АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР
ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ
INFORMATION AND COMMUNICATION TECHNOLOGIES

IT ЖАҢА ТРЕНДТЕРІ
НОВЫЕ ТРЕНДЫ IT
IT NEW TRENDS

DOI 10.51885/1561-4212_2025_1_221
IRSTI 28.23.25; 14.15.15

A.S. Tlebaldinova[1], S.K. Kumargazhanova[2], S.S. Smailova[3], Zh.T. Konyrbayeva[4], Zh.M. Seitakhmetova[5]
D. Serikbayev East-Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan
[1]E-mail: ATlebaldinova@edu.ektu.kz
*[2]E-mail: SKumargazhanova@edu.ektu.kz**
[3]E-mail: SSmailova@edu.ektu.kz
[4]E-mail: ZhKonurbayeva@edu.ektu.kz
[5]E-mail: Seitahmetova_zh@ukk.nis.edu.kz

CLUSTER ANALYSIS OF BEHAVIORAL FACTORS IN THE FORMATION OF STUDENTS' DIGITAL IDENTITY

БІЛІМ АЛУШЫЛАРДЫҢ ЦИФРЛЫҚ СӘЙКЕСТІГІН ҚАЛЫПТАСТЫРУДАҒЫ МІНЕЗ-ҚҰЛЫҚ ФАКТОРЛАРЫН КЛАСТЕРЛІК ТАЛДАУ

КЛАСТЕРНЫЙ АНАЛИЗ ПОВЕДЕНЧЕСКИХ ФАКТОРОВ В ФОРМИРОВАНИИ ЦИФРОВОЙ ИДЕНТИЧНОСТИ ОБУЧАЮЩИХСЯ

Abstract. *With the development of technology and the transition to online learning, ensuring the security of student data and their authentication in a digital environment are becoming important tasks for educational institutions. The purpose of the study is to use cluster analysis to analyze survey data in order to identify the features of students' digital identity. The general methodology of the study is presented, according to which preliminary data processing was carried out in order to prepare them for subsequent analysis. The paper uses hierarchical clustering using Ward's algorithm to analyze behavioral factors affecting the formation of students' digital identity. The study used survey data collected among 324 students from three regional universities of the Republic of Kazakhstan. The content of the questions includes digital activity and interaction of students in the online space, as well as questions about their academic, research and social activities. The Cronbach's α coefficient was used to assess the reliability and reliability of the data. As a result of the study, groups of students with different levels of digital identity ("Intensive digital activity", "Hybrid form of digital activity", "Limited digital activity").*

Keywords: *digital identity, survey data, hierarchical clustering, digital activity, academic activity, research activity, social activity.*

Аңдатпа. *Технологиялардың даму қарқынына және онлайн оқытуға көшуге байланысты студенттер деректерінің қауіпсіздігін қамтамасыз ету және оларды цифрлық ортада аутентификациялау сияқты мәселелер білім беру мекемелері үшін маңызды міндеттерге айналды. Зерттеудің мақсаты – білім алушылардың цифрлық сәйкестіктерінің ерекшеліктерін анықтау мақсатында сауалнама деректерін талдау үшін кластерлік талдауды қолдану. Зерттеудің жалпы әдістемесі ұсынылған, оған сәйкес оларды кейінгі талдауға дайындау мақсатында деректерге алдын ала өңдеу операциялары жүргізілді. Жұмыста студенттердің цифрлық сәйкестігін қалыптастыруға әсер ететін мінез-құлық факторларын талдау үшін Ward алгоритмі арқылы иерархиялық кластерлеу қолданылды. Зерттеу барысында Қазақстан Республикасының үш өңірлік жоғары оқу орындарынан 324*

студент арасында жиналған сауалнама деректері пайдаланылды. Сұрақтардың мазмұны цифрлық белсенділікті және онлайн-кеңістіктегі білім алушылардың өзара іс-қимылын, сондай-ақ олардың академиялық, зерттеу және әлеуметтік белсенділігі туралы сұрақтарды қамтиды. Деректердің сенімділігін бағалау үшін α Кронбах коэффициенті қолданылды. Зерттеу нәтижесінде білім алушылардың цифрлық сәйкестіктерінің әртүрлі деңгейлерін қамтитын («Қарқынды цифрлық белсенділік», «Цифрлық белсенділіктің гибридті түрі», «Шектеулі цифрлық белсенділік») топтар бөлінді.

