# EnArgus – Ontology Based Search

Michael Dembach
Fraunhofer Institute for Communication, Information Processing and Ergonomics FKIE
Fraunhoferstrasse 20
53343 Wachtberg, Germany
michael.dembach@fkie.fraunhofer.de

Lukas Sikorski
Fraunhofer Institute for Communication, Information Processing and Ergonomics FKIE
Fraunhoferstrasse 20
53343 Wachtberg, Germany
lukas.sikorski@fkie.fraunhofer.de

Rudolf Ruland
Fraunhofer Institute for Applied Information Technology FIT
Schloss Birlinghoven
53754 Sankt Augustin, Germany
rudolf.ruland@fit.fraunhofer.de

## ABSTRACT

This paper presents the EnArgus Project – a project which aims to make energy research funding more transparent and includes the use and development of an ontology. The structure of the paper is as follows: first, we will present the EnArgus project, its domain, and its main goals. Next, we will describe the domain-specific ontology which we developed for EnArgus. This will lead to a description of how that ontology has been constructed and evaluated. Finally, we will discuss describe how the whole system has been evaluated.

## Keywords

Ontology Research, Ontology Visualization, Semantic Search, Energy Research

## 1. THE ENARGUS PROJECT

The EnArgus Project is sponsored by the Federal Ministry of Economic Affairs and Energy (BMWi) by decision of the German Bundestag.[1] The project aims to make federal subsidies policy in the field of energy research more transparent by facilitating the assessment of technology development. EnArgus's approach includes the close collaboration of experts from energy research and information science.[2]

EnArgus is building a central information system for energy research projects funded by the Federal and State. Via this, system professionals – as well as interested members of the public – can receive consistent and central access to information about energy research in the Federal Republic of Germany. For the general public, the system will be available as "EnArgus Public". Professionals, who receive special authorization, may access "EnArgus Master". These two versions exist due to data security reasons: The system uses official databases for research funding which contain protected data and are therefore only accessible with special privileges.

The system works on a database of the BMWi which contains all energy projects funded by the Federal Republic of Germany. To use the system, the user may enter a single keyword. For example, in order to answer the question "Which of Germany's states fund how many projects for the development of wind power plants?", the keyword "Windkraftanlage" (wind power plant) should be entered. Just by the use of that keyword about 136 projects will be found. This number might be considered too low [1]. In that case, the system provides ways of increasing the search result, e.g., the use of synonyms which are already in the system. With the use of synonyms, the query for "Windkraftanlage" will deliver 1682 projects, mostly because the more common term "Windkraftanlage" will be used [2]. To handle this huge number of projects, the distribution may be shown in different clusters depending on parameters like "start of project", "amount of money granted" and, as was the original intent, "state".

## 2. THE DOMAIN-SPECIFIC ONTOLOGY

The backbone of any search process in EnArgus is its domain-specific ontology about energy research. In this ontology the knowledge from the fields of energy, energy research and energy research funding is formally stored, so it can be used by the whole system. When using EnArgus Public, the user is able to incorporate alternative terms (synonyms) suggested by the system into his query. EnArgus Master will additionally suggest terms which are semantically related to the keyword (hyponyms, hypernyms and terms resulting out of certain relations, e.g., meronyms). The user can choose which of those terms are to be

[1] *EnArgus* has been sponsored in two phases by the Federal Ministry for Economic Affairs and Energy at the tag 03ET1064 (Phase 1: July 2011 to June 2013) and at the tag 03ET4010 (Phase 2: July 2013 to June 2016 as EnArgus 2.0).

