КОНЦЕПЦІЯ DATA MINING ПОБУДОВИ ПОШУКОВИХ ЗАПИТІВ ПО ВЕЛИКІЙ КІЛЬКОСТІ КРИТЕРІЇВ

КОНЦЕПЦИЯ DATA MINING ПОСТРОЕНИЯ ПОИСКОВЫХ ЗАПРОСОВ ПО БОЛЬШОМУ КОЛИЧЕСТВУ КРИТЕРИЕВ

THE CONCEPT OF DATA MINING BUILDING SEARCH QUERIES FOR A LARGE NUMBER OF CRITERIA

Н.О. ШИБАЕВА, канд.техн.наук,
В.В. ВЫЧУЖАНИН, докт.техн.наук, **Д.С. ШИБАЕВ**
Одесский национальный морской университет, Украина

Паливна система судна, є однією з технічних систем, свідчення стану якої передаються з високою частотою. Це дозволяє отримувати інформацію про стан паливних резервуарів судна, а також контролювати перекачування палива ну судно.

Судновласники є зацікавленою стороною у процесах контролю за судновим витратою палива, його ефективністю використання, а також якісної заправці судна в порту. Для досягнення найбільшого контролю, використовуються цифрові системи контролю наповненості резервуарів, що складаються з великої кількості датчиків. Також, для досягнення більшої точності заправки судна, обладнуються датчиками паливні магістралі як з віддає боку, так і з приймаючою.

Інформація, одержувана від датчиків, зберігається в суднових базах даних, на головному судновому сервері, а також дублюються на берегові сервера компанії судновласника, для подальшого аналізу. Для подальшої обробки великого обсягу інформації використовуються сучасні засоби аналізу великого обсягу даних. Одним з таких методів є концепція data mining, що дозволяє побудувати сучасні пошукових запит з подальшою обробкою отриманої інформації з великої кількості критеріїв.

Ключові слова: *паливна система судна, великі дані, аналіз даних, суднові технології, контроль палива на судні.*

Топливная система судна, является одной из технических систем, показания состояния которой передаются с высокой частотой. Это позволяет получать информацию о состоянии топливных резервуаров судна, а также контролировать перекачку топлива ну судно.

Судовладельцы являются заинтересованной стороной в процессах контроля за судовым расходом топлива, его эффективностью использования, а также качественной заправке судна в порту.

Для достижения наибольшего контроля, используются цифровые системы контроля наполненности резервуаров, состоящих из большого количества датчиков. Также, для достижения большей точности заправки судна, датчиками оборудуются топливные магистрали как с отдающей стороны, так и с принимающей.

Информация, получаемая от датчиков, сохраняется в судовых базах данных, на главном судовом сервере, а также дублируются на береговые сервера компании судовладельца, для дальнейшего анализа. Для последующей обработки большого объема информации используются современные средства анализа большого объема данных. Одним из таких методов является концепция data mining, позволяющая построить современные поисковые запросы с последующей обработкой полученной информации по большому количеству критериев.

Ключевые слова: *топливная система судна, большие данные, анализ данных, судовые технологии, контроль топлива на судне.*

Fuel system the vessel is one of the technical systems, the indications of which are transmitted with high frequency. This allows you to obtain information about the state of the fuel tanks of the vessel, and to control the pumping well, the ship.

Shipowners are interested parties in the control of marine fuel consumption, efficiency, and quality refueling ship in the port. For best control, use digital control system of filling tanks, consisting of a large number of sensors. Also, to achieve greater accuracy of filling of the vessel, sensors are equipped with fuel line with both giving and receiving.

The information obtained from the sensors is stored in the ship's databases on the main ship server and duplicated on the coast, the servers of the company of the shipowner, for further analysis. For processing large amount of information using modern means of analyzing large amounts of data. One of these methods is the concept of data mining, which allows to build a sophisticated search query with subsequent processing of the received information on a large number of criteria.

Keywords: *the fuel system of a vessel, big data, data analysis, marine technology, control of fuel on the ship.*

Introduction. Fuel on the ship is the responsible maintenance operation that requires special attention from the officers. It must be carried out in strict accordance with the «safety Rules on the vessels» and under the control of the engineer.

Directly supervises the admission of fuel to a third mechanic. Before receiving fuel, a senior (third) mechanic should be familiar with the fuel passport, to verify his data with the standard and to establish compliance with the technical specifications for the marine engine.

