

Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0420U100992

Особливі позначки: відкрита

Дата реєстрації: 14-07-2020

Статус: Захищена

Реквізити наказу МОН / наказу закладу:



II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Куроп'ятник Олена Сергіївна

2. Kuropiatnyk Olena

Кваліфікація:

Ідентифікатор ORCID ID: Не застосовується

Вид дисертації: кандидат наук

Шифр наукової спеціальності: 01.05.02

Назва наукової спеціальності: Математичне моделювання та обчислювальні методи

Галузь / галузі знань: Не застосовується

Освітньо-наукова програма зі спеціальності: Не застосовується

Дата захисту: 22-06-2020

Спеціальність за освітою: Програмне забезпечення систем

Місце роботи здобувача: Дніпровський національний університет залізничного транспорту імені академіка В. Лазаряна

Код за ЄДРПОУ: 01116130

Місцезнаходження: вул. Лазаряна, 2, м. Дніпро, Дніпропетровський р-н., Дніпропетровська обл., 49010, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

Сектор науки: Не застосовується

III. Відомості про дисертацію

Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради): Д 08.084.01

Повне найменування юридичної особи: Національна металургійна академія України

Код за ЄДРПОУ: 02070766

Місцезнаходження: проспект Гагаріна, 4, м. Дніпро, Дніпропетровський р-н., Дніпропетровська обл., 49600, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

Сектор науки: Не застосовується

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

Повне найменування юридичної особи: Дніпровський національний університет залізничного транспорту імені академіка В. Лазаряна

Код за ЄДРПОУ: 01116130

Місцезнаходження: вул. Лазаряна, 2, м. Дніпро, Дніпропетровський р-н., Дніпропетровська обл., 49010, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

Сектор науки: Не застосовується

V. Відомості про дисертацію

Мова дисертації:

Коди тематичних рубрик: 27.03.21, 28.17.19

Тема дисертації:

1. Конструктивно-продукційні моделі природомовних текстів для виявлення запозичень у структурованих документах
2. Constructive- synthesizing models of natural language texts for text borrowings detection in structured documents

Реферат:

1. Дисертаційну роботу присвячено вирішенню актуальної науково-прикладної задачі розробки моделей природомовних текстів для виявлення запозичень у структурованих документах На основі розроблених конструктивно-продукційних моделей мови, мовних конструкцій (текстів) та їх графового представлення запропоновано метод і алгоритми зіставлення текстів та структурованих документів для виявлення за-

позичень. Запропонована модель процесу маскуванню запозичень дозволила формалізувати сценарії маскуванню, створити платформу для моделювання нових змін тексту, автоматизувати побудову тестів для систем антиплагіату. Розроблено програмні засоби виявлення запозичень у текстових фрагментах та структурованих документах. Розроблено програмний засіб автоматизованого формування тестів для перевірки здатності демаскування запозичень систем анти плагіату. Комплексне використання отриманих у роботі результатів дозволяє виконувати автоматизовану перевірку текстових фрагментів і структурованих документів на наявність запозичень; тестувати системи антиплагіату та постійно збільшувати тестову базу, будуючи нові сценарії маскувань.

2. The dissertation is devoted to solving the relevant scientific applied problem of development natural language texts models for detecting borrowings in structured documents. The dissertation reviews and analyzes existing models of natural languages, methods for processing language constructions (texts) and comparing natural language texts. Most models do not take into account the performer model and its features, and ignore human thinking processes. Most methods for preprocessing text do not allow restoring the input data after working with them. Methods based on simple comparisons (fingerprints, greedy comparisons, etc.) are sensitive to borrowings disguise mechanisms. Existing software, including those developed using these methods and algorithms, does not take into account the structure of the document to which the text belongs, when checking for borrowings. According to the results of the analysis of the current state and tendencies of development of linguistic constructions formalization methods and constructions comparison, the necessity of developing effective models and methods of natural language texts for borrowings detection in structured documents is shown. Constructive-synthesizing modeling (CSM) based on the use of formal languages and grammar apparatus, graph theory, methods and means of set theory, regression analysis were used to solve tasks of developed models of natural language texts and methods for its processing. In the framework of the CSM constructors and methods of their transformation (specialization, concretization, interpretation and realization) are used. Formalization of the forming human images processes by means of object-oriented modeling was performed. That allowed constructing a hierarchy of images based on the commonality of attributes in order to represent the meaning of the word and to reflect its connection with the objects of reality within the concept of word semantics. It was used in constructing the constructive-synthesizing languages model. Constructive-synthesizing and object-oriented models of the natural language and text, structured document model, process model of disguise of borrowings text were developed. Based on the developed models of the language, language constructions (texts) and their graph representation, method and algorithms are proposed for compare text fragments and structured documents to borrowings detection. Computer implementations of models the text graph representation and processes of disguise was created. The text-weighted graph compression method was developed to improve the performance of the computer implementation of the graph representation model, which made it possible to use the object serialization mechanism to form a structured documents database. These implementations are software for detecting borrowings in text fragments and structured documents and automated test generation to test the ability to unmask borrowings of anti-plagiarism systems. The developed models and tools form a single complex, which covers: language and speech - the creation of language constructions; lexical, syntactic, semantic components of language constructions; processes of disguise and detection of borrowings text. The time effectiveness of the implementation of models the graphical representation text is investigated. The check time for one document was from 11 to 65 sec for the database from 0.6 to 3.8 million characters. Restore graphs spent about 94% of the time. The influence of masking borrowings on increasing the originality of documents amounted to about 0.007%. Functional effectiveness metrics of developed software for borrowing in text-unstructured documents was compared to its counterpart (WCopyfind). The difference does not exceed 5%. The factors that cause the difference in the performance of the programs have been identified. The proposed model of the borrowing disguise process allows formalizing masking scenarios, creating a platform for modeling new text changes, and automating the construction of tests for anti-plagiarism systems. The integrated use obtained results allows performing an automated borrowings check of text fragments and structured documents; performing test anti-plagiarism systems and constantly increasing the test base by building new disguise scenarios. The developed

software allows completing the requirement of the Law of Ukraine "On Higher Education" regarding the academic plagiarism detection in the diploma works of students in the specialty 121 "Software Engineering" at Dnipro National University of Railway Transport named after Acad. V. Lazaryan.

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. Шинкаренко Віктор Іванович
2. Shynkarenko Viktor Ivanovych

Кваліфікація: 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

Сектор науки: Не застосовується

VII. Відомості про офіційних опонентів та рецензентів

Офіційні опоненти

Власне Прізвище Ім'я По-батькові:

1. Кіріченко Людмила Олегівна

2. Kirichenko Lyudmyla Olegivna

Кваліфікація: 01.05.02

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація:

Повне найменування юридичної особи:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

Сектор науки: Не застосовується

Власне Прізвище Ім'я По-батькові:

1. Шаронова Наталія Валеріївна

2. Sharonova Natalia Valeriyivna

Кваліфікація: 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

Сектор науки: Не застосовується

Власне Прізвище Ім'я По-батькові:

1. Кірія Руслан Вісаріонович

2. Kiriya Ruslan V.

Кваліфікація: 01.05.02

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація:

Повне найменування юридичної особи:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

Сектор науки: Не застосовується

Рецензенти

VIII. Заключні відомості

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

Гнатушенко Вікторія Володимирівна

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

Гнатушенко Вікторія Володимирівна

**Відповідальний за підготовку
облікових документів**

Реєстратор

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



Юрченко Т.А.