Geophysical Research Abstracts,
Vol. 10, EGU2008-A-08566, 2008
SRef-ID: 1607-7962/gra/EGU2008-A-08566
EGU General Assembly 2008

Ontology-based discovery of geoscientific information and services

D. Hilbring (1), T. Usländer (2)

(1) Fraunhofer IITB, Germany (desiree.hilbring@iitb.fraunhofer.de / Phone +49 721 6091 463), (2) Fraunhofer IITB, Germany (thomas.uslaender@iitb.fraunhofer.de / Phone +49 721 6091 480)

Motivation

In the information society of today, as well as in the geoscientific community, there is no lack of available data and information. Instead the problem is to discover the appropriate information one is searching for. In order to overcome this problem the geoscientific community has developed standard services called catalogues by which the required resources, being information or services, can be found. Geospatial catalogue services provide access to (meta-)information about available geo-spatial resources, e.g. topographical data or geo-statistical processing capabilities. In contrast to Internet search engines such as Google or Yahoo, these geospatial catalogue services explicitly take into account the specifics of geospatial information processing such as spatial queries and geo-referencing of resources. Widely used for this application field are catalogues that have been specified by the Open Geospatial Consortium (OGC) such as the OGC Catalogue Services Specification (Nebert and Whiteside, 2007).

The clients of OGC catalogue services are often application components or human users who know the appropriate search keywords. When used by thematic experts, this approach mostly leads to successful search results. The problem arises if the user does not have this exact pre-knowledge and simply wants to browse through available resources regarding a certain topic domain. In this case, the OGC catalogues may not return useful results as the used keywords often do not match with the meta-information stored in the catalogues.

This paper proposes a semantic extension of such geospatial catalogue services that aims at overcoming this limitation: the Semantic Catalogue realised in the project ORCHESTRA (http://www.eu-orchestra.org). ORCHESTRA is a European Integrated Project running between 2004 and early 2008. On of the key objectives and results of ORCHESTRA is the specification of an Open Service Architecture for Risk Management (Usländer (ed.), 2007). The ORCHESTRA Semantic Catalogue Service is an artefact of this architecture. It provides means for the semantic exploration of contents of conventional catalogue services (Hilbring and Coraboeuf, 2007).

Firstly, the architecture of the semantic extension is described. Secondly, the requirements for the ontology used as foundation for the semantic extension are discussed. Furthermore, the necessary extensions for the catalogue interface and a first approach of an ontology-based ranking of the search results are presented. The paper concludes with a description of the semantic catalogue as a central portal application that provides a uniform search interface to the ORCHESTRA pilot applications.

Architecture of the Semantic Catalogue

The ORCHESTRA Semantic Catalogue is composed of three architectural levels (Hilbring and Usländer, 2006): source meta-information systems (L1), cascaded catalogues (L2) and semantic catalogues (L3).

L1 comprises conventional catalogues and/or other meta-information sources. Examples for conventional catalogues are OGC-compliant catalogue services or the functionally-similar "conventional" interfaces of the ORCHESTRA catalogue service. In addition, an Internet search engine such as Google or Yahoo is also considered as a source of meta-information (e.g. about documents and web sites) on this architectural level. An ORCHESTRA catalogue service can be directly integrated at this level, while OGC Catalogue Services are made available for the semantic catalogue by means of an ORCHESTRA wrapper around the OGC Catalogue. Two wrappers around the two predominant variants of OGC Catalogue Services, one with a meta-information schema according to ISO19115/ISO10119 and one based on ebRIM (ISO/TS 15000-3), have been implemented. These catalogues all support search functionality using a typical geospatial catalogue query language such as OGC Filter Encoding based on XML filter expressions. A filter expression constrains property values to create a subset of a group of objects (OGC Doc.No. 04-095). Often the conventional catalogues also provide publishing functionality realized either with push and/or pull mechanisms. These functions, also called "harvesting" operations, enable the automatic update of the catalogue according to the change frequency of the resources underneath.

Level L2 combines the underlying L1 meta-information systems in a cascade whereby

each L2 system in this cascade provides an interface that is compliant to the ORCHESTRA Catalogue Service specification. Cascading means that the queries are individually propagated to known L1 meta-information systems whereas the query results are combined in one response.

L3 realizes the semantic extensions of the catalogue but still provides the conventional ORCHESTRA Catalogue Service interface in order to maintain the upwards compatibility. This approach ensures that all architectural levels can be accessed by conventional ORCHESTRA Catalogue Clients. Thus, only clients that want to exploit the semantic extensions need to adapt or extend their functionality (e.g. in the user interface).

The role of Ontologies in the Semantic Catalogue

The goal of the inclusion of ontologies into the catalogue is two-fold: On one hand, it is the basis for automatic or interactive query expansion, on the other hand it is used for the ranking of the search results. Both goals require that the ontology is context-specific, i.e. it should reflect the major concepts and relations of the thematic domain in which the user is currently interested. Context-specificity may be understood in several ways: either specific to a particular instance of the semantic catalogue service, specific to a user (personalisation) or even specific to a given query. Overall there is a need that the ontology used in the semantic catalogue is not fixed but may be configured, either by an administrator or by the user himself at installation or at run-time. As mentioned above, it is the goal of the Semantic Catalogue to integrate and re-use conventional catalogues in level L1 without any semantic adaption. However, there is a need to match the thematic concepts contained in ontologies with the keywords and queryables of the underlying conventional catalogues. The solution to this matching problem, as currently realised, is the addition of keyword-oriented labels in the definition of the ontological concepts.