If necessary, select a test sample in accordance with GOST, in two copies: one for the tank farm and to the analysis of the thermotechnical laboratory of the shipping company. The sample is fixed by the act, composed of representatives of the depots and vessels.

Analysis. Fuel can be divided into three periods: preparatory, fuel and final. In the process of preparing the third operator has to provide the correct fuel and fire safety:

− measure the remaining fuel in the ship's tanks;

− coordinate with the mate in charge of cargo operations, the order of filling of fuel tanks;

− prepare fuel system for receiving fuel;

− to determine the fuel temperature is taken;

− assign the watch minder in place of receiving fuel from the vessel;

− alert the watch to the Navigator about the upcoming receiving fuel;

− prepare for the effect of ship fire-fighting equipment;

− to install the suction piping a tray of sand and a fire extinguisher.

When receiving fuel control the filling of tanks and communicate with guaranteed timely object to change the mode of fuel delivery and the termination of bunkering.

The fuel supply is stopped to the limit of filling the tank, to avoid «pressing».

When taken from the shore medium viscosity fuels in the winter and high viscosity at any time of the year should take into account the possibility of formation of plugs of solidified fuel in the receive pipeline.

It is therefore necessary to maintain the temperature of fuel take at least 15-20 ° C above the pour point, and after bunkering for 20-30 minutes to leave open the suction valves to fuel the glass in the tank.

After the fuel empty the suction piping of the fuel tanker and the vessel and divide the pipelines [1].

After receiving fuel performs measurements in tanks, from which fuel was supplied to the ship and in the ship's fuel tanks, setting the number of the adopted fuel. A lot of the adopted fuel is determined from the temperature correction on density.

The amount of water trapped in the fuel should be removed from admissions documentation.

The receiving process ends with the drawing up of the waybill signed by the representative of the tank farm and the mechanic took the fuel. Passport of fuel and invoice the chief engineer is the shipping company together with the machine report.

The purpose of the study. Each of these processes can be automated and simplified through the use of modern technology. Most ship systems are fully covered by the monitoring system and the deposition per time unit. Fuel system the vessel is no exception.

The fuel tanks using several types of sensors for more precise identification of the amount of fuel. Such sensors include float fuel level sensors, submersible sensors control the viscosity of the fuel and temperature gauges and the incoming fuel in the tank.

The information every second. For greater accuracy the impact of fuel, additionally installed sensors on the tank return and fuel on the main pipelines through which fuel is supplied to the vessel. System triplicates indications of fuel transfer can significantly increase the accuracy of loading fuel for the ship.

Through the use of such a control system, the accuracy of the amount of fuel in the tanks every second saved group readings from all sensors involved in this process.

All these data form a large array of information that is stored in the ship database and having a non-relational structure.

Shipping companies are interested in more accurate forecast of the consumption and control of fuel over the period of a ship's voyage.

This is done the system analysis of the fuel consumption and energy efficiency of powertrains vessels. To ensure these calculations used the same sensors as in the process of pumping the fuel. However, to perform this control in real time is problematic due to the large load on the communication channel.

The results of the study. Optimal is a method of processing data in flight, to control the running modes of the vessel, and also proportional to the fuel consumption in these modes. As the scope of information is enough large, it is necessary to develop a comprehensive solution that is able to analyze non-relational ship database for a long period of time and finding the exact data for each specified injection.

Algorithmic methods of processing large amounts of information, can simplify the search for such information.

Big data characterized by volume, variety and speed with which structured and unstructured data coming at networks in processors and storage, along with the processes of turning this data into information [2].

In these circumstances, to increase the efficiency of analyzing large amounts of information there is a need of new and innovative methods of information processing by automated cognitive procedures using the database of facts and knowledge bases, automatic generation of hypotheses, procedures, explanations initial state database of facts for the justification and acceptance of the hypothesis [3]. At an earlier stage of development of these technologies, attempts were made to use this approach to interpret a large number of calculated data [4].

One of the methodologies, which helps to solve tasks of different classes of search patterns and the interpretation of the results is the methodology of data mining Data Mining.

It is used to detect and explore patterns in arrays of semi-structured information and building models describing the behavior of complex systems. Data Mining is the exploration and discovery of «machine» in the raw data of knowledge that were not previously known, non-trivial, practically useful and available for human interpretation [5-6].

