IEEE

mipro 2017

ISSN 1847-3938

organizer upro

40th

jubilee international convention

May 22-26, 2017, Opatija – Adriatic Coast, Croatia

Lampadem tradere

mipro - path to knowledge and innovation

mipro proceedings

MIPRO 2017

40th Jubilee International Convention

May 22 – 26, 2017
Opatija, Croatia

Proceedings

Conferences:

Microelectronics, Electronics and Electronic Technology /MEET

Distributed Computing, Visualization and Biomedical Engineering /DC VIS

Dew Computing /DEWCOM

Telecommunications & Information /CTI

Special Session on Modeling System Behaviour /MSB

Computers in Education /CE

Computers in Technical Systems /CTS

Intelligent Systems /CIS

Information Systems Security /ISS

Business Intelligence Systems /miproBIS

Digital Economy and Government, Local Government, Public Services / DE-GLGPS

MIPRO Junior - Student Papers /SP

Edited by:
Petar Biljanović

International Program Committee

Petar Biljanović, General Chair, Croatia
S. Amon, Slovenia
V. Anđelić, Croatia
M.E. Auer, Austria
S. Babić, Croatia
A. Badnjević, Bosnia and Herzegovina
M. Baranović, Croatia
B. Bebel, Poland
L. Bellatreche, France
E. Brenner, Austria
G. Brunetti, Italy
A. Budin, Croatia
Ž. Butković, Croatia
Ž. Car, Croatia
M. Colnarič, Slovenia
A. Cuzzocrea, Italy
M. Čičin-Šain, Croatia
M. Čupić, Croatia
M. Delimar, Croatia
T. Eavis, Canada
M. Ferrari, Italy
B. Fetaji, Macedonia
R. Filjar, Croatia
T. Galinac Grbac, Croatia
P. Garza, Italy
L. Gavrilovska, Macedonia
M. Golfarelli, Italy
S. Golubić, Croatia
F. Gregoretti, Italy
S. Groš, Croatia
N. Guid, Slovenia
J. Henno, Estonia
L. Hluchy, Slovakia
V. Hudek, Croatia
Ž. Hutinski, Croatia
M. Ivanda, Croatia
H. Jaakkola, Finland
L. Jelenković, Croatia
D. Jevtić, Croatia
R. Jones, Switzerland
P. Kacsuk, Hungary
A. Karaivanova, Bulgaria
M. Koričić, Croatia
T. Kosanović, Croatia
M. Mauher, Croatia
I. Mekjavić, Slovenia
B. Mikac, Croatia
V. Milutinović, Serbia
N. Mišković, Croatia
V. Mrvoš, Croatia
J.F. Novak, Croatia
J. Pardillo, Spain
N. Pavešić, Slovenia
V. Peršić, Croatia
S. Ribarić, Croatia
J. Rozman, Slovenia
K. Skala, Croatia
I. Sluganović, Croatia
M. Spremić, Croatia
V. Sruk, Croatia
S. Stafisso, Italy
U. Stanič, Slovenia
N. Stojadinović, Serbia
M. Stupičić, Croatia
J. Sunde, Australia
A. Szabo, IEEE Croatia Section
L. Szirmay-Kalos, Hungary
D. Šimunić, Croatia
Z. Šimunić, Croatia
D. Škvorc, Croatia
A. Teixeira, Portugal
E. Tijan, Croatia
A.M. Tjoa, Austria
R. Trobec, Slovenia
S. Uran, Croatia
T. Vámos, Hungary
M. Varga, Croatia
M. Vidas-Bubanja, Serbia
M. Vranić, Croatia
B. Vrdoljak, Croatia
D. Zazula, Slovenia

The Government of the Republic of Croatia is a Patron of the convention

organized by
MIPRO Croatian Society

technical cosponsorship
IEEE Region 8

under the auspices of
Ministry of Science and Education of the Republic of Croatia
Ministry of the Sea, Transport and Infrastructure of the Republic of Croatia
Ministry of Economy, Entrepreneurship and Crafts of the Republic of Croatia
Ministry of Public Administration of the Republic of Croatia
Central State Office for the Development of Digital Society
Croatian Chamber of Economy
Primorje-Gorski Kotar County
City of Rijeka
City of Opatija
Croatian Regulatory Authority for Network Industries
Croatian Power Exchange - CROPEX

