

Ontology-based Data Access for Energy Technology Forecasting

Alexey V. Mikheev

*Melentiev Energy Systems Institute of Siberian Branch of the Russian Academy of Sciences,
Lermontov str., 130
Irkutsk, Russia
E-mail: avmiheev@gmail.com*

Abstract

Ontology-Based Data Access (OBDA) is considered as a promising semantic approach to query various complex datasets for such weak-formalized activity as energy technology forecasting. OBDA uses an ontology to operate with complex energy technology data abstracting away from the technical schema-level details. Special mapping is required to connect the related data to ontology entities. OBDA approach automatically translates queries posed over the ontology into data-level queries which can be executed by the underlying database management system. The paper is focused on the main principles of OBDA applied to Energy Technology Database within technology forecasting information system.

Keywords: Energy technology forecasting, semantic model, ontology, ontology-based data access.

1. Introduction

Energy technology forecasting is one of key tasks of any national innovation system. In general, the technology forecasting implies “all decisive and systematic attempts to anticipate and understand the potential direction, rate, characteristics, and effects of technological changes, especially invention, innovation, adoption, and use” [1]. Objects under consideration in energy technology forecasting are newest emerging technologies of energy production, conversion, transportation, distribution, consumption, control and management. New derived technological structures such as hydrogen energy, distributed energy systems, smart grids are studied too. The use of new energy carriers, for example, hydrogen or methanol is also considered. The main result is an assessment of technological development in energy to maximize overall economic (productive) efficiency in the future under a set of technical and external constraints. The benefit from technological forecasting is identification of the best emerging technologies in the future and estimation of their applicability conditions for a certain time horizon.

The energy technology forecasting procedure is a comprehensive study based on an analysis of a large amount of heterogeneous complex data related to efficiency indices of certain technologies. The general challenges in energy technology forecasting are:

- requirement of sustainable energy development support;
- multidisciplinary effects growth and, therefore, increasing data variety in energy and related technology areas;
- acceleration of technological changes in the energy field caused by smart technologies, load-controlled consumers, individualization, autonomy and mobility, renewable energy advances, etc.

There are many different important factors taken into account such as advances in technology research and development, demands from energy markets, macro economy forecast, cost forecasts, availability of energy resources, external conditions (environment, CO₂, institutional barriers, energy security and so on), advances in neighboring technological fields. The data includes statistics from officials, international and national energy agencies and others organizations; energy technology parameters reflecting effectiveness, scalability, robustness, environmental restrictions, required resources, etc.; text corpus from science and technology sources related to technological development; socio-economic impact assessment data and so on.

The search, processing, reconciliation, analysis and integration of these data of different types and formats are very hard tasks for any domain expert. These tasks can be performed by means of semantic web - driven

technologies. One of them is ontology modelling. Thus, the elaborate ontology of energy technology domain is a good candidate to facilitate data operations for energy technology forecasting via implementation of ontology-based data access (OBDA) technique.

2. Data in Energy Technology Forecasting

The energy technology forecasting includes three main general steps demonstrated in Figure 1. The first step is monitoring of science and technology information sources like research and technical publications, patents, reports, news, blogs, official statistics, development scenarios and forecasts, commercial offers and so on. Undoubtedly, all data relate to an information source with a certain confidence level. The step includes a sequence of procedures such as data extraction, cleaning and reconciliation, clustering and analysis. These procedures are necessary to provide new energy technology data for the next step of the forecasting.

The second step is a system analysis of energy technologies development with different mathematical models. This is a form of case study to obtain key development parameters of a certain technology or technology set under consideration. This step operates

with the data obtained from the previous step. The energy technology data are organized according to multicomponent hierarchy representing a tree of child technologies. The upper high-level energy technology usually includes bottom low-level technologies. For example, solar energy technology has two child technologies based on photovoltaic effect and concentrated solar power.

Each energy technology as well as its child component has a set of characteristics and parameters related to different technical, economy, ecology efficiency and other dependent factors. A certain energy technology is characterized by used energy resources, processes and equipment, materials, applications, energy products, services, etc. These characteristics of a certain energy technology can be represented through organizational chart describing general redundant scheme of the energy conversion flows. These characteristics has own parameters and include other sets of complex data too.

