Methods and Technologies of Building an Intelligent Service for Energy Technology Forecasting

A.N. Kopaygorodsky
Melentiev Energy Systems Institute SB RAS
Irkutsk National Research Technical University
Irkutsk, Russia
e-mail: kopaygorodsky@mail.ru

Abstract[1]

This article reports the approach and software tools for decision-making support in forecasting the energy infrastructure development. The author considers the problem of searching for information from various open sources, technology of information searching, knowledge detection and classification. The author describes architecture of Intelligent information system. For experts this classification and integrated warehouse simplify search for the knowledge required.

1. Introduction

Future-oriented research of industry development is becoming particularly relevant because of complicated economic conditions worldwide. This is necessitated by the task (a) to predict potential results of innovative projects with high reliability and (b) to save money. Because the Energy industry is basic production of final products largely depends on its development. In addition, in the recent years, alternative energy technologies have been actively developed, for instance those related to ecology and energy conservation, e.g. Green Energy.

The technological forecasting of the RF energy sector up to year 2035 considers three different world energy development scenarios [1]. Fidelity of each scenario cannot be properly estimated. This leads to the high uncertainty in innovative development strategic planning because of high investment risks in the energy sector. Improving evaluation of technological development and rolling forecasts will let to focus on the amplification of certain technologies with better effect in future. Thus, providing researchers with topical information and knowledge is an important part of development of expert assessment of new technological solutions in energy infrastructure.

Proceedings of the 20th international workshop on computer science and information technologies CSIT'2018, Bulgaria, Varna, 2018

The team employed at the Melentiev Energy Systems Institute, Siberian Branch, Russian Academy of Sciences (ESI SB RAS) has been developing the intelligent information resources and support software tools for decision-making and collective expert activity for innovative development forecasting of energy infrastructure. The latter is truly important for work at the ESI. The experts from various fields such as Energy, Mathematics, Economy, Ecology and Informatics are involved in research.

2. Related works

Since 2003, the international workgroup of leading scientists from the USA, Europe and East Asia coordinates research on Future Oriented Technology Analysis [2]. Researchers are carried out on such closely related areas as Technology Monitoring, Watch and Alerts, Technical and Competitive Intelligence, Technological Forecasting, Strategic Technology Assessment, development Technology Roadmapping, Technology Foresight and other.

Forecasting as a method of research is used in the domain of Energy Infrastructure to study the development and the functioning. In [3], authors were considering the technological prospects of various directions of decisions of the problem of resource restrictions of the development of wind and solar energy.

Recently, more and more often analysis results of Big Data and Linked Open Data as information source has been used. For example, the Big Data analysis was used for Electricity Load Forecasting [4]. This model of Electricity Load was designed to make predictions for time series with specific properties (strong seasonal dependence and concept drift). The prediction results via Big Data analysis described the behavior of the real Energy System in the near future very well.

Text Mining is the process of deriving information from text. Some applications of this technique are Social Media Monitoring and Scientific Discovery. Authors of [5] propose a conceptual approach to the research into customer satisfaction based on a detailed analysis of consumer reviews written in natural languages.

Artificial Intelligence techniques successfully are used for forecasting the conduct of separate energy technologies. In [6], authors use patent indicators to predict the technological advances in Hydrogen Storage Materials (HSM). Patent analysis was been carried out using bibliometrics and Text Mining approaches in order to forecast the future trend of development. Authors were evaluating the technological life cycle stage, HSM class prominence and the role of different countries in HSM patenting.

3. Energy infrastructure development forecasting

The work on scientific and technological forecasting of Energy infrastructure is complex and requires solving the main problems:

- Need to support sustainable scientific and technological development;
- Acceleration of technological changes;
- Strengthening the impact of progress in science and technology on socio-economic processes;
- Increase in the complexity of facilities, counterparties and, as a consequence, the systems that unite them;
- Strengthening interdisciplinary effects, influence of interrelated areas of knowledge, including convergence of technologies;
- Need for a systematic adaptive forecast and mechanisms for "rolling" planning with continuous updating of scientific and technical information.

The ESI researchers are short of present data and topical knowledge to properly energy infrastructure forecasting.

Available basic methods of technological forecasting are currently used to study the Energy of RF:

1. Expert survey, e.g. Delphi and Foresight method. This is a simple cheap and quick solution. However, these methods have poor qualities as subjectivity, low validity and lack of responsibility.
2. Technological analysis by individual energy companies. This method is based on econometrics, monitoring, databases of development, operational analysis of problems and proposals.
3. System analysis of technology is research that requires a lot of time and effort of highly qualified scientific teams.

