Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 4 (34)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

UDC 004.358:681.518
DOI: https://doi.org/10.30837/2522-9818.2025.4.101

V. FILATOV, O. ZOLOTUKHIN, M. KUDRYAVTSEVA

INTELLECTUAL DATA ANALYSIS IN RELATIONAL INFORMATION AND ANALYTICAL SYSTEMS

The subject of the study is the methods of intellectual analysis, namely the construction of a decision tree, associative analysis, the identification of patterns between related events based on data presented by a relational model. **The purpose of the study** is to analyze the features of information units and data structures, using the example of relational systems that affect the technology of knowledge extraction. **Tasks:** the article solves the following tasks: to consider the relational data model as the most popular and effective data structure used in intelligent information systems for data processing and storage; to analyze the operations of relational algebra, the operational component of the relational data model regarding the application of aggregate functions; to develop a general formal statement of the problem of knowledge extraction from a relational database; to consider the concept of functional associative rules; the ID3 decision tree generation algorithm focused on data processing in relational systems is analyzed. **The following methods are implemented**: modern view and trends in the field of data mining; features of building information systems based on relational databases, relational algebra, theory of normalization of relations; analysis of literature on the topic of research; comparative analysis. **Results achieved:** the relational data model is considered as the most effective data structure used in intelligent information systems for data processing and storage. A group of aggregate functions of relational databases is identified and analyzed with respect to key attributes of the relation, which makes it possible to build logical dependencies between information units of the subject area being analyzed. The task of extracting knowledge from the database is formally formulated. The concept of functional associative rules is introduced. The ID3 decision tree generation algorithm focused on data processing in relational systems is carefully analyzed. The semantic network (SN), built on the basis of the proposed approach, allows to increase the efficiency of decision support systems. **Conclusions:** the universal approach proposed in the article to build a relational data model of an information system for searching for associative patterns in data allows to solve a whole class of typical tasks in which objects are connected by a "many-to-many" relationship or $M \rightarrow N$. The relational database model is proposed as a universal information structure for solving associative analysis tasks and presenting knowledge in the form of a semantic network. The examples given in the article confirm the effectiveness of the developed and considered approaches to solving the problem of data mining in the environment of relational systems. Solving the problem of identifying knowledge in data will allow to improve the quality of management decisions made.

Keywords: relational model, data mining, associative dependencies, database, decision tree, semantic network, information system.

Introduction

The main trends in the development of information systems and databases (DB) are, on the one hand, decentralization and distribution of information management resources, and on the other hand, heterogeneity of data storage structures of information systems. Thus, distributed automated systems become the basis of the information structure for data storage, processing, and management and ensure the informatization of society – the gradual creation of a unified information space. Given the expansion of regional, national, and international integration processes in the economy and business, the demands on the performance of distributed intelligent automated systems are growing [1].

The development of computing technology and the increase in the volume of stored information have led to the need to separate database technology into a separate field of intensive scientific research. Over more than 40 years of history, databases have evolved from automated workstations, so-called ARMs, file and client-server technologies to information spaces. Information management is undergoing changes in the following aspects: storage and access models, the scale of projected systems – from databases to intelligent decision support systems. New areas of research are rapidly developing in this field: integrated information systems and technologies, information spaces and communities, systems based on computational intelligence; issues related to information analysis remain relevant. The development of effective methods for analyzing large volumes of primary, or so-called "raw", data will enable the acquisition of new knowledge and the making of informed decisions.

There is currently a rapid growth in the number of software products that use new technologies, as well as in the types of tasks where their application provides

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 4 (34)

significant economic benefits. Elements of automatic data processing and analysis, known as *Data Mining* (knowledge extraction), are becoming an integral part of the concept of databases, electronic data warehouses, and the organization of intelligent computing [2, 3].

However, traditional mathematical statistics, which has long claimed to be the main tool for analyzing information, is effective for solving real-life complex problems. It operates with averaged sample characteristics, which are often fictitious values (such as the average temperature of patients in a hospital). Therefore, mathematical statistics methods are mostly useful for testing pre-formulated hypotheses. *Data Mining* is increasingly becoming a multidisciplinary field that has emerged on the basis of achievements in various sciences [4–6]. Hence, there are a significant number of methods and algorithms implemented in various existing *Data Mining* systems. Many of them integrate several approaches. However, each system usually has a key element on which the main focus is concentrated.

In the development of this direction, the use of such mathematical apparatus in database technology as aggregate functions is of interest. Aggregate functions are functions that determine the number of records in a table, count the number of values in a column or find the minimum and maximum values for it, and also sum up the data. Aggregate functions include *COUNT*, *SUM*, *MAX*, *MIN*, *AVG*, and probably others that may be offered by the developer of a particular system. To apply aggregate functions in calculations of a group of identical values, the *GROUP BY* parameter is used. It "compresses" identical values of a given attribute into a single row of summary results. For example, to find the average price of a part, you can formulate a query in *SQL* [7, 8].

This article is devoted to the study of information systems that combine relational databases as a source of primary information and means of intelligent analysis.

