

УДК 004.021

<https://doi.org/10.17721/1812-5409.2022/3.6>

Замрій І.В.¹, к.ф.-м.н., доц.,
Вишнівський В. В.², д.т.н., проф.

I. V. Zamrii¹, Cand. of Phys. and Math. Sc. (Ph.D.),
Associate Professor,
V. V. Vyshnivskyi², D.Sc., Professor

Метод ідентифікації атомарних елементів інформаційного простору за допомогою єдиного інформаційного простору

Method of identification of atomic elements of information space with the help of a unified information space

^{1,2} Державний університет телекомунікацій,
03141, м. Київ, вул. Солом'янська, 7

^{1,2} State University of Telecommunications, 03141,
Kyiv, Solomianska street, 7

e-mail: irinafraktal@gmail.com ²vish_vv@ukr.net

У статті розглядається метод, який дозволяє ідентифікувати вхідний атомарний елемент інформаційного простору шляхом об'єднання інформаційних просторів в єдиний інформаційний простір (це інформаційна модель складної предметної області, яка включає інформаційні об'єкти, зв'язки між ними, середовище простору та процеси, що супроводжують створення єдиного інформаційного простору). Даний метод дозволяє виконати уточнення відсутніх ознак для вхідного елемента для забезпечення можливості додавання його до єдиного інформаційного простору.

Ключові слова: єдиний інформаційний простір, інформаційна система, параметри елемента інформаційного простору, штучний інтелект.

The article considers a method that allows to identify the input atomic element of the information space by combining information spaces into a unified information space (this is an information model of a complex subject area, which includes information objects, relationships between them, space environment and processes). This method allows you to refine the missing features for the input element to enable it to be added to a unified information space.

Key words: unified information space, information system, parameters of the information space element, artificial intelligence.

Широке використання інтегрованих інформаційних систем управління даними обумовлене необхідністю та потребою в автоматизації та оптимізації роботи. На підприємствах з підвищеними ризиками небезпеки виключення участі людини у технологічних процесах є необхідністю, а швидкість та точність обміну інформацією є одним із ключових факторів успішного виконання технічних завдань. Тому сучасні підприємства та установи все більше застосовують інтегровані інформаційні системи управління даними, як для управління підприємства, так, і для комп'ютеризації технологічних, фінансових, логістичних та інших процесів. Основним та найбільш перспективним методом комп'ютеризованого управління складними динамічними системами та процесами в життєво важливих і критичних з точки зору безпеки та надійності галузях є диспетчерське

управління, моніторинг та диспетчеризація – SCADA (Supervisory Control And Data Acquisition) системи. На принципах SCADA-систем будуються великі автоматизовані системи в промисловості, енергетиці, транспорті, в космічній та військових галузях, при комп'ютеризації будівель та споруд.

Огляд останніх досліджень та публікацій. Інтеграція комп'ютеризованих систем усіх підрозділів підприємства в єдиний інформаційний простір сукупності всіх інформаційних систем є необхідністю для забезпечення стійкості технологічних, фінансових, логістичних та інших процесів та їх взаємозв'язків в єдиній системі. Актуальність даного питання в різних галузях підтверджується роботами [1]-[3].

У роботі [4] описано методику створення єдиного інформаційного простору на сучасному виробничому підприємстві з функціонально

стійким виробничим процесом. Сучасні промислові підприємства у великій мірі автоматизовані та здатні автономно виконувати багато технологічних процесів протягом наперед заданих часових інтервалів випускаючи продукцію з високими стандартами споживчої та експлуатаційної якості. Залежно від типу, призначення, способу організації технічної системи, нині гостро стоїть проблема мінімізації впливу людського фактору у виробничих процесах, розвитку критеріїв оцінки і методів підвищення функціональної стійкості для інформаційних систем підприємства; ретельне вивчення окремих класів технічних систем і визначення найбільш ефективних засобів підвищення їх функціональної стійкості [5].

