

Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0421U101633

Особливі позначки: відкрита

Дата реєстрації: 14-05-2021

Статус: Захищена

Реквізити наказу МОН / наказу закладу:



II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Лупей Максим Іванович

2. Lupei Maksym I

Кваліфікація:

Ідентифікатор ORCID ID: Не застосовується

Вид дисертації: кандидат наук

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 05.13.06

Назва наукової спеціальності: Інформаційні технології

Галузь / галузі знань: Не застосовується

Освітньо-наукова програма зі спеціальності: Не застосовується

Дата захисту: 26-04-2021

Спеціальність за освітою: Прикладна математика

Місце роботи здобувача: Державний вищий навчальний заклад "Ужгородський національний університет"

Код за ЄДРПОУ: 02070832

Місцезнаходження: вул. Підгірна, буд. 46, м. Ужгород, Ужгородський р-н., Закарпатська обл., 88000, 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): Д 35.101.01

Повне найменування юридичної особи: Українська академія друкарства

Код за ЄДРПОУ: 02071004

Місцезнаходження: вул. Під Голоском, буд. 19, м. Львів, Львівська обл., 79020, 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: Державний вищий навчальний заклад "Ужгородський національний університет"

Код за ЄДРПОУ: 02070832

Місцезнаходження: вул. Підгірна, буд. 46, м. Ужгород, Ужгородський р-н., Закарпатська обл., 88000, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

V. Відомості про дисертацію

Мова дисертації:

Коди тематичних рубрик: 28.23, 28.23.24

Тема дисертації:

1. Інформаційна технологія аналізу й визначення авторської та стильової належності українськомовних текстів
2. Information technology of analysis and determination of author's and stylistic affiliation of Ukrainian-language texts

Реферат:

1. У дисертаційній роботі запропоновано вирішення науково-прикладної задачі створення інформаційної технології для аналізу й визначення авторської та стильової належності українськомовних текстів. Для цього було проведено кілька етапів роботи із даними. Проведено аналіз особливостей українськомовних текстів різних стилів та розглянуто особливості української граматики стосовно формування кінцівок слів, що необхідно для проведення етапу попереднього опрацювання українськомовних текстів. В роботі використовується метод стемінгу, який спеціальним чином адаптований під українську мову. На наступному етапі проаналізовано існуючі методи векторизації українськомовного тексту, серед яких виділено метод, що базується на хеш-функціях, метод векторизації за основи зворотної частоти документів та метод векторизації,

що базується на частоті документів. Проведено класифікацію різних методів векторизації українськомовного тексту в поєднанні з різними типами машинного навчання та виявлено найкраще їх співвідношення для кожного типу досліджуваних завдань. Здійснено вибір методів машинного навчання для аналізу й визначення авторської та стильової належності українськомовних текстів, до яких можна віднести різні комбінації архітектур штучних нейронних мереж (MLP, SVR та SVC штучні нейронні мережі). На базі проведеної класифікації методів та підходів векторизації та класифікації, розроблено інформаційну технологію для аналізу й визначення авторської та стильової належності українськомовних текстів. В результаті визначено переваги методів машинного навчання та їхнє використання для створення інформаційних технологій. Проведено вибір методів векторизації текстів, їхню апробацію в рамках інформаційної технології для аналізу українськомовних текстів в комбінації із різними типами машинного навчання, під час якої визначено, що найкращі результати отримано саме при використанні векторизації з використанням зворотної частоти документу та методу векторизації, що базується на використанні хеш-функцій. Було удосконалено метод класифікації текстової інформації на базі багатозарового перцептрону за рахунок використання спеціалізованих процедур навчання та регуляризаційних процедур, що надає можливість зменшити час прийняття рішення без втрати точності.

2. In the dissertation work the decision of a scientific and applied problem of creation of information technology for the analysis and definition of author's and stylistic affiliation of the Ukrainian-language texts is offered. To do this, several stages of data processing were performed. The peculiarities of Ukrainian-language texts of different styles are analyzed and the peculiarities of Ukrainian grammar in relation to the formation of word endings are considered, which is necessary for the stage of preliminary processing of Ukrainian-language texts. The paper uses the method of steaming, which is specially adapted to the Ukrainian language. At the next stage, the existing methods of vectorization of Ukrainian-language text are analyzed, among which the method based on hash functions, the method of vectorization based on the inverse frequency of documents and the method of vectorization based on the frequency of documents are highlighted. The classification of different methods of vectorization of the Ukrainian language text in combination with different types of machine learning is carried out and their best ratio for each type of researched tasks is revealed. The choice of machine learning methods for the analysis and determination of the authorial and stylistic affiliation of Ukrainian-language texts, which include various combinations of architectures of artificial neural networks (MLP, SVR and SVC artificial neural networks). Based on the classification of methods and approaches of vectorization and classification, information technology has been developed for analysis and determination of the author's and stylistic affiliation of Ukrainian-language texts. As a result, the advantages of machine learning methods and their use to create information technology are identified. The choice of methods of vectorization of texts, their testing in the framework of information technology for the analysis of Ukrainian texts in combination with different types of machine learning, during which it was determined that the best results were obtained using vectorization using the inverse frequency of the document hash functions. The method of classifying textual information based on a multilayer perceptron has been improved through the use of specialized training procedures and regularization procedures, which makes it possible to reduce decision-making time without losing accuracy.

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 ОПІВ:

Впровадження результатів дисертації:

Зв'язок з науковими темами:

VI. Відомості про наукового керівника/керівників (консультанта)

Власне Прізвище Ім'я По-батькові:

1. Міца Олександр Володимирович

2. Miza Olexander V

Кваліфікація: к.т.н., 05.13.23

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. Сеньківський Всеволод Миколаевич

2. Senkivskyy Vsevolod M

Кваліфікація: д. т. н., 05.05.01

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. Машталір Сергій Володимирович

2. Mashtalir Serhii V

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

Рецензенти

Власне Прізвище Ім'я По-батькові:

1. Тимченко Олександр Володимирович

2. Timchenko Olexander V

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. Огірко Ігор Васильович

2. Ohirko Ihor V

Кваліфікація: д.ф.-м.н., 01.02.04

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація:

Повне найменування юридичної особи:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

