© University Press 2021
ISSN 2719-6550
ISSN 2719-7417 online

"Journal of Education, Technology and Computer Science"
No. 2(32)/2021
www.eti.rzeszow.pl

Received: 27.08.2021
Accepted for printing: 25.11.2021
Published: 28.12.2021
License: CC BY-SA 4.0

DOI: 10.15584/jetacomps.2021.2.15

EVGENIY E. KOVALEV

Application of Digital Data Analysis Technology in Education

ORCID: 0000-0002-3015-5084, Moscow State Pedagogical University (MPGU), Vice-Director of Institute of Mathematics and Informatics, Russia

Abstract

The main directions of digitalization of the education system are considered, the infrastructural and technological aspect of new strategic directions of development is analyzed. The author analyzed information systems and services in education, taking into account the possibility of using technologies for analytical data processing. The problem of using data analytics tools in education at present is the lack of uniform formats and tools for integrating information systems. Data in education accumulates in various solutions that do not share the results of the educational process. Based on this, it is impossible to build a clear system of end-to-end interaction between systems on common platforms. The author aims to solve this problem by modeling an integrated solution for data analysis at all levels of education. The article proposes technical and organizational solutions for the creation of integrated analytics services that allow accumulating and transmitting data on educational results during the transition of a student between learning levels. The analysis tools are based on the use of big data technologies.

Keywords: information systems and services in education, integration of digital resources and services, analysis of big data in education, interoperability of digital systems and data portability

Introduction

The development of the fourth industrial revolution entails not only advanced scientific and technical developments, but also a qualitative change in the culture of work (Osburg, 2016). In this regard, those employed in all spheres of the economy, regardless of the skill level, are required to:

- high level of mathematical, financial and digital literacy;
- soft skills;

– knowledge, skills and abilities in the field of formed thinking – algorithmic design, critical thinking.

Studies (Elliott, 2018) show that educational reforms carried out in recent decades do not always meet modern technological challenges.

In the process of digital transformation of education, it is necessary to form and implement new models of educational organizations, which are based on the development of digital tools, sources of information and services aimed at improving the organizational and infrastructural conditions for the implementation of the necessary transformations and the adoption of managerial organizational and pedagogical decisions. A flexible response of the education system to challenges is also necessary to transform the system based on the analysis of data on the state of society.

Materials and methods

The current state of information systems and services in education is characterized by the presence of a large number of discrete software products, each of which accumulates digital traces of the learning process and generates output data in its own, sometimes unregulated, formats. This makes it impossible to perceive the state of the education system as a whole and makes it difficult to exchange data between systems. Also, some of these systems do not generate data in formats suitable for reuse and use by third-party systems.

The study showed serious shortcomings of existing solutions in terms of a systematicapproach and the possibility of using data analytics:

– fragmented platforms;
– proprietary data presentation formats;
– impossibility of full reuse of data;
– lack of integration and the ability to exchange data between platforms without pre- processing and adaptation;
– lack of logical connections between assessment criteria at different levels of education;
– poor data visualization;
– poor opportunity for teamwork, crowdsourcing and replication of the results.

As a result, the education system at various levels weakly interacts both among themselves and with participants in the educational process. This leads to the impossibility of building in a single format a general picture of the state of the education system and the implementationof continuity between its levels.

Results of the research

The main obstacle standing in the way of creating a unified information and analytical space is the lack of the ability to transmit in electronic form and in

established formats information about learning outcomes in relation to a specific student, as well as to provide a reliable assessment of the transmitted learning outcomes. This does not allow creating a single technological platform for storing and exchanging information and introducing a completely electronic document flow between participants in educational relations. The situation is aggravated by the processes of mergers and acquisitions, leading to the inheritance of new information systems and applications, an extremely heterogeneous IT landscape containing applications and software components from different manufacturers, which are implemented on different platforms and often duplicate separate functions.

To achieve a single information and analytical space, it is necessary to integrate and interoperate data, ensure the availability of data, while information systems must interact with each other in the same language. A prerequisite for this is uniform rules for data interpretation and a single data ontology (information exchange model), taking into account the industry specificity of education. which will unify data management technologies. All information services and systems must connect to the data management and analysis infrastructure and exchange data according to uniform established rules. The data management infrastructure may not be designed to store the data itself; in this case, it performs technical and technological functions, storing only information about the data: their description (passports), data registers, accounting for data use, transfer rules and data quality control. In part, these functions should be performed by modernized information systems of the e-government infrastructure or departmental management information systems.