patrons
University of Zagreb, Croatia
University of Rijeka, Croatia
IEEE Croatia Section
IEEE Croatia Section Computer Chapter
IEEE Croatia Section Electron Devices/Solid-State Circuits Joint Chapter
IEEE Croatia Section Education Chapter
IEEE Croatia Section Communications Chapter
T-Croatian Telecom, Zagreb, Croatia
Ericsson Nikola Tesla, Zagreb, Croatia
Končar - Electrical Industries, Zagreb, Croatia
HEP - Croatian Electricity Company, Zagreb, Croatia
VIPnet, Zagreb, Croatia
University of Zagreb, Faculty of Electrical Engineering and Computing, Croatia
Ruđer Bošković Institute, Zagreb, Croatia
University of Rijeka, Faculty of Maritime Studies, Croatia
University of Rijeka, Faculty of Engineering, Croatia
University of Rijeka, Faculty of Economics, Croatia
University of Zagreb, Faculty of Organization and Informatics, Varaždin, Croatia
University of Rijeka, Faculty of Tourism and Hospitality Management, Opatija, Croatia
University of Applied Sciences, Croatia
EuroCloud Croatia
Croatian Regulatory Authority for Network Industries, Zagreb, Croatia
Selmet, Zagreb, Croatia
CISEx, Zagreb, Croatia
Kermas energija, Zagreb, Croatia
Business Center Silos, Rijeka, Croatia
River Publishers, Aalborg, Denmark

general sponsor
HEP - Croatian Electricity Company, Zagreb, Croatia

sponsors
Ericsson Nikola Tesla, Zagreb, Croatia
Končar-Electrical Industries, Zagreb, Croatia
T-Croatian Telecom, Zagreb, Croatia
City of Opatija
InfoDom, Zagreb, Croatia
Hewlett Packard Croatia, Zagreb, Croatia
IN2, Zagreb, Croatia
King-ICT, Zagreb, Croatia
Storm Computers, Zagreb,
Croatia Transmitters and Communications Company, Zagreb, Croatia
VIPnet, Zagreb, Croatia
Danieli Automation, Buttrio, Italy
Mjerne tehnologije, Zagreb, Croatia
Selmet, Zagreb, Croatia
Institute SDT, Ljubljana, Slovenia
Nomen, Rijeka, Croatia
EuroCloud, Croatia

donor
Erste&Steiermärkische bank, Rijeka, Croatia

All papers are published in their original form

For Publisher:

Petar Biljanović

Publisher:

Croatian Society for Information and Communication Technology,
Electronics and Microelectronics - **MIPRO**
Office: Kružna 8/II, P. O. Box 303, HR-51001 Rijeka, Croatia
Phone/Fax: (+385) 51 423 984

Printed by:

GRAFIK, Rijeka

ISBN 978-953-233-093-9

Moodle-based data mining potentials of MOOC systems at the University of Szeged

G. Kőrösi*, F. Havasi**
* University of Szeged, Institute of Informatics, Hungary
** University of Szeged, Department of Software Engineering, Hungary
*korosig@inf.u-szeged.hu, **havasi@inf.u-szeged.hu

Abstract - In today's world virtual online educational platforms emerge literally on daily bases and many offer MOOC-based courses. With the appearance of MOOC, educational platforms have gained an additional boost, a new aspect in their evolutionary process, which has opened a new field of research thanking to the extraction of logging information within the frames of data mining. It has become clear that educators will be able to tailor their courses by merging the two previously mentioned fields and by carrying out MOOC-based data mining, targeting pedagogical aspects. This field of research seems promising and important, thus a faculty at the University of Szeged has created its own MOOC educational platform which has been set to facilitate data mining by implementing a wide range of logging algorithms. The data would be processed through a complex Artificial Intelligence program, which, in the short term, could reveal new and exciting pedagogical findings, while in the long run, the supervisors could put together a platform that would help and notify educators about relevant information. It would become possible to create adaptive educational materials, as well. This work aims at clarifying how such platforms function and what the steps of data collection and evaluation are.

