Sofia University „St. Kliment Ohridski"

Faculty of Mathematics and Informatics
Information Technologies Department

Big Data Visualization

Dafinka Savova Miteva

ABSTRACT

of the PhD Dissertation

for the purpose of acquiring the educational and science degree
"Doctor" in the professional field of 4.6 Informatics and Computer Sciences,
Doctoral Program "Information Technologies
(Information and Communication Technologies)

Doctoral supervisor:
Prof. Eliza Petrova Stefanova, PhD

Sofia, 2023

Content

Note: In this abstract, where it is not indicated, all references to pages, chapters, appendices, cited literature, tables, and figures refer to the full text of the dissertation.

Acknowledgement

Confucius said that he who knows all the answers has not been asked all the questions.

My supervisor, Prof. Dr. Elisa Stefanova, taught me to constantly ask questions and seek their answers. To be creative and to pursue my goal to the end. I am grateful for her excellent professionalism, personal attitude, invaluable advices, always given at the right moment.

To Prof. Dr. Krasen Stefanov for the boundless energy with which he motivated me and for the interesting research projects through which he gave me the chance to present the results of this thesis at national and international conferences. Thank him for his wise guidance and the courage to finish this work.

To Prof. Dr. Slavi Stoyanov, who inspired me for one of the studies in this dissertation and helped me to realize it in a professional environment.

To Dr. Elitsa Peltekova, who with her youthful enthusiasm constantly encourages me not to stop fighting for everything I have chosen to achieve.

To the experts who shared their valuable experience with me - Prof. Dr. Elisa Stefanova, Prof. Dr. Maya Stoyanova, Prof. Dr. Krasen Stefanov, Prof. Dr. Pavel Boychev, Assoc. Dr. Mariana Atanasova, Assoc. Prof. Dr. Nikolina Nikolova, Assoc. Prof. Dr. Alexander Dimov, Assoc. Prof. Dr. Lyubka Alexieva, Asst. Prof. Irena Avdjieva, Asst. Maya Zhelyazkova, as. Albena Antonova, senior lecturer Pencho Mihnev, Diana Pencheva, Ioanna Atanasova, Dimitrinka Tomova, Svetlana Dimitrova and all the students who welcomed the challenge with excitement.

To the experts who, hidden behind anonymous participation, dared to be more critical and honest with me.

To all the teachers with whom we worked together in the studies and experiments.

To the Department of Information Technology at FMI and friends who make me feel like part of a great team!

To Prof. Dr. Pavel Boychev for editorial advice and constructive feedback.

To my family who respected my desire to write this work and supported me.

To all of you my deepest thanks!

Introduction

Actuality of the Problem

In today's world of information technology and big data, we can paraphrase the famous idiom "cannot see the forest for the trees" into "cannot see the results for the big data". Analysing large volumes of data brings with it great potentials, but in order to unleash them, users must first of all be able to interpret the results, extract useful information from them and use it. And in order to create a bridge between big data and decision makers, this data needs to 'speak' quickly and clearly, it needs to be visualised.

This dissertation explores how big data visualization can address challenges in education.

The focus of traditional education is the average learner, which increases the risk that at some point the choice of learning materials, teaching methods, assessment of knowledge to be designed for a 'non-existent' learner. Modern education therefore faces a number of challenges and issues: Can learning be customized? Can each learner follow his/her own individual curriculum adapted to his/her starting level, pace of learning, progress and specific needs? Can the learner's outcomes be predicted and efforts directed where he/she is having difficulty? Can the learner's interest be provoked and kept by responding to personal needs? How can learning be more effective and useful? How to evaluate in what direction to develop/improve teaching materials?

These questions can be answered by learning analytics (Learning Analytics, LA), formally defined in 2011 as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs" [1].

According to Siemens (Siemens et al., 2011) the use of LA could (1) reduce the number of dropouts by identifying at-risk students early and notifying teachers, (2) customize the learning process and content so that each learner receives resources according to his/her current knowledge, (3) notify teachers in a timely manner which learners need additional support and which teaching practices have the greatest impact, (4) contribute to the design of high-quality learning, (5) create interactive visualizations of complex information that will help find answers quickly, etc.

Modern learning is served by Learning Management Systems (LMS), which in addition to learning materials provide a technological apparatus for collecting and storing information about the user's activity. Within an educational institution, more than one learning systems can be used to serve learning process, specific courses, work on projects and tasks. Sometimes learning needs the use of specific tools and environment covering its particular requirements. Often, entire courses are conducted on platforms outside the administrative control of the primary educational institution. In order to provide a

---

[1] 1st International Conference Learning Analytics and Knowledge,
http://www.wikicfp.com/cfp/servlet/event.showcfp?eventid=11606

meaningful answer to the questions posed at the beginning, learning analytics from all systems supporting learning must be collected and analysed as a whole.

Object

The object of the presented research are learning analytics and their visualizations used in LMS.

Subject

The subject of the research are the types of reports that a LMS can provide in the learning analytics section and data visualization methods used for easier and proper data understanding.

Objective of the Dissertation

The purpose of the research is to design a learning analytics system in which, through modern visualization methods, ways to increase the effectiveness of e-learning can be outlined and to support learners, teachers and managers of educational institutions in making the right decisions at the right time.

Tasks of the Dissertation

In order to achieve the goal, set in the dissertation, the following tasks were completed:

Task 1. To investigate and analyze the challenges to the creation of a system for visualizing learning analytics, in the following aspects:

1.1.   Metrics and key performance indicators, learner data collection methods, models for learning analytics cycle, compliance with ethical norms and personal data protection;

1.2.   Modern methods for data visualization and big data visualization;

1.3.   Modern software platforms for creating a virtual dashboard.

Task 2. To study and analyze how methods and tools for LA can optimize the learning process and increase the success rate of learners.

Task 3. To explore and analyze the functionalities of the virtual dashboard as a tool of visualizing learning analytics in Learning Management Systems (LMS).

Task 4. To study and classify the expectations of LMS users regarding:

4.1.   the reports they expect to find and use in LA section;

4.2.   the report visualizations that users (learners, teachers and managers) find most intuitive and easy to read.

Task 5. To create a model of a learning analytics visualization system.

Task 6. To create a prototype of the learning analytics visualization system.

Task 7. To test and analyze the created prototype.

Structure and Content of the Dissertation

The dissertation consists of an introduction, seven chapters, a conclusion, an author's contributions, bibliography, and six appendices.

The introduction presents the relevance of the problem, defines the object, subject and purpose of the dissertation, outlines the tasks and structure.

Chapter 1 provides a literature review of learning analytics and its use in education. The metrics and key performance indicators are outlined. Data collection methods in LMS are explored. Ethical standards and data protection requirements are commented. An overview of current methods and tools for visualizing learning analytics is provided. Modern platforms for creating a virtual dashboard are analysed.

Chapter 2 describes experiments conducted as part of the PhD thesis, examining how the application of different LA can optimize the learning process by assisting lecturers and managers of learning institutions, and therefore the learners themselves.

Chapter 3 describes experiments approbating different visualization methods created and used in different research projects within the PhD thesis.

Chapter 4 presents the requirements for creating an integrated learning analytics system. Two studies conducted as part of this dissertation are described. The first study presents expert opinion on the use of big data to benefit learning by creating LA reports. The second study reveals the experts' vision of which data visualizations allow these large amounts of data to be read and perceived most quickly and accurately.

Chapter 5 proposed an architectural model of a visualization system for learning analytics. Its aim is to meet the current requirements for creating an integrated learning analytics system to support e-learning effectiveness through better visualization.

Chapter 6 describes the prototyping of the model. The choice of a virtual dashboard implementation tool is motivated. Data and metrics to be visualized in the prototype are described. The creation of different types of virtual dashboard is presented.

Chapter 7 describes the experimental validation of the model and prototype. Two testing methods are applied: functional testing and "Think aloud" protocol. The methodology of anonymizing the real data is described and the results obtained are illustrated.

The **Conclusion** presents a summary of the results and directions for future development.

The **Contributions** presents the scientific contributions, scientific and applied contributions, scientific publications and research projects related to the dissertation.

A **bibliography** is a list of used sources.

Appendix 1. Invitation letter to participate in a brainstorming activity to generate ideas on what kind of reports the modern user expects to find in the LMS.

Appendix 2. Invitation letter for participation in the activities "Grouping" and "Sorting" of the proposed reports for LMS.

Appendix 3: Social questions for participants in the Sorting and Grading task.

Appendix 4: All suggestions for LA reports proposed by the experts during the group concept map activity.

Appendix 5: The group concept map processing to select the final number of clusters.

Appendix 6: Reports by categories rated above average on the two criteria of "usefulness" and "relevance".

Chapter 1. Short Overview of the PhD Thesis

This chapter describes the work on Task 1, by examining and analyzing the challenges to creating a system for learning analytics visualizatins.

1.1. Learning Analytics in Education

This section describes the study and analysis of metrics and key performalce indicators measuring learner progress, methods for collecting learner data, learning analytics application models, ethical compliance, and privacy protection.

1.1.1. The Evolution of LA and their impact on learning

In 2010, George Siemens introduced learning analytics as " the use of intelligent data, learner-produced data, and analysis models to discover information and social connections, and to predict and advise on learning" (Siemens, 2010). One year later, during the 1st international Learning Analytics and Knowledge (LAK) conference, it was decided that learning analytics would focus on integrating technological and socio-pedagogical dimensions and the current definition was given, published on the conference website.

The rapid development of information technology in the 21st century has changed the units and scales for measuring information as well as the methods for processing data. Starting with bytes and kilobytes five decades ago, today we talk about exabytes (EB, 1018), zettabytes (ZB, 1021) and even yottabytes (YB, 1024) of data. Large data sets are being collected that "exceeds the reach of commonly used hardware environments and software tools to capture, manage, and process it with in a tolerable elapsed time for its user population"[2].

