

WEB-BASED BIG DATA PROCESSING SYSTEM AND STRUCTURE DEVELOPMENT**Khasanova Madina Utkirbek qizi**

Student, Ferghana branch of TUIT named after Muhammad Al-Khorazmiy

<https://doi.org/10.5281/zenodo.6825142>

ABSTRACT: Over the past five years, there has been an active development of automated information systems, in particular Internet network technologies, methods of artificial intelligence development, methods of storing data and knowledge, etc. The main problem of these tools is the focus on solving particular problems, and they are also designed for a certain class of systems. However, the level of their automation opens up the possibility of creating technical solutions containing all the most advanced tools for big data mining based on cloud resources and WEB systems technologies, including the structure, functions and methods of organization.

The result of this article is the creation of the concept of an integrated Internet system of an intelligent knowledge repository with an automated process of information support for managerial decision-making. The main ways to create a software system designed to improve the results of processing big data arrays for such intellectual analysis tasks as classification, training, forecasting.

Keywords: artificial intelligence, web system, Internet, algorithms, databases and knowledge, big data, big data processing, systematization, automation, intelligent analysis.

INTRODUCTION

Automated information systems have recently been rapidly developing in artificial intelligence methods, Internet network technologies, methods of storing and presenting knowledge, programming languages and tools [3, p56].

In turn, the development of artificial intelligence has had a significant impact on both the creation of new and the transformation of old classes of information systems, such as [1, p. 29]:

- Decision support systems
- Expert systems



- Data mining systems
- Intelligent information systems.

Most of the modern tools are designed to solve various particular problems, or are originally adapted to a specific class of systems. At the same time, with the existing level of automation of tools, it is possible to develop a “super-system” that would integrate the most advanced tools (in particular: technologies, algorithms, models, methods, approaches) through a kind of intelligent knowledge repository with an automated process of information support for the adoption of managerial solutions. Internet technologies based on data mining methods have been developed due to the need to increase the efficiency of using the collected information in electronic form (knowledge bases, repositories, data banks, etc.) by unifying and integrating storage formats and processing procedures [2, p. 204].

Description of the concept of an integrated Internet system

The concept of developing an integrated Internet system for an intelligent knowledge repository with an automated process of information support for managerial decision-making is becoming more and more popular and relevant. To identify the main ideas, technologies and recommendations necessary for the development of such a system, we will focus on the Big Data Intelligent Processing System (BDIPS) - the hardware and software component.

The characteristic features of the BDIPS influence both the technology and the methodology of its development. And since the technology for creating BDIPS has specific features, it differs significantly from the process of designing and developing other software and information systems. To a large extent, such differences are due to the fact that BDIPS is an intelligent information system (IIS), which is based on concepts, methods and artificial intelligence.

Data mining platform

The created platform can be classified as AaaS - Analysis as a Service or SaaS Business Intelligence (BI). The system assumes the presence of a client extension of functionality, as well as some platform-as-a-service (PaaS) features. Capabilities of the created system (SaaS BI):



- Visualization of analysis results.
- Import/export of initial data and learning outcomes.
- Tools for analyzing the effectiveness of training based on data.
- Means for constructing an automatic collective solution based on algorithms.
- A library of ready-made analysis subsystems and algorithms, as well as: clustering and classification, neural networks, building rules and decision trees, statistical and genetic algorithms, etc.
- How the workspace is provided as a service (authentication and authorization, tools for uploading and editing files).

Strategic development plan. BDIPS is a software and hardware system that is configured to use different classes of recognition, formats and languages for representing knowledge and data, an apparatus for analyzing and synthesizing model representations, including those aimed at providing services for solving applied and scientific problems of processing any data

Strategic plans include the possibility of expanding user functions for creating and editing individual sections of the library of algorithms (PaaS), as well as:

- Automatic data entry into BDIPS (working with the server via API).
- Integration with business processes of customers/users (upon completion of training or analysis, the server will automatically send specially configured data to the customer's server, resulting in automatic output of data to the customer's system).
- Ability to build scenarios using analysis algorithms.
- Possibility to publish algorithms in public access.
- Metalanguage-based editor for creating new algorithms. BDIPS can be described as follows:



- In terms of supported functions, the BDIPS will be extensible, which means that it can be adapted to different classes of tasks. BDIPS will include programming interfaces (APIs) and tools that will allow not only adding new modules and functions, but making them available to other users.

- The Internet will provide access to the system. Most of the system functions will be carried out through the browser web interface. If it is necessary to work with confidential information, the system will provide encryption of stored and transmitted information (access can be carried out using the HTTPS protocol).

