

# Облікова картка дисертації

## I. Загальні відомості

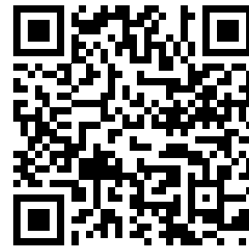
Державний обліковий номер: 0826U001174

Особливі позначки: відкрита

Дата реєстрації: 25-04-2026

Статус: Запланована

Реквізити наказу МОН / наказу закладу:



## II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Кириченко Євген Олександрович

2. Yevhen O. Kurychenko

Кваліфікація:

Ідентифікатор ORCID ID: 0009-0005-6150-5410

Вид дисертації: доктор філософії

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 121

Назва наукової спеціальності: Інженерія програмного забезпечення

Галузь / галузі знань:

Освітньо-наукова програма зі спеціальності: Інженерія програмного забезпечення

Дата захисту:

Спеціальність за освітою: 122 Комп'ютерні науки

Місце роботи здобувача: Чернівецький національний університет імені Юрія Федьковича

Код за ЄДРПОУ: 02071240

Місцезнаходження: вул. Коцюбинського, Чернівці, 58012, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

### **III. Відомості про організацію, де відбувся захист**

**Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради):** PhD 12668

**Повне найменування юридичної особи:** Чернівецький національний університет імені Юрія Федьковича

**Код за ЄДРПОУ:** 02071240

**Місцезнаходження:** вул. Коцюбинського, Чернівці, 58012, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

### **IV. Відомості про підприємство, установу, організацію, в якій було виконано дисертацію**

**Повне найменування юридичної особи:** Чернівецький національний університет імені Юрія Федьковича

**Код за ЄДРПОУ:** 02071240

**Місцезнаходження:** вул. Коцюбинського, Чернівці, 58012, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

### **V. Відомості про дисертацію**

**Мова дисертації:** Українська

**Коди тематичних рубрик:** 20.54.07

**Тема дисертації:**

1. Оптимізація структури гетерогенних даних в Big Data
2. Optimization of the Structure of Heterogeneous Data in Big Data

**Реферат:**

1. У дисертаційній роботі здійснено аналіз існуючих підходів до порівняння табличних наборів даних, виявлено їх обмеження щодо роботи з гетерогенними даними. Розроблено підхід створення компактних представлень даних для різних типів змінних, запропоновано зважену метрику структурної подібності даних, а також створено алгоритми побудови графів та дерев подібності для організації корпусів даних. Розроблено хмарну масштабовану архітектуру на базі AWS для обробки великих корпусів табличних даних та проведено експериментальні дослідження ефективності запропонованого підходу. Об'єктом дослідження є процес порівняння гетерогенних табличних наборів даних з слабо документованих джерел. Предмет дослідження є моделі та методи класифікації змінних, створення компактних представлень гетерогенних табличних наборів даних, побудова графів подібності для ефективного порівняння та структурування корпусів даних за мінімальних припущень щодо схеми. Метою дослідження є розробка уніфікованої типо-орієнтованої

