Batyrbek Bakytkereiuly
Master's student
batyrbek.bakytkereiuly@astanait.edu.kz, orcid.org/0009-0008-6716-2457
Astana IT University, Kazakhstan

Andrii Biloshchytskyi
Doctor of Technical Sciences, Professor, Vice-rector for Science and
Innovation
a.b@astanait.edu.kz, orcid.org/0000-0001-9548-1959
Astana IT University, Kazakhstan

Dilara Abzhanova
Master of Science, General manager, Center of Competence and Excellence
dilara.abzhanova@astanait.edu.kz, orcid.org/0000-0002-7988-3971
Astana IT University, Kazakhstan

DEVELOPMENT OF A METHOD FOR COMBINING DATA IN ORDER TO PREVENT DUPLICATION OF RECORDS IN THE DATABASE OF THE INFORMATION SYSTEM FOR THE DEVELOPMENT OF METHODOLOGICAL COMPETENCE OF TEACHERS OF IT DISCIPLINES

Abstract: The article assesses the outcomes of digitization on higher education in Kazakhstan. Given current trends, there is a need to establish a dedicated focus on the ongoing professional growth of educators. A crucial measure in this regard involves shifting from disjointed learning approaches to implementing a cohesive system for the continuous professional development of teaching staff.

It is crucial to highlight that in accordance with the Development Concept of Higher Education and Science in the Republic of Kazakhstan the duration of an individual's active economic engagement has extended from 35-40 years to 50-60 years. This shift underscores the growing necessity for continuous learning throughout one's life and underscores the significance of non-formal education.

The educator plays a pivotal role in influencing the educational landscape and significantly contributes to the outcomes of socio-economic changes in Kazakhstan. In the current era of socio-economic and digital transformations the teacher's proficiency has emerged as a crucial element that directly influences the quality of students' education.

The Ministry of Digital Development, Innovation, and Aerospace Industry of Kazakhstan has reported a yearly requirement for approximately 30,000 additional IT professionals. Consequently, the focus is on enhancing the methodological expertise of IT discipline educators, recognizing their pivotal role in influencing the quality of education and student achievements. Information systems aimed at enhancing this competence are emerging as a crucial element in elevating the overall effectiveness of education.

The article presents the rationale for the need to create a single space for additional professional education of teachers of IT disciplines. It is considered as an effective mechanism for the implementation of the state educational policy and innovation strategy in the field of ed-

ucation. The aim of this initiative is to ensure that the human resource is ready to successfully achieve the country's strategic goals in the context of digitalization processes.

The article focuses on the development of a structural model of an information system aimed at improving the methodological competence of teachers of IT disciplines using the principles of continuing education. Based on a conceptual model with microservices subsystems, this methodology represents a comprehensive approach to the development of the competence of teachers of IT disciplines, integrating technological and methodological aspects.

The development of a method for combining data in order to prevent duplication of records in the database of the information system is also considered. Methods of estimating the degree of similarity based on the types of attributes are proposed.

The article extensively delineates the models, architecture, and components constituting the software implementation of the envisioned information system. It scrutinizes the accessibility of the employed data sources and formulates conclusions regarding the future potential for advancing the subject matter.

Keywords: digitalization; methodological competence; information system; professional development; IT education; structural models of information systems; unification method; databases.

Introduction

Digitalization has emerged as a pivotal trend in the worldwide economy. Leading global economies have incorporated digitalization programs into their economic development strategies. Over the past decade or so, the implementation of digitalization at the national level has gained momentum, marking an active phase in the digital transformation of the global economy.

Information technologies are widely employed globally across diverse industries, becoming indispensable in contemporary times. The pressing challenge lies in the widespread adoption of these technologies across various facets of human endeavors.

Experts predict that in ten years, up to 70% of the world's products and services will be based on digital platform models. Under these conditions, Kazakhstan sets itself a strategic task – to become one of the largest digital hubs in the Eurasian space [1].

It is predicted that in the near future, 90% of jobs will require a certain level of digital professional skills. However, currently, even in Europe, only 53% of residents have the necessary competencies.

The digitalization process involves crucially developing and enhancing systems for knowledge formation. Various domains such as modern economics and science demand expert professionals with advanced skills to meet the evolving needs of human activities.

In July 2023, President of the Republic of Kazakhstan Kassymzhomart Tokayev named five key priorities for the country's digital transformation, one of which is strengthening human capital. Just two years ago, the global IT market needed 30 million specialists. Today, the demand is already 100 million. And by 2025 it will increase to 200 million [1].

Hence, a key benchmark outlined in the Action Plan to execute the Digital Transformation Concept, focusing on advancing the information and communication technologies as well as the cybersecurity industry from 2023 to 2029, involves the goal of educating a minimum of 100,000 highly skilled IT professionals by the year 2025 (Table 1).

Table 1. Target indicator 5. Training of highly qualified personnel [2]

Year	2023	2024	2025	2026	2027	2028	2029
Quantity	35k	35k	35k	35k	35k	35k	40k

From the perspective of this target indicator, the main result of educational activity is the training of a competent specialist.

