Assessment of the University Digital Ecosystem through Intelligent Analysis of Open-Ended Survey Responses

Aizhan Tlebaldinova, Natalya Denissova, Irina Dyomina, Ruslan Chettykbayev

D. Serikbayev East Kazakhstan Technical University
Digital Technologies and Artificial Intelligence School
D. Serikbayev Str. 19, 070004 Ust-Kamenogorsk, Kazakhstan;
e-mail: {atlebaldinova, ndenisova, idyomina, rchettykbayev}@edu.ektu.kz

Markhaba Karmenova

Sarsen Amanzholov East Kazakhstan University
Higher school IT and Natural Sciences
30th Guards Division Str. 34, 070004, Ust-Kamenogorsk, Kazakhstan;
e-mail: mkarmenova@vku.edu.kz

György Györök

Óbuda University, Alba Regia Technical Faculty, Budai út 45, H-8000 Székesfehérvár, Hungary; e-mail: gyorok.gyorgy@amk.uni-obuda.hu

Abstract: The use of open-ended questions in surveys for conducting research in various fields of activity allows you to obtain detailed and in-depth information that can be used to make important decisions. However, automatic processing of open questions is a complex task that requires an integrated approach. This study proposes a methodology based on sentiment analysis, topic modeling, and text networks to extract useful data on satisfaction with the university's digital ecosystem and the current situation. The methodology allows you to get a more complete picture of the level of satisfaction of students, lecturers and other participants in the university's digital ecosystem, as well as identify problematic issues and offer specific recommendations for their improvement. The main contribution of the article is the use of algorithms for modeling topics and network structures to complement the results of the analysis of the tone of the text. The methodology includes Data Gathering, Text Processing, Text Mining and Recommendations. The proposed methodology is applicable not only to the educational field, but can also be replicated for automatic processing of unstructured data in other areas.

Keywords: assessment of the digital ecosystem; open-ended questions; data mining; text mining; topic modeling; terms; network modeling

Introduction

Recently, the modern educational society has been undergoing various transformations. The active introduction of new learning solutions and technologies, including IT technologies, into the educational process, in turn, demonstrated significant changes in the field of higher education. Such changes also imply the development of digital technologies that contribute to increasing the level and support of teaching and learning processes, management processes of the entire digital ecosystem of higher education institutions (hereinafter referred to as universities) [1].

Today, many universities are striving to invest in digital technologies in order to provide high-quality services for learning, facilitate and transform teaching processes [2], as well as to increase the level of involvement of both lecturers and students in the educational process. In this regard, the need to assess and analyze the quality of services provided by universities for students makes it possible to identify important aspects and existing problems in the educational, organizational, administrative and other sectors of the digital ecosystem [3]. As is known, in practice, most researchers use various methods and technologies to evaluate the work of the existing digital system of universities, among which studies aimed at establishing feedback with subjects (lecturers, students, staff) of educational activities give good results and objective assessments. So, for example, conducting surveys, questionnaires [4]-[7] among students, lecturers and other staff are an effective way to organize data collection for further research evaluation of the university's work. The application of textual analysis methods and technologies to the collected data currently allows for a thorough and comprehensive study. The purpose of this article is to assess the digital ecosystem of the university through the use of intelligent analysis of open-type responses. Thus, this study is aimed at determining the main strategies for the formation and development of the university's digital ecosystem, as well as assessing its current state.

1 Related Works

High-quality education in any country depends on many well-known factors, among which the digital ecosystem plays an important role at different levels of education. In our case, the digital ecosystem of the university is considered, which provides various communication tools for all subjects of education and stakeholders, thereby meeting their needs related to educational, research and innovation activities.

The processes related to improving the quality of education, the introduction of new forms and technologies of education, allow universities to remain competitive in the market of services provided among many higher educational institutions. In this regard, receiving feedback from participants in the educational environment in the form of regular monitoring, polling, questionnaires, analysis and evaluation of accumulated data, make it possible to introduce new solutions and develop the digital ecosystem of universities. So, in [8], the authors present a study in which they analyze text data collected from university social media sites in order to assess their competitiveness. The study used data from five international universities and data from five local private universities (Lahore, Pakistan). As a research problem, the authors consider obtaining and disclosing answers to such questions as the use of social networks and its impact on attracting potential students. The data analysis uses an approach combining statistical analysis, sentiment analysis and text analysis. The application of text analysis based on comments showed the frequency of use of words such as "RIU", "Riphah", "Lahore", "Program", "Admission", "apply", thereby showing the attractiveness for admission to RIU International University. Based on the analysis obtained, the authors offered their recommendations that will allow universities in general to promote their brand and increase their attractiveness to potential students using social networks. In [9], a two-stage methodology (quantitative and qualitative data analysis) is considered to assess lecturers' opinions and visions of existing problems and achievements related to the use of digital technologies for teaching and learning in Latin American universities (LATAM). At the heart of the methodology, researchers use Text Mining (sentiment and emotional valence analysis) to analyze the opinions of university lecturers with the task of identifying hidden problems, as well as the reasons influencing the use of digital technologies for teaching and learning. In the research phase, the authors used the Kruskal-Wallis H-test to quantify the data on the collected questions. This type of analysis revealed the problems associated with the lack of training, infrastructure and resources, access to the Internet and digital platforms in the teaching and learning process. At the research stage, for qualitative data analysis, the authors applied the Text mining analysis (sentiment/ emotional valence) to identify the most commonly used terms that describe the problems associated with the use of digital technologies for teaching and learning in Latin America. Thus, the results of the analysis confirmed the main factors associated with limited training and resources, access to the Internet and infrastructure, thereby reflecting the problems in terms of the level of introduction of digital technologies in the education of universities in Latin America.

