UDC 65.015.12
DOI: 10.15587/1729-4061.2024.314129

ONTOLOGY MODELING FOR AUTOMATION OF QUESTIONNAIRE DATA PROCESSING

Kainizhamal Iklassova
PhD, Associate Professor
Department of Information and Communication Technologies**

Aliya Aitymova
Corresponding author
PhD, Senior Lecturer*
E-mail: aitimova_a_1985@mail.ru

Oxana Kopnova
Senior Lecturer
Department of Mathematics and Informatics**

Anna Shaporeva
PhD, Senior Lecturer, Head of Department
Department of "Building and design"**

Gulmira Abildinova
PhD, Associate Professor
Department of Informatics
N. Gumilyov Eurasian National University
Satbayev str., 2, Astana, Republic of Kazakhstan, 010000

Zhanat Nurbekova
Professor-Researcher
Non-Profit Limited Company "Abai Kazakh National Pedagogical University"
Dostyk str., 13, Almaty, Republic of Kazakhstan, 050010

Leila Almagambetova
Director
Branch of the Joint-Stock Company National Center of Professional Development «Orleu» Institute for Professional Development in North Kazakhstan Region»
Truda str., 36, Petropavlovsk, Republic of Kazakhstan, 150000

Alexey Gorokhov
Senior Lecturer, Master's Degree*

Zhanat Aitymov
Teacher
Municipal State-Owned Enterprise "Higher Construction and Economic College"
N. Nazarbayev str., 262, Petropavlovsk, Republic of Kazakhstan, 150008
*Department of Theory and Methods of Primary and Preschool Education**
**Non-Profit Limited Company
"Manash Kozybayev North Kazakhstan University"
Pushkin str., 86, Petropavlovsk, Republic of Kazakhstan, 150000

The object of this study is the analysis of questionnaire data using ontological modeling. The task relates to the fact that conventional methods for processing questionnaire data are often insufficiently effective when working with large volumes of information and do not make it possible to automate many analysis processes.

As a result of the study, an ontology was designed that structures and analyzes questionnaire data, which allows for a more accurate identification of hidden relationships between variables. Using these theoretical provisions, an information system for assessing the quality of assimilation of preschool children's competencies was built. 150 children from various preschool organizations were involved in the study as respondents. The data integration method proposed in this paper significantly facilitated the process of data analysis both for a group and for an individual respondent.

The key difference of the proposed methodology is the automation of routine data analysis operations based on the ontological structure, which significantly simplifies the processing of large volumes of information. This makes it possible to solve the problem of limitations in conventional analysis methods and makes data analysis more scalable and reproducible.

The practical application of the results is possible in marketing for analyzing customer satisfaction, market segmentation, and evaluating the effectiveness of advertising campaigns. In the educational domain, the ontology could be used to evaluate the quality of programs and analyze respondents' opinions, and in sociology – to analyze public opinion and conduct research on social phenomena.

Thus, the proposed ontology provides an effective tool for analyzing large volumes of questionnaire data, allowing organizations to make more informed decisions and improve their efficiency

Keywords: ontological modeling, questionnaire data, data integration, automation of decision-making systems, questionnaire data analysis, preschool education

Received date 12.07.2024
Accepted date 03.10.2024
Published date 31.10.2024

***How to Cite**: Iklassova, K., Aitymova, A., Kopnova, O., Shaporeva, A., Abildinova, G., Nurbekova, Z., Almagambetova, L., Gorokhov, A., Aitymov, Z. (2024). Ontology modeling for automation of questionnaire data processing. Eastern-European Journal of Enterprise Technologies, 5 (2 (131)), 36–52. https://doi.org/10.15587/1729-4061.2024.314129*

1. Introduction

The modern world is experiencing an era of digital transformation when information is becoming the basis for making management decisions in all areas of activity. One of the most important sources of such information is questionnaires, which are widely used in commercial structures, as well as in education, marketing, sociology, and other fields. Effective processing of questionnaire data is critical to enable organizations to extract useful information from this data, identify hidden patterns, and make informed decisions. However, conventional methods for analyzing questionnaire data are often ineffective. For example, computer surveys have shown that the percentage of missed responses was 1.7 %, compared to 3.3 % for paper surveys [1]. Mail questionnaires may receive a large number of responses, as evidenced by a study according to which 81 % of respondents recently responded to questionnaires [2]. Although questionnaires are a valuable tool, their effectiveness can be reduced by factors such as bias in the perception of questionnaires and respondent comprehension, which highlights the need for careful design and testing [3]. With a significant increase in the volume of data, the complexity of their processing increases, and the human factor can lead to interpretation errors. As a result, there is a need to devise new methods that would ensure the systematization, structuring, and automation of the analysis of questionnaire data.

As the volume of data grows, the complexity of its processing increases, the probability of human errors increases, which significantly complicates the interpretation of results. This makes it relevant to implement new methods, such as the ontological approach, which make it possible to formalize knowledge about the subject area, automating data processing, and increasing the accuracy of their analysis.

Ontologies are formalized structures that are used to represent and analyze knowledge in various subject areas. In the context of analyzing questionnaire data, the ontological approach helps identify hidden relationships between variables, systematize data, and automate tasks such as classification and report generation. Ontologies make it possible to solve problems associated with processing complex multidimensional data that are difficult to analyze using conventional methods.

Despite the active development of business analytics systems, their integration with ontologies for automating data analysis remains a pressing issue. The introduction of ontologies into such systems can significantly increase the efficiency of data analysis, revealing hidden patterns, and ensuring the systematization of large volumes of information. With the development of digital technologies and the increase in the amount of information, conventional methods for data analysis often prove ineffective. They require significant time and human resources, have a high error rate, and are limited in their ability to identify hidden relationships between data.

The use of ontological models makes it possible to solve these problems by providing a tool for structuring and systematizing data, automating the process of their processing, and increasing the accuracy and scalability of analysis. Ontologies provide deep formalization of knowledge about the subject area and facilitate the identification of relationships between data, which makes the analysis more accurate and reproducible. This is especially important when making management decisions, which require working with large volumes of data and the need for prompt analysis to improve the efficiency of processes. In addition, automation of data analysis make it possible to significantly reduce time costs, minimize human errors, and increase the accuracy of forecasts, which makes the implementation of ontological models relevant for a wide range of practical tasks in various fields of activity.

2. Literature review and problem statement

The rapid increase in interest in the development and application of new approaches to integrating disparate data for analysis and decision-making has led to the need to devise new approaches to data structuring. Thus, in [4], the authors propose a paradigm shift towards computational social science in the context of big data. However, the results of the study reveal a certain conceptual approach, but the questions of using specific tools or models for structuring and automating the processing of questionnaire data remain open.

Work [5] emphasizes that data collected from questionnaires and other sources become a valuable asset for companies that can monetize this data by providing it for analysis and forecasting market trends. Such data could be used for financial gain and making strategic decisions based on market analysis and consumer preferences. However, the developed methods focus on the financial domain, creating a gap in application to other areas, including education and competency assessment. In [6], the problems associated with the implementation of data analytics in business processes, competencies in the analytical field are considered. The experience in the context of questionnaire surveys is interesting, data analysis also becomes an important element for assessing the effectiveness of training programs, which is directly related to the level of assimilation of competencies. The authors note the lack of a clear understanding of business values from data analysis and emphasize the positive impact of competencies in data analytics on business value. In the paper, however, they do not consider the practical implementation of the proposed conceptual approaches and the use of models and technologies to solve the issue of automating the processing of questionnaire data.

Study [7] examines the impact of changing the number of user segments on the process of data personalization in information systems. The study showed that the correct setting of hyperparameters makes it possible to improve the quality of analysis and more accurately predict user behavior. In the context of analyzing questionnaire data, these methods could be used for a more detailed assessment of the assimilation of competencies based on various characteristics of respondents. However, the study is limited by the use of only one approach to personification (algorithmically generated personas). This requires further investigation of other forms of personification and modeling of users, scenarios or user profiles, and their impact on hyperparameter settings. Although the results could be useful for assessing competencies, the study is limited to one approach to personalization and does not take into account the capabilities of ontological models. It is also important to note the works concerning ontologies and their role in data analysis. In [8], a model for joint analysis of personal data using ontologies and software systems is proposed, which makes it possible to better manage data and their privacy. This experience is useful because, despite the small size of the group of users recruited for the trial, a good mix of gender and age groups was achieved. Moreover, on average, users did not have a particularly high level of relevant technical knowledge. However, the small sample and limited technical knowledge of the participants limit the generalizability of the results. Study [9] considers an abstract model of the personal data lifecycle (APDL), which aims to support data management and tracking of personal information, as well as ensuring compliance with legal regulations. However, the model does not provide specific tools for automating the processing of questionnaire data using ontologies.

To examine the gaps in the analysis of personal informatics, which is especially relevant in the context of analyzing questionnaire data, where flexibility in approaches to information processing is often required [10]. The authors identified the need to devise more flexible tools for the analysis of personal data, which would make it possible to take into account the diverse and dynamic needs of users. In the context of questionnaire data, this can be useful for designing more adaptable analysis systems that can respond to changes in responses and adjust conclusions based on them. However, specific solutions for the implementation of ontological models are not proposed.

In [11], the importance of a holistic approach to the design of business processes and data is discussed, emphasizing the complexity of their joint processing. In the field of questionnaire data, this approach could help optimize the processes of collecting and analyzing information, when it is necessary to simultaneously take into account various data flows coming from different sources. However, the issues of structuring questionnaire data using ontologies remain outside the scope of the study.

