

Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0825U001587

Особливі позначки: відкрита

Дата реєстрації: 08-05-2025

Статус: Наказ про видачу диплома

Реквізити наказу МОН / наказу закладу:



II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Бочарова Майя Юріївна

2. Maiia Bocharova

Кваліфікація: 122

Ідентифікатор ORCID ID: Не застосовується

Вид дисертації: доктор філософії

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 122

Назва наукової спеціальності: Комп'ютерні науки

Галузь / галузі знань: інформаційні технології

Освітньо-наукова програма зі спеціальності: Комп'ютерні науки

Дата захисту: 26-06-2025

Спеціальність за освітою: Комп'ютерні науки

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 8717

Повне найменування юридичної особи: Одеський національний університет імені І. І. Мечникова

Код за ЄДРПОУ: 02071091

Місцезнаходження: вул. Дворянська, буд. 2, Одеса, 65082, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

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

Повне найменування юридичної особи: Одеський національний університет імені І. І. Мечникова

Код за ЄДРПОУ: 02071091

Місцезнаходження: вул. Дворянська, буд. 2, Одеса, 65082, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

V. Відомості про дисертацію

Мова дисертації: Українська

Коди тематичних рубрик: 20.53, 20.53.19

Тема дисертації:

1. Методи та інформаційна технологія предметно-орієнтованого аналізу природномовних текстів.
2. Methods and information technology of subject-oriented analysis of natural language texts.

Реферат:

1. У дисертаційній роботі представлені результати проведених здобувачем досліджень, які виконують актуальне наукове завдання створення моделей та методів предметно-орієнтованого аналізу природномовних текстів, яке має істотне значення для розвитку інформаційних технологій. У вступі дисертації обґрунтовано актуальність дослідження за темою предметно-орієнтованого аналізу природномовних текстів, сформульовані мета, задачі та методи дослідження, наведено наукову новизну та практичне значення отриманих результатів, зазначено особистий внесок здобувача. У першому розділі роботи досліджено актуальний стан проблеми автоматизованого аналізу документів в сфері управління персоналом із застосуванням штучного інтелекту. Показано, що обробка резюме для витягнення ключової інформації, зіставлення вакансій та резюме є необхідним елементом підвищення ефективності рекрутингу і перспективним напрямком для подальшого удосконалення і розвитку. Показано, що застосування моделей, які використовують комп'ютерний зір, не є доцільним для обробки візуально насичених документів у сфері управління людськими ресурсами. На основі аналізу літературних джерел обґрунтована доцільність

використання контексту під час моделювання подань фраз. Висвітлена проблема обробки документів, обсяг яких перевищує ліміт моделей, які використовуються для їх обробки. Показано, що відсутність даних щодо впливу обсягу тренувальних зразків при автоматичній їх генерації (з використанням великих мовних моделей) на якість сумаризації документів у домені управління персоналом потребує дослідження в цьому напрямку. Потребують дослідження підходи некерованого попереднього тренування з використанням структури документів, а також функції втрат, які використовуються для попереднього тренування (зважена функція втрат). Відзначена важливість англійської мови для поширення набутих знань щодо використання штучного інтелекту у рекрутингу. Встановлено необхідність удосконалення крос-лінгвістичної дистиляції векторних подань для підвищення ефективності рекрутингу із застосуванням штучного інтелекту. Встановлено доцільність дослідження впливу дистиляції на показники швидкості та якості етапів технології обробки природномовних текстів щодо аналізу резюме та зіставлення з вимогами вакансій. У другому розділі розроблено методи та моделі для предметно-орієнтованої обробки природномовних текстів. В тому числі представлено новий метод безпосередньої інтеграції параметрів про стильові ознаки, де додаткові дискретні ознаки векторизуються і передаються в архітектуру "Трансформер" разом із позиційними і токенними поданнями. Запропоновано новий метод тренування подань назв посад, що базується на використанні фраз навичок, які зазначені в описі роботи. Цей метод базується на введенні спеціального токена для виділення та представлення кожної навички у поєднанні з контрастним тренуванням з метою зіставлення усередненого подання навичок та назви посади з одного опису роботи. Запропоновано новий метод некерованого навчання моделі з використанням структури документів. На відміну від традиційного методу, в якій позитивні пари для подальшого контрастного навчання вибирають з документу випадковим чином, запропонований метод базується на використанні структури документу. Запропоновано новий метод автоматичного створення датасету вакансія-резюме, який полягає у використанні структури документа і визначеного опису останньої ролі та перетворення цього запису на опис вакансій з використанням великої мовної моделі. Запропоновано метод скорочення тексту з урахуванням структури документу та ключових фраз. Цей метод полягає у скороченні кожної секції пропорційно до її відсоткового внеску у загальну довжину резюме на основі виділення ключових фраз. Розроблено метод некерованого попереднього тренування для сумаризації документів у сфері управління персоналом. Цей метод полягає у використанні секції "анотація" з резюме для некерованого тренування моделі сумаризації, а також у застосуванні зваженої функції втрат, яка підвищує вагомість для токенів, які формують ключові фрази. У третьому розділі представлена інформаційна технологія предметно-орієнтованого аналізу природномовних текстів, яка може бути застосована за двома напрямками: вироблення рекомендацій резюме в умовах відсутності рекрутера, та інтенсифікації процесу відбору та ранжування резюме рекрутером, що дає можливість рекрутерам швидко та зручно ознайомлюватися з рекомендованими кандидатами та відфільтровувати їх. Представлена технологія є послідовністю застосування наступних етапів: "Сегментація", "Парсинг", "Сумаризація", "Векторизація". В результаті застосування цих етапів документ перетворюється на сукупність атрибутів, анотації та векторного подання, які зберігаються у векторній базі даних. Показано, що для оцінювання етапів технології "AI ResJobFit" необхідно обчислювати наступні показники: F1, Recall@N, ???, MRR, nDCG, RougeN.

