fracturing indexes statistical characteristics have been defined. Key words: drillability rocks, fracturing, electric drive, rotation frequency, control, current, power, torque, correlation, structural scheme.

Keywords: rock drilled, fracture, electric, speed, control, current, power, torque, correlation, block diagram.

UDC 665.9

T. B. SHATOVSKA, associate Professor, National University of Radioelectronics, Kharkov;
A. A. ZAREMSKAYA, student, National University of Radioelectronics, Kharkov

ONTOLOGY MODELS RESEARCH AND DEVELOPMENT FOR DATA MINING REPOSITORY

Currently, there are a lot of data mining, as well as a large number of data sets stored in different repositories. A significant problem is the lack of analysis methods themselves in the repositories, in fact there is no connection to a specific data set method for the respective data sets. But in this work we offer the implementation of the system and model research ontology and development for our Data mining repository.

Keywords: DM, Ontology models, data, metadata, machine learning, SDMX Standards Version 2, repositories, dataset.

Implementation

Search algorithm is based on ontological models and ranking results. Implementation of access to intelligent agents based on web-services, implemented on the basis of programming paradigms such as dependency injection and control inversion. There were a variety of access through the implementation of web-projects and implemented a more flexible, robust architecture. During this work we analyzed and present ontology models for data mining repository.

The above mentioned drawbacks have a point to create a new public information repository to store datasets using intelligent agent and ontological approach for storing, conversion, search, add, description, selection of the required information for researchers' needs in the field of Data mining and Machine Learning. Using the Protégé 3.4 we created an ontology model of Data mining methods, an ontology model of the user, a model of the resource (W3C, 2004).

A base standard was chosen the standard SDMX Standards Version 2.0 and the main parameters of the Statistical European Repositories were taken. The interaction between the ontological models is based on intelligent agents: coordinator agent, resource agent, search agent, a user agent. The agent approach has been implemented by multitechnology JADEX. We use intelligent software agents. This is a new class of software systems, which acts either on behalf of the user, or on behalf of the system. They are, in fact, a new level of abstraction, different from the usual abstract type - classes, methods and functions. For practical implementation of these agents JADE offers to programmer-designer of agent systems the following possibilities: FIPA-compliant Agent Platform, which includes system agents AMS, ACC and DF; Multiple Domains support – DF agents and so on (IEEE Computer Society standards organization, 2006; Bellifemine, et al., 2006).

Ontology representation

In recent years the development of ontologies are formal descriptions of explicit terms for business and relations between them. In the World Wide Web became commonplace ontology. Ontology in the network range from large taxonomy that categorize websites (Yahoo! website) to products and their characteristics (like on the website Amazon.com). Consortium WWW (W3C) develops RDF (Resource Description Framework). The RDF is language of encoding knowledge on Web pages. It makes knowledge understandable to electronic agents to search information. Now many disciplines develop standard ontology that can be used by experts in subject areas to share and annotate information in their field. For example, in medicine large standard structured dictionaries such as semantic web unified medical language system (the Unified Medical

Language System) were created. Also large ontology appears general intent. For example, the UN Program for Development (the United Nations Development Program) and the company Dun & Bradstreet combined efforts to develop ontology UNSPSC. It provides terms for goods and services. Ontology defines a common vocabulary for researchers, who need to share information in the subject area. It includes machine-interpreted formulating the basic notions of domain and relations between them. Ontologies are developed for joint use by people or software agents common understanding of data structures for possible reuse of knowledge in the subject area, to make assumptions explicit in the subject area, to separate knowledge in subject area of operational knowledge, the analysis of knowledge in subject area.