In 2012, Gartner gave a definition of big data, also known as the three "V's": volume (how big the data is), velocity (how fast that data is processed), variety (how various it is in type). To this definition, some authors add more "V "s describing the big data model from a business perspective (Mediratta, 2015): *variability* – takes into account whether the data flow is regular and reliable, veracity - checks whether the data is trustworthy before any analysis is performed, and *value* - identifies the purpose for processing the data, the analysis must follow ethical norms, protect the reputation of users. Wang (R. Wang, 2012) adds *viscosity*, which measures the "resistance" of data when it arrives from different sources and at different speeds, and *virality*, which describes how quickly this information spreads among users in peer-to-peer networks

The processing of data generated in learning environments is gradually becoming a real challenge, necessitating the use of big data technologies and tools for their processing. In a review of state-of-the-art data mining and learning analytics tools in higher education, the techniques and tools for big data processing that can be used in various ways in learning analytics are outlined (Salihoun, 2020).

---

[2] Big Data Definition, Teradata, https://www.teradata.com/Glossary/What-is-Big-Data

1.1.2. Metrics and key performance indicators

Learning analysts use metrics and key performance indicators (KPI)[3], to measure and analyse learner progress.

In a systematic literature review (Schwendimann et al., 2017) the authors systematized over 200 indicators that are used in modern learning systems. They grouped them into 6 categories: (1) Action-related - measuring number of pages visited; number of files downloaded; time logged in and time spent in the system; (2) Learner-related – consider previous training and courses, competencies gained, age; (3) Content-related – monitoring attitude towards the content of the learning material - topics completed and skipped, sentiment captured from forum messages; (4) Result-related – calculating average grade, distribution of outcomes across the group; (5) Context-related - in what context the learning takes place, where it takes place; and (6) Social related - communications between group members.

According to (Hout, N., 2020) it can be summarized that the critical indicators in most modern EMSs boil down to: (1) course participation data, (2) number of learners who successfully completed and number of dropouts, (3) details of learner activity, (4) most frequently visited learning materials, (5) performance data, (6) time spent reviewing learning materials, (7) best and worst performing learners, (8) looking for patterns in learner behaviour, (9) attempts to complete assignments, and (10) where learners are physically located.

1.1.3. Methods for Data Collection

Learning analytics involves the use of a wide range of data analysis methods, combining knowledge in different fields, such as statistics, psychology, computer science, and education (Khalil & Ebner, 2016). These methods are reviewed and described in detail in a number of articles. In general, they can be grouped into two categories: (1) clickstream data - analysis of the data generated by users clicking on mouse buttons while browsing web pages () and Natural Language Processing (NLP).

1.1.4. Models for Applying Learning Analytics

The big data that are collected in e-learning systems need to be managed so that useful information can be extracted from them to support decision making.

In the full version of PhD dissertation, the following models for implementing learning analytics are reviewed: (1) Clow's lifecycle model (Clow, 2012) for implementing learning analytics; (2) Elias' model (Elias, 2011) for implementing learning analytics; (3) Chatti's reference model (Chatti et al., 2012) for implementing learning analytics; (4) Freitas' model (de Freitas et al., 2015) and Simens' model (Siemens, 2013).

1.1.5. Ethics and data privacy

In parallel with the growth of technological capabilities to analyse and extract information from learner data, the ethical question of 'should we' since 'we can' is also emerging. There are a lot of debates about the ethics of working with big data and artificial intelligence, concerning privacy. The potential risks also apply to education.

---

[3] Key Performance Indicators, https://www.kpi.org/kpi-basics

The storage of data in cloud and the use of advanced processing mechanisms in a networked environment required the creation of a common legal instrument for data protection, which enacted on 25 May 2018 for the European Union and is known as the General Data Protection Regulation (GDPR). It regulates the processing of personal data, distinguishes between public and private entities and operations, and provides for high penalties for violations. Ethically, the collection and processing of big data must strike a balance between the scope of the data, the utility, and the risks of granting access (Florea & Florea, 2020).

1.2. Data visualization

The purpose of this analysis is, at a later stage, to support the selection of visualizations for the designed system for learning analytics.

In a systematic literature review for the period 2000 – 2017, published in specialised visualization and learning analytics journals (Vieira et al., 2018), the authors conclude that (1) there are relatively few tools for visual analysis of learning analytics, (2) for visualization of learning analytics still mostly uses traditional statistical data representation techniques such as bar and dot plots, and (3) there is a lack of studies that go deeper into educational theories while applying more sophisticated advanced visualizations.

1.2.1. The power of data visualization

In a number of scientific studies, psychologists have shown experimentally that the human brain perceives visual information more easily and retains it more permanently than verbal information (Whitehouse et al., 2006).

The use of visualization challenges the "software" built into our brains to discover patterns, correlations, and trends that would be difficult to spot in traditional texts and tables (Iliinsky & Steele, 2011). It can reveal new challenges and inspire deeper analysis.

1.2.2. Data visualization methods

In 2010, Abela described a 10-step algorithm for visually influencing the user (Abela, 2010). It proposes a hierarchical flowchart-like diagram that, through guiding questions and answers related to the type of data available, assists the user to select an appropriate visualisation. At the first level, data are classified into 4 main categories: comparison, composition, distribution or relationship. At each subsequent level, the corresponding category is divided into sub-categories.

In 2015 Edo Van Dijk[4] added two more principles to Abela's model - deviation and trend. The purpose of the deviation is to show which values deviate from the norm through column, bar, linear and area plots, while the trend shows the expected development of some variables through linear, bar, point or area plots.

Methods for visualizing various data types

This section describes sample data visualization methods that can support the presentation of learning analytics (Митева & Стефанова, 2018).

---

[4] Which chart type works best for your data?, Blog post, http://www.chartlr.com, last access: 21.05.2022

For example, column and bar, circle and ring, point and area plots are used to *compare* data, and a histogram is used for a *distribution* of values over a range.. *Relationships* between objects can be shown by a concept map or a Venn diagram. For representing a *flow of data* flowchart diagrams are appropriate. For visualizing *multidimensional data*, radar-type diagrams are used. To illustrate a specific *numerical value*, a gauge is usually used. In semantic analysis, to visualize the *frequency of words* used, wordcloud is most suitable. An interactive heat and hierarchy map or sunburst chart can be used to represent *hierarchical data*. A streamgraph is useful for visualizing a *flow of data*, and chord diagrams are useful for visualizing *relationships* between objects.

Methods for static and interactive visualization

Static visualization focuses on a specific meaning of data and can show its value at a given time. Interactive visualization allows the user to manage how they will view the data. If standard visualization shows things "as they are", interactive visualization tries to answer the question "what if" (Rouse et al., 2017). There are a number of interactive techniques for user interaction with graphical visualisation, some of which are described by Khan(Khan & Khan, 2011):

- zoom in and zoom out, allowing the user to detail the visibility of objects to varying degrees;
- overview and detailing, showing both the location of the object and a detailed view of the selected area;
- focusing and detailing context (fisheye view), which allows focusing on a specific area and viewing it in detail without taking it out of the context of the whole image.

Methods for big data visualization

E-learning environments enable the collection of large amounts of data. Often this data comes from a variety of sources and is presented in a structured or unstructured format. In addition, data generated by social communication tools in the course is generated in real time and at high speed. Modern visualization methods must be able to represent dynamic data volumes effectively.

Bikakis (Bikakis, 2018) describes state-of-the-art approaches to data management, data mining and information visualization. Visualizing big data faces us with a number of challenges as described by Wang (L. Wang et al., 2015), such as loss of information when data is reduced; the appearance of visual noise when data is very close; the difficulty of perceiving large images that change at high speed; the high performance requirements of the computing technique with which they are processed.

Modern e-learning systems collect, store and process data on actually everything that happens in the educational process. We can get detailed information about every aspect of learning, but the biggest advantage of visualization is that everything can be displayed on one screen as a snapshot of the data. One such visual tool is the virtual dashboard.

1.2.3. Tools for learning analytics visualiziation

In the context of this PhD dissertation, learning analytics tools are examined and evaluated in terms of their proposed visualizations.

In a literature review of LA tools in higher education, the author (Salihoun, 2020) discusses about 40 tools commonly used for data collection and analytics in education.

With the widespread penetration and popularity of MOOC courses, the need for analytics to study learner 's behaviour while watching videos is coming to the fore. An example visual analytics system is VisMOOC, described in (Shi et al., 2014). At a later stage, one of the authors of VisMOOC presents two new tools for visualization of MOOC courses (Chen, 2018): PeakVizor и ViSeq. In a study (Caldarola & Rinaldi, 2017) the authors describe 36 of the most used tools and techniques for big data visualization, grouped into 4 categories: (1) information visualization; (2) data visualization; (3) research visualization and (4) business intelligence and visualization tools..

Dashboard

Stephen Few defines a virtual dashboard as „a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance" (Few, 2006). The original idea comes from the dashboard of vehicles or aircraft, which summarizes the status of all engine systems in a small panel. Pauwels (Pauwels et al., 2009) summarizes the goals of the virtual dashboard as (1) enforcing consistency in metrics and measurement procedures; (2) monitoring for effectiveness and early prediction of potential problems; (3) planning for future goals and strategies; and (4) communicating with important stakeholders, with evaluation done by selecting metrics/indicators.

The dashboard can be developed as a stand-alone application or a web-based analysis tool as part of the LMS. The design of the application and the performance indicators depend on the specific objectives.