The purpose of the study

The purpose of the study is to analyze the relational data model, namely the impact of the normalization level of the relational database schema on knowledge extraction technology using the example of developing a semantic model of knowledge representation and associative data analysis. The results achieved will make it possible to formulate conditions for effective collaboration with analytical processing systems based on *Data Mining* methods.

Relational database as an information storage environment in intelligent analytical systems

A relational DB is based on the relational model. This model is defined by a simple data structure, a user-friendly tabular representation, and the ability to use the formal apparatus of relational algebra and relational calculus for information processing [9–12].

Traditional means of specifying the relational data model are used to construct the structural diagram of databases.

The basic structural unit of data in the relational model is an n-ary relation, which is a finite subset of the Cartesian product of domains, i.e., sets of atomic values of data elements – attributes of the relation.

Let R be a finite subset of the names of relations in the database;

D – $\{D_1,...,D_i\}$ – a set of domains where each domain is a named set of atomic values of data elements;

A – final set of attribute names of a relation;

dom – reproduction from A to D, which determines from which domain the attribute values are selected.

A pair $\langle A_i, domA_i \rangle$, where $A_i \in A$, is called an attribute.

The structural diagram S_i of the relation $R_i (R_i \in R)$ can be presented in the form $R_i (A_1,...,A_n)$, in which all A_i are different. The relation r_i can be defined as an extension of the diagram S_i: $r_i \subseteq domA_1 \times ... \times domA_n$.

Reordering attributes in the schema does not generate a new extension, and the set $\{A_1,...A_n\}$ of relation attributes R_i defines the relation type. The expression $R_i = A_1...A_n$ is used to specify the composition of the medium. The structural schema U of a relational database is a specification of the form $(R_1,...,R_p)$, where $R_i \in R$ and all R_i are different [13].

The task of identifying knowledge in relational databases is considered, the effective solution of which will improve the quality of management decisions.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 4 (34)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Building a decision tree based on data presented by a relational model

Formally, the task of extracting knowledge from a DB can be demonstrated as follows. The subject area is reproduced in the form of a relational model, presented as a universal relation *R*, as a subset of tuples of the Cartesian product

$$R = (DX_1, DX_2, ..., DX_n, DY_1, ..., DY_m) = \\ = \{< x_1, ..., x_n, y_1, ..., y_m >\},$$

where x_i – values of input attributes X_i from the domain DX_i; y_i – values of output attributes Y_j from the domain DY_j; $P(x_1, ..., x_n, y_1, ..., y_m)$ – predicate as a condition for reproducing a specific subject area in a tuple of attribute values $< x_1, ..., x_n, y_1, ..., y_m >$.

The objective of the study is to develop a set of rules:

$$\{X_1, .X_2, ..., X_n\} \to \{Y_1, Y_2, ..., Y_m\},$$

which assign a set of target values $\{y_j = DY_j, j = \overline{1,m}\}$ to each incoming set of values

$$\{x_i = DX_i, i = \overline{1,n}\}.$$

The obtained functional dependencies

$$Y_j = F_j(X_1, X_2, ..., X_n), j = \overline{1,m}$$

must be correct for relation tuples and can be used in the process of finding output attributes Y_j for new input attribute values $X_i, i = \overline{1,n}$.

If the database is presented in the form of a strongly typed relation, for example, in third normal form, then for subsequent calculations it is necessary to switch to a universal relation. To transform the model, you can use the join operation contained in the basic operations of relational algebra [14].

The join is equivalent to the following sequence of relational operations:

– renaming of identical attributes in relations A and B;

– Cartesian product of relations A and B;

– selection by corresponding attribute values that had the same names in relations A and B;

– projection with the removal of duplicate attributes from relations A and B;

– renaming attributes of relations A and B, returning them to their original names.

The universal relation obtained by such a transformation is a set of facts, each of which is described by a finite set of discrete attributes [15].

For further data analysis, one of the *Data Mining* methods can be applied – building *Decision Trees*. Decision Trees are currently the most common approach to identifying and reproducing logical patterns in data. Among them are:

- ID3 (*Interactive Dichotomizer*);
- CART (*classification and regression trees*);
- CHAID (*chi square automatic interaction detection*).

Let's take a closer look at the process of building decision trees using the ID3 system as an example. The data table fully meets the requirements for the structure of the relational model discussed above, namely: the key attribute ***Day*** is selected, the tuples are unique, and the attributes of the relation are not repeated.

The ID3 decision tree generation algorithm is an algorithm that builds a tree from the root to the leaves, selecting the attribute that best classifies the data at each node. The input parameters of the algorithm are: Examples – the current training example – the target attribute, Attributes – a set of candidate attributes.

Let's consider the general scheme of the algorithm's operation using the example of a universal relation presented in Table 1. We select the target attribute – this is *Play Tennis*.

At this stage, the root node of the tree is formed. The *InformationGain* is calculated for each candidate attribute, and the one with the *InformationGain* that best matches expressions (1) and (2) is selected.