Program - instrumental method of implementation of the ontological model

The tool Protégé 3.4 for “data mining repositpry” was selected. It was developed at Stanford University (USA) (Gennari J.H. et al., 2002). Protégé 3.4 is meta-tool. It helps users to create a system of acquisition knowledge for a particular subject area and experts can use these systems to enter and view the information contained in electronic databases of knowledge. The modular architecture of Protégé 3.4 very expands class of systems that can be collected for certain tasks on the acquisition of knowledge and making the future of knowledge acquisition can be better adjusted in accordance with certain requirements of end users. The Protégé 3.4 developers say: "The system is open software. It is difficult to calculate the number of users ..." Now the list on the Protégé, nearly 9,000 subscribers, and website Protégé registered over 20,000 users (we can download the Protégé without registration). You can download 85 different plugins for Protege from the site. Protégé user community is very active and has representatives in more than 100 countries. The functional editor is inextricably linked to the specific for the ontology model and knowledge arising from the classification scheme vocabulary. The editor has a graphical interface that provides a visual edit mode. Graphical interface is implemented on the basis of standard software Object TreeView, a significant addition of additional functionality - mainly in the search, input and control logic. Ontology editor Functionality is:

• View and search: supports viewing grid, standard types of search time;

• editing (input, correction, deletion);

• logical control in the introduction: the introduction of technology almost completely eliminates the violation defined description schemes;

• functionality testing: writing queries;

• interaction with other ontology (import - export, mainly using communicative presentation formats).

Ontology source model (dataset ontology model)

Information about “data and metadata exchange repository” is stored in the ontological models form. One of the main classes of this model is «data set» (DataSet). Each separate instance of this class contains information about the data set to this information include name, analysis method, short description, information about its creators and more. This class contains several classes that belong to its structure: DataSetFile and Judge. Class DataSetFile contains information about the sample that covers this data set, but Judge class contains information about the evaluation of the different set of moderators.

Ontological source models development

The Ontologies are developed and can be used in solving various problems, including joint use of people or software agents to possible accumulation and reuse of knowledge in the subject area, to create models and programs that operate ontology, but not rigidly defined data structures, analysis of knowledge in the subject area. For a more intelligent synthesis of information systems section must define ontology, which should describe the terminology used in the contents of set rules for the use of these terms in the context of other terms.

The basic building block of dataset model is an assertion that represents: resource named property and value. In RDF terminology these three statements are respectively: subject,

predicate and object (W3C, 1999). Show description of the dataset source in the environment of ontologies Protégé 3.4. Classes and attributes of selected classes are created and presented in Figure 1. In the development of ontological models of the resource repository was allocated 3 classes. Here more detail the selection process classes. First of all area and scope of ontology were defined. Then important terms of source ontological model of "data and metadata exchange repository": sample, method of analysis, attribute, subject area, data set description, the dataset file, name, type, articles that refer to the dataset, keywords, author, date of creation. We highlighted three classes and a set of slots in the ontological model of the resource:

- DataSet;
- DataSetFile;
- Judge.

In this thesis ontology source model are described by three classes. After ontology model source determination Protégé system allows to convert Protégé project to RDF model.

Ontology data mining model

Ontology data mining model is the exact specification of the subject area. It provides a vocabulary for presenting and sharing knowledge about methods of analysis and methods of deduction and many relationships established between terms in the dictionary. One of the advantages of using this ontology is a systematic approach to the study of the subject area. It is achieved: systematic (ontology presents a holistic view of the subject area); monotony is material (represented in a single form is much better perceived and reproduced), scientific development (construction of ontology can restore missing logical links in their entirety).

Ontological data mining models development

There are two levels on which ontologies are used to support data processing: domain ontologies and task ontologies. Domain ontologies are used to describe knowledge from the domains relevant to the particular task (Fig. 1). The first step in the ontology development is the definition of the domain and scope of the ontology itself: in our scenario the ontology will cover the Data Mining domain. To build a consistent ontology model it is necessary to establish for what we are going to use the ontology and for what types of questions the information in the ontology should provide answer. The choice of how to structure ontology determines what a system can know and reason about. We have built our ontology through a characterization of data mining methods that is classified on the basis of some parameters useful to select the more ones method to solve a KDD problem. Repository determines characteristics of the data and of the desired mining result, and enumerates the DM processes that are valid for producing the desired result from the given data. Then the Repository assists the user in choosing processes to execute, for example, by ranking the process (heuristically) according to what is important to the user. Results will need to be ranked differently for different users. A different user may want to minimize run time, in order to get results quickly. There are other ranking criteria: accuracy, cost sensitivity, comprehensibility, etc., and many combinations thereof.