The basic algorithm for data integration and processing assumes:

1. Extraction of structured data (conforms to a data model, has a well--defined structure, follows a sequential order and can be easily accessed and used by a human or a computer program). Extraction of information based on ontologies, a terminological dictionary of synonyms/relationships

2. Cleansing of unstructured data (does not have a predefined data structure, or is not organized in an established order. Unstructured data is usually presented in the form of text, which can contain data such as dates, numbers and facts. This leads to difficulties in analysis, especially in the case of using traditional programs designed to work with structured data), the selection and removal of “noise”, the transformation of the maximum possible types of unstructured data, the selection of data suitable for analytics (Text files and documents. Photos, drawings and other graphic information. Biometric data).

3. Obtaining data in a machine-readable format that allows information systems to identify, process, transform such data and their constituent parts (elements) without human intervention, as well as provide ranked access to them for system users, including public access.

4. Data validation. Formation of a reliable assessment of the transferred learning outcomes (trust data register). Formation of metadata that makes it easier to extract the necessary data for analysis.

5. Selection of related data, which can store semantic queries and show data that affects the selection.

6. Obtaining analytical data.

7. Application of the Application Programming Interface (API).

8. Application of criteria for evaluating analytical data.

9. Formation of analytical data in formats suitable for the consumer and decision making, suitable for reuse, for accumulation in databases.

10. Uploading data in formats for exchange between systems, visualized data, generation of reports in established forms to support the electronic document management system.

Pre-design work involves taking into account the following factors:

a) study of the potential demand for relevant datasets by potential consumers;

b) an assessment of the degree of readiness, characterized by the availability of the necessary data in electronic form, as well as the readiness of the organizational, technical, technological and other means necessary for the publication of data sets;

c) publishing costs (financial, time, labor) required to publish datasets and keep them up to date. As a proposed solution, it is necessary to develop a single portal with entry points for system participants at various levels of education to download and exchange data and obtain statistical information. After the accumulation and purification of data, it is possible to identify inter-component groups of indicators and criteria for assessing education, which can be transferred and adapted between the levels of the education system. With the further development and refinement of the mechanisms of interaction between participants in educational relations and relations in the field of education, it will become possible to build a quality management system based on Deming principles adapted to assess the education system and continuous quality standards (TQM) (Kovalev, Kovaleva, 2018).

In the light of the indicated solutions for the modernization of existing systems, it seems necessary to use the technology of accumulation, processing and analysis of big data (big data). They allow you to: process large amounts of data in comparison with “standard” scenarios, often in different formats, work with quickly arriving data having a fast update time in very large volumes. Moreover, such data is experiencing constant growth, being able to work with structured and loosely structured data in parallel and in different aspects (Patil, 2011).

In addition, the relevance of the use of big data is confirmed by the fact that educational policy begins to be built on educational analytics, on new analytical and managerial methods (Fiofanova, 2020).

The main methods of big data analysis and their possible use cases in education management are given in Table 1 (EMC, 2015):

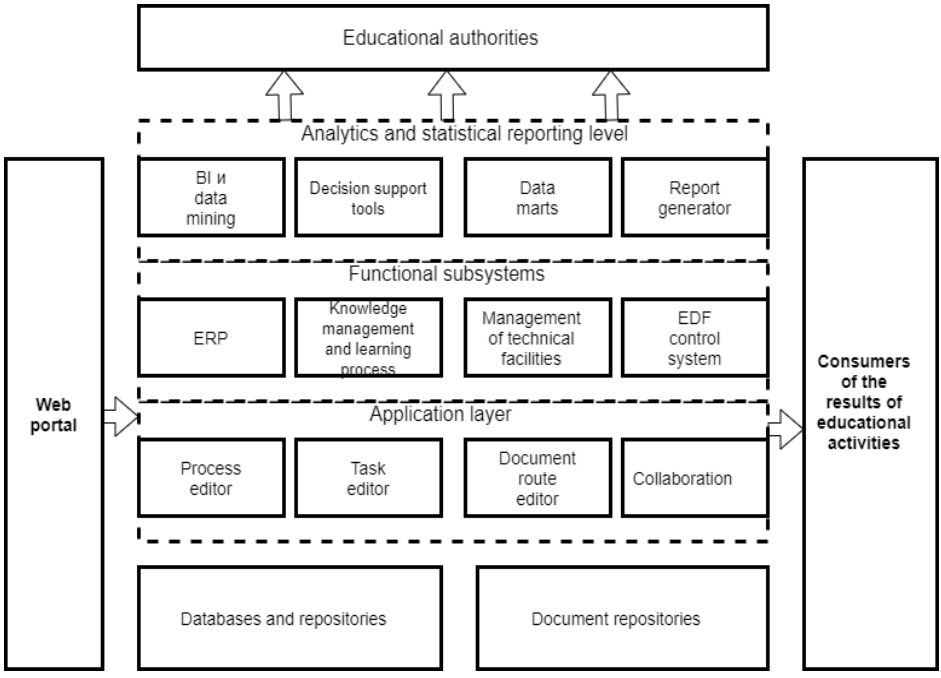
Table 1. Big data analysis methods and their possible use cases in education management