1.3. Tools for virtual dashboards

This section describes work on the third aspect of Task 1 - to explore and analyse current software platforms for creating a virtual dashboard. The purpose of this analysis is to support the creation of an interactive dashboard in the designed LA visualization system.

In 2022, the leading research company Gartner published a report evaluating 20 analytics and business intelligence (BI) platform providers. This type of platform, according to Gartner's definition, provide the ease and convenience for the user to "analyse, explore, share and manage data", enable him to visualize his own and discover others' insights, often based on artificial intelligence. In this study, the platforms were evaluated according to the following criteria: (1) security; (2) management; (3) cloud analytics, (4) connecting data sources; (5) data preparation; (6) catalogue; (7) automated insights; (8) data visualization; (9) natural language query; (10) data storytelling, (11) natural language generation, and (12) sharing. The rating is based on all published reviews, with older ratings halving every 12 months.

Based on these, Gartner defines its "magic quadrant" in which companies are categorized as "leaders", "challengers", " visionaries and "niche players". For 2022, the leadership position in combining vision and providing value is occupied by Microsoft. Power BI, combined with Azure and Office 365, offer a visual-based presentation of data, advanced analytics, and interactive virtual dashboards. The next position is occupied by Tableau and Qlik. Slightly above average in completeness of vision, but above average in executability, are the Google and Domo products.

The full version of the PhD dissertation discusses the advantages and challenges of each of the Microsoft Power BI, Tableau, Qlik, Google Data Studio and Domo platforms and concludes that Microsoft Power BI allows the connection of most types of data sources from an educational perspective, allows processing and local storage, an extremely rich set of visualizations, as well as the ability to customize existing and program new ones, offers a user-friendly and intuitive interface, supports a mobile version, can be used for free to create visualizations, and the price for sharing is not high.

Chapter 2. Experiments using different LA methods and approaches to optimize the learning process

This chapter describes the work on the second set task - to study and analyse how the methods and tools of LA can optimize the learning process and increase the learners' performance.

As part of this PhD dissertation, a research was conducted on how LA methods and tools increase learner performance (Miteva et al., 2017). The research method focused on 3 main objectives: (1) improving the quality of learning for students; (2) improving the quality of teaching by faculty; (3) improving the overall organization of course delivery at the faculty level.

2.1. Methodology

The research was based on three e-courses "Introduction to Computer Networks", "Routing and Switching Essentials" and "Scalable Networks", conducted during the academic year 2015/2016 with 98 first year students from the Master's programs "Distributed Systems and Mobile Technologies" and " Protection of Information in Computer Systems and Networks" at the Faculty of Mathematics and Informatics of Sofia University (SU).

For a more detailed analysis, data on learners' behaviour were collected from all possible sources related to the learning process. Organisational data was taken from the university's internal Student Information Management System; student activity data from the university's LMS (Moodle); and performance and activity data from the Instructure Canvas system. All collected data were stored in a centralised database (LA database) and further analysed using standard statistical software tools - in our case the R language and the R Studio environment. Finally, this database of results has been analysed by educational experts who have formulated a list of recommendations to improve the learning process from the perspective of all stakeholders.

2.2. Challenges and conclusions

Building a learning analytics system faces many challenges. Such a system should gain access to as large data as possible. An example challenge is to offer automatic analytics based on data from more than one learning course. LMS work at the course level, built-in logs, accessible for the teacher, show statistics concerning a single course. When we want to track how the curriculum is perceived by the learners and how they deal with the learning material, a more complex analysis is needed.

The same time, it is a technological challenge to collect, store and process data imported from external systems without having a predefined relational schema defining how to map data. Importing data from a learning system that is outside the administrative control of the institution and outside the authority to register a user account with special rights for data extraction.

The natural language is also a challenge. The university systems can store data in both Bulgarian and English languages. Learners are allowed to edit their profiles, change personal details such as names and email addresses. A way needs to be found on how these data to be matched.

This chapter describes an experimental analysis of the application of various methods and approaches for LA to optimize the learning process to improve the quality of learning for students; enhance the quality of teaching by faculty and improve the overall organization of learning at the level of the educational institution.

Chapter 3. Experiments with designed, developed and/or used visualization methods

This chapter describes the work on the third set task - to investigate and analyze the functionalities of the virtual dashboard as a visualization tool for learning analytics in the LMS. The results of this analysis are further used in the design of a virtual dashboard in the target system for visualising LA.

3.1. Approbated models of virtual dashboards

The following virtual dashboards are approbated:
- Moodle - one of the most popular and widespread open source LMS. For each course, a set of reports are provided that filter the data collected in the system's log files and extract information on request.
- Graphical Interactive Student Monitoring Tool (GISMO) is a Moodle plugin and provides a comprehensive system for monitoring and tracking learner activities in an online LMS course, such as course attendance, material review, assignment submission, etc.
- Canvas Analytics is approbated in the period from 2012 to 2019 within Sofia University "St. Kliment Ohridski" Cisco Networking Academy

- virtual dashboard of the Learning Analytics and Reflection & Awareness Environment (LARAe) system - created within the "Working Environment with Social and Personal Open Tools (weSPOT)" project to support learning analytics in playing out research scenarios using the inquiry-based learning.
- Mobile Inquiry Coordinator Interface[5] - designed to supervise and coordinate learners in mobile activities as they apply the inquiry-based learning.

3.2. Comparison of tools for creating virtual dashboards

Table 1 systematizes the key features of the described virtual dashboard creation tools. From their comparison, it can be concluded that none of the studied tools meets all required criteria.

Table 1, Comparison of virtual dashboard tools

Criterion	Moodle	GISMO	Canvas	LARAe	MICI
Visualisation of assessments	√	√	X	X	X
Graphical data visualization	X	√	√	√	√
User-friendly interface for request definition	X	√	√	√	√
Ability to compose a personalized query, beyond the built-in	частично	X	X	X	X
Mobile version	X	X	X	X	√
Version update	√	X	√	X	X

3.3. Conclusions and recommendations

The described LA tools provide a large set of predefined analytics reports, such as student activity, course participation, statistics from LMS log files, etc. This information is extremely valuable, but it is not sufficient to answer all questions that may arise during teaching and learning. The proposed reports are not flexible enough and cannot provide the desired freedom to the participants in the learning process. Often the data require further processing with external software to extract useful information, which means more mathematical, statistical and computer literacy, more working time, more software licenses.

One challenge for a modern LA visualization system is to offer a user-oriented interface allowing easy and quick query compilation.

The presented tools are designed to work with structured databases. For the purposes they were developed, and for the amount and types of data they were designed, these tools have done their job. However, a modern tool for processing large data sets must also be able to process unstructured data in real time. Such solutions have been used successfully in business. For example, Hadoop is the primary file system and task tracking tool in Amazon Web Services (https://aws.amazon.com), Cloudera (https://cloudera.com), etc.

---

[5] weSPOT MICI app, https://apkcombo.com/wespot-mici/els.wespot, last access: 01.02.2023

These tools, as well as the SLMs themselves, are oriented to a specific course or study (in the case of tools for serving the research approach to learning). The instructor can track the activities of each learner within a course, but cannot compare that learner's performance across different courses in the curriculum and draw more general conclusions.

A modern LA tool should also offer cross-course learning analytics.

These types of statistics can be used to improve the quality of education by empowering the instructor to self-reflect; by looking for the reasons for the contrast in learner outcomes from different runs of the same course or from different courses; and by assessing the impact of instructional methods on outcomes.

From all the reasons mentioned above, we can conclude that it is necessary to design and develop new technologies in terms of approaches, methodologies and software tools that are able to visualize the complex interrelationship between learners, teachers and courses. Such a tool could be an embedded unit of any LMS, or a stand-alone application.

Chapter 4. Requirements for an integrated learning analytics system

This chapter describes two studies related to the objectives (1) to explore and systematize the reports that users (learners, educators, and managers) expect to find and use in LA section of LMS and (2) to explore and analyze the visualizations of those reports that users find most intuitive and easy to read.

The first study presents expert opinion on the use of big data to benefit learning by creating learning analytics in the LMS. Its aim is to derive the main requirements on which to concentrate system development efforts.

The second study reveals the experts' vision of which visualisations allow these large volumes of data to be read and perceived most quickly and accurately. The results of this study formed the basis for the implementation of the learning analytics visualizations.

4.1. The expert view on learning analytics

As a result of the first study (Miteva & Stefanova, 2020a), the most valuable reports that LMS users expect to find in the LA section are summarized. The results show that experts with rich have higher requirements and expectations. Even for existing reports, experts suggest improved versions and adding missing details. Due to the critical evaluation, new requirements for LA are derived.