Establishing feedback with participants and students based on collecting their opinions, feedback and comments on all components and structural elements of the educational process of universities is an important aspect for determining levels of satisfaction and competitiveness. In [10], an extended hybrid multilevel approach of thematic modeling is proposed for evaluating a textual dataset that is obtained as a result of establishing feedback with students of academic institutions. The proposed approach is based on Aspect-based sentiment analysis (ABSA). This type of data analysis identifies a set of aspect terms and meanings from a text fragment. The proposed Aspect2Labels (A2L) approach divides the data analysis system into three levels. The layers of the system extract terms of high and low-level aspects, comparisons of extracted terms and classifications are carried out. At the initial level, various LDA variants were used for the aspect extraction process. At subsequent levels, various machine learning algorithms were applied to the extracted aspects for the purpose of classification. The authors were able to obtain 97% and 93% accuracy on the test dataset using the support vector machine (SVM) and artificial neural networks (ANN). In [11], statistical classifier methods are also used to extract suggestions from qualitative student reviews for seven undergraduate courses, where the decision tree method (C5.0), according to experiments, showed a good result.

Currently, topic modeling is one of the new methods in Text mining, as it allows you to discover hidden knowledge in datasets and search for relationships between text documents. Thus, in [12], a methodology is described that is based on extracting hidden information from textual survey data in the field of teacher self-assessment using thematic modeling. The proposed methodology uses algorithms for modeling the text network and thematic modeling of LDA. The methodology under consideration is implemented in 4 stages: creation of a text database; text analysis and topic modeling; network modeling; relevance of identified topics. The use of clustering algorithms in the structure of the methodology made it possible to supplement the results obtained when performing thematic modeling. As a result of the research on the developed methodology, the authors present a collection of basic strategies for lecturers to use in the classroom with the aim of improving student retention is presented. In [13], the process of using the resource of answers to open feedback questions using the thematic modeling approach is also considered. Thematic models are created using the LDA method. Methods of qualitative and quantitative assessment are used to verify the results of the topic. The proposed process allows lecturers to analyze the quality of teaching at the program level or at the whole institution or in individual courses with a very large number of students. The use of text analysis and machine learning methods has shown good effectiveness in extracting useful topics from the mass of open student reviews. The study in [14] also presents a tool (Palaute) for analyzing written student feedback using thematic modeling and emotion analysis. In [15], the strengths and weaknesses, opportunities and threats affecting the university's activities are investigated and identified based on the analysis of online student reviews using text analytics. The study uses thematic modeling, sentiment analysis, root cause analysis and SWOT analysis, where a group of thematic models of Latent Dirichlet allocation (E-LDA) is also used for thematic modeling and the predominant key topics discussed by students are determined.

In [16], a study and assessment of students' views on how distance learning has affected their academic level of progress and success was conducted. The main objective of the study was to extract important issues from the opinions of students during online learning. In conducting data analysis, the authors used statistical data analysis and a text mining approach. As a result of the study, text mining revealed the positive aspects of online learning in the form of frequent pairs of words like "flexible location", "flexible schedule", "social distance", "park pass". The most common pair of words that have been used to describe negative online learning experiences include "human interact," "due date," "distance learn," "real-time," "class synchrony." In [17], the degree of satisfaction of students with a massive open online course (MOOC) is assessed. The course analysis used a supervised machine learning algorithm, sentiment analysis, and hierarchical linear modeling. In [18], the factors influencing the satisfaction of students with MOOCs using Latent Dirichlet allocation (LDA) are also investigated and a qualitative comparative analysis (QCA) with a fuzzy set is carried out to analyze configurations of high and low high levels of satisfaction with learning in the field of computer science. Similar studies to improve the MOOC learning experience have been conducted in [19]-[20], which also use latent and Latent Dirichlet allocation (LDA) models to identify important topics from the collected student responses. The application of the statistical method of Latent Dirichlet allocation (LDA) is also used in [21] to identify aspects of students' satisfaction with courses based on the collected survey data. In [22], the LDA model is also used to analyze the text of the survey results to assess the self-perception of IT students.

The study of issues related to the assessment of the quality of educational services provided, which also assesses the information and digital components of the entire educational space, in turn require comprehensive research. As such studies, it is possible to conduct a survey of respondents to identify cause-and-effect relationships between the parameters characterizing the changes associated with the development and implementation of new educational management mechanisms. In the above review of research papers, the authors mainly used methods related to text data analysis, classification, as well as thematic modeling. Many studies use latent and Latent Dirichlet allocation (LDA) models to identify important terms from the collected responses of respondents. An analysis of these papers shows that in most cases, these studies mainly limit their results to identifying important topics and interpreting them, without establishing links between certain topics. It should also be noted that most of them are dominated by quantitative analysis without in-depth interpretation [8, 10], samples may be unrepresentative, and the influence of factors on the educational process is not analysed [9, 16]. There is a lack of verification of the proposed models, and management recommendations are not sufficiently developed, which reduces the practical significance of the findings [10, 13]. Our research has developed an approach based on the consistent application of sentiment analysis, thematic and network modeling. This approach allows for a deeper analysis of textual data, identifying hidden topics and finding connections between objects based on their content.

