

EnArgus: A Knowledge-Based Search Application for Energy Research Projects

Frederike Ohrem, Lukas Sikorski, Bastian Haarmann

Abstract—Often the users of a semantic search application are facing the problem that they do not find appropriate terms for their search. This holds especially if the data to be searched is from a technical field in which the user does not have expertise. In order to support the user finding the results he seeks, we developed a domain-specific ontology and implemented it into a search application. The ontology serves as a knowledge base, suggesting technical terms to the user which he can add to his query. In this paper, we present the search application and the underlying ontology as well as the project EnArgus in which the application was developed.

Keywords—Information system, knowledge representation, ontology, semantic search.

I. INTRODUCTION

RESEARCH on ontology is becoming increasingly important in different areas. E.g., information systems can use ontologies as an instrument to make knowledge accessible to the user. Especially in semantic search applications, ontologies can support the user in his manual search by suggesting reasonable additional keywords. This leads to an improvement of the search result. Simultaneously, the duration of the search process decreases for the user.

For example, an employee of a federal ministry searches all information about research projects on batteries. He starts a query in the database and the system only shows results for this keyword. Other terms like “accumulator” or “cells” will not be considered. If the employee is not an expert in the battery domain, the search is time-consuming or incomplete. In order to find a solution, the German Federal Ministry of Economics and Technology (BMWi) promotes a project called “EnArgus” including the use of an ontology for search result improvement. The objective of this project is an information system for the administration of energy research projects and a lookup platform to help non-experts finding detailed information about those energy research projects using an ontology-supported search algorithm.

To make knowledge accessible for employees of the ministry and other non-expert users, an ontology has been build and certain Java methods and classes which help to choose the appropriate search term have been implemented. The methods and classes look up information from the ontology and help making keyword suggestions to the user

which he can integrate in his search query.

A domain-specific ontology containing knowledge on the energy domain is hard to build for a computational linguist. Therefore, we asked fellow scientists carrying out energy research to fill the ontology with their knowledge. The next challenge was to formulate the knowledge into the ontology. We agreed that the domain experts provide the knowledge in the form of a collaborative text collection. These texts can be used to develop an ontology using automated information extraction. The resulting ontology from the texts then supports non-professional users seeking domain specific information by suggesting alternative and related terms to the keywords they entered.

In our previous paper, we presented a method to extract and formalize acquired knowledge structures from natural language texts and automatically generate an ontology out of the extracted information [1].

In the following, we will first introduce the EnArgus project, the wiki texts as well as the ontology. The information system itself is explained afterwards.

II. ENARGUS

The EnArgus® project is funded by the German Federal Ministry of Economics and Technology. It is an energy research project featuring a committee of 7 institutes from the energy research and computer science domain [2]. Energy experts from Bochum University, the Fraunhofer Institute for Systems and Innovation Research (ISI), the Fraunhofer Institute for Environmental, Safety and Energy Technology (UMSICHT) and the Helmholtz Association’s Institute of Energy and Climate Research (IEK) are supposed to write Wikipedia-like texts. Computational linguists from the Fraunhofer Institute for Communication, Information Processing and Ergonomics (FKIE) create an ontology out of this textual information. The text collection also serves as a glossary for the non-expert users of the EnArgus system which will be built by our fellow Fraunhofer Institute for Applied Information Technology (FIT) and the partner software company OrbiTeam Software.

III. WIKITEXTS

As a part of the EnArgus project, German energy experts create a text collection focusing on four energy subdomains. The major definitions and processes of these areas are explained and stored in a wiki (see Fig. 1). The wiki texts will be accessible for non-expert users, such as Parliament members or media representatives who may want to be

Frederike Ohrem, Lukas Sikorski, and Bastian Haarmann are with Fraunhofer Institute for Communication, Information Processing & Ergonomics, Fraunhofer Str. 20, 53343 Wachtberg, GERMANY (phone: +49 228 9435 722; fax: +49 228 9435; e-mail: frederike.ohrem | lukas.sikorski | bastian.haarmann@fkie.fraunhofer.de).

informed about specific research project details as well. The experts included links between the wiki texts helping the users to quickly navigate through the explanation of the terms and other terms related to them.