However, there is a shortage of personnel in the field of education, including qualified ones. The current shortage of teaching staff annually amounts to more than 5 thousand people. It will constantly grow due to the increase in the number of students. 20% of graduates of pedagogical specialties do not enter the profession every year. According to the Head of State of the Republic of Kazakhstan, the programs of pedagogical specialties in universities are largely based on outdated methods and methodologies that do not take into account modern realities [1].

Literature review

An analysis in the field of continuing education of teachers of IT disciplines has shown that technologies for the formation of modern competencies have not become the norm of professional activity of teachers, since there is no single space for professional development and methodological support (post-course support), for the following reasons:

- the lack of inclusion of aspects contributing to practical training in the curricula of universities and colleges is explained by a reduction in time, outdated methodology, subject-centrism and insufficient training of the teaching staff;

- the lack of a single space for professional development of teachers of IT disciplines is characterized by the lack of common principles, content requirements, as well as the lack of risk- and practice-oriented approaches.

- new technologies are not successfully integrated into the regular activities of a teacher due to the incompleteness of the innovation process, insufficient support for innovations and insufficient risk orientation.

- routine procedures that are not related to educational activities overload the teacher of IT disciplines, which reduces the automation of these procedures and leads to professional burnout, which makes it difficult to master new technologies.

- heads of educational organizations do not show willingness to integrate new technologies into the professional activities of teachers of IT disciplines, not considering them as the norm [3].

Thus, the existing model of professional growth of teaching staff of IT disciplines no longer provides in its current form the solution of strategic tasks of digitalization, and needs a radical transformation, since it is focused on the pragmatic task of passing certification for a category, and not on ensuring readiness to solve modern educational problems.

To overcome the identified difficulties, it is necessary to create a unified digital space that promotes the continuous development of methodological competencies of teachers of IT disciplines. This space should also perform the function of advanced professional development aimed at developing the competencies necessary to work in the face of modern challenges and constant changes.

The processes of storing and processing information are an important component of any information system. The growth of information needs caused by both internal and external factors can be satisfied only if there is consistent, relevant and correct data.

One of the most important parts of the database of the developed information system is a single user information model representing data about a teacher registered in the system. However, the evaluation of the database conducted by the authors showed that there are duplicate records of data about teachers in it. Based on a detailed analysis, typical situations leading to data duplication in the database were identified:

1. The minimum intersection of characteristics with the same values. For example, two persons have known the values of the characteristics "First_name", "Last_name", and "Birth_day", while the values of the first two characteristics are the same, and the values of the third characteristic are different. In such situations, it is difficult to make a decision about their identity.

2. Inconsistencies in attribute values, that is, the discrepancy between the actual and registered values. The following types of errors can be distinguished: typos when entering (for example, incorrect spelling of the surname), errors as a result of loss or distortion of data, errors related to the difference in time of updating values that occur, for example, when changing the surname or address.

In such conditions, the use of manual processing by the administrator can be completely excluded from the identification process, since this process is impractical and requires considerable time. An effective solution is to develop a method for detecting duplicates, which will serve as the basis for creating an information system.

In the research conducted by a group of authors [4-5], the structural model of the system for the development of methodological competence of teachers of higher educational institutions and methods for evaluating its effectiveness are considered.

However, despite the importance of these studies, there is some gap in the literature review covering the theory and practice of using a single digital space in the format of an information system to improve the methodological competence of teachers of IT disciplines. This aspect, in particular, requires additional attention, since it has been little studied in the context of the development of appropriate systems that are individually customized to the profile of competencies with subsequent tutor support.

Other studies devoted to the theoretical foundations of the formation of methodological competence of teachers of higher educational institutions [8-9] offer a comprehensive approach to the assessment and development of this competence. These works emphasize the importance of supporting and assisting the university administration in the process of forming the methodological readiness of teachers.

Special attention is paid to research on methods for identifying and eliminating duplicates in records conducted by foreign scientists such as Newcombe [10], Fellegi and Sunter [11], Winkler [12], Jaro [13], and others. These studies are relevant in the context of integrating various data warehouses, preprocessing data for analysis and solving data mining tasks.

Taking into account the literature review and the identified gaps, it can be concluded that despite active research in the field of methodological competence of teachers and the search for duplicates in records, there are no specific studies devoted to the development of a system combining these aspects.

This project involves the development of a proprietary method for merging duplicate records in the database, including attributes not only of the surname, but also other characteristics of teachers, such as first name, patronymic, date of birth, name of the completed educational institution, city, home address and others. It is assumed that the development of such a system will make it possible to effectively process duplicates and create more accurate and extensive profiles of teachers in the database.

This approach will include not only technical aspects of merging records, but also methodological ones aimed at optimizing the data management process and improving the accuracy of information about teachers. Taking into account existing problems such as errors in attribute values and the intersection of properties with matching values, the development of a merge method will be an important step in improving data quality and ensuring the effective functioning of the information system.