Each technology has technical parameters: net and partial efficiency, possible load ranges and operating conditions, consumption, unit capacity range, etc. A set of economic parameters of the energy technology contains capacity per unit costs, operating costs and

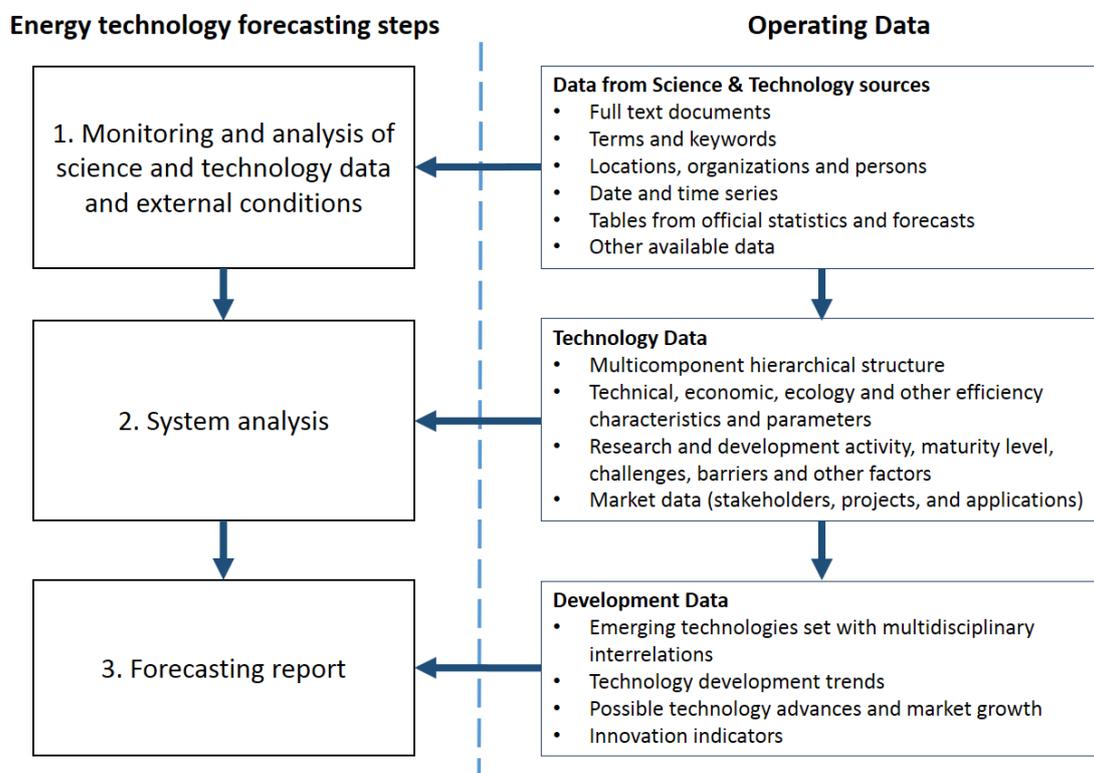


Fig. 1. Energy technology forecasting stages and data.

costs per unit of output. Ecological parameters include emissions per unit of output and emission composition. Another set of parameters represents research and development activity, maturity, current challenges and barriers hindering the progress of the technology. The last set of the data concerns market parameters such as a list of key stakeholders, implemented projects, application segments, investments, etc.

The energy technology data are often of complex types: tables, sets of ranged enumerations (for example, a maturity level), time series, interval values (for example, efficiency factors), and distributions over interval. The data also depend on one another and can be multivalued, incomplete, and probabilistic.

There are three important requirements of the energy technology data: (a) consistency, (b) ability to be aggregated bottom-up, (c) comparability.

The data of technology development in the third step are derivatives from previous steps. They include assessments of emerging technologies, development trends, possible technology advances and market growth. Forecasting results are based on innovation

indicators reflecting positive or negative changes in the efficiency of technology parameters over time. Examples of such indicators are R&D intensity increase, learning rate, spillovers, specific costs reduction, efficiency growth, etc. Such innovation indicators can have a multicomponent complex structure and often require special fuzzy-logic rules to obtain total indicator values.