4.1.1. Research Methodology

The Group Concept Mapping method (Kane & Rosas, 2017) was used to gather expert opinion and analyse the data. The dedicated online environment of Concept Systems Inc. Global MAX was used, which provides a simple and intuitive web-based

interface to arrange the main activities The study was conducted in the following sequence of activities (Figure 1).
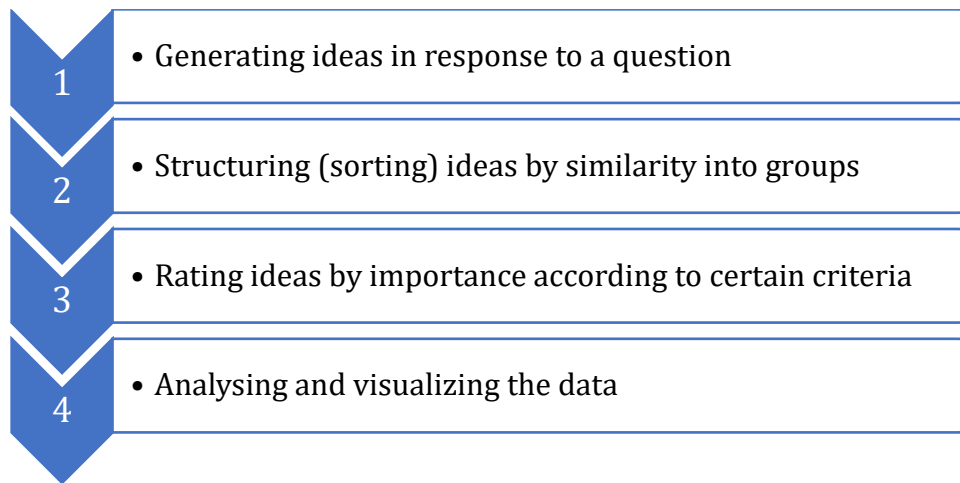

Figure 1. The process of extracting LMS reports

The study involves 30 professionals from the Faculty of Mathematics and Informatics, the Faculty of Education and the Centre of Information Society Technologies of St. Kliment Ohridski, philologists, science teachers, PhD students and undergraduates. The focus question in response to which experts had to brainstorming ideas during the first stage, was "*In Learning Analytics (LA) of LMS I would like to have reports for…*"

As a result of the study, 85 expert suggestions for reports were synthesized for the learning analytics part of the LMS.

4.1.2. Data analysis

The collected data was processed using two statistically methods: multi-dimensional scaling and hierarchical cluster analysis. The results of sorting by each participant are represented by a correlation matrix called a similarity matrix, in which for each two from 85 sentences is marked 1 – if they are sorted in the same group and 0 - if are allocated in different categories. Using the multidimensional scaling method, this matrix is visualized as a point map in which each idea is represented as a point in a plane. During the next phase of data analysis ideas had to be grouped into categories (clusters) by the hierarchical cluster analysis method. As a result, 8 clusters were obtained, named according to the main idea that unites the proposals in them. Figure 2 shows the list of categories, with each one showing in parentheses how many LMS reference proposals are included in it.

1. Teacher Evaluation (19)
2. Student Evaluation (17)
3. Grades (7)
4. Course Feedback (6)
5. LMS Reports (6)
6. Student Support (9)
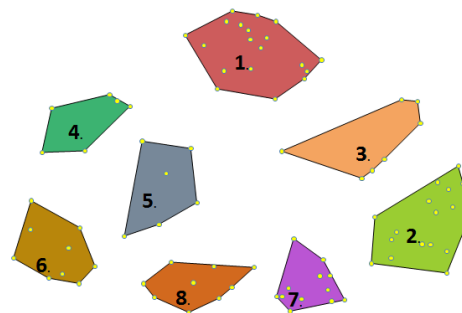7. Student Activity (12)
8. Course Management (9)

Figure 2. Final number of categories with reports

An indicator of how typical each idea is for the cluster it belongs is a parameter bridging value. It varies in interval [0, 1], with lower values indicating that the idea is typical for the cluster, while higher values indicate that the location of the idea is at the "boundary" of the cluster, i.e. if the number of the final groups was larger, it would most likely be part of another group. Among the ideas proposed by the experts, there is a representative with zero marginal value and a representative with maximum marginal value.

In addition to the individual ideas, an average boundary value is also calculated for the each of the 8 categories into which they are grouped. The smaller the cluster boundary value, the more unanimous the experts are that the ideas in the cluster should be together. Conversely, the higher a cluster's boundary value, the more "generalizable" it is by the relevant criterion.

4.1.3. Rating ideas by usefulness/significance

The range of average scores by criterion usefulness/significance is from 3.10 to 4.60. Two suggestions received the lowest rating: (1) *During a course to be visualized in percentage what part has already passed and what part remains* (M (mean)=3.10; SD (standard deviation) =1.3) and (2) *To be visualized statistics on teacher's activity in forums* (M=3.10; SD=0.8). As the most useful is esteemed one suggestion: *In teacher' view to have a graphical representation of schedule conflict (for tests, home works, and exams) between current course and the other courses for the same students* (M=4.6; SD=0.7).

From estimates of the individual ideas, an average score for each category of usefulness/significance can be calculated. Figure 3 shows that as the most useful is evaluated the category Course feedback with score 4.26 out of 5.00 and as the least useful - the categories Student evaluation and Student activity with score 3.81 out of 5.00.

(4) Course Evaluation (**4.26**)
(8) Course Management (**3.99**)
(3) Grades (**3.97**)
(5) LMS Reports (**3.96**)
(1) Teacher Evaluation (**3.92**)
(6) Student Support (**3.90**)
(2) Student Evaluation (**3.81**)
(7) Student Activity (**3.81**)
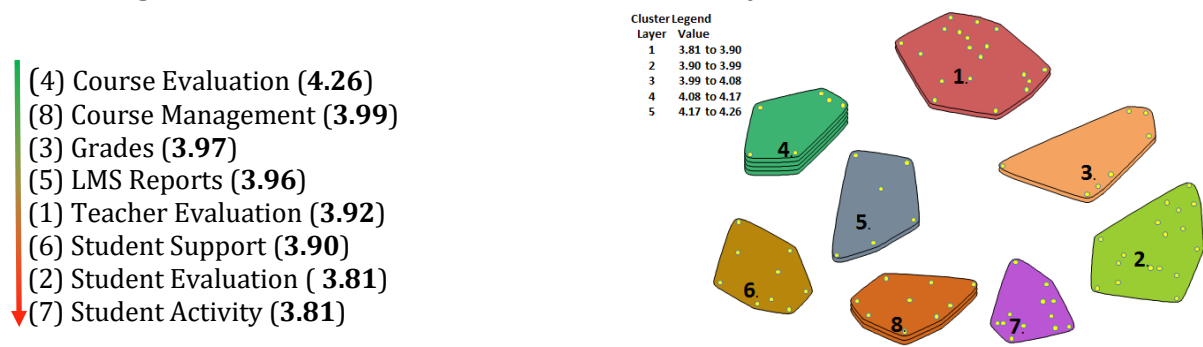
Figure 3. Rating the categories according to usefulness/significance

The results of the usefulness/significance evaluations at a later stage were used to select and prioritize the reports that the planned learning analytics system should offer.

4.1.4. Rating ideas by applicability/feasibility

The average ratings given by the experts for the second criterion of applicability/feasibility range from [2,95 – 4,45]. As the most difficult to perform is marked: *To visualize an estimated time for publishing results of a test/homework/exam* (M=2.95; SD=1.23) and as the easiest to implement is ticked: *For each assignment/activity to be visualized a list of all students already submitted it* (M=4.45; SD=1.05).

From estimates of the individual ideas, the average rating could also be calculated for each category. Figure 4 shows that the most feasible are reports in the category Grades with score 4.16 out of 5.00 and the most difficult to implement – in the category Student support with score 3.73 out of 5.00. Concerning groups, the difference between the min and max average scores is not very high.
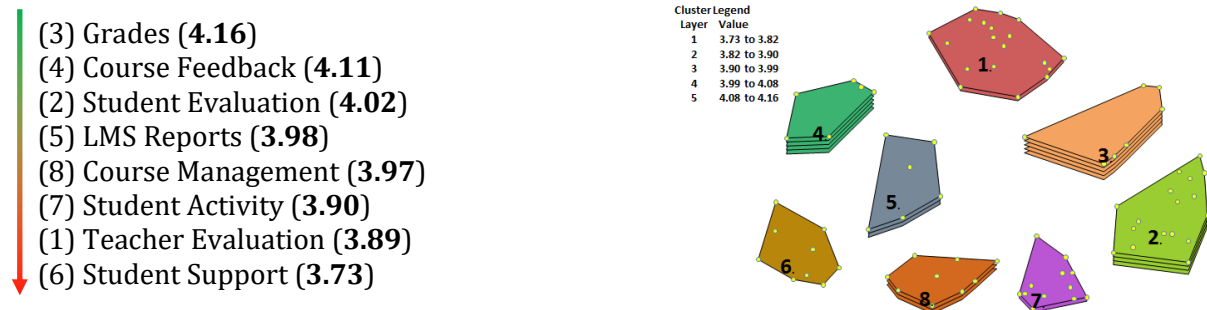
(3) Grades (**4.16**)
(4) Course Feedback (**4.11**)
(2) Student Evaluation (**4.02**)
(5) LMS Reports (**3.98**)
(8) Course Management (**3.97**)
(7) Student Activity (**3.90**)
(1) Teacher Evaluation (**3.89**)
(6) Student Support (**3.73**)

Figure 4. Rating categories according to the applicability/feasibility criterion

The results of the applicability/feasibility evaluations later are taken into account when selecting and prioritizing the reports that the planned learning analytics system should offer.

