МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ В ПРИРОДНИЧИХ НАУКАХ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

MATHEMATICAL MODELING IN NATURAL SCIENCES AND INFORMATION TECHNOLOGIES

DOI: 10.31319/2519-8106.1(54)2026.350837
UDC 519.688:004.622

Yalova Kateryna, Candidate of Technical Sciences, Associate Professor, Head of the Department of the Systems software
Ялова К.М., кандидат технічних наук, доцент, кафедра програмного забезпечення систем
ORCID: 0000-0002-2687-5863
e-mail: yalovakateryna@gmail.com

Ismailov Vitaliy, PhD student, Department of the Systems Software
Ісмаілов В.В., здобувач третього (доктора філософії) рівня освіти, кафедра програмного забезпечення систем

Dniprovsky State Technical University, Kamianske
Дніпровський державний технічний університет, м. Кам’янське

METHODS FOR HANDLING UNCERTAINY IN MULTIDIMANTIONAL DATA ANALYSIS IN RELATIONAL DATABASES

МЕТОДИ ВРАХУВАННЯ НЕВИЗНАЧЕНОСТІ В БАГАТОВИМІРНОМУ АНАЛІЗІ ДАНИХ РЕЛЯЦІЙНИХ БАЗ ДАНИХ

The paper addresses the problem of handling uncertainty in multidimensional data analysis of relational databases, which represents an important scientific and practical challenge driven by the growing volume, complexity, and heterogeneity of modern data. The study has a review character and focuses on analyzing the current state of the problem and existing approaches to its solution. The task of multidimensional analysis under uncertainty is formulated, and the main types of uncertainty in relational databases are systematized, including gaps, contradictions, fuzzy data, and semantic ambiguity of join operations. Uncertainty is considered an inherent property of data that significantly affects the reliability of analytical results.

Traditional multidimensional analysis methods, which rely on exact value matching or strict statistical assumptions, are often ineffective in the presence of incomplete or inconsistent data. Therefore, the paper reviews and compares modern methods for uncertainty handling, such as Rough Sets, Bayesian Belief Networks, Dempster–Shafer Theory, the EM algorithm, and Probabilistic Join. The analysis identifies the conditions under which each method is most effective and outlines promising directions for addressing uncertainty in multidimensional analysis of relational data.

***Keywords**: uncertainty, multidimensional data analysis, relational database.*

У статті розглянуто проблему обробки невизначеності під час багатовимірного аналізу даних реляційних баз даних, яка є актуальним науково-практичним завданням. Поява невизначеності в даних, розташованих в реляційних базах даних зумовлена зростанням обсягів, складності та *гетерогенності сучасних даних. У роботі здійснено постановку задачі багатовимірного аналізу в умовах невизначеності та систематизовано основні типи невизначеності, що виникають у даних реляційних баз даних, зокрема пропуски, суперечності, нечіткі значення та семантичну неоднозначність операцій з'єднання. Запропонований формальний опис цих типів дозволяє розглядати невизначеність як невід'ємну властивість даних, що впливає на коректність аналітичних результатів.*

Традиційні методи багатовимірного аналізу базуються на точному порівнянні значень або статистичних припущеннях про коректність даних, що робить їх недостатньо ефективними у присутності пропусків, суперечностей та альтернативних представлень одного об'єкта. У межах дослідження представлено та проаналізовано сучасні методи врахування невизначеності, серед яких теорія грубих множин, Баєсівські мережі довіри, теорія Демпстера–Шафера, алгоритм максимальної апроксимації очікувань та ймовірнісне з'єднання. Порівняльний аналіз показав, що кожен із зазначених методів має чітко визначену область доцільного застосування, обумовлену математичним апаратом та природою невизначеності. Встановлено, що для невизначеності типу «пропуск» найбільш ефективними є ймовірнісні підходи, зокрема ЕМ-алгоритм і Баєсівські мережі довіри, оскільки вони забезпечують імовірнісне відновлення відсутніх значень із урахуванням залежностей між атрибутами. Для обробки суперечливих даних доцільно застосовувати теорію грубих множин і теорію Демпстера–Шафера, які дозволяють локалізувати область конфлікту та кількісно оцінити ступінь невизначеності без необхідності введення жорстких імовірнісних припущень. У випадку нечітких даних найбільш адекватними є методи нечіткої логіки, що формалізують лінгвістичну та інтервальну невизначеність через функції належності. Для невизначеності з'єднання, пов'язаної з неоднозначністю зв'язків між таблицями, найкращі результати забезпечує ймовірнісне з'єднання, яке переносить невизначеність на рівень імовірнісної інтерпретації відповідностей.