У роботі [6] розроблено математичну модель інформаційного простору виробничого підприємства на основі пов'язаних даних та побудовано на її основі архітектуру розподіленої інформаційної системи, функціональність якої може бути збільшена шляхом додавання нових модулів та схем даних шляхом конфігурування керуючих компонентів. Запропонована модель єдиного інформаційного простору показує взаємодію користувача з інформаційним простором та забезпечує ефективне виконання завдань оперативного управління.

Під час використання інтегрованих інформаційних систем управління даними виникає багато питань стосовно розпізнавання даних та розподілу їх по різних системах і базах даних, встановлення всіх зв'язків, тобто параметрів. Так у роботі [7] запропоновано метод, що дозволяє відповідати вимогам щодо надання достовірної та послідовної інформації про структурні підрозділи компанії та призначати спеціальні ролі для основних інформаційних об'єктів організації. У роботі [8] розглядається підхід до виділення критеріїв для унікальної ідентифікації деяких важливих видів інформаційних об'єктів, незалежно від виду носія або конкретного кодування. Про ідентифікацію одночасно в декількох базах даних йдеться у роботі [9]. У роботі [10] запропоновано оптимізований алгоритм, який в 30 разів швидше виявляє об'єкти у реальному часі. Покращений метод на основі пошуку сукупності для налаштування параметрів ознак екстрактора, що дозволяє отримати найкраще значення в інформаційному та вартісному значенні параметрів функціонуючої системи розпізнавання зображень за кілька ітерацій роботи алгоритму відображено у роботі [11].

У роботі [12] розглядаються питання управління якістю та ризиками для зберігання, обробки та відновлення інформації з урахуванням витрат за допомогою різних різних ІТ-методів.

У роботі [13] розглянуто етапи ідентифікації інформаційного об'єкта в єдиному інформаційному просторі та запропоновано метод пошуку відсутніх ознак вхідного об'єкта шляхом реалізації взаємодії інформаційних об'єктів між собою всередині єдиного інформаційного простору, хоча не розглянуто вид топології для реалізації такого пошуку.

Розглянуто методи пошуку та ідентифікації інформаційних об'єктів у роботі [14]. Але розглянуті алгоритми при цьому є досить вузькими і не виходять за рамки простих інтерфейсів, що залишає відкритим питання про розроблення більш масштабованих алгоритмів мінімізації інтерфейсів.

Але не зважаючи на все ідентифікація елементів інформаційного простору, їх відновлення або відновлення параметрів елементів інформаційного простору є не вивченим повністю питанням, яке залежить від типів зв'язків, тобто топології; галузі дослідження, кожна з яких має свою особливість; методів дослідження, які можуть оптимізуватися у зв'язку із тотальною інформатизацією суспільства та іншим.

Мета статті. Метою даної роботи є удосконалення методів ідентифікації елементів інформаційного простору за допомогою єдиного інформаційного простору.

Основні результати досліджень. Розглянемо множину інформаційних просторів, таких як IS_1, IS_2, \dots, IS_n , де n — це кількість інформаційних просторів. Кожен такий інформаційний простір має свій набір інформаційних об'єктів — елементів інформаційного простору e_1, e_2, \dots, e_m , де m — кількість елементів інформаційного простору. Кожен елемент інформаційного простору має відповідний набір параметрів t_1, t_2, \dots, t_k , де k — кількість параметрів, і для кожного e_1, e_2, \dots, e_m кількість параметрів однакова (якщо деякі параметри відсутні, то їх значення дорівнює *NONE*). Вхідні елементи інформаційної системи, інформація про які зчитується за допомогою сенсорів, позначимо через $e_{b_1}, e_{b_2}, \dots, e_{b_i}$.

Зважаючи на те, що в інформаційному просторі всі його елементи різні, тобто $e_1 \neq e_2 \neq \dots \neq e_m$, він може бути представлений як

система, що постійно оновлюється та поповнюється даними. Для цього необхідно:

1. Розпізнати вхідні елементи інформаційної системи e_{b_i} не прив'язуючись до конкретного інформаційного елемента і таким чином сформуванати інформаційний простір.

2. Для кожного вхідного елемента інформаційної системи за допомогою сенсорів зчитати його характеристики, тобто параметри.

3. Провести ідентифікацію вхідного елемента інформаційної системи, тобто інформаційний простір повинен однозначно визначити чи існує ще один елемент з аналогічними параметрами.