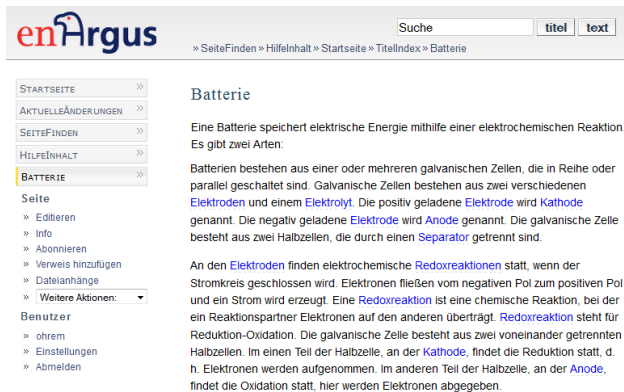


Fig. 1 An extract from a wiki text

IV. ONTOLOGY

According to Gruber [3], an ontology is an “explicit specification of a conceptualization” where conceptualization is defined as “an abstract, simplified view of the world that we wish to represent for some purpose”. Common knowledge often is taken granted by humans but nevertheless has to be represented so that a system can use it. In order to stress this point, Borst [4] modified the definition of Gruber, by adding the modifier “shared” to “conceptualization” and by substituting “explicit” by “formal” so that Borst proposes the definition: “An ontology is a formal specification of a shared conceptualization.”

In contrast to philosophy where ontology is the science of being as such [5], the definitions by Gruber and Borst refer to “ontology” as used in information science. Here, ontologies usually consist of a defined basic vocabulary structured as a taxonomy. This means that the terms refer to concepts and are put into a tree structure for which the structuring relation is the hyponymy relation “ISA” (= is a). This relation is chosen because it allows inheritance. From the implementation point of view, the concepts are classes. For example, there might be the class “battery” in the ontology as a subclass of “energy storage” (see Fig. 2). These classes might have instances denoting specific objects of the real world. Often, the instances are stored in a data base. In addition to the taxonomic structure, the classes are augmented with attributes as well as relations that hold among them. For example, the class “energy storage” might have attributes like “performance” or “energy density” which are inherited to its subclass “battery”. An instance of the class “battery” then has a value for the attribute “performance” which has to be a performance denotation, e.g. “10 watts”. Another example might be the relation “owner” by which for example instances of “energy storage” can be linked to instances of “organization” or “person”.

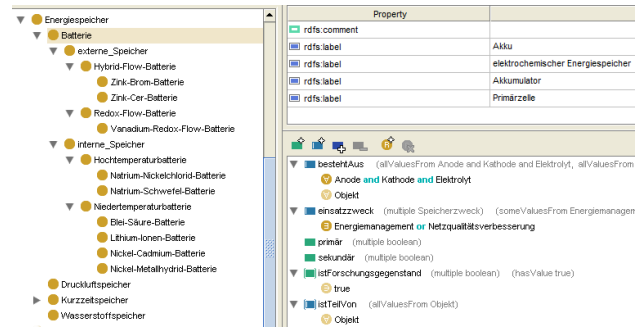


Fig. 2 A partial extension of the ontology

In the next section, we will describe how the basic EnArgus information system interacts with the ontology. We will also present the classes and methods we have implemented to allow querying and manipulating the concepts represented in the ontology and thus in particular the ontology enrichment of search terms by semantically related terms.

V. METHODOLOGY

One objective of the information system is to support the user in finding projects. If he enters a term, the system will generate keyword suggestions of related terms from the ontology. Related terms may be synonyms, superclasses, subclasses and special related classes i.e. classes connected by a part-of-relation. For these functions to be used, a communication between the basic EnArgus information system and the ontology system is required. The EnArgus information system is based on Python. There are no ontology libraries available for this environment. Therefore, the ontology system was implemented based on a Protégé-OWL API in Java [6]. In order to load and use the Java library with Python, a wrapper called JPytype was used (see Fig. 3). JPytype use memory-based transmission and transfers data between Python and Java with a hardly measurable runtime delay. An ontology service process is used to avoid reloading the ontology for each request. The service starts together with the EnArgus system. The EnArgus system contacts the ontology service via remote procedure calls (RPC).