***Ключові слова**: невизначеність, багатовимірний аналіз даних, реляційна база даних.*

Problem's formulation

In modern tasks of multidimensional data analysis in relational databases (RDB), problems related to uncertainty, incompleteness, and inconsistency of records frequently arise, significantly complicating the process of obtaining reliable results. Uncertainty may emerge due to measurement errors, missing values, variability of data sources, differences in data representation formats, as well as ambiguities or duplicate records. Ignoring these factors leads to a loss of accuracy, erroneous conclusions, and incorrect modeling of relationships between variables.

These issues become particularly pronounced in large-scale and heterogeneous databases, where values are stored in different formats, using diverse measurement units, abbreviations, or linguistic variations. For example, in financial, medical, geographical, or sociological databases, the same information may be represented in multiple ways: records may contain spelling errors, missing or redundant attributes, changes in the order of elements, ambiguities, and duplicates. Such characteristics substantially complicate automated data processing, classification, aggregation, and analysis.

Traditional methods of multidimensional analysis are based on exact value comparisons or statistical assumptions regarding data correctness, which makes them insufficiently effective in the presence of missing values, inconsistencies, and alternative representations of the same object. To overcome these limitations, modern approaches capable of accounting for uncertainty and imprecision are employed, including probabilistic models, fuzzy logic algorithms, robust methods, and hybrid techniques. These approaches enable data modeling while explicitly considering errors as well as syntactic and semantic variations, thereby providing increased reliability and robustness of analytical results.

Thus, the development and systematization of methods for multidimensional data analysis that take into account uncertainty, missing values, and inconsistencies in databases are critically important for improving the accuracy, stability, and interpretability of results, and have both theoretical and practical significance for a wide range of applied problems.

Analysis of recent research and publications

The problem of optimizing the storage and processing of RDB data under conditions of uncertainty is a relevant scientific and practical task, since the relational model remains the fundamental basis for organizing data warehouses [1—3].

Contemporary research focuses on the formalization of types of uncertainty and the development of methods for their handling. In the work by Channar and Vighio, a relational model for storing uncertain information is presented [1], while O. Pivert and H. Prade described models based on possibility theory for processing incomplete and inconsistent data [2]. The studies by A. Grenyer and G. Gorla analyze methods of multidimensional analysis and uncertainty assessment [3, 4]. The authors of [5] describe the results of developing a framework for the analysis and processing of grey data and data with inherent uncertainty. F. Parisi and J. Grant examined inconsistencies in databases [6], while M. Liao proposed a unified model that integrates data uncertainty and the relationships between data [7]. J. Zhu highlighted trends in the integration of artificial intelligence methods for data cleaning and analysis with consideration of uncertainty in [8]. F. Chacon-Gomez et al. demonstrate the application of Rough Sets to conflicting data [9], and S. Yang combined Bayesian networks with models of uncertain knowledge [10]. L. Fei applied Dempster–Shafer theory for the integration of heterogeneous sources [11], whereas K. Zhao implemented the EM algorithm within RDB for probabilistic value restoration [12]. The authors of [13] proposed a combined model using probabilistic intervals for multiple attributes.

Contemporary research on uncertainty in multidimensional RDB analysis spans from identifying and formalizing uncertainty types to applying classical and hybrid processing methods, highlighting the relevance of the problem and the need to integrate mathematically grounded approaches into practical analytical tools.