2 Methodology

Improving the competitiveness and effective operation of a modern university is currently becoming one of the most urgent tasks of higher education. Today, many studies show that the digital transformation of the higher education system has a significant impact on solving such problems. A comprehensive analysis and assessment of the university's digital ecosystem reveals many aspects related to management vision, material and technical equipment, the level of application of modern communication tools, the level of introduction of modern technologies into the educational process, etc. Receiving feedback from the subjects of the educational environment make it possible to introduce new solutions and develop the university's digital ecosystem.

In this regard, this article examines and analyzes the collected data sets of respondents (students, teachers, staff) to assess the quality of the digital ecosystem of three universities in East Kazakhstan (S.Amanzholov East Kazakhstan University, Kazakh-American Free University and D.Serikbayev East Kazakhstan Technical University (hereinafter the University). In addition, as a result of the study, it was assumed to identify factors and existing problems affecting the further development of the university's digital ecosystem as a whole.

An online survey method was chosen to assess user satisfaction with the university's digital ecosystem. The online survey was compiled in Russian. 723 people took part in the survey (43% of respondents belong to the category of "Lecturers", 45% - "Students" and 12% - "Developers"), from three regional universities of the Republic of Kazakhstan. The preparation of the online survey questions was carried out taking into account the characteristics of these subjects of education. The survey consisted of 17 questions on the digital ecosystem of the university, divided into three blocks: the main factors taken into account in the formation and development of digital ecosystems of universities (9), difficulties and problems of ecosystem formation based on a digital platform (4), recommendations for the formation and development of a digital ecosystem (4).

In this study, the proposed assessment methodology user satisfaction with the university's digital ecosystem is based on the analysis of open-ended questions, the contents of which are presented in Table 1. The respondents' answers to other types of questions such as closed questions and Likert scale questions were analysed using statistical methods and the results of this study are presented in [23].

Table 1
The structure of the questions

	Questions			
	1	2	3	4
	The impact of the University's digital ecosystem on various aspects of teaching staff activities	Digital competencies of teaching staff	Digital educational resources	Recommendations for the formation and improvement of the University's digital ecosystem
Lecturers	+	+	+	+
Developers	-	-	-	+
Students	-	-	+	+
	5	6	7	
	Difficulties in acquiring software and hardware	Personnel shortage	Integration with government systems and databases	
Lecturers	-	-	-	
Developers	+	+	+	
Students	-	-	-	

The proposed methodology (Figure 1) includes data collection and purification, data preprocessing, sentiment analysis, thematic and network modeling, interpretation of the results in the form of recommendations and conclusions.

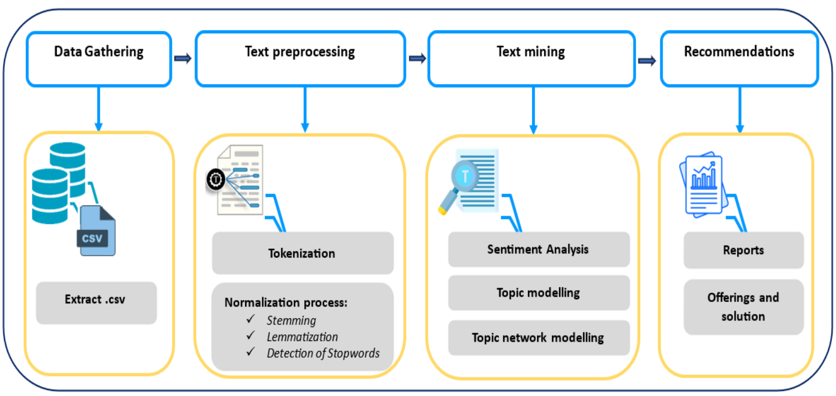

Figure 1
Research methodology

The first step describes the steps necessary to collect and clean up a text dataset. This means that after the collection procedure from the source data, it is necessary to carry out the process of removing irrelevant information, which further interferes with the analysis and distorts its results. The next stage includes the basic operations of processing a text dataset such as tokenization, normalization (lemmatization) and removal of stop words from the source text. All of the above operations are very important since the results of thematic modeling directly depend on how the preprocessing of texts was carried out [24]-[25].

The third and fourth stages are based on natural language analysis and machine learning to extract valuable information from unstructured data. First, a sentimental analysis is carried out to determine the tonality of the text, then thematic modeling is carried out to identify the main topics. At the last stage, network data modeling is used to study the relationships between the data and obtain more relevant research results. All these methods help you make informed decisions based on survey data and extract valuable insights from unstructured data.

2.1 Data Gathering

Data collection was carried out by distributing a link to an online survey among lecturers, students and developers. At the stage of data purification, specialized data processing modules were used, the purpose of which is to eliminate various factors that degrade the quality of data and prevent their correct analysis.

2.2 Text Preprocessing

In general, the data preprocessing stage is the most important operation since the success of the subsequent stages of work depends on the quality of the data [26]-[27]. Therefore, it is important to devote enough time and attention to this stage in order to have reliable and accurate material for analysis and drawing conclusions in the future.

The data preprocessing stage consists of several key steps. It begins with the removal of stop words, such as common prepositions and conjunctions that do not carry meaningful information. This is followed by the processes of lemmatization and stemming: within the framework of lemmatization, words are transformed into their grammatical bases, taking into account the peculiarities of the language, while stemming reduces words to their root forms without taking into account grammar. Noise removal completes the stage, which includes filtering out unwanted characters such as special characters and punctuation marks.