In [12], the Vadalog system for knowledge graph management, applicable in industry, is described. However, the application of this system to the analysis of questionnaire data is not considered. Study [13] discusses the importance of trust in managing big data chains and the need for effective governance mechanisms. However, the study does not consider ontological approaches to data structuring. The literature [14–17] form the basis for building mathematical models that facilitate understanding the processes of assimilation of competencies and optimization of the information infrastructure of educational institutions. However, the use of ontologies for the automation of processing of questionnaire data is not studied in the cited papers.

Thus, our review reveals that most existing studies do not provide comprehensive solutions for automating the processing of questionnaire data using ontologies. The main shortcomings include:

– the lack of specific methods and tools for structuring and analyzing questionnaire data using ontologies;

– insufficient flexibility of existing data analysis systems to adapt to changes in respondents' answers;

– limited research into the use of ontological models to automate and improve the accuracy of data analysis.

The task to devise an ontological model that automates the process of processing questionnaire data, structures key parameters and relationships, and improves the quality of analysis for various tasks, including assessing competencies at different stages, remains unresolved. The solution to this problem is relevant for improving decision-making processes based on questionnaire data and overcoming existing limitations in the field of data analysis.

3. The aim and objectives of the study

The objective of our study is to construct a mathematical model of ontology for analyzing questionnaire data using business analytics systems aimed at improving the efficiency and accuracy of data analysis for assessing respondents' competencies at various stages of training.

To achieve the goal, the following tasks were set:

– to systematize existing approaches to analyzing questionnaire data and constructing ontologies to form a theoretical basis for the study;

– to identify key relationships between variables that allow for a qualitative analysis of respondents' questionnaires and the formation of a mathematical model of data ontology;

– to devise methods for analyzing the dynamics of assimilation of respondents' competencies based on questionnaire data, covering the initial, intermediate, and final stages of training;

– to determine algorithms for forming ontologies that will automate data processing processes and adapt to various types of questionnaire surveys;

– to conduct an experimental test of the proposed model on real data and evaluate its efficiency, accuracy, and relevance for processing questionnaire data.

4. The study materials and methods

The object of our study is questionnaire data obtained during a survey of preschoolers.

The subject of the study is the process of analyzing questionnaire data aimed at assessing the competencies of respondents at various stages of training using business analytics systems and ontological modeling.

As an example of the implementation of the proposed model, questionnaire data are used, which make it possible to demonstrate the effectiveness of the ontological approach for identifying hidden relationships and improving the accuracy of the analysis. The purpose of the surveys is to assess the quality of assimilation of competencies at various stages of training: initial, intermediate, and final. The main hypothesis of the study assumes that the use of an ontological approach could enhance the accuracy of data analysis, improve data systematization, and automate the decision-making process based on questionnaire data. The questionnaire covers three key stages: initial, intermediate, and final, which make it possible to get a dynamic picture of changes in the level of knowledge of respondents in various competencies. Questionnaire questions are divided into three assessment levels: high, medium, and low, depending on the complexity and expected level of competence.

An ontological knowledge base was designed and used to store and analyze data, which structures information about competencies, survey results, and the dynamics of their change. Ontologies, unlike relational databases, are used to create and manage knowledge about a subject area. In this paper, the ontology described the relationships between key concepts such as respondent, competency, training stages, and competency assessment. Business analytics systems integrated with the ontological model were used as tools for data analysis, which made it possible to effectively analyze large volumes of data and track respondents' progress:

1. Mathematical model of competency analysis.

To assess the dynamics of competencies acquisition, a mathematical model was built, which is based on the survey results, structured by complexity levels (high, medium, low) and stages (initial, intermediate, final). The model is a system of equations, where the dynamics of change in the level of competence is described by the function $C(t)$, where t is the time of passing the stages of the survey, and C is the level of competence. For each respondent, changes in competence indicators are calculated at each stage:

$$\Delta C = C_{final} - C_{initial}. \quad (1)$$

$$C_{intermediate} = \frac{C_{beginner} + C_{final}}{2}. \qquad (2)$$

This makes it possible to evaluate both the average increase in the level of competence acquisition and deviations at intermediate stages.

2. Ontology formation algorithms.

The algorithm for preparing and analyzing questionnaire data included the following stages:

1. Defining the goals and objectives of the questionnaire – formulating the goals of assessing the level of competence for each stage.

2. Formulating questionnaire questions – the questions were devised taking into account the possibility of comparing them at different stages (initial, intermediate, final). The questions were classified in the ontology, which simplified their further processing.

3. Building a data ontology – based on the key concepts of the questionnaire and their interrelations, an ontological model was built, which was used to group data into classes.

4. Determining the number of control sections – analyzing data on control sections made it possible to identify the dynamics of competence acquisition among respondents at different stages.

5. Data analysis – statistical methods and machine learning (clustering, regression) were used to assess changes in the level of respondents' competencies at each stage.

The use of classification algorithms allows for automatic updating of the ontology as new data is received, which makes it possible to promptly analyze changes in the level of assimilation of competencies. Clustering algorithms were used as classification methods, in particular the k-means method, which made it possible to identify incorrect questionnaires to clean up the quality of the data obtained, and to divide the questionnaire results into three main groups according to the level of knowledge acquisition.

3. Data analysis using business analytics systems.

Business analytics systems integrated with the ontological knowledge base were used for data analysis. Power BI was used as the main tool, which made it possible to effectively visualize the dynamics of assimilation of competencies by respondents at various stages of training. In addition to visualization, correlation analysis methods were used to identify relationships between competencies at different stages.

The experimental part of the study was aimed at testing the developed mathematical model and algorithms in practice. The real results of the survey conducted among respondents of educational organizations were used as initial data. The resulting data were structured in an ontological knowledge base and analyzed using business analytics tools. It was expected that the use of the ontological approach would significantly improve the accuracy of data analysis and automate the processing of questionnaire data.

5. Results of constructing an ontology for analyzing questionnaire data

5. 1. Systematization of existing approaches to analyzing questionnaire data

When analyzing questionnaire surveys that are extended in time and aimed at assessing the formation of competencies, the effectiveness of methods largely depends on their ability to take into account the time dynamics, scalability, and accuracy of data analysis. The following methods were considered, which make it possible to work effectively under the listed conditions, taking into account their limitations:

1. Multivariate statistical analysis, which allows for the simultaneous consideration of several variables (questionnaire questions) and the assessment of their interrelationships, which is especially important in the context of questionnaires aimed at assessing competencies and their interrelations with each other (for example, skills, knowledge, abilities). This method works well with data collected at several stages (initial, intermediate, final) and make it possible to assess the dynamics of changes in competencies. In addition, it can be used to classify respondents by competency levels, identify hidden patterns, and analyze progress.

The limitation of this method is the need for a large volume of data to achieve statistical significance. In addition, the method does not make it possible to work with data that have complex or nonlinear dependences, which is typical for competency formation processes.

2. Factor analysis is used to reduce the number of variables and identify hidden factors that may influence the survey results. This method makes it possible to group survey questions by common factors, which helps more clearly define the key components of competencies. It is effective for analyzing complex data collected at various stages, making it possible to identify hidden relationships and reduce the number of variables, simplifying the interpretation of data. Factor analysis is useful for identifying key elements that influence the formation of competencies.

The limitation is the need to interpret factors, which can be subjective and vary depending on the researcher. In addition, the method is not always easily adapted to highly variable data at each stage of the survey.

3. Regression analysis is used to predict the values of the dependent variable (for example, the level of competence) based on independent variables (for example, survey results). It helps assess the influence of various factors on the formation of competencies. Regression analysis makes it possible to predict changes in competencies based on data collected at early stages and evaluate the effectiveness of the survey for each respondent. The method works well in long-term questionnaire surveys, as it can take into account time changes and predict results at subsequent stages.

The limitations of the method include a strong dependence on the quality of the data. With strong deviations or non-linear dependences between variables, the method can give distorted results since correct application requires the presence of a sufficient amount of data.

4. Clustering (e.g., k-means) makes it possible to group respondents by similar characteristics, which makes it possible to assess which groups show similar results in the development of competencies. The method works well with data collected at different time stages, as it makes it possible to track progress in each group of respondents. The method makes it possible to identify groups of respondents who demonstrate similar dynamics in the development of competencies, which is useful for targeted assessment of their effectiveness.

The limitations of this method include inefficiency with highly heterogeneous data or when respondents have significant individual differences that do not fit into simple clusters. The method also requires a predetermined number of clusters, which can limit its flexibility.

5. Ontological modeling makes it possible to formalize the structure of data and their relationships, which helps sys-

tematize the process of analyzing questionnaire data and automate it. This method works effectively with large volumes of data collected at different time stages and makes it possible to automate the analysis process. Ontological modeling makes it possible to identify hidden dependences between variables, improves the reproducibility and flexibility of the analysis. The use of this approach makes it possible to track how different elements of the questionnaire (for example, questions or respondents) are related to the development of competencies. Ontology helps clearly formalize the structure of data and the relationships between them. This makes it possible to precisely define how the data will be classified, structured, and analyzed. Building ontologies makes it possible to automate many routine analysis tasks, such as classifying responses, identifying connections between respondents and their responses. This is especially important when working with large volumes of data. The ontological model can be expanded and adapted to various types of questionnaires, which allows for more flexible and scalable data analysis. By structuring data based on ontology, one can identify hidden dependences between variables, which helps in making more informed decisions.

The limitation of this method is that it requires significant resources and time, as well as specialized knowledge in the field of semantic modeling. In addition, the ontology needs to be constantly updated to account for all changes and features of the data.