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (1)$$

$$Entropy(S) \equiv -p_+ \log_2 p_+ - p_- \log_2 p_-, \quad (2)$$

where p_+ – number of positive examples; p_- – number of negative examples; S – a set of attributes, $S_v = \{s \in S | A(S) = v\}$.

Table 1 presents a sample containing 14 tuples. To determine the root node, it is necessary to calculate the entropy of all the parameters under study – the attributes of the relation. Among all the attributes, the information threshold is highest in *Outlook*, which is defined on the set of values *Sunny*, *Overcast*, and *Rain*. According to the values of parameters (1) and (2), a tree letter is formed (Fig. 1).

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Table 1. *Weather conditions*

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

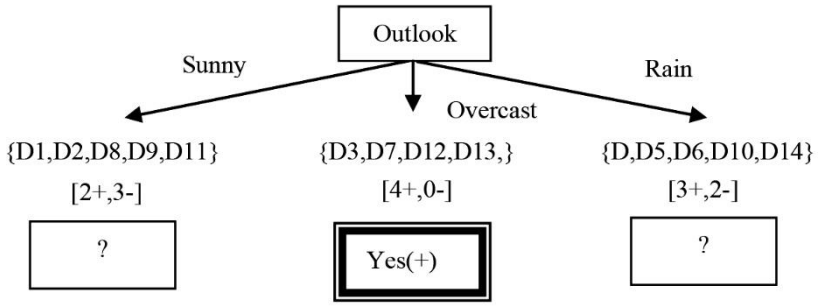

Fig. 1. Example of constructing a decision tree

We perform similar calculations until all the leaves of the tree are formed.

As a result, we obtain the desired decision tree, shown in Fig. 2.

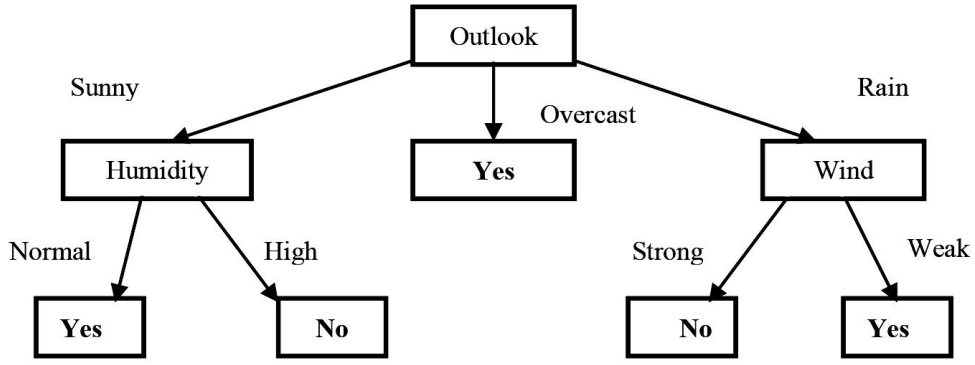

Fig. 2. Decision tree

The above algorithm for extracting and representing knowledge has a number of advantages, among which the following can be highlighted:

– high speed of the knowledge discovery process;

– generation of rules for subject areas in which it is not possible to obtain a formal model of knowledge representation by other methods;

– intuitively understandable classification model of the subject area.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 4 (34)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

The results obtained can be interpreted in the form of one of the classical models of knowledge representation – a semantic network. A semantic network is a directed graph whose vertices are concepts and whose edges are the relationships between them. Concepts are usually abstract or concrete objects, and relations are connections such as "genus-species", "part-whole", "class-subclass", etc. A distinctive feature of semantic networks is the mandatory presence of three types of relationships: "class – class element", "property – value", and "example – class element".

A semantic network provides the following basic functions:

– storage of information about objects and the relationships between them;

– search for objects by various properties;

– replenishment and correction of the system's knowledge during its training;

– implementation of various procedures for generalizing and specifying knowledge.

In general, a semantic network is understood as an expression in the form of $S = \langle O, R \rangle$, where $R = \{R_j, j = \overline{1,k}\}$, $O = \{O_i, i = \overline{1,n}\}$, $O_i, i = \overline{1,n}$ a set of objects in a specific subject area; $R_j, j = \overline{1,k}$ – a set of relations between objects; j – type of relationship.

Let's imagine a decision tree obtained by applying the ID3 algorithm in the form of a semantic network. The root of the tree, *Outlook*, is linked to the child nodes by the relationships *Sunny*, *Overcast*, and *Rain*. By analyzing all the components of the decision tree – links and relationships – we can represent a set of objects in the subject domain as follows:

Outlook (Sunny, Overcast, Rain)

Humidity (Normal, High)

Wind (Strong, Weak)

Thus, for example, the problem situation "how weather affects tennis" can be described as a semantic network, the structure of which is shown in Fig. 3.