Thus, the upcoming research aims to create an innovative method of merging duplicate records, which, based on a thorough analysis of the literature, will contribute to more effective data management in the educational environment and improve the accuracy of information about teachers.

Research methodology

Theoretical analysis of scientific papers; content analysis of periodical materials, regulatory documents and static data on the research problem; analysis and communication of existing experience in designing similar information systems; system analysis in order to determine the scientific basis for improving the method of combining records in databases in information systems.

Main part

The proposed architecture of the system under consideration is a carefully designed solution aimed at optimizing effective interaction between the components of the backend, interface and database (Figure 1). Consistency between these components is the foundation for a system specifically adapted for educational purposes and is highly adaptable to a multitude of simultaneous user requests.

In the context of the architecture of the server part of the system, special attention is paid to details when using Java Spring Boot, an integrated platform known for its professionalism in developing server applications. This choice of architecture provides the system with elegant processing of a significant number of parallel user requests, which is an integral characteristic in the context of an educational information system. Taking advantage of the capabilities of Java Spring Boot, the system provides smooth and responsive interaction with users, fully meeting the main goals of the educational system [14].

In a similar vein, the system interface is carefully designed using React JS, a technology noted for its ability to create reliable and user-friendly user interfaces. The intuitiveness of the interface provided by React JS is of paramount importance, especially in an educational system where ease of use plays a key role. The interface seamlessly integrates various system functions, including, but not limited to, training, forecasting, and post-course support. All these characteristics together contribute to the creation of a user environment focused on an effective educational experience [15].

Moving to the next level of the architectural spectrum, the database component of the system is implemented using PostgreSQL, widely known for its scalability and reliability in data management. The role of the database in the system is a key function in providing an extensive data warehouse covering information about users, courses and more. With a special focus on data integrity and confidentiality, PostgreSQL ensures that confidential information in the system remains securely protected from unauthorized access [16].

Within this complex architectural ecosystem, the interaction between the backend and the interface is organized using an application programming interface (API), which ensures efficient transfer and processing of critical data between these main components. This API acts as a channel for user requests sent to the server side for subsequent careful processing and provision of necessary responses. It is important to note that the database actively interacts with the backend, providing meaningful data, while the backend interacts with each other, updating the database with any new generated information.

Upon a detailed examination of the server component, it becomes obvious that its functionality is based on a variety of services, each of which plays a separate but integral role in ensuring the successful operation of the system.

For example, the file storage service is responsible for efficiently handling downloaded media by acting as storage for these files. This provides convenient user access to a shared repository of important documents and multimedia resources.

On the other hand, the user service represents a key element of the web application structure, carefully managing user accounts. Its role includes user registration, maintaining and updating user data, verifying email addresses to enhance security, assigning roles for precise

access control, user search functionality, and user data extraction. The user service is an essential element that ensures smooth and secure interaction with users, carefully controlling account management and access in the application.
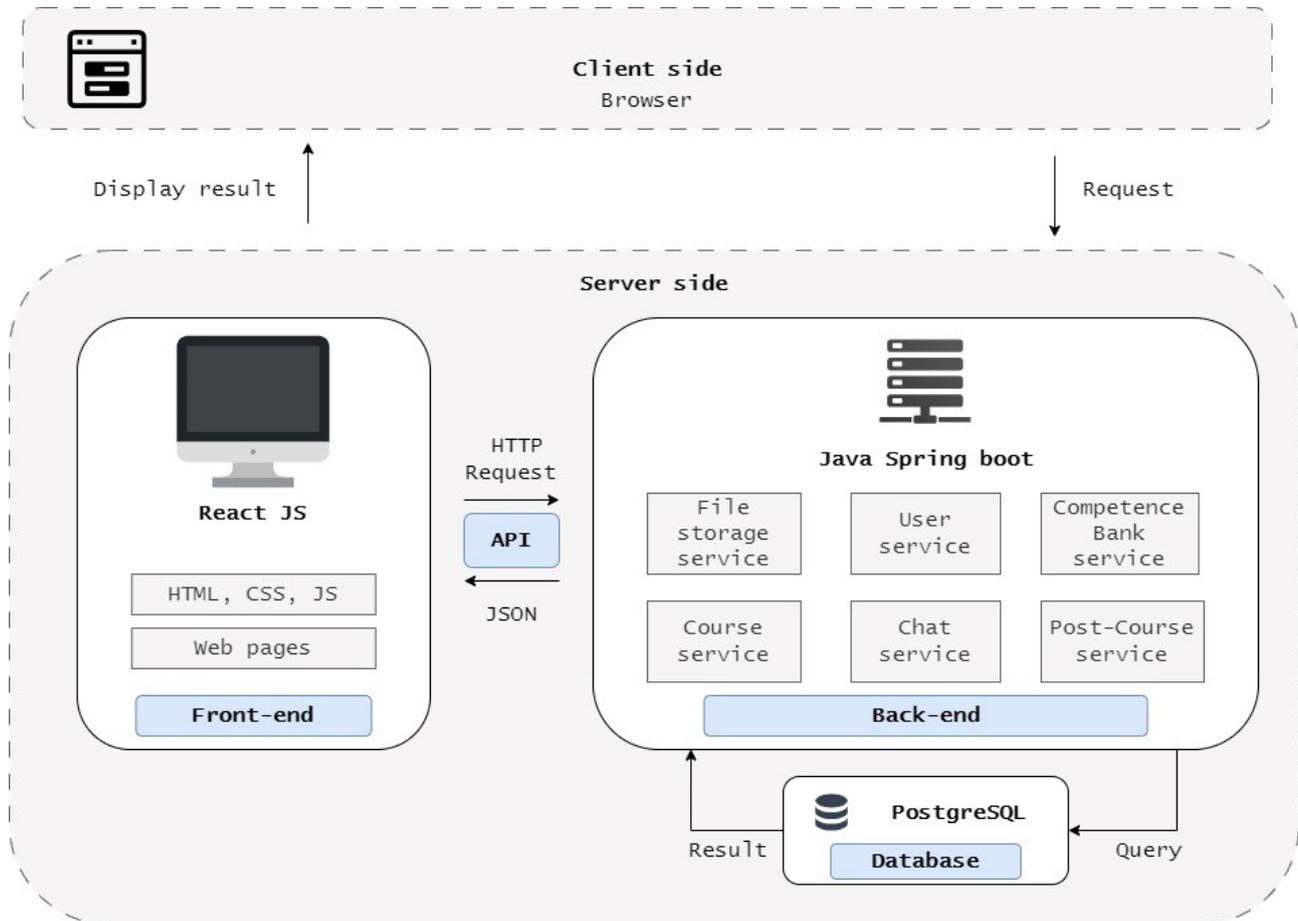