***Түйін сөздер**: цифрлық сәйкестілік, сауалнама деректері, иерархиялық кластерлеу, цифрлық белсенділік, академиялық қызмет, зерттеу қызметі, әлеуметтік қызмет.*

***Аннотация**. С развитием технологий и переходом к онлайн-обучению, обеспечение безопасности данных студентов и их аутентификация в цифровой среде становятся важными задачами для образовательных учреждений. Цель исследования заключается в применении кластерного анализа для анализа опросных данных с целью выявления особенностей цифровой идентичности обучающихся. Представлена общая методология исследования, согласно которой была проведена предварительная обработка данных с целью их подготовки для последующего анализа. В работе была использована иерархическая кластеризация по алгоритму Ward для анализа поведенческих факторов, влияющих на формирование цифровой идентичности студентов. В исследовании были использованы опросные данные, собранные среди 324 студентов из трех региональных вузов Республики Казахстан. Содержание вопросов включает в себя цифровую активность и взаимодействие обучающихся в онлайн-пространстве, а также вопросы об их академической, исследовательской и социальной активности. Для оценки достоверности и надежности данных был использован коэффициент α Кронбаха. В результате исследования были выделены группы обучающихся с различным уровнем цифровой идентичности («Интенсивная цифровая активность», «Гибридная форма цифровой активности», «Ограниченная цифровая активность»).*

***Ключевые слова**: цифровая идентичность, данные опроса, иерархическая кластеризация, цифровая активность, академическая деятельность, исследовательская деятельность, социальная активность*

Introduction. In our information age, permeated by the rapid development of digital technologies, the digital identity of students is becoming an integral part of their social life and academic activities.

This study focuses on identifying the features of the digital identity of students of regional universities in East Kazakhstan, taking into account their academic, scientific and social activities. The analysis of these aspects will provide a better understanding of how students interact with digital technologies and information as part of their learning, research, and social interactions in the learning environment, which can then be used to analyze students ' behavioral and preferred characteristics in an online environment. This will help them identify their digital identity in the context of developing a digital profile and identify educational needs and weaknesses, which in turn will allow them to create a personalized learning trajectory in the future.

In this regard, the purpose of this study is to use cluster analysis to analyze survey data in order to identify the features of students ' digital identity.

The methodology and analysis were based on the followingих researchих questionsах: What behavioral factors have a significant impact on the formation and expression of students ' digital identity? How can these factors be identified and systematized using cluster analysis methods based on online survey data?

The structure of the article is presented as follows: Section 1 contains an analysis of previous studies, section 2 describes the methodology, including the stages of data collection and pre-processing, as well as the application of cluster analysis. Section 3 includes a description of the course and results of the study, as well as a discussion of the results obtained.

1 Overview of previous works. In recent years, digital identity has become a key aspect of modern life, especially in the context of education and online interactions. A review of scientific papers on the definition of this concept reveals a variety of approaches and concepts used by

researchers. For example, identity, without specifying it as "digital", describes a set of personality characteristics that help distinguish one person from others (Kovaleva, 2019). This process of identity formation is closely related to social interaction and depends on social norms and practices (Tomlinson & Jackson, 2021, Batool & Ghayas, 2020). In today's world, people are spending more and more time in an online environment that constantly generates a huge amount of content. This encourages us to consider the concept of identity in the digital space as well. Thus, the concept of digital identity arises, defined in (Korać, Damjanović & Simić, 2022, Borrás-Gené, Serrano-Luján & Díez, 2022, Moskovskaya & Perova, 2022, Feher, 2021) as the relationship between a person's self-perception and the identity formed by them and other participants in the online space.

Of course, the issue of the formation and development of digital identity is an integral part of modern education and society as a whole. Understanding and taking into account the peculiarities of digital identity helps to effectively integrate digital tools into the educational process, creating conditions for successful personal development and preparation for a future career in the digital age (Borrás-Gené, Serrano-Luján & Díez, 2022, Borrás-Gené & Blázquez-Sevilla, 2020, González, 2024, Hernández-Orellana, Pérez-Garcias & Roco-Videla, 2021, Muslov, 2023). In turn, one of such tools described in (Borrás-Gené, Serrano-Luján & Díez, 2022) is the self-reflection tool "Digital Identity Analysis Canvas". It is designed to raise students' awareness of the importance of their digital identity. The purpose of creating this tool is to analyze the student's digital footprint on the web, that is, his online activities, which will subsequently serve as a starting point for providing recommendations on the competent management of his professional digital identity. The canvas structure includes five blocks: Activity, Community, Focus, Privacy, Communication. The basis for the creation of this tool is an online questionnaire, which consists of two blocks: 1 block - "Student activity in digital social networks" (5 questions), 2 block – "Use of other educational digital networks and platforms" (15 questions). The content of the questions covers the reflection of the student's digital identity, his perception of his own online presence as a digital resume and its impact on job search, as well as the level of interest in actively managing his professional digital identity.