When building ontologies for analyzing questionnaire data, it is important to consider the following aspects:

– identifying key concepts. For each type of questionnaire, it is necessary to identify the main concepts (e.g., respondent, question, answer, competence level). These concepts form the basis of the ontology;

– defining relationships. After identifying concepts, it is necessary to determine their relationships. For example, a respondent fills out a questionnaire, the questionnaire includes questions, the respondent's answers assess their competencies;

– formalizing rules and restrictions. The ontology should include rules and restrictions describing possible values of variables, the relationships between them and permissible actions (e.g., the respondent's age must be within the described age group, the competence level can be "low", "medium" or "high");

– selection of tools for building an ontology: various specialized languages and tools can be used to create ontologies, such as OWL (Web Ontology Language) and Protégé. These tools allow the ontology to be formalized and visualized, ensuring its use in business analytics systems.

Ontologies can be integrated with BI systems to create powerful analytical tools. Integrating an ontology with a BI system makes it possible to automatically generate reports and visualize data based on structured domain knowledge.

An ontological model makes it possible to analyze data from different perspectives, compare results across multiple dimensions (e.g. age, gender, competency level), and identify complex dependences.

6. Limitations of existing methods:

– sensitivity to data quality. Many methods, such as regression analysis and clustering, require high-quality source data. Gaps, outliers, or errors in the data can greatly affect the results and interpretation. This is especially important for long-term studies where changes in data format or inconsistencies between survey stages are possible;

– limited adaptability to changes over time. Some methods, such as conventional correlation analysis, are not adapted to take into account time dynamics, which reduces their effectiveness in analyzing data collected at several stages. These methods may miss important changes that occur between survey stages;

– difficulty in interpreting results. Methods such as factor analysis and clustering may produce results that are difficult to interpret, especially if there are complex nonlinear relationships between variables or significant heterogeneity in the data;

– dependence on data volume. Methods such as machine learning and time series analysis require large amounts of data to provide accurate and reliable results. Such methods may be less effective when the number of respondents is limited, or surveys are sparse at different stages.

For the analysis of questionnaire data collected at several time stages, the most effective methods are multivariate analysis, regression models, and ontological modeling, as they allow for the dynamics of competence development to be taken into account and provide flexibility of analysis. However, it is important to consider the limitations of each method, such as the requirement for high-quality data and the need for a sufficient sample size. Based on the above, we have chosen the ontological method of data analysis as the most optimal and easily scalable.

5. 2. Determining key relationships between variables and forming a data ontology

To form key relationships between variables, an ontological model of information systems has been developed that describes the relationships of information objects that automate the organization's activities. The formal ontology model is represented by a tuple:

$$O=\langle C,A,G,MA,MC,R,I\rangle, \tag{3}$$

where C is a set of concepts;

A is a set of attributes of concepts;

G is a set of constraints imposed on attributes;

$M_A{:}A\to G$ is a mapping that specifies constraints on attributes;

R is a set of relations;

I is an interpretation function that maps each ontology concept to a set of elements of the object scheme of the information system.

An information system using an ontology O is represented as:

$$U^O=\langle O,U,M_U,M_R\rangle,$$

where U is a set of elements of the object scheme of the information system (IS);

$M_U{:}U\to C$ is a mapping that maps an element of the object scheme to its concept;

$M_R{:}U\times U\to R$ is a mapping that defines the correspondence between the links between the elements of the object scheme to their relations in the ontology.

For any element $u\in U$, the following condition is satisfied: $\{a{:}\langle a,d\rangle\in u\}=M_C(MU(u))$ that is, the set of attributes of an element of the object scheme corresponds to the attributes of its concept.

Sets of concepts, attributes, constraints (Table 1) and relations for the system for assessing the quality of assimilation of competencies based on the proposed model with respondents and competencies, including 3 levels of assessment (low, medium, high).

Table 1

Conceptualization of ontology

Concept (C)	Attributes (A)	Restrictions (G)	Relations (R)
Respondent	Respondent ID, name, age, gender, group	Unique respondent ID, age >=4 and <7	A questionnaire is underway (connection with the concept of «Questionnaire»); evaluated according to competence (connection with the concept of «Competence»)
Competence	Competency name, competency description	The name of the competency is unique	Evaluated by the respondent at different stages (connection with the concept of «Assessment»); included in the questionnaire (connection with the concept «Questionnaire»)
Survey	Questionnaire ID, questionnaire name, date of implementation	Date>current date	Includes competencies (link to the concept of «Competence»); to be filled in by the respondent (link to the concept «Respondent»)
Evaluation	Assessment level (1 – low, 2 – medium, 3 – high), assessment stage (initial, intermediate, final), Assessment date	The value of the assessment ∈{1, 2, 3}, evaluation stage ∈{initial, intermediate, final}	Evaluates the respondent (connection with the concept «Respondent»); related to a competence (connection with the concept of «Competence»)

The interpretation function I makes it possible to link concepts to specific objects in the database, which simplifies the data analysis process:

I(Respondent)→{Table «Respondents»,Field «Respondent ID»,Field «Name»,Field «Age»,Field «Gender»,Field «Group»}.

Example of a database record:
Respondent ID: 1001
Name: Ivan Ivanov
Age: 21
Gender: Male
Group: Group 101.

I(Assessment)→{Table "Assessments",Field "Assessment Level",Field "Assessment Stage",Field "Assessment Date"}.

Database entry:
Assessment Level: 2
Assessment Stage: Initial Stage
Assessment Date: 2024-09-01.

I(Questionnaire)→{Table «Questionnaire»,Field «Questionnaire ID»,Field «Questionnaire Name»,Field «Date of implementation»}.

Database entry:
Questionnaire ID: 301
Questionnaire Name: Assessment of first-year competencies
Date of implementation: 2024-09-01.

Attitude: is being surveyed:

– interpretation: THIS attitude describes the relationship between the respondent and the survey. The respondent is participating in a survey that evaluates his/her competencies;

– presentation of the interpretation:

I(Taking_survey)→{Table "Respondents", Table "Survey", Field "Respondent ID", Field "Survey ID"}

I(Taking_survey)→{Table "Respondents",Table "Survey",Field "Respondent ID",Field "Survey ID"}.

Database entry:
Respondent ID: 1001
Survey ID: 301.

Relationship: assessed by competence:

– interpretation: this relationship describes the relationship between the respondent and his/her assessment of competence at a certain stage;

– presentation of the interpretation:

I(Assessed by competency)→{Table "Assessments",Table "Competencies",Field "Respondent ID",Field "Competency ID",Field "Assessment Level",Field "Assessment Stage"}.

Database entry:
Respondent ID: 1001
Name of competency: Data Analysis
Assessment Level: 2 (Intermediate)
Assessment Stage: Initial.

The interpretation function makes it possible to link abstract concepts such as respondent, competence, assessment, and questionnaire with specific objects in the information system database. This makes it possible to automate the process of data analysis, storage, and subsequent interpretation to assess the acquisition of competencies at various stages. The ontological model provides a high degree of automation and structured data storage, which makes it possible not only to significantly speed up data processing but also improve the accuracy of competency analysis. The model is easily scalable and can be adapted to various types of questionnaire data and requirements for competency analysis, which expands its applicability and flexibility.

5. 3. Methods for analyzing the dynamics of assimilation of respondents' competencies based on questionnaire data

To implement the task of analyzing the dynamics of assimilation of respondents' competencies based on questionnaire data, a mathematical model is proposed that takes into account three key stages of training: initial, intermediate, and final. This model makes it possible to track changes in the level of respondents' competencies and predict their progress. The proposed ontology should take into account three-level control (initial, intermediate, and final) and the dynamics of assimilation of respondents' competencies based on questionnaire data. For each questionnaire i, a data set X_i is collected at each stage of analysis:

$X_i^{(0)}$ – data at the initial stage (input slice);

$X_i^{(1)}$ – data at the intermediate stage;

$X_i^{(2)}$ – data at the final stage.

To process data $X_i^{(k)}$, where $k \in \{0,1,2\}$, they are classified according to the selected criteria.

To assess changes between slices, the difference $X_i^{(0)}$, $X_i^{(1)}$, $X_i^{(2)}$. between them is used. Difference in the values of responses i at the initial and intermediate stages:

$$\Delta X_i^{(0,1)} = X_i^{(1)} - X_i^{(0)}. \quad (4)$$

And between the intermediate and final stages:

$$\Delta X_i^{(1,2)} = X_i^{(2)} - X_i^{(1)}. \quad (5)$$

To analyze changes, one can use mathematical processing $\Delta X_i^{(k,k+1)}$, for example, through the construction of regression models or other methods for time series analysis.

Linear regression model for assessing the impact of initial competency indicators on the results of the final stage:

$$X_i^{(2)} = \beta_0 + \beta_1 X_i^{(0)} + \beta_2 X_i^{(1)} + \varepsilon_i \quad (6)$$

where β_0, β_1, β_2 are regression coefficients (model parameters showing how each stage affects the final result), ε_i is a random error.

Time series analysis for predicting respondent progress using indicators from previous stages. This method makes it possible to predict the value of competencies at the final stage based on data from the initial and intermediate stages.

The calculated values ΔX_i^k, for each respondent can be grouped by levels (low, medium, high), which makes it possible to classify respondents based on their progress. This classification helps identify the following.

High progress groups – respondents who demonstrate a stable increase in indicators.

Low progress groups – respondents whose level of competencies remains unchanged or decreases.

In practice, this model can be used to assess the success of educational programs based on survey data. For example, data at each stage make it possible to identify key factors that contribute to the development of competencies and adjust programs to improve respondents' performance.

5. 4. Algorithm for forming ontologies that make it possible to automate data processing processes

The algorithm for preparing and further analyzing data is shown in Fig. 1.