- Only registered users will be able to access most of the resources and services of the system. However, some of the system resources will be open for mass access and registration will not be required to use them. For example, well-known information and reference materials can act as open resources.

- In order to gain access to the entire set of resources and services of the system, you will need to register. The status of the user will determine both the set of services and resources provided by the system, and the corresponding quantitative and qualitative characteristics, including various restrictions. The characteristics of resources and services, depending on the tasks and requirements of the user, can change.

- The concept of “projects” will be implemented in the system. The idea is that in order to solve their problems, the user needs, through the system software, to create the required number of projects. For each project, the user will have the opportunity to rent and use the resources and services that he needs to solve a particular problem. In this case, if desired, the user will be able to provide access to his project to other users or their groups. The set of rights will determine the capabilities and allowed operations when working on this project.

- Third-party applications and websites through the programming interface (Web API) will be able to interact with the system via the HTTPS protocol and, according to certain procedures, use its resources.



- For users who need a high level of security and/or for other reasons cannot or do not want to store and transfer their data in the system, they have the opportunity to obtain client software and install it on their computers. This software will allow you to perform the main functions of BDIPS and perform data processing and analysis locally, offline (without connecting to the Internet).

The main functions of the BDIPS. BDIPS is a set of tools designed to solve problems:

- extraction of knowledge from data (data mining) and texts (text mining);
- forecasting (determination of trends in the development of processes);
- identification (determination of the distinguishing features of the studied objects);
- clustering (classification without training);
- pattern recognition (classification with learning);
- statistical data processing (covariance and variance analysis, regression and correlation analysis, etc.). Since the system has an open architecture and API, the set of supported task classes can be easily supplemented, including modules developed by third-party developers.

SIODB may have the ability to support the import of data from various sources, such as:

- relational databases;
- spreadsheets in OpenDocument and Excel format;
- CSV, XML and HTML files;
- Web applications and services, such as Google Spreadsheets, MyOffice, Airtable, Yandex Documents.

In addition, directly through the system interface, data can be entered, edited and changed. Data and results can be exported to files of various formats, for example: XML, CSV, PDF, HTML,



Excel, RTF, PNG, JPEG. It will be possible to integrate other import / export converters into the system, which will provide work with non-standard data sources and their formats.

The system will have the necessary tools that provide opportunities for graphical representation and visualization of the initial data and the results of their processing in various forms, including various charts and graphs. The indicated tools can also be used to create certain reports that can be published within the system or exported for further use.

BDIPS will contain various reference books, interactive tutorials and test modules on the subject of intelligent data processing, which are intended both to teach users how to effectively solve problems using the system, and to increase their level of knowledge about data processing methods and related models. In addition, using the system, it will be possible to organize e-learning courses and conduct automatic knowledge testing for different subject areas.

BDIPS automatically creates a personal account for each user. Projects that operate within the system are created in the personal account; other users of the system can also apply to this project. The created projects will allow aggregating and publishing various kinds of materials that do not contradict the system rules, and, if necessary, the user will be able to restrict access to them.

In order for users to communicate with each other and receive feedback from them, the system will support communication services. This includes the exchange of electronic messages (for example, real-time chat) and mailing lists, subscription to inform about changes in the system (for example, changes on the sites of any users) and thematic discussions and forums, voting and polls.

Control system. The site management system will be a software package that will automate the process of managing both the site as a whole and entities within the site, such as data output templates, page layouts, content, structure, access rights and users. In particular, additional services will be provided: statistics, mailing lists, means of interaction with users, search, etc.

The considered design solutions for the creation of BDIPS together with AI methods, data mining technologies, electronic databases in various subject areas, information presentation



models are a tool to improve the efficiency of educational, scientific and innovative activities in various fields.

CONCLUSIONS

The implementation of this project in practice will make it possible to use it as an adaptive, self-adjusting, open intelligent information system with built-in expert system functions and subsystems for data mining (knowledge extraction). The proposed mathematical apparatus and the methodology for its application in the process of data analysis is a tool for scientific evaluation and formal justification of decisions made in various areas of classification, forecasting, and decision making.

REFERENCES:

1. Bozhenyuk A.V., Kotov E.M., Tselykh A.A. Intelligent internet technologies. M: Phoenix, 2009. ss. 204-219. ISBN 978-5-222-15978-1
2. Agrawal A., Hans D., Goldfarb A. Artificial intelligence in the service of business. How machine forecasting helps to make decisions. M: Mann, Ivanov and Ferber, 2019. pp. 29-35. ISBN 978-5-00117-881-1
3. Norenkov I.P. Automated information systems: textbook. allowance. M.: Publishing house of MSTU im. N.E. Bauman, 2011. pp.56-58. ISBN 978-5-7038-3446-