

Leveraging Ontologies for Environmental Information Systems

Ulrich Bügel¹, Martin Schmieder¹, Boris Schnebel¹,
Thorsten Schlachter² and Renate Ebel³,

¹ Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB),
Fraunhoferstraße 1, 76131 Karlsruhe, Germany
{Ulrich.Buegel, Martin.Schmieder, Boris.Schnebel}@iosb.fraunhofer.de

² Karlsruhe Institute of Technology (KIT), Institute for Applied Computer Science,
Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany
Thorsten.Schlachter@kit.edu

³ Baden-Württemberg State Institute for the Environment, Measurements and Nature
Conservation (LUBW), Griebßbachstraße 1-3, 76185 Karlsruhe, Germany
Renate.Ebel@lubw.bwl.de

Abstract. The provision of accurate, comprehensive and condensed information contained in distributed environmental information systems via public search interfaces raises several technological challenges. Our approach to tackle these challenges is based on a consequent use of ontologies. Starting with an analysis of requirements resulting from semantic search scenarios, we explain the advantages of using ontologies based on standards and aim to reuse and transform terminological systems available in the environmental domain into ontologies. We develop an architecture guided by the premise of exerting a minimum of influence on existing search infrastructures. As a consequence of using a (possibly large) number of ontologies, tools for ontology management are needed. A key argument for using ontologies is that nowadays – as an outcome of the Semantic Web initiative - very powerful processing tools are available. We elaborate ontology mapping as an example and outline how a comprehensive ontology management can be achieved.

Keywords: Public Search Portals, Environmental Terminology, Semantic Technologies, Ontology Management, Ontology Mapping, Life Circumstance Ontology

1 Introduction

Guided by environmental directives at the regional, national and international levels, many environmental agencies nowadays offer free access to distributed environmental information by means of public portals. Typically the underlying Environmental Information Systems (EIS) were primarily designed to provide highly detailed

information and sophisticated user interfaces to thematic experts. They have to be re-focused to provide information in a more restricted and aggregated form to a new user community: the public user, e.g. the citizen who wants environmental information about his surroundings. The provision of accurate, comprehensive and condensed information contained in EIS via public search interfaces is a technological challenge for several reasons:

- EIS usually comprise many sub-systems which hold information in different representation forms and may have dedicated built-in search functions. When retrieving environmental information, it is helpful to link structured information in databases with unstructured information contained in documents.
- For public users, keyword-based search is the most favoured method for information search. In order to integrate responses of query-based subsystems into the search results, text-to-query approaches need to be integrated.
- Public users often do not know the correct search keywords that exactly match stored information which may result in a poor recall.

Our approach to tackle these challenges is based on the utilization of available environmental terminologies as a basic asset. The development of terminologies has been influenced by technological progress and observed trends. The most advanced specification techniques for describing a knowledge domain are developed in the context of the Semantic Web initiative. The World Wide Web Consortium (W3C) has approved standards by which ontologies can be specified in terms of formal logic with high expressivity. In the following sections, we suggest a consequent use of ontologies and explain why this yields considerable advantages. Moreover, nowadays very powerful processing and management tools are available.

2 Enhancing Search Facilities - How Ontologies can Help

2.1 Semantic Search Scenarios – Why are Full-Text Engines not enough?

Nowadays, full-text search engines provide access to various types of information and exhibit excellent performance. However, they still have deficits as soon as an intelligent combination of different information types and information content is required. Users of environmental portals often search for topics with a concrete location reference, e.g. “air quality in the Upper Rhine Valley” or “measurement values for respirable dust in my environment”. Moreover, users with a concrete objective in mind generally need to submit a series of requests to (one or several) EIS and collect and select “manually” the retrieved information which is relevant to take a decision. Consider a citizen who looks for “building land” where he would like to build a house. He is interested in getting an overview about environmental conditions (e.g. transport connections, noise disturbance, air quality, land contamination) in a given geographical area. An ontology describing life circumstances may define a concept structure “build house”/“find location”/“living conditions”/“environmental conditions” that helps to provide links to all content of interest as the result of a single search. The idea is to use keywords entered by a user to identify corresponding

concepts defined in an uploaded ontology. Starting from a concept, other concepts that reside in a “semantic bounding box” determined by ontological relationships (parent, children and property relationships) are additionally looked up in the ontology. This context can be used to provide automatic query expansion and context-based navigation through domain knowledge before submission of the final query.