A characteristic feature of the data analysis methods Data Mining is the use of various algorithms for finding patterns in the data. Expansion of the set of data-mining models in various algorithmic nature can be productive in the class of problems where not accurately work classical methods: statistical, analytical or deterministic.

Each stage of the research data, we can build a finite number of hypotheses that can be confirmed or not be confirmed subsequently.

The more constructed models and descriptions are close to the hypotheses, the more we have the right to assume the accuracy of the result.

An important factor in the development of such systems is the consideration of its needs to equipment and data lines.

Due to this, the best solution is to develop a web-system which does not require installation and does not load the custom hardware does not use the high-speed data transmission, which simplifies its operation through VSAT technology, used as a standard marine satellite communication system to communicate with the Internet.

The process of searching for information online consists of the following steps:

− collection of information on the Internet from various sites;

− research on request;

− ranking results;

− indexing resources.

Despite the fact that every search engine has its own characteristics in search algorithms and ranking of the results, the principles of operation of all search engines in common.

Big data characterized by volume, variety and speed with which structured and unstructured data coming at networks in processors and storage, along with the processes to transform these data into information.

One of the architectures for processing data used in the application of search algorithms are the relational database management system (DBMS).

The use of these DBMSs yield the results the result of using a search algorithm with the system of constant circulation:

− waiver from strict consistency;

− care from the normalization and implementation of redundancy;

− the need for software simulation of functions of the SQL language;

− complexity of the client software;

− the difficulty of maintaining the health and resiliency of the resulting solution.

Disadvantages of using relational database systems has led to the development of adaptive-disciplined architecture, able to expand and scale as needed, with constant increase of data.

Such technologies include the use of NoSQL [9; 10].

Model data warehouse architecture NoSQL are divided into 4 categories depending on the conditions of use.

One of the most popular models used in cluster architecture is the storage of key-value. A distinctive feature is a simple data model, an associative array or dictionary, which allows to work with data by key.

The main objective of such storage – maximum capacity.

If the basis of the data used in the databases, the data are text entries that have a meaningful logical structure, use of document storage. The data model such storage allows you to combine a plurality of pairs key-value in an abstraction called «document». Documents can have nested structure, and to unite in collection. Work with documents produced by key enabling queries on attribute values.

If the database architecture is classic, mixed architecture, it uses columnar storage. Most similar to traditional relational DBMS. Data model and storage of this type involves storing the values as uninterpreted arrays of bytes, addressable tuples.

A special case when building database can be considered a repository on graphs. Such storage used with data that are naturally represented by graphs. The data model consists of vertices, edges, and properties. Working with data is done by graph traversal in the ribs with the desired properties.

Conclusion. Develop the concept of an algorithm processing a large volume of information generated during a fixed period, is an important issue. Solution to this problem may be to construct a universal algorithmic solutions able to analyze large amounts of information available in databases of different types.

The main factor of this method should be the processing speed and the accuracy of the forecasts resulting information. This should be considered dynamically-variable number of stored data in vessel databases, the coastal centres of information processing and intermediate cloud storage.

<div align="center">

REFERENCES

</div>

1. *Кончаков Е.И. Техническая диагностика судовых энергетических установок [Текст] / Е.И Кончаков. – Владивосток: ДВГ-ТУ, 2007. – 112 с.*
2. *Tom White. Hadoop: The Definitive Guide, 3rd Edition. O`Reilly Media, 2012. – 688 p.*
3. *Арский Ю.М. Принципы конструирования интеллектуальных систем / Ю.М. Арский, В.К. Финн // Информационные технологии и вычислительные системы. – № 4. – М., 2008. – С. 4-37.*

4. *Геловани В.А. Программное обеспечение системы моделирования / В.А. Геловани, С.И. Болоткин, В.Б. Бритков. − Препринт. – М.: ВНИИСИ, 1978. – 26 с.*

5. *Бритков В.Б. Интеллектуальные методы в решении задач использования океанологической информации / В.Б. Бритков, Н.А. Маленкова // Информационные технологии Государственного фонда данных о состоянии окружающей природной среды. Труды ВНИИГМИ-МЦД. − Вып. 170. − СПб.: Гидрометеоиздат. – Обнинск: ВНИИГМИ-МЦД, 2002. – С. 252-277.*

6. *Барсегян А.А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод // БХВ-Петербург, 2008. – 173 с.*