Formulation of the study purpose

The aim of this study is to provide a systematic review and analysis of contemporary methods for accounting for uncertainty during the processing and analysis of multidimensional data in databases containing missing values, contradictions, and imprecise entries. The research focuses on the classification, comparison, and evaluation of the effectiveness of approaches, including probabilistic models, fuzzy set methods, Bayesian techniques, which enable the modeling of data while considering errors, missing or conflicting values, as well as variations in formats and representations of information.

The authors do not claim scientific novelty, as the paper has a review character and is aimed at identifying and substantiating the most effective directions for solving the problem of multidimensional data analysis stored in RDB under conditions of uncertainty.

Presenting main materials

Multidimensional data analysis is a set of methods and algorithms aimed at investigating dependencies, structures, and patterns in data characterized by a large number of attributes, while simultaneously accounting for the interrelationships among them. In the context of databases, multidimensional analysis is applied for classification, clustering, forecasting, anomaly detection, and information summarization. Uncertainty in the context of data representation in RDB is a property of the data manifested as absence, incompleteness, ambiguity, or imprecision of attribute values, as well as in the ambiguity of relationships between tuples of different tables, which significantly complicates their analysis and processing.

A relational database *DB* is considered as a set of interrelated tables, each of which describes a specific entity or process:

$$DB = \{T_1, T_2, \ldots, T_k\}, \tag{1}$$

where T_l is each individual table of the database, which is defined by the tuple:

$$T_l = \langle U_l, A_l, V_l, f_l \rangle, \tag{2}$$

where $U_l=\{r_{l1}, r_{l2}, \ldots, r_{lm}\}$ is the set of rows of the table, $A_l=\{a_{l1}, a_{l2}, \ldots, a_{ln}\}$ is the set of attributes forming the feature space, $V_l = \bigcup_{j=1}^{n} V_{lj}$ is the domain of values, V_{lj} is the set of permissible values for the *j*-th column, and $f_l : U_l \times A_l \to V_l \cup \{\perp\}$ is the information function that determines the content of each cell.

The transition from the general structure of tables to their specific content allows for the classification of types of uncertainty that arise directly in attribute values. A mathematical description of uncertainty at the level of table cells in a database can be expressed through a function $f_l(r_i, a_j)$, which maps the state of the data at the intersection of a row and a column. In real-world conditions, the result of this mapping can take various forms depending on the nature of the information deficit.

The primary types of uncertainty in database data are:

1. Gaps. They represent a situation characterized by the complete absence of an attribute value for a given tuple. Such uncertainty arises due to incomplete data collection, loss of information, or the impossibility of recording it at the time of entry. Formally, a gap is modeled by the special value *NULL*, which does not belong to the attribute's domain and cannot be interpreted as a specific value. Mathematically, this is described as:

$$f_l(r_i, a_j) = \perp, \quad \perp \notin V_{lj}\,. \tag{3}$$

The presence of gaps makes it impossible to apply standard comparison and aggregation operations without additional assumptions.

2. Contradictions represent a situation in which, for the same object, mutually exclusive values of a single attribute are obtained from different sources or at different points in time. Such uncertainty is typical for the integration of heterogeneous data sources and cannot be resolved by simply selecting a single value without loss of information. In this case, a cell containing conflicting values is formally considered as a set of values:

$$f_l(r_i, a_j) = \{v^{(1)}, v^{(2)}, \ldots, v^{(p)}\}, \;\; v^{(s)} \in V_{lj}, \quad p > 1\,. \tag{4}$$

3. Fuzzy data. Within RDB, fuzzy data arise when an attribute value cannot be represented as a discrete element of the domain V_{lj} and is instead described by a linguistic term with vague boundaries. This implies that, for a given attribute $a_{lj} \in A_l$, the set of permissible values V_{lj} is extended to a set of fuzzy subsets, where each fuzzy value $\tilde{v}$ is defined by a membership function:

$$\mu_{\tilde{v}} : V_{lj} \to [0,1]\,. \tag{5}$$

Thus, an attribute value is not represented as a point in the feature space, but rather as a distribution of membership degrees, which formalizes linguistic and interval uncertainty within the relational data model. Such uncertainty is characteristic of qualitative attributes and subjective assessments.