2.2 Architectural Aspects – Spicing up the Established Search Infrastructure

Technologies emerging in the Semantic Web area still have some academic flavour and their integration into established infrastructures for conventional information retrieval is a challenge. In order to achieve acceptance by organisations that operate an EIS, we favour an approach by which components of the deployed IT infrastructure (e.g. catalogues, search engines) do not need to be replaced. Hence, we propose an architecture which foresees semantic add-ons in terms of pre-processing of user queries before their execution through the deployed EIS search facilities, and post-processing of search results before their final presentation to the user.

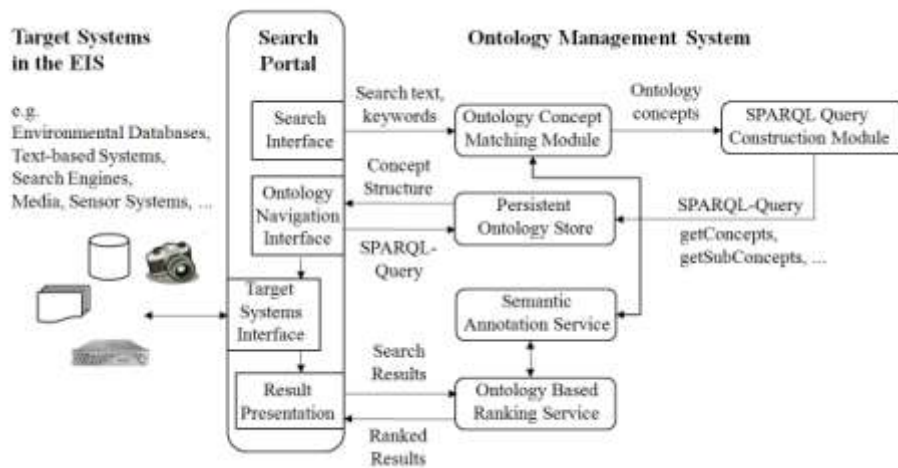


Fig. 1. Pre- and post-processing of keyword based queries via public portals by an Ontology Management System.

The architectural approach is outlined in Fig. 1. We distinguish three main architectural components: the search portal, the target systems in the EIS and the Ontology Management System providing the semantic add-on functionality.

In the first step, search strings entered via the portal are processed by a text-to-query system. In order to resolve relationships between the input text in natural language with ontological concepts, we suggest to use automatic semantic annotation tools as described in [6]. Once corresponding concepts have been identified, queries to the ontology store (e.g. SPARQL queries) can be composed automatically. These queries return the “semantic bounding box” of the identified concepts. The portal

software transforms this ontological structure into a navigation structure which allows the user to explore the semantic context of his query and then issue his request more precisely, e.g. by narrowing or widening his initial request.

In the next step, the portal software constructs requests to the various EIS target systems. As the ontological elements have been defined independently from the interfaces of the target systems, we face another matching problem here; this can be resolved by attaching target system information to the ontological elements. This crucial issue is not further elaborated here; a generic approach based on OpenSearch descriptions is described in a further contribution to ISESS 2011 [18].

Once the results from the target systems have been received by the portal, ontology based post-processing can be applied before the results are finally presented to the user. One way to improve the quality of the search results is to exert influence on the ranking of the results by means of the ontology, i.e. the results are ordered according to their semantic closeness to the search item. For instance, if a user searches for “environmental conditions”, entries about “air quality” are ranked higher despite the fact that the terms “air” and “quality” were not search terms [15].

Finally, the search results from target systems with various forms of information representation (e.g. text-based systems, databases) can be presented uniformly and combined by means of relationships defined in the ontology.

2.3 Ontologies – Where do they come from?

We have validated the developed architecture in various projects, e.g. as frontend/backend of conventional OGC Catalogue Systems [12] and in a project “Semantische Suche nach Umweltinformation (SUI)” of the German environmental cooperation network KEWA [16] for an integrated semantic search over environmental systems in Baden-Württemberg [1][5][18]. Here, we have made two important observations:

- We usually have to deal with a number of ontologies from different sources, rather than a single ontology covering all aspects.
- As ontology development is costly and time consuming, whenever possible ontologies should not be developed from scratch. Especially in the area of EIS, the information to be covered in many cases is available in terms of terminological systems which have formerly been developed under huge effort. Thus, we favour to reuse these systems and spend new effort mainly for development of methods and tools which transform these systems into ontologies and which help to automate the process of updating.