Research based on the analysis of online survey data is also conducted in order to study the characteristics of students' academic digital identity (Hernández-Orellana, Pérez-Garcias & Roco-Videla, 2021). The authors propose, based on the identification of the characteristics of the academic digital identity of students, to redefine their learning space, which allows integrating all those technological tools that the student actually uses. The sample consisted of 509 students who voluntarily filled out an online questionnaire. To identify the characteristics, a conglomerate analysis was performed using a hierarchical dissociative method.

According to recent studies, the study of students' digital identity problems is usually limited to analyzing only one aspect, which does not always allow us to fully understand their diverse online activity. The present study attempts to eliminate this shortcoming by expanding the analysis to behavioral factors related not only to the academic, but also to the research and social spheres of students.

2 Metodology. The general scheme of the research methodology is shown in Figure 1.

As can be seen, the proposed research methodology consists of three main stages:

- data collection.
- pre-processing of data;
- data processing (cluster analysis).

A more detailed description of each stage is provided below.

2.1 Data collection. As you know, the digital identity of a student includes various aspects covering academic, research and social spheres of activity. To implement the student's Digital Profile service, a survey was conducted among students. The survey included questions related to each of these areas.

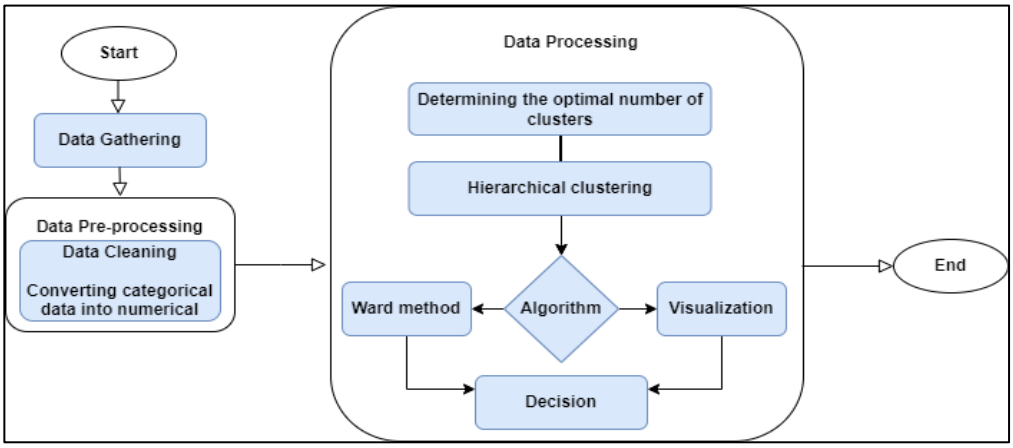


Figure 1. General scheme of research methodology

Note – compiled by the authors

The survey was conducted online via a Google form. 324 people from three regional universities of the Republic of Kazakhstan took part in the survey: D. Serikbayev East Kazakhstan Technical University, S. Amanzholov East Kazakhstan University and the Kazakh-American Free University. There are 203 men and 121 women among the participants. The survey questions were previously tested for reliability and validity with ten experts in the field of digitalization of education and IT employees of enterprises. The results were analyzed in order to assess reliability using the Cronbach's a statistical method (Muslov, 2023), which showed a coefficient of 0.85. This indicates a fairly high reliability of the internal consistency of the survey data.

The survey consisted of 27 questions divided into four blocks: general questions (G) – 4, questions on digital activity and interaction in the online space (D) – 7, questions on academic activity (A) – 6, questions on research activity (R) – 6 and on social activity (S) – 4. The basic design of the study is presented in Table 1.