I. Introduction

In today's world virtual online educational platforms emerge literally on daily bases and many offer MOOC-based courses. Mushrooming as a scalable lifelong learning paradigm, Massive Open Online Courses (MOOCs) have enjoyed significant limelight in recent years, both in industry and academia [10]. With the appearance of MOOC, educational platforms have gained an additional boost, a new aspect in their evolutionary process, which has opened a new field of research thanks to the extraction of logging information within the frames of data mining. Plenty of research at various institutions in the last 8 years seem more than promising with a significant breakthrough. Its origin dates back to the Educational Data Mining conference in 2008, where the idea of educational data mining of MOOC courses first emerged. All the findings since then have been published in the form of research papers and scientific journals underpinning that it is worth digging even deeper.

It might be asked what the reasons are that brought this interdisciplinary field of sciences to life. The main reason is that MOOCs are open education platforms, in which the participants are self-motivated to complete courses [24]. However, learning outside the framework of an educational institution and the supervision of a teacher may bring about certain obstacles. Learning in a MOOC requires that students apply self-regulation [23]. Among the most debatable issues are the high drop-out rates, this has been proven by dozens of researches. [3][7][9][11][13][23][25]. From this standpoint, one could easily question the existence of MOOC courses; nevertheless, there is pressure in higher educational institutions to provide up-to-date information to achieve institutional effectiveness [20]. Online platforms are among the most important tools to gain insight into how online education functions.

Ironically, the problem can be solved through the online platform itself because its structure allows all around logging of student activities, which may lead us to some so far unknown tools of pedagogical research. This idea has been grounded in the investigation of Romero & Ventura who think that learning management systems accumulate a great deal of log data about students' activities [20]. The system can automatically record whatever student activities are involved, such as reading, writing, taking tests, performing various tasks, and even communicating with peers [26]. In short, MOOC big data is a gold mine for analytics [18]. On the other hand, data mining technology has been proved effective in CMS pedagogical research as well [13].

Science that deals with verifying such data is called EDM (Educational Data Mining), with many prominent names of the field and outstanding research achievements. Berland et. al suggest that EDM may have the potential to support research that is meaningful and useful both to researchers working actively in the constructionist tradition but also to wider communities [16]. Data collected from learning systems can be aggregated over large numbers of students and can contain many variables that data mining algorithms and techniques can explore for model building. [22] Working from student data can help educators both track academic progress and understand which instructional practices are effective [5].

Educational data mining (EDM) is a research area which utilizes data mining techniques and research approaches for understanding how students learn [22]. In recent years, there has been an increasing interest in the use of data mining to investigate scientific questions within educational research, an area of inquiry termed educational data mining [2]. The scope of educational data mining includes areas that directly impact students [19]. The emerging field of educational data mining (EDM) examines the unique ways of applying data mining methods to solve educationally related problems [19].

It must be stated that the primary goal of research is not just to obtain information but to keep as many students as possible signed up to our courses. Through his investigation into the relevant papers, Huebner reveals works that suggest how learners can be kept in the learning environment, efficient educational techniques, and better course books in the future which may help reducing the drop-out rate in a predictive way [19]. Along this line of thought, researchers at Bowie State University have assigned risk factor points to each learner which demonstrated who would have difficulties [6].

The authors' research takes this route to test e-learning platforms by putting together two MOOC courses (Conscious and Safe Internet Usage - TÉBIA, Database Management) a logging platform to register online activities. This paper has been written to investigate EDM opportunities and to develop the authors' own logging system dealing with how the steps of data recording, cleaning, and pre-processing are done.

II. What is EDM and why is it important?

While thousands of students have been attracted to large online classes, keeping them motivated has become a challenging endeavor [13]. Thus, it is of paramount importance to understand student motivation or why it is lost. A tool to gain access to such answers in Data Mining. In order to find out what this notion is one may quote Baker who states that Data mining, also called Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from large amounts of data [2]. Knowing that EDM has existed for only a decade, it is advisable to take a close look at it to reveal what it really means. Its meaning depends on how it is defined, however a common meeting point has been established, which seems to be digital education. Educational data mining is a research area which utilizes data mining techniques and research approaches for understanding how students learn [22]. EDM is an emerging tool and technique used to comprehend and represent educationally related data [19]. Furthermore, data mining is a series of tools and techniques for uncovering hidden patterns and relationships among data [27]. Data mining is a multidisciplinary area in which several computing paradigms converge: decision tree construction, rule induction, artificial neural networks, instance-based learning, Bayesian learning, logic programming, statistical algorithms, etc. And some of the most useful data mining tasks and methods are: statistics, visualization, clustering, classification, association rule mining, sequential pattern mining, text mining, etc. [21]. Educational data mining is an emerging discipline that focuses on applying data mining tools and techniques to educationally related data [1]. A large number of researchers within EDM focus directly on course management systems and how they can be improved to support student learning outcomes and student success [19]. Data mining is the process of efficient discovery of non-obvious valuable patterns from a large collection of data [12].

