

# Облікова картка дисертації

## I. Загальні відомості

Державний обліковий номер: 0822U100531

Особливі позначки: відкрита

Дата реєстрації: 01-02-2022

Статус: Захищена

Реквізити наказу МОН / наказу закладу:



## II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Крамов Артем Андрійович

2. Kramov Artem Andriiovych

Кваліфікація:

Ідентифікатор ORCID ID: Не застосовується

Вид дисертації: доктор філософії

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 123

Назва наукової спеціальності: Комп'ютерна інженерія

Галузь / галузі знань:

Освітньо-наукова програма зі спеціальності: Не застосовується

Дата захисту: 18-01-2022

Спеціальність за освітою: Комп'ютерні системи та мережі

Місце роботи здобувача: ПРИВАТНЕ ПІДПРИЄМСТВО "ХЕЛП МІКРО"

Код за ЄДРПОУ: 31565308

Місцезнаходження: вулиця Колекторна, будинок 3-а, м. Київ, 02260, Україна

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

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

**Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради):** ДФ 26.001.242

**Повне найменування юридичної особи:** Київський національний університет імені Тараса Шевченка

**Код за ЄДРПОУ:** 02070944

**Місцезнаходження:** вул. Володимирська, буд. 60, м. Київ, 01033, 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:** Київський національний університет імені Тараса Шевченка

**Код за ЄДРПОУ:** 02070944

**Місцезнаходження:** вул. Володимирська, буд. 60, м. Київ, 01033, Україна

**Форма власності:**

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

**Ідентифікатор ROR:** Не застосовується

### **V. Відомості про дисертацію**

**Мова дисертації:**

**Коди тематичних рубрик:** 28.23.37

**Тема дисертації:**

1. Проектування і застосування нейронних мереж для створення методів оцінки когерентності україномовних текстів
2. Design and application of neural networks for the creation of the methods of the coherence estimation of Ukrainian-language texts

**Реферат:**

1. Робота присвячена дослідженню проектування і застосування різномісних нейронних мереж для створення методів оцінки когерентності україномовних текстів. Здійснення оцінки когерентності документу в автоматизований спосіб відноситься до категорії задач обробки природної мови. Формування вихідної оцінки когерентності передбачає аналіз тематичної цілісності всіх частин тексту на основі відповідності змісту документу фоновим знанням читача та різномісності узгодженості текстових фрагментів. Врахування наведених властивостей тексту протягом обчислення оцінки його когерентності обумовлює застосування відповідних методів розрахунку цілісності документу для вирішення задач різних галузей (медичної діагностики, пошукових систем тощо). Варто зазначити, що аналіз методів оцінки когерентності для текстів

української та інших слов'янських мов знаходиться на початковому етапі. Доцільним є здійснення водночас експериментальної перевірки ефективності застосування існуючих методів для україномовного корпусу, а також створення нових моделей для детального аналізу процесу формування оцінки когерентності текстів української мови. Відсутність фіксованої структури тексту, сформованого згідно з процесом мислення конкретної особи, а також складність однозначної оцінки когерентності документу (суб'єктивність сприйняття тексту читачем на основі власних вражень та фонових знань), унеможливають розрахунок зазначеного показника тематичної цілісності тексту за допомогою визначеного набору інструкцій. Наведені проблеми оцінки когерентності тексту обумовлюють доцільність застосування різних засобів машинного навчання з метою вирішення поставлених задач. У зв'язку з приростом обчислювальної потужності пристроїв та розвитком технологій паралельних і розподілених обчислень, сучасні методи оцінки когерентності тексту ґрунтуються на проектуванні багатопланових нейронних мереж різної архітектури. Застосування моделей глибокого навчання дозволяє підвищити точність вирішення поставленої задачі за допомогою узагальнення різноманітних властивостей текстів вхідного корпусу. Однак у зв'язку зі складністю однозначного визначення критеріїв когерентного документу, крім досягнення встановленої точності відповідних методів доцільно з'ясувати причини отримання вихідного результату. Актуальною задачею є дослідження формування ознак вхідного тексту та проектування елементів нейронних мереж з метою виявлення характеристик тексту та типів зв'язку між його елементами, що аналізуються протягом обчислення вихідної оцінки когерентності україномовного документу. В роботі здійснено аналіз існуючих методів оцінки когерентності текстів на основі застосування методології машинного навчання. Детально розглянуто принцип роботи методів Entity Grid, Entity Graph та теорії риторичних структур, що полягає в аналізі закономірності зміни синтаксичних та дискурсивних ролей в межах сусідніх речень та всього тексту в цілому. Основними недоліками застосування зазначених методів є відсутність врахування інших елементів тексту, семантичних властивостей компонентів; залежність точності вирішення задач від зовнішніх інструментів. Наведено аналіз методів оцінки локальної та глобальної когерентності документу на основі використання різноманітних нейронних мереж для формалізованого представлення елементів тексту та обчислення його цілісності на рівні семантичної узгодженості речень. Зазначено доцільність використання рекурентних та згорткових шарів для проектування моделі нейронної мережі водночас для оцінки локальної та глобальної когерентності документу. На основі проведеного порівняльного аналізу методів показано ефективність застосування моделей нейронних мереж для вирішення поставлених задач оцінки когерентності тексту.

