

ВИСНОВОК
ПРО НАУКОВУ НОВИЗНУ, ТЕОРЕТИЧНЕ ТА ПРАКТИЧНЕ
ЗНАЧЕННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЇ

Демидович Інни Миколаївни на тему
«Розвиток методів та засобів для визначення авторства україномовних
текстів на основі конструктивно-продукційного моделювання»,
що подана на здобуття ступеня доктора філософії
за спеціальністю 122-комп'ютерні науки

Дисертаційна робота Демидович Інни Миколаївни на тему «Розвиток методів та засобів для визначення авторства україномовних текстів на основі конструктивно-продукційного моделювання» виконана на кафедрі «Комп'ютерні інформаційні технології» факультету «Комп'ютерних технологій і систем» Українського державного університету науки і технологій, подана на здобуття ступеня доктора філософії зі спеціальності 122-комп'ютерні науки. Тема дисертації затверджена на засіданні вченої ради Дніпровського національного університету залізничного транспорту імені академіка Лазаряна (протокол №3 від 28 жовтня 2019 року) з уточненням на засіданні вченої ради Українського державного університету науки і технологій (протокол № 2 від 4 жовтня 2023 року).

**1. Ступінь актуальності теми дисертації та її зв'язок з планами
наукових робіт університету**

Стрімке зростання обсягів природномовних текстів у різних глобальних мережах супроводжується проявами неправомірного їх використання. Зокрема видання скопійованого тексту, або його частини без зазначення або зазначення хибного, найчастіше свого (копіювальника).

У зв'язку з цим виникає ціла низка задач встановлення автора тексту. Наприклад: є декілька претендентів на авторство, один з яких дійсно автор, а інші – копіювальники. Є зразки текстів всіх або деяких претендентів, авторство яких точно встановлене. Треба встановити хто автор тексту, або відсіяти частину претендентів. Частіш за все мова йдеться про двох претендентів, один з яких автор.

Інша задача: є деякий текст заявленого автора. Слід знайти цей текст, або його частини у наявному банку текстів і встановити ступінь оригінальності.

Кожна з цих, або інших задач, автоматизовано можуть бути вирішені лише з деякою похибкою. Для першої з вказаних задач впливає варіативність стилю автора з часом та набуттям ним нових знань, навичок, зміною емоційного стану і т.і. Для другої задачі неможливо мати повний банк текстів. Він є багатомовним і постійно поповнюється.

Визначення авторства тексту досі є не вирішеною проблемою для багатьох сфер діяльності людини, таких як освіта, юриспруденція, літературознавство, історія тощо.

В даний час для атрибуції текстів (один з атрибутів – автор) застосовуються підходи з теорії розпізнавання образів, математичної статистики та теорії ймовірностей, алгоритми нейронних мереж та кластерного аналізу та багато інших. Проте, всі такі методи не мають достатньої ефективності та не можуть працювати з текстами різних мов та напрямів, а також не працюють зі стилістичними особливостями автора в достатній мірі.

Основне досягнення цієї роботи є співставлення та встановлення статистичного зв'язку між рішеннями наведених вище двох задач. Це дає можливість поєднання методів розв'язів цих задач, взаємного підсилення, а в деяких випадках і заміни одного іншим.

Зв'язок роботи з науковими програмами, планами, темами.

Дисертаційна робота виконана відповідно до стратегії АТ «Укрзалізниця» на 2019–2023 роки, зокрема за напрямками інтеграції та стандартизації інформаційних систем, підвищення достовірності даних та автоматизації бізнес-процесів, а також директиви Євросоюзу 2008/57/ЕС.

Наукові дослідження, викладені в дисертації, виконані згідно з напрямом наукової роботи кафедри «Комп'ютерні інформаційні технології» Українського державного університету науки і технологій, а саме є частиною

науково-дослідних робіт «Інструментальна підтримка систем обробки природньомовних документів» (2022 р. № держреєстрації 0122U002086), у яких дисертант приймала участь у якості виконавця.