4. Join Uncertainty in RDB is the semantic ambiguity of the result of a *JOIN* operation that arises when a foreign key FK does not define a unique correspondence between tuples of two relations. A join operation between tables T_l and T_k is semantically uncertain if, for a given row $r_{li} \in U_l$, the set of rows $r_{kt} \in U_k$, that satisfy the join condition has a cardinality different from one:

$$\left|\{r_{kt} \in U_k \,|\, \varphi_l(r_{li}, a_{lj}) = \varphi_k(r_{kt}, a_{kp})\}\right| \neq 1\,. \tag{6}$$

In this case, the *JOIN* operation generates a set of alternative correspondences between rows of the tables, which makes the join result formally correct but semantically ambiguous. Join Uncertainty reflects a violation of the functional determinacy of the relationship between tables and is a consequence of incomplete, multivalued, or insufficiently constrained key attributes in the relational data model.

The presence of uncertainties in RDB data violates the basic assumptions of classical multidimensional analysis methods regarding data completeness and determinacy, necessitating the application of specialized mathematical models and processing methods capable of correctly handling missing, contradictory, and imprecise values. To enhance the reliability and accuracy of results, various methods are applied, which can be classified according to the mathematical framework employed to compensate for information deficits.

Rough Sets Theory is an effective formal framework for the analysis of conflicting and incomplete data in RDBs, particularly in situations where objects with identical attribute values exhibit different target characteristics. Such conflict often arises due to the limited or fragmented nature of information. Mathematically, for a subset of attributes $B \subseteq A$, an indiscernibility relation $IND(B)$, is defined, which partitions the universe of objects U into equivalence classes $[r_i]_B$. Each such class contains objects that cannot be distinguished based on the attributes in the set B.

Any analytical category (class) $X \subseteq U$ is described by a pair of approximations. The lower approximation represents the set of objects that certainly belong to X:

$$\underline{B}X = \{r_i \in U \,|\, [r_i]_B \subseteq X\}. \tag{7}$$

The upper approximation encompasses objects that may belong to this class:

$$\overline{B}X = \{r_i \in U | [r_i]_B \cap X \neq 0\}. \quad (8)$$

Thus, considering (7)—(8), the uncertainty caused by contradictions is localized within the boundary region, which serves as a quantitative measure of incompleteness and conflict in the database information and allows the evaluation of classification quality without relying on probabilistic assumptions. This is mathematically described as:

$$BN_B(X) = \overline{B}X \setminus \underline{B}X . \quad (9)$$

Bayesian Belief Networks provide probabilistic modeling of dependencies between attributes of relational tables and are effectively applied for handling missing values and structural uncertainty arising during the integration of data from multiple sources. Each attribute in such a model is treated as a random variable, and the relationships between them are described by conditional probabilities. To restore a missing value, as described in (3), the posterior probability is computed:

$$P(a_i | A_{obs}) = \frac{P(A_{obs} | a_i) P(a_i)}{P(A_{obs})}, \quad (10)$$

where A_{obs} are the observed attributes based on which the distribution is computed.

The joint probability of a multidimensional object is factorized according to the structure of the network:

$$P(a_1, a_2, \ldots, a_n) = \prod_{j=1}^{n} P(a_j | parents(a_j)). \quad (11)$$

This enables probabilistic inference even in cases where relationships between tables are lost or key values are missing.

Dempster–Shafer Theory is intended for the formalized combination of conflicting data originating from different relational tables or sources. Each information source is associated with a mass function $m(A)$, which reflects the degree of belief in the attribute value. The combination of conflicting data from independent sources is performed according to Dempster's rule:

$$(m_1 \oplus m_2)(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_1(B) m_2(C), \quad (12)$$

where K — the measure of conflict between sources, is defined as:

$$K = \sum_{B \cap C = 0} m_1(B) m_2(C) . \quad (13)$$

Thus, the influence of conflicting records during information aggregation is reduced.