4.1.5. Comparison of scores on both criteria

It is interesting to compare the average scores of the proposed reports by category on the two criteria. Some of the categories received almost the same average rating, e.g., "LMS Reports" (usefulness/significance: 3.96 out of 5 - applicability/feasibility: 3.98 out of 5) or "Course Management" (usefulness/significance: 3.99 out of 5 - applicability/feasibility: 3.97 out of 5), while the reports in other categories were rated as considerably more applicable than useful, e.g., "Grades" (usefulness/significance: 3.97 out of 5 - applicability/feasibility: 4.16 out of 5) or "Student Evaluation" (usefulness/value: 3.83 out of 5 - applicability/feasibility: 4.02 out of 5) or more useful than easily applicable, e.g. "Student Support" (usefulness/value: 3.90 out of 5 - applicability/feasibility: 3.73 out of 5). In general, usefulness scores are higher than feasibility scores (Figure 5).
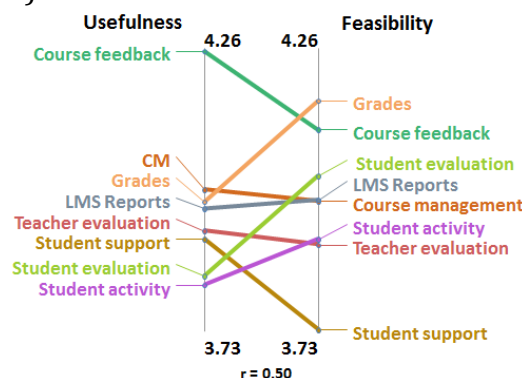

Figure 5. Categories rating comparison

4.1.6. Results

The evaluations of the suggested reports can be summarized in go-zone diagrams, dividing the area into 4 zones, according to the average values of the evaluations of the

two criteria: usefulness and feasibility. In the green zone or quadrant I are visualized the reports whose scores are above the average value for both criteria, in the orange zone or quadrant II are positioned the reports that have scored above the average value for feasibility, but below the average value for usefulness, in the grey area or quadrant III are the reports that scored below average on both criteria and in the yellow area or quadrant IV are the reports that scored above average on the usefulness criterion but below average on the feasibility criterion. In addition to visualizing the distribution of queries in each category as a whole, these charts can also filter the results by role: manager, teacher, and student. Thus, on the one hand to see the reports that each role has prioritized, and on the other hand to see the cross-sections, i.e. the reports that should be developed for the users of the LMS in all roles.

The first version of the analytics visualization system will implement the reports that have been rated above average by experts on both criteria. Reports that scored above the mean on only one of the two criteria will be revised and implemented in a second iteration. Reports that are scored below the average on both criteria will be reviewed, revised and proposed for re-evaluation. The number of reports in each category by zone is shown in Table 2.

Table 2. Number of reports by category and role

Category	Manager				Teacher				Student			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
Teacher evaluation	9	2	7	1	11	3	5	1	6	3	5	5
Student evaluation	9	4	2	2	4	5	3	5	5	5	6	1
Grades	4	2	0	1	3	0	4	0	3	1	2	1
Course feedback	3	1	2	0	3	2	1	0	1	2	1	2
LMS reports	2	1	3	0	1	2	0	3	1	2	1	2
Student support	3	1	3	2	2	3	3	1	4	2	1	2
Student activity	4	2	3	3	5	1	4	2	5	1	4	2
Course management	3	2	3	1	4	2	2	1	4	0	3	2

In this chapter, the main requirements for an integrated for LA system are extracted in two aspects: (1) a list of all the reports that stakeholders would like to find in LMS, (2) for each type of data presented in these reports, the visualization identified as the most intuitive, easy to read and quick to understand is selected.

4.2. Experts' perspective for selecting learning analytics visualizations

This section describes a survey (Miteva & Stefanova, 2020b) of expert opinion on the types of visualizations for the LMS reports described in the previous section.

4.2.1. Research Methodology

The research involved 48 experts with different background, experience and LMS role, related to Sofia University. Participants filled out an online survey „Data visualization

in LMS" sharing their professional opinion on which visualizations presents the most appropriately and accurately data in the main types of reports.

The list of reports for the learning analytics section of the LMS defined in 4.1 was further analysed with respect to the data types of the results and the purpose of their visualization: composition, comparison, relation or disaggregation. In accordance with Abela's classification, 12 main data types are extracted for illustration. For each of them, a section in the questionnaire has been defined with a brief description of what type of data is expected to be visualized, to whom and for what purpose. From 2 to 6 sample visualizations are proposed for evaluation from 1 (non-suitable) to 5 (suitable). Each section ends with an option for comments and new ideas.

The collected data were processed, analysed and categorized according to the system roles of the respondents in order to propose virtual dashboards with visualizations that would satisfy the largest possible number of LMS users.

4.2.2. Data analysis

This section provides details of the proposed visualizations in the survey and the scores achieved according to the respondent's role in the LMS. Recommendations made to the visualizations are also commented.

4.2.3. Results

The target LA visualization system is assumed to suggest a proper analytics dashboard for each LMS role: manager, teacher or student. When choosing the most effective visualizations for its elements, two simple rules were followed:

- If the experts from an LMS role have rated the visualization of data output with a score above the range average, this type of report will be included in their dashboard.
- If respondents gave scores below the range average, this visualization will be omitted in the first version of the system and further research will be done.

4.2.4. Conclusions

Among the proposed visualizations for expert evaluation, there are reports and statistics presented in some available LMSs and missing in others. Comments in the survey show that attitudes towards familiar visualizations are quite extreme - one group is used to them, accepts them and recommends them, while another group is dissatisfied and wants the same results to be presented with newer, more intuitive methods. On the other hand, even if don't agree with well-known charts, respondents can hardly imagine how exactly new visualizations will work in real time and with real data. Meanwhile, more experienced experts offer practical ideas that enrich the survey results. It can be concluded that the new types of visuals should be presented in a more attractive way and in a working environment then the study should be repeated. This idea is set out in the plan for future development of the system. For better visualization and interpretation of the data, advanced methods and techniques have been applied to present big data, such as data reduction, hierarchical visualization, progressive display of results, incremental and adaptive processing, caching and pre-retrieval. At a future point, 3D visualizations or even virtual reality may be added.

The next step is to create a model for a learning analytics system that uses contemporary visualization methods to help stakeholders make the right decisions at the right time.

Chapter 5. A model of a learning analytics visualization system eAnalytics

This chapter describes an architectural model of a learning analytics visualization system called eAnalytics. The aim of the model is to meet as much as possible the requirements of users using the LMS in the three main roles: student, teacher or manager of an educational institution. Figure 6 presents the elements of the model.
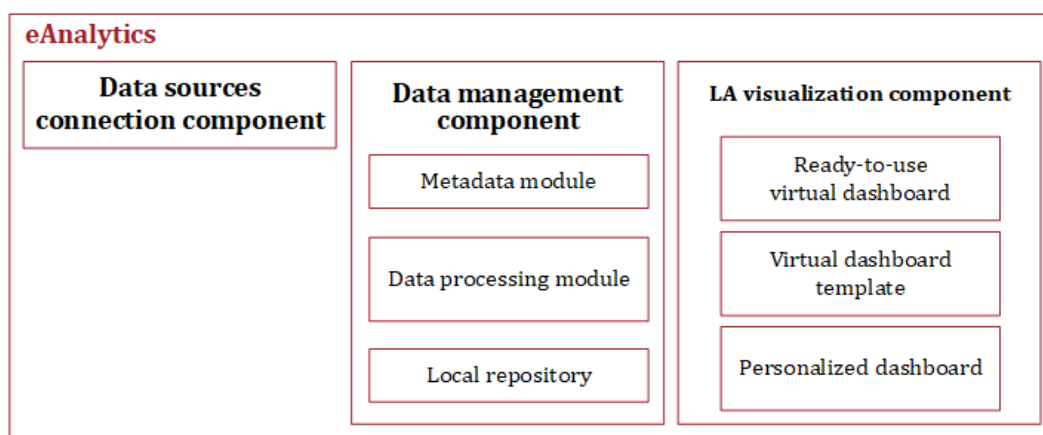

Figure 6. Model architecture

5.1. Users and roles

eAnalytics recognizes several default system roles, which are granted different permissions on the available functions:

- **Manager**. Designed for the institution's management team to derive analytics on both learner and educator participation in a course, overall performance, and student opinion and evaluation of their instructors.
- **Teacher**. Designed for lecturers to monitor the overall conduction of a course, to observe the individual performance of each learner, to create a complex analyses of student performance.
- **Student**. A role for learners to monitor their performance, successes and weaknesses and receive feedback on how to improve their own performance.
- **Administrator**. He is responsible for the installation, configuration and management of the system, is granted full access.
  The system allows for definition of new roles and delegation of respective rights.

5.2. Components of the system

eAnalytics is a web-based system with a hub architecture between connected learning systems that stores data on teaching, learning and learning management

processes to give a holistic view of learner progress. Collecting more data for more accurate evaluation is realized by merging data from multiple learning analytics sources. Using APIs, the system connects to different LMSs and databases and uses metadata to describe which analytics they will provide.

The system consists of three main components: (1) Data sources connection component, (2) Data management component, and (3) LA visualization component.

5.2.1. Data sources connection component

The data source connection component is responsible for providing connection and access to the sources according to the specific user, his roles and permissions in the respective LMS. The metadata for this access is stored in the system and is used to retrieve data from it.

5.2.2. Data management component

This component includes three modules: (1) metadata module, (2) data processing module, and (3) local repository.

Metadata module

Once metadata is extracted and associated relationships described, this module builds a hierarchical metadata model structured by system role, course access, and data provided. The basic unit into which metadata is grouped in eAnalytics is a course, and the access to analytics depends on the user's role in the system.

Data processing module

For some of the analytics, mainly related to combining data from different sources, additional processing is required in the system itself. eAnalytics offers a set of standard formulas, mathematical and statistical functions for data processing and transformation, for text analysis, for data parameterization.

Local repository

Extractions from big data can be imported and stored in local storage. Such can be used when the data transmission from the source to the system does not need to be implemented in real time. In such cases csv or xlsx files can be exported and imported.