Stage 1 – defining the goals and objectives of the survey. The goals of the survey for each of the sections are specified. This may include assessing the level of knowledge or skills at each stage. Clearly formulating the goals makes it possible to determine which aspects will be reflected in the ontology. For example, if the goal is to assess respondents' satisfaction with the educational process, the ontology will contain the entities: respondent, group, teacher, satisfaction, and their relationships.

Stage 2 – formulating questions. The questions in the survey should be developed so that they can be compared between stages (initial, intermediate, and final). The questions should be formulated in such a way as to reflect the aspects being studied as fully as possible. It is important to take into account the type of response (open, closed, Likert scale, etc.).

Stage 3 – defining key concepts and identifying classes. Each concept or response in the survey is classified to simplify further processing. At this stage, the main entities, and the relationships between them in the subject area are identified. For example, if we study respondent satisfaction, the key concepts may be respondent, group, teacher, task, assessment.

Stage 4 – identifying classes to form an ontology. Each key concept becomes a class in the ontology. Forming a primary set of concepts corresponding to questionnaire elements, on the basis of which the basic ontology is built.

Stage 5 – adjusting questionnaire questions and defining answer types. Sometimes, after identifying classes, it becomes clear that some questions need to be clarified or reformulated. Answer types determine how the information will be presented in the ontology. If, at the stage of defining control slices, it is revealed that the questionnaire requires more detailed questions or other approaches to data collection (for example, different answer types for different groups of respondents), it is necessary to make changes to the wording of the questionnaire questions and revise the answer types.

Stage 6 – building a data ontology. It is used to group data into classes, which will facilitate further analysis. At this stage, a formal model is built representing knowledge about the subject area. Special languages are used to describe ontologies, such as OWL. The algorithm establishes links between concepts, such as "Respondent is taking a survey", "Question is related to competence", "Answer evaluates competence". These links determine how the survey data should be organized and interpreted. Use of association rules and classification trees to automatically generate links based on the structure and logic of survey questions.

Stage 7 – Data analysis. Comparison of data at each stage, using statistical methods to identify significant changes. Control sections make it possible to analyze data from different points of view. For example, one can analyze results by respondent groups and by task types. If difficulties arise with interpreting or structuring data during data analysis, this may indicate insufficient development of key concepts and relationships. In this case, one needs to return to the stage of defining key concepts to review and clarify them. If data analysis reveals problems with their structuring or insufficient flexibility of the ontological model, one needs to return to the stage of building the data ontology to review the classes and relationships between them.

The final stage 8 is decision making based on the identified changes. The model can be used to predict success or identify critical factors affecting the survey result.

Adaptation of the ontology to new types of questionnaires.

The proposed ontological model is flexible and adapted to different types of questionnaires. For this purpose, if new elements appear in the questionnaire structure (for example, new categories of questions or competencies), the model is updated taking into account the changes (Fig. 1). Changes in the data structure (for example, identifying new competencies or types of questions), which make it possible to dynamically adapt the ontology to new types of questionnaires, do not affect the ontology and the designed DB. The simplicity and stability of DB has another advantage, making it possible to track changes made during the experiment. Since the experiment is extended in time, such indicators can adjust the parameters of the linear regression coefficients (6). For example, in the process of the experiment, new instructions came on the formation of competencies in respondents. The learning process is rebuilt, and surveys are introduced into the questionnaire, making it possible to understand the formation of new competencies.

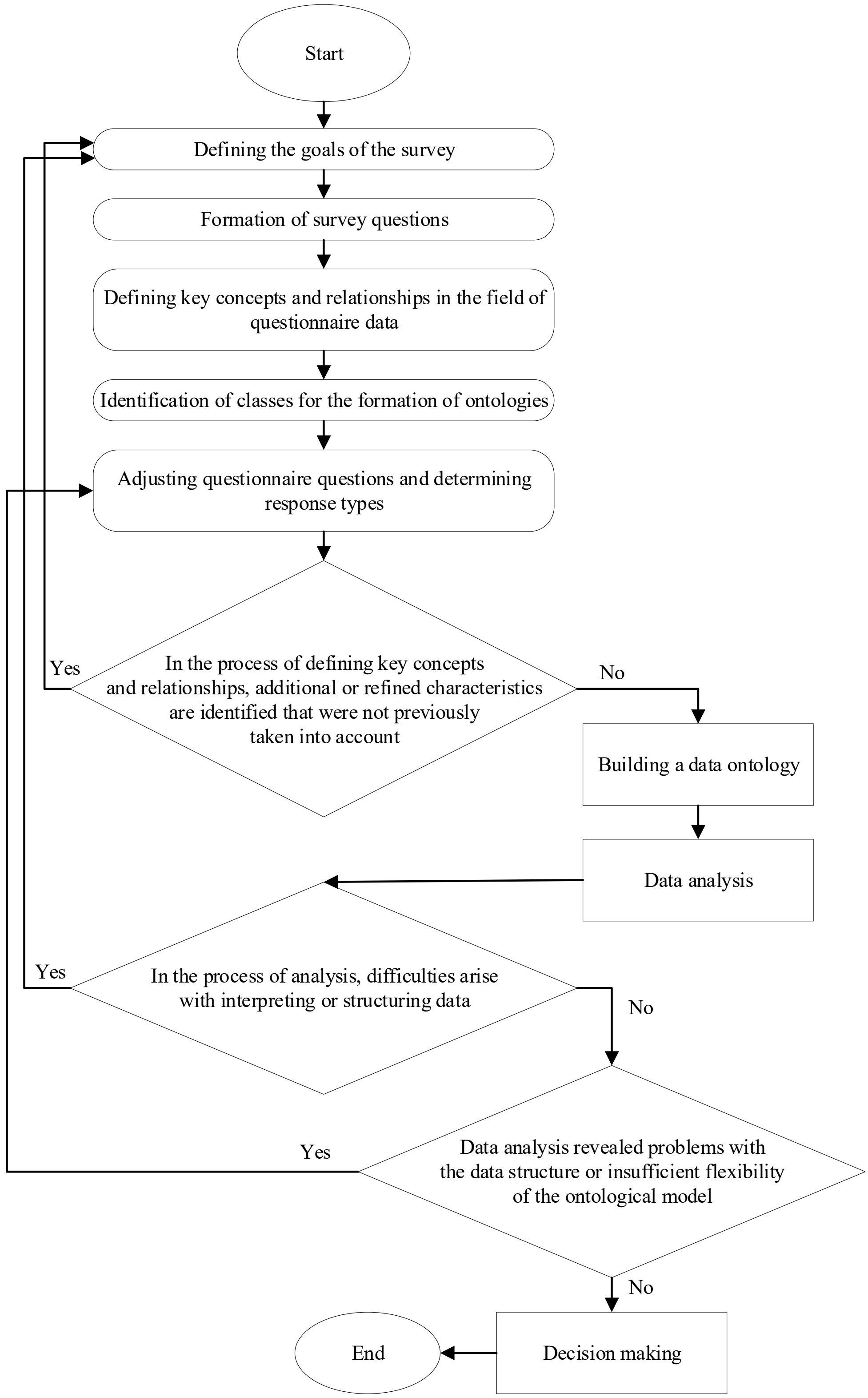


Fig. 1. Algorithm for forming a questionnaire and analyzing data

Next, there are two options for analyzing the level of competence development.

The first is to take the previous level of development of this competence as zero and not make any changes to DB. Then the dynamics of assimilation will be positive. To adjust the general indicators, we change the coefficients β_i (formula (6)).

The second is to evaluate the level of participation of the "new" competence relative to those that were formed at the previous stage and derive a formula for its participation based on the competence formed. One can also use a linear regression model, but the number of variables in this case will depend on the number of competences formed (each with its own attention coefficient) at previous stages of training. After that, using R-procedures, we form additional records relative to the level of formation and save this value in DB. For example, if we have a new competence "ability to write Arabic numerals", which was not tested in the previous stages – the competence to write numbers. However, the competences of drawing lines and other graphic objects were tested, the concept of "Number" was introduced, and the competence of addition was formed. Then, by placing weighting coefficients relative to the formed skills, we can obtain the initial value of the level of proficiency in the competence.

Description of the algorithm for constructing and processing the results of questionnaires and investigations:

The set of all questionnaires as feature vectors:

$$X=\{X_1,X_2,...,X_n\}. \tag{7}$$

For each slice:

$$X^{(k)}=\left\{X_1^{(k)},X_2^{(k)},\ldots,X_n^{(k)}\right\},\quad k\in\{0,1,2\}. \tag{8}$$

For each individual i, the comparison can be made according to the formulas of change:

$$\Delta X_i^{(0,1)}=X_i^{(1)}-X_i^{(0)},\quad \Delta X_i^{(1,2)}=X_i^{(2)}-X_i^{(1)}. \tag{9}$$

For global analysis, we shall use:

$$\bar{X}^{(k)}=\frac{1}{n}\sum_{i=1}^{n}X_i^{(k)}, \tag{10}$$

and calculate the difference between the average values for the groups:

$$\Delta\bar{X}^{(k,k+1)}=\bar{X}^{(k+1)}-\bar{X}^{(k)}. \tag{11}$$

Let's take a closer look at stage 6 of ontology construction:

1. Modeling competency levels.

For each individual i and for each competency j, at each stage $k\in\{0,1,2\}$, the competency level $C_{ij}^{(k)}$, is estimated, which can take one of the following values:

– "Low" (L) will take the value 1 in the data table;

– "Medium" (M) will take the value 2 in the data table;

– "High" (H) will take the value 3 in the data table.

Thus, the value $C_{i,j}^{(k)}$ takes values from the set {1, 2, 3}.