For the processing of fuzzy data, where the boundaries between object states are vague, methods based on membership functions $\mu_j(x_{ij}) \in [0,1]$ are applied, corresponding to fuzzy clustering techniques such as Fuzzy C-Means. In this case, the objective function of multidimensional analysis minimizes the weighted distance to the cluster centers:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} (u_{ij})^m \| r_i - c_j \|^2 , \quad (14)$$

where u_{ij} is the degree of membership of the i-th object to the j-th cluster, allowing a single tuple with imprecise values to partially belong to multiple analytical categories.

For datasets with a large number of missing values that exhibit a stochastic nature, the Expectation-Maximization (EM) algorithm is applied. The process consists of two main steps:

1. E–step: Computation of the expected value of the log–likelihood for the missing data X_{mis} given the current parameter vector $\theta^{(t)}$, which is mathematically expressed as:

$$Q(\theta | \theta^{(t)}) = E[\ln L(\theta; X_{obs}, X_{mis}) | X_{obs}, \theta^{(t)}]. \quad (15)$$

2. M–step: Updating the parameters by maximizing the function Q, which can be mathematically represented as:

$$\theta^{(t+1)} = \arg\max Q(\theta | \theta^{(t)}). \quad (16)$$

The Probabilistic Join method is appropriate when the standard algorithm ignores records that satisfy the join condition *C* of the classical *JOIN* operation described in (6), particularly when the attributes involved contain missing or multivalued conflicting data. Probabilistic Join constructs a probabilistic correspondence function $P(r_{ai} \approx r_{bj})$, which defines the probability that a row r_i from table T_a corresponds to a row r_j from table T_b, instead of using the binary predicate function of the classical join condition. The result of the Probabilistic Join operation is a set of row pairs with associated probabilities:

$$T_a \bowtie pT_b = \{(r_i, r_j, P(r_i, r_j)) | P(r_i, r_j) > 0\}. \quad (17)$$

Probabilistic Join allows the preservation of information regarding all potentially relevant correspondences between tables, without discarding records with incomplete or conflicting keys, while transferring the uncertainty to the level of a probabilistic interpretation of the results. This provides a more flexible and information-rich analysis of the data.

To systematize the properties of the considered methods and facilitate their selection depending on the specific type of uncertainty, a comparative analysis is appropriate. Tabl. 1 presents a comparative characterization of uncertainty processing methods according to key criteria: their mathematical foundation, functional advantages, and existing limitations in the context of RDBs.

Table 1. Comparative characterization of processing methods

Method	Core Mathematical Concept	Main Advantage	Limitation
Rough Sets	Approximation (lower and upper bounds) through the indiscernibility relation	Does not require knowledge of the probability distribution; works with conflicting rules	Sensitive to significant noise in numerical data
Bayesian Belief Networks	Modeling conditional dependencies via directed acyclic graphs	Allows incorporation of prior knowledge; flexible in missing value restoration	Requires large datasets for accurate learning of the graph structure
Dempster-Shafer Theory	Combination of mass (belief) functions for conflict resolution	Efficient integration of data from multiple independent tables/sources	Computational complexity increases with the number of hypotheses (states)
EM-algorithm	Iterative search for maximum likelihood parameter estimates	High mathematical accuracy in restoring the structure of the distribution	May converge to a local maximum instead of the global maximum
Probabilistic Join	Table join based on calculated correspondence weights of records	Prevents data loss when join keys (Foreign Keys) are incomplete or conflicting	Requires complex similarity metrics for record comparison

The generalized architecture of multidimensional data analysis in RDB under conditions of uncertainty consists of three main components, as illustrated in Fig. 1. In the first stage, raw data from relational tables are processed, where various types of uncertainty are identified. The second component is the computational unit, which applies uncertainty-handling methods to integrate and process these data. Finally, the processing results are transmitted to the multidimensional analysis module, which provides analytical evaluation of the data through classification, clustering, and prediction, producing the final analytical outcome.

