

Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0826U001064

Особливі позначки: відкрита

Дата реєстрації: 16-04-2026

Статус: Запланована

Реквізити наказу МОН / наказу закладу:



II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Власюк Євгеній Романович

2. Yevhenii R. Vlasiuk

Кваліфікація:

Ідентифікатор ORCID ID: 0009-0007-9036-1510

Вид дисертації: доктор філософії

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 126

Назва наукової спеціальності: Інформаційні системи та технології

Галузь / галузі знань:

Освітньо-наукова програма зі спеціальності: Інформаційні системи та технології

Дата захисту:

Спеціальність за освітою: Комп'ютеризовані системи управління та автоматика

Місце роботи здобувача:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

III. Відомості про організацію, де відбувся захист

Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради): PhD 12987

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

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

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

V. Відомості про дисертацію

Мова дисертації: Українська

Коди тематичних рубрик: 20.23.17, 50.41.21, 20.54.03

Тема дисертації:

1. Методи ідентифікації дублювання даних у галузево-розподілених інформаційних системах на основі онтологій та аналізу структурних залежностей
2. Data quality methods for identifying data duplication based on ontologies and structural dependency analysis in Data Mesh platforms

Реферат:

1. Дисертаційна робота присвячена вирішенню проблеми ідентифікації дублювання даних в галузево-розподілених інформаційних системах. Галузево-розподілені інформаційні системи оброблення та аналізу даних застосовуються компаніями за потреби оброблення та аналізу даних, які мають різну природу, походження, відносяться до різних сфер та галузей. Для оброблення, аналізу та оцінки якості, різних за своєю суттю, даних, потрібні спеціалізовані правила і політики, які враховують специфіку даних. Необхідність в поєднанні доменних правил оцінки та оброблення даних з глобальними політиками, що мають виконуватись незалежно від специфіки даних, забезпечується федеративною моделлю управління галузево-розподіленими системами. Існуючі методи оцінки якості даних і зокрема, ідентифікації

дублювання даних, потребують адаптації і оптимізації до федеративної моделі, яка застосовується в галузево-розподілених інформаційних системах оброблення та аналізу даних. Для забезпечення підвищення ефективності ідентифікації дублювання даних, в роботі запропоновано метод ідентифікації дублювання об'єктів даних з використанням онтологій. Створено математичну модель методу з використанням описових, структурних атрибутів онтології, а також методу з комбінуванням описових і структурних атрибутів. Для комплексного вирішення питання ідентифікації дублювання даних в галузево-розподілених системах оброблення та аналізу даних розглянуто і вирішено проблему ідентифікації дублювання продуктів даних. На відміну від оригінальних даних, отриманих з систем джерел даних, продукти даних представляють собою набори даних, що створені шляхом трансформації даних в рамках галузево-розподілених систем оброблення даних. Саме продукти даних використовуються кінцевими користувачами, а тому недостатній рівень якості цих даних безпосередньо вплине на сприйняття даних і правильність прийняття бізнес-рішень. Для вирішення задачі виявлення дублювання продуктів даних запропоновано метод ідентифікації дублювання продуктів даних на основі аналізу структурних залежностей. Розроблено математичну модель цього методу, яка базується на математичному апараті теорії графів. Процес виявлення дублювання продуктів даних складається з етапу побудови графів залежностей продуктів даних і етапу порівняння графів залежностей продуктів даних. Проаналізовано методи збору метаданих для побудови графів залежностей продуктів даних, зокрема метод на основі журналів повідомлень. Виявлений недолік даного методу, який полягає в низькій точності побудови графів залежностей при використанні недостатнього проміжку часу для аналізу журналів повідомлень. Для вирішення виявленого недоліку запропоновано метод побудови графів залежностей продуктів даних на основі синтаксичних дерев коду. Запропонований метод відрізняється від існуючого побудовою синтаксичних дерев коду на основі вихідного програмного коду застосунків оброблення даних і ідентифікацією залежних об'єктів зберігання даних, що дає змогу не використовувати журнал повідомлень і уникнути проблеми низької точності графів залежностей при використанні недостатньої кількості записів журналів повідомлень. Експериментальні дослідження запропонованого методу побудови графів залежностей на основі синтаксичних дерев коду демонструють точність побудови графів залежностей на рівні 90% з можливістю подальшого збільшення за рахунок покращення програмного модуля аналізатора коду. Для оцінки дублювання продуктів даних, проаналізовано існуючі методи порівняння направлених ациклічних графів, як представляють залежності між застосунками оброблення даних і об'єктами зберігання даних. Виявлено недолік існуючого методу, що полягає в погіршенні точності порівняння графів за умови великої кількості залежних об'єктів, що мають однакові імена, але є різними сутностями. Для покращення точності порівняння направлених ациклічних графів продуктів даних в середовищі галузево-розподілених систем оброблення та аналізу даних, де існує велика кількість об'єктів даних з однаковими іменами, які представляють різні сутності, запропоновано комбінований метод порівняння направлених ациклічних графів залежностей продуктів даних. Розроблений метод відрізняється від існуючого інтеграцією з методом ідентифікації дублювання об'єктів даних, а саме використанням списку дубльованих об'єктів даних при порівнянні відповідних вершин графів залежностей продуктів даних, а також врахуванням рівня вершини. Це дозволяє досягти на 6% кращої ефективності в порівнянні з існуючим методом порівняння направлених ациклічних графів залежностей продуктів даних. Розроблені методи ідентифікації дублювання об'єктів даних і продуктів даних використано при розробленні інформаційної технології оцінки якості даних в галузево-розподілених системах оброблення та аналізу даних.