Interactive e-learning methods and tools have opened up opportunities to collect and scrutinize student data, to ascertain patterns and trends in those data, and to formulate new discoveries and test assumptions about how students learn [22]. Researchers have found that they can apply data mining to rich educational data sets that come from course management systems such as Angel, Blackboard, WebCT, and Moodle. Numerous studies have shown that data mining can be used to discover at-risk students and help institutions become much more proactive in identifying and responding to those students [14]. Educational data mining is defined as the area of scientific inquiry centered around the development of methods for making discoveries within the unique kinds of data that come from educational settings, and using those methods to better understand students and the settings which they learn in [2]. Online learning systems log student data that can be mined to detect student behaviors that correlate with learning [22].

Four main axes can be identified along which EDM methods may be helpful for constructionist research:

- EDM methods do not require constructionists to abandon deep qualitative analysis for simplistic summative or confirmatory quantitative analysis;
- EDM methods can generate different and complementary new analyses to support qualitative research;
- By enabling precise formative assessments of complex constructs, EDM methods can support an increase in methodological rigor and replicability;
- EDM can be used to present comprehensible and actionable data to learners and teachers in situ.
- In order to investigate those axes, the first step is to describe one's perspective on compatibilities and incompatibilities between constructionism and EDM [16].

The strengths of EDM systems can be traced back to their tools, primarily logging methods that provide information to researchers, who would in turn reveal so far unknown pedagogical conclusions. Baker sums up (Table I.) what is known about those tools and result [2].

TABLE I. THE PRIMARY CATEGORIES OF EDUCATIONAL DATA MINING

Category of Method	**Goal of Method**	**Key applications**
Prediction	Develop a model which can infer a single aspect of the data (predicted variable) from some combination of other aspects of the data (predictor variables)	Detecting student behaviors (e.g. gaming the system, off-task behavior, slipping); Developing domain models; Predicting and understanding student educational outcomes
Clustering	Find data points that naturally group together, splitting the full data set into a set of categories	Discovery of new student behavior patterns; Investigating similarities and differences between schools
Relationship Mining	Discover relationships between variables	Discovery of curricular associations in course sequences; Discovering which pedagogical strategies lead to more effective/robust learning
Discovery	A model of a	Discovery of relationships

with Models	phenomenon developed with prediction, clustering, or knowledge engineering, is used as a component in further prediction or relationship mining.	between student behaviors, and student characteristics or contextual variables; Analysis of research question across wide variety of contexts
Distillation of Data for Human Judgment	Data is distilled to enable a human to quickly identify or classify features of the data.	Human identification of patterns in student learning, behavior, or collaboration; Labeling data for use in later development of prediction model

III. E-LEARNING LOGGER MODULE

The system presented in this paper was built on the basis of Moodle which is an open-source, free, well supported, popular e-learning platform. It has a long history, given that the first version came out in 2002. The platform is known for its robustness, though its user interface is little less modern than it is expected in these days. This is the underlying reason for completely replacing its font-end and develop a new one, which calls the Moodle’s back-end. One module of its front-end is responsible for logging, which is the focus of this article.

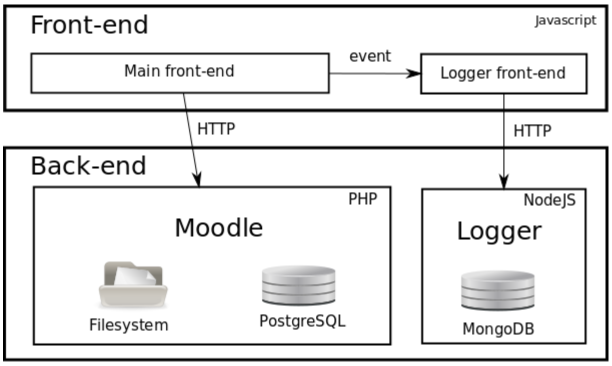


Figure 1. Front-end logging system

This logger front-end collects and process events, and calls its back-end part via HTTP to store them. This back-end part – which is completely independent from the back-end of the Moodle - is developed in NodeJS and uses MongoDB to store events (Figure 1.).