Method, its description	Possibility of application in the education management system
Data Mining (data mining, data mining, data mining).	The analysis of educational outcomes and socio--economic information in the field of education for clustering and classification, identification of rules and parameters dependencies, regression analysis andthe detection and analysis of variance for the adoption of organizational and educational decisions
Mixing and integration of data (data fusion andintegration)	Integration of heterogeneous data from a variety of sources in order to conduct in-depth analysis (for example, data streams from social networks, educational networks, digital campuses, digital natural language processing, including tone analysis, etc.)
Machine learning, including learning with andwithout a teacher	Statistical analysis or machine learning to obtain complex predictions of the main criteria for the development of education
Predictive analytics	Calculation of the student's life value, an indicator that helps to understand the value will be brought by the student throughout the life cycle (including future admissions), in particular in the management of educational services. Development of optimal offers / recommendations based on the data obtained on user behavior on the site. Formation of an analytical forecast of tasks, which the student with a certain degree of probability will perform in the future. Forecasting the outflow of trainees. Building a trainee profile based on a sample of data
Visualization of analytical data, presentation of information in the form of figures, diagrams, using interactive features and animations both for obtaining results, and for use as initial data for further analysis.	Creating dashboards and visual indicators that prompt in real time the state of indicators and theirdevelopment trends.

The system model of a single information resource based on the integration of information services of various levels of education and the proposed algorithm in Pictures 1.

Technically, the application integration approach is based on the use of middleware class software and open systems methodology. Modern systems of the middleware class are able to process messages based on universal formats and provide multi-channel message transfer between all application components. The essence of the methodology of open systems is that when they are built,

docking should be provided using standard interfaces between all components of the systems. The preferred solution for managing and creating specifications for the protocol of interaction of multi-format systems is the use of XML technology.



Pictures 1. The system model

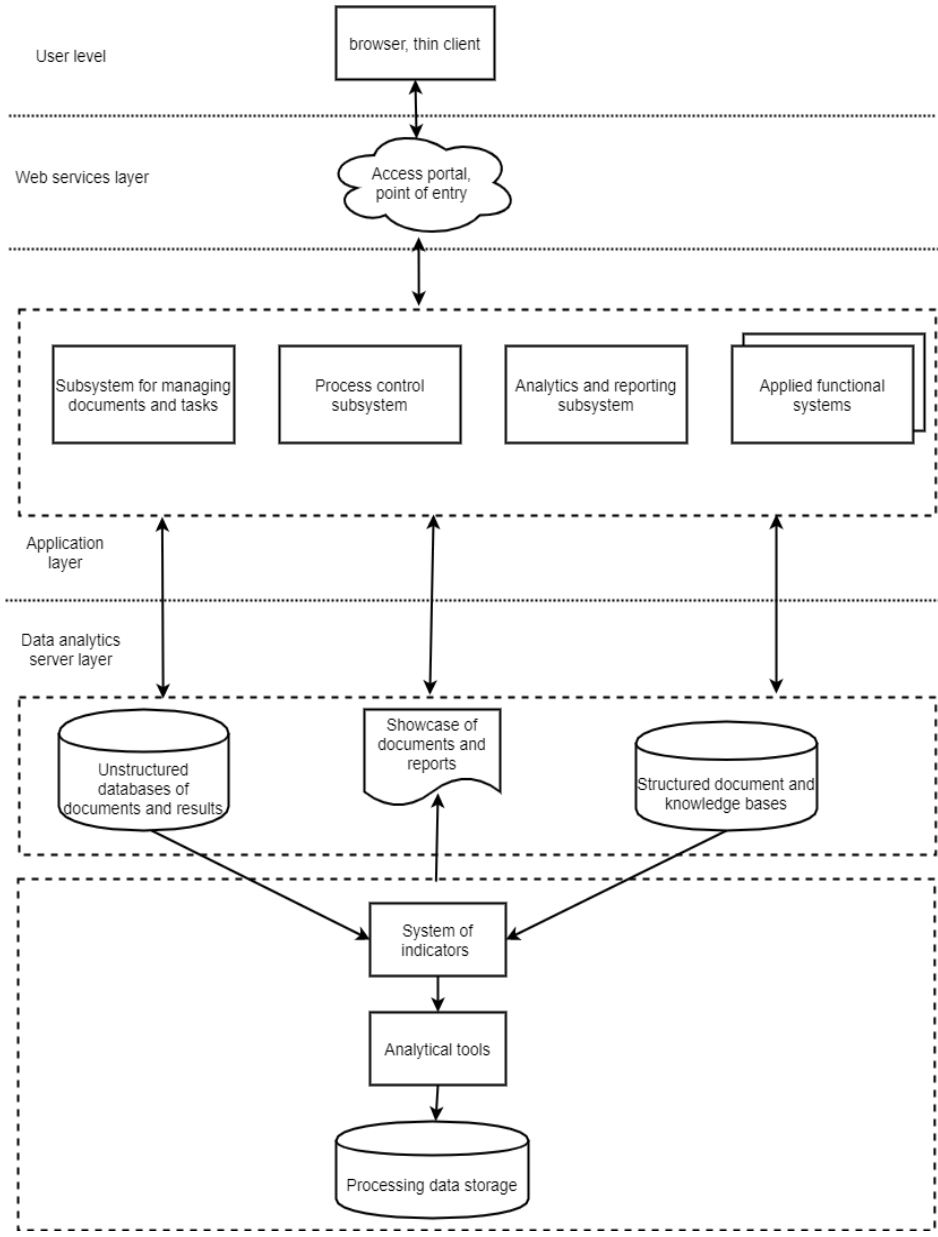
The main components of systems of this class are:

– an integration broker playing the role of a service bus (performs the functions of data reformatting, message routing, transaction management, monitoring and control of application interactions);

– a set of adapters that allow various applications to connect to an integration broker.

The integration costs of such a solution can be minimized by the use of the open systems methodology, which implies the allocation of an interface part in the system that provides communication with other systems or subsystems. To combine systems, it is enough to have information only about the interface parts of the mating objects, made in accordance with certain standards.

The ecosystem of integration of information services in education based on the analysis of big data is shown, and the levels of integration of solutions when implementing the proposed approach are shown in Pictures 2.



Pictures 2. The ecosystem of integration of information services in education based on the analysis of big data

Discussion of the research results

The following cases can be examples of using the obtained analytical results:

Analysis of student performance data. The system, using regression analysis of data, allows to identify students who are in the so-called "risk group" (missed classes or showed low and unsatisfactory results), to predict visits and future successes, trends in learning outcomes. The system aggregates all the grades and achievements of students, finds problems, including potential ones. It will also be possible to analyze the participation of each student in a variety of activities conducted by the educational institution and evaluate its electronic portfolio. This parameter can be considered key in terms of academic success. The school, therefore, monitors the frequency of attending various events using virtual identification cards: if the student's involvement decreases, the school staff identifies the reason and can offer ways to solve the problem and increase the level of student involvement, and for those who show good results – recommend as potential applicants and recommend to participate in olympiads, contests. The entry point can be a student's virtual card, which collects data on the location, time spent studying, participating in other educational and scientific events, is a single ID for entering the library and social services of an educational nature.

Personalization of training. One popular learning personalization strategy is to offer an additional online course to a lagging student. As the student answers questions, the analytic system will be able to predict his readiness for new topics and analyze the time spent on topics studied and necessary for new topics.

Predictive modeling. Educational institutions can form a model of potential applicants with respect to the selected parameters (such as, for example, academic performance in a number of subjects, final works, grades, availability of certificates and documents on participation in various events). With such a selection of the target group, it is possible to build targeted interaction and point offers for applicants and/or graduates, for example, when interacting with a bank of job vacancies.

Conclusion

The conducted research on data analysis in education can become the basis for the formation of documents on the management of analytical systems and data in education. They should describe the general target data management, determine the procedures for the formation of data analytics. The requirements for managing specific data sets and formats should be a set of techniques used for each specific task. The methods should be refined and refined constantly, in the framework of the implementation of the principle of flexible approach and continuous quality improvement. The development of methods should correspond to the current needs of the information and analytical system and should be accompanied by training of personnel to work with the information system at all levels.

References

Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data (2015). EMC Education Services.

Elliott, S.W (2017). *Computers and the Future of Skill Demand*. OECD. Retrieved from: http://http://www.oecd.org/edu/computers-and-the-future-of-skilldemand-9789264284395--en.htm (12.08.2021).

Fiofanova, O.A. (2020). Analysis of the current state of research in the field of education management based on data. *Values and Meanings*, *1*(65), 71–83.

Kovalev, E.E., Kovaleva, N.A. (2018). Implementation of Models for Assessing Professional Competencies Using ICT Tools. *Edukacja – Technika – Informatyka*, *4*(26), 276–282.

Osburg, T. (2015). *Industry 4.0 Needs Education 4.0*. Retrieved from: http://www.linkedin.com/pulse/industry-40-needs-education-thomas- osburg+&cd=1&hl=ru&ct=clnk&gl=en (9.08.2021).

Patil, D.J. (2011). *Building Data Science Teams. O'Reilly*. Retrieved from: http://cdn.oreilly.com/radar/2011/09/Building-DataScience-Teams.pdf (3.08.2021).