

# Облікова картка дисертації

## I. Загальні відомості

**Державний обліковий номер:** 0823U100106

**Особливі позначки:** відкрита

**Дата реєстрації:** 21-02-2023

**Статус:** Захищена

**Реквізити наказу МОН / наказу закладу:**



## II. Відомості про здобувача

**Власне Прізвище Ім'я По-батькові:**

1. Юсин Яків Олексійович

2. Yusyn Yakiv Oleksiiovich

**Кваліфікація:**

**Ідентифікатор ORCID ID:** Не застосовується

**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:** Не застосовується

**Дата захисту:** 20-02-2023

**Спеціальність за освітою:** Інженерія програмного забезпечення

**Місце роботи здобувача:** Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

**Код за ЄДРПОУ:** 02070921

**Місцезнаходження:** проспект Перемоги, буд. 37, м. Київ, 03056, Україна

**Форма власності:**

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

**Ідентифікатор ROR:** Не застосовується

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

**Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради):** ДФ 26.002.06

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

**Код за ЄДРПОУ:** 02070921

**Місцезнаходження:** проспект Перемоги, буд. 37, м. 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 Перемоги, буд. 37, м. Київ, 03056, Україна

**Форма власності:**

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

**Ідентифікатор ROR:** Не застосовується

### **V. Відомості про дисертацію**

**Мова дисертації:**

**Коди тематичних рубрик:** 50.41, 50.41.25

**Тема дисертації:**

1. Методи та програмні засоби метаморфічного тестування програмних систем автоматичної кластеризації природномовних текстових даних
2. Methods and software tools for metamorphic testing of software systems for automatic clustering of natural language text data

**Реферат:**

1. Робота присвячена вирішенню наукової задачі вдосконалення теоретичних (методів) та практичних (програмних засобів) основ тестування програмних систем автоматичної кластеризації природномовних текстових даних. У дисертаційній роботі отримано низку нових наукових результатів, зокрема, уперше розроблено метод метаморфічного тестування програмних систем автоматичної кластеризації природномовних текстових даних МЕЕТС, який, на відміну від існуючих методів, є застосовним до програмних реалізацій будь-яких детермінованих методів кластеризації текстів без вхідного параметру кількості кластерів та забезпечує ефективність мутаційного тестування за показником оцінки мутації у 81-100%. Уперше розроблено метод метаморфічного тестування програмних систем автоматичної кластеризації

природномовних текстових даних MEETC-k, який, на відміну від існуючих методів, є застосовним до програмних реалізацій будь-яких детермінованих методів кластеризації текстів з вхідним параметром кількості кластерів та забезпечує ефективність мутаційного тестування за показником оцінки мутації у 86-100%. Уперше запропоновано узагальнену архітектуру програмного забезпечення для метаморфічного тестування на основі використання моделі безсерверних обчислень, що дозволяє спростити процес розроблення програмних засобів для метаморфічного тестування та досягти прискорення виконання тестування програмних систем кластеризації текстів у 34-50%, в порівнянні з існуючими архітектурами. Уперше розроблено архітектурний шаблон проектування програмного забезпечення «Metamorphic Testing-as-a-Service», характерною рисою якого є декомпозиція метаморфічного зв'язку на окремі складові та кодогенерація тіл метаморфічних зв'язків і безсерверних функцій. Використання запропонованого шаблону дозволяє зменшити дублювання і зв'язність програмного коду при розробленні програмних засобів для безсерверного метаморфічного тестування та покращує стандартні метрики якості коду: кількість рядків виконуваного коду та зв'язність класів в цілому, цикломатичну складність та індекс підтримуваності для окремих компонент. Уперше розроблено сімейство методів генерування корпусів текстових даних CorDeGen, характерною рисою яких є детермінованість та легкість апіорного опису структури отриманого корпусу та які, на відміну від існуючих методів генерування корпусів, приймають на вхід мінімально можливу кількість параметрів, таким чином спрощуючи опис, зберігання та відтворення результатів: 1. базовий метод CorDeGen – на відміну від інших методів сімейства, забезпечує найвищу швидкість генерування корпусів за рахунок використання більш складного апіорного опису структури отриманого корпусу, що викликано видаленням частини генерованих термів методами попереднього оброблення природномовних текстових даних; 2. метод CorDeGen+ – на відміну від інших методів сімейства, уникає видалення частини генерованих термів з корпусу, завдяки чому полегшується апіорний опис структури отриманого корпусу; 3. метод SemCorDeGen – на відміну від інших методів сімейства, генерує корпуси текстів, які можливо використовувати разом з методами оброблення природномовних текстових даних на основі застосування семантичних моделей за рахунок розширення кількості вхідних параметрів. Практичне значення одержаних результатів полягає у спрощенні процесу розроблення програмних засобів для метаморфічного тестування в цілому (за рахунок використання узагальненої безсерверної архітектури та шаблону проектування «MTaaS») та, зокрема, для тестування програмних систем автоматичної кластеризації природномовних текстових даних (за рахунок використання сімейства методів генерування корпусів текстів CorDeGen). Крім того, сімейство методів CorDeGen також може використовуватись на практиці в наукових дослідженнях для підвищення їх відтворюваності. Розроблені програмні засоби опубліковано у відкритому доступі, для їх частини створено та опубліковано загальнодоступні NuGet пакети, які можуть під'єднуватись та використовуватись сторонніми розробниками.