2.3 Sentiment Analysis

The sentiment analysis of the respondents' responses was performed using a polynomial naive Bayesian classifier. The polynomial naive Bayesian classifier is one of the popular classification algorithms that is widely used in sentiment analysis. The algorithm evaluates the probabilities that the text belongs to certain classes (positive, neutral or negative mood) and then assigns the text the class with the highest probability [28]-[29]. The ratio of training and test data is 70:30. The result of the classification is shown in Figure 2.

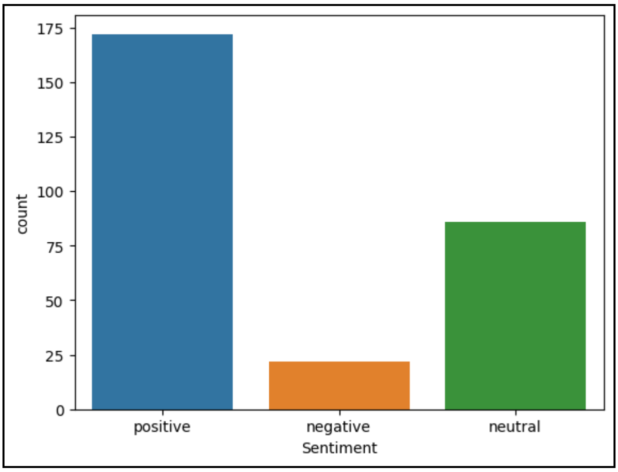

Figure 2
Classification result

The performance of the classifier is evaluated using the confusion matrix, which is shown in Figure 3. The Confusion matrix shows the number of correctly and incorrectly classified examples for each class.

Since, the dataset does not have an equal number of labels for the classes under consideration, there is an issue with an unbalanced dataset and this may lead to poor classification quality. Therefore, balancing methods such as SMOTE and ROSE were applied to the dataset. Based on these methods, the amount of data for classes that are not frequently encountered while preserving the statistical properties of the sample was increased and data normalization was applied. Applying a polynomial naive Bayesian classifier to analyse the dataset, also requires an evaluation of the classifier performance. The quality evaluation of the classifier model is presented in Table 2, which shows the high performance of the algorithm.

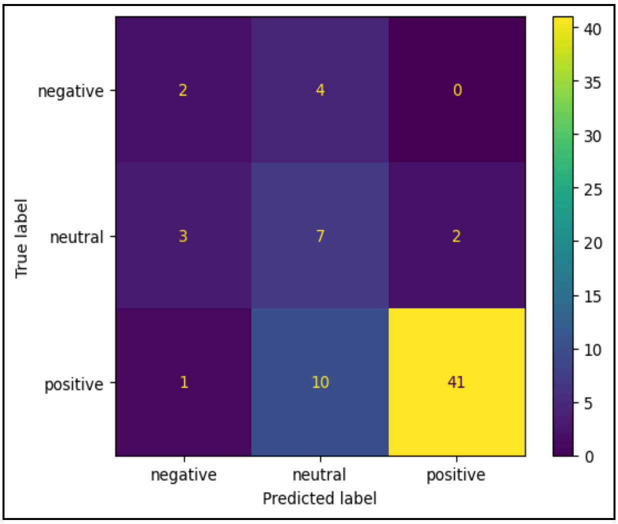

Figure 3
Confusion matrix

Table 2
Classifier performance evaluation

Algorithm	Accuracy	Precision	Recall	F1-score
PolyNB Classifier	95.5	95	97	96

2.4 Topic Modeling

Thematic modeling is a popular analytical tool for evaluating textual data, as well as many other data sources [30]-[31]. There are various methods of thematic modeling. One of the most powerful thematic modeling algorithms designed for the intelligent analysis of texts, the detection of hidden topics and the search for connections between them is the Latent Dirichlet allocation (LDA) [32]. The LDA algorithm treats each document as a mixture of topics, and each topic as a mixture of words. This allows documents to "overlap" each other in content, rather than being divided into separate groups, which reflects the typical use of natural language [32].

The main parameters of the LDA are:

- the number of topics extracted from the text corpus. This parameter can be configured depending on the specific task and data;

- alpha parameter is a hyperparameter that determines the distribution of topics on documents. A higher value of the alpha parameter means that each document will contain more topics;

- beta parameter is a hyperparameter that determines the distribution of words on topics. A higher value of the beta parameter means that each topic will contain more words.

After completing the stages of preliminary data processing of the proposed methodology, and bringing them to a form convenient for analysis, the algorithm parameters were adjusted for its further training.

As can be seen from the matrix, the classifier determines the majority of respondents' answers correctly. The classifier shows low accuracy rates within a class such as neutral.

2.5 Topic Network Modeling

The construction of network models was carried out in order to study and clarify the relationships between the identified topics and their corresponding terms.

The application of network data modeling shows the structure and connections, each of which represents topics relevant to this research area. Thus, network modeling made it possible to build a network where network nodes represent terms and relationships between them derived from cluster groups. The resulting network does not take into account the directions of connections since the network of topics is built to represent the structure of knowledge relations.