Every log entry is an event. Each of the events classified the system into different types, and depending on the type, they store different parameters for it. For example, a “textinput” event has a parameter, which stores the typed text, called text:

```
{
  type: "textinput",
  data: {
      target: "search-target",
      text: "database"
   },
  time: "2017.01.23. 16:01:28.242",
  page: "https://...",
  userid: 1876
}
```

There are some parameters, which are stored for all events:

- userid: the ID of the user, who has executed the operation, or 0, if there was an anonymous user
- time: the date of the event
- page: the URL of the page, where the event was happened
- type: the type of the event (see below)

The type of the events, and their parameters can be the following ones:

- load, unload, focus, blur: they are generated in the case of loading or unloading of the page, and getting and losing the focus.
- resize: it means resizing the browser window. It has two parameters: x, y (the new size of the windows).
- click: it represents a mouse click. Its parameters are x and y.
- testClick: it signifies a mouse click to an answer of a quiz. It is a preprocessed event: this javascript event handler automatically recognizes if the mouse click happened over an answer, and generates a testClick event, not a simple click event. Its parameters are: question, answer, correct, choiceCleared.
- download: it is generated in the case of downloading a file. Parameter: filename.
- textinput: this event represents a change of a text input field. Parameters: target (id of the text element), text (actual value of the text element).
- textinput_focus, textinput_blur: they are generated when a text input gains or loses the focus. Parameters: target (id of the text element), text (actual value of the text element).
- passwordinput, passwordinput_focus, passwordinput_blur: similar to the previous ones, but because of security consideration, the value of the password text input is not stored. Parameters: target (id of the text element), and length (of the text element).
- mousemove: mouse moving event. Parameters: x, y, xDistance, yDistance, realDistance. The system stores only two mouse events in a second.
- scroll: means scrolling the page. Parameter: top. The system stores only two mouse events in a second.

There are video events, as well. The system supports two kinds of video: html5 video element and embedded YouTube video. Events:

- videoSeek: means seeking in the html5 video element. Parameters: seekTime, videoId, totalTime, src.

- videoPlay, videoPause, fullscreenOn, fullscreenOff: html5 video playing events. Parameters: actualTime, videoId, totalTime, src
- volumeChange: html5 video element volume change. Parameters: actualTime, videoId, totalTime, src, newVolume.
- youtubePlay, youtubeEnd, youtubePause, youtubeBuffering: youtube video playing events. Parameters: actualTime, videoId, totalTime, src.
- youtubeQuality. changing youtube video quality settings. Parameters actualTime, videoId, totalTime, src, quality
- youtubeRate. parameters: actualTime, videoId, totalTime, src, rate

IV. DATA

Two courses have been created in order to test the logging platform.. In the first part of the research, a pilot study was conducted between the dates of March 1 and May 30, 2016, while the second study was recorded in the interval of October 1 to December 10, 2016. Altogether 163 students took part in the pilot study and 347 student signed up for the two courses in the Autumn semester. The details about the course are presented in Table II below. The learning material for both courses comprised a three week study period. One of the courses, which ran under the name 'TÉBIA,' included 4 + 1 (embedded) videos, while the other course, with the name 'Databases' had 7 (embedded/Youtube links) videos with attached embedded texts, or external links.

The primary point of interest for the researchers lay not in the drop-out rate, instead the aim was to discern how the platform functioned and how the learners would behave. It can thus be concluded that 99.8% of the learners who had signed up for the course, had also completed it.

TABLE II. COURSE CONTENTS

Course name	TÉBIA	Databases
Content	Basics of Conscious and Safe Internet Usage	Basics of Databases
Time frame	3 weeks	3 weeks
Parts of the Learning Material	Introduction: Video (3.37 min., Embed);	The concepts of databases: Video (2.12 min, Embed); Video (2.12 min, Youtube link); HTML embedded text;
	Digital footprint: Video (14.04 min, Embed); HTML embedded text;	Database handling systems: Video (3.52 min, Embed); Video (3.52 min, Youtube link); HTML embedded text; HTML embedded text;
	Conscious and Safe Internet Usage: Video (13.07 min, Embed); HTML embedded text; External link;	Transaction, closing methods: Video (3.28 min, Embed); Video (3.28 min, Youtube link); HTML embedded text;
	Online bullying: Video(13.31 min, Embed); HTML embedded text;; Extra video (11.55 min, Embed);	Basics of handling databases: Video (3.24 min, Embed); Video (3.24 min, Youtube link); HTML embedded text;
		Database models: Video (2.51 min, Embed); Video (2.51 min, Youtube link); Video (4.11 min, Embed); Video (4.11 min, Youtube link); HTML embedded text;
		Relational database models: Video (3.17 min, Embed); Video (3.17 min, Youtube link); HTML embedded text;;