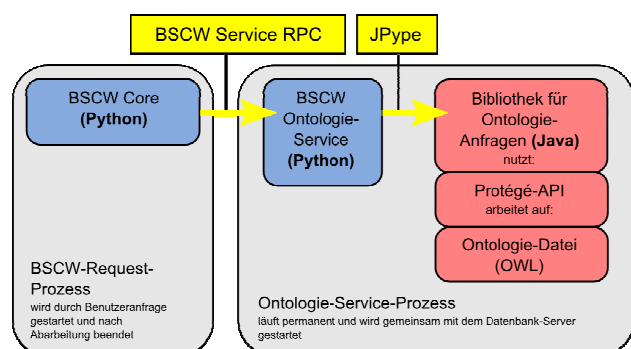


Fig. 3 The wrapper JPytype

The ontology service provides a lot of getter methods. These methods can be used by the system to obtain suggestions for the seeker. We implemented several getter methods that we will show below.

The method “public String[] getClass(String input);” returns the ontology class for a given individual or label. For example getClass(“accumulator”) returns “battery”. The method “public String[] getClasses();” returns all classes mentioned in the ontology. The method “public String[] getEquivalentClasses(String input);” returns ontology classes of the same level. For example for “battery” we get “compressed air reservoir”, “short-term memory” and “hydrogen storage”. The method “public String[] getLabels(String input);” returns labels for a given ontology resource. For example for “battery” we get the labels “accumulator”, “primary cell” and “electrochemical energy storage”. The method “public String[] getLabels(String input);” returns all labels mentioned in the ontology. The method “public String[] getInstances(String input, boolean includingSubclasses, int count);” returns instances/individuals for a given class or label. The parameter *includingSubclasses* can be used to include labels from the subclasses and *count* can be used to specify the number of results. The method “public String[] getRestrictions(String input, String property, boolean includingSuperclassRestrictions);” returns restrictions for a given class and its property. For example, for the class “wind turbine” and the property “consistsOf” we get “rotor, hub, brake, generator, tower, foundation and gear”. The parameter *includingSuperclassRestrictions* can be used to include superclass restrictions. The method “public String[] getHyponyms(String input, int level);” returns hyponyms and the method “public String[] getHypernyms(String input, int level);” returns hypernyms for a given class or label. The parameter *level* can be used to specify the search depth. We also implemented a method for spell checking. This method is called “public String[] getRecommendations(String word, int maxDistance);” and returns all the words of the ontology of a predetermined maximum distance (*maxDistance*).

VI. IMPLEMENTATION

As described above, EnArgus consists of different components like wiki texts, an ontology, a methodology and a query. The following figure (see Fig. 4) shows the architecture of the system.

All components are assembled on a webpage. The database to be searched contains partly sensitive data, so the information system is divided into two parts: EnArgus.public for the public and EnArgus.master only for registered users [7]. Both have priority in the search query. Therefore, the search can be narrowed down to a time period, a federal state, the government-funded sum or the customer.

Below, the two different web pages are depicted.

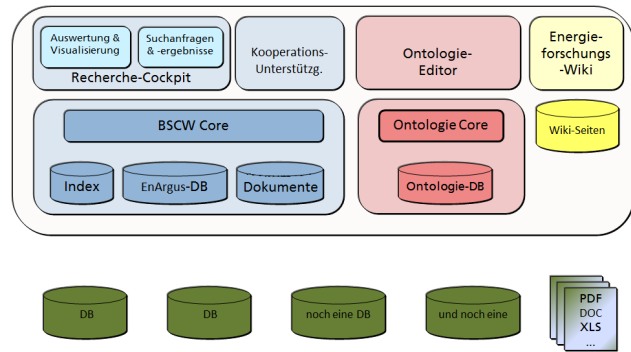


Fig. 4 System architecture

A. EnArgus.public

EnArgus.public is open to the public such as journalists, delegates or energy-interested users. The webpage is without registration. Therefore, users obtain a constrained functionality. EnArgus.public provides a simple search form to enter a term. Users can decide between search with (see Fig. 5) or without (see Fig. 6) “synonyms”. The term “synonyms” includes subclasses, superclasses and labels of the ontology. As depicted in both figures, the search with synonyms gets more research projects (488 results) than the search without synonyms (430 results).