Because the last method very expensive and require a lot of information and explicit knowledge usually first two methods are applied for developing forecasts. And moreover using a second method available only within energy companies. Application of semantic technologies facilitates scientific and technological forecasting in the energy sector. This procedure includes:

- Semantic Information Search;
- Analysis of weakly structured texts and data, including the extraction of terms, concepts, their alignment and aggregation;
- Statistical processing and analysis of obtained semantically structured data, including computing of quantitative indicators and indicators of technology development;
- Cognitive technologies of visualization;
- Ontological modelling.

A new approach to extract and analyze knowledge and data lets us to improve quality of scientific results. The uncertainty can be reduced via Large Data Analytics (LDA) applied to global science, technology and innovation data sources. When applying LDA, it is feasible to identify the existing and raise new development trends, as well as foresee the technological breakthroughs by a comprehensive understanding of continuous innovation processes [7].

All possible sources of data and knowledge should be used for solving the problem of development forecasting. Even if the ESI has its own data sources and previous results, it is not enough to fulfill high-quality forecasts. Researchers should attract additional knowledge and data to improve the quality and accuracy of forecasts. Different sources contain information in diverse forms and information have varied level of reliability. Open government data and other open data may be used as additional information sources. For example, to improve the quality of results and probability of energy development forecasting models the researchers may apply to the database of patents and inventions or news feeds. Our research team stick to the governmental information systems, which integrate information about scientific and technical projects and developments, e.g. CITIS, RFBR and FIPS, and various commercial systems, e.g. SCOPUS, Web of Science, RSCI and Science Index as the sources for Tech Mining. The information in such sources is commonly represented in text form. The Tech Mining is a special form of LDA based on the intelligent text mining of electronic information sources. Such Intelligent technologies are capable to reduce large information flow and to classify the information items. It is possible to estimate qualitative and quantitative innovation indicators of potential technological growth in the areas of interest.

The task of elaboration software tools for energy infrastructure development forecasting consists of two main parts:

- Creating an Intelligent Information Environment for collaborative expert work and knowledge accumulation;
- Improving the existing and creating new methods of information search and forecasting.

Researchers use two classes of models to forecast: (1) Qualitative models show the relationship and influence of various factors; they are employed to determine some trends; (2) Quantitative model. The qualitative analysis as the first stage of forecasting development makes it possible to quickly and cheaply solve the problem. In some cases, quantitative modelling can be used as a more expensive approach, but giving better results in addition to the qualitative solution obtained. The main drawback of quantitative models is a long process of their construction and data preparation. The quantitative models are commonly used to compute the consequences of events and strategic decisions. In any case, researchers need to make relevant changes in the model to obtain adequate results, when researching forecast of energy development.

4. Technology of information searching

The most important task is providing relevant information to researchers. They apply some existing models and create new ones in the study process. Fresh information and knowledge permit the existing models to be updated. A large stream of poorly structured information significantly slows down and complicates building and updating of models. This is why researchers have to spend a lot of time searching for and identifying the facts of interest. This is extremely inefficient. Automatic classification is required to effectively extract knowledge from large data sets. Knowledge in explicit form may be described by set of knowledge element (digest). Each element from the set may be mapped to some concept from ontology. This can be represented as

$$F_K: T \to \{K_I\},\quad F_O: K_I \to O_I,\ O_I \subset O'$$, where

F_K presents the text fragment T to set of knowledge element K; $\{K_I\}$ is digest of T; F_O is a mapping for all knowledge element K_I to concepts O_I. And O' is a generalized ontology. Application of Ontology system allows us to simplify navigation and search of information.

Figure 1 exhibits the process of extracting interesting knowledge from the general information flow and their subsequent processing. Knowledge for energy infrastructure development forecasting arrives from various sources. Each information flow is unified and the text fragments are extracted. Some elements may contain keywords, abstracts, descriptions in explicit forms. The other elements may contain only raw text. At the stage of classification two important tasks are performed: (a) definition of word-set and (b) mapping with concepts of the Ontology system.

Methodical approaches to knowledge collection from text sources are employed to solve some tasks:

- to obtain raw semi-structured text data;
- to clean, identify and extract the text;
- to classify the text; draft the digest for each obtained element.