інформаційної системи для ефективного порівняння гетерогенних табличних наборів даних з невідомих або слабо документованих джерел за допомогою створення компактних представлень даних з використанням зваженої метрики структурної подібності інформації даних та побудова масштабованої хмарної інформаційної системи для формування графів та дерев подібності корпусів. Наукова новизна одержаних результатів наступна: – вперше розроблено уніфіковану методологію конструювання компактних представлень даних для гетерогенних табличних змінних; – запропоновано нову метрику структурної подібності даних, що використовують одиничні метрики або векторні представлення для забезпечення принципово обґрунтованого порівняння характеристик різнотипних змінних; – розроблено новий метод побудови ієрархічних структур подібності корпусів табличних даних на основі графів суміжності; – розроблено та реалізовано наскрізну хмарну масштабовану інформаційну технологію обробки великих корпусів табличних даних; – модифіковано підхід до кластеризації гетерогенних даних на основі компактних представлень та матриці суміжності. Практичне значення отриманих результатів. Розроблено архітектуру інформаційної технології для порівняння гетерогенних табличних наборів даних, процедуру конструювання компактних представлень даних, метрики обчислення подібності та методи побудови структур корпусів даних для використання в умовах інформаційних систем великого масштабу з невідомими джерелами. Експериментальні дослідження продемонстрували зменшення обсягу даних завдяки використанню компактних представлень. Застосування розподіленої архітектури забезпечило прискорення обробки приблизно на 40–60% порівняно з традиційними підходами повного сканування таблиць. Дисертація складається зі вступу, чотирьох розділів, висновків, переліку використаних джерел та чотирьох додатків. У вступі обґрунтовано актуальність теми дослідження, сформульовано мету, завдання, предмет, об'єкт та методи дослідження, вказано наукову новизну, подано та проаналізовано зв'язок роботи з науковими темами. Зазначено особистий внесок здобувача, а також наведено відомості про апробацію та публікації основних результатів дисертації. Перший розділ дисертації містить теоретичні засади аналізу великих та гетерогенних даних, обґрунтовує застосування спектральних методів та апарату випадкових матриць для дослідження їхньої структури. У другому розділі запропоновано узагальнену модель компактного представлення (CDR), що замінює повні набори даних їх стислими дескрипторами, які зберігають інформаційно значущі статистичні та структурні характеристики. Розроблено алгоритм автоматичної типізації змінних. Побудовано та обґрунтовано конвеєрну систему аналізу даних. У третьому розділі здійснено аналіз сучасних хмарних технологій та принципів побудови масштабованих обчислювальних систем. Розроблено багаторівневу модульну архітектуру інформаційної системи. У четвертому розділі проведено експериментальну перевірку інформаційної технології моделювання та структурного аналізу гетерогенних табличних даних, що базується на методі компактного представлення CDR та матриці структурної подібності DISS. У висновках підсумовано основні результати дисертаційного дослідження. У додатках подано наукові публікації, відомості про апробацію результатів дисертації, лістинг частини коду програмного забезпечення. Запропоновані підходи використовуються у роботі ТОВ «Кодерс ПРО» та ТОВ «Палетний сервіс». Результати теоретичних та практичних досліджень використовуються у навчальному процесі кафедр математичних проблем управління і кібернетики та програмного забезпечення комп'ютерних систем Чернівецького національного університету імені Юрія Федьковича. Ключові слова: набори даних, модель, моделювання, класифікація, точність, машинне навчання, аналіз даних, штучний інтелект, інформаційні технології / системи, програмне забезпечення, програмна система, нейронні мережі, Python, хмарна платформа, кластерний аналіз.

2. The dissertation analyzes existing approaches to comparing tabular datasets and identifies their limitations when applied to heterogeneous data. An approach for constructing compact data representations for different types of variables has been developed, a weighted metric for structural data similarity has been proposed, and algorithms for building similarity graphs and trees to organize data corpora have been developed. In addition, a scalable cloud-based architecture built on AWS has been developed for processing large tabular data corpora, and experimental studies have been conducted to evaluate the effectiveness of the proposed approach. The research objective is to compare heterogeneous tabular datasets obtained from poorly documented sources. The research