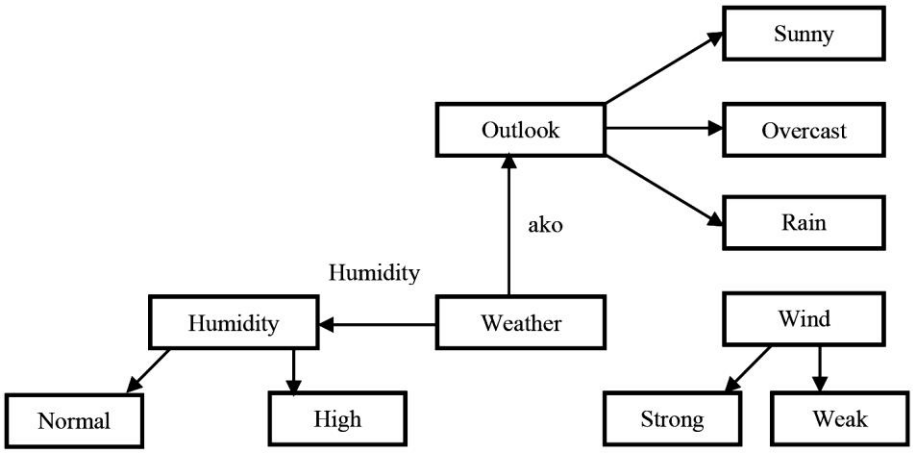

Fig. 3. Semantic network

With the help of a semantic network, it is possible to provide a formal decision-making procedure and pre-record network elements in the form of predicates. From the analysis of the decision tree obtained using the ID3 algorithm, we can draw the disappointing conclusion that three weather parameters follow from this decision: *Outlook*, *Humidity*, *Wind*. The predicate will look like this: let's go play (yes/no) (*Outlook*, *Humidity*, *Wind*).

So,

If we go play

Yes (*Outlook* = *Sunny*, *Humidity* = *Normal*)

Yes (*Outlook* = *Rain*, *Wind* = *Weak*)

Yes (*Outlook* = *Overcast*)

If we are not going to play

No (*Outlook* = *Sunny*, *Humidity* = *High*)

No (*Outlook* = *Rain*, *Wind* = *Strong*)

Based on the predicates formulated above, a semantic network can be developed, the appearance of which is shown in Fig. 4.

It should be noted that the approach to intelligent data analysis and knowledge representation in the form of a semantic network proposed in this article can be applied to large-scale databases, the schema of which may contain more than 1,000 attributes.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 4 (34)

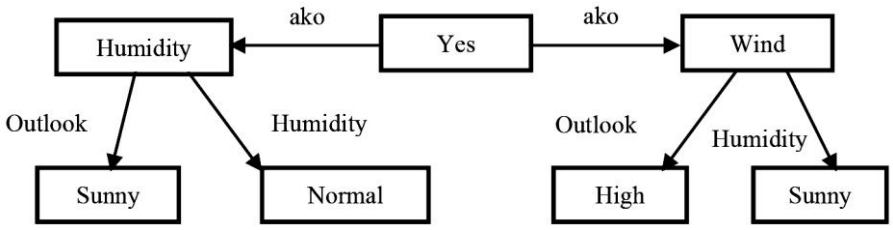

Fig. 4. Semantic network construction

Searching for associative rules in data represented by a relational model

Most often, the method of evaluating the associative properties of data is used in the following areas:

– retail (determining which products should be promoted together; choosing the location of products in the store; analyzing the consumer basket; forecasting demand);

– marketing (searching for market segments, trends in consumer behavior);

– customer segmentation (identifying common characteristics of the company's customers, identifying buyer groups);

– catalog design, sales campaign analysis, determining customer purchase sequences (which purchase will follow the purchase of product A);

– analysis of web logs.

One of the most frequently cited examples of associative rule mining is the Market-Basket Problem. The essence of this problem is to identify products that people buy together. This is necessary so that marketing specialists can place these products in the store in an appropriate manner to increase sales, as well as make other effective decisions. It is the ability to discover hidden rules that makes associative rule mining valuable and contributes to knowledge discovery.

Let us consider this task in a generalized form (3)–(11). Let us denote the objects that make up the sets under study by the set

$$I = \{i_1, i_2, ...i_r, ...i_m\}, \quad (3)$$

where i – objects contained in the sets being analyzed; n – total number of objects.

Transactions are called sets of objects located in the database that will be analyzed. Let's describe a transaction as a subset I:

$$T = \{i_j \mid i_j \in I\}. \quad (4)$$

The set of transactions for which information is available for analysis will be presented as a set D

$$D = \{T_1, T_2, ..., T_r, ..., T_m\}, \quad (5)$$

where m – number of transactions available for analysis.