Figure 2 exhibits this process. Main steps of process are the extraction of text fragments and the building document digest. Thus process may be divided between two components: data search engine and data analysis engine. In this case data search engine encapsulate all exclusivity of data source. And data analysis engine perform common algorithm of classification by text fragment.

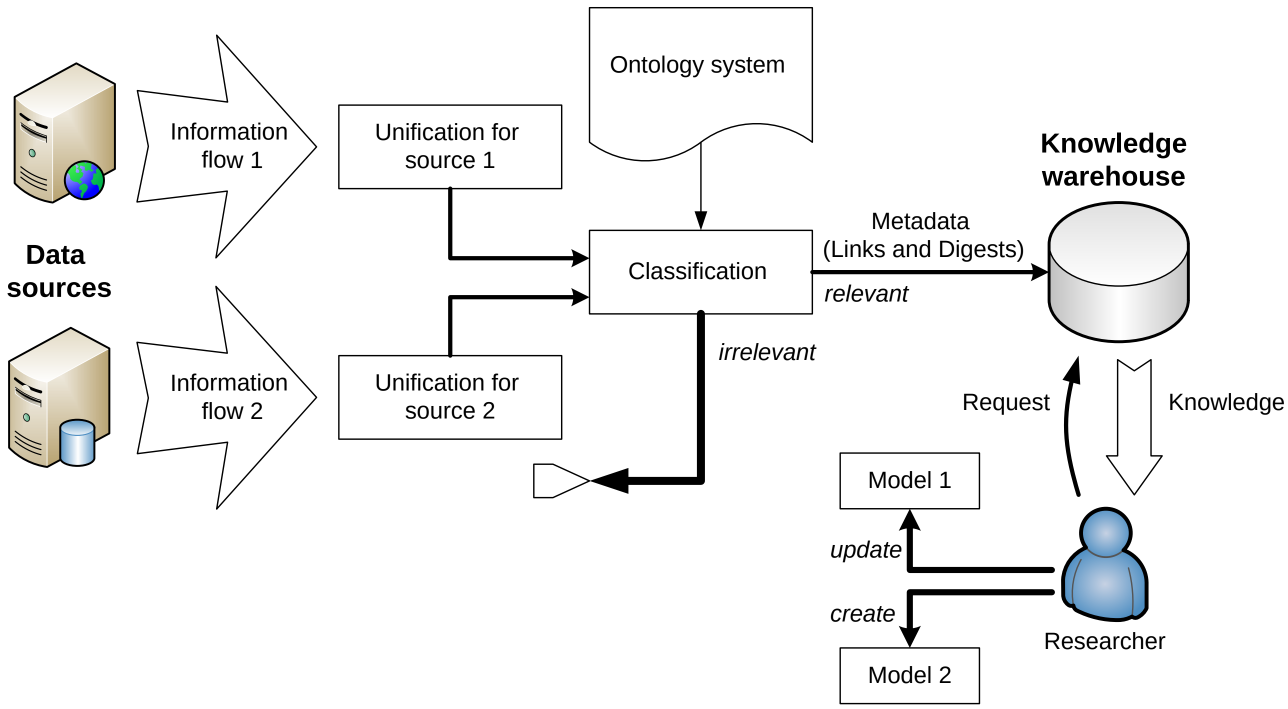

Fig. 1. The process of extracting knowledge from the sources, transformation and using for updating research models.

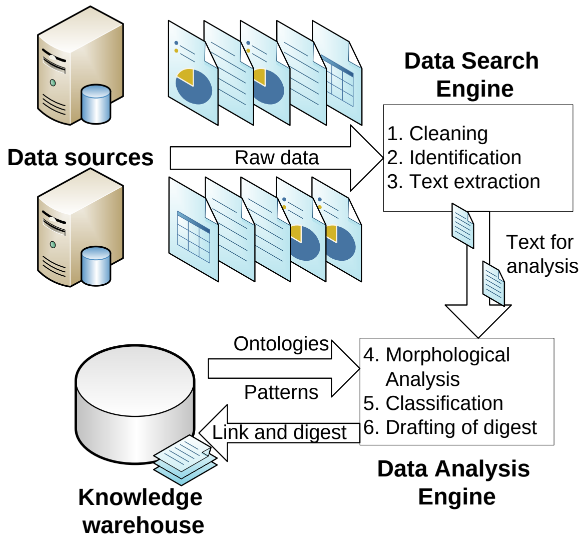

Fig. 2. The process of document classification based on analysis of text fragments.

