The Semantic Network Service – Supporting Heterogeneous Environmental Information Systems

Matthias Menger
Umweltbundesamt, Dessau, Germany
(Federal Environment Agency of Germany)
Tel: +49 340 2103 2037
Fax:+ 49 340 2104 2037
matthias.menger@uba.de
http://www.umweltbundesamt.de

Maria Rüther
Umweltbundesamt, Dessau, Germany
(Federal Environment Agency of Germany)
Tel: +49 340 2103 2391
Fax:+ 49 340 2104 2391
maria.ruether@uba.de
http://www.umweltbundesamt.de

Abstract:
This paper describes aim, technology and prospects of the 'Semantic Network Service – SNS'[1] which has been developed by the Umweltbundesamt (UBA, Federal Environment Agency of Germany) and is integrated in and used by several environmental and geographical information systems. SNS offers the possibility to share controlled environmental vocabulary and to boost the search & retrieval performance of environmental information systems.

Keywords:
Semantic Network Service, harmonized vocabulary, environmental information system, environmental thesaurus, environmental Web Services, topic map, semantic indexing

1. Goal and Background of Semantic Network Service SNS

The aim of SNS is to improve the search and retrieval performance of environmental information systems and make dispersed data better available for distributed information gathering across different information systems in the web.
SNS supports the user in finding appropriate terms to get more precise and useful information out of environmental information systems. On the other hand SNS should help to support a controlled terminology for the environmental community, which is and will always be very heterogeneous due to its cross-sector relevant data and information.
Access to the vocabulary and the related methods is given by Web Services. The idea was to provide the terminology, its maintenance and the technology only 'once for all'. Therefore, information suppliers don't have to integrate and care for this in each application, but have only to integrate the offered services.

The development of the Semantic Network Service SNS of the German Umweltbundesamt (German Federal Environment Agency) started in 2001. It has been built on the experience of the German Environmental Information Network gein® which has provided fully-automated semantic indexing of documents on the Internet. Based on this experience improved methods has been implemented in the SNS. Since, it has been developed continuously to a modular approach providing services to support the environmental community with a common controlled vocabulary to describe their data and information provided via their diverse information systems with metadata.
The controlled vocabulary shared by SNS is based on three main components: an

[1] http://www.semantic-network.de

https://doi.org/10.23106/dcmi.952108606

environmental thesaurus, a national gazetteer and an environmental chronicle. These components are interrelated and realized in a 'semantic network' - a topic map. Further components, like more detailed vocabulary of a specific environmental field might be added, e.g. as micro thesaurus, and thus allows building on a growing shared ontology which used to describe and index environmental information.

2. Components and Technology

The 'semantic network', interlinked controlled vocabulary of environmental terms, locations and events by integrating these three components in a semantic model by a topic map (see figure 1). The topic map is based on the XML Interchange format XTM 1.0 (XML Topic Maps, ISO/IEC 13250). The semantic network is established by setting up associations between events and locations (where happened) and events and descriptions (what happened).

