AN ONTOLOGY TO SUPPORT KNOWLEDGE ENABLED SERVICES ON EARTH OBSERVATION

A. Pereira(1), R. A. Ribeiro(1), P. A. Sousa(1), M. Pantoquilho(1), P. Bravo(1), A. Falcão(1), S. D'Elia(2)

(1)UNINOVA- CRI/CA3, Caparica 2829-516, Portugal, Email: rar@uninova.pt

(2 ESA/ESRIN, Frascatti,Italy Email: sergio.delia@esa.int

ABSTRACT

This paper discusses an ontology-aided approach to EO product search. A specific user domain ontology, an EO domain ontology are proposed and linked. The user domain ontology covers the case studies of water quality and maritime security while the EO domain covers oil-spills, algae bloom, ship detection and wind and waves information.

This work is based on AI technologies such as ontologies, knowledge based systems, knowledge discovery and data mining.

The ontology described is part of the project *EO-KES-B - Earth Observation domain-specific Knowledge Enabled Services* under development for the ESA/ESRIN (European Space Agency/ European Space Research Institute).

1. INTRODUCTION

A quick analysis of the Earth Observation (EO) domain would provide a straightforward and trivial conclusion: this area of research and applications suffer from the same "data glut" problem found across other domains. Data acquisition capabilities have increased in impressive ways (e.g. by way of new instruments, higher data rates, better precision and coverage) where as data analysis has only marginally improved by comparison. The transformation of raw products into meaningful information, such as detection of an oil-spill or warning of an impending algae bloom is, for the most part, performed by experts in slow and expensive ways, hard to repeat and automate. EO needs rightly belong in the same overall motivational backdrop that led to the appearance of data mining and Knowledge Discovery in Databases (KDD) in the 1990's [2-1].

In parallel with other domains, two categories of research can be identified in EO; one is the automated production of EO products at higher levels of abstraction, closer to the concerns of "automatic extraction of patterns in data" of data mining and KDD [3]; the second includes searching for relevant products, events, resources associated with a given user query, a possible reformulation of information retrieval goals in terms of EO specifics. In this paper we focus in the problem of searching for ESA products on an ontology-aided approach to EO product search.

1.1. The knowledge gap from user domain to EO domain

To use EO products (e.g. a SAR (Synthetic Aperture Radar) image or a photo in the visible spectrum range) for a concrete problem is far from a trivial issue of acquiring EO data at a distributor. The final-user, in most cases, needs a processed product and never the original base EO product. For example, to detect an oil-spill a SAR image must be used to pinpoint a potential oil-spill area. Additional data about winds and waves is required and only after merging all the products (the SAR image and the winds and waves data) and with some expert assistance can the oil-spill be identified with a reasonable accuracy. Therefore, if a user has no expertise in EO terminology and available products just the initial task of searching for relevant products is considerably challenging.

To summarize this, a useful way of thinking about the problematic of EO product search is structured in Fig. 1.

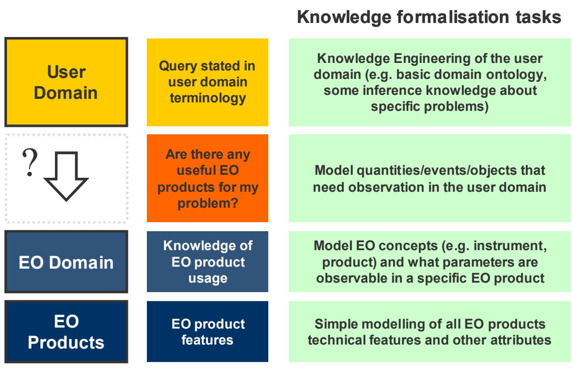

Fig. 1. The user domain to EO domain knowledge gap.

An initial query, stated in some user domain terminology, with no a priori connection to EO domain, must be somehow "translated" into associated EO products which were the target of a focused product search. Note how our proposal suggests overcoming the knowledge gap using knowledge of user domain and EO domain semantics, available after explicit knowledge formalisation. Moreover, some inference knowledge may also be needed, for instance when modelling the consequences to natural resources of an oil-spill. This knowledge is used afterwards to make inferences about the input query and find associated concepts that indirectly connect the user query to EO products. Other approaches attempt to induce terms and relations between them using a corpus of documents [10]. However several factors influenced our decision of a classical knowledge engineering approach:

- clear requirement/goal of developing a collaborative environment where experts and non-user experts can share their knowledge over time
- facilitating agreement on terminology for EO product usage
- support for users with no previous exposure to EO terminology
- lack of a representative corpus of EO documents to use.