2. The dissertation presents the results of the research conducted by the applicant, which fulfills the urgent scientific task of creating models and methods for subject-oriented analysis of natural language texts, which is essential for the development of information technology. The introduction of the dissertation substantiates the relevance of the research on the subject-oriented analysis of natural language texts, formulates the purpose, objectives and methods of the study, presents the scientific novelty and practical significance of the results obtained, and indicates the personal contribution of the applicant. The first section of the paper investigates the current state of the problem of automated document analysis in the field of human resources management using artificial intelligence. It is shown that processing resumes to extract key information, matching vacancies and resumes is a necessary element in improving the efficiency of recruiting and a promising area for further improvement and development. It is shown that the use of models that use computer vision is not appropriate for processing visually rich documents in the field of human resource management. Based on the analysis of literature

sources, the expediency of using context in modeling phrase representations is substantiated. The problem of processing documents whose volume exceeds the limit of the models used for their processing is highlighted. It is shown that the lack of data on the impact of the volume of training samples during their automatic generation (using large language models) on the quality of document summarization in the HR domain requires research in this direction. The approaches of unsupervised pre-training using the structure of 10 documents, as well as the loss functions used for pre-training (weighted loss function) need to be investigated. The importance of the English language for the dissemination of the acquired knowledge on the use of artificial intelligence in recruiting is noted. The need to improve the cross-linguistic distillation of vector representations to increase the efficiency of recruiting using artificial intelligence is determined. The expediency of studying the impact of distillation on the speed and quality of the stages of natural language text processing technology for analyzing resumes and comparing them with job requirements has been established. In the second section, we develop methods and models for subject-oriented processing of natural language texts. In particular, a new method of direct integration of parameters about style features is presented, where additional discrete features are vectorized and transferred to the Transformer architecture along with positional and token representations. A new method of training job title representations based on the use of skill phrases specified in the job description is proposed. This method is based on the introduction of a special token to highlight and represent each skill, combined with contrast training to compare the average skill representation and job title from one job description. A new method of unsupervised model training using document structure is proposed. In contrast to the traditional method, in which positive pairs for further contrast training are randomly selected from the document, the proposed method is based on the use of the document structure. A new method of automatic creation of a job-resume dataset is proposed, which consists in using the document structure and a specific description of the last role and converting this record into a job description using a large language model. A method of text reduction based on the document structure and key phrases is proposed. This method consists in shortening each section in proportion to its percentage contribution to the total length of the resume based on the selection of key phrases. 11 An unsupervised pre-training method was developed for summarizing documents in the field of human resources management. This method consists of using the “annotation” section of the resume for unsupervised training of the summarization model, as well as applying a weighted loss function that increases the weight for tokens that form key phrases. The 3rd chapter presents the information technology of subject-oriented analysis of natural language texts, which can be applied in two directions: making resume recommendations in the absence of a recruiter, and intensifying the process of selection and ranking of resumes by a recruiter, which allows recruiters to quickly and conveniently familiarize themselves with recommended candidates and filter them. The presented technology is a sequence of the following stages: “Segmentation, Parsing, Summarization, Vectorization.