The first stage of text analysis should be language detection. Using a method of character frequency allow to spot code page and text alphabet and then draw a conclusion about language.

On the next stage may be applied stemming algorithm [8] and improved TF-IDF metrics [9] to detect whole Term Frequency. This approach provide a simple way to define document class. But in this case estimate is rough.

For a better definition of Terms and their relations may be used specialized tools such as KLAN system [10, 11]. Context relation between words may be identified via analysis of word sequences. The result of this step is the set of normalized word-sets (patterns). The groups of patterns are mapped to keywords and concepts of ontology space. Thus, the text fragments of elements are mapped to some set of concepts and numerical characteristics (frequency). As a result of classification at the last step the digest contains keywords related to the Ontology system concepts only. This element holds information about the meaning of the text fragment of the knowledge element. The Knowledge warehouse keeps digests, links on raw documents and ontologies. If the text fragment has not been correlated with the key concepts of the Ontology system, it is not stored in the Knowledge warehouse. Thus, the warehouse contains only the elements of knowledge correlated with the concepts of ontologies. Researchers can access structured knowledge for model adjustment and research.

The texts of scientific and specialized literature were used for learning and testing the classification subsystem of text. Besides, the other texts were used to determine common phrases of the Russian language (words of general vocabulary). The databases of patterns consist of two parts: (1) general vocabulary (about 2000 elements); (2) scientific vocabulary (about 1500 elements). Maximum length of element of scientific vocabulary is 4 words.

Figure 3 exhibits two ways of knowledge warehouse enrichment. The first is collecting documents from structured sources such as government information system, paper collection et al. Source-oriented data adapters may extract information from such sources.

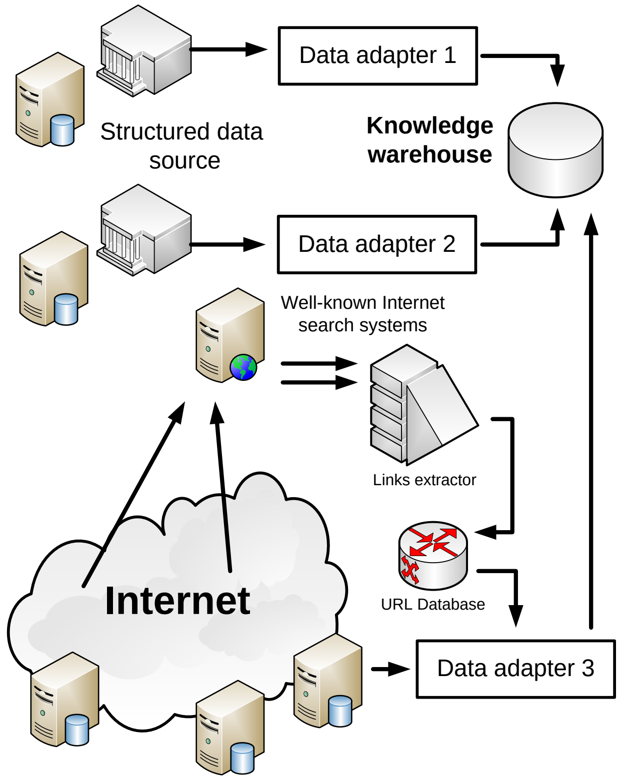

Fig. 3. The scheme of collecting information and enrichment of Warehouse.

The second is a blind Internet searching. Special constructed crawlers extract information from Internet resources and news feeds. Crawlers start job from top level link then looking through HTML extract new links from next level and add to work list. There is a problem of creating initial link set for supply crawlers. The most rational approach was to obtain relevant links directly from search engines by predefined queries. The link extractor interacts with the search system and generates an answer file, the format of which is presented below

```
BEGIN name_of_engine timestamp search_query
url 1
title 1
description 1
…
url n
title n
description n
END
```

The first string contain keyword “BEGIN”, name of search engine, time stamp of request and query. Further there are 3*N rows. Each 3 rows contain information about URL, page title and description.

This format has benefits: all packages are related to time and Internet search system. This makes it possible to observe change in mediaspace in time.