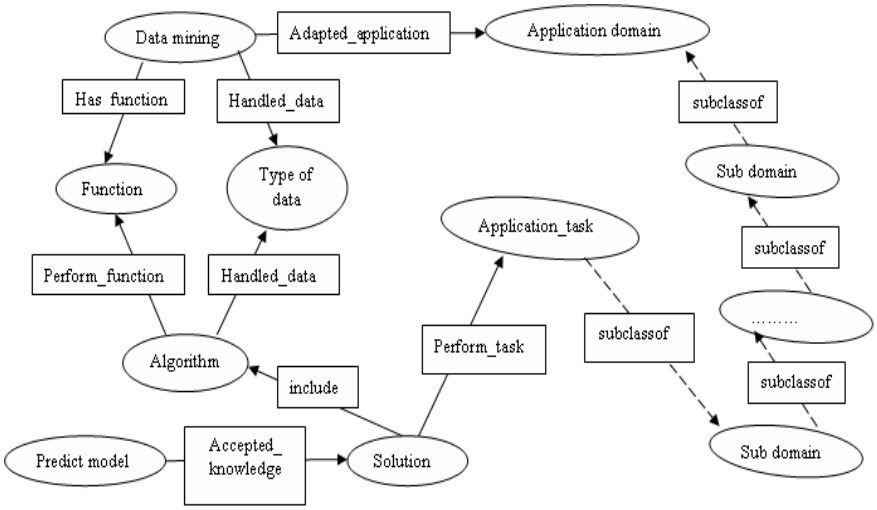

Fig 1 - Data mining ontology, where property relation of concepts, Subclass relation of concepts

To solve problems related to data analysis in the presence of random and unpredictable effects, mathematicians and other researchers over the last two hundred years produced a powerful and flexible arsenal of methods, collectively called mathematical statistics. During this time extensive experience was gained in the successful application of these methods in different spheres of human activity from economics to space research. And under certain conditions these methods allow for the optimal solution. For example, one of the problems solved in the radiolocation is the known signal detection in background additive interference in the form of white noise. Mathematical statistics methods solve this problem successfully. It is difficult to imagine the need for other approaches to solving this problem.

Because knowledge is personal in nature, the same subject area can be described by different ontologies. This is particularly true of domains that are not formalized or when there are many contentious issues. In this work one of the problems is the task of ontology development methods is data mining. Certainly a good practice is to use already existing ontologies and a good specialist should be able to quickly find existing and already proven any ontology or an algorithm, rather than spend time on developing new. The fact is that ontologies are not clearly structured and formalized. Now a lot of online ontologies and of course they are all correct. But research of existing data mining ontologies did not give a satisfactory result.

Therefore a new ontology was developed. Analysis of knowledge in the subject field of Data Mining is quite possible because there is a declarative specification of terms. Formal analysis of the terms will be extremely valuable as when to reuse the developed ontology so in its expansion. The reason for the development of ontology data analysis provides an Analysis Method slot in DataSet class of developed ontological model resource.

It contains data mining methods that are under all this set of statistics. Ontology with a set of individual instances of classes forms a knowledge base. In fact, in this case it is difficult to determine where the ontology ends and where the start of knowledge base. Ontological model was presented in Fig. 2.