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:

- Bocharova M. Y. VacancySBERT: the approach for representation of titles and skills for semantic similarity search in the recruitment domain / M. Y. Bocharova, E. V. Malakhov, V. I. Mezhuiev // Applied Aspects of Information Technology. – 2023. – Vol. 6, no. 1. – P. 52–59.
- Bocharova M. General-purpose text embeddings learning for Ukrainian language / M. Bocharova, E. Malakhov // Advanced Information Technology. – 2024. – No 1(3). – P. 6–12.
- Bocharova M. Improving information theory of context-aware phrase embeddings in HR domain / M. Bocharova, E. Malakhov // Eastern-European Journal of Enterprise Technologies. – 2024. – Vol. 5, iss. 2(131).

– P. 53–60.

- Bocharova M. Y. ResJobFit – end-to-end artificial neural networks based technology for job-resume matching / M. Y. Bocharova, E. V. Malakhov // Applied Aspects of Information Technology. – 2024. – Vol. 7, no. 4. – P. 378–391.

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

Соціально-економічна спрямованість: забезпечення промисловості чи населення новим видом інформаційно-комунікаційних послуг

Охоронні документи на ОПІВ:

Впровадження результатів дисертації: Впроваджено

Зв'язок з науковими темами: 0121U111663

VI. Відомості про наукового керівника/керівників (консультанта)

Власне Прізвище Ім'я По-батькові:

1. Малахов Євгеній Валерійович
2. Eugene Malakhov

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

Ідентифікатор ORCID ID: 0000-0002-9314-6062

Додаткова інформація:

Повне найменування юридичної особи: Одеський національний університет імені І. І. Мечникова

Код за ЄДРПОУ: 02071091

Місцезнаходження: вул. Дворянська, буд. 2, Одеса, 65082, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

VII. Відомості про офіційних опонентів та рецензентів

Офіційні опоненти

Власне Прізвище Ім'я По-батькові:

1. Арсірій Олена Олександрівна
2. Olena Arsirii

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

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація: Scopus Author ID: 54419480900

Повне найменування юридичної особи: Національний університет "Одеська політехніка"

Код за ЄДРПОУ: 43861328

Місцезнаходження: пр. Шевченка, буд. 1, Одеса, 65044, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Власне Прізвище Ім'я По-батькові:

1. Висоцька Вікторія Анатоліївна

2. Viktoria Vysotska

Кваліфікація: д.т.н., доцент, 10.02.21

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація:

Повне найменування юридичної особи: Національний університет "Львівська політехніка"

Код за ЄДРПОУ: 02071010

Місцезнаходження: вул. Степана Бандери, буд. 12, Львів, 79013, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Власне Прізвище Ім'я По-батькові:

1. Бісікало Олег Володимирович

2. Oleh Bisikalo

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

Ідентифікатор ORCID ID: 0000-0002-7607-1943

Додаткова інформація:

Повне найменування юридичної особи: Вінницький національний технічний університет

Код за ЄДРПОУ: 02070693

Місцезнаходження: вул. Хмельницьке шосе, буд. 95, Вінниця, Вінницький р-н., 21021, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Рецензенти

Власне Прізвище Ім'я По-батькові:

1. Шпінарева Ірина Михайлівна

2. Iryna Shpinareva

Кваліфікація: к. ф.-м. н., доцент, 01.04.01

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація:

Повне найменування юридичної особи: Одеський національний університет імені І. І. Мечникова

Код за ЄДРПОУ: 02071091

Місцезнаходження: вул. Дворянська, буд. 2, Одеса, 65082, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

VIII. Заключні відомості

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

Мещеряков Володимир Іванович

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

Мещеряков Володимир Іванович

**Відповідальний за підготовку
облікових документів**

Лукашук Світлана Борисівна

Реєстратор

УкрІНТЕІ

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



Юрченко Тетяна Анатоліївна