5.2.3. Learning analytics visualization component

This component provides a choice of several methods for visualizing analytics: (1) via a ready-to-use virtual dashboard, (2) via a virtual dashboard template, or (3) by composing personalized analytics. In the first method, the virtual dashboard contains information about the data sources and can be shared in a view-only or editable manner. The second option enables the user to select/specify the data sources themselves in order to obtain the corresponding visualizations. The resulting virtual dashboard remains in the user's profile. The third option directly links to a custom analytics composition screen where the data sources must first be selected, then the data itself, and finally the visual objects that will represent the data. Based on the selected data types, "appropriate" visualizations will be recommended, but the choice is left entirely to the user (Figure 7).
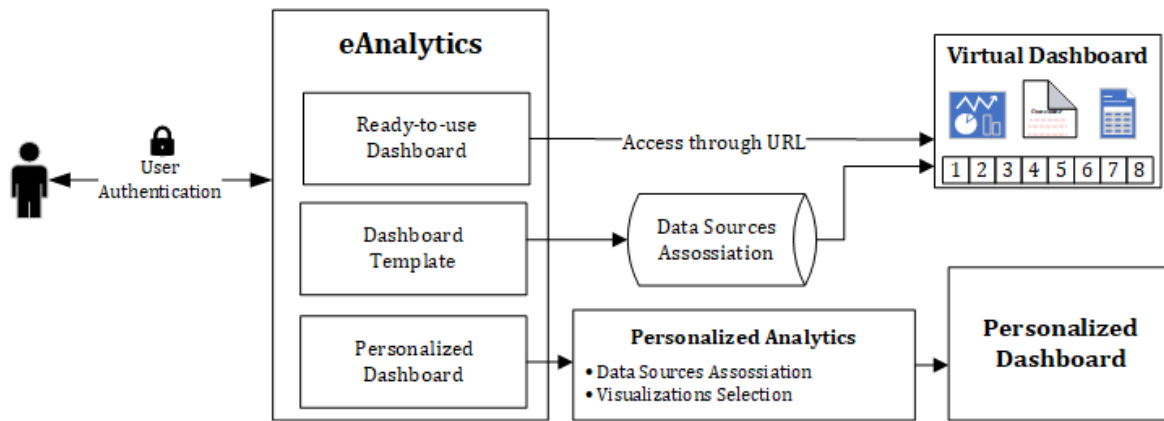
Figure 7. Learning analytics visualization component

Visualization via a ready-made virtual dashboard

The virtual dashboard that the user accesses is fully ready to use. Its main advantage is that it does not require any additional settings, nor knowledge of the data structure and the relationships between them. Its content varies depending on the system role of the user. If the user has more than one role assigned, he can choose which one he participates in at any given time.

The dashboard consists of eight sections corresponding to the categories described in Chapter 4: (1) Teacher Evaluation, (2) Learner Activity, (3) Learner Evaluation, (4) Course Management, (5) Grades, (6) Student Support, (7) LMS Report, and (8) Course Feedback. The list of analytics and visualizations in each category varies depending on the current role and expert perspective on selecting visualizations described in 4.2.

The LA visualization component allows switching between the default visualizations developed for different roles, as well as reflecting the settings in the user's profile. Additionally, options are provided to filter the data and export the results in tabular and/or graphical format.

Visualization via virtual dashboard template

The template is structured as a ready-to-use virtual dashboard, contains all the analytics and the appropriate visualizations for them (Chapter 4), keeps the metadata model and the relationships between them, but has no connectivity to specific data sources and the local repository is empty. In order to visualize analytics, the template must first be associated with data sources. Then, the data sources connection component and the metadata module complete and update the metadata. This is followed by extracting, processing the data and displaying the relevant analytics in the virtual dashboard.

This visualization method has a number of advantages. The virtual dashboard is versatile, it can work with different sources that support a single data model. For eAnalytics, this means that developed for one specific institutional structure, it can be applied to multiple analogous structures.

Visualization via composing personalized analytics

Creating a personalized virtual dashboard allows the user to define what analytics they want to include and what visualizations to use for them. The personal virtual dashboard is created in the following sequence of actions in the system: (1) selecting data from the metadata module, (2) selecting visualizations, (3) customizing parameters.

The data that is provided to compose analytics is described clearly and intuitively in the system and gives a precise idea of what information it includes. They can be accessed from different perspectives: student data, faculty data, course data.

Each user can personalize the layout of the system for his profile. The administrator role defines the global layout settings for the entire system.

5.3. Mobility, Security and Ethics

No matter if users are connecting from a desktop, laptop, tablet or smartphone, the analytics reports are offered/adapted in the best possible format for the respective device. The system is available via Android and iOS apps.

In order to reduce the risk of vulnerability in the storing sensitive data, the model envisages addressing the relevant LMS only upon request and for retrieving only the requested data provided per selected account. When results need to be published to third parties, data is anonymised.

The model envisages the application of a code of ethics, in accordance with the requirements of the organization that will use the relevant learning analytics system and the rules to be published explicitly in the system.

Chapter 6. Prototype Implementation

This chapter presents the work on Task 6 - prototypeprototyping the LA visualization system. An implementation of the model is demonstrated for system role manager and ready-to-use virtual dashboard for the category "Teacher Evaluation". A dashboard template is also created in the prototype. The template stores information about the data model, the visualizations created in it, but does not contain specific data.

6.1. Choosing a Tool for Virtual Dashboard Implementation

There are a number of modern technologies and tools for visualizing big data increasingly penetrating the field of education. Some of them offer an easy and user-friendly interface, others require considerable technological and programming knowledge. Some of these tools are closely specialized for particular types of visualizations, e.g. geographic maps, others provide a variety and are open to adding new ones.

Based on the advantages and disadvantages of the virtual dashboard tools described in Chapter 1, a choice was made to use the Microsoft Power BI platform to create the prototype, as it covered the most of the described platform selection criteria.

6.2. Data and Metrics used in the Prototype

The prototype of the model is implemented on a real problem with the processing and visualization of large volumes of data related to the students' feedback in the courses at Sofia University "St. Kliment Ohridski.

At the end of each academic term at Sofia University a course evaluation survey is conducted, through which students give feedback on the courses they have studied, the attendance of classes, the work of their lecturers and assistants, share their opinions and recommendations. In this questionnaire, students' rate on a scale of 1 to 6 or yes/no/cannot decide the performance of lecturers and teaching assistants - whether they present the teaching material in an accessible way, whether they encourage students and stimulate creative thinking, whether they hold regular classes, whether students approve teaching methods, whether students would recommend the course to their colleagues. Each of the questions is evaluated with a certain weight and according to a regulated by SU formula forms part of the teacher's attestation evaluation. So far, the results of this questionnaire have been processed manually and this made it difficult to analyse and extract meaningful information.

The virtual dashboard described in this chapter was developed according to the data model of these surveys.

6.3. Importing the data and binding it to the model

For the purpose of the demonstration, the data is obtained as "Excel Workbook" and "Text/CSV", pre-formed into 3 different groups simulating 3 different data sources.

A metadata model has been built that specifies the relationships between the different datasets. The possible relations are (1) one-to-one, (2) one-to-many, (3) many-to-one, and (4) many-to-many. The system provides the ability to define relationships visually through drag-and-drop functionality.

6.4. Creating a virtual dashboard and selecting visualizations

According to the eAnalytics model, the virtual dashboard for "Teacher Evaluation", accessible for the manager role provides 8 reports to the user. The visualizations are grouped into four sections/pages, and a start page has been added with explanations what the user should expect from this virtual dashboard and what the evaluation criteria are.

6.5. Creating a Virtual Dashboard Template

To allow the created virtual dashboard to be reused many times, but with different data sources supporting the same format and data model, the system offers the opportunity to create a virtual dashboard template. If the data source is set as a parameter in this template, then the next time the template is opened, the user will be able to choose what data to be used with the template.

In the prototype, the data sources are set as parameters. At start-up, parameter values are set and all data gets a real path to the corresponding source.

6.6. Creating an app and sharing in an online store

The eAnalytics application (eAnalytics_app) was created for the purpose of the prototype, which was published in Power BI apps and shared with specific users from Sofia University.

Chapter 7. Testing the prototype

This chapter describes the experimental testing of the created eAnalytics prototype in response to Task 7. Two types of tests were conducted: functional tests for compliance of the implemented with the set requirements for the prototype and user tests by applying the "Think aloud protocol" (Lewis, 1982).

7.1. Preliminary data preparation

Prior to processing, the data were cleaned of incorrect or partially completed records. Reports generated with data from the LMS contain sensitive data and for the purposes of the demonstration de-identification has been applied to mask personal data and minimise the risk of inadvertent disclosure of faculty and their information. The open source tool Amnesia[6] which follows GDPR guidelines for pseudo-anonymisation, was used. Through it, the link to the original dataset is broken. The identifiers first name, surname, last name and subject were pseudo-anonymized, because in combination or separately they would help to identify the evaluated teacher.

7.2. Functional testing of the prototype

This section describes the tests performed on the developed eAnalytics functionalities: (1) opening a virtual dashboard template and configuring it to work with real data sources, (2) testing a ready-to-use virtual dashboard, and (3) creating a custom virtual dashboard.

7.2.1. Virtual dashboard template testing

When opening the template, a dialog box first appears prompting the user to select (or enter a web address of) real data sources to be visualized in the virtual dashboard. In the tested template, the user is facilitated to directly select from a drop-down list of predefined LMSs. The model reads the specified addresses, connects to the sources and loads the data.

First, the initial screen is displayed (Figure 8), describing the analysed period, the criteria for assessing student opinion and the types of analytics provided, with links to their presentations.

---

[6] Amnesia Anonymization Tool, https://amnesia.openaire.eu/index.html, last access: 01.02.2023

Figure 8. Virtual dashboard "Teacher Evaluation"

By following the links or navigation between the separate pages, all sections of the virtual dashboard can be visited and analysed.

7.2.2. Testing a ready-made virtual dashboard

The ready-to-use virtual dashboard, published as an application in the online store, was tested and used by a registered Sofia University user of the Power BI service. It was accessed (1) via a URL link provided upon sharing, (2) by scanning a QR code, and (3) by searching for the name such as "eAnalytics_app" in the Power BI apps online store.

