E3S Web of Conferences **1**, 18004 (2013)
DOI: 10.1051/e3sconf/20130118004

Pollution and Health terminology into the EARTh thesaurus

S. Di Franco, V. De Santis and P. Plini

[1] National Research Council, Institute of Atmospheric Pollution Research, Environmental Knowledge Organisation Laboratory, Rome, 1 Research Area, ***difranco@iia.cnr.it, vds@iia.cnr.it, plini@iia.cnr.it***

Abstract. Highly structured and refined, but flexible tools are needed to deal with issues as the information management on the web and the semantic interoperability. Many issues in the field of knowledge organisation and information management can be traced back to meaning delimitation. There are opposite needs and tendencies: on one hand there's the necessity to share a common and stable meaning of the terms in order to guarantee communication within a community. On the other hand, openness to further exploration of meaning should also be ensured to do not impoverish its richness and complexity. The term "benzene" could be defined in different ways. A biologist may consider its toxicity and the different routes on which it can enter an organism. An engineer would consider it as a fuel for combustion engine. A physicist may see it as a volatile and inflammable liquid. A fire brigade can consider benzene as a dangerous inflammable substance that cannot be extinguished by water, but only with foam. A chemist may see it as the precursor of a class of chemical compounds, etc. Benzene could be thus defined in several different ways depending on the context in which it is considered. But we should also underline that all these definitions share a common premise: "benzene" is first of all a substance (that can have toxic effects, be used as fuel, that can also cause accidents, etc.). This semantic trait seems to be unavoidable. Starting from this premise and referring also to suggestions coming from the development of applied ontologies a new environmental thesaurus format containing some innovative elements has been designed and is currently available. The goal is the development of a thesaurus aiming to become an advanced tool to be applied for environmental information management, to sustain environmental policy and research. To achieve this result EARTh has been conceived as a tool able to deal with the specific features of the environment sector (multidisciplinary character, high complexity, bio-cultural implications, etc.). In this paper the EARTh component dealing with environmental pollution and heath is presented.

Key words: Thesaurus, Environmental pollution, Air, Soil, Water, Health

Introduction

The Environmental Knowledge Organisation Laboratory (EKOLab) of CNR-IIA started its activities in the field of environmental knowledge organisation in 1992, aiming at designing and maintaining of mono- and multilingual thesauri, classification schemes and terminological systems for the environment.

Thesauri are controlled and dynamic vocabularies of terms, where semantic relations between terms (hierarchical, associative, equivalence) are explicitly declared. Thesauri are tools for the semantic control of the language. In the environmental field they are used for indexing, classification, cataloguing and retrieval of information in environmental databases and more recently as part of a network to expose, share, and connect pieces of data, information, and knowledge on the Semantic Web.

In 1999 EKOLab in collaboration with the German Federal Environment Agency developed GEMET, the GEneral Multilingual Environmental Thesaurus, as an indexing, retrieval and control tool for the European Topic Centre on Catalogue of Data Sources (ETC/CDS) and the European Environment Agency (EEA).

In 2001, EKOLab started to design and develop EARTh (Environmental Applications Reference Thesaurus); the project aimed at improving/expanding the 1999 version of GEMET to revise and refine its categorical and thematic structure; to update the content; to ensure the diffusion of the thesaurus using the more updated technologies.

The first versions of EARTh were developed following the ISO standards available at that time, 2788:1986 and 5965:1985.

New version of EARTh should be released every 6 months. New versions will be made available in two formats: web interface and RDF/SKOS.

The semantic model

EARTh-Environmental Application Reference Thesaurus (Mazzocchi et al., 2007) is based on a

multidimensional classificatory and semantic model.

The "vertical structure" of the Thesaurus has been built through a deductive (top-down) – inductive (bottom-up) approach, it is basically mono-hierarchical, has been developed according to a tree semantic model and is based on a system of categories. The first level of categories corresponds to Entities, Attributes, Dynamic aspects, Dimensions.

The vertical structure analyses the primary meaning of the terms and places them in the classificatory-hierarchical tree aiming to orientate the users towards the most "essential" characteristics of terms' semantics.

A thematic organization of terms has been elaborated. A theme or a subject is here conceived as a sector of interest that reassembles the terms linked to it despite their semantic meaning.

Thematic organisation, as it was conceived, is developed according to the specific needs of the applicative context like the classification of terms for the management of information in the field of environmental policy.

"Traditional" thesauri typically provide a poorly differentiated set of relationships between terms, distinguishing only among hierarchical relationships, associative relationships and equivalence relationships. In the EARTh project the standard relationships are being arranged into richer subtypes, whose semantic content is specified. This work is particularly useful dealing with the associative relations (RTs). Typically RTs include a heterogeneous and undifferentiated set of relations, expressing many kinds of association between terms that are not hierarchically based. In EARTh RTs are differentiated into subtypes, thus strengthening the transversal relational structure.

