang, National Archives, Korea

September 2008

Introduction (Background)

NAK, the National Archives of Korea, is a government service agency that accumulates and preserves historical records and data. It then makes this available through various means. In January of 2004, the agency began an electronic organization of the records. A record system model of the central records management system was developed in 2006 and the technological analysis system and electronic document system were connected. All of the records in the NAK and the national government records system were interlinked and a new search engine was developed to search through both databases. It is the goal of the system to provide an information delivery system to the public.

General Description

Challenge

Presently, the NAK does not have a standard metadata strategy. Metadata pertaining to job classifications, department classifications or contents types (text, books, periodicals, etc.) are absent. Depending on the usage objective, various metadata implementations are being created on a case by case basis. Because the NAK is associated with various information sources, semantic interoperability takes priority over volume.

The NAK, because of the challenges referred to above, has recognized the need to develop a large capacity semantic information retrieval system and construct Semantic MetaData Repository (SMDR) that will keep 70 million non-current national records. In order to build a collective search engine, there is a need to resolve the inhomogeneities between metadata entities' names, attributes and relations. In this project, the ISO/IEC 11179 MDR Meta model needs to be executed to output over 10 million indexed RDF triples and provide real-time query capacity.

The Solution

The NAK's search engine construction needs to be approached from a top-down, ISO-11179 standard ontology MDR (MetaData Repository) type. This is to confirm the interoperability and to construct domain ontology from a bottom-up approach. Over 12 million metadata statements are used in the resulting semantic IR System.

The key to semantic IR is semantic annotation; this skill is essential to metadata search. Therefore we used a text-mining solution [IN2]TMS to extract NAK's relationships then took ontology based mapping solution, [IN2]Semano to execute the automated Semantic annotation. Additionally, to search the large capacity metadata, the [IN2]SOR framework and metadata storage AllegroGraph was utilized to realize a reasonable capability of the semantic metadata. The rules in the reasoning engine of OntoBroker are appropriate to produce semantic search results. This increased customer satisfaction dramatically.

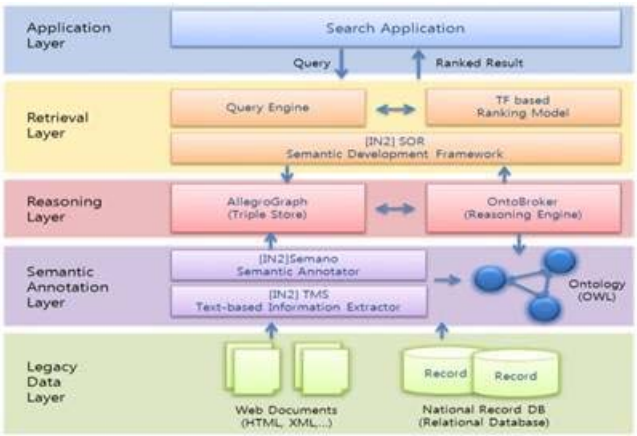

Figure 1: System Architecture

Key Benefits of Using Semantic Web Technology

The previous keyword search system, directory system and collaborative search system took the keyword matching approach. These limitations were overcome by construction of NAK metadata relationship and reasoning based search engine service that resolves the user's information approach and ambiguity.

Additionally, benefits were realized using Web 2.0 technology, including an improved user interface and improved retrieval results.

Text-mining technology improved the user's experience in searching for appropriate information. Because the search execution is now approachable from information- and subject-based types of search, even information that was previously difficult to understand has been simplified. Therefore there is an increase in the usage and the effectiveness of the system.

Figure 2: Screen dump of semantic search system at `http://search.archives.go.kr/searcher_semantic_form.htm`

Next Steps

By participation in the FP7 LarKC project of the EU, large capacity metadata reasoning technology is to be improved by semantic annotation technology. Through further development of semantic search engines, a large capacity semantic information retrieval service is to be developed.