2. The dissertation work is devoted to solving the problem of identifying data duplication in domain-oriented Data Mesh platforms. Domain-oriented Data Mesh platforms for data processing and analysis are used by organizations for the needs of processing and analysis of data which have a different nature, origin, and belong to different domains and industries. Existing methods for assessing data quality and, in particular, identifying data duplication, require adaptation and optimization to the federated model used in domain-oriented Data Mesh platforms. To ensure an increase in the efficiency of data duplication identification, the paper proposes a method for identifying data object duplication using ontologies. A mathematical model of the method using descriptive, structural attributes of the ontology, as well as a method combining descriptive and structural attributes, is created. An

ontology model for the information and service business industry is developed. To assess the quality of the ontology, the created ontology is compared with the existing open-source ontology. The qualitative characteristics of the ontology are sufficient for conducting an experimental study of the method for identifying data object duplication using ontology. For a comprehensive solution to the issue of identifying data duplication in domain-oriented Data Mesh platforms, the problem of identifying data product duplication has been considered and solved. Unlike the original raw data obtained from data source systems, data products are data sets created by transforming data within domain-oriented data processing systems. It is data products that are used by end users, and therefore the insufficient level of quality of these data will directly affect the perception of data and the correctness of business decisions. To solve the problem of detecting data product duplication, a method for identifying data product duplication based on the analysis of structural dependencies has been proposed. A mathematical model of this method has been developed, which is based on the mathematical apparatus of graph theory. The process of detecting data product duplication consists of the stage of constructing dependency graphs of data products and the stage of comparing dependency graphs of data products. Methods of collecting metadata for constructing dependency graphs of data products have been analyzed, in particular, a method based on audit logs. A drawback of this method has been identified, which resides in the low accuracy of constructing dependency graphs when using an insufficient time period for analyzing audit logs. To solve the identified drawback, a method for constructing dependency graphs of data products based on syntax code trees is proposed. The proposed method differs from the existing one by constructing syntax code trees based on the source code of data processing applications and identifying dependent data storage objects. The developed method makes it possible not to use the audit log and avoid the problem of low accuracy of dependency graphs when using an insufficient number of audit log records. Experimental studies of the proposed method for constructing dependency graphs based on syntax code trees demonstrate the accuracy of constructing dependency graphs at the level of 90% with the possibility of further increase by improving the code analyzer software module. To assess the duplication of data products, existing methods for comparing directed acyclic graphs, which represent dependencies between data processing applications and data storage objects, are analyzed. The disadvantage of the existing method is revealed, which consists in worsening the accuracy of graph comparison under the condition of a large number of dependent objects that have the same names, but represent different entities. To improve the accuracy of comparing directed acyclic graphs of data products in the environment of domain-oriented data processing and analysis systems, where there are a large number of data objects with the same names, representing different entities, a combined method for comparing directed acyclic graphs of data product dependencies is proposed. The developed method differs from the existing one by integration with the method for identifying data object duplication, namely, using a list of duplicate data objects when comparing the corresponding vertices of data product dependency graphs, as well as taking into account the vertex level. This allows achieving 6% better efficiency compared to the existing method for comparing directed acyclic graphs of data product dependencies. The developed methods for identifying duplication of data objects and data products were used in the development of information technology for assessing data quality in domain-oriented data processing and analysis Data Mesh platforms.