Figure 1. High-level architectural model of the system

The interaction between the interface and server systems (Figure 2) is a key component of the integrated architecture of the information technology education platform, providing an effective channel for operational data transmission. This interaction forms the basic structure of the system's functionality, expressed in an enriched and user-oriented experience. The importance of the server system in this context is undeniable, since it functions as an extensive repository of key information, including user profiles, teacher data, courses, data after their completion, the creation of competence maps, statistics and questionnaires. On the other hand, the interface system is responsible for interacting with the competence map and providing system data to users [18].

The fundamental data transfer mechanism is organized using the REST API (a representative interface for transmitting application state). This REST API acts as a link between the two systems, providing the interface with the initiation of interaction with the backend. To implement this interaction, the interface uses HTTP requests directed to the backend with specific requests for data extraction. In response to these requests, the server part processes the information and transmits the results in the form of JSON data (JavaScript object notation).

© Batyrbek Bakytkereiuly, Andrii Biloshchytskyi,
   Dilara Abzhanova
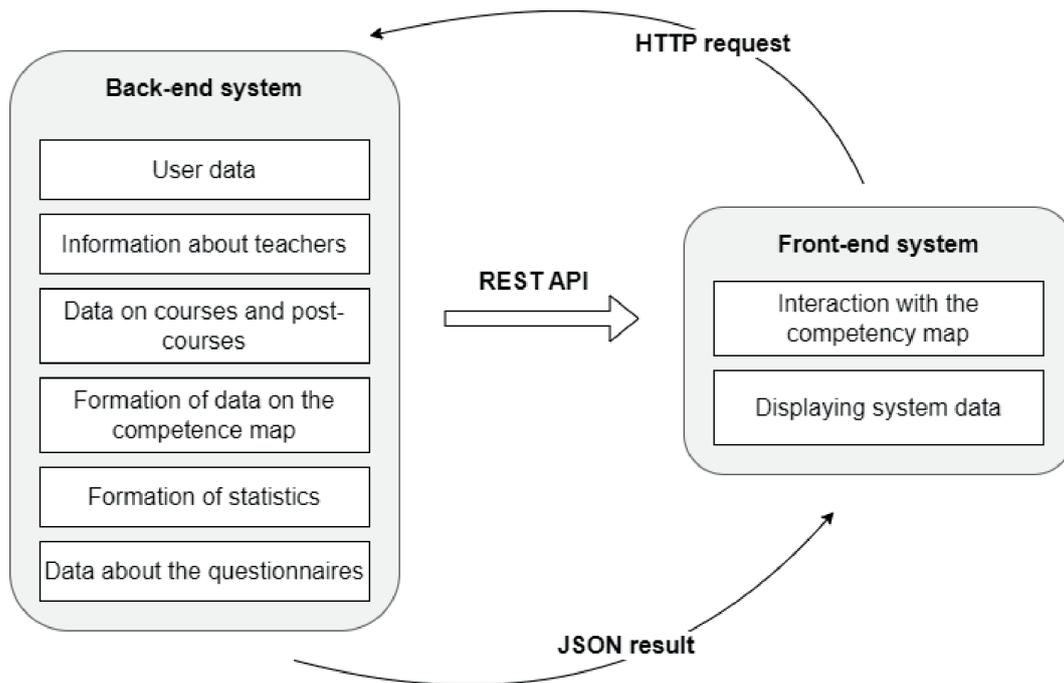
Figure 2. The structure of data transfer between the interface and server systems