In SUI, we use the following ontologies:

- Environmental thesauri: We use GEMET (General Multilingual Environmental Thesaurus [9]) because it covers a comprehensive base of search terms. GEMET is available as SKOS ontology (Simple Knowledge Organisation Systems [22]).
- Feature Type Catalogues: In order to facilitate data exchange in integrated EIS such as WIBAS [13], so-called feature type catalogues are being defined for all domain areas represented by features that describe real world phenomena. This information is often available as structured data in environmental databases. Although not originally designed for the task of keyword based search, such catalogues provide valuable information, especially for accessing structured

information (e.g. object type codes). We have transformed the WIBAS feature type catalogue into a SKOS ontology such that it can be used in the search and navigation interface.

- Life circumstance ontologies: We have referred to this in the “build house” scenario in section 2.1. Portals of various federal state authorities in Germany offer a navigation through their web pages according to a structure of defined life circumstances, e.g. “Service-BW” operated by the Ministry of Interior of Baden-Württemberg [19] or the “Directory to Administrations” operated by the Bavarian Ministry of Interior [2]. We have extended the structure of Service-BW with concepts that specialise on environmental information issues and made it applicable to semantic search.

The list is open, i.e. there are plans to transform further terminological systems into ontologies in order to feed the semantic search interface. A candidate is the GSBL (Gemeinsamer Stoffdatenpool Bund/Länder), a database of substances operated by the German Federal Government and the Federal States [8].

3 Fostering Integrated Domain Knowledge - Ontology Mapping

Up to now, we have explained that ontologies can guide users to explore available domain knowledge. Naturally, each of the deployed ontologies offers specialised knowledge focussed on the common view of a particular domain. Cross references between domains are not explicated and the terms used to express similar concepts in different domains may differ. However, the ontology system should provide means to make these relationships transparent to the user.

The necessity for ontology mapping arises because query expansions computed for search terms should have a range across multiple ontologies. Consider for instance a search for “building land”: the corresponding concept is found in GEMET and a bounding box will be constructed only from ontological elements contained in GEMET. If we relate the concept “building land” in GEMET to the concept “build house” as described above, the bounding box will additionally contain ontological elements from the life circumstance ontology.

Such mappings can be constructed by means of automatic ontology mapping tools. In the SUI project, we use a tool developed in the German Research Program THESEUS [21] which utilizes Particle Swarm Optimization (PSO) to search for the optimal alignment of ontologies [4]. It was specially designed for parallel execution in distributed systems and exhibits excellent performance especially for large ontologies.

While tools for automatic ontology mappings are available (e.g. in the form of console programs) and can be selected according to specific needs, integration of these tools into a concrete application environment requires additional functionality and customization. Bilateral mappings between ontologies have to be developed and deployed. Moreover, it is important to provide means for manual post-processing of automatically generated mappings; although they can be very helpful and some tools exhibit excellent precision/recall, we do not recommend uncritical confidence in these automatic findings and we recommend to check the consequences in the context of the entire application.

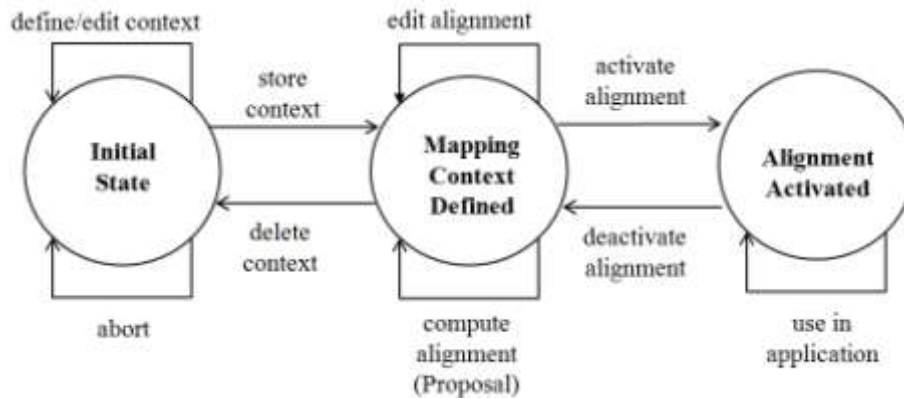


Fig. 2. State Transition Diagram for a Mapping Workflow.