Формально таку модель можна описати у вигляді:

$$IS = \langle E, T, Z, f \rangle,$$

де E – множина елементів інформаційного простору; T – множина параметрів елементів інформаційного простору; Z – множина типів зв'язків, яка знаходиться як об'єднання множини зв'язків між елементами інформаційного простору та множини зв'язків між параметрами елементів інформаційного простору; f – відображення, що задає конкретне відношення з множини типів зв'язків Z між елементами інформаційного простору.

Єдиний інформаційний простір, що по суті є об'єднанням всіх інформаційних просторів, можна подати у вигляді графа:

$$UIS = \langle IS(E, T, Z, f), \tilde{Z}, \tilde{f} \rangle,$$

де IS – множина інформаційних просторів; \tilde{Z} – множина типів зв'язків, яка знаходиться також як об'єднання множини зв'язків між елементами різних інформаційних просторів та множини зв'язків між параметрами елементів різних інформаційних просторів); \tilde{f} – відображення, що задає конкретне відношення з множини типів \tilde{Z} між множиною інформаційних просторів IS .

Підмножина параметрів T_{ij} визначається для вхідного елемента інформаційного простору e_{b_i} , що подається на вхід інформаційного простору та вже існуючого в інформаційному просторі елемента e_j . Таким чином для кожного вхідного елемента інформаційного простору e_{b_i} є два можливі варіанти: 1) якщо $T_{ij} = T_i \cap T_j$ не співпадає з набором параметрів e_{b_i} та e_j , то такий вхідний елемент вважається новим і включається до інформаційного простору IS , при цьому додаються зв'язки Z_{ij} між новим елементом та вже існуючими; 2) якщо $T_{ij} = T_i \cap T_j$ співпадає з набором параметрів e_{b_i} та e_j , то такий вхідний

елемент вважається ідентифікованим і не включається до інформаційного простору IS .

Таким чином однозначно представляється кожен елемент інформаційного простору в інформаційному просторі.

Розглянемо метод паралельної ідентифікації, який полягає в тому, що пошук елемента інформаційного простору здійснюється незалежно і паралельно у всіх інформаційних просторах. Суть даного методу полягає у наступних кроках:

Крок 1. Потрібно сформуванати незалежні інформаційні простори, при цьому можуть виникати повторення елементів інформаційних просторів в різних просторах і навіть в одному й тому ж самому.

Крок 2. Необхідно виключити у кожному сформованому інформаційному просторі всі елементи інформаційних просторів, які повторюються, але в різних інформаційних просторах однакові елементи можуть залишитися.

Крок 3. Проводиться процес порівняння елементів інформаційних просторів у кожному з інформаційних просторів із деяким вхідним елементом інформаційного простору e_{b_i} з набором параметрів t_1, t_2, \dots, t_k . Вхідний елемент e_{b_i} однозначно ідентифікується $e_{b_i} = IS_i(e_{b_j})$, якщо значення кожного з параметрів вхідного елемента потрапляє в допустимий інтервал значень $M - D \leq t_k \leq M + D$ для відповідного параметру деякого елемента інформаційного простору. Елементи інформаційного простору мають локальну пам'ять параметрів. Коли відбувається однозначна ідентифікація вхідного елемента, то значення його параметрів додаються в пам'ять відповідних параметрів елемента інформаційного простору. Потім, з множини значень кожного параметру елемента інформаційного простору, визначаються статистичні характеристики, які описують цей параметр – математичне сподівання M та дисперсія D , і чим більше вхідних елементів ідентифікується інформаційним простором, тим більш точними вони стають.

Крок 4. Знайдені елементи інформаційного простору, що відповідають критеріям пошуку виводяться, при цьому уточнюється з якого інформаційного простору.

Крок 5. Проводиться повторне зчитування сенсорами для встановлення відсутніх параметрів вхідного елемента інформаційного простору. При вдалому зчитуванні параметр стає відомим, а при

невдалому — приймає значення *NULL* та ініціюється його пошук.