The architectural component that plays a key role in the system is the Entity Relationship Diagram (ERD) (Figure 3). This diagram visualizes the primary data objects and their relationships within the system, including IT teachers, trainers, users, courses, a Competency Map, post-course training, analytics, and a competency bank. For example, the teacher database contains a variety of information, providing a holistic profile of teachers, including general information, education, professional development and scientific activities.
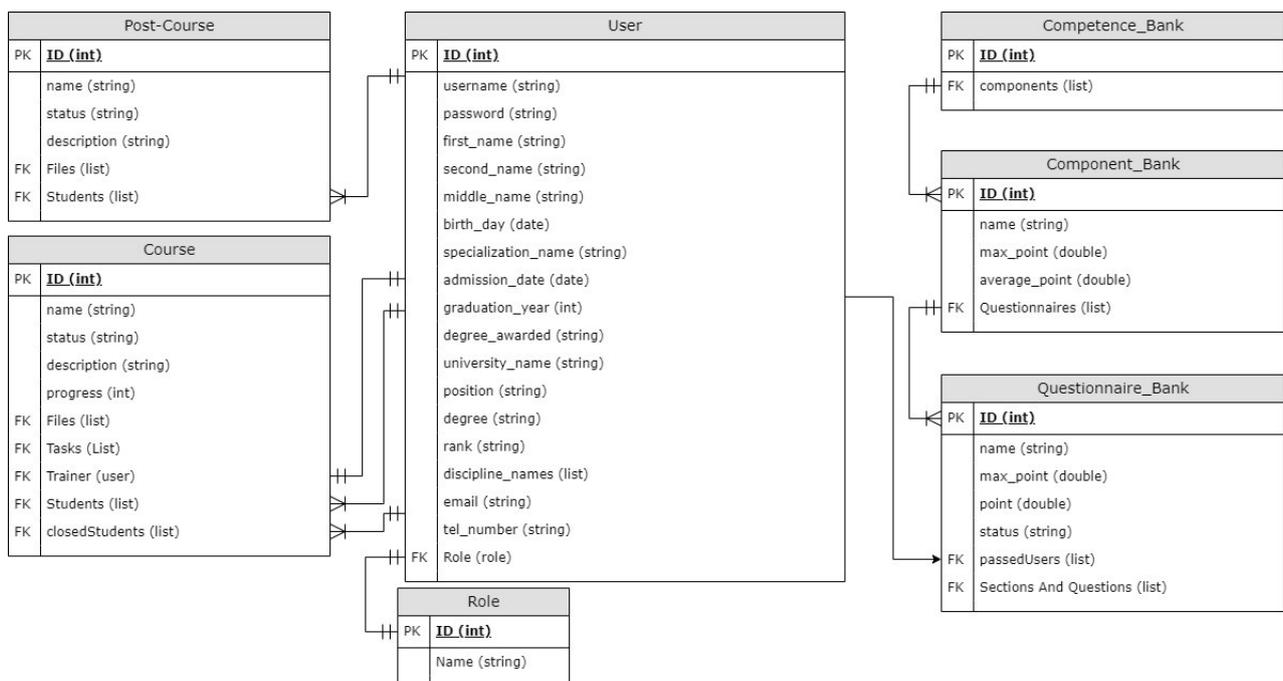
Figure 3. A fragment of the structural model of the information system database

Solution methods and discussions

Understanding the ERD and the objects of the system plays an important role in building a dynamic educational environment. In this context, the standard record merging system (Figure 4) proposed in [19] builds a link between architecture and database.
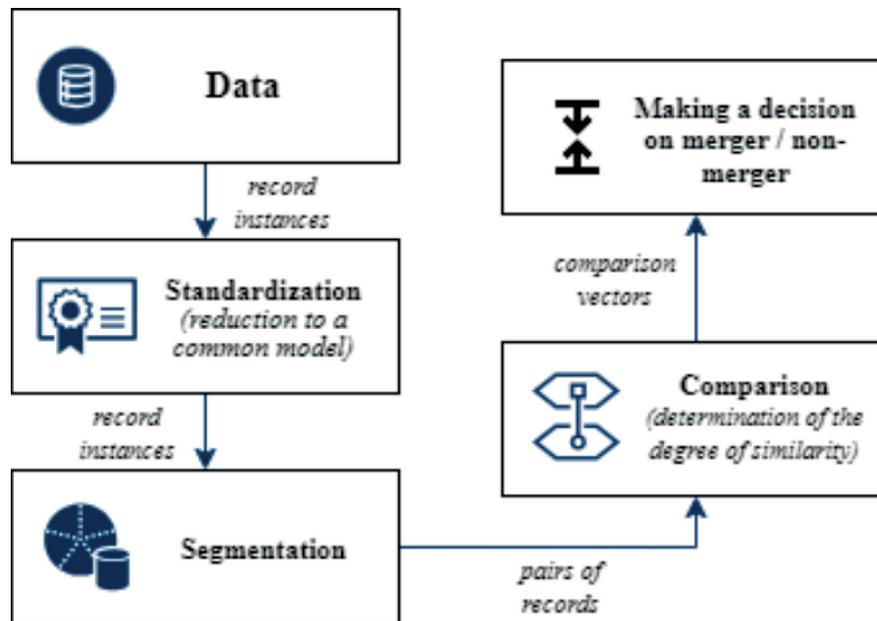
Figure 4. The process of identifying and deleting duplicate records in the database