subject comprises models and methods for variable classification, the construction of compact representations of heterogeneous tabular datasets, and the construction of similarity graphs for efficient comparison and structuring of data corpora under minimal assumptions about the schema. The research aims to develop a unified type-oriented information system for comparing heterogeneous tabular datasets from unknown or poorly documented sources by constructing compact data representations using a weighted metric of structural data similarity, and to build a scalable cloud information system for generating graphs and trees of corpus similarity. The scientific novelty is as follows: – for the first time, a unified methodology for constructing compact data representations for heterogeneous tabular variables has been developed; – a new metric of structural data similarity has been proposed, combining individual metrics and vector representations to provide a principled comparison of variables of different types; – a new method has been developed for constructing hierarchical similarity structures of tabular data corpora based on adjacency graphs; – An end-to-end scalable cloud information technology for processing large corpora of tabular data has been developed and implemented. – The approach to clustering heterogeneous data based on compact representations and an adjacency matrix has been improved. The practical significance of the obtained results lies in the development of an information technology architecture for comparing heterogeneous tabular datasets, a procedure for constructing compact data representations, similarity computation metrics, and methods for building corpus structures for use in large-scale information systems with unknown data sources. Experimental studies demonstrated a reduction in data volume by using compact representations. The application of a distributed architecture resulted in 40–60% faster processing than traditional scanning. The dissertation consists of an introduction, four chapters, a conclusion, references, and four appendices. The introduction substantiates the relevance of the research topic, formulates the aim, objectives, subject, and object, identifies the scientific novelty, and analyzes the relationship of the work to broader scientific themes. It also outlines the author's contribution and provides information on the dissertation's approval and publications of its results. The first chapter presents the theoretical foundations for analyzing heterogeneous data and substantiates the use of spectral methods and random matrix theory to study their structure. The second chapter proposes a generalized compact representation model (CDR), which replaces complete datasets with compressed descriptors that preserve statistically and structurally significant information. An algorithm for automatic variable typing has been developed. A pipeline-based data analysis system has also been designed and substantiated. The third chapter analyzes modern cloud technologies and the principles of building scalable computing systems. A modular architecture of the information system has been developed. The fourth chapter presents an experimental validation of the information technology for modeling and structural analysis of heterogeneous tabular data based on the CDR compact representation method and the DISS structural similarity metric. The conclusions summarize the dissertation's main results. The appendices contain scientific publications, information on the approval of the dissertation results, and a listing of part of the software code. The proposed approaches are used in Coders PRO LLC and Paletnyi Servis LLC. The results of the theoretical and practical research are also used in the educational process of the Departments of Mathematical Problems of Control and Cybernetics and Software of Computer Systems at Yuriy Fedkovych Chernivtsi National University. Keywords: data sets, model, modeling, classification, accuracy, machine learning, data analysis, artificial intelligence, information technology / system, software, software system, neural networks, Python, cloud platform, cluster analysis.

**Державний реєстраційний номер ДіР:**

**Пріоритетний напрям розвитку науки і техніки:** Інформаційні та комунікаційні технології

**Стратегічний пріоритетний напрям інноваційної діяльності:** Розвиток сучасних інформаційних, комунікаційних технологій, робототехніки

**Підсумки дослідження:** Теоретичне узагальнення і вирішення важливої наукової проблеми

**Публікації:**

- Кириченко О., Кириченко Є. Асимптотичні властивості випадкових матриць = Asymptotic Properties of Randommatrices. Вісник Київського національного університету імені Тараса Шевченка. Фізико-математичні науки. 2024. 78(1). С. 41–44. (Scopus)
- Кириченко Є. О., Кириченко О. Л. Побудова моделей нейронних мереж у хмарному середовищі засобами AWS EMR. Вчені записки Таврійського національного університету імені В.І. Вернадського. Серія : Технічні науки. 2024. 35(74), №6. С. 75-79.
- Kyrychenko Y., Malyk I. Information Technology for Data Compression and Transformation by Means of Apache Airflow and Amazon EMR. Security of Infocommunication Systems and Internet of Things. 2025. Vol. 3, no. 1. P. 01004.
- Кириченко Є., Малик, І. Один результат кластеризації на основі випадкових матриць. Information Technology: Computer Science, Software Engineering and Cyber Security. 2025. №3. С. 257-270.
- Malyk I. V., Kyrychenko Y., Gorbatenko M., Lukashiv T. Data Optimization through Compression Methods Using Information Technology. International Journal of Information Technology and Computer Science. 2025. Vol.17, No.5. P. 84-99. (Scopus)