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: Нове вирішення актуального наукового завдання

Публікації:

- Vlasiuk Y., Onyshchenko V. Data Mesh as Distributed Data Platform for Large Enterprise Companies /Hu, Z., Dychka, I., He, M. (eds) // Advances in Computer Science for Engineering and Education VI. – 2023 – №181. –

С. 183–192

- Власюк Є.Р., Онищенко В.В. Автоматизований спосіб визначення користувачів продуктів даних в розподіленій системі сіток даних /Є.Р. Власюк, В.В. Онищенко// Адаптивні системи автоматичного управління. – 2023 – №43. – С. 113–123
- Онищенко В.В., Власюк Є.Р. Аналіз зміни процесу операціоналізації життєвого циклу моделей машинного навчання в розподілених системах сіток даних /В.В. Онищенко, Є.Р. Власюк // Телекомунікаційні та інформаційні технології. – 2024 – №2(83). – С. 143–149
- Онищенко В.В., Власюк Є.Р. (2025) Процес трансформації інформаційної системи обробки та аналізу даних від монолітної до децентралізованої та розподіленої. XIII Міжнародна науково-практична конференція з інформаційних систем та технологій. 21 - 22 травня 2025р.

Наукова (науково-технічна) продукція: технології; програмні продукти, програмно-технологічна документація

Соціально-економічна спрямованість: підвищення продуктивності праці; забезпечення промисловості чи населення новим видом інформаційно-комунікаційних послуг

Охоронні документи на ОПВ:

Впровадження результатів дисертації: Впроваджено

Зв'язок з науковими темами: 0121U110810

VI. Відомості про наукового керівника/керівників (консультанта)

Власне Прізвище Ім'я По-батькові:

1. Ролік Олександр Іванович
2. Oleksandr I. Rolik

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

Ідентифікатор ORCID ID: 0000-0001-8829-4645

Додаткова інформація:

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

VII. Відомості про офіційних опонентів та рецензентів

Офіційні опоненти

Власне Прізвище Ім'я По-батькові:

1. Любченко Віра Вікторівна

2. Vira V. Liubchenko

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

Ідентифікатор ORCID ID: 0000-0002-4611-7832

Додаткова інформація:

;https://scholar.google.com/citations?user=igCycUUAAAAJ;https://www.scopus.com/authid/detail.uri?authorId=56667638800;https://orcid.org/0000-0002-4611-7832

Повне найменування юридичної особи: Національний університет "Одеська політехніка"

Код за ЄДРПОУ: 43861328

Місцезнаходження: пр. Шевченка, Одеса, 65044, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Власне Прізвище Ім'я По-батькові:

1. Циганок Віталій Володимирович

2. Vitaliy V. Tsyganok

Кваліфікація: д.т.н., професор, 01.05.04

Ідентифікатор ORCID ID: 0000-0002-0821-4877

Додаткова інформація:

Повне найменування юридичної особи: Інститут проблем реєстрації інформації Національної академії наук України

Код за ЄДРПОУ: 03771755

Місцезнаходження: вул. М. Шпака, Київ, 03113, Україна

Форма власності: Державна

Сфера управління: Національна академія наук України

Ідентифікатор ROR:

Рецензенти

Власне Прізвище Ім'я По-батькові:

1. Ульяницька Ксенія Олександрівна

2. Ksenia O. Ulianitska

Кваліфікація: к. т. н., доц., 05.13.06

Ідентифікатор ORCID ID: 0000-0003-0240-6250

Додаткова інформація:

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

Власне Прізвище Ім'я По-батькові:

1. Гавриленко Олена Валеріївна

2. Olena V. Gavrilenko

Кваліфікація: к. ф.-м. н., доц., 01.02.05

Ідентифікатор ORCID ID: 0000-0003-0413-6274

Додаткова інформація:

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

VIII. Заключні відомості

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

Теленик Сергій Федорович

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

Теленик Сергій Федорович

**Відповідальний за підготовку
облікових документів**

Власюк Євгеній Романович

Реєстратор

Юрченко Тетяна Анатоліївна

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



Юрченко Тетяна Анатоліївна