Prod-Trees: Semantic Search for Earth Observation Products

M. Karpathiotaki[1](✉), K. Dogani[1], M. Koubarakis[1], B. Valentin[2],
P. Mazzetti[3], M. Santoro[3], and S. Di Franco[3]

[1] National and Kapodistrian University of Athens, Athens, Greece
{mkarpat,koubarak}@di.uoa.gr
[2] Space Applications Services, Zaventem, Belgium
[3] Institute of Atmospheric Pollution Research - National Research Council,
Rome, Italy

Abstract. Access to Earth Observation products remains difficult for
end users in most domains. Although various search engines have been
developed, they neither satisfy the needs of scientific communities for
advanced search of EO products, nor they use standardized vocabu-
laries reusable from other organizations. To address this, we present
the Prod-Trees platform, a semantically-enabled search engine for EO
products enhanced with EO-netCDF, a new standard for accessing Earth
Observation products.

1 Introduction and Motivation

The demand for aerial and satellite imagery, and products derived from them
has been increasing over the years, in parallel with technological advances that
allow producing a bigger variety of data with an increasing quality and accuracy.
As a consequence of these advances, and the multiplication of deployed sensors,
the amount of Earth Observation (EO) data collected and stored has exploded.

However, access to EO products remains difficult for end users in most sci-
entific domains. Various search engines for EO products, generally accessible
through Web portals, have been developed. For example, see the interfaces
offered by the European Space Agency portal for accessing data of Coperni-
cus, the new satellite programme of the European Union[1] or the EOWEB por-
tal of the German Aerospace Center (DLR)[2]. Typically these search engines
allow searching for EO products by selecting some high level categories (e.g.,
the mission from which the product was generated, the satellite instrument that
was used etc.) and specifying basic geographical and temporal filtering criteria.
Although this might suit the needs of very advanced users that know exactly
what dataset they are looking for, other scientific communities or the general
public require more application-oriented means to find EO products.

---

[1] http://gmesdata.esa.int/web/gsc/home
[2] https://centaurus.caf.dlr.de:8443/eoweb-ng/template/default/welcome/entryPage.
vm

In this demo paper, we present a semantically-enabled search engine for EO-products currently under development by the project Prod-Trees funded by the European Space Agency. The system uses semantic technologies to allow users to search for EO products in an application-oriented way using free-text keywords (as in search engines like Google), their own domain terms or both, in conjuction with the well-known interfaces already available for expert users. A specific innovation of the presented system is the use of a new standard called EO-netCDF, currently under development in Prod-Trees and expected to be submitted to OGC, for accessing EO products annotated with netCDF. netCDF is a well-known standard consisting of set of self-describing, machine-independent data formats and software libraries that support the creation, access, and sharing of array-oriented scientific data.[3]

The Prod-Trees system has been developed using state of the art semantic technologies developed by the partners of the project: the company Space Applications Services, the National and Kapodistrian University of Athens and the research institute CNR.

2   The Prod-Trees Platform

The Prod-Trees platform is a semantically-enabled EO products search engine. It allows end-users to search for EO products using filtering criteria provided by the EO-netCDF specification and the EO vocabulary designed and implemented in the Prod-Trees project[4]. Figure 1 depicts the architecture of the platform, which partially re-uses components from the RARE platform[5].

The web interface of the Prod-Trees platform allows the users to submit free-text queries, navigate to the ontology browser, select applications terms defined in the supported ontologies and finally, search for EO product by specifing EO-netCDF parameters and controlled (bounding box, time, range) search criteria.
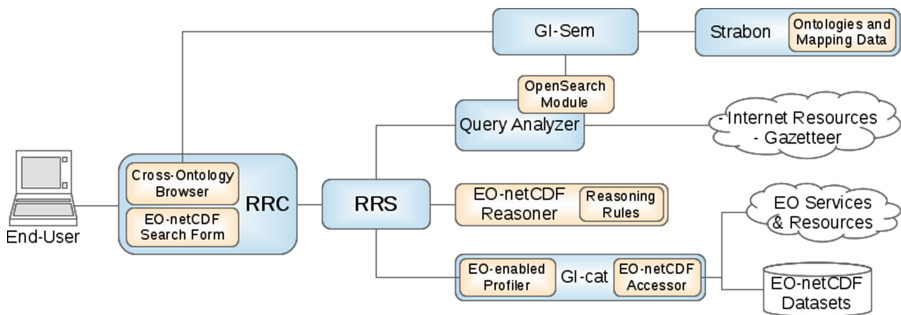