Methods based on approximations, probabilistic models, or combinations of belief functions exhibit varying sensitivity to noise, sample size, and computational complexity, which determines the appropriateness of their use in specific scenarios. An analysis of their properties, limitations, and advantages has allowed the formulation of generalized recommendations for selecting an uncertainty processing method corresponding to a particular type of uncertainty, as presented in Tabl. 2.

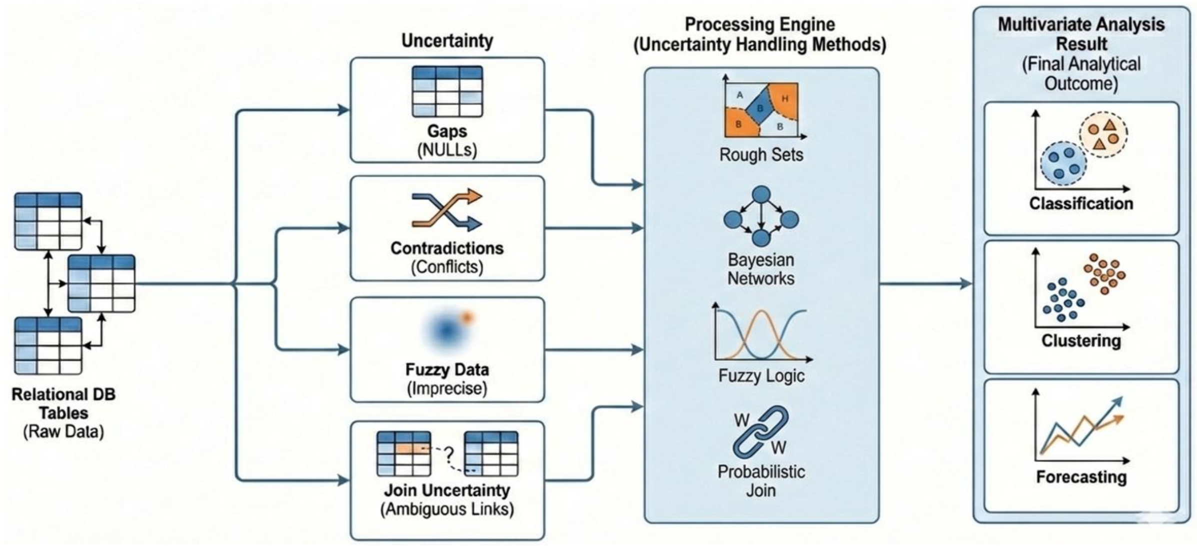

Fig. 1. Generalized architecture of multidimensional data analysis with uncertainty relational data

Table 2. Recommended methods for handling uncertainty in RDBs

Type of Uncertainty	Recommended Method	Mathematical Outcome
Gaps	EM-algorithm, Bayesian Belief Networks	Probabilistic imputation of missing values
Contradictions	Rough Sets, Dempster-Shafer Theory	Determination of the boundary region and conflict measure
Fuzzy Data	Fuzzy Logic	Determination of membership degrees μ of term-sets
Join uncertainty	Probabilistic Join	Formation of an analytical space with confidence weights

The combination of a comparative characterization and a recommendation-based approach enables the justified selection of methods for processing missing data, conflicting or fuzzy records, as well as uncertain relationships between tables, providing effective construction of an analytical space that takes into account the specifics of RDBs.

Conclusions

The conducted study confirms that modern multidimensional analysis in RDBs is not feasible without considering factors of uncertainty, which manifest at different levels of the structure —from individual cells to complex inter–table relationships.

Based on the mathematical formalization of relational model components, four fundamental types of uncertainty were identified and described in detail, namely Gaps, Contradictions, Fuzzy Data, and Join Uncertainty, which allowed clear delineation of the applicability domains of various mathematical frameworks. The analysis of contemporary approaches, including Rough Sets Theory, Bayesian Belief Networks, Dempster–Shafer Theory, and probabilistic join algorithms, demonstrated their capability to effectively compensate for information deficits while preserving the semantic integrity of the data.