1.2. The ESA KES-B project

The overall context of the work presented here is the ESA/ESRIN (European Space Agency/ European Space Research Institute) project KES-B: *EO-KES - Earth Observation domain-specific Knowledge Enabled Service,* code*:* EOKES_PROP_001_1-0OF02/4456**.** This project was developed by the consortium UNINOVA (PT), GTD (SP), STARLAB (SP), UTV (IT).

The project deals with the artificial and (semi-) automatic reproduction (in an Earth Observation oriented domain) of several of the following human being capabilities and processes: knowledge capture; knowledge reception; knowledge archive; knowledge retrieval; knowledge organization; and knowledge application. Hence, the 'services' (i.e., the transformations) provided and/or supported by the EO-KES system are referred to as 'Knowledge Enabled' ones. The goals are to develop:[i] Specific 'knowledge access' interfaces - thus assuming that it shall be feasible to standardize knowledge formalization. De-facto standards to represent wide types of knowledge are now available, from rule and complex ontologies to neural networks. [ii] 'Knowledge services', categorized as general purpose to grant the effective handling of the knowledge.[iii] 'Domain Knowledge Enabled Services', which are those applications specific constructed with all the general purpose available services (data, information and knowledge ones). [iv] 'Applications: knowledge exploitation and formalization'. Resulting from the combination of knowledge services as well as applications which allow the system to be 'instructed', support (supervised/unsupervised) learning or - in general - bring knowledge into the system. [v] 'HMI knowledge formalization and application'.

The reader is referred to [16] for further details on the KES-B project.

2. OVERVIEW OF ONTOLOGIES FOR INFORMATION SYSTEMS

2.1. Why use ontologies

In the computer science community the term ontology has a more constrained meaning, connected with knowledge sharing and reuse. Gruber speaks of “an explicit specification of a conceptualisation” [6] Others might be more specific, for example requiring the conceptualization to be formal and shared [15] or as Guarino puts it “a logical theory accounting for the intended meaning of a formal vocabulary, i.e. its ontological commitment to a particular conceptualisation of the world” [7].

An ontology is similar to a dictionary or glossary, but with greater detail and structure that enables computers to process its content. It consists of a set of concepts, axioms, and relationships that describe a domain of interest [5].

The simplest form of an ontology is a taxonomy. However, ontologies do not define a simple set of keywords: they structure the information. With structured information it is possible to use ontologies for [8]:

- *consistency checking*: if ontologies contain information about properties and value restrictions on the properties, then type checking can be done within applications;
- to provide *completion*: an application may obtain a small amount of information from a user, such as the fact that she is looking for a high-resolution screen on a pc, and then have the ontology expand the exact pixel range that is to be expected; or it can be adaptive;
- to provide *interoperability support:* we may have a complete operational definition for how one term relates to another term and thus, we can use equality axioms or mappings to express one term precisely in terms of another and thereby support more “intelligent” interoperability;
- to support *validation and verification testing of data (and schemas)*: if an ontology contains class descriptions, these definitions may be used as queries to databases to discover what kind of coverage currently exists in datasets ;
- to support *structured, comparative, and customized search*: if an ontology contains mark-up information it can be used to prune comparative searches and to point which properties are most useful to present in comparative analyses so that users may have

concise descriptions of the products instead of comparisons in complete detail;

- to exploit *generalization/specialization information*. If a search application finds that a user's query generates too many answers, one may dissect the query to see if any terms in it appear in an ontology, and if so, then the search application may suggest specializing that term.

The above list is not exhaustive and its purpose is only the illustration of some ways that ontologies have been used to support intelligent applications. More information on these topics can be found on [0].

The use of ontologies greatly surpasses the simple use of keywords, due to the structured and adaptive information representation supported by an ontology. Moreover, ontologies allow the comparison and inference of conformed knowledge based on other ontologies, providing the means to global, transparent information sharing.