Table 1. The basic design of the study and the corresponding tags of the survey questions

Question tag	G	D	A	R	Question tag	D	A	R	S
Sex	+				Degree of engagement in DLE			+	
Age	+				Desire to engage in research			+	
Specialty/Position	+				Assessment of DLE			+	
University/Company	+				Motivation for participation in DLE			+	
Virtuality degree		+			Form/type of participation in DLE			+	
Preferred environment		+			Online learning platforms		+		
Communication preferences		+			Educational resources offered by the university		+		
Internet activity		+			Protection of educational achievements on educational platforms		+		
Learning resources			+		Additional source of knowledge		+		
Social activity		+			Recognition of non-formal learning outcomes		+		
Online privacy		+			Participation in student events				+

Awareness digital security		+			Membership in student organization				+
Benefits of participating in DLE				+	Participation in university sports events				+
					Preferred university events				+
Note – compiled by the authors									

Determining the main characteristics of a student's digital identity is based on the analysis of various types of questions in the questionnaire, such as open, closed, Likert scale and multiple-choice questions.

2.2 Pre-processing of data. Pre-processing of data after collecting survey materials is a necessary and important step (Naseem, Razzak & Eklund, 2021). The purpose of these activities is to prepare data for subsequent analysis and application of machine learning algorithms. To begin with, work has been done on data cleaning and preprocessing. This stage included deleting or filling in missing values, processing outliers, and converting categorical data to numeric values using the Label Encoding method. Label Encoding is a method of categorical data conversion in which each unique category is assigned a unique integer value. The advantage of Label Encoding is that it is easy to use and can be effective if there is a certain order between the categories (Dahouda & Joe, 2021). At the final stage of preliminary data processing, the aggregation of identical responses of respondents for each question was carried out. Thus, the data was prepared for subsequent analysis and application of machine learning algorithms. All data processing operations were implemented using the R programming language.

2.3 Applying cluster analysis. Cluster data analysis (or clustering) is a data analysis method that aims to divide a dataset into groups, or clusters, so that objects within one cluster are more similar to each other than to objects from other clusters (Sari, Al-Khowarizmi & Batubara, 2021, Wójcik et al., 2021, Shahapure & Nicholas, 2020). The purpose of cluster analysis is to identify the internal structure in the data and identify natural groups of objects that have similar characteristics or behavior in the provided dataset.

In general, many clustering algorithms can be divided into iterative and hierarchical ones.

Iterative clustering methods are effective if the number of segments is known in advance. These algorithms recalculate the distance matrix at each stage until all objects are assigned to their clusters. Centroids, representing typical and central examples of each cluster, are formed during this method.

Hierarchical methods, on the contrary, consistently recalculate the distances between all objects, combining them into clusters on the dendrogram. These methods are suitable for situations where the number of groups is unknown in advance, which is relevant for our case.

In this study, the Euclidean distance is used as a measure of the similarity of observations. The distances between clusters are calculated using the Ward method (Sharma et al., 2019, Randriamihamison, Vialaneix & Neuvial, 2021).

The Ward method. Let the partition of the set *V* be understood as a set of elements represented in the form

$$K = \{V_1, \dots, V_i, \dots, V_N\}, \quad (1)$$

where *N* is the number of clusters, V_N – and Vn is the set of vertices that make up the cluster with number k, $k = 1, \dots, N$. The objective function for K is defined as the sum of the intracluster variances:

$$Z(K) = \sum_{i=1}^{N} \sum_{j \in V_i} d^2\,(j, z_i), \quad (2)$$

where i – cluster number, z_i – the center of cluster i.

This division is formed from K when two clusters with numbers $i_1, i_2 \in \{1, \dots, N\}, i_1 \neq i_1$, are combined by (i_1, i_2); let

$$\Delta(i_1, i_2) \coloneqq Z\big(K(i_1, i_2)\big) - Z(K). \quad (3)$$

looking at all pairs $i_1, i_2 \in \{1, \dots, N\}, i_1 \neq i_2$, we find numbers i_1^* and i_2^* such that

$$\Delta(i_1^*, i_2^*) = min_{i_1, i_2 \in \{1, \dots, N\}} \Delta(i_1, i_2). \quad (4)$$

Therefore, the Ward method does not know the number of clusters in advance.

In the Ward method, the increase in the sum of squared distances of objects to cluster centers is measured as a result of combining them.

This method seeks to minimize the variance within clusters at each step of hierarchical clustering.

Next, the nearest clusters are sequentially combined to minimize the increase in the sum of squared distances. At each step of the algorithm, objects are grouped into clusters, and then the distances between clusters are recalculated. This process is repeated until the specified number of clusters is reached or until a satisfactory clustering structure is obtained. Thus, the Ward method is focused on forming compact and homogeneous clusters in the data.

3 Results and discussion. As a result of the preprocessing stage, the total number of observations was 135. Before applying cluster analysis, the optimal choice of the number of clusters was made using the Elbow method, since the number of clusters in the Ward method is unknown in advance.