2. Determining changes in competency levels.

Changes in the competency level between stages are estimated by the difference in levels for each individual:

$$\Delta C_{i,j}^{(0,1)}=C_{i,j}^{(1)}-C_{i,j}^{(0)},\quad \Delta C_{i,j}^{(1,2)}=C_{i,j}^{(2)}-C_{i,j}^{(1)}. \tag{12}$$

If $\Delta C_{i,j}^{(k,k+1)}>0,$ this means an improvement in the level of competence at stage k+1.

If $\Delta C_{i,j}^{(k,k+1)}<0,$ this means a decrease in the level of competence.

If $\Delta C_{i,j}^{(k,k+1)}=0,$ the level of competence has not changed.

3. Analysis of the dynamics in competence acquisition.

For each stage, the average value of the level of competence can be calculated for all individuals by competence:

$$\bar{C}_j^{(k)}=\frac{1}{n}\sum_{i=1}^{n}C_{i,j}^{(k)}, \tag{13}$$

where n is the number of survey participants. This average value can be used to assess the overall level of competence acquisition at each stage. The change in the average competence level between stages:

$$\bar{C}_j^{(k,k+1)}=\bar{C}_j^{(k+1)}-\bar{C}_j^{(k)}. \tag{14}$$

This indicator is used to assess the overall progress or regression in learning for each competency.

4. Model of transitions between competency levels.

A function $P\left(X_i^{(k)}\right)$ is introduced – belonging of the answer i to a certain class (for example, people, professions, classes, etc., as shown in the first diagram).

To analyze the probability of transition between competency levels, a Markov chain model could be used. The probability that an individual with a level $C_{i,j}^{(k)}=x$ moves to a level $C_{i,j}^{(k+1)}=y$ at the next stage:

$$P_{i,j}=P\left(C_{i,j}^{(k+1)}\right)=y\setminus C_{i,j}^{(k)}=x. \tag{15}$$

A transition matrix can be built for each stage. For example, for the transition between the initial and intermediate stage (for one competence), the transition matrix will take the following form (Table 2).

Table 2

General view of the transition matrix

–	L (low)	M (medium)	H (high)
L	P_{LL}	P_{LM}	P_{LH}
M	PML	P_{MM}	P_{MH}
H	PHL	P_{HM}	P_{HH}

Note: P_{xy} is the probability of transition from level x to level y.

5. Evaluation of progress by competency groups.

For a general evaluation of competencies, an integral progress indicator can be calculated for each individual by summing up the changes in all competencies j:

$$I_i^{(k,k+1)}=\sum_{j=1}^{m}\Delta C_{i,j}^{(k,k+1)}, \tag{16}$$

m – number of competencies.

6. Global analysis by groups.

To analyze the overall dynamics, individuals can be grouped by levels at each stage and the proportion of respondents with different levels can be compared:

$$P(L)=\frac{\left|\left\{i:C_{i,j}^{(k)}=1\right\}\right|}{n}, \tag{17}$$

$$P(M)=\frac{\left|\left\{i:C_{i,j}^{(k)}=2\right\}\right|}{n}, \tag{18}$$

$$P(H)=\frac{\left|\left\{i:C_{i,j}^{(k)}=3\right\}\right|}{n}. \tag{19}$$

Thus, it is possible to track the dynamics in the distribution of levels in the group at each stage and use statistical methods to identify significant changes. Such a mathematical model makes it possible to track changes in competency levels, predicting transitions between levels and assessing the effectiveness of training at different stages. Thus, as a result of the study of respondents, measuring the level of formation of competencies, a significant saving of time resources was noted. Considering that the study was conducted using the described ontological method of data analysis. It should be noted that the time spent on data collection remained the same. For a more visual comparison of conventional methods and the ontological approach, the following table can be proposed, the data in which are obtained based on the results of analysis (Table 3).

Table 3

Comparison table for experimental data

Indicator	Traditional analysis methods	Ontology approach
Time costs (per 100 questionnaires)	20–30 hours	5–8 hours through automation
Accuracy of data analysis	75 %	92 %
Error in forecasts	High (10–15 %)	Low (3–5%)
Scalability	Limited by data volume	High
Complexity of relationship analysis	It is difficult to identify hidden relationships	Identifying complex relationships

It should also be noted that under the conditions of a dynamically changing environment, taking into account the extended time of the study, the ontological approach makes it possible to analyze research data not only in dynamics but also under the condition of changes in the learning environment.

5. 5. Implementation of the data processing model using ontologies

To track the dynamics in assimilation of competencies by specific respondents based on the presented data ontology, a database can be used. The database contains information on each of the three assessment stages (initial, intermediate, and final) and the level of assimilation of competencies ("high", "medium", "low"). Below is a step-by-step algorithm for tracking the dynamics of assimilation of competencies using a relational database:

1. Structure of the database:

Table "People"
person_id: unique identifier of the individual.
name: name or other data of the individual.

Table «Competencies»
competency_id: unique identifier of the competency.
competency_name: name of the competency).

Table «Competency Assessments» (Competency_Assessments)
person_id: foreign key to the «Persons» table.
competency_id: foreign key to the «Competencies» table.
stage: assessment stage (values: «initial», «intermediate», «final»).
competency_level: level of competence acquisition (values: «1» – low, «2» – average, «3» – high).
assessment_date: assessment date (optional if assessments were conducted on specific dates).

«other information» assumes that in addition to the basic information that will be used within the framework of this system, the tables may contain a number of additional information that can be used in other studies or predictive models.

2. Query for tracking the dynamics in the respondent's competencies.

In order to track the dynamics in the acquisition of a competence by a particular respondent, it is necessary to obtain data on his/her level at each stage for each competence.

SQL query for one respondent for one competency

```
SELECT c.competency_name, ca.stage, ca.competency_level
FROM Competency_Assessments ca
JOIN Competencies c ON ca.competency_id = c.competency_id
WHERE ca.person_id = :person_id
ORDER BY c.competency_id, ca.stage;
```

In this query: person_id is the identifier of the respondent to be tracked. It will return a list of competencies and their levels at each stage for this respondent.

SQL query for dynamics for all respondent competencies.

To analyze the dynamics, one can use a query that will show changes in the level of competence from stage to stage:

```
SELECT c.competency_name,
SUM(CASE WHEN ca.stage = 'Промежуточный' THEN ca.competency_level - ca_init.competency_level ELSE 0 END) AS delta_initial_intermediate,
SUM(CASE WHEN ca.stage = 'Итоговый' THEN ca.competency_level - ca_int.competency_level ELSE 0 END) AS delta_intermediate_final
FROM Competency_Assessments ca
JOIN Competencies c ON ca.competency_id = c.competency_id
JOIN (SELECT competency_id, person_id, competency_level
FROM Competency_Assessments WHERE stage = 'Начальный') ca_init
ON ca.competency_id = ca_init.competency_id AND ca.person_id = ca_init.person_id
JOIN (SELECT competency_id, person_id, competency_level
FROM Competency_Assessments WHERE stage = 'Промежуточный') ca_int
ON ca.competency_id = ca_int.competency_id AND ca.person_id = ca_int.person_id
WHERE ca.person_id = :person_id
GROUP BY c.competency_name;
```

This query calculates the difference in competence levels between the initial and intermediate stages, as well as between the intermediate and final stages for all respondent competencies.

This will help us understand how much the respondent's competence level has improved or worsened at each stage.

3. Mathematical description of dynamics.

The following metrics can be used to analyze the dynamics of competence acquisition:

– average increase in competence.

The average increase in the respondent's competence for all competencies can be expressed as follows:

$$\Delta C_i = \frac{1}{m}\sum_{j=1}^{m}\left(C_{i,j}^{(2)} - C_{i,j}^{(0)}\right), \quad (20)$$

where the total number of competencies $C_{i,j}^{(2)}$ is the level of competence j at the final stage; $C_{i,j}^{(0)}$ is the level of competence j at the initial stage.

This indicator provides an overall assessment of how much the level of assimilation of all competencies has changed for a given respondent from the initial stage to the final stage:

– assessment of individual progress.

To analyze progress for each competence j for respondent i, a relative increase can be introduced:

$$R_{i,j} = \frac{C_{i,j}^{(2)} - C_{i,j}^{(0)}}{C_{i,j}^{(0)}}, \quad (21)$$

where $R\left(X_i^{(k)}\right)$ – compliance of the answer with the criteria.

This value reflects how much the level of assimilation of a specific competency has increased (or decreased) relative to the initial level.

4. Conclusions and recommendations.

This model make it possible to track and analyze changes in competency levels at each stage.

Using SQL queries, one can extract data for a specific individual or for the entire group and conduct an analysis based on dynamics.

The mathematical description makes it possible to calculate the average changes in competency levels, which helps determine which competencies are being learned better or worse.

Such a model can be expanded using statistical analysis or machine learning methods to predict the dynamics of competency acquisition in the future.

To analyze the dynamics of competency acquisition based on the presented database and three assessment levels (initial, intermediate, final), one can use the following steps and methods:

Step 1: data preparation.