Within the scope of this study, a generalized architecture of a multidimensional data analysis system in RDBs operating under conditions of uncertainty was developed and presented. The proposed structure ensures a logical transition from the level of raw relational tables to the final analytical outcome, passing through filters of specialized uncertainty-processing methods. The mathematical frameworks of leading approaches — including Rough Sets Theory, Bayesian networks, Dempster–Shafer Theory, the EM–algorithm, and Probabilistic Join — were described and analyzed in detail. Special attention was given to the analysis of the advantages and limitations of each method, which enabled the determination of their effectiveness depending on data volume and noise level, as well as the establishment of correspondences between types of uncertainty and the most relevant processing methods.

The study demonstrates that Gaps-type uncertainty is most effectively processed using probabilistic methods such as the EM-algorithm and Bayesian Belief Networks, which allow probabilistic imputation of missing values while accounting for dependencies between attributes. For Conflicting data, Rough Sets and Dempster–Shafer Theory are the most adequate, as they allow localization of the conflict region and quantitative assessment of uncertainty without information loss. In the case of Fuzzy Data, fuzzy logic methods formalize linguistic and interval uncertainty through membership degrees, whereas for Join Uncertainty, the optimal approach is Probabilistic Join, which constructs an analytical space with confidence weights and preserves all potentially relevant correspondences between tables.

References

[1] Channar, S.H., & Vighio, M.S. (2022). A relational data model for uncertain data. International Journal of Emerging Multidisciplinaries: Computer Science& Artificial Intelligence, 1(2), 46—55. doi: https://doi.org/ 10.54938/ijemdcsai.2022.01.2.141.
[2] Pivert, O., & Prade, H. (2018). Handling uncertainty in relational databases with possibility theory — A survey of different modelings. Proceedings of the 12th International Conference on Scalable Uncertainty Management (SUM `18), Oct 2018, Milan, Italy. pp.396—404. doi: ff10.1007/978-3-030-00461-3_30ff.
[3] Grenyer, A., Erkoyuncu, J.A., Zhao, Y., & Roy, R. (2021). A systematic review of multivariate uncertainty quantification for engineering systems. CIRP Journal of Manufacturing Science and Technology, 33, 188—208. doi: https://doi.org/10.1016/j.cirpj.2021.03.004.
[4] Gorla, G., Taborelli, P., & Giussani, B. (2023). A multivariate analysis-driven workflow to tackle uncertainties in miniturized NIR data. Molecules, 28(24):7999, 1—19. doi: 10.3390/molecules28247999.
[5] Yang, Y., Liu, S. & Xie, N. (2019). Unceraintу and grey data analytics. Marine economics and management, 2(2), 73—86. doi: https://doi.org/10.1108/MAEM-08-2019-0006.
[6] Parisi, F., & Grant, J. (2023). How many inconsistencies are in your database? CEUR Workshop Proceeding, 3478, 1—11.
[7] Liao, M., Shen, D., & Lv, P. (2023). A unified model of data uncertainty and data relation uncertainty. Knowledge-Based Systems, 278. doi: https://doi.org/10.1016/j.knosys.2023.110811.
[8] Zhu, J., Zhao, X., Sun, Y., Song, S., & Yuan, X. (2025). Relational data cleaning meets artificial intelligence: a survey. Data Science and Engineering, 10, 147—174. doi: https://doi.org/10.1007/s41019-024-00266-7
[9] Chacon-Gomez, F., Cornejo, M.E., Medina, J., & Ramirez-Poussa, E. (2023). Rough set decision algorithms for modeling with uncertainty. Journal of Computational and Applied Mathematics 437(3):115413. doi: 10.1016/j.cam.2023.115413.
[10] Yang, S., Lin, L., Liu, C., Zhang, M., & Yang, H. (2025). Prediction of implicit relationships using uncertain knowledge graph embedding and Bayesian networks. Neural Processing Letters, 1, 57—72. doi: https://doi.org/10.1007/s11063-025-11794-2.
[11] Fei, L., Li, T., & Ding, W. (2024). Dempster-Shafer theory-based information fucion for natural disaster emergency menagement: A systematic literature review, 112, 102585. doi: https://doi.org/10.1016/j.inffus.2024.102585.
[12] Zhao, K., Yu, X. J., Rong, Y., Liao, M. & Huang J. (2021). Towards Expectation-Maximization by SQL in RDBMS. arXiv:2101.09094. doi: https://doi.org/10.48550/arXiv.2101.09094.
[13] Nguen, H., & Tran, T.N. (2024). A relational database model with probability intervals for uncertain set-valued attributes. Malaysian Journal of Science and Advanced Technology, 4(4), 456—463. doi: https://doi.org/10.56532/mjsat.v4i4.329.