2.2. From Ontology to Knowledge Base

A knowledge base is an informal term for a collection of information that includes an ontology as one component. Besides an ontology, a knowledge base may contain information specified in a declarative language such as logic or expert-system rules, but it may also include unstructured or unformalized information expressed in natural language or procedural code [14].

A knowledge base should provide sufficient expressive power to represent human knowledge as well as an efficient, powerful, and understandable reasoning mechanism.

3. AN ONTOLOGY-AIDED APPROACH TO EO PRODUCT SEARCH

3.1. A design pattern for ontology design

As stated previously, in this paper's scope we will mainly be concerned with using an ontology to improve the capabilities of EO product search. Note how this already implies several types of roles, according to McGuiness classification. The query system will use some user domain ontology and an EO domain ontology to initially perform query completion (by supplying additional associated terms, query expansion (by adding associated terms to the initial ones). After settling in a set of terms to search and potentially a set of derived queries, the search process uses the ontology to infer relationships between concepts, for instance when using the captured relation that a SAR image can be used to detect an oil-spill and suggests that product from an input query of simply "oil-spill".

Before tackling the issue of how to use an ontology to improve the search process we must first consider the task of modelling expert knowledge and build the starting ontology. The experience of conducting knowledge engineering has demonstrated the need for a structured approach to it, one that attempts to find opportunities for modularity and reuse, discarding the view of a set of rules as structured by simply assuming rules as modular in themselves. The broad design principle is well demonstrated by the appearance of established knowledge engineering methodologies and the discovery of templates for knowledge intensive tasks [12].

A design pattern for structuring a knowledge base, inspired in Gruber's design principles [6] is helpful in guiding us through the rationale of our model (v.d. Fig. 2 **Error! Reference source not found.**). In this design pattern, a start-up generic KB contains all the constructs we use to build an ontology (e.g. some kind of class mechanism with formal is-a relations, a generic relation, time, space, etc). This part is composed of a state dependent part ("core" KB) and a state-independent part (generic ontology). The generic ontology is called meta-ontology or top-level ontology in other design patterns. Top-level ontologies are currently the focus of several proposals to adopt a specific top-level ontology as a standard, in the expectation that this will promote wider knowledge sharing and reuse (e.g. the CyC project [4]). These are the building blocks in which a domain expert can express a particular domain ontology. As shown in Fig. 2 the generic KB is, in this example, shared by a large set of scientists, including scientists from completely different domains.

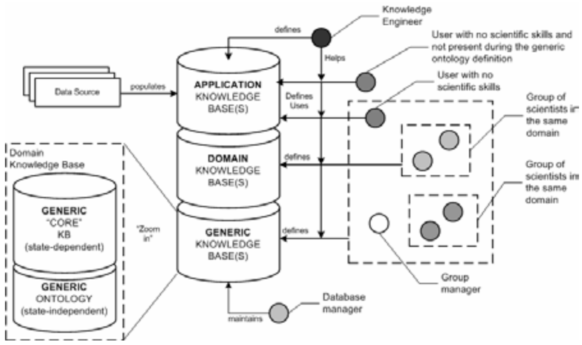

Fig. 2. A design pattern for ontology design using Gruber's design principles.

Next the domain KB contains essentially a set of domain ontologies. Each domain ontology is shared inside a group of scientists and is designed to be reusable outside a specific information system. This materializes a true-shared conceptualization and provides the most important knowledge for improving a search procedure over simple keyword search. Finally, application-tied concepts are maintained in the application KB, which is the least reusable component, andis obviously connected to a specific information system's goal and respective data sources.

This design pattern is not without its issues, since we are not presenting here the existence of intra-domain relations. These are important since they allow us to navigate from a specific user domain ontology to the EO domain ontology to find relevant products. Therefore no clean modularity will exist in real system and there is always some level of interaction. Nevertheless this discussion is important for us to conclude on the needed components for an ontology-aided EO product search: a generic ontology with the base concepts; a concrete user domain ontology and an EO domain ontology; a specific ontology for the information system (in our case KES-B) supporting the search.