Set of transactions to which i_j-object belongs,

$$D = \{T_r \mid i_j \subseteq T_r; j = 1...n; r = 1...m\} \subseteq D. \quad (6)$$

We will define the general set of objects as follows:

$$F = \{i_j \mid i_j \in I; j = 1...n\}. \quad (7)$$

The set of transactions to which the set F belongs will be represented as

$$D_F = \{T_r \mid F \subseteq T_r; r = 1...m\} \subseteq D. \quad (8)$$

The ratio of the number of transactions to which the set F belongs to the total number of transactions is defined as the *support* of the set and denoted by $Supp(F)$:

$$Supp(F) = |D_F| / |D|. \quad (9)$$

During the search process, the analyst can specify the minimum support value for the sets that interest him $Supp_{\min}$. A set is called large (*large itemset*) if its support value exceeds the minimum support value specified by the user:

$$Supp(F) > Supp_{\min}. \quad (10)$$

Therefore, when searching for associative rules, it is necessary to find a set in the sets:

$$L = \{F \mid Supp(F) > Supp_{\min}\}. \quad (11)$$

Development of a database model for solving associative analysis problems

Let us consider the options for relational database (RDB) structures that meet the requirements of tasks (3)–(11) for searching for associative rules [16]. Expression (3) describes the objects of intellectual analysis. According to the rules of RDB design, the following model can be considered for representing objects from different subject areas: the "Objects"

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 4 (34)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

relation has a key attribute "**Object ID**" and a list of attributes $A_1, ..., A_n$. An example of the relationship between "Objects" in general terms and the practical "SUPPLY" is given.

Fig. 5 shows a diagram of the "Objects" relationship. As an example, a fragment of a database for displaying a list of goods is given, which in theoretical-set form looks like this: $SUPPLY = \{ID, NAME, PRICE\}$. In the methodology discussed above (3)–(11), a transaction is understood to be an object or set of second-level objects that are determined by the integration properties of previously located objects. Therefore, the second-level object relationship diagram will be identical to the diagram shown in Fig. 6.

"Objects"

Object ID
Attribute 1
Attribute 2
Attribute 3
Attribute n

"SUPPLY"

ID	**Name**	**Price**
101101	LG HD Phone	1452
101102	PCHP321	5687
101103	PRINTER	540

Fig. 5. Example of the "Objects" relations

Tax invoices

ID_N	**Delivery**	**Supplier**
101	10-11-2009	MKS
102	11-11-2009	FOX
103	12-11-2009	EPICENTER

Fig. 6. Relation "Tax invoices"

Let us determine the type of relationships between first- and second-level objects. Note that a single transaction may contain several first-level objects and, accordingly, there cannot be duplicate objects in a single transaction.

The type of relationship considered corresponds to the many-to-many multiple type, or $M \to N$. In a relational data model, this type of relationship is implemented through an additional relation [17]. An example of a transaction is a real second-level object with integration properties – a delivery note containing objects (goods). The RDB schema "Waybills_Goods" is shown in Fig. 7.

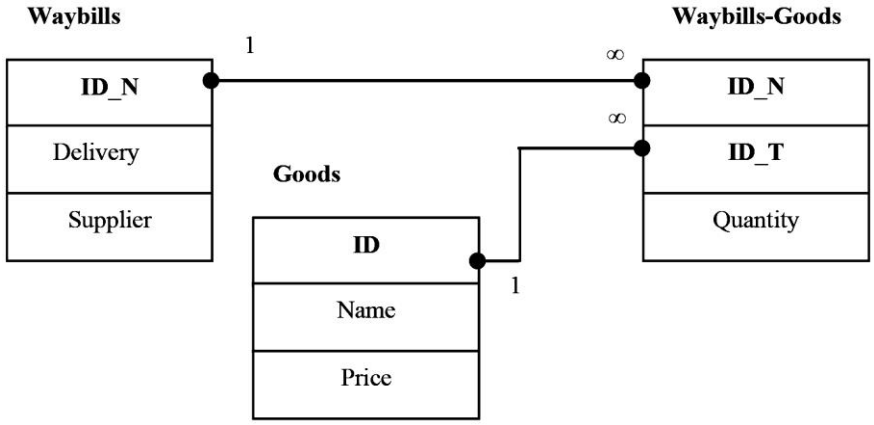

Fig. 7. Database "Waybills_Goods" schema

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 4 (34)

The "Waybills_Goods" database schema contains three relations: "Waybills", "Goods", and "Waybills Goods". Key attributes are marked in bold. A double composite key is used for "Waybills_Goods". This solution allows us to maintain the integrity constraint formulated at the stage of setting the task of synthesizing the database schema: one item can be added twice to one waybill. An example of the "Waybills_Goods" relation is shown in Fig. 8.

Waybills_Goods

ID_N	Supplier	**Goods ID**	Name	Quantity
101	MKS	**101102**	PC HP321	1
101	MKS	**101103**	PRINTER	1
102	FOX	**101103**	PRINTER	2
103	EPICENTER	**101101**	LG HD Phone	1
103	EPICENTER	**101102**	PC HP321	1
103	EPICENTER	**101103**	PRINTER	1

Fig. 8. Database "Waybills_Goods" schema

The above relationship "Invoices_Goods" meets the normalization conditions: goods PRINTER and PC HP321 are added to invoice No. 101, invoice No. 102 contains one item PRINTER, and invoice No. 103 contains three goods: PRINTER, PC HP321, and LG HD phone.

An arbitrary set of objects added to the analyzed ones forms a set for associative analysis. This parameter is variable and is a subset of the set:

$$I = \{LG\ HD\ phone, PC\ HP321, PRINTER\}.$$

Let $F = \{LG\ HD\ phone\}$, then the content of the associative analysis task (9)–(11) will be as follows: determine how many times an object $LG\ HD\ phone$ occurs in the total number of goods received according to all invoices $Supp(LG\ HD\ phone) = 1/6$.