3 Results and Discussion

An effective assessment of the optimal number of topics was carried out using the connectivity method proposed in [33]. The connectivity method takes into account the connectivity of semantic information in the model, that is, the conditional likelihood is estimated for the combination of words for which the document simultaneously contains the first n words of the topic. Therefore, for the consistency method, the higher the probability that words with the highest thematic rating can appear in a document at the same time, the better the classification effect of the model [33].

As a result of, calculating consistency estimates for each thematic model, a graph of connectivity indicators for a different number of topics was constructed. As can be seen from Figure 4, the optimal number of topics maximizing the connectivity score is 4.

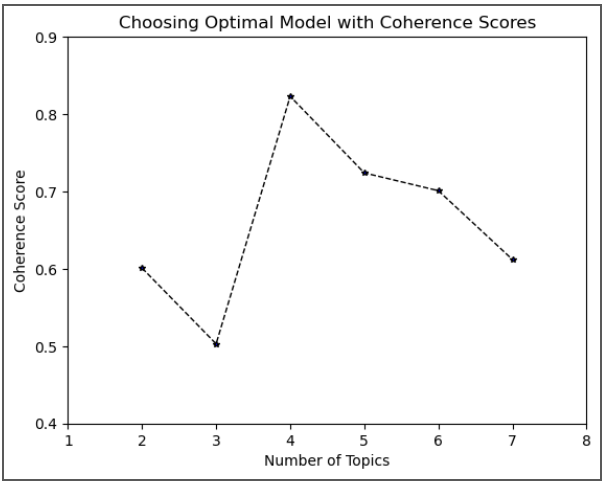

Figure 4
Choosing Optimal Model with Coherence Scores

Visualization of the output of the results of the thematic model (LDA) in the form of a word cloud is shown in Figure 5.

Figure 5
Wordclouds of top words in each topic

Also, to facilitate the interpretability of the generated topics, their key topics and importance have been studied. Accordingly, the results of calculating the number of terms and their importance are shown in Figure 6.

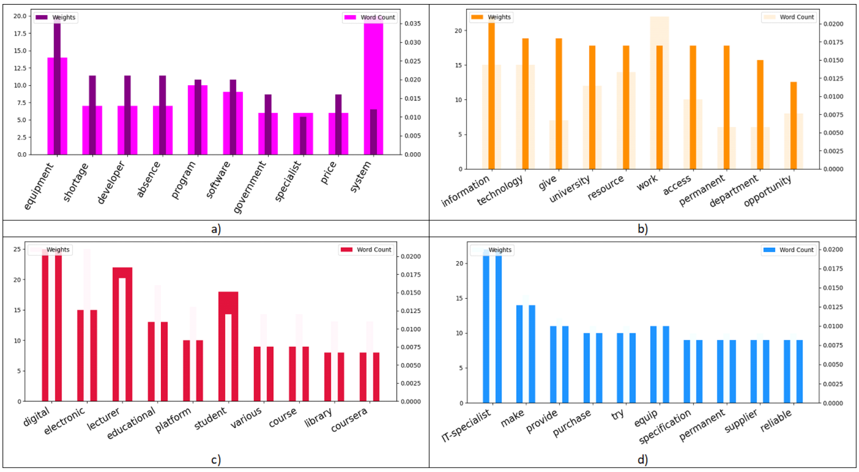

Figure 6

a)-d): Importance (weight) of terms by each topic and total frequency of terms

Table 3 shows the calculation of 4 topics using the LDA model, as well as the words corresponding to each topic. The ten most popular terms are listed for each topic.

Table 3

LDA Topics List

Topic 0	equipment (0.037), shortage (0.021), developer (0.021), absence (0.021), program (0.020), software (0.020), government (0.016), specialist (0.010), price (0.016) and system (0.012)
Topic 1	information (0.021), technology (0.018), give (0.018), university (0.017), resource (0.017), work (0.017), access (0.017), permanent (0.017), department (0.015) and opportunity (0.012)
Topic 2	digital (0.021), electronic (0.021), lecturer (0.017), educational (0.016), platform (0.013), student (0.012), various (0.012), course (0.012), library (0.011) and coursera (0.011)
Topic 3	IT specialist (0.020), make (0.014), provide (0.011), purchase (0.010), try (0.010), equip (0.010), specification (0.009), permanent (0.009), supplier (0.009) and 0.009 reliable (0.009)

Topic 0 deserves special attention since the collected terms of this cluster reveal the problems and challenges related to the formation and development of the university ecosystem based on a digital platform. Terms such as "shortage", "absence", "price" and "government" may indicate problems related to financing and availability of hardware and software. It is assumed that the key terms "developer", "specialist" and "system" are related to the skills and knowledge of specialists necessary for the development and maintenance of hardware and software.

It is interesting to note that the terms collected on Topic 1 are generally related to the educational process and technological resources. Key terms such as "information", "technology", "university" and "department" may indicate topics related to the use of information technology in higher education. The terms "resource", "access", "permanent" and "opportunity" can be indicated on topics related to providing access to information technologies and resources for teaching and research. While the terms "work" and "give" may be related to the use of technology in the work environment or teaching students how to work with technological resources.

The key terms of Topic 2 include terms related to digital and electronic educational platforms (such as Coursera), lecturers, students, various courses and libraries. Together, these keywords can be used to analyze the topics of educational resources and processes used in the educational process of the university.