All aforementioned data excluding full texts from science and technology information sources are stored in specially developed Energy Technology Database (ETDB). Now ETDB is a normalized relational database operated by PostgreSQL engine.

3. OBDA for Energy Technology Database

A semantic-based approach is proposed to organize and formalize (where possible) on a unified basis the steps of energy technology forecasting [2]. This approach is similar to the one applied to the energy security assessment [3]. The main useful features of this approach are:

- supporting and automating the monitoring and

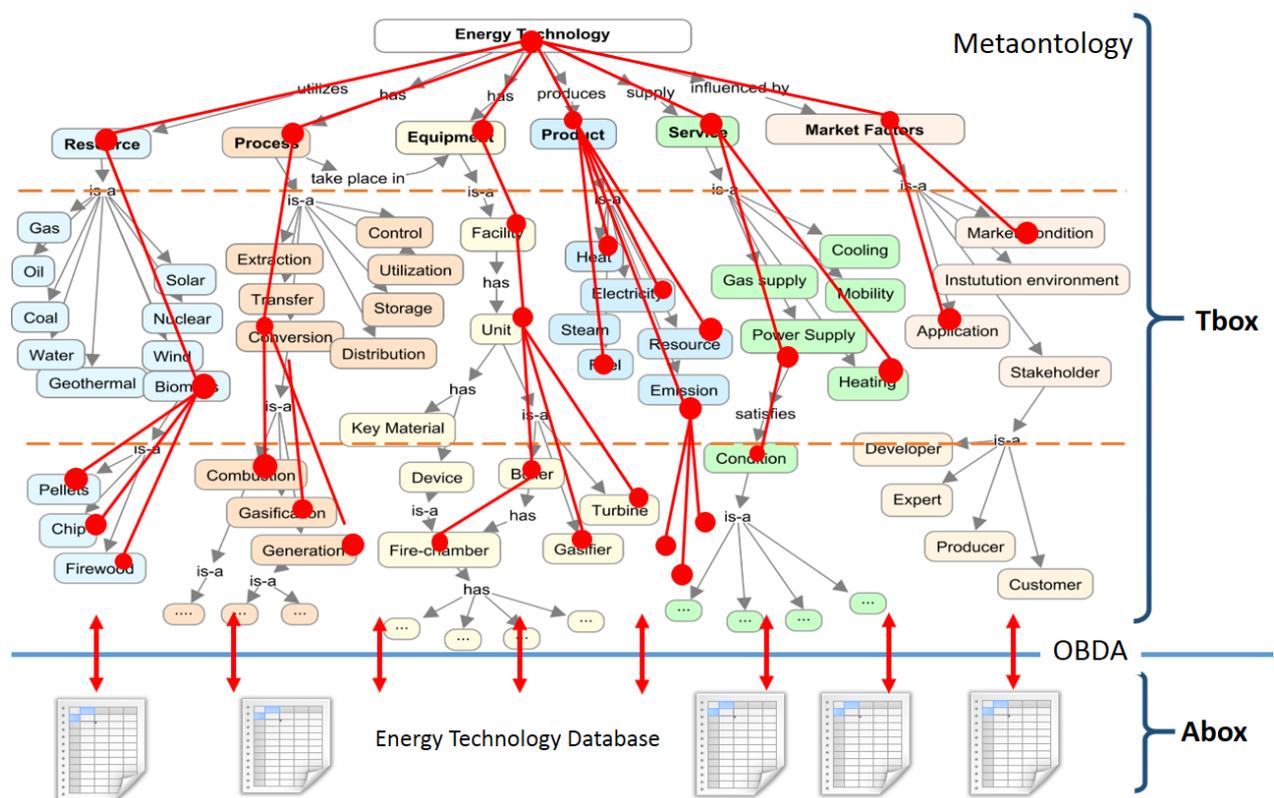


Fig. 2. Multilevel representation of generalized fragment of energy technology domain ontology. Red circles and lines show concepts and relations for the energy technology “direct biomass combustion”.

analysis of large data sets from science and technology information sources related to energy;

- organizing energy technology datasets for system analysis step;
- supporting structure and calculation rules of innovation indicators;
- visualization ability of research landscape reflecting current and forecasted state of research and development, i.e. putting information into comfortable, easily understood form.