The standardization stage includes bringing the data to a common structure and types. The segmentation stage is performed in cases where, from the point of view of efficiency, it is impractical to compare the entire array of pairs of records. Instead, they are divided into segments according to a certain criterion, and then an assessment of the similarity within each selected segment is carried out.

The process of comparing and deciding whether to merge records is the most difficult stage. There are many methods such as neural networks, cluster analysis, and others. Probabilistic algorithms seem to be the most suitable for solving this problem, which allow you to obtain an interval estimate of similarity based on the analysis of attribute values. These algorithms are widely described in foreign literature and have been successfully used in conducting population censuses, integrating medical and postal databases. The main advantages of such algorithms are the presence of a formal apparatus, the relative simplicity of implementation and the ability to process missing and erroneous values.

This database is designed to store information about users and manage the learning process. The main entities include "User", "Role", "Competence_Bank", "Component_Bank", "Questionnaire_Bank" and "Course". Each user is described by a unique identification number, username, password, personal information, date of birth, education information, email address, phone number and role (defined in the "Role" table). Roles define user access levels. The competence bank includes components that assess the level of competence, and the questionnaires contain questions, maximum scores and status. The users who completed the questionnaires are linked to the corresponding entries in the "User" table. The courses include information about the title, status, description, progress, coach and students. Each course is linked to users via foreign keys, providing communication between participants.

During the registration of new users, the database collects a variety of information necessary for the full use of the platform. The user must provide the following information: user-

name, password, full name, date of birth, email address and contact phone number. All entered data is entered in the corresponding fields of the database attributes.

As a methodological basis for further implementation in a real information system, an abstract mathematical model of the probabilistic method of combining records was developed, corresponding to the above requirements.

Let two sets of objects be given (users are registered teachers in the information system) X and Y, the elements of which are denoted as x and y, respectively.

Consider the set:

$$X \times Y = \{(x, y); x \epsilon X, y \epsilon Y\},$$

as two subsets of A and B. If x and y in an element of the set $X \times Y$ coincide (they are the same object), then this element belongs to A, otherwise it belongs to B.

Suppose that when entering information about elements X and Y into the database, the Z_x and Z_y record files were created. Let's denote the entries corresponding to the elements of the sets X and Y as $\chi(x)$ and $\gamma(y)$. The result of comparing two records is a comparison vector, which we denote as **v**. It includes elements such as "Name matches", "Date of birth does not match" and etc.

The comparison vector **v** is defined as a vector function over $\chi(x)$ and $\gamma(y)$ and has the following representation:

$$v\,[\chi(x),\, \gamma(y)] = \{v^{\,1}[\chi(x),\, \gamma(y)], \ldots,\, v^{n}[\chi(x),\, \gamma(y)]\}.$$

Next, we will use the notation $v\,(x,y)$, $v\,[\chi,\, \gamma]$ or simply v.

If a pair of records represents the same object and the comparison vector m reflects the coincidence of values, then we denote the probability (P) of this event for this object:

$$a(v) = P(v|(x,y) \in A).$$

However, if a pair of records represent different objects and the comparison vector v reflects the coincidence of values, then we denote the probability (P) of this event:

$$b(v) = P(v|(x,y) \in B).$$

Consider the ratio $a(v) / b(v)$ as the degree of similarity of records. Nevertheless, with a significant number of v implementations, it becomes necessary to make certain assumptions. In accordance with the methodology proposed by the authors [11], let's assume that the components of v can be ordered and displayed as follows:

$$a(v) = a_1(v^1) * a_2(v^2) * \ldots * a_n(v^n),$$
$$b(v) = b_1(v^1) * b_2(v^2) * \ldots * b_n(v^n).$$

Imagine that v^1 can include the attributes of the "First_name" field, and v^2 can include the attributes of the "Birth_day" field. Let's apply the logarithm to ensure the additivity of the weights:

$$w^i\left(v^i\right) = log\left(\frac{a(v^i)}{b(v^i)}\right).$$

Thus, we can write $w(v)$ as the weight of the compound and write as follows:

$$w(v) = w^1 + w^2 + \ldots + w^n.$$

The degree of similarity of the records, expressed by the weight of the connection, represents the probability that two records reflect the same object.

Suppose that one of the attributes of the records is the "Last_name" field, let's build a list of all the unique values of the surname without errors. Next, we will calculate the number of

teachers who have each of these surnames in some test sample. Let's denote the proportion of the occurrence of the k-th surname in this set as M_k.