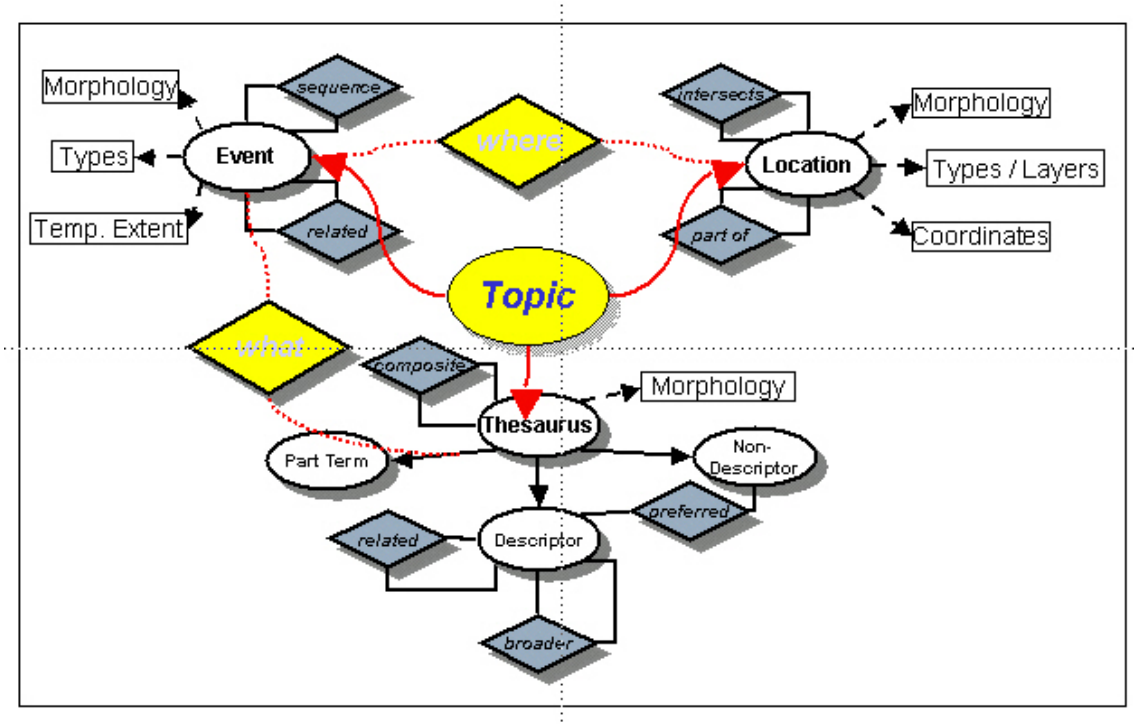

Figure 1: Structure of the Semantic Network

term: The source of the 40,000 environmental terms is the thesaurus UmThes® of the Umweltbundesamt and it is maintained by an editorial group of the Agency.

location: The maintenance of the 20,000 geographical names based on a work of the Federal Agency of Cartography and Geodesy (BKG)[2], extended by several environmental location types by the Umweltbundesamt. Actually the BKG is establishing a Gazetteer Service, which will be available from mid 2006. So the continuous maintenance of the geographic names is ensured.

event: The chronicle of environmental events is continuously maintained by the SNS project.

