

Голові спеціалізованої вченої ради
ДФ 74.122.2025
Одеського національного університету імені І.І. Мечникова
доктору технічних наук, професору
Мещерякову Володимирі Івановичу

Відгук
офіційного опонента

Висоцької Вікторії Анатоліївни, доктора технічних наук, доцента, доцента кафедри інформаційних систем та мереж Національного університету «Львівська політехніка» про дисертацію **Бочарової Майї Юріївни «Моделі та інформаційна технологія предметно-орієнтованої обробки природномовних текстів»**, подану на здобуття наукового ступеня доктора філософії за спеціальністю 122 Комп'ютерні науки, галузь знань 12 Інформаційні технології

(Одеса, 2025)

1. Актуальність дисертаційного дослідження. Актуальність дослідження зумовлена швидкими змінами на ринку праці, що викликані глобальними технологічними трансформаціями, цифровізацією та міграційними процесами. Автоматизація процесів рекрутингу, зокрема аналізу резюме та вакансій, є одним із ключових напрямів удосконалення управління персоналом. Сучасні методи штучного інтелекту, зокрема технології глибинного навчання та семантичного пошуку, дозволяють значно підвищити ефективність цього процесу, зменшуючи час на обробку документів і покращуючи якість відповідності між кандидатами та вакансіями.

Розробка інформаційної технології, що інтегрує передові методи штучного інтелекту для автоматизованого аналізу резюме та вакансій, є важливим кроком у вдосконаленні рекрутингових процесів. Це сприяє ефективному управлінню людськими ресурсами, підвищенню конкурентоспроможності компаній та розвитку персоналу. Враховуючи глобальні тенденції та зростаючий попит на високоточні технології для аналізу

великих обсягів текстових даних, дослідження є надзвичайно актуальним як для України, так і для міжнародного співтовариства.

2. Ступінь обґрунтованості наукових положень, висновків і рекомендацій, сформульованих у дисертації.

Об'єктом дослідження є обробка природної мови в рекрутингових документах (резюме, вакансії). *Предметом* дослідження є методи глибокого навчання для ідентифікації та екстракції даних у візуально насичених документах, моделі їх подання, контекстно-освічені методи фразового подання та методи дистиляції векторних подань.

Мета дослідження полягала в підвищенні ефективності рекрутингу за рахунок інтенсифікації процесів відбору, ранжування резюме та вироблення рекомендацій шляхом створення методів та інформаційної технології предметно-орієнтованого аналізу природномовних текстів.

Мета та основні завдання дисертаційної роботи сформульовані чітко логічно, визначаються науковою новизною. Чіткі дослідницькі завдання, що представлені у дисертації, є адекватними характеру обраної теми, предметно окреслюють наукову позицію автора, створюють можливість всебічного й об'єктивного розкриття даної теми.

У *вступі* обґрунтовано актуальність дослідження предметно-орієнтованого аналізу природномовних текстів, сформульовано мету, задачі, методи дослідження, наукову новизну та практичне значення результатів.

У *розділі 1* проаналізовано стан автоматизованого аналізу документів у сфері управління персоналом із застосуванням ШІ. Виявлено проблеми обробки візуально насичених документів, подання фраз із контекстом, обробки довгих текстів та впливу обсягу тренувальних даних на сумаризацію. Відзначено важливість англійської мови для поширення знань та вдосконалення крос-лінгвістичної дистиляції.

У *розділі 2* запропоновано нові методи обробки текстів: інтеграції параметрів стильових ознак, навчання подань посад на основі фраз навичок, некерованого навчання з урахуванням структури документів, автоматичного

створення датасету вакансія-резюме та скорочення тексту з урахуванням структури.

У розділі 3 представлено інформаційну технологію аналізу текстів для рекрутингу, що включає сегментацію, парсинг, сумаризацію та векторизацію.

У розділі 4 обґрунтовано ефективність методів обробки текстів, зокрема інтеграції параметрів стильових ознак та контекстно-орієнтованого вирівнювання фраз. Доведено перевагу косинусної функції втрат для дистиляції векторних подань.

У розділі 5 проведено оцінку ефективності технології обробки текстів у сфері управління персоналом та вплив окремих етапів на швидкість і якість.