Let's introduce the following notation:

- p_X and p_Y are the probabilities of an erroneous surname value in Z_X or Z_Y.

- p_e – the probability that the values of the teacher's last name in Z_X and Z_Y are different, provided that they are written down without errors (for example, in the case of a change in a person's last name).

Rule 1: provided that the surname matches or the k-th surname is on the list:

$$w = log\left(\frac{1}{M_k}\right), \tag{1}$$

Rule 2: provided that the last name does not match, we calculate:

$$w = log\left(\frac{p_x + p_y + p_e}{1 - \sum_k M_k^2}\right), \tag{2}$$

Rule 3: provided that the last name is not specified in any record:

$$w = 0 \tag{3}$$

It follows from the above-mentioned rules (1-3) that a match by surname leads to a positive weight, and the weight increases with a decrease in the frequency of occurrence of the surname; a mismatch causes the appearance of a negative weight, which decreases taking into account the errors p_X, p_Y, p_e; and in the absence of a surname in one of the entries, the weight will be zero. Figure 5 shows a conceptual schema of a database for storing proportions, which helps to increase the effectiveness of the method (Fig. 5)
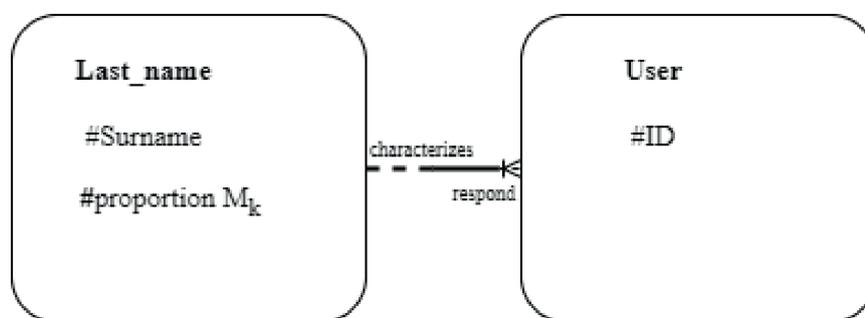

Figure 5. Conceptual diagram of a database for storing proportions

The initial proportions should be calculated based on samples with minimal redundancy. With the arrival of new records in the database, at regular time intervals or when a certain amount of data is reached, the list of proportions must be updated.

An important practical task is to set the values of p_X, p_Y and p_e. A characteristic feature of many organizations is the presence of multiple sources of updating objects of the same type. Each of the sources has a different degree of reliability of the information entered, therefore, the sets Z_X and Z_Y are a set of subsets, for each of which it is possible to determine the eigenvalues p_X, p_Y and p_e in relation to a particular attribute. For example, let's imagine the set Z_Y, meaning the set of teachers registered in the database, in the form $Z_{y_1} \cup Z_{y_2} \cup ... \cup Z_{y_k}$, where Z_{y_k} is a subset of personalities actualized by the k-th type of accounting. Since each type of accounting can be compared with the probability of an error in the value of a property, then, for example, for a surname $P_y^{ln} = \{p_{y_1}^{ln}, p_{y_2}^{ln}, ..., p_{y_i}^{ln}\}$.

For a variety of Z_X personalities, the following formulations can be applied when:

1. creating a new identity: $Z_X = \{z'\}$, p_X – corresponds to the type of accounting that performs the creation operation;

2. searching for duplicates among registered personalities: $Z_X = Z_Y$;

3. importing data from an offline source into the system database: $Z_X = \{z_1, z_2, ..., z_l\}$, p_X must be defined for this source.

Thus, the algorithm for comparing pairs of records included in $Z_X \times Z_Y$ can be represented as a task for comparing pairs included in the sets $Z_X \times Z_1, Z_X \times Z_2, ..., Z_X \times Z_{y_x}$ at the corresponding error levels p_X and $p_{y_1}, p_{y_2}, ..., p_{Y_x}$ defined for for each attribute

Conclusion

This method of calculating weights provides a universal and flexible mechanism applicable not only to the attributes of the surname, but also to all other relevant fields characterizing the teacher. Attributes such as first name, patronymic, date of birth, name of the completed educational institution, city, home address, and many others influence the uniqueness of the teacher's identification in the system.

The application of this method to these attributes enriches and deepens the identification process considering the many characteristics of the teacher's personality. In particular, taking into account additional attributes allows you to mitigate the effects of possible errors in one of the attributes, providing the system with more information for accurate matching of records.

Thus, the developed method is not limited only to the surname but provides expanded opportunities for effective identification of teachers in the system, while ensuring high accuracy and reliability of the process. This has made the method scalable and applicable in various contexts, providing operational solutions to the problem of duplicate records in the information system under development and similar systems.