In addition to search function, users can read the wiki texts to inform more about energy research or to define their search more precisely.



Fig. 5 EnArgus.public with using the synonyms



Fig. 6 EnArgus.public without using the synonyms

B. EnArgus.master

EnArgus. master provides an expert search for a limited audience, among them the lead project partners, ministers and specific energy experts. They all need a registration because the system contains sensitive data from different databases.

Therefore, users can start an expert search with many functions to create a complex query. As contrasted with EnArgus.public, users see related terms like synonyms, subclasses and subclasses and an additional function is to involve certain terms (= "synonyms") in search query. Each term contains a question mark-symbol to redirect to the respective wiki text directly (see Fig. 7). In addition, users can analyze values statistically and display graphs.

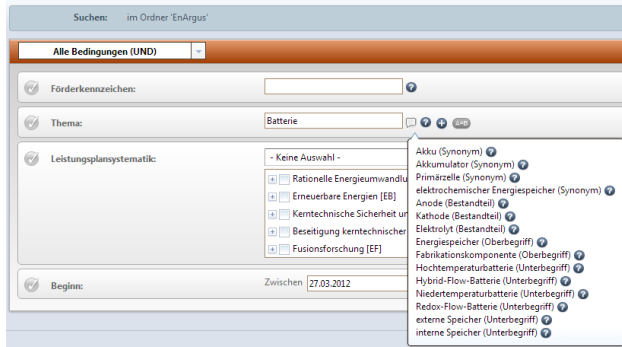


Fig. 7 Search query at EnArgus.master

VII. EVALUATION

The presented information system was tested in a workshop by several energy-experts. They familiarized with the homepage, checked out the search mask and searched different terms. The experts tested both EnArgus.public and EnArgus.master. At the end of the workshop, experts filled in a questionnaire.

The consensus of the experts indicates an improved search in databases. A comparison of results between search with or without ontology shows more research projects using the ontology-based search (see Figs. 5 and 6). In particular, the search for an employee of the ministry in the database for research projects on "battery" also shows research projects with the terms "accumulator" and "cell". It shows that research projects with related terms were also found. This leads to an improved and time-saving search result, even if the user is not an expert.

Currently, the EnArgus project includes four focus topics in the ontology. Due to the good search results, the energy topics are likely to be expanded. This should be done in a follow-up project called "EnArgus2.0" in the near future.

VIII. CONCLUSION

EnArgus is an energy research project to improve searching in databases of ministry. Energy experts and computer scientist works closely together to develop an information system. Energy experts wrote wiki texts with definitions and processes in energy research. Computational linguists built an ontology and provided different methodologies to look up information in the energy-ontology. Information scientists created two webpages with a search box application. All works were combined into an information system for energy-interested

users. In practice, users enter a term, such as "battery" in the search form. In the background, the search starts in database with information of the ontology and synonyms like "accumulator" or related classes like "energy storage" are taken into account, too.

We assume that any business or project that relies on storing information in databases can benefit from an ontology-based search system like the information system of EnArgus. We also hope that our work and research can contribute to the forth going process of search result optimization.

REFERENCES

- [1] Bastian Haarmann & Lukas Sikorski & Frederike Gottsmann. "Knowledge for Information Systems Acquired from a Co-Created Text Collection". Proceedings of the World Academy of Science, Engineering & Technology. Issue 66. Art. 193. pp. 1092-1095. Paris, FR. 2012. ISSN: 2010-376X.
- [2] EnArgus: Central information system for energy research funding. <http://www.enargus.de/>
- [3] Thomas R. Gruber. "A Translation Approach to Portable Ontology Specifications". In: *Applied Artificial Intelligence*, 5(2):199-220, 1993.
- [4] Willem Borst "Construction of Engineering Ontologies for Knowledge Sharing and Reuse". Ph.D. Dissertation, University of Twente, 1997.
- [5] Uwe Meixner. „Einführung in die Ontologie“. Wissenschaftl. Buchges. Darmstadt 2004.
- [6] Protégé-OWL API: www.protege.stanford.edu/plugins/owl/api/
- [7] EnArgus: www.enargus.de/pubilc; www.enargus.de/master