Крок 6. У разі відсутності ознак для однозначної ідентифікації вхідного елемента необхідно звернутися до так би мовити «глобальної пам'яті» кожного елемента інформаційного простору, які зберігають траєкторії взаємодії та всі разом об'єднуються в єдине сховище даних, тобто необхідно запросити необхідні відсутні параметри.

Крок 7. Якщо ідентифікація не відбулася, тобто вхідний елемент не співпав з жодним елементом інформаційного простору у всіх просторах, то він вважається новим і додається до інформаційного простору в якості ще одного елемента інформаційного простору, якщо значення всіх його параметрів відомі. Для коректної класифікації з метою додавання нових елементів до інформаційного простору використаємо багатовимірний дискримінантний аналіз. Побудуємо функцію, що задається наступним рівнянням:

$$f_{IS_r}^j = \delta_0^r + \delta_1^r t_{1r}^j + \dots + \delta_k^r t_{kr}^j,$$

де $f_{IS_r}^j$ – значення функції для j -го елемента інформаційного простору в IS_r ; $t_{\varphi r}^j$ ($\varphi = 1, \dots, k$) – параметри, які відмінні від *NONE*, для e_j в інформаційному просторі IS_r ; δ_0^r – вільний коефіцієнт, який відповідає простору IS_r ; δ_φ^r ($\varphi = 1, \dots, k$) – коефіцієнти, що відповідають параметрам, відмінним від *NONE* для кожного IS .

Обчислюємо значення функції $f_{IS_r}^j$ для кожного e_j ($j = 1, \dots, q$). Потім на основі отриманих значень $f_{IS_r}^j$ для кожного IS_r обчислюємо математичне сподівання $M(f_{IS_r})$ та дисперсію $D(f_{IS_r})$, які дозволяють визначити допустимий інтервал значень і визначити приналежність елемента e_j до деякого інформаційного простору. Таким чином, для вхідного елемента e_{b_i} обчислюється значення функції f та порівнюється із статистичними характеристиками кожного інформаційного простору. Якщо виконується умова

$$M(f_{IS_r}) - D(f_{IS_r}) \leq f \leq M(f_{IS_r}) + D(f_{IS_r}),$$

то вхідний елемент e_{b_i} вважається новим і додається до інформаційного простору у вигляді e_{m+1} . Такий підхід дозволяє однозначно визначити до якого саме інформаційного простору необхідно додати новий елемент.

Блок-схему даного методу наведено на рис.1.

Узагальнимо даний метод ідентифікації вхідного елемента шляхом об'єднання інформаційних просторів в єдиний інформаційний простір. Основою розроблюваного методу є процес об'єднання окремих інформаційних просторів в єдиний інформаційний простір. При цьому, всі інформаційні елементи e_j з усіх інформаційних просторів IS об'єднуються в єдиний інформаційний простір UIS для подальшого пошуку в ньому вхідного елемента e_{b_i} .

Розглянемо основні етапи даного методу:

Крок 1. Формування кожного окремого інформаційного простору незалежно один від одного.

Крок 2. Всі елементи інформаційного простору з усіх інформаційних просторів об'єднуються в єдиний інформаційний простір UIS .

Крок 3. В єдиному інформаційному просторі виключаються елементи інформаційного простору, що повторюються, тобто перебудовується єдиний інформаційний простір UIS_γ .

Крок 4. Вхідний елемент e_{b_i} , що має набір параметрів t_1, t_2, \dots, t_k , порівнюється з кожним елементом з UIS_γ . Якщо значення кожного параметра потрапляє в допустимий інтервал значень $M - D \leq t_i \leq M + D$ відповідного параметру деякого елемента e_j , то він однозначно ідентифікується.

Крок 5. Інформаційні елементи, які відповідають критеріям виводяться.

Крок 6. Проводиться повторне зчитування сенсорами для встановлення параметрів вхідного елемента.

Крок 7. Вхідний елемент додається до єдиного інформаційного простору в якості нового елемента інформаційного простору, якщо значення всіх його параметрів є зчитаними і відомими, використовуючи багатовимірний дискримінантний аналіз.

Блок-схему узагальненого методу наведено на рис.2.