A detailed ontology describing energy technology domain and forecasting process is being developed now. The energy technology ontology is a core component to build an information system supporting the three steps of energy technology forecasting. The domain ontology has several levels, and this model was used to develop and manage ETDB. The ontology includes concepts tree to describe full hierarchy of energy technologies. Each concept of the ontology has a set of attributes corresponding to energy technology characteristics and parameters. Two additional data models are added to the ontology. One data model describes original information documents about energy technologies and its characteristics. Another data model contains a description of innovation indicators structure with aggregation and calculation rules. The ontology is also saved in RDF triple store as part of ETDB. A

generalized fragment of the ontology model of energy technology domain is presented in Figure 2. The ultimate goal of the domain ontology development is abstracting data structure away from specific relational database scheme to make it understandable for end-user researcher. Moreover, since energy technology forecasting is a weak-formalized problem there is additional requirement for the ontology to provide user-oriented data management over relational database – ETDB.

Ontology-based data access (OBDA) is a new paradigm of organizing access using description logic to various types of data. OBDA is considered as key component in the new generation of information systems including Web applications [4]. This technique implies an application of additional declarative specifications for describing a set of maps that specify the translation of a common set of ontology structures into a relational database. Principles of the designing effective information systems to link a data to the ontology are discussed in detail in [5].

OBDA has several advantages. First of them is that end-user has a single point of semantic data access. Secondly, OBDA provides queries running over any data sources in terms of a user-oriented conceptual model. In addition, OBDA expresses information needs in their own terms without any prior knowledge about

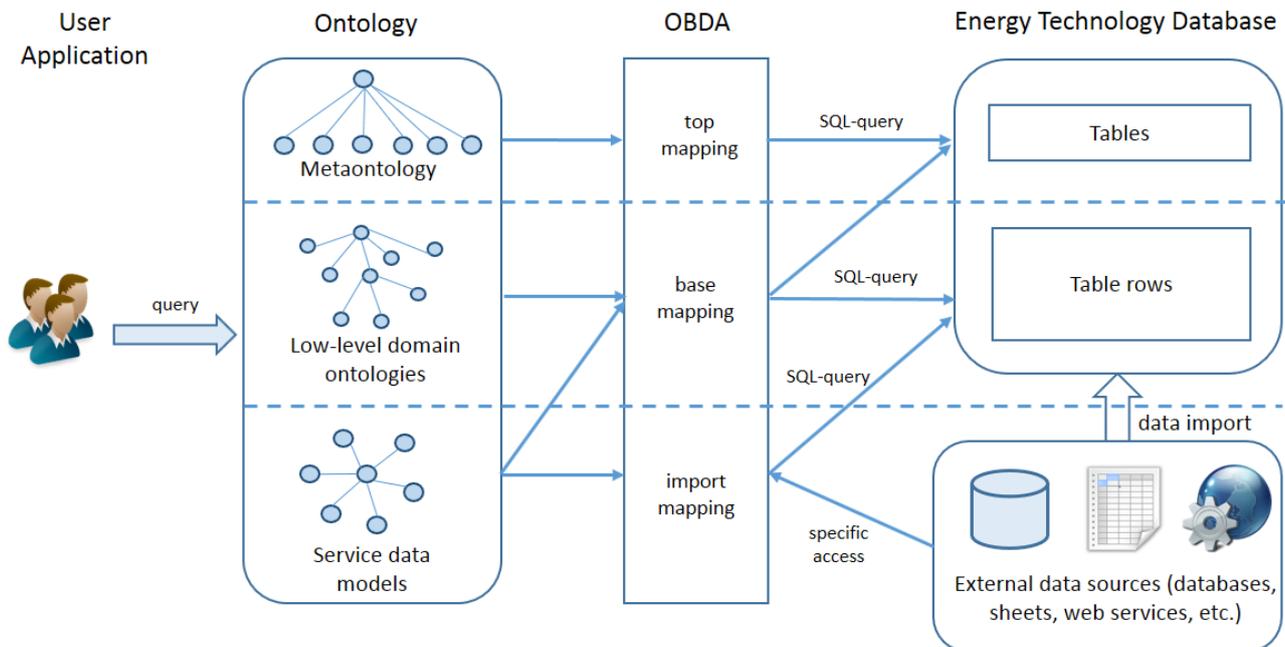


Fig. 3. General view of ontology-driven energy technology data management.