To support expert work and substantiate recommended solutions the author develops Intelligent Information Environment of Energy Innovative Development Forecasting (IIEIDF). They apply the tool set of IIEIDF to perform own studies. IIEIDF supplies interconnections between applications and stores and transforms the data and knowledge.

5. Architecture of Intelligent Information Environment

Intelligent Information Environment of Energy Innovative Development Forecasting consists of several components (Fig. 4): data search and data analysis engines, knowledge and data warehouses; FishTail server; Web-server supporting the server-side components of Web-applications. IIEIDF applies the some existing components of Intellectual Collective Expert Environment (ICEE) [12]. The data search engine automatically performs scheduled searching, obtains knowledge elements and then extracts the text content. The data analysis engine performs lexical and morphological analysis of the text fragments and mapping to concepts.

Knowledge and data warehouse describes not only scientific knowledge, but also contains initial information for investigation. In addition, the knowledge warehouse store digests and links the knowledge elements from open Internet data and sources. Access to data sources and separate servers of database management systems is reached by virtual integration methods [13]. These database management systems contain information and results of some other ESI SB RAS investigations.

The FishTail Server developed at ESI SB RAS is applied to store and manage RDF-triples associated with some knowledge and concepts of Ontology space. The FishTail is employed for enrichment by the new RDF-triples based on some predefined rules and existing RDF-triples. Each rule has body and head. Head contain one or more groups of triplets, body has only one triplets. If head is true and body is absent then FishTail adds body triplet in the storage.

Figure 5 exhibits placing a IIEIDF services. The most valuable services are located in ESI SB RAS. Data adapters and crawlers required fast Internet connection and low latency. Therefore, this components was be placed in Datacenter. Data exchange between sites is not large by reason of applying local copy of Ontology database. One important consequence of the selected architecture is a easy increasing count work-servers

Implementation of Web-based application components IIEIDF are performed in a Private Cloud on the basis of approach of Rich Internet Applications (RIA). RIA Web-application is a full functionality of traditional desktop applications. The implementation of the RIA client Web-based applications is performed on the Java platform as the most suitable and convenient one for development. The technology of Java Web Start (JWS) and Java Network Launching Protocol (JNLP) lets downloading and running quazilocal applications that are automatically updated through the Internet and work outside of the Web-browser environment.

Researchers commonly apply relevant modelling tools to support the proposed methodical approach, e.g. OntoMap, CognitiveMap, EventMap and BayNet. The modelling software tools for qualitative analysis are based on graphical core GrModeling [14]. This approach provides shared space for building and transforming models.

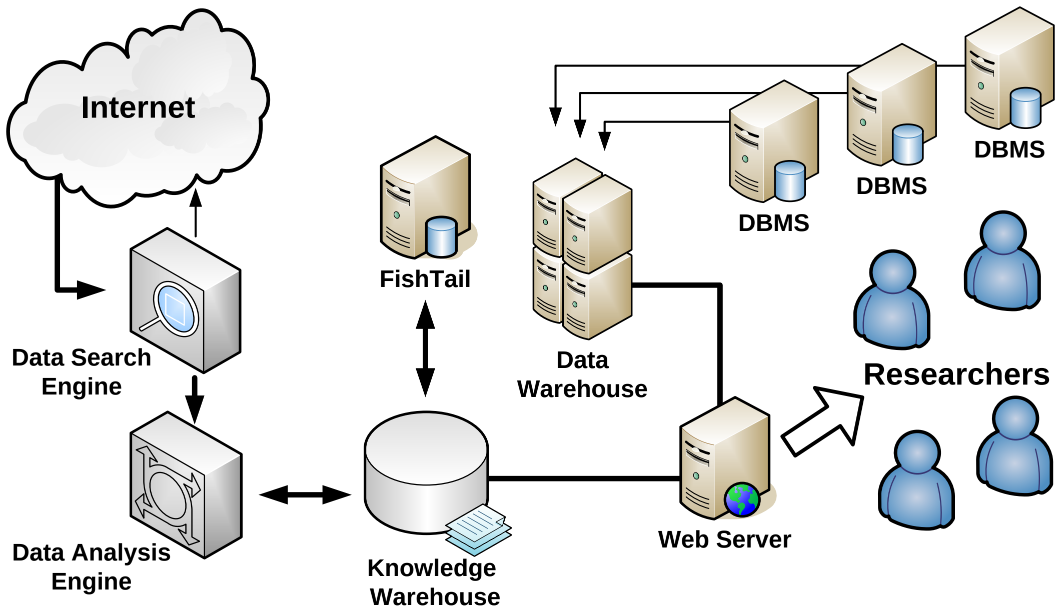

Fig. 4. Logical Architecture of Intelligent Information Environment of Energy Innovative Development Forecasting.