Fig. 2 illustrates a workflow for the management of mappings as applied in the SUI project. Via the administration GUI of the SUI ontology system mapping contexts for bilateral mappings can be defined and stored. For a stored context, a run of the automatic mapping process can be configured and started. It produces an alignment, i.e. a number of correspondences between entities of two ontologies in the format of the Alignment API [7]. Dependent on the size of the ontologies, this process may consume a considerable amount of time. The computed alignment is treated as an initial suggestion of correspondences which can be manually post-processed via the GUI: correspondences can be accepted, rejected, added and changed. Finally, the alignment can be activated. Activated alignments are involved in the execution of the SPARQL queries: the resulting semantic bounding is spawned across ontologies.

4 Driving the Change – A Comprehensive Ontology Management

We have elaborated the aspect of ontology mapping as we conceive it as a key component to support semantic search scenarios. Beyond that, ontologies can provide excellent support for further applications needed in EIS, e.g. navigation in content, interpretation and analysis of content, visualization of relationships or combination/orchestration of environmental information services. The effective utilization of the expressive power of ontologies can be further extended by consecutive incorporation of ongoing developments of domain ontologies, geo-ontologies, top level ontologies and ontologies constructed from any available representation format. This can open up outstanding opportunities to excavate treasures of information hidden deeply in a distributed EIS.

However, there are no benefits without costs: the boosting impact of ontologies can only be leveraged if we can provide the means for a comprehensive ontology management. One important motivation to favour ontologies as the unique representation form is that the Semantic Web is a driving force not only for standardization, but also for the development of the needed management tools. Ontology management covers a wide range of tools, namely for ontology design,

evolution, persistent storage, querying, mapping, reasoning, visualization and provenance management. Moreover, there is a growing number of available tools to extract instance data from various sources such as databases, text-based systems, web pages, media, sensors systems or maps.

In the SUI project we mainly integrate tools developed in the context of the THESEUS program. The THESEUS Core Technology Cluster (CTC) focuses on the development of generic components to be used in several defined use cases. Especially CTC-WP3 is concerned with all aspects of ontology management. At Fraunhofer IOSB we integrate tools developed in WP3 (and other WPs) into our development platform WebGenesis® [22]. We apply it in the shape of an Internet ontology workbench in several projects in the area of EIS. In SUI, we have reused and customized the following components:

- HARMONIA, the THESEUS component for ontology mapping (cf. section 3).
- MNEMOSYNE, a component which addresses the fact that keeping large ontologies in main memory is no longer adequate and an OWL based alternative to the persistent storage in RDF triple stores is needed. It provides a persistency layer for ontologies based on the de-facto standard OWL-API [17]. This standard is used in the well known ontology development tool Protégé in its version 4. MNEMOSYNE has been integrated into Protégé [11].
- PYTHIA, which can convert SPARQL queries into the query language of the OWL-API. We have customized PYTHIA by adding a front end for graphic based construction of queries and integrated it into the ontology workbench.

We plan to integrate further CTC components into the SUI ontology system. The near-term focus will include components for text-to-query conversion, disambiguation of geo-spatial references [14], reasoning brokerage [3] and integration of growing pools of web data (“Linked Open Data” [10]).

5 Conclusions

Finally, it can be stated that semantic search technologies and ontology management tools can help to “open-up the door” for public users to information contained in EIS which was originally designed for domain experts. For implementation purposes we recommend to use technologies, tools and standards (e.g. OWL-API, Alignment API, SPARQL) developed in the context of the Semantic Web initiative. Deployment of innovative tools within an existing EIS infrastructure, even when being state-of-the-art and easily available, requires careful selection, adaptation and operation in order for them to be accepted by the EIS community.

Acknowledgements. The presented work has been funded by the German Federal Ministry of Economics (BMWi) in the THESEUS research programme under grant 01MQ07019 and by the environmental research network KEWA led by the Ministry of the Environment, Nature Conservation and Transport of Baden-Württemberg, Germany.