The application is downloaded to the user's account, installed and available for online use at any time (Figure 9).
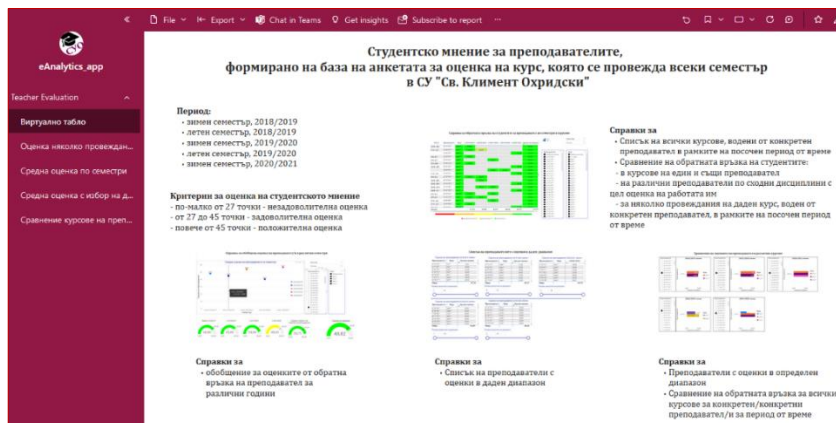
A) eAnalytics accessed via web

B) Mobile version of eAnalytics

Figure 9. eAnalytics application

7.2.3. Testing building a custom virtual dashboard

eAnalytics is shared with a test user with edit permissions. This user can create their personal analytics and custom visualizations with the provided data. The user created a new report in which he (1) marked to use the data for the number of faculty with average positive, satisfactory and unsatisfactory attestation for all analyzed terms, (2) selected a pie chart as the visual object, (3) customized the visualization settings and illustrated the proportion of all faculty according to attestation for the entire analyzed period (Figure 10).

Figure 10. Testing the creation of a personalized dashboard

From the tests performed, it can be concluded that the implemented features of the prototype described in Chapter 6 work as intended in the prototype.

7.3. User tests with "Think Aloud" protocol

In this type of testing of eAnalytics were involved 10 experts from the Faculty of Mathematics and Informatics of Sofia University who declared they have been actively using LMS in their work for more than 10 years. One of them is a member of the FMI attestation committee. The application was shared from the Power BI online store with them in advance and each participant had to complete the following list of tasks to test the web and mobile version:

1. Find and download the eAnalytics app from the online store.
2. Launch the application.
3. Browse all pages, view the suggested analytics and use the navigation.
4. Switch between the different types of visualizations available for the same analytics.
5. Finding the highest and lowest faculty member's grade for each of the 5 semesters displayed. (Without specifying which report to search for the information.)
6. Retrieve three different lists of lecturers: with unsatisfactory, satisfactory and positive attestation for a randomly selected semester. (Without specifying which reference to search for the information.)

Each participant was instructed to share aloud any thoughts that occurred to him or her while completing the assigned tasks.

Results of observation

The first and second tasks were done very quickly. Navigation was found easy and intuitive, and switching between the different visualizations was findable and logical. Tasks 5 and 6 were completed with ease. The only comment was related to the explicit display of the download button, not only when the object is in focus, i.e. the mouse is over it.

First impressions were expressed with positive enthusiasm like "Oh, perfect, this will be very useful for bosses!", "Wow, the interface is adapted for users with special needs!", "The idea with the colours is very good, you don't need to look at the numbers to

orient yourself!", "Oh, this is very useful too!", etc. After the initial euphoria, a more critical evaluation and list of recommendations was made.

After the web app, the mobile version was suggested for testing. It was evaluated as more intuitive, easier to use and with a better design.

Testing of the prototype shows experimentally that big data, statistics and graphic visualization of the data speed up the perception of the data many times and enable the participants in education to answer at least some of the questions asked at the beginning.

Conclusions

The dissertation "Visualization of Big Data" studies the possibilities of increasing the efficiency of learning, teaching and learning management processes by using modern methods for visualization of learning analytics. Its goal is to design a LA visualization system that, through appropriate visualizations, supports learners, teachers and managers of learning institutions to make the right decisions at the right time.

Summary of the results

The contributions of the dissertation are as follows:

The use of learning analytics in education is explored (Chapter 1). The metrics and KPI used in LMS are analysed. Current methods of collecting data from LMS describing learners' success and progress, their activity and responsibility towards the tasks, their attitude towards learning and social activity are studied. The more well-known models for applying LA in training are discussed. The ethical rules for the collection and processing of personal data regulated by the GDPR legal instrument are commented.Modern visualization methods for a variety of data types: static, interactive, and big data. Visualization tools for learning analytics actively used in MOOC courses are discussed.

Modern visualization methods for different categories of data are analyzed: static, interactive and big data. Visualization tools for learning analytics actively used in MOOC courses are discussed.

Platforms for creating a virtual dashboard are explored. The leading analytics and business intelligence platforms according to Gartner's 2022 rankings are examined.

Experiments have been conducted using different LA methods and approaches to optimize the learning process (Chapter 2) with the aim of (1) improving the quality of learning for students, (2) improving the quality of teaching by faculty, and (3) improving the overall organization of learning at the level of the teaching organization.

Experiments have been carried out applying different methods for visualizing learning analytics (Chapter 3), such as Moodle, GISMO, Canvas analytics, LARAe, MICI (the last two were created by the weSPOT project). As a result, a set of challenges have been identified that need to be addressed in the design of a modern visualisation system..

The main requirements for an integrated learning analytics system are defined in two aspects: (1) on the one hand, a list of all the analytics that stakeholders (learners, teachers and managers of learning institutions) would like to find in the learning analytics

part is derived, (2) on the other hand, for each type of data presented in these reports, the visualization is identified as the most intuitive, easy to read and quick to understand (Chapter 4). The list of reports is separated into eight categories, and the list of rated reports and visualizations vary for different LMS user roles.

Based on the expert-defined requirements, a model of a learning analytics visualization system called eAnalytics was created. It assumes the collecting and combining of learner data from more than one LMS to allow more comprehensive analysis and more accurate results on the status of the individual learner and educator. Visualized properly, these outcomes support good decision-making and responding to individual support or encouragement for better performance (Chapter 4).

Based on the model, a prototype has been created that implements a virtual dashboard visualizing the teacher's evaluation from students useful for manager role. In parallel, an application was created and shared through the Power BI apps online store.

Finally, successful functional and user tests of the template and the online application were performed.

All the objectives set in the dissertation have been achieved, and thereby the goal of creating a system that supports learners, teachers and managers in improving the quality of learning through advanced visualizations for big data has been met.

Directions for Future Work

The LA visualization system described in this dissertation will continue to be developed in several directions.

First, it is based on a survey of what learning analytics visualization experts find useful and necessary. For some of these reports, learning systems collect and store enough data to be implemented. For other reports, additional planning is needed on the one hand what data to collect, on the other hand how, by what means.

Another study on the use of advanced data visualization methods is to be conducted, as noted in Chapter 4. In this study, the latest big data visualization methods applied to learning analytics will be presented in a real-world setting so that respondents can better appreciate the benefits and challenges of each. The current version of eAnalytics will be able to be used for this study.

Given the sensitivity of the data in the LMS that is processed and visualized in the learning analytics, one of the directions for further development of eAnalitics is to add a local server for the university to host all the virtual dashboards. Thanks to it, the data processed and visualized by eAnalytics will not leave the Sofia University network. Their sharing will be with users from the university itself or, in the case of external users, secure protocols will be used to ensure data security and protection. A 'gateway' application will be added to the server to allow data to be kept on campus even when a virtual dashboard is published to the cloud. In addition, a service will be added to embed the eAnalytics virtual dashboards within the SU systems.

Modern big data processing and image visualization technologies are constantly evolving. The requirements for the LMS are constantly increasing, new functionalities are developed, more and more data are collected. Expectations for learning analytics

visualizations will also grow. To be useful, eAnalytics will be keep up-to-date to meet the requirements of its users: educators, students and managers.

Contributions

Scientific Contributions

1. The paper explores learning analytics in education, the metrics and key performance indicators used, methods for collecting learner data, and models for applying learning analytics and their role in enhancing the quality of teaching and learning. Ethical standards and privacy protections in the application of learning analytics are commented on (Chapter 1).
2. Methods for visualizing large amounts of data and tools for visualizing learning analytics are explored (Chapter 1, Publication 1, Publication 2).
3. Current platforms for creating a virtual dashboard are analyzed (Chapter 1).
4. An experimental analysis of the application of different LA methods and approaches to optimize the learning process is performed and the challenges of creating a system for visualizing LA are extracted (Chapter 2, Publication 3).
5. Tools for visualizing learning analytics are explored (Chapter 3, Publication 4).
6. The LA reports that LMS users expect to find in the learning analytics section are analysed and systematized (Chapter 4, Publication 5).
7. The visualizations of learning analytics that LMS users in different roles (learners, teachers and managers) find most intuitive and easy to read are analyzed and systematized (Chapter 4, Publication 6).

Scientific and Applied Contributions

- A model for a learning analytics visualisation system has been developed to best meet the requirements of LMS users (Chapter 5).
- A prototype model of a learning analytics visualization system has been created. The implementation of the "Teacher Evaluation" category for the system role "manager" is demonstrated (Chapter 6).
- The prototype of virtual dashboard, part of the learning analytics system, was tested experimentally (Chapter 7).