За допомогою теоретико-множинного формалізму відбувається пошук елемента інформаційного простору за його параметрами. Для цього множина всіх інформаційних просторів позначається S , яка складається з елементів e_i ($i = \overline{1, m}$), а множина всіх параметрів позначається T і складається з параметрів елементів $t_u(S)$ ($u = \overline{1, n}$), де n – це кількість допустимих ознак.

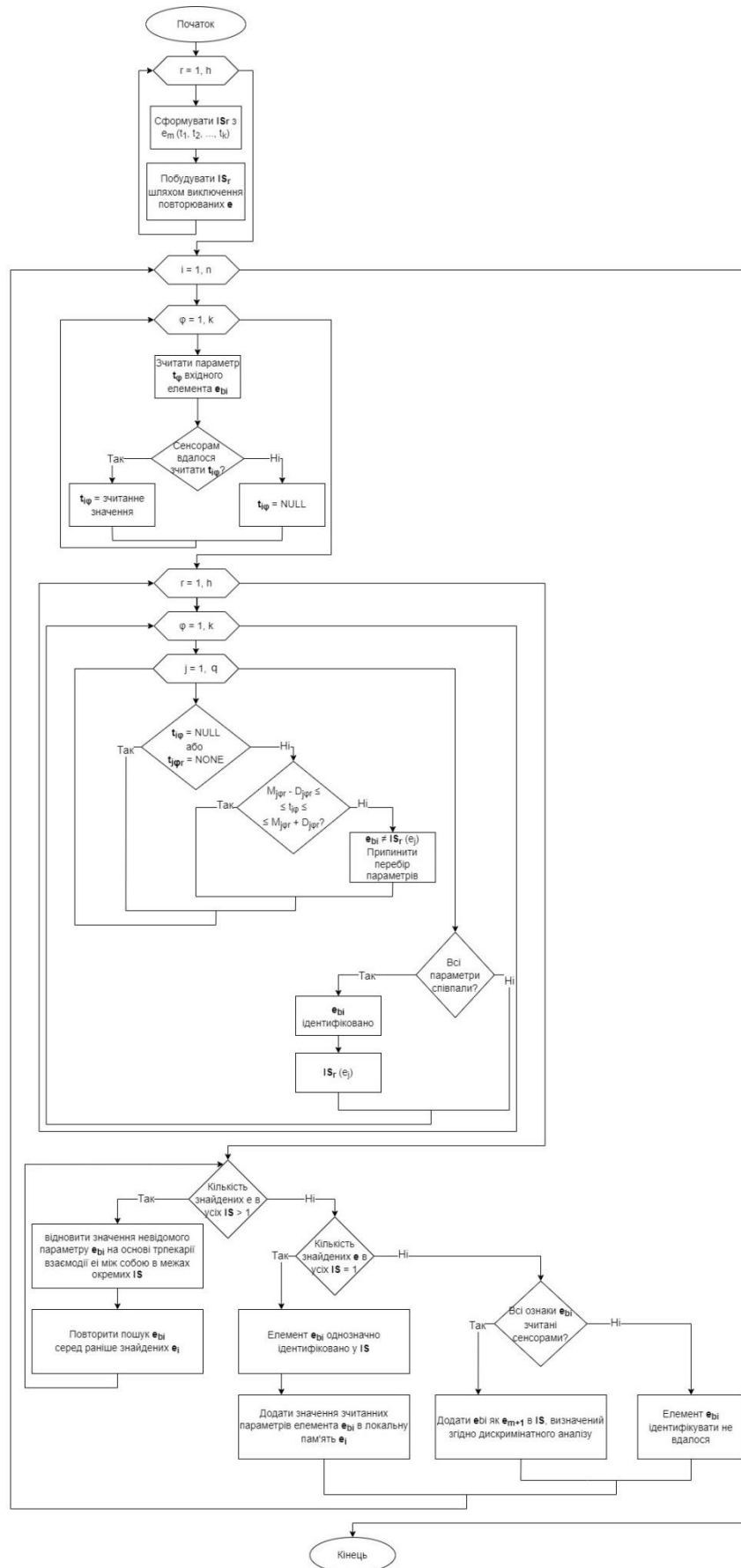


Рис. 1. Схема методу паралельної ідентифікації вхідного елемента

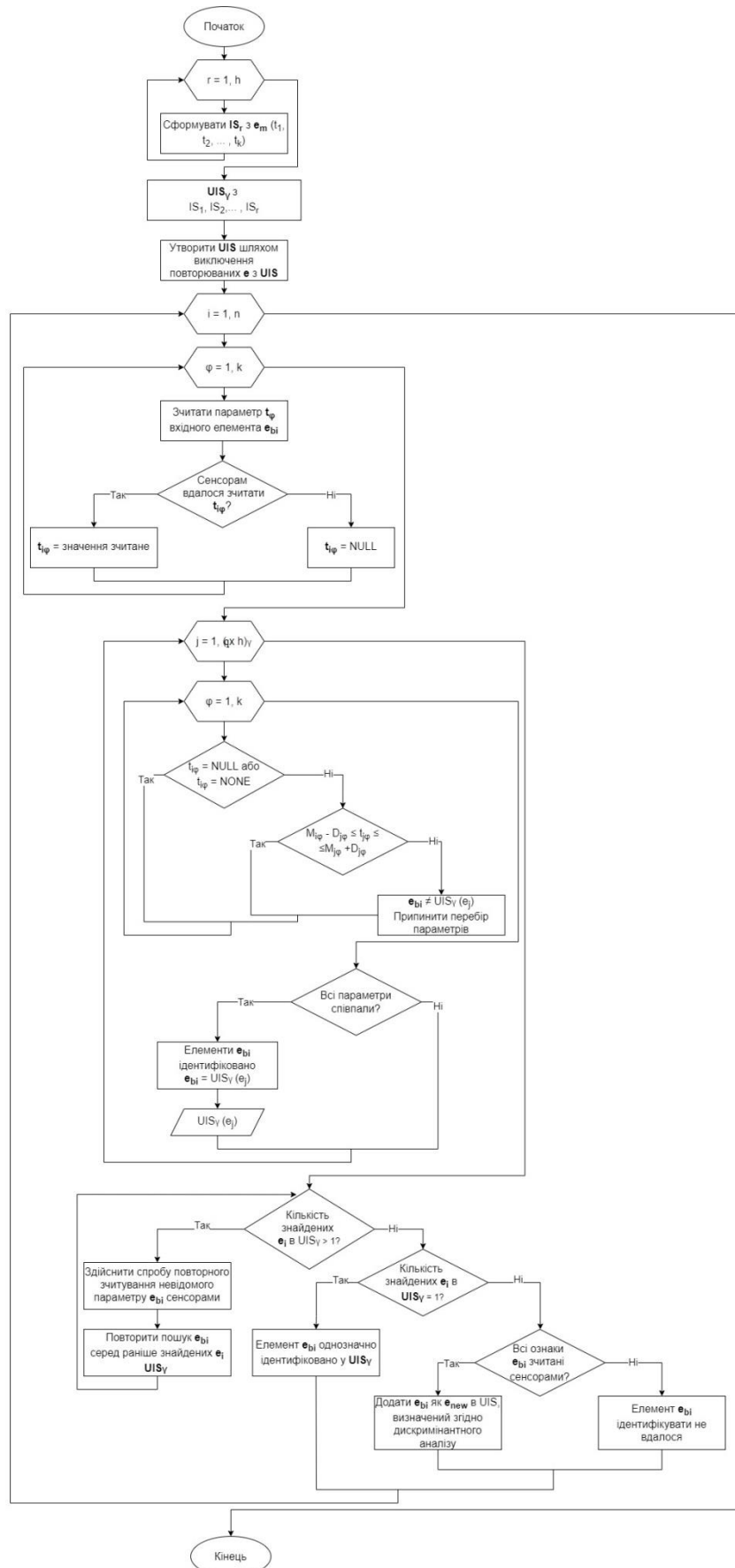


Рис. 2. Схема методу ідентифікації вхідного елемента за допомогою єдиного інформаційного простору

Для параметру t_{u_j} можливі значення $\{t_{u_{j_1}}, \dots, t_{u_{j_k}}\}$, де k – кількість допустимих значень параметру деякого елемента.