[2] The project partners in the first phase were Fraunhofer Institute for Applied Information Technology (FIT), Fraunhofer Institute for Communication, Information Processing and Ergonomics (FKIE), Fraunhofer Institute for Systems and Innovation Research (ISI), Fraunhofer Institute for Environmental, Safety and Energy Technology (UMSICHT), Ruhr University Bochum's Chair for Energy Systems and Energy Economy (LEE), Forschungszentrum Jülich's Institute of Energy and Climate Research, section Technology Development (IEK-STE), and OrbiTeam Software GmbH (Bonn). IEK-STE had dropped out after phase one. In the current second phase the Institute for Water Supply, Wastewater Technology, Material Flow Management and Resource Economy and Spatial and Infrastructure Planning (IWAR) of the TU Darmstadt, the Materials Testing Institute (MPA) of the University of Stuttgart, the *Zentrum für Beratungssysteme in der Technik* (ZEDO) and bense.com joined as new project partners. Fraunhofer FIT is responsible for the coordination and Project Management Jülich supervises the project in the name of the Federal Ministry for Economic Affairs and Energy.

included in the query. The connection of those terms to the initial keyword is called **semantic relation** [4][5].

Especially useful are the taxonomic relations (hyponym and hypernym) and alternative terms (synonyms). With the help of synonyms it is possible to find a relevant element even if it uses a different term than the keyword. The taxonomic relations, which are always given in an ontology, are also useful because a project which refers to B also refers to A when B is a hyponym to A. For example, when searching for projects which refer to "wind power plants", it is also useful to look for "lift-based wind turbines", which are a special kind of "wind power plants". Another kind of alternative terms are translations into different languages (especially from German to English) which are stored just like synonyms.

The next important semantic relation is the meronymy (part-of relation) which appears in certain parts of the ontology. Meronymy is used in classes of the concrete objects and in classes of the processes. With respect to the semantic search, meronymy fulfills a function similar to hyponymy. A typical example is a project in which the characteristics of specific membranes are examined and these membranes are parts of batteries. This project belongs to projects for the improvement of batteries, even if the project title is about the membranes and the concept "battery" isn't found in the project description.

As a rule, further semantic relations are only defined between specific classes and stored in the ontology. One example is "use" (actually "use_as_energy_source"). This relation exists between power stations and energy sources. So the energy source "sun" (also represented in the ontology as a class) is assigned to the classes of the solar power plants. By this representation the energy source used, in this case "sun", is transmitted to all subclasses and individuals of "solar power plant". Thus the concept "sun" can be offered to the user whenever the name of a subclass of solar power plant or the name of an appropriate individual is entered.

# 3. CONSTRUCTION AND EVALUATION OF THE ENARGUS ONTOLOGY

The construction of the domain-specific ontology is the part where the collaboration between experts from information science and energy research gets important. There are several different ways of developing an ontology [3]. Our basic idea is that the experts for energy research hold the relevant knowledge while the experts for information science know how this knowledge can be integrated into the system in a helpful way.

Both parties work together in the following way: the experts for energy research write short articles similar to those in Wikipedia. These articles serve as the source of knowledge for the experts from information science. With regard to syntax and vocabulary, these texts make use of simple structures and words of common understanding, because they are also stored in the EnArgus system and will later serve as a source of further information non experts. For the information scientists, it is important that those texts include the semantic relations mentioned in section 2 because they create the ontology out of the texts. Additionally, the energy experts draw mind maps containing those concepts they consider crucial for the ontology.

In the next step, the energy experts evaluate the ontology to make sure that the knowledge has been integrated correctly. For that step the ontology is visualized as hyperbolic tree (hyper tree). The term in question is at the center of the visualization and the related terms are emerging from this root like branches. In the case of a wider evaluation, the root can be changed at any point by navigation through the branches. This changes the focus and provides new branches.

The hyper trees allow flexible search depth: one can choose depth "1" if one only wants to see direct relations, or a depth with higher value to see and evaluate relations which are more indirect. In other words, it is possible to determine the semantic radius for a term such that the user may adjust the visualization according to her/his own demands and preferences. This is a significant advantage when evaluating the ontology. The visualization is also integrated into the system and will be open to the public.