The generic part is usually already available and too abstract to serve us in describing the search process. The application-oriented is related with software engineering issues and requirements of a particular information system. These two parts support the domain ontologies where most of the search process occurs. The end-user has selected for the KES-B prototype two user domains: water quality (with oil-spill and algae bloom detection) and maritime security (with ship detection and winds & waves information). This totals four different case studies, however due to the tight connection between the selected areas, a single domain ontology can capture the most important concepts of the two domains. In fact, they were selected because a small set of EO product types are used in all of them.

In the next section, we describe a simplified view of the user domain ontology, the EO domain ontology and the relations between them.

3.2. An ontology of the EO domain and a case study user domain ontology

To construct the ontology we used the Protégé 2000 system ontology editor [11]. The model is presented here using an object oriented notation. The curved lines are regular associations, which are implemented using the association class mechanism. They are our relations in the ontology. We are only describing classes and their names in this object-oriented model, so the concrete instances and their relations with other instances are not shown. In a sense we are providing a partial view of the full ontology that only contains the essential concepts and structure.

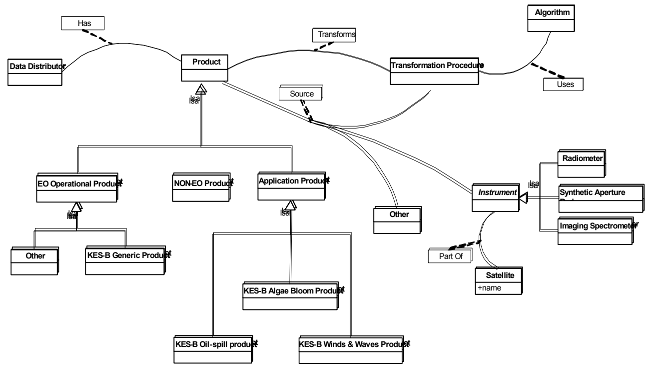

Fig. 3. A simple EO domain ontology.

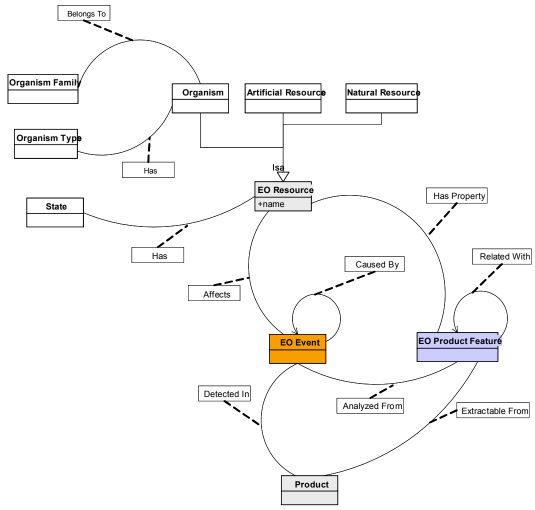

Fig. 4. A set of key concepts from a user domain connected with the EO

Fig. 4 exemplifies a base EO domain ontology with a modelling of EO products and their sources. Following this, Fig. 3 exemplifies how a small set of user domain concepts can be defined and interlinked with the EO domain. Finally, Fig. 5, shows a view of how the full model looks like. To better understand this sequence, a model guided tour is provided in the next section.