The logging system during the two courses registered 4.663.120 logs, out of which 26 variables were generated and assigned to the users. These were the following:

Data, Page, Pid, Time, Type, User, Data.realDistance, Data.x, Data.xDistance, Data.y, Data.yDistance,

Data.Text, Data.Top, Data.Target, Data.Filename, Data.Length, Data.ActualTime, Data.Scr, Data.TotalTime, Data.VideoId, Data.SeekTime, Data.NewVolume, Data.Ip Adress, Data.Quality, IP

V. PRE-PROCESSING

The e-learning platform of the University of Szeged is a website which provides a wide variety of services, including video lessons to every courses, however it does not have a complex platform such as Coursera or edX. This is the reason why is was not possible to close the e-learning portal after the pilot study and the two courses. Thus, as a consequence, the platform used for this research could not only be accessed by those students who had signed up, but it was accessible to but external users, as well. This led to the recording of 1,443,817 (404 Mb) logs, while during the second time the system had 3.219.303 (936 MB) recorded logs. The portal recorded

1229 students and lecturers, out of which only 513 were relevant. The contaminated raw data had to be put to serious data cleaning procedures. The file was cleaned and sorted out by examining students' IDs, user behavior, and sign-in tendencies. Table III demonstrates a simple but effective algorithm that filters relevant and usable data:

TABLE III. USER BEHAVIOUR, AND SIGN-IN TENDENCIES

Dataset	Lectures	Lecture Videos	Video Length (min)	Quizzes	Users	Clickstream Events
Pilot	11	11	65.74	3	163	1,443,817
Autumn	11	11	65.74	1	347	3.219.303

A deeper study of this JSON file revealed further important information. This includes the user activities showing that 54% of the logs were generated by 10% of the users. The comprehensive diagram of user activities shows the distribution of active and passive participants (Figure 2).

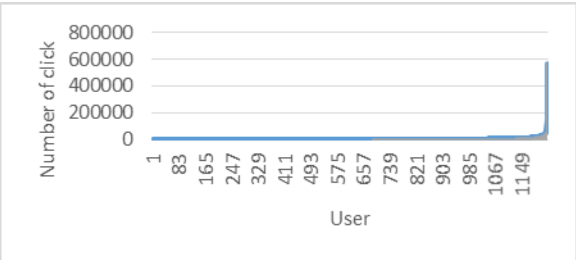
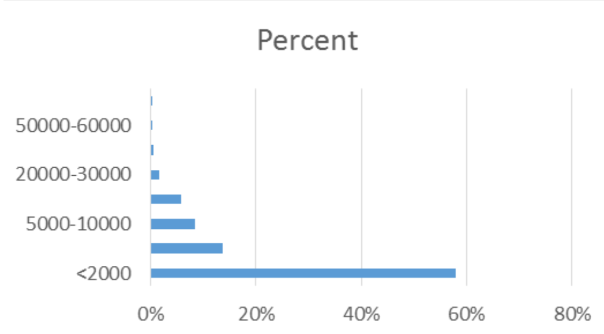


Figure 2. Number of clicks by user

This fact is not surprising since one of the side effects of a MOOC course is the uneven distribution of active and passive students. Anderson et al. created 5 categories [28]:

- Viewers, in the left mode of the plot, primarily watch lectures, handing in few if any assignments.
- Solvers, in the right mode, primarily hand in assignments for a grade, viewing few if any lectures.
- All-rounders, in the middle mode, balance the watching of lectures with the handing in of assignments.
- Collectors, also in the left mode of the plot, primarily download lectures, handing in few assignments, if any. Unlike the Viewers, they may or may not be actually watching the lectures.
- Bystanders registered for the course but their total activity is below a very low threshold.