The enrichment of thesaurus relationships and the increased semantic clarification of the relations could enable a better semantic description of Web resources and guide users in meaningful information discovery on the Web (Soergel et al., 2004).

EARTh contains at present more than 15.000 terms in English and Italian. Its terminological content is derived from various mono- and multi-lingual sources of controlled environmental terminology:

- GEMET (1999)
- UN Environment and Development (1992)
- Italian Thesaurus of Earth Sciences (2000)
- Inland Water terminology (2001)
- Emergency Management Terms Thesaurus (1998/2003)
- other reference documents in specific fields. Between others the IPCC (Intergovernmental Panel on Climate Change) terminology and glossaries.

Terminology related to pollution and health

The discussion on pollution and health topics requires a clear comprehension of terms that have a technical meaning.

An increasing interest about pollution impact on human and environmental health necessitates the use of a specific terminology supporting not only a better understanding of concepts but also the sharing of information avoiding errors and misunderstanding.

EARTh includes several terms connected to pollution and related topics such as health, water, soil and atmosphere having a special meaning to those who work in this field.

Starting from the EARTh's system of themes a query was performed in order to extract terms related to pollution, air, health, soil, water and climate; the query selected about 3.500 terms out of 15.000.

Being expanding issues, related terminology is continuously growing and changing; the attribution of themes to Thesaurus' terms is in a work-in-progress phase as-well-as the input of new terms concerning health problems and air and heavy metal pollution field, so abovementioned results will increase as long as the work continues.

Applications

EARTh represents an environmental semantic map that could be utilized for different purposes, being able to combine the search of stable logical and conceptual basis with a flexibility towards different applications, to represent a semantic map of the environmental domain and to ensure an optimal conceptual coverage to be aware of the cultural dimension of knowledge organization, to allow different levels of comprehensibility and applicability for users with different expertise and finally to ensure the porting of the thesaurus into different technological applications.

It is important to stress that the main use of any thesaurus is indexing and classification. The presence of linguistic equivalents and/or definitions can be considered as an added value but not as the main elements of the thesaurus.

EARTh has been adopted by the Italian Ministry for the Environment as a glossary for the management of information on waste management (http://www.osservatorionazionalerifiuti.it/), by the Italian Institute for Environmental Protection and Research for the indexing of technical and scientific documents relevant for the environment (http://www.envidocnet.isprambiente.it/INDEKS/public/welcome.do), by European projects (Nature-SDI, NESIS, EGIDA) and by national and international research projects (GIIDA, GEO-GMOS). This was also made possible by the development of a RDF/SKOS version of the thesaurus developed by CNR-IMATI (http://linkeddata.ge.imati.cnr.it:2020/directory/EARTh).

Nowadays networked information access to heterogeneous data sources requires interoperability of controlled vocabularies. Different thesauri are created with different points of view and can be based on different ways of conceptualization. Their development reflects different scopes and can imply different levels of abstraction and detail. Switching between thesauri, thus being able to create dynamic and semantically based correspondences among different vocabularies is urgently needed (Bandholtz et al., 2009).

Future steps

Following the publication of the ISO 25964-1:2011 an overall revision of the thesaurus structure and content is currently undergoing.

It is foreseen to increase the number of terms also as a consequence of the use of EARTh by environmental stakeholders.

The mapping with other environment-related thesauri will be strengthened also through the Linked Open Data network.

References

ISO, 2011. ISO 25964-1 Information and documentation - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval.

Hudon, M., 1997. Multilingual thesaurus construction: Integrating the view of different cultures in one gateway to knowledge and concepts. Knowledge Organization, Vol. 24 (2), pp. 84-91.

Mazzocchi F., De Santis B., Tiberi M., Plini, P., 2007. Relational Semantics in thesauri: Some Remarks at Theoretical and Practical Levels. Knowledge Organization, vol. 34 (4). pp. 197-214.

Felluga, B., Batschi W.D. (Eds.), 1999. GEMET, General European Multilingual Environmental Thesaurus. Version 2.0. European Environmental Agency, Copenhagen.

Bandholtz T., Fock J., Legat R., Nagy M., Schleidt K., Plini P., 2009. Shared Terminology for the Shared Environmental Information System. Environmental Informatics and Industrial Environmental Protection: Concepts, Methods and Tools, 23rd International Conference on Informatics for Environmental Protection., Volume 1. Shaker, Aachen, pp. 123-127.