[2] http://www.bkg.bund.de

https://doi.org/10.23106/dcmi.952108606

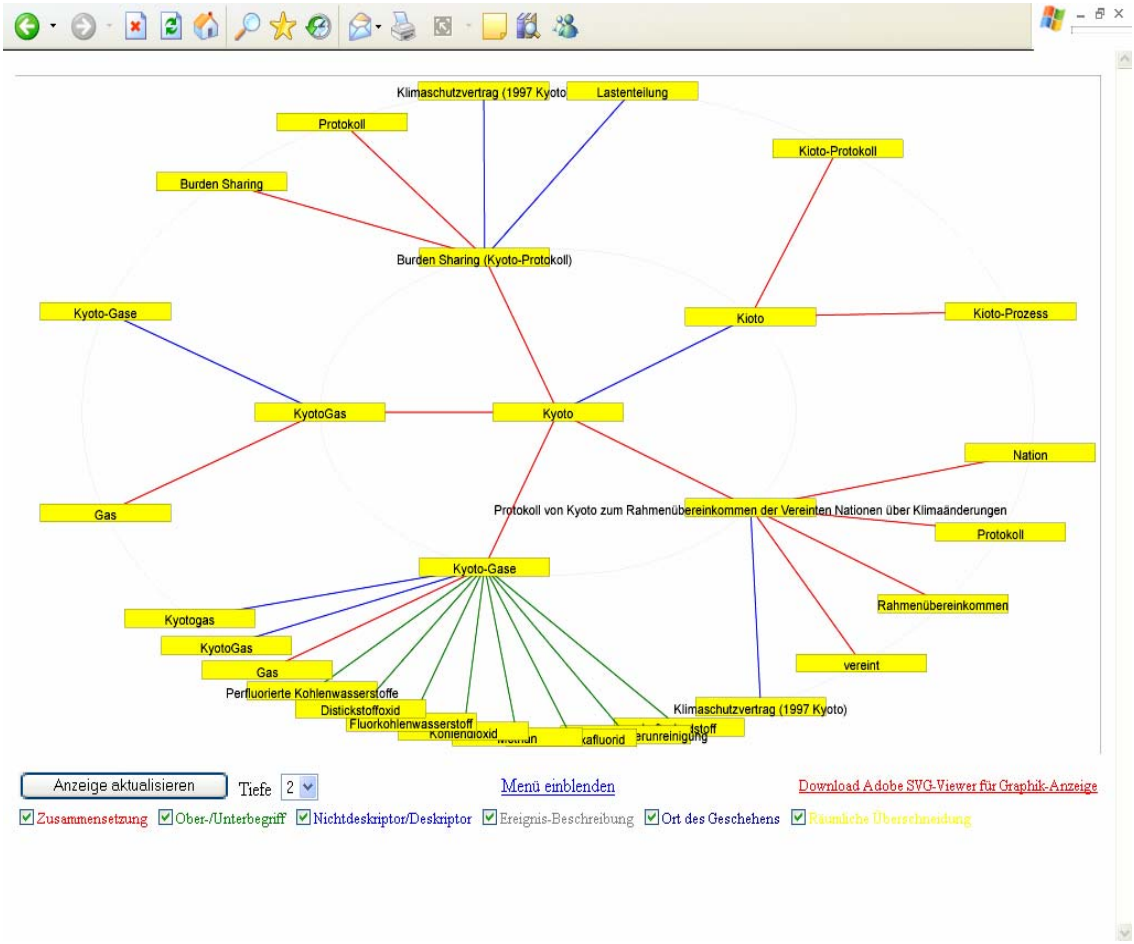

Figure 2: Graphical demo search result for search term `Kyoto`

These semantic associations allow browsing and searching in a complex network of terms, which is done automatically by services. A graphical user interface, available via www.semantic-network.de, demonstrates the complexity of this mechanism (see figure 2).

The idea has been to provide services accessing the topic map (interrelated vocabularies) and use it for automatic searching and indexing of documents. The services are realized as web services accordingly to W3C Web Service standards and implemented as Java servlets. The following services have been realized so far:

findTopics searches topics by names and topics types

getPSI (PSI="Published Subject Identifier") reference of topic characteristics and associations (browsing terms).

autoClassify provides automatic classification indexing of html, xhtml and PDF documents. The text to be indexed can be passed by a document or by a web address. The result is a list of significant topics that classify the document.

getSimilarTerms returns "somehow" similar terms for a given search term as recommendation for a more successful search & retrieval

https://doi.org/10.23106/dcmi.952108606

findEvents finds events matching the given search term.

anniversary finds events in the environmental chronicle happened x years ago by a reference date as a reminder.

The service *autoClassify* (see figure 3) for automatic classification and indexing of environmental information is currently the mainly used service today.