References

1. Abecker, A. et al.: SUI – Ein Demonstrator zur semantischen Suche im Umweltportal Baden-Württemberg. In: Mayer-Föll et. al. (Eds.): F+E Vorhaben KEWA , Phase IV 2008/2009, Forschungszentrum Karlsruhe, Wissenschaftliche Berichte FZKA 7500, S. 157-166
2. Behördenwegweiser Bayern, see. <http://www.behördenwegweiser.bayern.de/>
3. Bock, J., Tserendorj, T., Xu, Y., Wissmann, J., Grimm, S.: A Reasoning Broker Framework for OWL. 5th Int. Workshop on OWL: Experiences and Directions (OWLED 2009), Chantilly, Virginia, USA (October 2009).
4. Bock, J., Hettenhausen, J.: Discrete particle swarm optimization for ontology alignment. Information Sciences (2010). doi:10.1016/j.ins.2010.08.013
5. Bügel, U. et al.: SUI –Weiterentwicklung der diensteorientierten UIS Infrastruktur für die semantische Suche. In: Mayer-Föll et. al. (Eds.): F+E Vorhaben KEWA, Phase V 2009/2010, Karlsruhe Institute of Technology, KIT Scientific Reports 7544, 43-50 (2010)
6. Bügel, U., Usländer, Th.: Discovery and analysis of environmental information based on formalised terminology. Towards eEnvironment. European Conference of the Czech Presidency of the Council of the EU 2009: Opportunities of SEIS and SISE: Integrating Environmental Knowledge in Europe; March 25-27, 2009; Prague, Czech Republic
7. Euzenat, J.: An API for ontology alignment, in: Proc. 3rd conference on international semantic web conference (ISWC), Hiroshima (JP), Lecture notes in computer science 3298:698-712, 2004
8. Gemeinsamer zentraler Stoffdatenpool Bund/Länder (GSBL), <http://www.umweltbundesamt.de/chemikalien/gefahrstoffe/gsbl.htm>
9. GEMET, <http://www.eionet.europa.eu/gemet>
10. Haase, P., Eberhart, A., Godelet, S., Mathäß, T., Tran, T., Ladwig, G., Wagner, A.: The Information Workbench – Interacting with the Web of Data. Forschungsbericht KIT. 2009.
11. Henß, J., Kleb, J., Grimm, S., Bock J.: A Database Backend for OWL. 5th Int. Workshop on OWL: Experiences and Directions (OWLED 2009), Chantilly, Virginia, USA (2009)
12. Hilbring, D., Usländer, T., Ontology-Based Discovery of Geoscientific Information and Services. Session on Semantic Interoperability, Knowledge and Ontologies in the European Geosciences Union General Assembly in Vienna, 13-18 April 2008
13. Informationssystem Wasser, Immissionsschutz, Boden, Abfall, Arbeitsschutz (WIBAS), <http://www.lubw.baden-wuerttemberg.de/servlet/is/23889/>
14. Kleb, J., Volz, R.: Ontology based Entity Disambiguation with Natural Language Patterns. In: B. Grosky, F. Andres, P. Pichappan (Eds.). Proceedings of Fourth IEEE International Conference on Digital Information Management, ICDIM 2009, November 1-4, 2009, University of Michigan, Ann Arbor, Michigan, USA. pp 19-26
15. Madlener, A.: Entwurf und Implementierung eines Verfahrens für Ontologie-basiertes Ranking von Suchergebnissen. Master Thesis, Karlsruhe Institute of Technology / Fraunhofer IOSB, Karlsruhe, 2010
16. Mayer-Föll, R., Keitel, A., Geiger, W. (Eds.). F+E Vorhaben KEWA Phase V 2009/10, <http://www.fachdokumente.lubw.baden-wuerttemberg.de/servlet/is/5121/2009>.
17. OWL API, <http://owlapi.sourceforge.net/>
18. Schlachter, T. et al.: Towards a Universal Search in EIS. International Symposium on Environmental Software Systems (ISESS), Brno, Czech Republic, 27-29 June 2011
19. Service-BW, see <http://www.service-bw.de/>
20. SKOS Simple Knowledge Organisation Systems, <http://www.w3.org/2004/02/skos/>
21. The THESEUS Programme, <http://theseus-programm.de/home/default.aspx>
22. WebGenesis[®], <http://www.iosb.fraunhofer.de/servlet/is/2223/>