**Наукова (науково-технічна) продукція:** методи, теорії, гіпотези; програмні продукти, програмно-технологічна документація

**Соціально-економічна спрямованість:**

**Охоронні документи на ОПІВ:**

**Впровадження результатів дисертації:** Впроваджено

**Зв'язок з науковими темами:** 0121U109232, 0121U109159

## **VI. Відомості про наукового керівника/керівників (консультанта)**

**Власне Прізвище Ім'я По-батькові:**

1. Малик Ігор Володимирович
2. Ihor V. Malyk

**Кваліфікація:** д. ф.-м. н., професор, 01.05.01

**Ідентифікатор ORCID ID:** 0000-0002-1291-9167

**Додаткова інформація:**

**Повне найменування юридичної особи:** Чернівецький національний університет імені Юрія Федьковича

**Код за ЄДРПОУ:** 02071240

**Місцезнаходження:** вул. Коцюбинського, Чернівці, 58012, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

## **VII. Відомості про офіційних опонентів та рецензентів**

**Офіційні опоненти**

**Власне Прізвище Ім'я По-батькові:**

1. Виклюк Ярослав Ігорович
2. Yaroslav I. Vyklyuk

**Кваліфікація:** д.т.н., професор, 01.05.02**Ідентифікатор ORCID ID:** 0000-0003-4766-4659**Додаткова інформація:****Повне найменування юридичної особи:** Національний університет "Львівська політехніка"**Код за ЄДРПОУ:** 02071010**Місцезнаходження:** вул. Степана Бандери, Львів, 79013, Україна**Форма власності:** Державна**Сфера управління:** Міністерство освіти і науки України**Ідентифікатор ROR:****Власне Прізвище Ім'я По-батькові:**

1. Криворучко Олена Володимирівна
2. Olena V. Kryvoruchko

**Кваліфікація:** д. т. н., професор, 05.13.22**Ідентифікатор ORCID ID:** 0000-0002-7661-9227**Додаткова інформація:****Повне найменування юридичної особи:** Національний університет біоресурсів і природокористування України**Код за ЄДРПОУ:** 00493706**Місцезнаходження:** вул. Героїв Оборони, Київ, 03041, Україна**Форма власності:** Державна**Сфера управління:** Міністерство освіти і науки України**Ідентифікатор ROR:****Рецензенти****Власне Прізвище Ім'я По-батькові:**

1. Баловсяк Сергій Васильович
2. Sergiy V. Balovsyak

**Кваліфікація:** д. т. н., доц., 05.13.05**Ідентифікатор ORCID ID:** Не застосовується**Додаткова інформація:****Повне найменування юридичної особи:** Чернівецький національний університет імені Юрія Федьковича

**Код за ЄДРПОУ:** 02071240

**Місцезнаходження:** вул. Коцюбинського, Чернівці, 58012, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

**Власне Прізвище Ім'я По-батькові:**

1. Угрин Дмитро Ілліч

2. Dmytro Uhryn

**Кваліфікація:** д. т. н., професор, 05.13.06

**Ідентифікатор ORCID ID:** 0000-0003-4858-4511

**Додаткова інформація:**

**Повне найменування юридичної особи:** Чернівецький національний університет імені Юрія Федьковича

**Код за ЄДРПОУ:** 02071240

**Місцезнаходження:** вул. Коцюбинського, Чернівці, 58012, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

## VIII. Заключні відомості

**Власне Прізвище Ім'я По-батькові  
голови ради**

Ушенко Юрій Олександрович

**Власне Прізвище Ім'я По-батькові  
головуючого на засіданні**

Ушенко Юрій Олександрович

**Відповідальний за підготовку  
облікових документів**

Кириченко Євген Олександрович

**Реєстратор**

Юрченко Тетяна Анатоліївна

**Керівник відділу УкрІНТЕІ, що є  
відповідальним за реєстрацію наукової  
діяльності**



Юрченко Тетяна Анатоліївна