The Elbow method is a graphical method that is used to determine the optimal number of clusters in cluster data analysis. The name of the method comes from its visual interpretation, where a curve resembling an elbow on the arm is formed on the graph of the dependence of the sum of squares of distances from points to cluster centroids on the number of clusters. The optimal number of clusters is selected based on this bend - the point after which the decrease in the sum of squared distances becomes less significant, and changes in the data become more gradual (Liu & Deng, 2021). Thus, as can be seen in Figure 2, the greatest decrease in the steepness of the curve is observed at a value of 3.

For better visualization of the data structure, in addition to the dendrogram, a phylogenetic tree of hierarchical clustering was constructed (see Figure 3). The results clearly demonstrate a clear division between clusters, and their mutual disjointment emphasizes the pronounced structure and uniformity of data within each cluster. This highlights the effectiveness of the analysis performed by clearly identifying different groups in the data set under consideration.

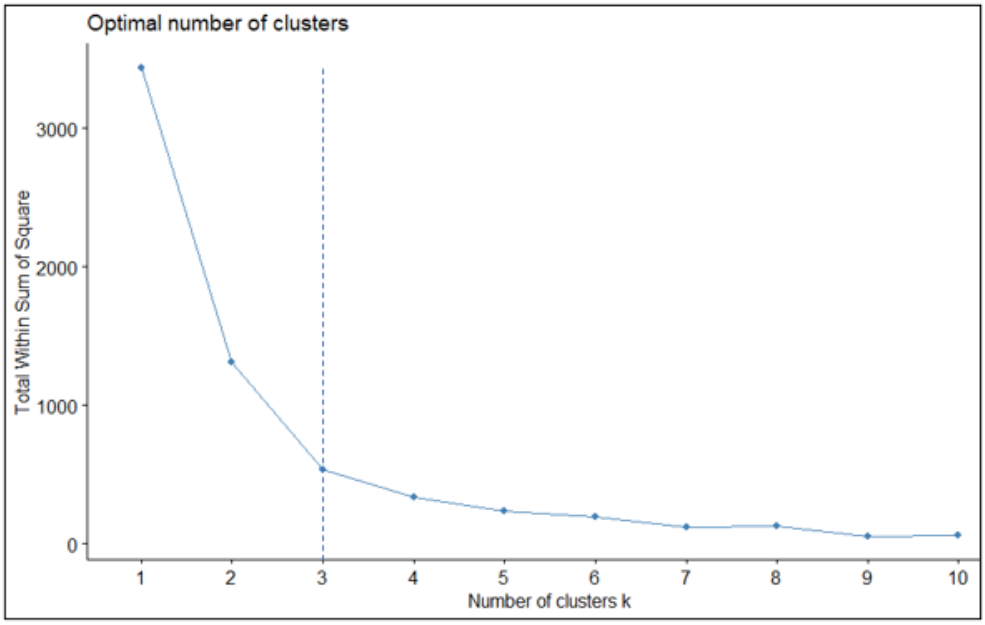
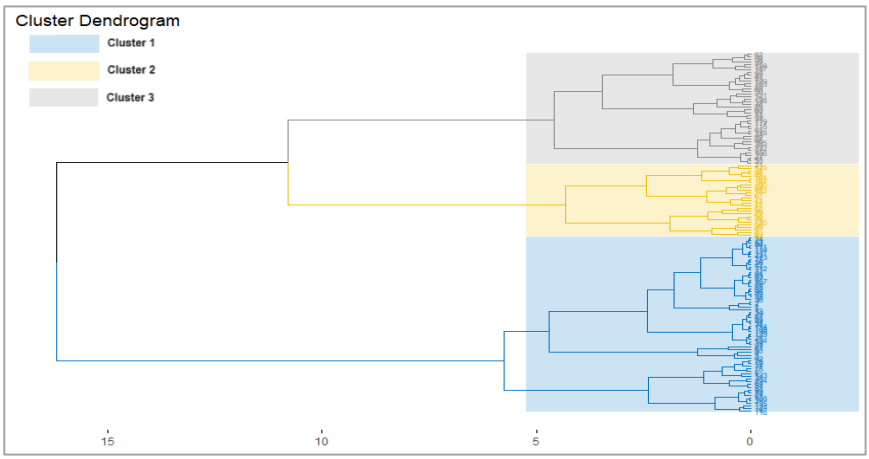


