

European Environmental Topic Map [EETM]

by Thomas Bandholtz, Sema

Written contribution to the Preparatory Workshop on

Intelligent Systems for Environmental applications

FP6: Ambient Intelligence Systems for the Citizen

Brussels (B), April 27th 2001.

1 Problem definition

In the 80s and 90s of the past century it was difficult to find and access environmental information. Most of the information was distributed among 1000s of organizations, and there was no general interface to collect contributions to a global or European knowledge base. Most of the information management efforts focused on *data catalogues* containing metadata about the original information objects which could only be accessed via personal contact to some authorized person. While the catalogue systems developed quite rapidly, the same can not be said about the catalogue content. Metadata content was collected manually and discontinuously - once filled, the catalogue was not up to date any more. Metadata had to be very descriptive (verbose), because it was the only source of information available to decide, whether it would be worthy to gain the effort to access the real data behind.

Today, environmental information is published in masses in the internet. Everybody can access this information with his internet browser from his office, home, from a public internet cafe, or using mobile media in the near future. But most of this information you will not find listed in the catalogues, as it develops very fast and unpredictable, and the catalogues are too slow. Still the information is distributed among many places, and each information provider follows his own rules and preferences. There is no harmonization but being "in the internet", and there is (almost) no localization support but common search engines. Thus the recipient, though amidst a wide variety of environmental information, just *happens* to find some of it without being able to compare it to what could be found.

2 Proposed RTD or support activity

The growth of the internet brought up the idea of the *Semantic Web*, searchable by means of semantically meaningful metadata (and not only by character strings). Early implementations can be found as national (i.e. German Environmental Information Network) or European (EEA) environmental *portals* that serve as centralized access points, not only as passive catalogues like before.

Metadata can be kept up to date by intelligent agents that find and index any relevant piece of information ("Web Mining"). Metadata needs not to be that verbose any longer, as the information itself is just "one click away" and can easily be accessed immediately to have a first look. Systems like these are "light", fast, and they offer a *complete* catalogue of everything accessible in the internet.

Portals like these will play a most important role in the "global environment intelligence information, communication and service infrastructure" as they support basic environmental knowledge management features to learn about and communicate with "interoperable & harmonized environmental software and databases" or "distributed sensor systems and networked environmental applications".

The new semantic system benefits from the inheritance of the early catalogues: thesauri and metadata standards. They use GEMET or the different national thesaurus sources to build an index in a well structured database.

The new semantic system is always up to date, as it is self-maintaining by an automatic process that scans the web and analyzes content permanently.

In 1998, ISO has standardized a vision of a general structure for semantic systems like these, called Topic Map [ISO/IEC 13250:2000]. Basically, a Topic Map is a network of *topics*, *associations* between topics, and of *occurrences* (information) of these topics. In a map like this, the *distance* between two topics is defined as the number of associations to be followed from one to the other.

The vision is that each Topic Map in the world can be automatically merged with another because all of them follow the structural standard.

So the proposal is to develop a multilingual European topic map for the environmental scope, based on the catalogue inheritance (like GEMET) and on the web mining features of the early semantic portals (like GEIN).

The main work to be done will be:

- transform GEMET to conform the ISO Topic Map structure,
- add a geographic orientation system, containing the pre-calculated intersections between all relevant European locations,
- add a synopsis of environmental events like accidents, conferences etc.,
- model associations between all instances of those three "meta-topics" (thesaurus terms, locations and events),
- develop sophisticated methods for auto-classification of web information as occurrence of topics ("semantic engine"),
- build an easy to use generic interface for accessing the environmental topic map both for semantic engine services and for information retrieval.
- support national and international portals to integrate the features of the environmental topic map and act as autonomous sub-maps that can be merged by an application.

All these implementations should be done based on the Extensible Markup Language (XML). The EETM is part of the general proposal of the Environmental Markup Language (see attached presentation).

3 Potential benefit/exploitation/markets

The European Environmental Topic Map will build and maintain an information network of all environmental information published in the internet. That is why information will be accessible and comparable, and so: sustainable.

The benefit will be that existing information is not longer hidden and does not seem to be unstructured any more. All the information will act as a self service pool for reporting and investigation. It can be easily navigated using the associations of the Topic Map, including document-to-document comparison that is supported by running the semantic engine twice, on the documents each, even if documents are in different languages.

There will be a broad usage in the field of public information following the Aarhus Convention, but also there may be a commercial service for company concerns.

4 Major actors potentially interested

- Environmental Authorities of all European Countries and elsewhere in the world,
- Global Environmental Organizations like (UNEP/Infoterra),
- Standardization Committees like ISO/IEC 13250 or XTM (www.topicmaps.org),
- Companies that offer content management systems dealing with topic maps (like Ontopia, Empolis, and Sema).

Contact:

Thomas Bandholtz
XML Competence Center
Sema GmbH
Kaltenbornweg 3
D50679 Köln/Cologne
++49 (0)221 8299 264
thomas.bandholtz@de.sema.com