First, one needs to extract data from the database. This example uses a SQL query to retrieve information about the competencies assessed at different stages:

```
SELECT
p.name AS person_name,
c.competency_name,
a.stage,
a.competency_level,
a.assessment_date
FROM
Competency_Assessments a
JOIN
People p ON a.person_id = p.person_id
JOIN
Competencies c ON a.competency_id = c.competency_id
ORDER BY
p.name, c.competency_name, a.stage;
```

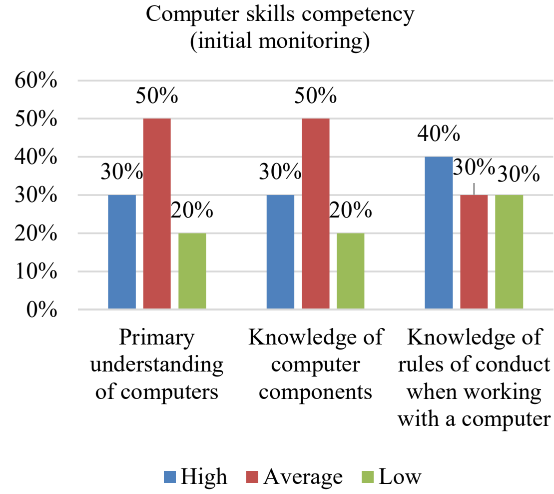
Step 2: data analysis.

The following methods can be used to analyze the data:

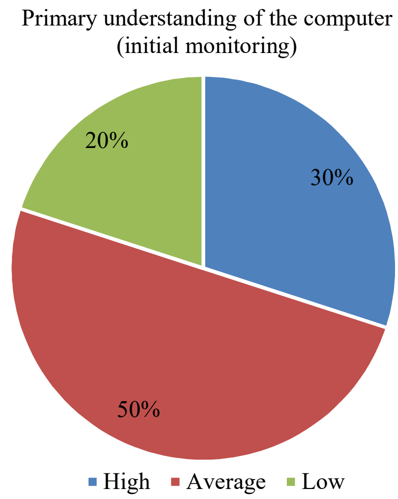
– competency level comparison: compare the level of competency acquisition at the initial, intermediate, and final stages for each respondent. This will help identify the dynamics of changes in levels;

– visualization of changes: build graphs that show changes in competency acquisition levels. For example, line graphs can display changes by stage for each competency for each respondent. Or, if a specific group is analyzed, the data for the group as a whole is visualized, calculating the average value (Fig. 2);

– statistical analysis: calculate the average competency levels for each stage. This will help determine how the overall group overcomes barriers to competency acquisition.



a



b

Fig. 2. An example of data representation for a group (initial monitoring): *a* – average data by criteria; *b* – general representation of the level of development

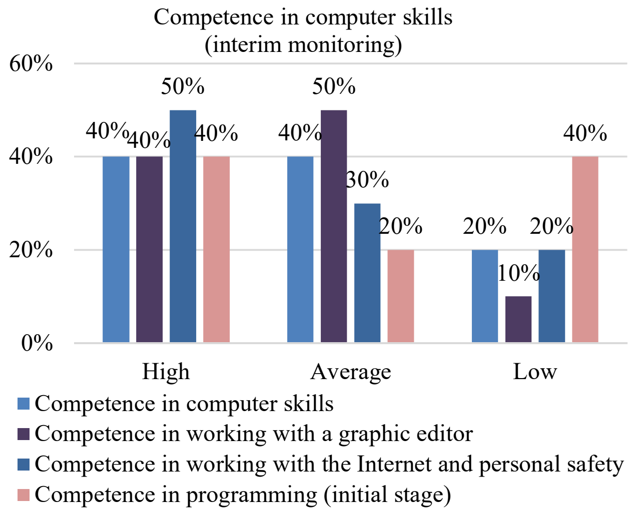
In the context of monitoring the development of IT competencies in preschool children, the use of business intelligence systems (BIS) allows for a deeper understanding of the dynamics of knowledge and skills acquisition at various stages of the educational process. Below are the key aspects of data analysis at the initial, intermediate, and final stages and their significance in the context of the effectiveness of educational programs.

1. Initial monitoring (Fig. 2).

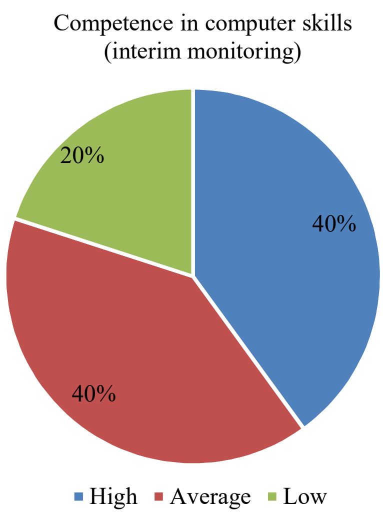
At the initial stage of monitoring, the goal is to establish the initial level of IT competencies in children. It is important to determine the basic knowledge and skills, which will make it possible to build individual educational plans and goals. At this stage, a limited number of criteria can be used, which is due to several factors:

– focus on basic skills: at the beginning of the process, it is enough to determine the basic skills and knowledge that will be developed in the future. An excess of criteria can complicate the initial analysis and complicate the development of educational plans;

– simplification of analysis: a limited number of criteria simplifies the interpretation of data, which is especially important when working with preschool children, where the initial level of knowledge may be uneven.



a



b

Fig. 3. Example of data representation for a group (interim monitoring): *a* — average data by criteria; *b* — general representation of the level of development

2. Intermediate monitoring (Fig. 3).

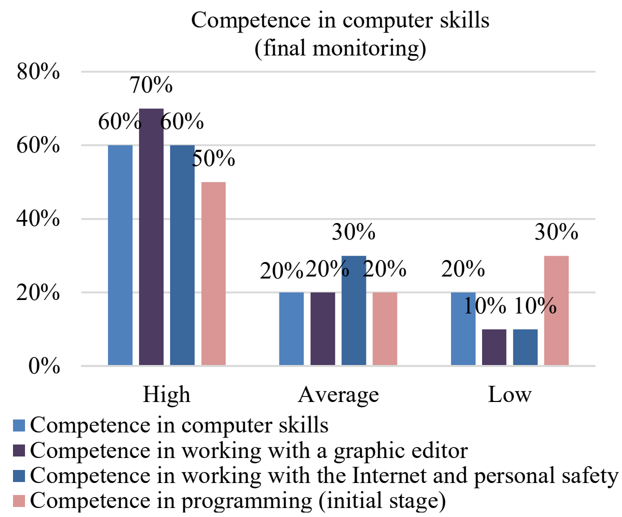
Intermediate monitoring is used to assess progress in acquiring IT competencies and allows for adjustments to the educational process. It is important to use the same criteria as at the initial stage for several reasons:

– data comparability. Maintaining the same criteria at the intermediate stage ensures data consistency, which makes it possible to accurately track the dynamics of changes in the acquisition of competencies;

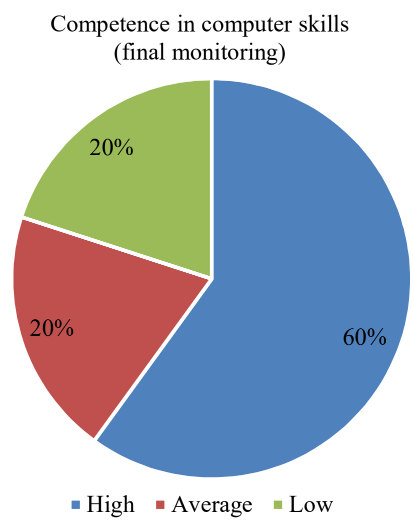
– effectiveness assessment. Using identical criteria helps in assessing the effectiveness of educational methods and identifying areas for improvement. Intermediate analysis makes it possible to understand which aspects of the program are working well and which require adjustments.

3. Final monitoring.

Final monitoring (Fig. 4) is aimed at assessing the final level of IT competencies after the completion of the educational process. It provides a complete picture of the results achieved and helps assess how successfully the program was implemented.



a



b

Fig. 4. Example of data representation by group (final monitoring): *a* — average data by criteria; *b* — general representation of the level of development

The results obtained during the testing of the ontological model for the analysis of questionnaire data showed its high efficiency in assessing the respondents' competencies at various stages of training. The use of ontology made it possible to systematize and structure the data, which significantly increased the accuracy of the analysis. Significant changes in the levels of competencies were identified, especially at the intermediate and final stages. This allows for a more in-depth assessment of the respondents' progress and conclusions about the need to adjust educational programs. The use of the ontological model ensures automation of the analysis process, which significantly reduces time costs and minimizes errors associated with the human factor. This approach allows for a more accurate analysis of data, revealing hidden relationships and adapting the methodology to changing conditions.

6. Discussion of the results of using ontology for data analysis

Conventional methods for analyzing questionnaire data demonstrate their limitations in terms of scalability, automation, and accuracy. Research shows that when assessing the effectiveness of training programs and analyzing competencies, conventional methods are not automated enough, which leads to an increase in time costs and a decrease in the accuracy of the analysis. The study showed that the use of an ontological model (3) for analyzing questionnaire data allows for the systematization and structuring of educational data, which significantly improves the process of processing and interpretation. The results of our study confirm that the proposed model ensures the automation of routine operations, such as data classification, information search and report generation, which allows for efficient work with large volumes of data (Fig. 2–4).

Compared with conventional methods for analyzing questionnaire data, which require significant time and human resources for processing and interpretation, the proposed ontological model has a number of significant advantages:

1. Automation of the analysis process. Conventional methods require manual data processing, which increases the likelihood of errors and reduces the speed of analysis. The ontological approach makes it possible to automate a significant part of the process, including classification and analysis of data.

2. Flexibility and scalability. Unlike existing analogs, the proposed model is easily adapted to increasing data volumes and can be applied in various fields, such as education, marketing, and sociology. This makes it possible to scale data analysis to new objects and organizations, which is difficult for conventional methods.

3. Accuracy of analysis. Unlike standard approaches, ontology makes it possible to take into account complex relationships between data, which significantly improves the quality of conclusions. This is especially important when analyzing multidimensional data, where conventional methods may be insufficient.