the way the data is organized. There are several examples of successful OBDA engineering in different fields. In [6] OBDA approach is applied to integration of real-world data sources. Industrial use of OBDA for petroleum company Statoil based on universal Optique platform is presented in [7]. Further, novel techniques of uniform querying via OBDA of both streaming and static data in Siemens are presented in [8].

Application of OBDA mapping to integrate energy technology ontology and database satisfies the requirements:

1. Domain ontology reflects the structure of energy technology database and part of data. This means that all available data can be retrieved from the database via query over domain ontology.

2. Data from external sources such as official statistics, web services, or spreadsheets are obtained via special service data model incorporated into the parent ontology. In this case, a special data provider is required to import the external data into ETDB or to integrate the data into the query result.

Figure 3 presents a general view of the ontology-driven data management with a three-level mapping scheme. Concepts from top level of ontology or metaontology from Fig. 2 are directly reflected to respective tables of ETDB. Connections between low levels of domain ontology and database are realized by means of standard OBDA specifications using the additional data tables: technology characteristics and its values, technology parameters and its values, innovation indicators, parameter types, data types, ranged enumerations, time series and so on. The third mapping level includes special service data models to manage data import or data receiving from external sources. Such a three-level approach reduces complexity of OBDA mapping design.

OBDA mediating interface between ontology and ETDB is now under development and implementation.

4. Conclusions

Ontology-Based Data Access (OBDA) is considered as a feasible technique to build a unified query formulation procedure with natural semantic and good flexibility over complex energy technology data. In this research, a three-level OBDA approach is proposed to reduce the complexity of mapping between ontology and database.

Acknowledgements

The research is funded by the Russian Foundation for Basic Research (project No. 17-07-01341) and Project III.17.1.4 of Fundamental Research Program of Siberian Branch of the Russian Academy of Sciences.

References

1. V. Coates, M. Faroque, R. Klavins et al, On the future of technological forecasting. *Technol. Forecast. Soc. Change*, **67**(1) (2001) 1–17. DOI: 10.1016/S0040-1625(00)00122-0
2. A. V. Mikheev, Semantic-based approach for energy technology forecasting, *Information and mathematical technologies in science and management*, 4-1 (2016) 7–15.
3. L. V. Massel, A. G. Massel, Intelligent IT-environment for integration of semantic models to energy research, in *Proc. on 24th International Crimean Conference Microwave and Telecommunication Technology*, 2014. Article no.: 6959469. pp. 441–442. DOI: 10.1109/CRMICO.2014.6959469
4. M. R. Kogalovsky, Ontology-Based Data Access Systems, *Programming and Computer Software*, Vol. 38, No. 4 (2012) 167–182. DOI: 10.1134/S0361768812040032
5. Poggi A., Lembo D., Calvanese D., De Giacomo G., Lenzerini M., Rosati R., Linking Data to Ontologies. *Journal on Data Semantics X. Lecture Notes in Computer Science*, vol 4900. (Springer, Berlin, Heidelberg, 2008) pp. 133–173.
6. M. G. Skjæveland, M. Giese, D. Hovland, E. H. Lian, A. Waaler, Engineering ontology-based access to real-world data sources. *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol. 33 (2015) 112–140. DOI: 10.1016/j.websem.2015.03.002
7. E. Kharlamov, D. Hovland, M. G. Skjæveland, Ontology Based Data Access in Statoil. *Journal of Web Semantics*, Vol. 44 (2017) 3–36. DOI: 10.1016/j.websem.2017.05.005
8. E. Kharlamov et al. Semantic access to streaming and static data at Siemens. *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol. 44 (2017) 54–74. DOI: 10.1016/j.websem.2017.02.001