# 4. EVALUATION OF THE ENARGUS SYSTEM

At the end of the first project phase, the EnArgus system was evaluated in two expert workshops. During these workshops, external energy experts were asked to test various sorts of queries using the system. Some search problems were pre-defined, and the experts tackled these problems on the one hand with a standard search, and on the other hand with a search supported by the ontology. As expected, it turned out that the searches supported by the ontology found significantly more relevant projects (see [2] for details on these results). In addition, the experts did searches on self-defined problems. In order to get valuable hints and ideas for improvement, the experts were asked to express their suggestions and requirements. Finally, they were asked to answer usability questions on a questionnaire.

Most of the experts' criticism was aimed at the wiki-texts. Some criticized that the wiki-texts were more suitable for laymen than for experts; however, such criticism was unwarranted as the texts were intended for a general audience. Another point of criticism concerned the comparison between the use of the public version and the use of the master version. While the public version was rated as very understandable and intuitive to use, numerous suggestions for improvement were presented for the use of the master version. Most of these suggestions have now been implemented. The general principle of the EnArgus system was rated as worthwhile and sensible. The system was recommended for completion (in order to cover more fields of energy research) and rollout.

As a further indicator of EnArgus' quality, the cover rate was evaluated, i.e., how many current projects assigned formally to the area of energy research in the database of the BMWi were found by queries with the concepts supported by the ontology. This rate was 86%, which may be rated as a positive result, since energy research is a very broad and diverse field, and since, at the time of the above-mentioned evaluation (at the end of the first project phase), not all areas of energy research were represented in detail in the ontology. Again, the evaluation led to additions and improvements, while more notations and abbreviations were added to the concepts. It was decided to carry out cover analyses regularly to ensure the quality of the ontology.

The work on EnArgus proves that the fields of energy and energy research are extraordinarily wide in scope and complexity, and often require very specific knowledge. Therefore, it will always be necessary for the system to be re-evaluated and adapted to the latest developments. At the same time, developers and users must be aware that not every detail can be represented.

## 5. OUTLOOK

The EnArgus Project is still running and its ontology – currently holding about 2.400 classes – is expected to grow to over 3,000 by the end of 2016. We are confident that EnArgus is a valuable contribution to the field of energy research, since this area will grow and become even more important in the future[6], and that our approach of having experts from energy research and information science working closely together provides a high level of informational integrity.

## 6. LITERATURE


[1] Hoppe, T. (2014). Modellierung des Sprachraums von Unternehmen – Was man nicht beschreiben kann, das kann man auch nicht finden. In: Humm, B., Reibold, A. & Ege, B. (Hrsg.), Corporate Semantic Web. Berlin: Springer.

[2] Schade, U., Bense, B., Dembach, M., Sikorski, L. 2015: Semantische Suche im Bereich der Energieforschungsförderung. Nutzen, Entwicklung und Evaluation einer Fachontologie. In: *Corporate Semantic Web. Wie semantische Anwendungen in Unternehmen Nutzen stiften*, Hrsg. B. Humm, B. Ege, und A. Reibold. Berlin: Springer.

[3] Schäfermeier, R. 2015. Verteilte und agile Ontologieentwicklung. In: *Corporate Semantic Web. Wie semantische Anwendungen in Unternehmen Nutzen stiften*, Hrsg. B. Humm, B. Ege, und A. Reibold. Berlin: Springer.

[4] Staab, S. & Studer, R. (Eds.) (2004). Handbook on Ontologies in Information Systems. Berlin: Springer.

[5] Studer, R., Benjamins, R. & Fensel, D. (1998). Knowledge Engineering: Principles and Methods. Data & Knowledge Engineering, 25, 161-198.

[6] Wietschel, M., Arens, M., Dötsch, C., Herkel, S., Krewitt, W., Markewitz, P., Möst, D., Scheufen, M. (2010). Energietechnologie 2050 – Schwerpunkte für Forschung und Entwicklung. ISI-Schriftenreihe Innovationspotenziale. Karlsruhe: Fraunhofer ISI.