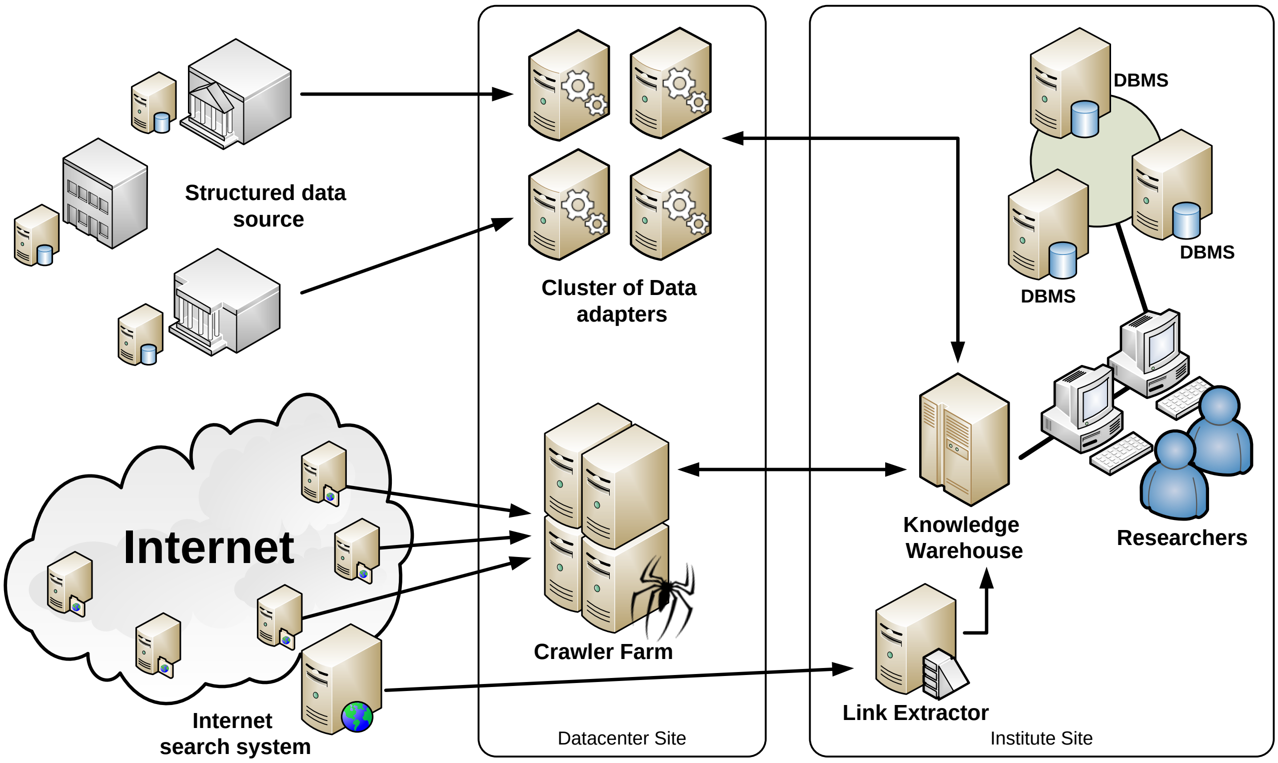

Fig. 5. Placing a services of Intelligent Information Environment of Energy Innovative Development Forecasting.

Implementation of native Web-application that contained not so many various functions is produced with such technologies as JavaScript (jQuery, Bootstrap et al.) and HTML5. Server-side components are developed in PHP and Java. To simplify the creation of PHP-applications is using its own MVC framework and such technology as TWIG.

6. Conclusion

The article considers methods, technology, algorithms and software tools to performing searching and analysis information from various sources:

- The Open data, Internet and other commercial systems may be employed as the additional knowledge sources for the purpose of the Energy Infrastructure Development Forecasting.
- The process of searching and extracting new knowledge is based on lexical and morphological analysis, TD-IDF method and frequencies analysis.
- The analysis of text fragments and creation of ontology-based digest improve the efficiency and quality of forecast estimation via simplified access to knowledge. The results of qualitative modelling are transferred into the process of numerical computing.
- The Intelligent Information Environment of Energy Innovative Development Forecasting includes the mechanisms for data search and analysis. Researchers use Web-applications to access knowledge and data warehouse, they also use the RIA-tools for creating own models.