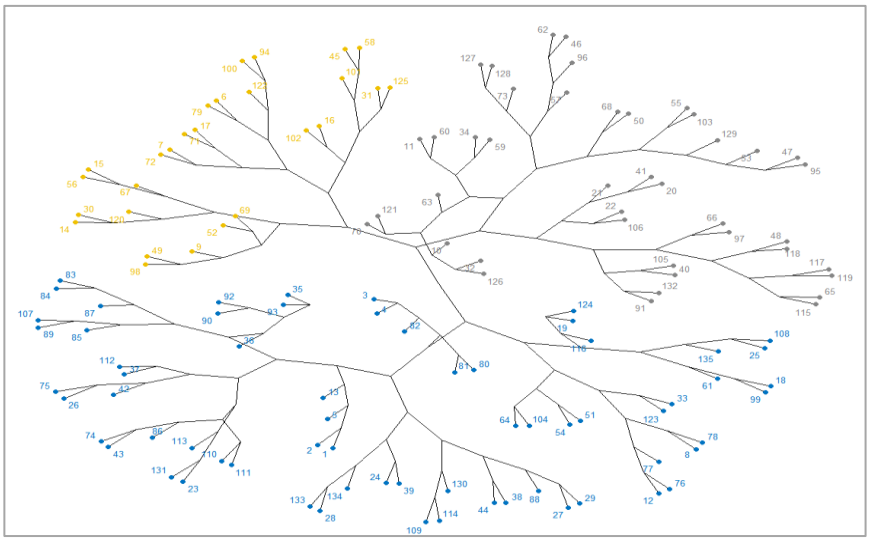
Figure 2. Optimal number of clusters

Note – compiled by the authors

The analysis of the obtained clusters shows that the students of the first (blue) cluster (66 observations) demonstrate a high digital identity, showing comfort in a virtual environment. They prefer online communication, use a variety of digital platforms, and show a preference for written digital communication. Their interests on the Internet are diverse – from educational to professional. The flexibility in learning is reflected in the use of various online resources. The perception of scientific activity as prestigious indicates a desire for academic growth. Participation in various forms of scientific activity, startups and conferences emphasizes their active digital scientific component. Students also show a wide interest in the university environment, using digital resources in student organizations and educational events. In general, the digital identity of students in this cluster is characterized by confidence and comfort in a virtual environment, active use of digital technologies in teaching and a wide range of interests in the online space, while combining them with active social and educational activities in real life.



a) Dendrogram



b) Phylogenetic tree

Figure 3. a, b – Results of cluster analysis using hierarchical clustering

Note – compiled by the authors

The students of the second (yellow) cluster (27 observations) prefer a hybrid environment, combining real and virtual space. They prefer live conversation and use the Internet to search for information and entertainment. Messengers and educational platforms are actively used in training. They are aware of their digital footprint and evaluate scientific activities, although they participate in it only in a limited way. They prefer sports activity outside the university and do not show interest in university events. In general, they show flexibility in digital technologies, but prefer more individual forms of activity and limited participation in scientific activities.

The analysis of the digital identity of students in the third (gray) cluster (42 observations) allows us to identify key characteristics. Students show a preference for relaxed digital communication through text messages and the phone, indicating the multidimensional use of the Internet for study, entertainment and mood maintenance. However, limited awareness of the digital footprint and ignorance of the relevant term indicate a lack of awareness of digital security. A negative attitude towards scientific activity is accompanied by pragmatic motives for participation, while the assessment of the prestige of research is multidirectional. The training mainly takes place through official documentation and communication with classmates. Dissatisfaction with the protection of academic achievements on educational platforms may reflect a perception of injustice. Academic activities remain a priority, although students recognize the importance of student life by showing interest in creative and scientific events at the university.

A thorough analysis of the clusters made it possible to identify the main characteristics of the digital identity of students in the context of their digital, academic, scientific and social activities. Summary characteristics of each cluster are presented in Table 2.

The results of the cluster analysis evaluation allow us to conclude that the largest cluster, cluster 1, can be characterized as "Intensive Digital Activity". Participants in this cluster show a high level of involvement in digital practices in various fields, including academic, scientific and social activities. Cluster 2, in turn, is a "Hybrid form of Digital Activity". In this group, participants combine different aspects of digital activity, maintaining a balance between traditional and digital forms of interaction. Finally, Cluster 3 is described as "Limited Digital

Activity". Participants in this cluster show a more limited interest in digital practices, preferring more traditional ways of interacting. These characteristics not only help in understanding the digital activity of different groups of participants, but also provide a basis for further research and development of appropriate support strategies in the context of the digital environment.