2. Наукова новизна, теоретичне та практичне значення результатів дисертації

Наукова новизна одержаних результатів. Наукова новизна отриманих результатів. В роботі вперше:

1) розроблено модель природньомовного тексту у вигляді множини правил стохастичних граматик та метод порівняння текстів на основі порівняння цих правил, на відміну від існуючих моделей вона дозволяє враховувати синтаксичні та стилістичні особливості тексту автора;

2) розроблено метод визначення авторства тексту на основі комбінованих показників складності тексту статистичного, рекурентного аналізу та конструктивно-продукційного аналізу. Це забезпечує всебічний аналіз з врахуванням значної кількості параметрів тексту;

3) розроблено метод встановлення профілю автора який визначає найбільш суттєві показники аналізу текстів притаманні певному автору, що спрощує та полегшує подальші розрахунки та скорочує потрібний час та ресурси розрахунку;

4) встановлено статистично значимий зв'язок результатів рішення задач виявлення запозичень та встановлення авторству текстів

Поліпшено:

5) метод багатокритеріальної оптимізації на основі генетичного алгоритму;

Отримали подальший розвиток:

6) методи конструктивно-продукційного моделювання у частині використання параметризованих конструкторів, зв'язків між конструкторами та конструювання конструкторів;

7) метод розпізнавання образів за критерієм мінімальної відстані в частині формування образу текстів;

80 методи рекурентного аналізу для роботи з природньомовними текстами.

Практичне значення отриманих результатів. Результати експериментів визначили значення раціонального параметру – мінімальну довжину текстових фрагментів (у кількості слів) яку слід вважати запозиченням, він дорівнює п'яти словам. Це слід вважати рекомендацією для використання будь-яких програм виявлення запозичень.

Запропонований метод може використовуватись як для вирішення проблем пошуку запозичень, так і для встановлення вірогідного авторства тексту. Може використовуватись як додатковий інструмент для комплексної перевірки текстів на плагіат. Особливістю є можливість використання запропонованого метода як самостійно для вирішення відповідних задач, так і комплексі з іншими інструментами для покращення їх роботи та отримання більш достовірних результатів.

Розроблений інструментарій на основі запропонованих методів та моделей використовувалося для пошуку прямих збіжностей та подібностей стилю написання дипломних робіт студентів ОКР «Бакалавр» за напрямом 6.050103 «Програмна інженерія» ДНУЗТ–2018 року випуску.

Практичне значення підтверджується впровадженням навчальному процесі Українського державного університету науки і технологій (акт впровадження від 11.12.2023).

Особистий внесок здобувача. Дисертаційна робота є одноосібно виконаною працею. Усі наукові результати, викладені в дисертації, отримані автором самостійно. З наукових праць, опублікованих у співавторстві, у роботі використані лише ті положення та ідеї, які є результатом особистих досліджень здобувача.

Оцінка мови та стилю дисертації. Дисертація написана діловою українською мовою з дотримання стилю викладення результатів проведених досліджень відповідно до встановлених вимог. Стиль викладання наукових

положень і висновків забезпечує доступність сприйняття та осмислення матеріалів дослідження.

3. Наукові публікації, у яких висвітлені основні наукові результати дисертації, та повнота опублікування результатів дисертації

Результати дисертаційної роботи опубліковано в 13 наукових працях. Статті у фахових та рекомендованих Міністерством освіти і науки (МОН) України для публікації результатів дисертацій – 4. Матеріали міжнародних конференцій, що індексуються МНМБ Scopus – 3. У тезах доповідей міжнародних та всеукраїнських конференцій – 6.

Публікації в наукових виданнях, включених до переліку наукових фахових видань України, затверджених МОН України:

Праці у фахових виданнях затверджених МОН України:

1. Shynkarenko, V. I., and Demidovich, I. M.. "Determination of the attributes of authorship of natural texts." *Artificial intelligence* 3 (2018): 27-35.
2. Shynkarenko, V. I., Demidovich, I. M., and Kuropiatnyk, O. S. "A Dual Approach to Establishing the Authority of Technical Natural Language Texts and Their Components." *Science and Transport Progress* 2 (102) (2023): 71-85. doi: 10.15802/stp2023/288958.
3. Shynkarenko, V. I., and Demydovych, I. M. "Methods and software for significant indicators determination of the natural language texts author profile." *Problems in programming* 3 (2023): 22-29. doi: 10.15407/pp2023.03.22
4. Shynkarenko, Viktor, and Demidovich, Inna. "Constructive-synthesizing modeling of natural language texts." *Computer systems and information technologies* 32023, p. 81-91. doi: 10.31891/csit-2023-3-10

Праці включені до міжнародних наукометричних баз (МНБД) Scopus та Web of Science:

5. Shynkarenko, Viktor, and Demidovich, Inna. "Natural Language Texts Authorship Establishing Based on the Sentences Structure." *COLINS*, 2022, p. 328-337.

6. Shynkarenko, Viktor I., et al. " Processing Words Effectiveness Analysis in Solving the Natural Language Texts Authorship Determination Task." IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT), 2021, p. 48-51. doi: 10.1109/CSIT52700.2021.9648829.