У висновках логічно наведено головні підсумки дослідження. Вони відповідають поставленим завданням, підтверджують наукову новизну та повною мірою відображають результати дослідження. Вторинно успішно вирішено актуальну науково-практичну задачу предметно-орієнтованого аналізу природномовних текстів для автоматизації аналізу та інтерпретації великих масивів даних у сфері управління персоналом. Розроблена інформаційна технологія дозволяє підвищити ефективність рекрутингу шляхом поліпшення точності зіставлення кваліфікацій кандидатів з вимогами вакансій, що підтверджується значним покращенням метрик якості та швидкості обробки документів. Особливо цінним є вдосконалення моделей представлення токенів та фраз, а також методів нормалізації назв посад, що дозволило досягти суттєвих покращень у порівнянні з базовими моделями. Практичне впровадження результатів у компаніях Data Science UA та Daxtra Technologies підтверджує їх ефективність та доцільність.

Список використаних джерел та літератури відповідає виду і проблематиці роботи, є інформативним та логічно структурованим.

3. Наукова новизна одержаних результатів.

Дисертаційна робота вирішує актуальну науково-практичну задачу підвищення ефективності аналізу документів у сфері управління персоналом за рахунок вдосконалення методів та створення інформаційної технології

обробки природномовних текстів. Можна погодитись, що вона містить раніше незахищені наукові положення та отримані автором нові науково обґрунтовані результати. А саме:

- удосконалено модель подання токенів для візуального насичених документів, яка відрізняється від існуючих безпосередньою інтеграцією параметрів про стильові ознаки, що дозволяє підвищити якість подань токенів таких документів без використання методів комп'ютерного зору;

- *вперше запропоновано* метод подання фраз у контексті, що базується на використанні спеціальних маркерів для виділення фраз, які моделюються, що дозволяє значно пришвидшити процес подання фраз за рахунок використання лише однієї моделі та покращити якість в порівнянні з базовими методами;

- *вперше запропоновано* метод структурування документа для некерованого навчання подань текстів у сфері управління персоналом, що дозволяє адаптувати модель до домену та як наслідок підвищити якість сумаризації документів;

- удосконалено метод зменшення обсягу тексту, що оброблятиметься моделлю, який відрізняється від існуючих урахуванням структури документа та ключових фраз, для подальшої сумаризації, що дозволяє підвищити якість сумаризації довгих документів;

- *вперше запропоновано* модель векторизації документів у сфері управління людськими ресурсами на основі векторних подань секцій та механізму самоуваги разом із абсолютним позиційним кодуванням, що дозволяє покращити якість подань документів, довжина яких перевершує ліміт токенів, притаманний моделі.

Таким чином, наукові положення представлені в дисертаційній роботі здобувача дозволяють розширити клас важливих для теорії та практики задач обробки природномовних документів з використанням моделей на основі архітектури типу “Трансформер”.

У тексті дисертації відсутні порушення академічної доброчесності.

4. Повнота викладу основних результатів дисертації в опублікованих публікаціях, зарахованих за темою дисертації.

Основні положення та результати дослідження пройшли апробацію в межах міжнародних науково-практичних і наукових конференцій (Малазія, Україна).

Теоретичні та практичні напрацювання дослідження знайшли відображення в п'яти наукових публікаціях М.Ю.Бочарової. Три статті опубліковано у фахових виданнях України, одна — у виданні, проіндексованому в базі даних Скопус.

5. Практичне значення одержаних результатів.

Практичне значення дослідження полягає у вдосконаленні та адаптації технологій обробки природної мови для україномовних текстів та їх впровадженні в реальних умовах.

По-перше, розроблені методи багатомовної дистиляції текстових подань з використанням косинусної функції втрат дозволяють оптимізувати процес навчання моделей студентів на основі моделей вчителів, що відкриває нові можливості для ефективної крос-лінгвістичної дистиляції знань. Практична цінність підтверджується створенням еталонного набору даних для оцінювання векторних подань текстів українською мовою, який був оприлюднений на GitHub.

По-друге, запропонована технологія рекомендацій резюме в умовах відсутності рекрутера, базована на архітектурах типу "Трансформер" та інтеграції сегментації, парсингу й векторизації документів, була успішно апробована у компанії Daxtra Technologies, що підтверджено відповідним актом впровадження.

Крім того, розроблена технологія інтенсифікації процесу відбору та ранжування резюме з використанням нейронних мереж, яка включає етап сумаризації для підвищення ефективності зіставлення вакансій та резюме, також була впроваджена в Daxtra Technologies.