These techniques surely improve quality of forecast results in the studies of energy infrastructure development.

Acknowledgments

The author is grateful to the Russian Foundation for Basic Research (RFBR) for partial financial support of the research via projects 17-07-01341, 16-07-00569. The results were obtained during the implementation of the basis scientific project of the fundamental research programs of SB RAS III.17.2.1, reg. № AAAA-A17-117030310444-2, III.17.1.4, reg. № AAAA-A17-117030310436-7.

This research was performed using the servers of the multi-access scientific center "High Temperature Circuit" of Melentiev Energy Systems Institute SB RAS.

References

1. The technological forecast of sectors development of the Russian fuel and energy complex for period till 2035. Ministry of energy RF, 2016, Available at http://minenergo.gov.ru/system/download-pdf/6365/66647.

2. Cagnin C., Keenan M., Johnston R., Scapolo F., Barre R. “Future-oriented technology analysis. strategic intelligence for an innovative economy”, Springer, 2008, 170 p., DOI: 10.1007/978-3-540-68811-2.

3. Nizhegorodtsev R. M., Ratner S. V., “Trends in the Development of Industrially Assimilated Renewable Energy: The Problem of Resource Restrictions”, *Thermal Engineering*, Springer, 2016, pp. 197-207, DOI:10.1134/S0040601516030083.

4. Grmanova G., Laurinec P., Rozinajova V. et al., “Incremental Ensemble Learning for Electricity Load Forecasting”, *Acta Polytech Hungarica*, vol. 13, no. 2, 2016, pp. 97-117. DOI: 10.12700/APH.13.2. 2016.2.6.

5. Yussupova N., George Kovacs G., Maxim Boyko M. et al., “Models and Methods for Quality Management Based on Artificial Intelligence Applications”, *Acta Polytech Hungarica*, vol. 13, no. 3, 2016, pp. 45-60. DOI: 10.12700/APH.13.3.2016.3.3.

6. Chanchetti L. F., Diaz S. M. O., Leiva D. R. et al., “Technological Forecasting of Hydrogen Storage Materials using Patent Indicators”, *International Journal of Hydrogen Energy*, vol. 41, Elsevier, 2016, pp. 18301-18310, DOI:10.1016/j.ijhydene.2016.08. 137.

7. Daim T.U., Chiavetta D., Porter A.L., Saritas O., “Anticipating future innovation pathways through large data analysis”, *Innovation, technology, and knowledge management*, Springer, 2016, 360 p., DOI 10.1007/978-3-319-39056-7.

8. Willett P., “The Porter stemming algorithm: then and now”, In: *Program: Electronic Library and Information Systems*, 2006, pp. 219-223.

9. Huilong Fan, Yongbin Qin., “Research on Text Classification Based on Improved TF-IDF Algorithm”, *The International Conference on Network, Communication, Computer Engineering*, 2018, pp. 501-506.

10. Sidorova E. “Multipurpose dictionary subsystem for extraction of subject lexicon”, *Proc. International conference Computer linguistics and intellectual technologies (Dialogue’2008)*, Moscow, 2008, pp. 475-481.

11. Kononenko I.S., Sidorova E.A. “Language Resources in Ontology-Driven Information Systems”, *Proc. First Russia and Pacific Conference on Computer Technology and Applications, 6-9 September*, Vladivostok, Russia, 2010, pp. 18-23.

12. Kopaygorodskiy A.N. “Knowledge management via intellectual collective expert environment”, *Proceeding of the 18th Workshop on Computer Science and Information Technologies (CSIT’2016)*, Vol. 1. USATU, Ufa, Russia, 2016, pp. 65-69.

13. Kopaygorodsky A.N. “Virtual integration of distributed data of energy research”, *Proceedings of the XIII Russian conference of the Baikal Information and mathematical technologies in science and management*, ESI, Irkutsk, Russia, 2008, Vol. 3, pp. 260-266.

14. Kopaygorodsky A.N. “Design and implementation of a graphical modeling system”, *Proceedings of the XV Russian conference of the Baikal Information and mathematical technologies in science and management*, ESI, Irkutsk, Russia, 2010, pp. 22-28.