2. Thesis is devoted to solving the scientific problem of improving the theoretical (methods) and practical (software) foundations of testing software systems for automatic clustering of natural language text data. The dissertation provides several new scientific results, in particular, for the first time, the method of metamorphic testing of software systems for automatic clustering of natural language text data MEETC was developed, which, unlike existing methods, is applicable to software implementations of any deterministic methods of text clustering without the input parameter of the number of clusters and ensures the effectiveness of mutation testing according to the mutation score in 81-100%. For the first time, the method of metamorphic testing of software systems for automatic clustering of natural language text data MEETC-k was developed, which, unlike existing methods, is applicable to software implementations of any deterministic text clustering methods with the input parameter of the number of clusters and ensures the effectiveness of mutation testing according to the mutation score in 86-100%. For the first time, the basic software architecture for metamorphic testing based on the use of a serverless computing model is proposed, which allows to simplify the process of developing software tools for metamorphic testing and to achieve an acceleration of the testing of text clustering software systems by 34-50%, compared to existing architectures. For the first time, the architectural software design pattern "Metamorphic Testing-as-a-Service" was developed, the characteristic feature of which is the decomposition of the metamorphic relation into

separate components and the code generation of the bodies of metamorphic relations and serverless functions. The use of the proposed pattern makes it possible to reduce the duplication and coupling of the software code when developing software tools for serverless metamorphic testing and improves the standard metrics of code quality: the number of lines of executable code and the class coupling as a whole, the cyclomatic complexity and the maintainability index for individual components. For the first time, the family of methods for generating corpora of text data CorDeGen has been developed, the characteristic feature of which is the determinism and ease of a priori description of the structure of the received corpus, and which, unlike the existing corpus generation methods, accept as input the minimum possible number of parameters, thus simplifying the description, storage and reproduction of results: 1. the basic CorDeGen method – unlike other methods of the family, provides the highest speed of corpora generation due to the use of a more complex a priori description of the structure of the received corpus, which is caused by the removal of a part of the generated terms by methods of pre-processing of natural language text data; 2. the CorDeGen+ method – unlike other methods of the family, it avoids removing part of the generated terms from the corpus, which facilitates the a priori description of the structure of the resulting corpus; 3. the SemCorDeGen method – unlike other methods of the family, generates corpora of texts that can be used together with natural language text data processing methods based on the application of semantic models due to the expansion of the number of input parameters. The practical significance of the obtained results lies in the simplification of the process of developing software tools for metamorphic testing in general (due to the use of the basic serverless architecture and the "MTaaS" design pattern) and, in particular, for testing software systems for automatic clustering of natural language text data (due to the use of the family of text corpora generation methods CorDeGen). In addition, the CorDeGen family of methods can also be used in practice in scientific research, to increase their reproducibility. The developed software tools are published in open access, for some of them publicly available NuGet packages have been created and published, which can be connected and used by third-party developers.

**Державний реєстраційний номер ДіР:**

**Пріоритетний напрям розвитку науки і техніки:**

**Стратегічний пріоритетний напрям інноваційної діяльності:**

**Підсумки дослідження:**

**Публікації:**

**Наукова (науково-технічна) продукція:**

**Соціально-економічна спрямованість:**

**Охоронні документи на ОПВ:**

**Впровадження результатів дисертації:**

**Зв'язок з науковими темами:**

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

**Власне Прізвище Ім'я По-батькові:**

1. Заболотня Тетяна Миколаївна

2. Zabolotnia Tetiana Mykolaivna

**Кваліфікація:** к. т. н., 05.13.05

**Ідентифікатор ORCID ID:** Не застосовується

**Додаткова інформація:**

**Повне найменування юридичної особи:**

**Код за ЄДРПОУ:**

**Місцезнаходження:**

**Форма власності:**

**Сфера управління:**

**Ідентифікатор ROR:** Не застосовується

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

### **Офіційні опоненти**

**Власне Прізвище Ім'я По-батькові:**

1. Яшина Оксана Миколаївна

2. Yashyna Oksana Mykolaivna

**Кваліфікація:** к. т. н., 05.13.06

**Ідентифікатор ORCID ID:** Не застосовується

**Додаткова інформація:**

**Повне найменування юридичної особи:**

**Код за ЄДРПОУ:**

**Місцезнаходження:**

**Форма власності:**

**Сфера управління:**

**Ідентифікатор ROR:** Не застосовується

**Власне Прізвище Ім'я По-батькові:**

1. Шинкаренко Віктор Іванович

2. Shynkarenko Viktor Ivanovych

**Кваліфікація:** д. т. н., 01.05.03

**Ідентифікатор ORCID ID:** Не застосовується

**Додаткова інформація:**

**Повне найменування юридичної особи:**

**Код за ЄДРПОУ:**

**Місцезнаходження:**

**Форма власності:**

**Сфера управління:**

**Ідентифікатор ROR:** Не застосовується

## **Рецензенти**

### **Власне Прізвище Ім'я По-батькові:**

1. Ролік Олександр Іванович
2. Rolik Oleksandr Ivanovych

**Кваліфікація:** д.т.н., 05.13.06

**Ідентифікатор ORCID ID:** Не застосовується

**Додаткова інформація:**

**Повне найменування юридичної особи:**

**Код за ЄДРПОУ:**

**Місцезнаходження:**

**Форма власності:**

**Сфера управління:**

**Ідентифікатор ROR:** Не застосовується

### **Власне Прізвище Ім'я По-батькові:**

1. Петрашенко Андрій Васильович
2. Petrashenko Andryi Vasylovych

**Кваліфікація:** к. т. н., 01.05.03

**Ідентифікатор ORCID ID:** Не застосовується

**Додаткова інформація:**

**Повне найменування юридичної особи:**

**Код за ЄДРПОУ:**

**Місцезнаходження:**

**Форма власності:**

**Сфера управління:**

**Ідентифікатор ROR:** Не застосовується

## **VIII. Заключні відомості**

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

Жаріков Едуард В'ячеславович

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

Жаріков Едуард В'ячеславович

**Відповідальний за підготовку  
облікових документів**

**Реєстратор**

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



Юрченко Т.А.