The universal approach to building a relational data model for an information system that searches for associative patterns in data, as proposed in this article, makes it possible to solve a whole class of typical problems in which objects are related by a "many-to-many" relationship, or $M \rightarrow N$ [18]. Examples of such tasks include: a polyclinic (this task involves two entities in a $M \rightarrow N$ relationship: "Patients" and "Doctors"); library ("Readers" and "Literature"); audio and video media rental ("Customers" and "Discs"), etc.

As an example, let's take the task of searching for associative patterns in the subject area "POSTGRADUATE STUDIES" of a higher education institution (HEI) in Ukraine. Let's consider one of the most important tasks performed by this department. PhD and doctoral dissertations are defended by applicants in specialized councils, which include scientists who are leading experts in scientific fields in their specialties. A specialized council may consist of 5 (PhD defense) to 25 specialists, if it is a specialized doctoral council. In addition, according to the requirements, one scientist-specialist in a specific field of research cannot participate in the work of more than two specialized councils. The universal relationship of the relational database is shown in Fig. 9.

Specialist_Council_Specialty

id	Last Name	HEI	Tel	**Council**	Address	Specialty
1	Ivchenko	NURE	33-22	Д.О1.001	KhPI	05.13.06
1	Ivchenko	NURE	33-22	Д.02.011	KhAI	05.13.06
2	Petrenko	NURE	33-22	Д.О1.001	KhPI	05.13.23
3	Skidan	KhPI	11-44	Д.О1.001	KhPI	05.13.06
3	Skidan	KhPI	11-44	Д.02.011	KhAI	05.13.23

Fig. 9. Relationship "Specialist_Council_Specialty"

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 4 (34)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

As a result of normalizing the universal relation shown in Fig. 9, the relational DB schema can be transformed and take the form shown in Fig. 10.

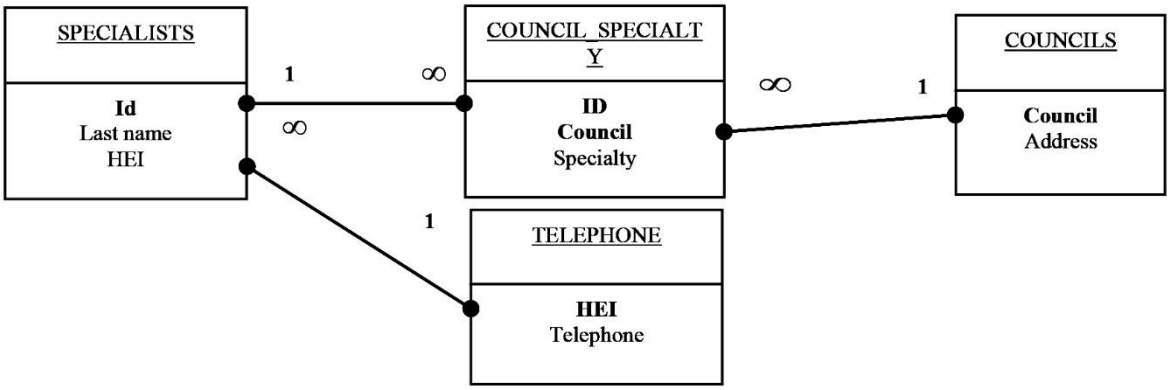

Fig. 10. Schema of the relational DB "Specialist_Council_Specialty"

The COUNCIL-SPECIALITY relation fully satisfies the requirements for a data structure oriented towards the application of associative pattern search technology (3)–(11).

The example considered confirms that normalized relations of the relational data model, reduced to 3 normal form, containing a double composite key, allow for the most effective application of associative data analysis methods [19].

Conclusions

The scientific novelty of the research lies in a comprehensive analysis of the relational data model, namely, identifying the impact of the normalization level of the RDB schema on knowledge extraction technology using the example of developing a semantic model of knowledge representation and associative data analysis. The results obtained make it possible to formulate conditions for effective collaboration with analytical processing systems based on *Data Mining* methods.

The article examines the features of relational algebra operations regarding the application of aggregate functions and considers the concept of functional associative rules, which contributes to the implementation of the classic ID3 decision tree generation algorithm focused on information processing in relational systems. The universal approach to building a relational data model of an information system for searching for associative patterns in data, proposed in the article, makes it possible to solve a whole class of typical tasks in which objects are related by a "many-to-many" relationship, or $M \to N$. The semantic network, built on the basis of the proposed approach, helps to improve the efficiency of decision support systems.

The relational DB model, proposed as a universal information structure for solving associative analysis tasks, will improve the efficiency of the analyzed approach through its implementation in modern information systems.

The examples given in the article confirm the effectiveness of the developed and considered approaches to performing the task of intelligent data analysis in the environment of relational systems. Solving the task of identifying knowledge in data will contribute to improving the quality of management decisions.