Table 2. Summary characteristics of clusters

Behavioral factors of students	Characteristics	Cluster 1	Cluster 2	Cluster 3
Digital activity				
Online activity	the Degree of similarity between the real and the virtual "I"	all the values of the scale	-	-
	Preferred medium of expression of oneself	virtual	hybrid	-
	Preferred method of discussion of important topics	email, video conference	live conversation	text messages, phone call
	The purpose of the Internet visit	Diverse interests	to search for information and entertainment	for studying and cheering up
	The frequency of the information exchange in the network	-	nothing puts	Posts regularly
the Academic activity				
Digital educational preferences	Used digital educational platforms and other online services for learning	different sites	only educational platform	social network forums
Digital security	Awareness of the security of educational achievements on educational platforms	difficult to answer	yes	no
	knowledge of the articulation of the results of informal education	-	yes	no
	Awareness about digital footprint	difficult to answer	yes	no
Research activity				
Involvement in the scientific community in the digital space,	Level of involvement in scientific research	Participated several times	Never participated	Participated once
	Desire to engage in research work (SRW)	Yes	Yes	no, I'm not doing research but perhaps it would be interesting
	Shape/form of	in startups, in	in contests/	-

	participation in research	conferences, in scientific circles	competitions	
	Reason of participating in SRA	opportunity to make new important contacts	additional benefits for employment, the possibility of a successful defense of the thesis (project)	more chances to enter a master/ doctoral program
Social activity program				
Social activity in the digital space	Participation in student events	yes	can definitely do it if you want, don't think it's necessary to participate everywhere	no. The main purpose of visiting the university is to study and gain knowledge
	Membership in student organizations	consist	not consist	consist
	Attending various university events	have not yet participated, but in the near future I think I will	no, I have not attended	various events
	Sports activity	yes, I am a member of one of the national teams of the university	I play sports outside university walls	I am far from sports
Note – compiled by the authors				

Conclusions. The results of the study allowed us to identify different groups of students with different levels of digital identity. The study of the influence of behavioral factors on the processes of formation and expression of digital identity among students has led to interesting conclusions about the relationship between individual preferences and the use of digital technologies. The results of the study emphasize the importance of taking this relationship into account when developing teaching methods, as well as integrating digital tools into the learning process. Based on the conducted research, it is proposed to develop a digital profile of the student, taking into account his level of digital identity, as well as planning personalized learning trajectories in order to optimize their academic and social development.

Acknowledgements. This work was supported by the Ministry of Science and Higher Education of the Republic of Kazakhstan under the grant financing program for the 2023-2025 years by the program AP19680002 «Student's digital identity formation methodology in the circuit continuing education of universities of the Republic of Kazakhstan».

Список литературы

Ковалева А.И. Разновидности социальной идентичности: подходы к классификации. – 2019. – № 4. – С. 89-103. https://doi.org/ 10.17805/zpu.2019.4.7// Kovaleva A.I. Raznovidnosti sotsial'noy identichnosti: podkhody k klassifikatsii. Sotsiologiya i zhizn'. 2019. – № 4. – C. 89-103. https://doi.org/ 10.17805/zpu.2019.4.7

Tomlinson M., Jackson D. Professional identity formation in contemporary higher education students. Studies in Higher Education. – 2021. – Т. 46. – №. 4. – С. 885-900.

Batool S. S., Ghayas S. Process of career identity formation among adolescents: components and factors. Heliyon. – 2020. – Т. 6. – №. 9.

Korać D., Damjanović B., Simić D. A model of digital identity for better information security in e-learning systems. The Journal of Supercomputing. – 2022. – С. 1-30.

Borrás-Gené O., Serrano-Luján L., Díez R.M. Professional and Academic Digital Identity Workshop for Higher Education Students. Information 2022, 13, 490. https://doi.org/10.3390/info13100490

Московская Н.Л., Перова Е.А. Формирование цифровой идентичности обучающегося в контексте аутентичной педагогики иноязычного образования. Вестник Северо-Кавказского федерального университета. 2022. – № 6 (93). – С. 120-127. https://doi.org/10.37493/2307-907X.2022.6.14 // Moskovskaya N.L., Perova E.A. Formation of a student's digital identity in the context of authentic pedagogy of foreign language education. Bulletin of the North Caucasus Federal University, 2022, №6 (93), p.120-127. https://doi.org/10.37493/2307-907X.2022.6.14

Feher K. Digital identity and the online self: Footprint strategies – An exploratory and comparative research study. Journal of Information Science, 2021, 47(2), 192-205. https://doi.org/10.1177/0165551519879702

Borrás-Gené O., Blázquez-Sevilla A. Apuntes para El Taller de Identidad Digital 2020 #idUPM. Presented at the Taller de Identidad Digital, Madrid. 2020.