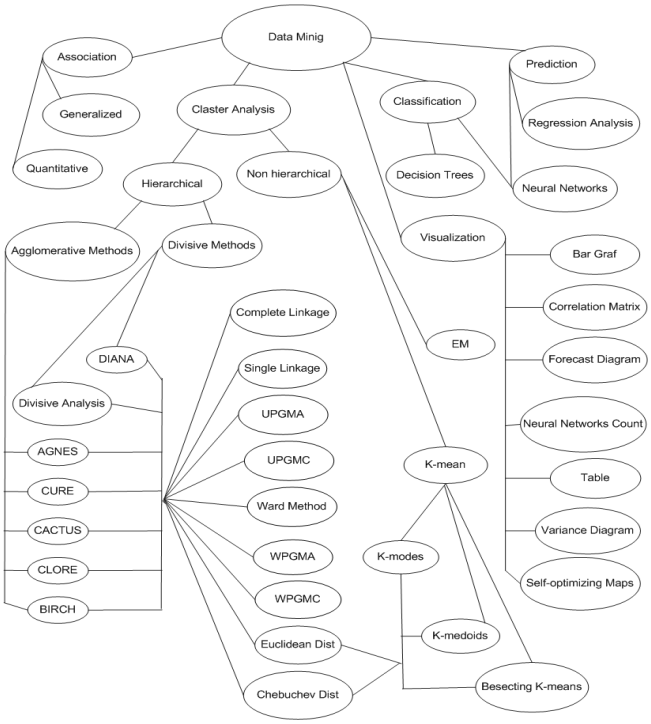

Fig 2 - Partial Data mining ontology model

Ontology user model

Ontological approach is offered for creation of model of user for intellectual repository “data mining repository”. This approach allows taking into account the collection of concepts and connections between them, having a place at interaction of the user with our repository. Ontology user model is the model for data structuring. It stores information about user. User model is obviously for our repository with different levels of training for work with a computer, with a variety of mental, psychological and physiological capabilities (Cargar, 2008; Waltz, 2008).

Ontological user models development

Protégé system has the following possibilities: tabs for ontology replenishment, functional expansion modules, generation of knowledge acquisition module requests and the logical deduction module.

This ontological model includes two abstract classes: Account and Person. Class Account represents the user as the logical nature of the user system. Class Person represents the user as the

person using the system. Experienced and Beginner classes are beginner and advanced user respectively. Admin slots class match Experienced slot class.

The ontological user models of classes description.

In Table.1 the description of Address slots class is shown.

Table 1 - Adress slots class

Attribute	Type	Pover	Presence	Description
country	String	1	Mandatory	country
city	String	1	Optional	city

In Table 2 University slots class is present. It works for base University description.

Table 2 - University slots class

Attribute	Type	Power	Presence	Description
name	String	1	Mandatory	University name
address	Address	1	Optional	University address

In Table 3 Preference slots class is present. It works for user interest and search requests description.

Table 3 - Preference slots class

Attribute	Type	Power	Presence	Description
interest	DataMiningMethod	*	Optional	Data format
search	String	*	Optional	A lot of user search requests
searchHistory	SearchHistory	*	Optional	A lot of user search requests results

The abstract Account slots class is present in Table 4.

Table 4 - Slots of abstract class Account

Attribute	Type	Power	Presence	Description
password	String	1	Mandatory	password
created	String	1	Mandatory	Date of creation
email	String	1	Mandatory	e-mail
preferences	Preference	1	Optional	Information about preferences
title	String	1	Optional	display name

Table 5 - Slots of abstract class Person

Attribute	Type	Power	Presence	description
first_name	String	1	Optional	Name
last_name	String	1	Optional	Surname
gender	Symbol (Male, Female)	1	Optional	Sex (male\female)
university	University	1	Optional	Information about university

The Account class is user representation base.

The slots of abstract class Person are present in Table 5. Its base class is Account. Class Person is base for Beginner and Experienced classes. Beginner class has the same slots as class Person. Slots of Experienced class are present in Table 6.

Table 6- slots of Experienced class

Attribute	Type	Power	Presence	description
speciality	String	*	Mandatory	speciality

Conclusion

Currently, there are many repositories of scientific datasets. The main disadvantages occurred in these systems are: text-only format is not convenient to use and to change the format of files, not user-friendly interface, and the search is only by one of many criteria, i.e. not allowed to combine the search for a number of conditions, poor search.