Fig. 1. The Prod-Trees platform architecture

When the user has filled the search form, the Query Analyzer is responsible for displaying a number of different interpretations for the inserted free-text. After the user has selected the semantics she wants to be used for the search, the backend service is called, generates one or more queries and sends them to GI-cat through its EO-netCDF Profiler. GI-cat searches for the matching EO products and returns back the metadata. Depending on the nature of each product (JPG, XML, HDF, etc.), this may be either visualized on-line or downloaded on the local system. The following paragraphs describe in more detail the components of the Prod-Trees architecture and their interaction.

The ***Rapid Response***[6] ***Client (RRC)*** provides the user interface to the Prod-Trees platform and communicates with several backend services. It displays a search form, where a user can give as input EO-specific search criteria or free text and can navigate to the supported ontologies through the ***Cross-Ontology Browser***. This component is a browser for ontologies expressed in SKOS that allows the users to exploit the knowledge contained in the supported ontologies. It provides relevant information for each concept and highlights the connections between different (but related) concepts belonging to the same or other ontologies. Its role is to support the user in the query creation phase, as a disambiguation and discovery tool. The browser is accessed via the RRC search page.

GI-Sem [4] is a middleware which is in charge of interconnecting heterogeneous and distributed components. Its main role in the Prod-Trees platform is to create a connection between the Cross-Ontology Browser and the supported ontologies. GI-Sem performs remote queries to Strabon and returns the results to the Cross-Ontology Browser. It can also be omitted from the system by using a version of the Cross Ontology Browser that calls Strabon directly.

Strabon [3] is a well-known spatiotemporal RDF store. It holds the supported ontologies and the cross-ontology mappings appropriately encoded in RDF. The supported SKOS ontologies are the GSCDA, GEOSS, GEMET and NASA GCMD. The mappings between these ontologies were created using an algorithm developed in the scope of Prod-Trees [2].

All the interactions with the backend modules go through the ***Rapid Response Server (RRS)***. In case a query string entered by the user need to be disambiguated, the RRS invokes the ***Query Analyzer (QA)***. The QA processes the query string, identifying the words that may be mapped to application terms, location names (toponyms), time constraints, or other types of named entities. In order to carry out this task, the QA interacts with GI-Sem (using an OpenSearch[7] interface), Internet Resources such as gazetteers, as well as external databases such as Wordnet.

After the disambiguation process, if the user has selected an ontology concept, the RRS interacts with the ***EO-netCDF Reasoner*** to obtain the filter criteria

for the search. The reasoner uses reasoning rules to map an ontology concept to EO-netCDF search criteria. These rules have been built manually with the consultation of experts in the context of the project Prod-Trees and the previous project RARE. RSS uses the returned results to build an appropriate query that is sent to GI-cat.

GI-cat [1] is an implementation of a catalogue service, which can be used to access various distributed sources of Earth Observation products. In Prod-Trees, it has been extended to support products compliant with the EO-netCDF convention. Thus, it provides an EO-netCDF enabled discovery and access engine, so that products annotated with EO-netCDF are searchable and accessible to the users.

3   Demonstration Overview

We will now present the core scenarios that we plan to demonstrate at ESWC.