Theoretical and practical issues of automatic or automated construction (generation) of the structure and model of an expert system knowledge base, based, for example, on a production model and, therefore, the integration of information system data designed on a relational data model and an expert system, can be considered as a prospect for further development of this area of research.

References

1. Xuanhe, Z., Chengliang, C., Guoliang, L., Ji, S. (2022), "Database Meets Artificial Intelligence: A Survey". *IEEE Transactions on Knowledge and Data Engineering*, Vol. 34, No. 3, P. 1096–1116. DOI: 10.1109/TKDE.2020.2994641
2. Chen, J., Sun, J., Wang, G. (2022), "From Unmanned Systems to Autonomous Intelligent Systems". *Engineering*, 12(5), P. 16–19. DOI: 10.1016/j.eng.2021.10.007

3. Zhang, F., Yuan, N. J., Lian, D., Xie, X., Ma, W.-Y. (2016), "Collaborative knowledge base embedding for recommender systems". *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, P. 353–362. DOI: 10.1145/2939672.2939673
4. Avrunin, O., Vlasov, O., Filatov, V. (2020), "Model of semantic integration of information systems properties in relay database reengineering problems". *Innovative Technologies and Scientific Solutions for Industries*, 4 (14), P. 5–12. DOI: 10.30837/itssi.2020.14.005
5. Cappuzzo, R., Papotti, P., Thirumuruganathan, S. (2020), "Creating embeddings of heterogeneous relational datasets for data integration tasks". *In Proceedings of the 2020 ACM SIGMOD international conference on management of data*. P. 1335 – 1349. DOI: 10.1145/3318464.3389742
6. Shi, C., Li, Y., Zhang, J., Sun, Y., Yu, P. S. (2017), "A Survey of Heterogeneous Information Network Analysis". *IEEE Transactions on Knowledge and Data Engineering*, 29(1), P. 17–37. DOI: 10.1109/TKDE.2016.2598561
7. Parciak, M., Weytjens, S., Hens, N., Neven, F., Peeters, L. M., Vansummeren, S. (2025), "Measuring approximate functional dependencies: a comparative study". *The VLDB Journal*, 34(4). DOI: 10.1007/s00778-025-00931-x
8. Filatov, V., Doskalenko, S. (2018), "On the Approach to Searching for Functional Dependences of Data in Relational Systems". *Innovative Technologies and Scientific Solutions for Industries*, 1 (3), P. 54–58. DOI: 10.30837/2522-9818.2018.3.054
9. Filatov, V., Semenets, V., Zolotukhin, O. (2019), "Synthesis of Semantic Model of Subject Area at Integration of Relational Databases". *2019 IEEE 8th International Conference on Advanced Optoelectronics and Lasers (CAOL).* DOI: 10.1109/caol46282.2019.9019532
10. Filatov V., Kovalenko A. (2019), "Fuzzy Systems in Data Mining Tasks". *Advances in Spatio-Temporal Segmentation of Visual Data. Studies in Computational Intelligence*, Vol 876. Springer, Cham P. 243–274. DOI: 10.1007/978-3-030-35480-0_6
11. Glava, M., Malakhov, V. (2018), "Information Systems Reengineering Approach Based on the Model of Information Systems Domains". *International Journal of Software Engineering and Computer Systems (IJSECS).* Vol. 4, P. 95–105. DOI: 10.15282/ijsecs.4.1.2018.8.0041
12. Codd, E. F. (1983), "A relational model of data for large shared data banks". *Communications of the ACM*. Vol. 26, No. 1. P. 64–69. DOI: 10.1145/357980.358007
13. Maier, D. (1983), *The theory of relational databases*. London: Pitman, 637 p.
14. Wan, X., Han, X., Wang, J., Li, J. (2024), "Efficient Discovery of Functional Dependencies on Massive Data". *IEEE Transactions on Knowledge and Data Engineering*, 36(1), P. 107–121. DOI: 10.1109/tkde.2023.3288209
15. Wang, Y. (2025), "Design and Implementation of a General Data Collection System Architecture Based on Relational Database Technology. In: Xu, Z., Alrabaee, S., Loyola-González, O., Ab Rahman, N.H. (eds) *Cyber Security Intelligence and Analytics. CSIA 2024. Lecture Notes in Networks and Systems*, Vol 1351. Springer, Cham. DOI: 10.1007/978-3-031-88287-6_53
16. Date, C. J. (2003), *Introduction to database systems.* Pearson Education, Limited. 1024 p.
17. Hector Garcia-Molina, Jennifer Widom, Jeffrey D. Ullman (2013), *Database systems the complete book*. Pearson India Education, 1139 p.
18. Sliusarenko, T., Filatov, V. (2023), "Relational vs non-relational databases". *Grail of science.* No. 23. P. 269–271. DOI: 10.36074/grail-of-science.23.12.2022.41
19. Filatov, V. O., Yerokhin, A. L., Zolotukhin, O. V., Kudryavtseva, M. S. (2019), "Information space model in tasks of distributed mobile objects managing". *Information Extraction and Processing*, 2019(47), P. 80–86. DOI: 10.15407/vidbir2019.47.080