Список використаної літератури

1. Channar S.H., Vighio M.S. A relational data model for uncertain data. International Journal of Emerging Multidisciplinaries: *Computer Science & Artificial Intelligence*. 2022. № 1(2). P. 46—55. doi: https://doi.org/ 10.54938/ijemdcsai.2022.01.2.141.
2. Pivert O., Prade H. Handling uncertainty in relational databases with possibility theory — A survey of different modelings. *Proceedings of the 12th International Conference on Scalable Uncertainty Management* (SUM `18), Oct 2018, Milan, Italy. pp. 396 — 404. doi: ff10.1007/978-3-030-00461-3_30ff.
3. Grenyer A., Erkoyuncu J.A., Zhao Y., Roy R. A systematic review of multivariate uncertainty quantification for engineering systems. *CIRP Journal of Manufacturing Science and Technology*. 2021. №33. P. 188—208. doi: https://doi.org/10.1016/j.cirpj.2021.03.004.
4. Gorla G., Taborelli P., Giussani B. A multivariate analysis-driven workflow to tackle uncertainties in miniturized NIR data. *Molecules*. 2023. № 28(24):7999. P. 1—19. doi: 10.3390/molecules28247999.
5. Yang Y., Liu S. Xie N. Uncerainty and grey data analytics. *Marine economics and management*. 2019. № 2(2). P. 73—86. doi: https://doi.org/10.1108/MAEM-08-2019-0006.
6. Parisi F., Grant J. How many inconsistencies are in your database? *CEUR Workshop Proceeding*. 2023. № 3478. P. 1—11.
7. Liao M., Shen D., Lv P. A unified model of data uncertainty and data relation uncertainty. *Knowledge-Based Systems*. 2023. № 278. P. 1—12. doi: 10.1016/j.knosys.2023. 110811.
8. Zhu J., Zhao X., Sun Y., Song S., Yuan X. Relational data cleaning meets artificial intelligence: a survey. *Data Science and Engineering*. 2025. № 10. P. 147—174. doi: https://doi.org/10.1007/s41019-024-00266-7
9. Chacon-Gomez F., Cornejo M. E., Medina J., Ramirez-Poussa E. Rough set decision algorithms for modeling with uncertainty. *Journal of Computational and Applied Mathematics*. 2023. № 437(3):115413. doi: 10.1016/j.cam.2023.115413.
10. Yang S., Lin L., Liu C., Zhang M., Yang H. Prediction of implicit relationships using uncertain knowledge graph embedding and Bayesian networks. *Neural Processing Letters*. 2025. № 1. 57—72. doi: https://doi.org/10.1007/s11063-025-11794-2.
11. Fei L., Li T., Ding W. Dempster-Shafer theory-based information fucion for natural disaster emergency menagement: A systematic literature review. *Information Fusion*. 2024. № 112. ID 102585. doi: https://doi.org/10.1016/j.inffus.2024.102585.
12. Zhao K., Yu X. J., Rong Y., Liao M., Huang J. Towards Expectation-Maximization by SQL in RDBMS. *arXiv*. 2021. № 2101. ID 09094. doi: https://doi.org/10.48550/arXiv.2101.09094.
13. Nguen H., & Tran T.N. A relational database model with probability intervals for uncertain set-valued attributes. *Malaysian Journal of Science and Advanced Technology*. 2024. № 4(4), P. 456—463. doi: https://doi.org/10.56532/mjsat.v4i4.329.

Надійшла до редколегії 26.11.2025
Прийнята після рецензування 10.01.2026
Опублікована 22.01.2026