This analysis reveals that conventional methods for analyzing questionnaire data have the following limitations:

– high time costs and low automation, which reduces the overall efficiency of data processing;

– limited flexibility in personalization and segmentation of data, which reduces the accuracy of results;

– insufficient development of privacy and data management issues in conventional approaches;

– difficulties in scalability of methods with increasing data volumes.

Based on the above analysis, we can conclude that ontological models provide more accurate, flexible, and automated solutions for analyzing questionnaire data, which reduces time costs and increases the accuracy of analysis. The ontological approach copes better with data privacy issues, offering more reliable information protection mechanisms, which makes it preferable for use in areas requiring the processing of personal data. The introduction of ontological models for data analysis increases the flexibility and scalability of processes, which is especially important when analyzing large amounts of information and makes it possible to more accurately segment data for personalized analysis. Thus, the transition to an ontological approach could significantly improve the process of analyzing questionnaire data, increasing its efficiency, accuracy, and adaptability to changing conditions. This study solves the following problems of questionnaire data analysis:

1. Structuring and systematizing questionnaire data. The paper proposes a mathematical model of ontology (3), which formalizes the key parameters and relationships of questionnaire data (4). This made it possible to build a data structure that significantly improves the process of their analysis and makes it reproducible and scalable (6).

2. Automation of data processing. A clustering and classification algorithm (7) has been developed, which automates the update of the ontological database and allows for automatic distribution of questionnaire data into categories (8), (9). This speeds up the analysis process and increases the accuracy of data processing.

3. Analysis of the dynamics of assimilation of competencies. The paper solves the problem of tracking the progress of respondents by competency levels using ontologies (12). The model makes it possible to analyze changes in competencies at different stages of training (initial, intermediate, final), which gives a more accurate picture of the dynamics of knowledge acquisition.

4. Improving the accuracy of data analysis. By using the ontological approach, the paper proposes more accurate methods for data analysis, which solves the problem of limited effectiveness of conventional questionnaire processing methods. The ontological structure makes it possible to reveal hidden relationships between variables and classify respondents more accurately.

Thus, the problem of insufficient efficiency of conventional methods for analyzing questionnaire data, especially under conditions of working with large volumes of information, makes it necessary to conduct applied research on the development and application of ontological models for the automation of these processes. Further research in this direction is of great practical importance for improving the processes of making management decisions in various fields of activity. The expected effect of achieving the goal of the study consists of the following aspects:

1. Improving the accuracy of data analysis: the use of an ontological model will make it possible to structure the survey data, which will lead to a more accurate identification of hidden relationships between variables. This is especially important when working with large volumes of questionnaire data, where conventional analysis methods may miss significant correlations or lead to incorrect conclusions.

2. Automation of the analysis process (16). By implementing an ontological approach, the study involves the automation of routine operations for processing questionnaire data, such as classifying responses, searching for information, and generating reports. This will significantly reduce time and human costs, increasing the productivity of analysis (Table 3).

3. Systematization of knowledge about the subject area: formalization of knowledge using ontologies will allow for a deeper understanding of the nature of the relationships

between the parameters assessed in the questionnaires. This will lead to the construction of a structured knowledge base that could be used not only for current analysis but also for further research.

4. Scalability and reproducibility of analysis: the devised methodology will be easily scalable to process larger amounts of data, which will make it applicable in various contexts and will make it possible to analyze data from various sources. The reproducibility of results will also increase since formalized analysis models will ensure unambiguous interpretation of data.

5. Making informed management decisions: increased accuracy of data analysis and automation of the process will lead to the fact that organizations will be able to make more informed management decisions. This will allow them to respond more quickly to changes in data and improve the effectiveness of strategic planning, which is especially important in areas such as education, marketing, and sociology.

6. Evaluation of the dynamics of competencies (17) to (19): in educational organizations, the ontological model will help evaluate the development of respondents' competencies at various stages of training. This will make it possible to more accurately monitor the progress of students and adjust educational programs in a timely manner to improve their effectiveness.

Ontology generation algorithms, such as the k-means clustering method, successfully automate the update of the ontology database as new questionnaires are received. This makes it possible for the ontology to quickly adapt to changes in the data and automatically classify them, improving the analysis process. The results showed that the use of these algorithms significantly simplifies data processing and speeds up the analysis process, which is an important advantage compared to existing approaches that require manual intervention and significant time costs for data processing.

In practice, the proposed model could be used in various fields:

– educational organizations can use the ontology to assess students' competencies at different stages of education. For example, educational institutions can use the model to monitor and adjust training programs based on questionnaire data, allowing them to more accurately assess the dynamics of student competency development and make changes to educational programs;

– marketing research can use the ontology to analyze customer satisfaction, market segmentation, and evaluate the effectiveness of advertising campaigns. In this case, the ontology makes it possible to automate the process of analyzing customer surveys and identify hidden patterns that are difficult to detect with manual data processing;

– sociological research can use the ontological model to analyze public opinion, which will make it possible to structure the results of questionnaire surveys and automatically generate reports on various social groups.

In general, the results of our study make it possible to improve the quality and accuracy of assessing the competencies of students; to provide a personalized approach to learning; to optimize educational processes.

The designed ontology and mathematical model provide a more accurate, scalable, and automated solution for analyzing questionnaire data. These methods allow for better systematization and formalization of data, which increases the accuracy of analysis and simplifies the processing of large volumes of information. Implementation of the proposed solutions in educational and corporate systems can significantly increase the efficiency of questionnaire data processing and improve informed decision-making based on the results obtained.

Using the same criteria as in the previous stages has the following advantages:

– holistic assessment (20): using the same criteria allows for an objective assessment of the achieved level of knowledge and skills, which is critical for analyzing the overall success of the program;

– identifying trends: comparing the final data with the initial and intermediate levels makes it possible to identify long-term trends and determine how successfully the educational goals were achieved.

Conclusions from data analysis.

Analysis of the dynamics of assimilation of IT competencies in preschool children makes it possible to draw several key conclusions (21):

– individual differences: initial monitoring can reveal significant differences in the initial levels of knowledge of children. This makes it possible to adapt educational programs to individual needs;

– progress assessment: midterm analysis helps assess how effectively children are progressing in mastering competencies and make the necessary adjustments to the educational process.

The use of business analytics systems in this context makes it possible to systematize and structure data, which contributes to a more in-depth and objective assessment of educational programs. The consistency of criteria at all stages of monitoring ensures the accuracy of the analysis and helps in developing more effective educational strategies for preschool children.

The results of our study related to the construction of an ontology for analyzing data from corporate information systems (CIS) in educational activities demonstrate significant opportunities for systematizing and analyzing educational data, which is explained by the effective structuring of key objects, attributes, and their relationships (Fig. 1). The proposed model makes it possible to systematize educational data, optimize the process of creating questionnaires for testing individual mastery of competencies, and take into account three levels of data analysis. This structure simplifies the processes of data processing and analysis due to the formalized approach to the construction of ontology.

A distinctive feature of the proposed approach is a high degree of detailing of the evaluation criteria when constructing and analyzing the results. The proposed ontology could be easily expanded and adapted for any enterprise by detailing the research data. The key feature of the method is the use of ontology for formalizing and structuring questionnaire data. This allows for more accurate modeling of the subject area and taking into account the relationships between various parameters, which ensures high accuracy of the analysis. Unlike conventional methods, the proposed method not only processes data but also structures them in the form of a formalized model that allows for automatic analysis and adaptation to changes in the data. The constructed mathematical model of the ontology (3) makes it possible to create a clear data structure and formalize the relationships between the key parameters of the questionnaire. This improves the quality of the analysis due to a more detailed representation of the data, which distinguishes this method from classical approaches. Such a model allows for taking into account complex relationships that are often ignored in other analysis methods, which leads to increased accuracy of conclusions. However, like any other solution, our ontology has its limitations. The approach described in the paper makes it possible to evaluate the formation of competencies for a specific respondent. The proposed methods for data storage and processing are focused on the fact that the data is stored in a specific database and compared with previous results. If anonymous analysis of questionnaires is required, this approach will not work. Compared with existing approaches, such as conventional methods for competency assessment, the proposed ontology provides more accurate modeling of the subject area and make it possible to take into account complex relationships between various parameters, which is confirmed by the findings reported in [6, 7, 9], in which data analysis is carried out through structuring their various sources. Unlike these methods, the proposed approach not only systematizes the data but also automates their processing, which makes it possible to adapt to changes in the data. The use of such a model increases the accuracy of the analysis, which was also noted in works related to the implementation of ontologies in the digital transformation of business [8].

The disadvantages of the study include the lack of integration with data on external factors affecting the educational process, such as the lack of data on additional development programs in which the respondent may participate. This disadvantage can be eliminated by expanding the database and applying models that take into account external parameters, which is also noted in study [12], in which extended graph models are used for data management.

Possible development of this study may include expanding the capabilities of the ontology by integrating data on the academic performance of respondents and their activities outside the core curriculum, which will allow for a deeper analysis of the long-term effects of educational programs. This will require solving the problems of standardizing heterogeneous data and ensuring their confidentiality, which was also emphasized in work [14], in which data protection issues were considered in the context of big data. One of the methodological difficulties may be ensuring the adaptability of the system to changes in the structure and format of incoming data, which will require further experimental testing and optimization of the model.

Further development of the study may include expanding the capabilities of the ontology by integrating data on respondents' academic performance in the learning process. This will make it possible to analyze the effectiveness of educational programs in terms of their long-term impact on the formation of IT competencies. A special feature of the method is its integration with business analytics systems, which makes it possible to expand the capabilities of data analysis and automate a number of routine operations. The use of such systems makes it possible to apply machine learning methods and analytical models to identify hidden patterns, which improves the quality and relevance of the results obtained.