Якщо для e_i задано хоча б одне значення параметра t_{u_j} , то множина заданих значень задається: $\{t_{u_{j_1}}(e_i), \dots, t_{u_{j_k}}(e_i)\}$, при цьому очевидно, що $\{t_{u_{j_1}}(e_i), \dots, t_{u_{j_k}}(e_i)\} \subset \{t_{u_{j_1}}, \dots, t_{u_{j_k}}\}$.

Якщо для e_i не задано значення параметра t_{u_j} , то множина наборів значень для параметрів визначається: $\{t_{u_{j_1}}(e_i), \dots, t_{u_{j_n}}(e_i)\}$, $t_u \neq \emptyset$.

Таким чином, формула пошуку елемента за його параметрами має вигляд:

$$FT(e_i) = \left\{ \begin{array}{l} \{t_{u_{j_1}}(e_i), \dots, t_{u_{j_k}}(e_i)\} \subset T \\ \{t_{u_{j_1}}(e_i), \dots, t_{u_{j_n}}(e_i)\} \subset \{t_{u_{j_1}}, \dots, t_{u_{j_k}}\}, t_u(e_i) \neq \emptyset \end{array} \right\}$$

В процесі покроково задаються значення підмножини параметрів для побудови формули пошуку за параметрами, таким чином формується запит з метою отримання шуканої вибірки інформаційних об'єктів.

Висновки. Розроблено метод ідентифікації вхідного елемента на основі об'єднання інформаційних просторів в єдиний інформаційний простір для подальшого пошуку в ньому цього ж елемента. Запропоновано формулу пошуку елемента за параметрами на основі теоретико-множинного формалізму.

Список використаних джерел

1. *Berson A.* Master Data Management and Customer Data Integration for a Global Enterprise / Berson A., Dubov L. // McGraw-Hill, 2007.
2. *Maciaszek L.A.* Managing Complexity of Enterprise Information Systems. In: Seruca I., Cordeiro J., Hammoudi S., Filipe J. (eds). Enterprise Information Systems VI. Springer, Dordrecht. 2006. Pp. 30-36.
3. *Vlasyk H.* The Method of Solving Problems of Optimal Restoration of Telecommunication Signals / Vlasyk H., Zamrii I., Shkapa V., Laptiev S., Kalyniuk A., Laptieva T. // 2021 IEEE 3th International Conference on Advanced Trends in Information Theory (ATIT), December 15-17, 2021, pp. 71-75.
4. *Собчук В.В.* Методика створення єдиного інформаційного простору на виробничому підприємстві з функціонально стійким виробничим процесом / Собчук В.В. // Наукове періодичне видання «Системи управління, навігації та зв'язку», 2019, Вип. 6 (58), с. 84-91.
5. *Sobchuk V.* Strategies for Control Automated Production Centers to Ensure the Functional Stability of Enterprise Information Systems / Sobchuk V., Zamrii I., Vlasyk H., Tsvietkova Yu. // 2021 IEEE 3th International Conference on Advanced Trends in Information Theory (ATIT), 2021, pp. 61-66.
6. *Завгородний В.В.* Единое информационное пространство производственных предприятий на основе связанных данных / Завгородний В.В., Щербак С.С. // Системи обробки інформації, 2013, № 2, с. 275–278.
7. *Schmidt A.* A Method for the Identification and Definition of Information Objects / Schmidt A., Otto B. // In M. Pamela Neely, Leo Pipino, John P. Slone, editors, Proceedings of the 13th International Conference on Information Quality, MIT, Cambridge, MA, USA, 2008, pp. 214-228.
8. *Doerr M.* Information Carriers and Identification of Information Objects: An Ontological Approach / Doerr M., Tzitzikas Ya. // Computer Science, 2012, pp. 1-30.
9. *Neiling M.* Identification of Real-world Objects in Multiple Databases. In: Spiliopoulou, M., Kruse, R., Borgelt, C., Nürnberger, A., Gaul, W. (eds) From Data and Information Analysis to Knowledge Engineering. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg, 2006.
10. *Senjian A.* Efficient algorithms for subwindow search in object detection and localization / Senjian A., Peursum P., Wanquan L., Venkatesh S. // IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami, Florida, USA, June 20– 25, 2009: proceeding, IEEE, Washington, D. C, USA, pp. 264– 271.
11. *Москаленко В.В.* Інформаційно-екстремальне машинне навчання системи ідентифікації об'єктів на місцевості / Москаленко В.В., Коробов А.Г. // Журнал інженерних наук, 2016, Вип. 3, №. 1, с. 1-7.