2. The paper is devoted to the investigation of the designing and applying of different neural networks for the creation of the methods of the coherence estimation of Ukrainian-language texts. The coherence estimation of a document falls into the category of natural language processing tasks. The generation of an output coherence estimation value involves the analysis of the thematic integrity of all text's parts basing on the correspondence of the document's content to the background knowledge of a reader and the different types of the consistency of text's spans. The consideration of the listed features of a text during the estimation of its coherence leads to the applying of the correspondent coherence estimation methods for the solving of the tasks that belong to different areas (medical diagnostics, search engines, etc.). It should be mentioned that the analysis of the coherence evaluation of the documents of Ukrainian and other Slavic languages is still at the initial stage. It is advisable to perform the experimental verification of the effectiveness of state-of-the-art methods on Ukrainian-language corpora; moreover, an important task consists in the creation of new models for the detailed analysis of the formation process of the coherence estimation value of Ukrainian-language texts. The lack of the fixed text's structure formed by a person according to its thoughts and the complexity of the unambiguous assessment of the coherence of a document (subjectivity of the reader's perception of the text based on their own impressions and background knowledge) complicate the evaluation of the mentioned metric of the thematic integrity of a text by utilizing a pre-defined set of instructions. The listed issues of the coherence estimation of a text cause the expediency of the applying of different machine learning techniques for the solving of assigned tasks. Taking into account the increase of devices' computational power and the development of the parallelized and distributed computing, the state-of-the-art methods of the coherence evaluation of a text are based on the designing of the

multilayer neural network of different architecture. The applying of deep learning models allows the achieving of accuracy increase according to an assigned task by means of the generalization of the different features of the texts of an input corpus. However, taking into account the complexity of the unambiguous definition of the criteria of a coherent document, besides the achievement of the desired accuracy of correspondent methods it is advisable to find out the reasons for the retrieving of an output result. An important task is to investigate the formation of the features of an input text and the connection types between its elements that are analyzed during the evaluation of the output coherence value of a Ukrainian-language document. In the paper, the analysis of current coherence estimation methods based on machine learning techniques has been performed. The principle of work of Entity Grid, Entity Graph, and an RST-based method that consists in the analysis of the regularity of syntactic and discourse roles change within adjacent sentences and a whole document at all has been considered in details. The key disadvantages of the usage of the mentioned methods are the neglecting of other text's elements, the semantic properties of components; the dependency of the accuracy of the solving of assigned tasks on external instruments. The analysis of the methods of the local and global coherence estimation of a document based on the usage of different neural networks for the representation of text's elements and the evaluation of its coherence at the level of the semantic consistency of sentences has been performed. The appropriateness of the utilizing of recurrent and convolutional layers for the designing of a neural network model for the estimation of both local and global coherence of a document has been shown. Basing on the performed comparative analysis of methods, the effectiveness of the applying of neural network models for the solving of the assigned tasks of the coherence estimation of a text has been demonstrated.

**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. Погорілий Сергій Дем'янович
2. Pohorilyi Serhii Demianovych

**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

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

### **Офіційні опоненти**

**Власне Прізвище Ім'я По-батькові:**

1. Бідюк Петро Іванович

2. Bidiuk Petro Ivanovych

**Кваліфікація:** д.т.н., 05.13.03

**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. Глибовець Микола Миколайович

2. Hlybovets Mykola Mykolaiovych

**Кваліфікація:** д.ф.-м.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. Бойко Юрій Володимирович

2. Boiko Yurii Volodymyrovych

**Кваліфікація:** к. ф.-м. н., 01.04.10

**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. Марченко Олександр Олександрович

2. Marchenko Oleksandr Oleksandrovyeh

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

**Ідентифікатор ORCID ID:** Не застосовується

**Додаткова інформація:**

**Повне найменування юридичної особи:**

**Код за ЄДРПОУ:**

**Місцезнаходження:**

**Форма власності:**

**Сфера управління:**

**Ідентифікатор ROR:** Не застосовується

## **VIII. Заключні відомості**

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

Кравченко Юрій Васильович

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

Кравченко Юрій Васильович

**Відповідальний за підготовку  
облікових документів**

**Реєстратор**

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



Юрченко Т.А.