Received (Надійшла) 06.06.2025
Accepted for publication (Прийнята до друку) 30.11.2025
Publication date (Дата публікації) 28.12.2025

Відомості про авторів / About the Authors

Filatov Valentin – Doctor of Sciences (Engineering), Professor, Kharkiv National University of Radio Electronics, Professor of Artificial Intelligence Department, Kharkiv, Ukraine; e-mail: valentin.filatov@nure.ua; ORCID ID: https://orcid.org/0000-0002-3718-2077; Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=56911938100

Zolotukhin Oleh – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Dean of Computer Science Faculty, Kharkiv, Ukraine; e-mail: oleg.zolotukhin@nure.ua; ORCID ID: https://orcid.org/0000-0002-0152-7600; Scopus ID: https://www.scopus.com/authid/detail.uri?origin=resultslist&authorId=57207774022

Kudryavtseva Maryna – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Professor of Artificial Intelligence Department, Kharkiv, Ukraine; e-mail: maryna.kudryavtseva@nure.ua; ORCID ID: https://orcid.org/0000-0003-0524-5528; Scopus ID: https://www.scopus.com/authid/detail.uri?origin=resultslist&authorId=57207765829

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 4 (34)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Філатов Валентин Олександрович – доктор технічних наук, професор, Харківський національний університет радіоелектроніки, професор кафедри штучного інтелекту, Харків, Україна.

Золотухін Олег Вікторович – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, декан факультету комп'ютерних наук, Харків, Україна.

Кудрявцева Марина Сергіївна – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, професор кафедри штучного інтелекту, Харків, Україна.

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ У РЕЛЯЦІЙНИХ ІНФОРМАЦІЙНО-АНАЛІТИЧНИХ СИСТЕМАХ

Предметом дослідження є методи інтелектуального аналізу, а саме побудова дерева рішень, асоціативний аналіз, виявлення закономірностей між пов'язаними подіями на основі даних, які подано реляційною моделлю. **Мета** – проаналізувати реляційну модель даних, зокрема вплив рівня нормалізації схеми реляційної бази даних на технологію видобування знань на прикладі розроблення семантичної моделі подання знань і асоціативного аналізу даних. У статті необхідно виконати такі **завдання:** розглянути реляційну модель даних як найбільш популярну й ефективну структуру, яка використовується в інтелектуальних інформаційних системах оброблення та зберігання даних; проаналізувати особливості операції реляційної алгебри щодо застосування агрегатних функцій; розробити загальну формальну постановку завдання видобування знань з бази даних реляційного типу; розглянути поняття функціональних асоціативних правил, алгоритм генерації дерев рішень ID3, орієнтований на оброблення даних у реляційних системах. Упроваджено такі **методи:** реляційна алгебра, теорія нормалізації відношень, порівняльний аналіз. **Досягнуті результати**. Досліджено реляційну модель даних як найбільш ефективну структуру, що використовується в інтелектуальних інформаційних системах оброблення та зберігання даних. Виокремлено та проаналізовано групу агрегатних функцій реляційних баз даних щодо ключових атрибутів відношення, що дає змогу будувати логічні залежності між інформаційними одиницями предметної галузі, яка аналізується. Формально сформульовано задачу видобування знань з бази даних. Запропоновано поняття функціональних асоціативних правил. Ретельно проаналізовано алгоритм генерації дерев рішень ID3, орієнтований на оброблення даних у реляційних системах. Семантична мережа, побудована на основі запропонованого підходу, сприяє підвищенню ефективності систем підтримки прийняття рішень. **Висновки**. Запропонований у статті універсальний підхід до побудови реляційної моделі даних інформаційної системи пошуку асоціативних закономірностей у даних дає змогу розв'язувати цілий клас типових завдань, в яких об'єкти пов'язані відношенням "багато до багатьох", або M → N. Реляційна модель бази даних запропонована як універсальна інформаційна структура для виконання завдань асоціативного аналізу та подання знань у вигляді семантичної мережі. Наведені в статті приклади підтверджують ефективність розроблених і розглянутих підходів до розв'язання задачі інтелектуального аналізу даних у середовищі реляційних систем. Виконання поставленої задачі виявлення знань у даних дасть змогу підвищити якість прийнятих управлінських рішень.

Ключові слова: реляційна модель; інтелектуальний аналіз даних; асоціативні залежності; база даних; дерево рішень; семантична мережа; інформаційна система.

Бібліографічні описи / Bibliographic descriptions

Філатов В. О., Золотухін О. В., Кудрявцева М. С. Інтелектуальний аналіз даних у реляційних інформаційно-аналітичних системах. *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 4 (34). С. 101–111. DOI: https://doi.org/10.30837/2522-9818.2025.4.101

Filatov, V., Zolotukhin, O., Kudryavtseva, M. (2025), "Intellectual data analysis in relational information and analytical systems", *Innovative Technologies and Scientific Solutions for Industries*, No. 4 (34), P. 101–111. DOI: https://doi.org/10.30837/2522-9818.2025.4.101