From the received terms of Topic 3, it can be assumed that they relate to recommendations and proposals for the development of the university's digital ecosystem. Suggestions and recommendations are related to the work of an IT specialist, to the procurement and provision of IT equipment, efforts to ensure continuous operation and reliability of equipment, as well as efforts to select reliable suppliers and compliance with technical specifications. In general, these keywords indicate the topic of managing the IT infrastructure and ensuring its effective operation.

At the previous stage of the methodology, the use of LDAP for thematic modeling made it possible to identify certain aspects in the form of 4 cluster groups, where the algorithm clearly detected and separated the main terms from the studied data. Following the results obtained, by computing the extended connections of the nodes in the data network, we can identify those nodes that have an impact on the whole network. In addition, we also analyzed centrality by proximity. This measure evaluates each node based on its "proximity" to all other nodes in the network. As a result of, the network analysis, the central nodes and their connections have been established. As central nodes, nodes such as ‘student’, ‘lecturer’ and ‘developers’ were identified to show the relationships between the key terms, thus revealing the relationships between the actors of the university's digital ecosystem.

Connections exist when there are joint tools and a systematic approach to the implementation of learning between the subjects of educational activity. Next, Figure 7 shows a visualization of the connections between the central nodes of the data network and their relationships.

The relationship between the central nodes "student" and "teacher" in Figure 8, highlighted the relationship based on a positive assessment of the work of the university portal, the implementation of massive open online courses (MOOC) like Coursera, EDX.

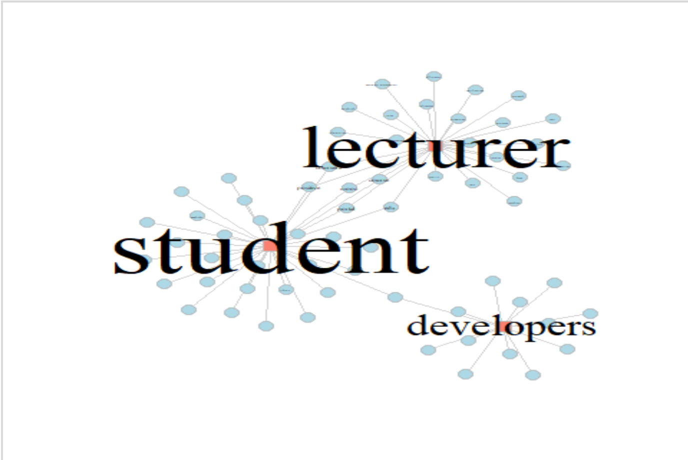

Figure 7
Connections between the central nodes of the data network and their relationships

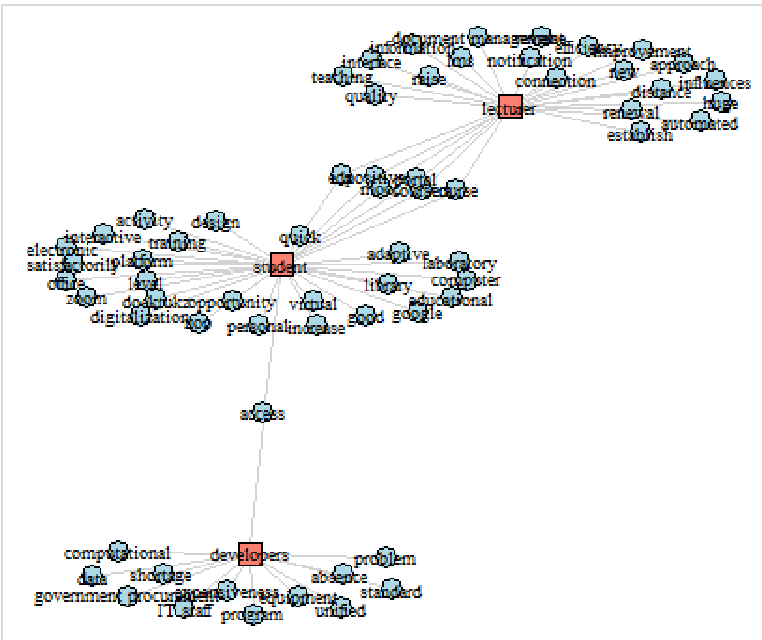

Figure 8
Network data analysis

The following relationship between the central nodes "student" and "developers" in Figure 8, also highlighted the relationship based on the keyword as "access". This allows us to interpret the properties of the university's digital ecosystem based on openness, accessibility and integration of internal modules and components of the entire system. In this way, the network analysis in this study allowed for a more in-depth and detailed assessment of the work of the university's digital ecosystem.

Conclusions

The intellectual analysis of open-ended answers allowed us to obtain a deeper analysis of the answers, identify hidden problems, as well as understand the real needs and expectations of subjects of educational activity. In addition, open-ended questions help survey participants express their opinions and perspectives, which contributes to building a trusting relationship between the university and their community. However, the analysis of open issues requires considerable effort and presents certain difficulties. Despite the small size of the dataset, the study provided important primary data and may serve as a starting point for more extensive research in the future.

Experimental studies demonstrate that the developed methodology involves several stages of research, where in the initial stage, sentiment analysis was performed using a polynomial naive Bayesian classifier. From the classification results, it is observed that this algorithm showed high performance with an accuracy of 95.5%. In addition, thematic modelling and text network analysis were applied in the next stages of the study, which represent an effective tool for automated processing of open-ended questions and extracting valuable information to help assess the level of satisfaction with the university ecosystem.

