A Taxi in Knowledge Land

a Case Study about implemented XML Knowledge Technology. XML Topic Map Engine as a Web Service.

Authors:

Michael Angrick, Richard Bös, Maria Rüther (UBA DE); Thomas Bandholtz (SchlumbergerSema)

C/o: thomas.bandholtz@koeln.sema.slb.com

In 1998 the European Union released the Aarhus convention on "access to information, public participation in decision-making and access to justice in environmental matters". This turned out to be a unheard-of challenge to the information policies and technologies of all the governmental authorities concerned. In Germany the Federal Environmental Agency ("Umweltbundesamt", UBA) decided to develop an "information broker" as a single-point access to any public information of the authorities on the federal and state ("Laender") level. We implemented an XML based Web solution, the German Environmental Information Network (www.gein.de/index_en.html). GEIN was presented as the customer's contribution to the EXPO 2000 and has become one of the favorite models for governmental information portals in Europe.

From a technical point of view, GEIN has been buildt on “XML anywhere”. XML has been used as a universal data format for storage (Tamino), retrieval (Xpath), processing (DOM) and communications. We were one of the international wide spread projects who invented the "wheel" of communicating XML in the body of HTTP requests and responses, which short time later was standardized as the Simple Object Access Protocol (SOAP) by W3C. In all, GEIN demonstrates ".. state-of-the-art web techniques such as XML, Java Servlets, and HTTP communication, ... providing a fast and dynamic, thesaurus-based knowledge management" (from the UBA marketing flyer).

In the field of Knowledge Management, GEIN uses three semantic structures:

- a thesaurus of about 30.000 environmental terms,
- a gazetteer including the intersections between over 50.000 geographical objects of all kinds,
- a synopsis of historical and contemporary events that affected the environment.

GEIN already has implemented a linguistic analysis of documents to suggest keywords that characterize each document. This analysis can also be used to process search criteria, so search criteria may be any kind of free text, including the complete content of a given document. After having implemented this we felt very comfortable while navigating in this “knowledge land” spontaneously – just like using a perfect taxi.

After more than one year of successful production now we are succeeding with an **XML Topic Map Engine** (XTME), and we named it the **KnowledgeTaxi®**, a general solution for web based knowledge navigation. XTME is included in the **Semantic Network Services** (SNS)[1], a follow-up of GEIN. The three semantic structures mentioned above are integrated into one single Topic Map by setting up associations between thesaurus descriptors, events and locations. The XML model of this Topic Map is based on ISO/IEC 13250, but it does not use the XTM 1.0 DTD. We preferred to develop an XML Schema for Topic Maps after having reviewed the basic ISO SGML DTD, the XTM XML DTD, and Martin Bryan’s early XML Schema proposal.

Trusty to the good tradition of XML anywhere**, KnowledgeTaxi®** is designed to be a "plug-in" that can be integrated with any Document/Content/Knowledge Management solution by means of a Web Service interface to the three operations:

(1) Topic matching – searching for any kind of occurrence of a given character string in the Topic characteristics,

(2) Topic navigation – moving from topic to topic using associations as stepping stones,

(3) Auto-Classification – using a new topic-map-based linguistic analysis to extract the most characteristic Topics of a given document.

---

[1] Research Project UFOPLAN-Ref. No. 20111612, promoted by BMU/UBA, Germany.