Scientific Publications related to the Dissertation Work

Dafinka Miteva's publications connected with the dissertation:
Chapter 1:

1. Miteva, D., Stefanova, E. (2018). From Big Data to Big Decisions via the Bridge of Learning Analytics Visualization, Eight International Scientific Conference COMPUTER SCIENCE' 2018, pages:85–91, ISBN:978-619-167-359-9
2. Митева, Д, Стефанова, Е., Визуализиране на учебни аналитики за по-ефективно обучение, Седма национална конференция "Електронното обучение във висшите училища", издателство: Университетско

издателство "Св. Климент Охридски", 2018, стр.:143-150, ISBN:978-954-07-4509-1

Chapter 2:

3. Miteva, D., Stefanova, E., Stefanov, K. (2017) e-Analytics for e-Learning, International Journal of Human Capital and Information Technology Professionals (IJHCITP), vol:8, issue:4, 2017, pages:1-13, ISSN (print):1947-3478, ISSN (online):1947-3486, https://doi.org/10.4018/IJHCITP.2017100101, SJR (0.28 – 2017)

Chapter 3:

4. Miteva, D., Stefanova, E. (2016), Driving teaching interventions with analytics, International Conference END 2016 Book of Proceedings, editor/s:Mafalda Carmo, Publisher:World Institute for Advanced Research and Science (WIARS), Portugal, 2016, pages:349-353, ISBN:978-989-99389-8-4 http://end-educationconference.org/wp-content/uploads/2016/07/END-2016_Book-of-Proceedings.pdf

Chapter 4:

5. Miteva, D., Stefanova, E. (2020). Design of Learning Analytics Tool: The Experts' Eyes View. Proceedings of the 12th International Conference on Computer Supported Education – Volume 2: CSEDU, 307–314. https://doi.org/10.5220/0009395503070314, ISBN978-989-758-417-6

6. Miteva, D., Stefanova, E. (2020). Experts' perspective for selecting learning analytics visualizations. 2020 International Conference Automatics and Informatics (ICAI), 1–6. https://doi.org/10.1109/ICAI50593.2020.9311341

Participation in Research Projects, related to the PhD Work

1. Improving online and blended learning with educational data analytics (ILEDA), ERASMUS+ 2021-1-BG01-KA220-HED-000031121, 2021-2023
2. The use of high-tech tools in building competency models in education, Sofia University Science Fund, №80-10-18/18.03.2020,
3. The National Science Program „Information and Communication Technologies for Unified Digital Market in Science, Education and Security", https://npict.bg/, Ministry of Education and Science of Republic of Bulgaria, 2018-2021
4. Use of high-tech tools in building competencies for applying an inquiry-based learning in STEM disciplines in high school, Sofia University Science Fund, 80-10-90/19.04.2018
5. Inquiry-based learning in the field of high technologies as an application of modern information technologies, Sofia University Science Fund, 80-10-217/24.04.2017
6. Study of the applicability in practice of new technological tools and approaches to teaching and learning process, Sofia University Science Fund, 30/31.03.2015, 2015
7. ASSETS, Methods and technologies to improve access to electronic services, funded by the National Science Fund, Ministry of Education and Science in Bulgaria – Part of the research team, November 2014 – April 2016

8. Working Environment with Social and Personal Open Tools for inquiry based learning (weSPOT), FP7-ICT-2011-8-318499, 2012-2015

Declaration of Originality

I declare that the presented in connection with the procedure for obtaining the educational and scientific degree "Doctor" at Sofia University "St. Kliment Ohridski" dissertation on the topic: "Big Data Visualization" is my own original work. It was done during 2014-2023.

Citation of all sources of information, text, figures, tables, images and others are made according to the standards.

The results and contributions of the dissertation research are original and are not borrowed from research and publications in which I do not participate.

Bibliography

1. Abela, A. V. (2010). *The presentation: A story about communicating successfully with very few slides*. CreateSpace.
2. Bikakis, N. (2018). *Big Data Visualization Tools*.
3. Caldarola, E. G., & Rinaldi, A. M. (2017). Big Data Visualization Tools: A Survey. *Proceedings of the 6th International Conference on Data Science, Technology and Applications*, 296–305. https://doi.org/10.5220/0006484102960305
4. Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A Reference Model for Learning Analytics. *Int. J. Technol. Enhanc. Learn.*, *4*(5/6), 318–331. https://doi.org/10.1504/IJTEL.2012.051815
5. Chen, Q. (2018). *Visual analytics and storytelling of data from massive open online courses* [The Hong Kong University of Science and Technology]. https://repository.hkust.edu.hk/ir/Record/1783.1-95778
6. Clow, D. (2012). The Learning Analytics Cycle: Closing the Loop Effectively. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 134–138. https://doi.org/10.1145/2330601.2330636
7. de Freitas, S., Gibson, D., Du Plessis, C., Halloran, P., Williams, E., Ambrose, M., Dunwell, I., & Arnab, S. (2015). Foundations of dynamic learning analytics: Using university student data to increase retention. *British Journal of Educational Technology*, *46*(6), 1175–1188. https://doi.org/10.1111/bjet.12212
8. Elias, T. (2011). *Learning Analytics: Definitions, Processes and Potential*.
9. Few, S. (2006). *Information Dashboard Design: The Effective Visual Communication of Data*. O'Reilly Media, Incorporated.
10. Florea, D., & Florea, S. (2020). Big Data and the Ethical Implications of Data Privacy in Higher Education Research. *Sustainability*, *12*(20). https://doi.org/10.3390/su12208744
11. Hout, N. (2020, July 6). *10 Accurate Training Metrics To Always Include In Your Learning Analytics Report*. https://elearningindustry.com/training-metrics-to-include-in-learning-analytics-report
12. Iliinsky, N., & Steele, J. (2011). *Designing Data Visualizations: Representing Informational Relationships 1st Edition* (1st edition). O'Reilly Media; 1st edition. https://www.amazon.com/Designing-Data-Visualizations-Informational-Relationships/dp/1449312284
13. Kane, M. A., & Rosas, S. R. (2017). *Conversations About Group Concept Mapping: Applications, Examples, and Enhancements* (Paperback). SAGE Publications, Inc. https://lead.to/amazon/com/?op=bt&la=en&cu=usd&key=1506329187

14. Khalil, M., & Ebner, M. (2016, March). *What is Learning Analytics about? A Survey of Different Methods Used in 2013-2015.*
15. Khan, M., & Khan, S. S. (2011). Data and Information Visualization Methods, and Interactive Mechanisms: A Survey. *International Journal of Computer Applications*, *34*(1), 1–14.
16. Lewis, C. (1982). *Using the "thinking-aloud" method in cognitive interface design.* IBM Thomas J. Watson Research Division. https://books.google.bg/books?id=F5AKHQAACAAJ
17. Mediratta, A. (2015). Big Data: Terms, definitions, and applications. *EMC Proven Professional Knowledge Sharing*.
18. Miteva, D., & Stefanova, E. (2020a). Design of Learning Analytics Tool: The Experts' Eyes View. *Proceedings of the 12th International Conference on Computer Supported Education - Volume 2: CSEDU,* 307–314. https://doi.org/10.5220/0009395503070314
19. Miteva, D., & Stefanova, E. (2020b). Experts' perspective for selecting learning analytics visualizations. *2020 International Conference Automatics and Informatics (ICAI)*, 1–6. https://doi.org/10.1109/ICAI50593.2020.9311341
20. Pauwels, K., Ambler, T., Clark, B. H., LaPointe, P., Reibstein, D., Skiera, B., Wierenga, B., & Wiesel, T. (2009). Dashboards as a Service: Why, What, How, and What Research Is Needed? *Journal of Service Research*, *12*(2), 175–189. https://doi.org/10.1177/1094670509344213
21. Rouse, W. B., Pennock, M. J., Oghbaie, M., & Liu, C. (2017). Interactive visualizations for decision support: Application of Rasmussen's abstraction-aggregation hierarchy. *Applied Ergonomics*, *59*(Pt B), 541–553. https://doi.org/10.1016/j.apergo.2016.03.006
22. Salihoun, M. (2020). State of Art of Data Mining and Learning Analytics Tools in Higher Education. *International Journal of Emerging Technology in Learning*, *15*(21). https://www.learntechlib.org/p/218354/
23. Schwendimann, B. A., Rodríguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., & Dillenbourg, P. (2017). Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research. *IEEE Transactions on Learning Technologies*, *10*(1), 30–41. https://doi.org/10.1109/TLT.2016.2599522
24. Shi, C., Fu, S., Chen, Q., & Qu, H. (2014). VisMOOC: Visualizing video clickstream data from massive open online courses. *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, 277–278. https://doi.org/10.1109/VAST.2014.7042528
25. Siemens, G. (2010). What are Learning Analytics? [Internet Archive]. *ELEARNSPACE*. https://web.archive.org/web/20180630115729/http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics
26. Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, *57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851
27. Siemens, G., Gašević, D., Haythornthwaite, C., Dawson, S., Buckingham Shum, S., Ferguson, R., Duval, E., Verbert, K., & Baker, R. S. (2011). *Open Learning Analytics: An integrated modularized platform* [WorkingPaper].
28. Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education*, *122*, 119–135. https://doi.org/10.1016/j.compedu.2018.03.018
29. Wang, L., Wang, G., & Alexander, C. A. (2015). Big Data and Visualization: Methods, Challenges and Technology Progress. *Digital Technologies*, *1*(1), 33–38. https://doi.org/10.12691/dt-1-1-7
30. Wang, R. (2012). *Monday's Musings: Beyond The Three V's of Big Data—Viscosity and Virality*.
31. Whitehouse, A. J. O., Maybery, M. T., & Durkin, K. (2006). The development of the picture-superiority effect. *British Journal of Developmental Psychology*, *24*(4), 767–773. https://doi.org/10.1348/026151005X74153
32. Митева, Д., & Стефанова, Е. (2018). *Визуализиране на учебни аналитики за по-ефективно обучение*. 143–150.