The proposed methodology also made it possible to identify the main problems of the development of the university's digital ecosystem, as well as to propose ways to solve them. The significance of this result lies in the opportunity not only to test our initial hypothesis, but also to offer specific recommendations to improve the existing situation. This ensures the practical significance of the research and its potential for real application in educational institutions. The study reflects the current state of the digital ecosystem of universities, which opens perspectives for studying dynamic changes over time. Future research may include expanding the sample and examining changes over time to better assess the evolution of the digital ecosystem of universities. It is important to note that the proposed methodology is adapted not only for the educational sphere, but also has the potential to be applied in other areas. The discussion of the results obtained with professional experts in the field of education and the IT sector confirmed the significance and reliability of the conclusions obtained during the study.

Acknowledgement

The work was supported by grant financing of scientific and (or) scientific and technical projects for 2023-2025 of the Ministry of Science and Higher Education of the Republic of Kazakhstan (grant No. AP19677501 "Development of intelligent services of a smart campus with integration into the information ecosystem of the university").

References

[1] G. R. El Said, "How Did the COVID-19 Pandemic Affect Higher Education Learning Experience? An Empirical Investigation of Learners' Academic Performance at a University in a Developing Country," Advances in Human-Computer Interaction, pp. 1-10, 2021, https://doi.org/10.1155/2021/6649524

[2] L. Amhag, L. Hellström, & M. Stigmar, "Teacher educators' use of digital tools and needs for digital competence in higher education," Journal of Digital Learning in Teacher Education, vol. 35(4), pp. 203-220, 2019, https://doi.org/10.1080/21532974.2019.1646169

[3] B. Bygstad, E. Øvrelid, S. Ludvigsen, M. Dæhlen, "From dual digitalization to digital learning space: Exploring the digital transformation of higher education," Computers & Education, Vol. 182, pp. 1-16, 2022, https://doi.org/10.1016/j.compedu.2022.104463

[4] L. South, D. Saffo, O. Vitek, C. Dunne, M. A. Borkin, "Effective Use of Likert Scales in Visualization Evaluations: A Systematic Review," Computer Graphics Forum, Vol. 41(3), pp. 43-55, 2022, https://doi.org/10.31219/osf.io/6f3zs

[5] S. Gottipati, V. Shankararaman & S. Gan, "A conceptual framework for analyzing students' feedback," IEEE Frontiers in Education Conference (FIE), pp. 1-8, 2017, https://doi.org/10.1109/FIE.2017.8190703

[6] S. Cunningham-Nelson, M. Baktashmotlagh and W. Boles, "Visually exploring sentiment and keywords for analysing student satisfaction data," 29th Australasian Association for Engineering Education Conference, pp. 132-138, 2018

[7] S. Gottipati, V. Shankararaman, and J. Lin, "Latent Dirichlet Allocation for textual student feedback analysis," Proceedings of the 26th International Conference on Computers in Education, pp. 220-227, 2018, Retrieved from: https://ink.library.smu.edu.sg/sis_research/4215, Accessed 8 Nov 2023

[8] Sh. Mehmood, I. Ahmad, M. A. Khan, F. Khan and T. Whangbo, "Sentiment Analysis in Social Media for Competitive Environment Using Content Analysis," Computers Materials & Continua, pp. 5603-5618, 2022, https://doi.org/10.32604/cmc.2022.023785

[9] K. Okoye, H. Hussein, A. Arrona-Palacios and et al., “Impact of digital technologies upon teaching and learning in higher education in Latin America: an outlook on the reach, barriers, and bottlenecks,” Education and Information Technologies, pp. 2291-2360, 2023, https://doi.org/10.1007/s10639-022-11214-1

[10] Sh. Hussain, M. Ayoub, G. Jilani, Y. Yu, A. Khan, J. A. Wahid, M. F. A. Butt, G. Yang, D. P. F. Moller, H. Weiyan. “Aspect2Labels: A novelistic decision support system for higher educational institutions by using multi-layer topic modelling approach,”Expert Systems with Applications, Vol. 209, 2022, https://doi.org/10.1016/j.eswa.2022.118119

[11] S. Gottipati, V. Shankararaman, and J. R. Lin, ‘‘Text analytics approach to extract course improvement suggestions from students’ feedback,’’ Research and Practice in Technology Enhanced Learning, Vol. 13, No. 1, pp. 2-19, 2018, https://telrp.springeropen.com/articles/10.1186/s41039-018-0073-0

[12] D. Buenaño-Fernandez, M. González, D. Gil and S. Luján-Mora, “Text Mining of Open-Ended Questions in Self-Assessment of University Teachers: An LDA Topic Modeling Approach,” IEEE Access, Vol. 8, pp. 35318-35330, 2020, https://doi.org/10.1109/ACCESS.2020.2974983

[13] M. Hujala, A. Knutas, T. Hynninen, H. Arminen, “Improving the quality of teaching by utilising written student feedback: A streamlined process,” Computers & Education, Vol 157, 2020, https://doi.org/10.1016/j.compedu.2020.103965

[14] N. Grönberg, A. Knutas, T. Hynninen and M. Hujala, “Palaute: An Online Text Mining Tool for Analyzing Written Student Course Feedback,” IEEE Access, Vol. 9, pp. 134518-134529, 2021, https://doi.org/10.1109/ACCESS.2021.3116425