7. Shynkarenko, Viktor I., and Demidovich, Inna. "Authorship Determination of Natural Language Texts by Several Classes of Indicators with Customizable Weights." 5th International Conference on Computational Linguistics and Intelligent Systems (COLINS), 2021, p. 832-844.

Матеріали міжнародних наукових конференцій:

8. Шинкаренко, В.І, та Демидович, І.М. . “Статистичний та рекурентний аналіз природньомовних. текстів”. Збірка тез XII Міжнародної науково-практичної конференції “Сучасні інформаційні та комунікаційні технології на транспорті, в промисловості та освіті”, Дніпропетровський нац. ун-т залізничного транспорту ім. акад. В. Лазаряна, 2018, с. 120.

9. Шинкаренко, В.И., та Демидович, И.Н. “Рекуррентный анализ естественно-языковых текстов”. Збірка тез Всеукраїнської науково-методичної конференції «Проблеми математичного моделювання». Дніпропетровський державний технічний ун-т, 2018, с. 40-43.

10. Шинкаренко, В.І, та Демидович, І.М. “Використання генетичного алгоритму для покращення визначення авторства природньомовних текстів”. Збірка тез XIV Міжнародної науково-практичної конференції “Сучасні інформаційні та комунікаційні технології на транспорті, в промисловості та освіті”, Дніпропетровський нац. ун-т залізничного транспорту ім. акад. В. Лазаряна, 2020, с. 127.

11. Шинкаренко, В.І, та Демидович, І.М. “Показатель структурного сходства естественно языкового литературного текста”. Збірка тез XV Міжнародної науково-практичної конференції “Сучасні інформаційні та комунікаційні технології на транспорті, в промисловості та освіті”, Дніпропетровський нац. ун-т залізничного транспорту ім. акад. В. Лазаряна, 2021, с. .

12. Шинкаренко, В.І, та Демидович І.М. . “Застосування формальних стохастичних граматики при визначенні авторства текстів” Матеріали Міжнародної науково-технічної конференції Інформаційні технології в металургії та машинобудуванні. Український державний університет науки і технологій, 2022, с. 293

13. Шинкаренко, В.І, та Демидович, І.М. «Застосування конструктивного моделювання при визначенні авторства текстів» Матеріали Міжнародної науково-технічної конференції Інформаційні технології в металургії та машинобудуванні. Український державний університет науки і технологій, 2023, с.394.

Апробація результатів дисертації. Результати дисертаційної роботи представлено на всеукраїнських та міжнародних науково-практичних конференціях: Інформаційні технології в металургії та машинобудуванні (Дніпро, 2022, 2023), Сучасні інформаційні та комунікаційні технології на транспорті, в промисловості та освіті (Дніпро, 2018, 2020, 2021), Проблеми математичного моделювання (Кам’янське, 2022), International Conference on Computer Sciences and Information Technologies (Львів, 2021), International Conference on Computational Linguistics and Intelligent Systems (Харків, Польща, 2021, 2022).

4. Дані про відсутність текстових запозичень та порушень академічної доброчесності

Під час виконання дисертації аспірантка Демидович І.М. дотримувалась принципів академічної доброчесності, що підтверджено сервісом перевірки робіт на виявлення схожості текстів Unicheck. За результатами перевірки та аналізу матеріалів дисертації не було виявлено ознак академічного плагіату, самоплагіату, фабрикації, фальсифікації.

ВИСНОВОК:

Ознайомившись з дисертаційною роботою Демидович Інни Миколаївни на тему «Розвиток методів та засобів для визначення авторства

україномовних текстів на основі конструктивно-продукційного моделювання» за спеціальності 122-комп'ютерні науки та науковими публікаціями, у яких висвітлено наукові результати, а також враховуючи результат апробації дисертаційної роботи, вважаємо, що:

- 1) дисертаційна робота «Розвиток методів та засобів для визначення авторства україномовних текстів на основі конструктивно-продукційного моделювання» за актуальністю, ступенем новизни, обґрунтованістю та практичною придатністю здобутих результатів є закінченим фундаментальними дослідженням, яке відповідає ОНП «Комп'ютерні науки»;
- 2) дисертаційна робота «Розвиток методів та засобів для визначення авторства україномовних текстів на основі конструктивно-продукційного моделювання» відповідає вимогам Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження доктора філософії, затвердженого постановою Кабінету Міністрів України від 12 січня 2022 р. № 44, і рекомендується до разового захисту у спеціалізованій вченій раді.

Голова засідання,
завідувач кафедри КІТ,
к.т.н., доцент



Вадим ГОРЯЧКІН

Науковий керівник,
д.т.н., професор



Віктор ШИНКАРЕНКО

Секретар, старший
викладач кафедри КІТ



Ірина ШАПОВАЛ