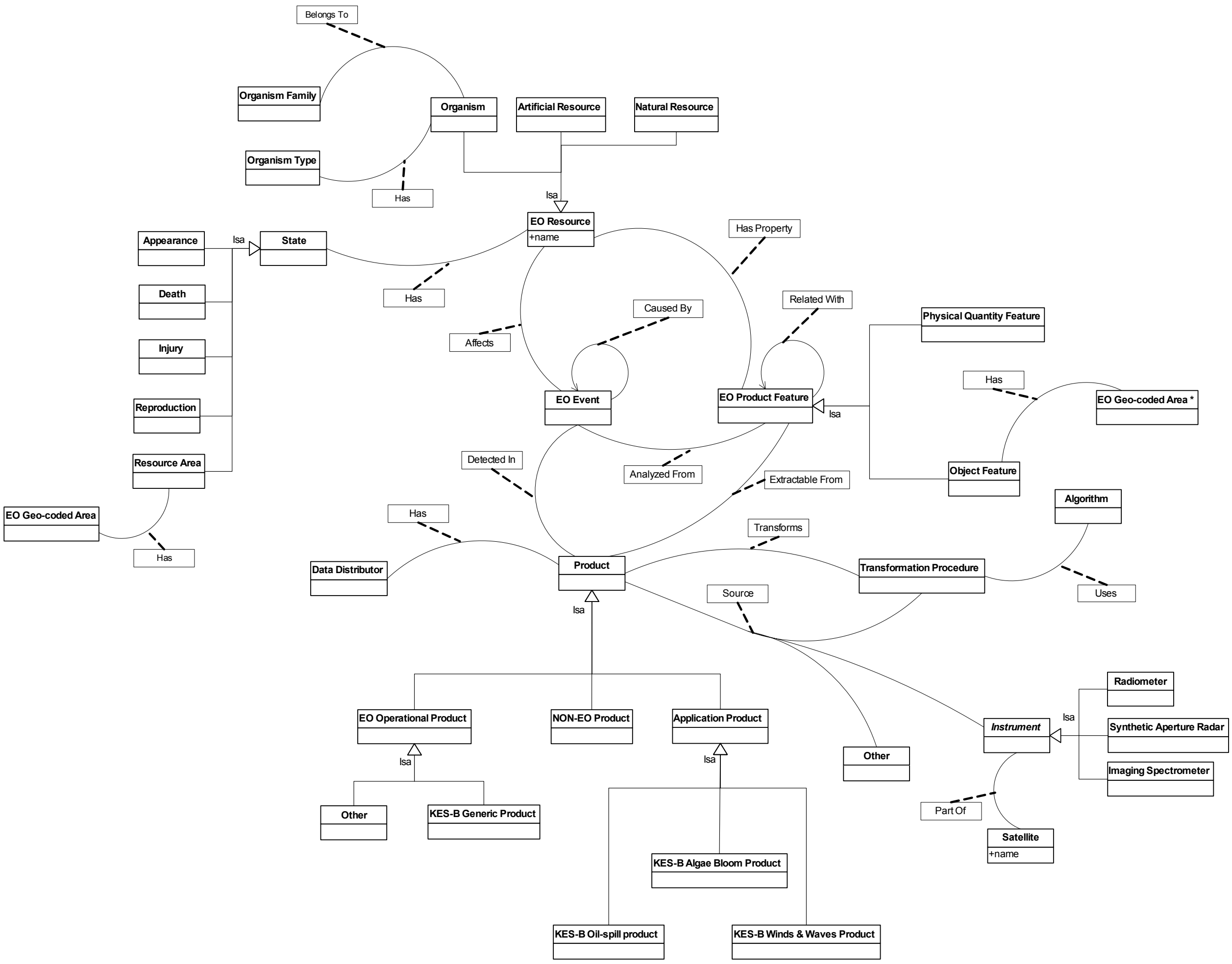

Fig. 5. A view of the full model.

3.3. Model guided tour

To facilitate the model description the notations used are: for concepts we use bold and for relations underscore.

A **product** can have three types of <u>sources</u>**:** it can be from an **instrument** or from a **transformation procedure** (when it is a product originated by a specific algorithm) or from any **other source**. A. A product can be of four types: a transformed one (**transformation procedure** <u>transforms</u> input products <u>using</u> **algorithms**); an **EO operational product**; **application products** (e.g. oil-spill products, algae bloom products and winds & waves products); and **non-EO products**. In addition a **product** always <u>has</u> a **data distributor**.

The concept of **EO product feature** is <u>related-with</u> other features; it is organized in **physical quantity features** or **object features** and it is <u>extracted-from</u> **product**. In addition, **object features** have an **EO geo-coded area** associated with it.