Extensions of the Catalogue Interface

For the query expansion, a new operation is made available through the ORCHESTRA Catalogue Service interface - the improveQuery operation. This operation receives a conventional catalogue search query as input parameter. The keyword contained in the query is used to identify the corresponding concept in the ontology. Starting from this concept, other concepts that reside in a kind of “semantic bounding box” determined by subsumption relationships (parents, children and property relationships) are looked up in the ontology and given back in the response to the improveQuery operation. The depth of the semantic bounding box may be determined by the user, e.g. a semantic bounding box of depth 1 will only return the direct parent and child concepts as well as those concepts that are directly related to the starting concept by a given property

relation. The improveQuery operation enables both automatic and interactive query expansion. In the automatic mode, the catalogue expands the terms of the query according to the depth of the semantic bounding box and sends individual queries for each of the found concepts (resp. the keywords attached to them within their labels) without any further user interaction. It returns a semantic report sorted firstly by keywords and secondly by type of resource (service, data or document). The interactive mode allows the user to select the expanded concepts that should be used for the query. It then issues the query for the selected concepts (resp. the keywords) and assembles the results into a common result set.

A further functional extension of the semantic catalogue comprises operations for ontology management. By means of the operations setOntology and getOntology it is possible for semantic catalogue clients to upload and select the ontology to be used by the Semantic Catalogue for all or for the current queries. This approach allows the user to personalise the usage of the catalogue.

Ontology Based Ranking of Catalogue Search Results

The Semantic Catalogue uses the ontology also for an optional search extension - the ranking of search results. Currently the following simple ranking algorithm is implemented. Firstly, the ranking extension uses the improveQuery operation to identify semantically related keywords to literals contained in the original catalogue query. Secondly, it searches for newly identified keywords in the conventional catalogue search result and calculates weightings for hits. Thirdly, the search result is sorted according to the assigned weightings. Finally the ontology based ranked result is given back by the search operation.

Pilot Implementation of the Semantic Catalogue in the ORCHESTRA Project

The presented approach has been successfully validated in an ORCHESTRA pilot application that provides a portal for people who are interested in exploring the achievements of the ORCHESTRA project. This installation of the semantic catalogue provides a cascaded access to all catalogues realized in the application pilots of the ORCHESTRA project, and, in addition, to the Internet via an adaptor to the Yahoo search engine. It thus acts as a federated pilot that provides uniform search access to geospatial catalogues that address different risk management issues (e.g. forest fire risks, marine risks, floods) as well as to information accessible in the Web (e.g. ORCHESTRA publications) which may be thematically associated to the pilots. Thus the ontology used in this semantic catalogue installation contains domain concepts dedicated to risk management, ORCHESTRA architectural concepts (e.g. service and interface types) as well as organisational concepts that reflect the structure of the ORCHESTRA project. Guided by this ontology, a user can now explore the project structure and the

risk management topics addressed without having to be familiar with the project and the thematic domain of risk management beforehand. By means of query expansion and query improvement the user is led to the concepts and related keywords which have a high chance to be matched with related entries in the underlying conventional catalogues. Associated accompanying information may be found in the Web using the same user interface.

Conclusion

This paper shows the practical usage of ontologies for an integrating multi-level geospatial catalogue service that has been specified and implemented in the ORCHESTRA project. The resulting semantic catalogue is a functional extension of the OGC catalogue service and provides ontology-based query expansion and a first trial of ontology-based result ranking. It has been validated in an portal for environmental risk management applications. Further experiments of its usefulness will be applied in the area of sensor service networks in the European Integarted Project SANY (Sensors Anywhere, http://sany-ip.eu). Here, resource discovery is required for sensors, its observation offerings and features of interests. Further research work will focus on the enhancement of the ontology-based ranking by investigating the applicability of similar work for Web documents (Fang et al, 2007) and the usage of the semantic catalogue as a supporting tool for the design of geospatial applications.

References

Fang, J. L. Guo, X. Dong and N. Yang, 2007. Ontology-based Automatic Classification and Ranking for Web Documents. International Conference on Fuzzy Systems and Knowldegde Discovery (FSKD 2007), IEEE Computer Society.

Nebert, D. and A. Whiteside. 2005: OpenGIS Catalogue Services Specification, OpenGIS Implementation Specification, OGC 07-006r1

Hilbring, D. and D. Corabouef, 2007. ORCHESTRA, OA-Service Specification, Specification of the Catalogue Service, http://www.eu-orchestra.org/docs/OA-Specs/Catalogue_Service_Specification_v1.1-IITB.pdf

Hilbring, D. and T. Usländer, 2006. Catalogue Services Enabling Syntactical and Semantic Interoperability in Environmental Risk Management Architectures. Proceedings of the 20th International Conference on Informatics for Environmental Protection (EnviroInfo 2006), September 6-8, 2006, Graz, Austria, ISBN-10:3-8322-5321-1, pp 39-46.

Usländer, T. (ed.), 2007. Reference Model for the ORCHESTRA Architecture Version 2 (Rev. 2.1). OGC Best Practices Document 07-097.

http://portal.opengeospatial.org/files/?artifact_id=23286.