In the first scenario the user inserts a free-text query, for example "water". The system replies by presenting a number of different interpretations for the inserted text, which are provided by the Query Analyzer during the disambiguation phase. This way it is clear for the user what are the semantics of the text on which the search will be based. The default interpretation for "water" maps this text to the concept "water" of GSCDA ontology. In case the user is not satisfied with this interpretantion, she can select another one from a proposed list, for example "water use", "water temperature", "ocean level" and more. Another option is to use the inserted text without any specific interpretation. In this case, a simple text-based search will be performed. The EO-netCDF reasoner is used to map the concept "water" of GSCDA to EO-netCDF parameters with specific values. This is done using appropriate mapping rules which allow us to connect concepts of an ontology (in this case water of GSCDA) to EO-netCDF parameters with specific values (in this case combinations of satellite sensor type, resolution, polarization etc.). As a result, GI-cat returns only the EO products that include EO-netCDF parameters with these values. Figure 2 displays the first two results of the keyword search for "water".
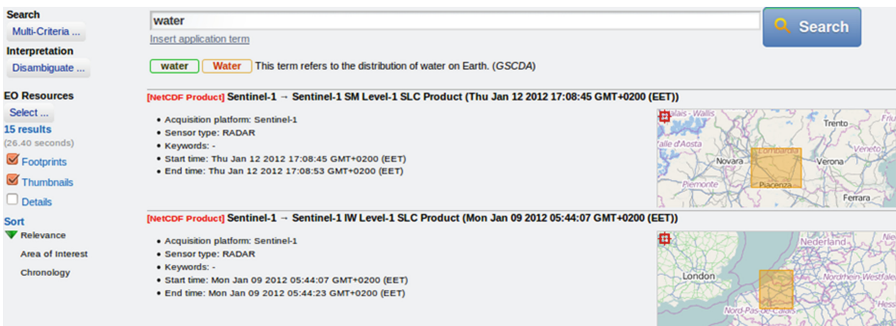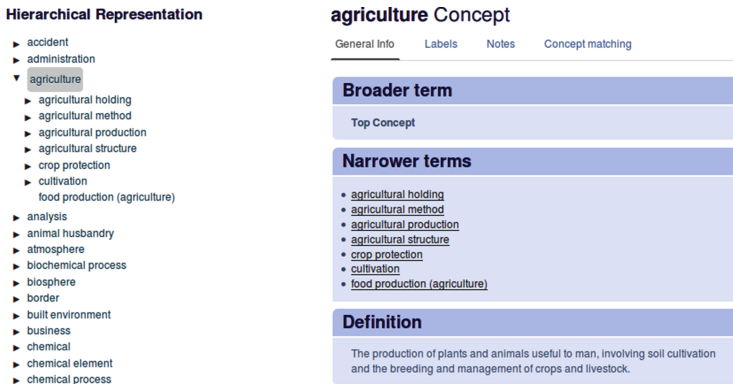

Fig. 2. Search results for the keyword "water"

Fig. 3. The Cross-Ontology Browser displaying the GEMET ontology

Instead of the text queries, the user can also use the ontology browser to select terms he wants to search for. Figure 3 displays the interface of the browser. The selected concept is copied back to the initial text area. Assuming the user has selected the concept "agriculture" of GEOSS ontology, she can add then more keywords (toponyms, date etc.) to the text area in order to restrict the search, for example "agriculture Bahamas 2010". Keywords with toponyms are also disambiguated using the Geonames gazetteer. Afterwards, the workflow is similar to the one described above.

Finally, the third scenario will show how to search using EO-related search criteria. This option might be more appropriate for expert users. In particular, the user can search using specific metadata attributes such as sensor type, bounding box, time, etc. and by specifying one or more EO-netCDF parameters. The search will be based on these attributes and will return only EO products that satisfy them. For example, selecting the parameter "Sensor Type" an optional value would be "optical" or "radar". As the EO-netCDF parameter is provided directly by the user, the EO-netCDF reasoner is bypassed and only the GI-cat component is invoked to return the relevant resources.

A video demonstrating the above functionality is available at http://bit.ly/ProdTreesPlatform.

References

1. Boldrini, E., Nativi, S., Papeschi, F., Santoro, M., Bigagli, L., Vitale, F., Angelini, V., Mazzetti, P.: GI-cat Catalog Service ver. 6.0 Specification. Draft Specification
2. Karpathiotaki, M., Dogani, K., Koubarakis, M.: Ontology Mapping on netCDF. Prod-Trees Technical Note
3. Kyzirakos, K., Karpathiotakis, M., Koubarakis, M.: Strabon: a semantic geospatial DBMS. In: Cudré-Mauroux, P., et al. (eds.) ISWC 2012, Part I. LNCS, vol. 7649, pp. 295–311. Springer, Heidelberg (2012)
4. Santoro, M., Mazzetti, P., Nativi, S., Fugazza, C., Granell, C., Diaz, L.: Methodologies for augmented discovery of geospatial resources. In: Discovery of Geospatial Resources: Methodologies, Technologies and Emergent Applications, chap. 9