An **EO event** (e.g. oil spill, dead-fish) <u>affects</u> an **EO resource,** is <u>analysed-from</u> **EO product features,** is <u>detected-in</u> **product** and an **EO event** can also be <u>caused-by</u> other **EO events**. Each **resource** has a set of **states** that can be altered by the occurrence of an **event**, which are: **appearance** e.g. the colour of the water; **death** e.g. dead fish; **injury** e.g. some algae may cause injuries in the skin of some fish; **reproduction** e.g. increase in fish stocks; and the **resource area** is given by the **EO geo-coded area**. An **EO resource** also <u>has-property</u>, denoted as **EO product-features** (e.g. ocean colour**,** chlorophyll**)**. The **EO resources** are organized in three categories: **Organism**, **Artificial** and **Natural**. The **Organism** is of a specific type (*algae*, *animalia*, *fungi* or *plantae*) and belongs to a **family**, e.g., the algae can be of the *rhodophyta* family. The **Natural** category of **resources** can include water, land and atmospheric resources, such as rivers and oceans. The **Artificial** category contains all other resources that are "man-tempered".

3.4. EO product search

With a user domain ontology linked with an EO domain ontology as presented in the previous section we can finally approach the task of improving a user query using the knowledge captured in the ontology.

A first straightforward way is simply to use the model as it is and show it to the user. With proper visualization software, an end-user, expert or non-expert can navigate through it and select terms to add to the query. This is an enhanced version of using a list of pre-defined keywords. The KES-B project is developing such visualizations and prototyping this strategy (see also [16]). However, visualization and selection of terms though simple and useful, with the added benefit of potentially building shared conceptualizations over time, cannot help a user without prior knowledge of EO domain. This is the case, unless the user discovers a path to a relevant EO product. Therefore, given an input user query, that will (partially) correspond to concepts in the ontology; a search algorithm must be applied for relevant EO products to be discovered. The search process can be split in two main parts:

- Semantic Matching

  The initial query string must be matched with the ontology contents. After this step, we will have a set of concepts, terms and relations selected as a starting point. Further, we shall have a set of nodes activated in the ontology which bear some connection to the user query.

  In summary this step involves three tasks: simple string preprocessing; detection of compound noun expressions; search in synonymy relations to detect if the user has used a synonym of a known concept in the ontology. The two last tasks are supported by WordNet, a lexical database of English with semantic relations between words [17].

- Domain ontology search

  With a tentative set of concepts selected, the next step is a classical definition of a search algorithm task. The set of activated nodes in the ontology semantic directed graph constitute an initial state. The relations between concepts allow the transition from one state to the next. The goal is to find instances of the Product class, an equivalent to an end-state. A generic graph-based search algorithm, such as A* [13], can be applied to derive the relevant EO products.

  Analysis of user requirements has also revealed that finding related concepts, even if they are not directly connected with instances of products, is interesting. In this context, our search algorithm also suggests as output, related concepts found when traversing the semantic graph.

3.5. User adaptation strategies

The KES-B project provides a complex scenario of several groups of cooperative researchers, each one belonging to a core domain, but possibly assigned to multiple domains of interest. Users can be experts in a given domain, in principle capable of participating in a domain ontology definition. However, for the most part, they have some knowledge of their domains and very little of the EO one. The behaviour of a particular user or of a group of users is important information since not all concepts and products are equally important. The behaviour of a user is captured as a minimum in his query history, or in a more advanced way in the overall activity when interacting with the system.

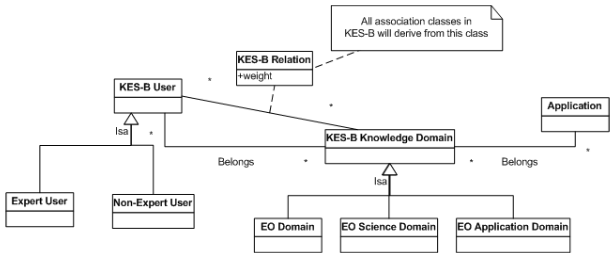

Fig. 6. Any relation is implicitly also associating a user and a domain pair.

The goal of user adaptation is to use the user interaction historical data to modulate the search process with the expectation of, after a large enough period of time, improve query results: a better order of the output result set; a new user might benefit from the accumulated interaction of a group of experts with the system as she is guided first to the most used concepts and queried products.