In many systems, there is no any understanding for what tasks you can use this dataset, there is also insufficient information on the data. Currently, the agent technologies are widespread, where the main part is the agent - a software entity capable of such qualities as autonomy, activity, commitment, mobility, sociability. The creation of ontologies is a prospective direction of up-to-date research in processing of information provided in natural language. One of the advantages of using ontologies as a tool for learning is a systematic approach to the study of the subject area. Meanwhile achieved: regularity - Ontology provides a holistic view of the subject area, uniformity - the material presented in a unified format is much better perceived and reproduced; scientific - Building the ontology allows to restore the missing logical link in their entirety. Also, ontologies allow the use the great volumes of data from different systems, due to the fact they creating the semantic description of data.

References: **1.** W3C Recommendation (2004). Web Ontology Language (OWL) overview, viewed <http://www.w3.org/TR/owl-features/>, (090812) **2**. *Ratushin, U., Polenok, S., Tkachenko,* S. (2001). Information society ontology at the network. University book, 256. **3.** W3C Proposed Recommendation (1999). Resource Description Framework (RDF) Model and Syntax Specification, viewed < http://www.w3.org/TR/PR-rdf-syntax/>, (090812) **4.** *Wooldridge, M., Jennings, N.* (1995). Intelligent agents: Theory and practice. The Knowledge Engineering Review 10(2), 115-152. **5.** *Russell, S., Norvig, P*. (2006). Russian translation of Artificial Intelligence: A Modern Approach, 2nd Edition, Translated by Ptitsyn K. Moscow: Williams Publishing, ISBN Press, 356. **6.** SDMX Standards: Version 2.0 (2007), ZIP File, viewed <http://sdmx.org/?page_id=16#package>, (090812) **7**. *Blake,C. L., Merz, C. J*. (2001). UCI repository of machine learning databases, viewed <http://www.ics.uci.edu/~mlearn/ML - Repository.html>, (090812) **8**. *Cortez, P., Morais, A.*(2007). A Data Mining Approach to Predict Forest Fires using Meteorological Data. In *Neves, J., Santos, M. F., Machado, J. Eds.*, New Trends in Artificial Intelligence, Proceedings of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence, Guimarres, Portugal, 512-523, viewed <http://www3.dsi.uminho.pt/pcortez/fires.pdf>, (090812)

Надійшла до редколегії 20.12.2012

УДК 665.9

Ontology models research and development for data mining repository/ T. B. Shatovska, A. A. Zaremskaya // Вісник НТУ «ХПІ». Серія: Нові рішення в сучасних технологіях. – Х: НТУ «ХПІ», – 2012. - № 68 (974). – С. 27-32. – Бібліогр.:8 назв.

В настоящее время существует много интеллектуального анализа данных, а также большое количество наборов данных, хранящихся в различных хранилищах. Существенной проблемой является отсутствие самих методов анализа в репозиториях, более того нет связи наборов данных с конкретным методом для соответствующих наборов данных. Но в этой работе мы представляем структуру системы и онтологические модели для нашего хранилища данных.

Ключевые слова: модели онтологии, данные, метаданные, машинное изучение, репозиторий, имитирующие данные, наборы данных.

В даний час існує багато інтелектуального аналізу даних, а також велика кількість наборів даних, що зберігаються в різних сховищах. Суттєвою проблемою є відсутність самих методів аналізу в репозиторіях, більш того немає зв'язку наборів даних з конкретним методом для відповідних наборів даних. Але в цій роботі ми представляємо структуру системи і онтологічні моделі для нашого сховища даних.

Ключові слова: моделі онтології, дані, метадані, машинне вивчення, репозиторій, імітуючи дані, набори даних.

УДК 044.89; 681.513.8

М. В. ЕВЛАНОВ, канд. техн. наук, доц., ХНУРЭ, Харьков

КОНЦЕПЦИЯ ПРЕДСТАВЛЕНИЯ ТРЕБОВАНИЙ К ИНФОРМАЦИОННОЙ СИСТЕМЕ

Предложена концепция представления требований к информационной системе, которая, в отличие от существующих, позволяет формализовать задачу выбора методологий и информационных технологий,