[15] Sh. Srinivas, S. Rajendran, “Topic-based knowledge mining of online student reviews for strategic planning in universities,” Computers & Industrial Engineering, Vol. 128, 2019, pp. 974-984, https://doi.org/10.1016/j.cie.2018.06.034

[16] W. Al-Mawee, K. M. Kwayu, T. Gharaibeh, “Student's perspective on distance learning during COVID-19 pandemic: A case study of Western Michigan University, United States”, International Journal of Educational Research Open, Vol. 2, 2021, https://doi.org/10.1016/j.ijedro.2021.100080

[17] K. F. Hew, X. Hu, C. Qiao, and Y. Tang, ‘‘What predicts Student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach,’’ Computers & Education, Vol. 145, No. 103724, 2020, https://doi.org/10.1016/j.compedu.2019.103724

[18] H. Fu, Y. Xiao, I.K. Mensah et al., “Exploring the configurations of learner satisfaction with MOOCs designed for computer science courses based on

integrated LDA-QCA method," Education and Information Technologies, 2023, https://doi.org/10.1007/s10639-023-12185-7

[19] G. Nanda, K. A. Douglas, D. R. Waller, H. E. Merzdorf and D. Goldwasser, "Analyzing Large Collections of Open-Ended Feedback From MOOC Learners Using LDA Topic Modeling and Qualitative Analysis," IEEE Transactions on Learning Technologies, Vol. 14, No. 2, pp. 146-160, 2021, https://doi.org/10.1109/TLT.2021.3064798

[20] M. Nilashi, R.A. Abumalloh, M. Zibarzani and et al., "What Factors Influence Students Satisfaction in Massive Open Online Courses? Findings from User-Generated Content Using Educational Data Mining," Education and Information Technologies, Vol. 27, pp. 9401-9435, 2022, https://doi.org/10.1007/s10639-022-10997-7

[21] S. S. Cunningham-Nelson, M. Baktashmotlagh and W. Boles, "Visualizing Student Opinion Through Text Analysis," IEEE Transactions on Education, Vol. 62, No. 4, pp. 305-311, 2019, https://doi.org/10.1109/TE.2019.2924385

[22] Shi, Dingpu and et al, "Characteristics of Students' Learning Behavior Preferences - an Analysis of Self-commentary Data Based on the LDA Model," Journal of Intelligent & Fuzzy Systems, Vol. Pre-press, No. Pre-press, pp. 1-15, 2023, https://doi.org/10.3233/JIFS-2329

[23] N. Denissova, I. Dyomina, A. Tlebaldinova, R. Chettykbayev, D. Muratuly, V. Zuev, "Using structural equation modeling methods to assess the university's digital ecosystem," Scientific Journal of Astana IT University, Vol. 17, pp. 95-105, 2024, https://doi.org/10.37943/17CCXJ5272

[24] W. Wojcik, M. Karmenova, S. Smailova, A. Tlebaldinova, and A. Belbeubaev, "Development of Data-mining Technique for Seismic Vulnerability Assessment," International Journal of Electronics and Telecommunications, Vol. 67, No. 2, pp. 261-266, 2021, https://doi.org/10.24425/ijet.2021.135974

[25] A. Tlebaldinova, M. Karmenova, E. Ponkina and A. Bondarovich, "CNN-based Approaches for Weed Detection," *2022 10th International Scientific Conference on Computer Science (COMSCI)*, Sofia, Bulgaria, 2022, pp. 1-4, doi: 10.1109/COMSCI55378.2022.9912593

[26] C. P. Chai, "Comparison of text preprocessing methods," Natural Language Engineering, Vol. 29(3), pp. 509-553, 2023, https://doi.org/10.1017/S1351324922000213

[27] A. Majid, D. Nugraha, F. D. Adhinata, "Sentiment Analysis on Tiktok Application Reviews Using Natural Language Processing Approach," Journal of Embedded System Security and Intelligent Systems, Vol. 4, No. 1, pp. 32-38, 2023, https://doi.org/10.26858/jessi.v4i1.41897

[28] L. Yue, W. Chen, X. Li, and et al., "A survey of sentiment analysis in social media," Knowledge and Information Systems, Vol. 60, pp. 617-663, 2019, https://doi.org/10.1007/s10115-018-1236-4

[29] I. Vayansky and S. A. P. Kumar, "A review of topic modeling methods," Information Systems, 2020, https://doi.org/10.1016/j.is.2020.101582

[30] H. Jelodar, Y. Wang, C. Yuan, and et al., "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," Multimedia Tools and Applications, Vol. 78, pp. 15169-15211, 2019, https://doi.org/10.1007/s11042-018-6894-4

[31] Sh. H. Mohammed, S. Al-augby, "LSA & LDA topic modeling classification: comparison study on E-books," Indonesian Journal of Electrical Engineering and Computer Science, Vol. 19, No. 1, pp. 353-362, 2020, https://doi.org/10.11591/ijeecs.v19.i1.pp353-362

[32] J. Gan, Y. Qi, "Selection of the Optimal Number of Topics for LDA Topic Model-Taking Patent Policy Analysis as an Example," Entropy (Basel), Vol. 23(10):1301, 2021, https://doi.org/10.3390/e23101301

[33] S. Peng, Y. Zhou, L. Cao, S. Yu, J. Niu, W. Jia, "Influence analysis in social networks: A survey," Journal of Network and Computer Applications, Vol. 106, pp. 17-32, 2018, https://doi.org/10.1016/j.jnca.2018.01.005