González D. Canvas de identidad digital. Diana González. 10.04.2024. Available online: https://dianagonzalezgonzalez.com/canvas-identidad-digital

Hernández-Orellana M., Pérez-Garcias A., Roco-Videla Á. Characterization of the Digital Identity of Chilean University Students Considering Their Personal Learning Environments. Future Internet 2021, 13, 74. https://doi.org/10.3390/fi13030074

Муслов С.А. Вычисление α Кронбаха при тестировании и опросах по исследованию качества жизни на занятиях по статистике в медицинских вузах. Сб. материалов XI Международной научно-практической конференции «Современные тенденции развития науки и мирового сообщества в эпоху цифровизации». – Москва, 2023. – С. 651-657. // Muslov S. A. Calculation α of the Cronbach's conjecture in testing and surveys on the study of quality of life in statistics classes in medical universities. Collection of materials of the XI Internationaloй Scientificoй and Practical Conference "Modern trends in the development of science and the world community in the era of digitalization". Moscow, 2023, pp. 651-657.

Naseem U., Razzak I. & Eklund P.W. A survey of pre-processing techniques to improve short-text quality: a case study on hate speech detection on twitter. Multimed Tools Appl 80, 35239–35266 (2021). https://doi.org/10.1007/s11042-020-10082-6

M.K. Dahouda and I. Joe. A Deep-Learned Embedding Technique for Categorical Features Encoding in IEEE Access, vol. 9, pp. 114381-114391, 2021, https://doi.org/10.1109/ACCESS.2021.3104357.

I.P. Sari, Al-Khowarizmi, I. H. Batubara. Cluster Analysis Using K-Means Algorithm and Fuzzy C-Means Clustering for Grouping Students' Abilities in Online Learning Process. Journal of Computer Science, Information Technology and Telecommunication Engineering (JCoSITTE). – Vol. 2. – No. 1, March 2021. – Pp. 139-144. https://doi.org/10.30596/jcositte.v2i1.6504

Wójcik W., Karmenova M., Smailova S., Tlebaldinova A., Belbeubaev A. Development of data-mining technique for seismic vulnerability assessment. International Journal of Electronics and Telecommunications, 2021. – Vol. 67. – No. 2. – Pp. 261-266, DOI: 10.24425/ijet.2021.135974

K.R. Shahapure and C. Nicholas. Cluster Quality Analysis Using Silhouette Score, 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), Sydney, NSW, Australia, 2020, pp. 747-748, doi: 10.1109/DSAA49011.2020.00096.

Sharma S. et al. Comparative study of single linkage, complete linkage, and ward method of agglomerative clustering. 2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon). – IEEE, 2019. – С. 568-573.

Randriamihamison N., Vialaneix N. & Neuvial P. Applicability and Interpretability of Ward’s Hierarchical Agglomerative Clustering With or Without Contiguity Constraints. J Classif 38, 363–389 (2021). https://doi.org/10.1007/s00357-020-09377-y

F. Liu and Y. Deng. Determine the Number of Unknown Targets in Open World Based on Elbow Method. IEEE Transactions on Fuzzy Systems. – Vol. 29. – No. 5. – Pp. 986-995, May 2021, https://doi.org/10.1109/TFUZZ.2020.2966182.

Information about authors

Tlebaldinova Aizhan Soltangaliyevna – PhD, Associate Professor, D. Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan, ATlebaldinova@edu.ektu.kz, +77057521522

Kumargazhanova Saule Kumargazhanovna – Candidate of Technical Sciences, Associate Professor, D. Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan saule.kumargazhanova@edu.ektu.kz, +77771839377

Smailova Saule Sansyzbaevna – PhD, D.Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan, ssmailova@edu.ektu.kz, +77771839752

Konurbayeva Zhadyra Tusuphanovna – Candidate of Economic Sciences, D. Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan, ZhKonurbayeva@edu.ektu.kz, +77771827250

Seitahmetova Zhanat Maratovna – PhD, D. Serikbayev East Kazakhstan University, Ust-Kamenogorsk, Kazakhstan, seitahmetova_zh@ukk.nis.edu.kz, +77053152123