7. Conclusions

1. As a result of our systematization of existing methods for processing questionnaire data and approaches to creating ontologies, key trends and problems in the field of automation of data processing have been identified. The systematization makes it possible to generalize and analyze various approaches used for structuring data, which led to the following:

– definition of key parameters and relationships: it was found that successful automation of processing questionnaire data requires the development of clear relationships between the main parameters of the data, such as respondents' competencies and the stages of their assessment. This made it possible to develop an ontological structure that provides precise formalization of these relationships;

– advantages of the ontological approach: comparison of conventional methods with the ontological approach showed that the ontological model significantly increases the accuracy and efficiency of data analysis. It automates the process of classifying questionnaire data and simplifies their interpretation, which leads to improved accuracy of results and reproducibility of the analysis;

– scalability of analysis: the ontological model demonstrated high flexibility in working with large volumes of data, which is especially important when increasing the volume of questionnaires. This has been identified as one of the key advantages of the ontological approach, which is better able to adapt to new conditions compared to conventional analysis methods.

During the study, a detailed systematization of existing methods for analyzing questionnaire data was carried out, such as descriptive and correlation analysis, regression methods, clustering, and other approaches. Particular attention was paid to the ontological approach, which makes it possible to structure and automate the data analysis process. This systematization provided a theoretical basis for further development of the model.

2. The relationships between key parameters, criteria, and results of the questionnaire data analysis have been identified. Three assessment levels (high, medium, low) were identified for each stage of the survey (initial, intermediate, final), which allow for clear tracking of the dynamics of changes in respondents' competencies. Based on this, it can be considered reasonable that the developed model is a system of equations describing changes in the level of competencies over time and allows for a quantitative assessment of knowledge gains and deviations at different stages.

An important feature of such a model is its ability to take into account complex relationships between survey parameters, which significantly increases the accuracy of data analysis. Compared with conventional methods, the model provides deeper systematization and automation of data processing, which improves their reproducibility and scalability.

3. Based on the proposed data analysis methods, graphical representations of the acquisition of competencies were obtained, both in terms of the group and for a specific respondent. In general, for the group, it can be concluded that the used methodology for the formation of IT competencies of preschool children is effective. However, there are still a number of respondents with a low level of formed competence. For further analysis, it is necessary to review the individual cards of respondents with a low level of forced competence. Depending on the goals of the organization, various decisions can be made, for example, if it is necessary to completely exclude the number of respondents with a low level of competence development – apply individual training or change the formation methodology.

4. The developed algorithm for the formation and analysis of questionnaire data includes several key stages, which

ensures its uniqueness and high efficiency. The main stages of the algorithm:

– defining the goals and objectives of the survey: at this stage, the objectives of the study are specified, which allows for the correct structuring of the data and the laying of the foundation for their further analysis. This helps set the correct parameters for data classification and to form key concepts for the ontology;

– formulating survey questions: questions are composed in such a way that they can be compared between different stages of the survey (initial, intermediate, final). This ensures the comparability of results at all stages of the study and sets precise parameters for automatic data processing;

– data classification and structuring: all collected data is classified based on the key concepts identified during the formation of the ontology. This step ensures flexibility of analysis and reduces labor costs for manual data processing;

– automation of processing: an important feature of the algorithm is its ability to automatically update the ontological base as new data is received. This simplifies working with large volumes of information and speeds up the analysis process;

– analysis of changes and decision-making: during the analysis, the algorithm identifies changes in the level of respondents' competencies at each stage of the survey, which allows for prompt adjustment of educational programs or other research parameters based on the results obtained.

5. Our testing of the devised ontological model based on real data obtained from educational institutions showed its high efficiency compared to conventional analysis methods. The use of ontology made it possible to systematize the data and identify key changes in the dynamics of assimilation of competencies at various stages of training (initial, intermediate, and final stages). The model effectively tracked dynamics in the formation of competencies and identified respondents with a low level of knowledge acquisition, which made it possible to adapt educational programs to eliminate these gaps. The results showed that the automation of data analysis using an ontological model can significantly reduce the time spent on data processing. While conventional methods for processing 100 questionnaires required 20–30 hours, our ontological approach reduced this time to 5–8 hours. At the same time, the accuracy of the analysis increased to 92 %, while conventional methods gave an accuracy of about 75 %. The error in forecasts also decreased from 10–15 % to 3–5 %. These results demonstrate that the use of an ontological model not only increases the efficiency of data analysis but also makes the process more accurate and reproducible, especially when working with large volumes of information. Consequently, the developed ontology and algorithms can be used in educational organizations to assess students' competencies, in marketing research to analyze consumer demand, and in sociological research to analyze public opinion. However, the study has limitations associated with the need for constant data updating and possible difficulties in applying the ontology in organizations with low technological infrastructure. The model built can be successfully integrated with a business analytics system, which will allow for effective data visualization and generating reports based on survey results. This will significantly improve the decision-making process since it has become possible to quickly analyze the dynamics of competencies acquisition and evaluate the effectiveness of educational programs based on real data.

Conflicts of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study, as well as the results reported in this paper.

Funding

This research is funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No.AP15473432).

Data availability

The manuscript has associated data in the data warehouse.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

References

1. Hanscom, B., Lurie, J. D., Homa, K., Weinstein, J. N. (2002). Computerized Questionnaires and the Quality of Survey Data. Spine, 27 (16), 1797–1801. https://doi.org/10.1097/00007632-200208150-00020
2. Cartwright, A. (1986). Who responds to postal questionnaires? Journal of Epidemiology & Community Health, 40 (3), 267–273. https://doi.org/10.1136/jech.40.3.267
3. Martin, E., Polivka, A. E. (1995). Diagnostics for Redesigning Survey Questionnaires: Measuring Work in the Current Population Survey. Public Opinion Quarterly, 59 (4), 547–567. https://doi.org/10.1086/269493
4. Chang, R. M., Kauffman, R. J., Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. Decision Support Systems, 63, 67–80. https://doi.org/10.1016/j.dss.2013.08.008
5. Feijóo, C., Gómez-Barroso, J. L., Voigt, P. (2014). Exploring the economic value of personal information from firms' financial statements. International Journal of Information Management, 34 (2), 248–256. https://doi.org/10.1016/j.ijinfomgt.2013.12.005
6. Klee, S., Janson, A., Leimeister, J. M. (2021). How Data Analytics Competencies Can Foster Business Value– A Systematic Review and Way Forward. Information Systems Management, 38 (3), 200–217. https://doi.org/10.1080/10580530.2021.1894515
7. Jansen, B. J., Jung, S., Salminen, J. (2023). Finetuning Analytics Information Systems for a Better Understanding of Users: Evidence of Personification Bias on Multiple Digital Channels. Information Systems Frontiers, 26(2), 775–798. https://doi.org/10.1007/s10796-023-10395-5

8. Tuovinen, L., Smeaton, A. F. (2022). Privacy-aware sharing and collaborative analysis of personal wellness data: Process model, domain ontology, software system and user trial. PLOS ONE, 17 (4), e0265997. https://doi.org/10.1371/journal.pone.0265997
9. Alshammari, M., Simpson, A. (2018). Personal Data Management: An Abstract Personal Data Lifecycle Model. Business Process Management Workshops, 685–697. https://doi.org/10.1007/978-3-319-74030-0_55
10. Moore, J., Goffin, P., Wiese, J., Meyer, M. (2022). Exploring the Personal Informatics Analysis Gap: "There's a Lot of Bacon." IEEE Transactions on Visualization and Computer Graphics, 28 (1), 96–106. https://doi.org/10.1109/tvcg.2021.3114798
11. Deutch, D., Milo, T. (2011). A quest for beauty and wealth (or, business processes for database researchers). Proceedings of the thirtieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (PODS '11). Association for Computing Machinery. New York. https://doi.org/10.1145/1989284.1989286
12. Bellomarini, L., Benedetto, D., Gottlob, G., Sallinger, E. (2022). Vadalog: A modern architecture for automated reasoning with large knowledge graphs. Information Systems, 105, 101528. https://doi.org/10.1016/j.is.2020.101528
13. Janssen, M., van der Voort, H., Wahyudi, A. (2017). Factors influencing big data decision-making quality. Journal of Business Research, 70, 338–345. https://doi.org/10.1016/j.jbusres.2016.08.007
14. Iklassova, K., Aitymova, A., Kopnova, O., Sarzhanova, A., Abildinova, G., Kushumbayev, A., Aitymov, Z. (2024). Construction of a fuzzy model for managing the process of forming IT-competences. Eastern-European Journal of Enterprise Technologies, 3 (3 (129)), 32–43. https://doi.org/10.15587/1729-4061.2024.306183
15. Aitymova, A., Shaporeva, A., Kopnova, O., Kushumbayev, A., Aitymov, Z. (2022). Development and modeling of combined components of the information environment. Eastern-European Journal of Enterprise Technologies, 2 (2 (116)), 51–60. https://doi.org/10.15587/1729-4061.2022.255084
16. Aitymova, A., Iklassova, K., Abildinova, G., Shaporeva, A., Kopnova, O., Kushumbayev, A. et al. (2023). Development of a model of information process management in the information and educational environment of preschool education organizations. Eastern-European Journal of Enterprise Technologies, 2 (3 (122)), 95–105. https://doi.org/10.15587/1729-4061.2023.276253
17. Kulikov, V., Iklassova, K., Kazanbayeva, A. (2020). Entropy based decision making method in managing the development of a socioinformational system. Journal of Theoretical and Applied Information Technology, 98 (1), 92–102.