12. Thome G. Grundlagen und Modelle des Information Lifecycle Management / Thome G., Sollbach W. // Springer. Berlin, 2007.
13. Dodonov A. Method of searching for information objects in unified information space / Dodonov A., Mukhin V., Zavgorodnii V., Kornaga Ya., Zavgorodnya A. // System research and information technologies, 2021, № 1, pp. 34–46.
14. Gershkovich M.M. The tasks of identification of informational objects in area-spread data arrays / Gershkovich M.M., Biryukova T.K. // Systems and Means of Informatics, 2014, № 24(1), pp. 224–243.
7. SCHMIDT A., OTTO B. A Method for the Identification and Definition of Information Objects // In M. Pamela Neely, Leo Pipino, John P. Slone, editors, Proceedings of the 13th International Conference on Information Quality, MIT, Cambridge, MA, USA, 2008, pp. 214–228.
8. DOERR M., TZITZIKAS Ya. Information Carriers and Identification of Information Objects: An Ontological Approach // Computer Science, 2012, pp. 1–30.
9. NEILING M. Identification of Real-world Objects in Multiple Databases. In: Spiliopoulou, M., Kruse, R., Borgelt, C., Nürnberger, A., Gaul, W. (eds) From Data and Information Analysis to Knowledge Engineering. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg, 2006.
10. SENJIAN A., PEURSUM P., WANQUAN L., VENKATESH S. Efficient algorithms for subwindow search in object detection and localization // IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami, Florida, USA, June 20–25, 2009: proceeding, IEEE, Washington, D. C, USA, pp. 264– 271.
11. MOSKALENKO V. V., KOROBV A. G. Informacijno-ekstremal'ne mashynne navchannja systemy identyfikacii' ob'ektiv na miscevosti // Zhurnal inzhenernyh nauk, 2016, Vol. 3, No. 1, pp. 1–7.

References

1. BERSON A., DUBOV L. Master Data Management and Customer Data Integration for a Global Enterprise // McGraw-Hill, 2007.
2. MACIASZEK L.A. Managing Complexity of Enterprise Information Systems. In: Seruca I., Cordeiro J., Hammoudi S., Filipe J. (eds). Enterprise Information Systems VI. Springer, Dordrecht. 2006. Pp. 30–36.
3. VLASYK H., ZAMRII I., SHKAPA V., LAPTIEV S., KALYNIUK A., LAPTIEVA T. The Method of Solving Problems of Optimal Restoration of Telecommunication Signals // 2021 IEEE 3th International Conference on Advanced Trends in Information Theory (ATIT), December 15–17, 2021, pp. 71–75.
4. SOBCHUK V.V. Metodyka stvorennia yedynoho informatsiinoho prostoru na vyrobnychomu pidpriemstvi z funktsionalno stiikym vyrobnychym protsesom // Naukove periodychne vydannia «Systemy upravlinnia, navihatsii ta zviazku», 2019, Vol. 6 (58), pp. 84–91.
5. SOBCHUK V., ZAMRII I., VLASYK H., TSVIETKOVA Yu. Strategies for Control Automated Production Centers to Ensure the Functional Stability of Enterprise Information Systems // 2021 IEEE 3th International Conference on Advanced Trends in Information Theory (ATIT), 2021, pp. 61–66.
6. ZAVGORODNIY V.V., SHCHERBAK S.S. Edinoe informatsionnoye prostranstvo proizvodstvennykh predpriyatij na osnove svyazannykh dannyakh // Systemy obrobky informatsii, 2013, № 2, pp. 275–278.

Надійшла до редколегії