If the student activities and other data are converted to percentages, one would gain deeper knowledge in this topic. Figure 3 shows these findings. In order to complete a test, an average user generates approximately 400-2000 log files, which demonstrates that 58% of the learners do not aim at having a thorough understanding of the material but want to complete to course as soon as possible, while only 10% can be categorized as superusers according to Zhu et al.[24].



	<2000	2000-5000	5000-10000	10000-20000	20000-30000	30000-40000	50000-60000	60000-70000
Percent	58%	14%	8.30%	5.70%	1.63%	0.49%	0.24%	0.24%

Figure 3. Student activities/ logging data

VI. CONLUSION

At this stage of the studies, the authors were able to add a functioning logging system to a Moodle platform with weak tools of analyses, which would serve well for similar portals to live up to current measuring requirements. The data obtained and analyzed from students' logging could reveal some unexpected pedagogical aspects thus helping educators and learners in the process of course planning and learning. The analysis of the pile of data amounting to millions of logs recorded during the pilot studies and the autumn courses could help the authors design an artificial intelligence (machine learning) that would automatically process input data without human intervention and which could intervene if extreme values emerge.

The aim of the redesigned and modified website is to enhance student motivation, learning achievements, and output results. After examining the relevant literature, the authors were able to sort out errors and potential opportunities that were unknown to them, like predictive analysis through clickstream [4], [7], [8], [15], [17] or feedback buttons by Chang et al. [11] which improved student concentration in the long run. A total of 1.2 GB of data was collected, enabling the authors to make the next step of designing a suitable mathematical model to their MOOC system. This would help to provide a full-scale predictive background support to educators who would upload their learning materials, and would help learners who sign up to a course. It is expected that the spring semester of 2017 will bring more users to this platform, which would double the amount of collected data. After quantitative and qualitative data analysis the new findings will also be published in a research paper.

REFERENCES

[1] R. Baker, K. Yacef, The State of Educational Data mining in 2009: A Review and Future Visions, Journal of Educational Data Mining vol. 1, no. 1, pp. 1–14, fall 2009.

[2] R. Baker, Data Mining for Education. In McGaw, B., Peterson, P., Baker, E. (Eds.) International Encyclopedia of Education (3rd edition), vol. 7, pp. 112-118. Oxford, UK: Elsevier, 2010.

[3] Y. Bergner, S. Droschler, G. Kortemeyer, S. Rayyan, D Seaton,D. Pritchard, E Model-Based Collaborative Filtering Analysis of Student Response Data: Machine-Learning Item Response Theory. International Educational Data Mining Society, Paper presented at the International Conference on Educational Data Mining (EDM), pp. 95–102 (5th, Chania, Greece, Jun 19-21, 2012)

[4] C. G. Brinton, R. Rill, S. Ha, M. Chiang, R. Smith, and W. Ju, “Individualization for Education at Scale: MIIC Design and Preliminary Evaluation,” Transactions on Learning Technologies, vol. 8, no. 1, pp. 136–148, 2015.

[5] L. Cao, B. RAMESH, M. ROSSI, Are Domain Specific Models Easier to Maintain Than UML Models? IEEE Software, vol. 26, pp. 19–21, 2009.

[6] F. Chacon, D. Spicer, A. Valbuena, Analytics in Support of Student Retention and Success (Research Bulletin 3, 2012 ed.). Louisville, CO: Educause Center for Applied Research, pp. 1–9, 2012.

[7] C. G. Brinton ,Mung Chiang, Mooc performance prediction via clickstream data and social learning networks, In: Computer Communications (INFOCOM), 2015 IEEE Conference on Computer Communications, IEEE, pp. 2299–2307, 2015.

[8] C. G. Brinton, S. Buccapatnam, M. Chiang, H. V. Poor, "Mining MOOC Clickstreams: Video-Watching Behavior vs. In-Video Quiz Performance", Signal Processing IEEE Transactions on, vol. 64, pp. 3677–3692, 2016.

[9] J. Guan, W. Nunez, J. Welsh, Institutional strategy and information support: the role of data warehousing in higher education, Campus-Wide Information Systems, vol. 19, no. 5, pp.168–174, 2002.