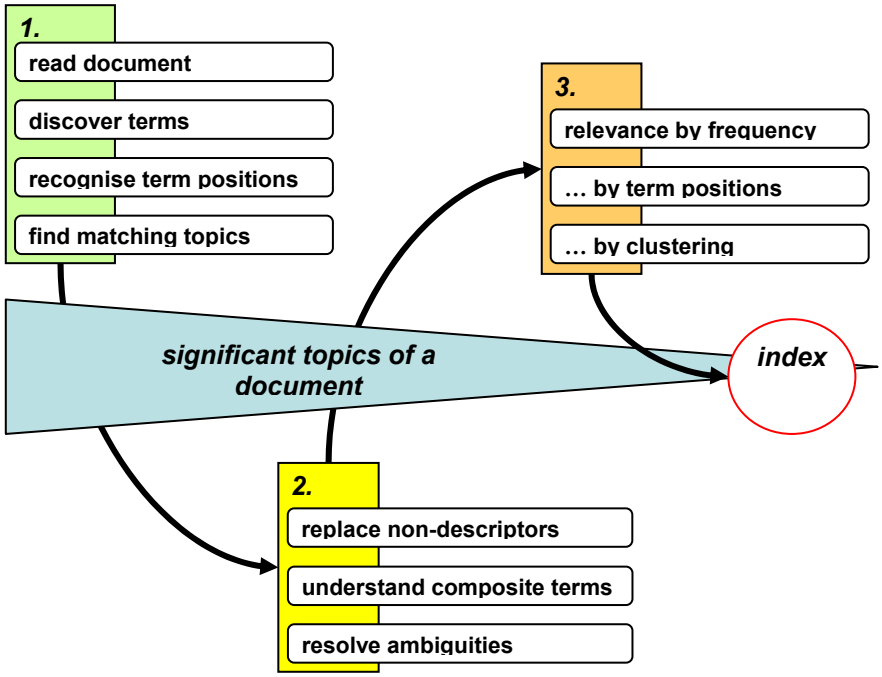

Figure 3: *autoClassify* - Automatic Content Classification

The services have been successively implemented as Web Services. Main user of the SNS is the German Environmental Portal 'PortalU'[3], launched end of May 2006 as the succession portal of the 'German Environmental Information Network - gein®', which has integrated all services in its application. In addition the 'German Portal of Spatial Data Infrastructure (SDI) GeoPortal.Bund'[4] and several SDI portals at the state level are using the services, too. Other applications as the Environmental Portal of the State of Baden-Württemberg are going to integrate the services in 2006.
The services are used by the general public via the several systems using these services. Most systems or applications are provided by governmental organizations to provide better information for the public. Therefore, the services are used by information professionals, government employees, and of course all users of these systems

Especially the methods *autoClassify* and *getSimilarTerms* support the user by formulating the query or retrieving more relevant results.
The method ***autoClassify*** proposes preferred terms for a given "document" to be used in a query (in case of using to general or common search terms by an user) or to index html-, xhtml or pdf-documents (e.g. websites) with specific metadata structured by theme, time and

[3] http://www.portalu.de
[4] http://www.geoportal.bund.de

https://doi.org/10.23106/dcmi.952108606

geographic reference. These additional metadata could be used to extend a given search engine index or to write (create) the metadata directly on a webpage to offer better, i.e. more relevant, search results.
Available in a first version since 2006, the new method ***getSimilarTerms*** however proposes "somehow" similar terms for a given term to be recommended as full text search terms. The SNS project is currently discussing and improving this method together with two information suppliers in a project co-operation. The main discussion point and challenge is how to organise the synonym rings of the vocabulary to support the user best. The improved method should be available in autumn 2006.

3. Prospects of SNS

The idea of integrating different environmental information systems via a simple and centrally available shared terminology is not new but is still lacking support and implementations are rare. The benefit of sharing harmonized vocabularies for indexing data resources with metadata and for a better search and retrieval performance is evident. Users and data/information providers are all gaining from this concept. Adding more semantic to data and information is a win-win situation.
For this reason the modular approach of SNS offering simply accessible semantic services is 'key' for an environmental information network with the aim to provide useful information.
The ability to extend the controlled vocabularies with further detailed micro thesauri of specific environmental themes, like water, air emission etc., allows to build a growing framework of semantics used in the environment sector. This semantics would benefit each environmental information systems.

Even more interesting might be the extension to non-environmental themes and thesauri, like administration terminology. Only one example might be the integration of SNS in a Document Management System (DMS) used by an environmental authority to manage their internal workflow and documents. Furthermore, it would support the processing of data to user specific information which is one of the essential and growing tasks not only in the environmental sector.
The example shows the extensive potential of integrating services of SNS in further systems.
SNS is intended for non-commercial use. It is available ready for use and integration with a downloadable developer tool kit. Therefore, SNS is one key component providing semantic services in the framework of service oriented architectures, although this framework has to be established yet.

4. References:

1. Thomas Bandholtz and Maria Rüther. Application of Environmental Semantics. In Informatics for Environmental Protection. Place: Brno/CZ, J. Hrebicek, J. Racek, 2005, pp. 44-48.
2. Maria Rüther. Sharing Environmental Vocabulary. In Informatics for Environmental Protection. Place: Geneva/CH, CERN, 2004, vol. 2 pp. 293-295

https://doi.org/10.23106/dcmi.952108606