The directed graph structure of the ontology has an obvious choice a weighted approach. The weights in a graph act as natural search heuristic and can also be used to sort the final result set. The model presented here already associates every relation with a user/domain pair (**Error! Reference source not found.**). This association is given a weight and by updating the weights using some learning algorithm and historical data it will adapt the search over time. Some weights might be initially set to different magnitudes, since some semantic relations are already known to be of higher importance than others (e.g. connecting a feature of interest in a domain with a product that contains it); nevertheless in most situations the weights have to be adjusted using user data.

Currently we are investigating the use of Hebbian learning to change the weights [9]. In this learning rule, activation of a relation (an arc in the directed graph) will increase its strength. The nodes also need a weight (consider weighting some products more than others) than can be subjected to an update rule. When not activated, the weight has a simple decay process.

Finally, interesting uses of using a weighted directed graph and adapt it to user queries also under research include: conflating several users to get an average of a domain; use the information about domain of other users to assign a user to a domain

4. CONCLUSION

In this paper we have proposed an ontology-aided approach to EO product search. The modelling of a specific user domain ontology enables us to go from user terminology into EO terminology and search for relevant products even in the case where the user has no EO expertise. We have presented a user domain ontology that covers water quality and maritime security case studies, linked with an EO domain ontology.

This work is being developed in the context of the ESA KES-B project. At publication time, the model presented here is under review and validation from domain experts.

5. ACKNOWLEDGEMENTS

We wish to thank our project partners of GTD, STARLAB and UTV, in particular Olivier Germain and Luca Salvatori for their useful domain-experts contributions.

We also wish to thank the fruitful collaboration of Pedro Santana, Nuno Henriques and Ricardo Raminhos.

6. REFERENCES

1. Berson, A., Smith, S.J. (1997), *Data warehousing, data mining & OLAP*. Series on data warehousing and data management. McGraw-Hill.

2. Chen, Z., (2000) *Computational Intelligence for Decision Support*, International Series on Computational Intelligence, CRC Press.

3. Cios, K.J., Pedrycz, W., Swiniarski, R. (1998), *Data Mining Methods for Knowledge Discovery*. Kluwer.

4. OpenCyc Advances Starting Line for Intelligent Software. (2003), Cycorp company http://www.cyc.com

5. IEEE (2001). The IEEE Standard Upper Ontology Working Group (P1600.1), http://suo.ieee.org

6. Gruber, T. R. (1993), A Translation Approach to Portable Ontology Specifications, **Academic Press**. June http://www-ksl.stanford.edu/kst/what-is-an-ontology.html

7. Guarino, N., (1998), Formal Ontology and Information Systems, Proceedings of FOIS'98, Trento, Italy, 6-8 June, IOS Press pp. 3-15

8. McGuinness, D. (2002), Ontologies Come of Age, In: Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. Editors: Dieter Fensel, J im Hendler, Henry Lieberman, and Wolfgang Wahlster.. **MIT Press**.

9. Mitchell Tom M., (1997), *Machine Learning*. MacGraw-Hill.

10. Navigli, R., Velardi, P., Gangemi, A., (2003), Ontology Learning and Its Application to Automated Terminology Translation, *IEEE Intelligent Systems*, January/February, pp 22-31

11. Protégé (2004), http://protege.stanford.edu/index.html

12. Schreiber, G., Akkermans, H., Anjewierden, A., Hoog, R., Shadbolt, N., Van de Velde, W., Wielinga, B. (2000), *Knowledge Engineering and Management: The CommonKADS Methodology*. Bradford Books, **MIT Press**.

13. Nilsson, N. J. (1998). *Artificial intelligence: A New Synthesis*, Morgan Kaufmann.

14. Sowa, J. F. (2001), Building, Sharing, and Merging Ontologies, http://www.jfsowa.com/ontology/ontoshar.htm#s1

15. Studer, R., Benjamins, V. R., Fensel, D., (1998) Knowledge Engineering: Principles and methods. Data and Knowledge Engineering (DKE).

16. Varas, J., Busto, J., Torguet, R., (2004), EO Domain Specific Knowledge Enabled Services (KES-B), in ESA-EUSC Conference on Theory and Applications of Knowledge driven Image Information Mining, with focus on Earth Observation, Madrid, Spain, 17-18 March.

17. WordNet - An Electronic Lexical Database (1998), Edited by Christiane Fellbaum, Bradford books. http://www.cogsci.princeton.edu/~wn/