[10] S. Haggard, S. Brown, R. Mills, A. Tait, S. Warburton, W. Lawton, T. Angulo, Haggard S, The maturing of the MOOC: Literature review of massive open online courses and other forms of online distance learning. Department for Business, Innovation and Skills, UK Government. 2013 Sep 20.

[11] J. Cheng, C. Kulkarni, S. Klemmer, Tools for predicting drop-off in large online classes. In: Proceedings of the 2013 conference on Computer supported cooperative work companion. ACM, pp. 121–124, 2013.

[12] W. Klosgen, J. Zytkow, M. Jan, Knowledge discovery in databases: the purpose, necessity, and challenges. In: Handbook of data mining and knowledge discovery. Oxford University Press, Inc., pp. 1–9, 2002.

[13] D. Liang, J. Jia, X. Wu, J. Miao, J. Wang, Analysis of learners’ behaviors and learning outcomes in a massive open online course. Knowledge Management & E-Learning, vol. 6, no. 3, pp. 281–298, 2014.

[14] J. Luan, Data Mining and Knowledge Management in Higher Education – Potential Applications. Paper presented at the Annual Forum for the Association for Institutional Research, Toronto, Ontario, Canada, pp. 1–13, 2002.

[15] M. Speiser, G. Antonini, A. Labbi, J. Sutanto, On nested palindromes in clickstream data, KDD '12 Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining p 1460-1468, 2012.

[16] M. Berland, R. S. Baker, and P. Blikstein, Educational Data Mining and Learning Analytics: Applications to Constructionist Research, Technology, Knowledge and Learning, July 2014, vol. 19, no. 1, pp. 205–220, 2014.

[17] R. Mazza and C. Milani, Exploring usage analysis in learning systems: Gaining insights from visualisations. In Workshop on Usage analysis in learning systems at 12th International Conference on Artificial Intelligence in Education, New York, USA, pp. 1–6, 2005.

[18] U. O'Reilly and K.Veeramachaneni, Technology for Mining the Big Data of MOOCs, Research & Practice in Assessment, vol. 9, pp. 29–37, 2014.

[19] R. A. Huebner, A Survey of Educational Data-Mining Research, Research in Higher Education Journal, vol. 19, pp. 1–19, April 2013.

[20] C. Romero and S.Ventura, Educational Data Mining: A Review of the State of the Art. Systems, Man, and Cybernetics Part C: Applications and Reviews, IEEE Transactionson, vol. 40, no. 6, pp. 601–618, 2010, doi: 10.1109/tsmcc.2010.2053532.

[21] C. Romero, S. Ventura, and E. García, Data mining in course management systems: Moodle case study and tutorial. Computers & Education, vol. 51, no. 1, pp. 368–384, 2007, doi: 10.1016/j.compedu.2007.05.016.

[22] S. Lakshmi Prabha and A.R.Mohamed Shanavas, Educational data mining applications ,Operations Research and Applications: An International Journal (ORAJ), vol. 1, no. 1, pp. 1–6, August 2014.

[23] T. Sinha, P. Jermann, N. Li and P. Dillenbourg, Your click decides your fate: Inferring Information Processing and Attrition Behavior from MOOC Video Clickstream interactions, EMNLP Workshop on Modelling Large Scale Social Interaction in Massive Open Online Courses, pp. 3–14, 2014.

[24] T. Zhu, W. Wang, W. Zhao and Riming Zhang, Participation Prediction and Opinion Formation in MOOC Discussion Forum, International Journal of Information and Education Technology, vol. 7, no. 6, pp. 417–423, June 2017.

[25] P. Esztelecki, G. Korosi, N. Namesztovszki, L. Major, The comparison of impact offline and online presentation on student achievements: A case study, 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2016, pp. 802–806, DOI: 10.1109/MIPRO.2016.7522249 IEEE Conference Publications

[26] J. Mostow, J. Beck, H. Cen, A. Cuneo, E. Gouvea and C. Heiner, An educational data mining tool to browse tutor–student interactions: Time will tell! In Proceedings of the workshop on educational data mining, pp. 15–22, 2005.

[27] M.H. Dunham, Data mining introductory and advanced topics. Upper SaddleRiver, NJ: Pearson Education, Inc., 2003.

[28] A. Anderson, D. Huttenlocher, J. Kleinberg and J. Leskovec (Engaging with Massive Online Courses, WWW '14 Proceedings of the 23rd international conference on World wide web, pp. 687–698, 2014.