Особливу увагу слід приділити алгоритму обробки природної мови для візуально насичених документів, який дозволяє ефективно опрацьовувати документи з комплексною структурою та успішно використовується в Daxtra Technologies.

Розроблені некеровані методи навчання векторних подань назв посад та текстів у сфері управління персоналом сприяють покращенню кластеризації та нормалізації даних. Вони також були впроваджені в компанії Daxtra Technologies, а створені еталонні набори даних доступні на GitHub.

Також варто зазначити успішне застосування контекстно-освіченого методу подання важливих фраз, розробленого для підвищення точності представлення навичок кандидатів у сфері управління персоналом. Цей метод був апробований у компаніях Daxtra Technologies та Data Science UA, що підтверджується актами апробації.

Результати дисертаційної роботи знайшли практичне застосування в науково-дослідних роботах, зокрема у проєкті "Методи, моделі, інформаційні технології розподілених систем підтримки прийняття організаційних рішень" (ДР 0121U111663). Крім того, розроблені методи використовуються в навчальному процесі кафедри Математичного забезпечення комп'ютерних систем для викладання дисципліни "Методи обробки текстів природної мови", що підтверджує їхню освітню цінність.

Таким чином, результати дослідження мають значний прикладний потенціал як у науковій, так і в комерційній сфері, сприяючи вдосконаленню технологій аналізу природномовних текстів та автоматизації процесів управління персоналом.

6. Зауваження та дискусійні положення до дисертації.

- не коректна назва дисертації, так як інформаційна технологія це сукупність методів і засобів;
- авторкою досліджені моделі на основі архітектури типу "Трансформер" з класичним механізмом самоуваги (Multihead self attention),

проте було б доцільно проаналізувати можливість використання альтернативних методів самоуваги, зокрема, Flash Attention, Paged Attention;

- у роботі розглянуто питання прискорення процесу сприйняття рекрутером резюме, але такі аспекти, як врахування вподобань споживачів (наприклад, форматування та структура анотації) не розкрито повною мірою. Було б доцільно розглянути вплив різних факторів на результат сприйняття резюме рекрутером;

- розділ 4 роботи є перевантаженим за обсягом – для кращого сприйняття було б доцільно розділити його на два, наприклад: основні етапи технології та опціональні;

- згідно назви, мети та деяких задач дослідження полягає ще в розробленні інформаційної технології, яка, в свою чергу, передбачає розроблення засобів (програм, системи, модулів тощо), але в роботі мало приділено увагу цим розробленим засобам, опису їх технічних характеристик, інтерфейсів, архітектури тощо;

- в роботі виявлено ряд технічних помилок граматичного та синтаксичного характеру.

7. Висновок про відповідність дисертації вимогам.

Перелічені зауваження не є принциповими і такими, що піддають сумніву результати дослідження та ніяким чином не зменшують наукову і практичну цінність роботи та позитивне враження від впровадження її результатів.

Дисертаційна робота, яка представлена на відгук, є закінченою кваліфікаційною науковою працею, яка містить раніше не захищені наукові положення і одержані автором нові науково обґрунтовані результати в області підвищення ефективності обробки природномовних текстів. Виклад матеріалу дослідження логічний та послідовний, висновки та рекомендації обґрунтовані, результати достовірні. Аналіз роботи дає підстави для висновку, що дисертація Бочарової Майї Юріївни є завершеним, самостійним науковим

дослідженням, яке має наукову новизну та практичне значення для теоретичних та практичних досліджень в галузі сучасної інженерії програмного забезпечення.

За актуальністю, науковою та практичною цінністю здобутих результатів дисертація «Моделі та інформаційна технологія предметно-орієнтованої обробки природномовних текстів» відповідає вимогам «Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», затвердженого постановою Кабінету Міністрів України від 12 січня 2022 року №44, із змінами.

Враховуючи наведені аспекти, авторка дисертації Бочарова Майя Юріївна заслуговує на присудження їй наукового ступеня доктора філософії у галузі знань 12 — Інформаційні технології за спеціальністю 122 — Комп'ютерні науки.

Офіційний опонент,

професор кафедри інформаційних систем та мереж

Національного університету

«Львівська політехніка»

доктор технічних наук, доцент

Вікторія ВИСОЦЬКА

«Підпис Висоцької В.А. засвідчую»:

Вчений секретар

Національного університету

«Львівська політехніка», в.т.н., доцент

Роман БРИЛИНСЬКИЙ