References

[1] ILS "Adilet." (2023, March 28). *Ob utverzhdenii Koncepcii razvitija vysshego obrazovanija i nauki v Respublike Kazahstan na 2023 – 2029 gody.* https://adilet.zan.kz/rus/docs/P2300000248/info

[2] ILS "Adilet." (2023, March 28). *Ob utverzhdenii Koncepcii cifrovoj transformacii, razvitija otrasli informacionno-kommunikacionnyh tehnologij i kiberbezopasnosti na 2023 - 2029 gody.* https://adilet.zan.kz/rus/docs/P2300000269

[3] Kassenov, K., Mukhatayev, A., Omirbayev, S., Omarova, S., & Toxanov, S. (2022). Domestic and international experience in organizing the process of professional competence development of teachers. *Vestnik Torajgyrov Universiteta*, 4.2022, 331–342. https://doi.org/10.48081/jbrv7653

[4] Biloshchytskyi, A., Omirbayev, S., Mukhatayev, A., Biloshchytska, S., Toxanov, S., & Faizullin, A. (2023). IT Project Administration Tools on Creation of Information and Training Platforms. *SIST 2023 - 2023 IEEE International Conference on Smart Information Systems and Technologies.* https://doi.org/10.1109/sist58284.2023.10223592

[5] Biloshchytskyi, A., Omirbayev, S., Mukhatayev, A., Faizullin, A., Toxanov, S., & Kassenov, K. (2020). Research on the Formation Level of Methodological Competence of it Disciplines Teachers. *ATIT 2020 - Proceedings: 2020 2nd IEEE International Conference on Advanced Trends in Information Theory.* https://doi.org/10.1109/atit50783.2020.9349337

[6] Omirbayev, S., Mukhatayev, A., Biloshchytskyi, A., Toxanov, S., Faizullin, A., & Biloshchytska, S. (2022). Structural model of the system of development of methodological competence of it-discipline teachers on the basis of continuing education. *Scientific Journal of Astana IT University*, 122–138. https://doi.org/10.37943/12tyuy5628

[7] Biloshchytskyi, A., Omirbayev, S., Mukhatayev, A., Kuchanskyi, O., Biloshchytska, S., Andrashko, Y., Toxanov, S., & Faizullin, A. (2023). A structural model for building a system for the development

of methodological competence and methods for evaluating its effectiveness. *Eastern-European Journal of Enterprise Technologies*, 5, 6-22. https://doi.org/10.15587/1729-4061.2023.289045

[8] Sharifbaeva, K., Niyazova, G., Abdurazzakova, D., Abdurashidov, I., & Alimardonov, R. (2022). Formation of methodical competence of special subjects teachers in technical universities. *AIP Conference Proceedings*. https://doi.org/10.1063/5.0089618

[9] Konovalov, A. (2023). Methodical Competence Deficits of Vocational Training Teachers. *Education & Self Development*, 18 (2), 81–99. https://doi.org/10.26907/esd.18.2.07

[10] Newcombe, H. B., Kennedy, J. G., Axford, S., & James, A. P. (1959). Automatic linkage of vital records. *Science*, 130(3381), 954–959. https://doi.org/10.1126/science.130.3381.954

[11] Fellegi, I.P., & Sunter, A.B. (1969). A theory for record linkage. *Journal of the American Statistical Association*, 64(328), 1183–1210. https://doi.org/10.1080/01621459.1969.10501049

[12] Winkler, W. E. (2000). Frequency-based matching in Fellegi-Sunter model of record language.

[13] Jaro, M.A. (2002). Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida.

[14] Ouadoud, M., Rida, N., & Chafiq, T. (2021). Overview of E-learning platforms for teaching and learning. *International Journal of Recent Contributions from Engineering Science & IT*, 9(1), 50. https://doi.org/10.3991/ijes.v9i1.21111

[15] Colace, F., De Santo, M., & Vento, M. (2003). Evaluating on-line learning platforms: a case study. *36th Annual Hawaii International Conference on System Sciences*. https://doi.org/10.1109/hicss.2003.1174342

[16] Tungpantong, C., Nilsook, P., & Wannapiroon, P. (2021). A Conceptual Framework of Factors for Information Systems Success to Digital Transformation in Higher Education Institutions. *9th International Conference on Information and Education Technology*, 57–62. https://doi.org/10.1109/iciet51873.2021.9419596

[17] Fernández-Díaz, E., Rodríguez-Hoyos, C., & Calvo Salvador, A. (2017). The pedagogic architecture of MOOC: A research project on educational courses in Spanish. *International Review of Research in Open and Distributed Learning*, 18(6), 18-35. https://doi.org/10.19173/irrodl.v18i6.2964

[18] Gorbunovs, A., Kapenieks, A., & Kudina, I. (2013). Competence development in a combined assessment and collaborative e-portfolio information system. *Procedia Computer Science*, 26, 79-100. https://doi.org/10.1016/j.procs.2013.12.009

[19] Gu, L., Baxter, R.A., Vickers, D., & Rainsford, C.P. (2003). CSIRO Mathematical and Information Sciences Technical Report. *CSIRO Mathematical and Information Sciences*. https://www.researchgate.net/profile/Lifang_Gu/publication/2479819_Record_Linkage_Current_Practice_and_Future_Directions/links/0deec52969a26cf1b0000000.pdf