

## РЕЦЕНЗІЯ

на дисертацію здобувача Демидович І. М.

«Розвиток методів та засобів для визначення авторства україномовних текстів на основі конструктивно-продукційного моделювання» подану на здобуття наукового ступеня доктора філософії в галузі знань 12 «Інформаційні технології» за спеціальністю 122 «Комп'ютерні науки»

### **1. Актуальність обраної теми дисертації**

Автоматична атрибуція авторства – це аналіз деякого тексту за допомогою застосування статистичних, математичних та комп'ютерних підходів для визначення автора даного тексту. Вона відноситься до дисципліни, спрямованої на розрізнення автора або автора певного тексту шляхом обробки та аналізу ознак, виділених із вмісту, що розглядається.

За останні два десятиліття автоматична ідентифікація авторства перетворилася на дуже динамічний дослідницький напрямок, що використовує останні досягнення в низці галузей, таких як штучний інтелект, лінгвістика та обчислювальна техніка. Крім того, дослідження автоматичної ідентифікації авторства зараз займаються не лише проблемами авторства в широкій галузі гуманітарних наук (література, історія, богослов'я), але також із застосуванням у різних правоохоронних органах, таких як розвідка, криміналістика.

Для визначення автора тексту часто використовують експертний аналіз, під час якого експерти визначають автора невідомого тексту або встановлюють, чи належить твір іншому автору, використовуючи характерні мовні особливості та стилістичні прийоми. Цей процес ґрунтується на вивченні інших творів цього автора та зіставленні його «почерку».

Незважаючи на те, що безліч підходів застосовні до такого завдання, їх

ефективність залежить від вибору показників, обсягу досліджуваних текстів, розміру викидів та багатьох інших. Через складність самого процесу було створено досить багато підходів та інструментів для вирішення завдання, але жоден з них не призвів до стовідсоткових результатів.

Саме тому глибоке вивчення особливостей українських природньомовних текстів є особливо актуальним, а результати виконаного Демидович І.М., дисертаційного дослідження є цікавими та інформативними з теоретичної та практичної точки зору.

## **2. Зв'язок теми дисертації з державними і галузевими науковими програмами**

Дисертація виконана згідно з планом наукових досліджень Українського державного університету науки та технологій і є частиною наукової роботи кафедри Комп'ютерних інформаційних технологій «Інструментальна підтримка систем обробки природномовних документів» (номер державної реєстрації 0122U002086) та «Моделювання в задачах розробки програмного забезпечення» (номер державної реєстрації 0121U109167), що виконувались впродовж 2021 року. Здобувач була виконавцем фрагменту цієї роботи.

## **3. Новизна дослідження та одержаних результатів**

У дисертаційній роботі Демидович І.М. «Розвиток методів та засобів для визначення авторства україномовних текстів на основі конструктивно-продукційного моделювання» вперше було 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.

Авторкою дисертації вперше розроблено метод встановлення профілю автора який визначає найбільш суттєві показники аналізу текстів, що притаманні певному автору. Цей метод спрощує та полегшує задачу визначення авторства та скорочує потрібний на це час та ресурси.

Уперше 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. Також встановлено статистично значимий зв'язок результатів рішення задач виявлення запозичень та встановлення авторству текстів.

Проведені авторкою експериментальні дослідження та отримані результати підтверджують ефективність запропонованих рішень, висновків та рекомендацій, які були розроблені.

Це свідчення підтримує обґрунтованість та достовірність наукових положень, висновків і рекомендацій, викладених у дисертаційній роботі.

#### **4. Теоретичне та практичне значення одержаних результатів**

Розроблені моделі, методи і технології при комплексному застосуванні дозволяють автоматизувати аналіз природньомовних текстів, а також правильно інтерпретувати та використовувати вирахуванні атрибути авторського стилю для визначення авторства в природньомовних текстах.

Розроблені методи та моделі можуть використовуватись для вирішення практичних завдань в галузі лінгвістики та інших галузях, що потребують роботи з природньомовними текстами та їх аналіз. Зокрема, ці результати можуть бути корисні для пошуку плагіату, автоматичної обробки текстів та їх класифікації при визначенні чи пошуку авторства.

Основні положення та висновки дисертаційної роботи впроваджені у навчальний процес та у наукову роботу кафедри Комп'ютерних інформаційних технологій, що підтверджено відповідним Актом впровадження.

#### **5. Ступінь обґрунтованості наукових положень, висновків і рекомендацій, сформульованих у дисертації**

Достовірність наукових положень, висновків і рекомендацій, сформульованих у дисертаційній роботі Демидович І. М. зумовлено адекватно обраними науково-методичними підходами до вирішення поставленої мети і завдань, достатнім обсягом дослідженого матеріалу, використанням комплексу правильно підібраних сучасних високоінформативних взаємодоповнюючих методів дослідження та відповідним статистичним аналізом отриманих числових показників.

Дослідження проведене на достатній, для отримання достовірних результатів, кількості матеріалу у навчальній та контрольній вибірках.

Ілюстративне підтвердження текстового матеріалу є цілком достатнім. Таблиці та рисунки відображають обсяг проведених досліджень і є високоінформативними.

Аналіз та узагальнення результатів дослідження проведено ґрунтовно у порівнянні з достатньою кількістю вітчизняних та зарубіжних джерел наукової літератури.

Основні результати наукові положення і висновки дисертаційної роботи висвітлені у 13 наукових працях, з яких 4 фахових та рекомендованих Міністерством освіти і науки (МОН) України для публікації результатів дисертацій виданнях, 3 матеріалах міжнародних конференцій, що індексуються МНМБ Scopus та 6 тезах доповідей міжнародних та всеукраїнських конференцій.

Наведене свідчить, що отримані результати проведених досліджень наукові положення, висновки та рекомендації, сформульовані у дисертації, є обґрунтованими та достовірними.

Дисертація викладена на 130 сторінках друкованого тексту, із них 110 сторінок основної частини. Дисертація містить наступні розділи: вступ, аналітичний огляд літератури, методи та моделі досліджень, експериментальні результати, розділ, присвячений розробці інструментарію, що використовувався для проведення експериментів, висновки, список використаних джерел та додатки. Робота ілюстрована 15 рисунками та 22 таблицями. Список використаних джерел складається з 148 найменувань.

Анотація подана українською та англійською мовами відображає основні положення дисертаційної роботи.

У вступі на шести сторінках авторкою дисертації обґрунтовано актуальність дослідження, сформульовано його мету, визначено предмет та об'єкт дослідження, підкреслено нові положення, теоретичне та практичне значення отриманих результатів. Для досягнення мети визначений та окресленого конкретні завдання наведено перелік методів дослідження

Огляд літератури присвячений, аналізу сучасних літературних джерел, які містять дані щодо особливостей підходів, методів та інструментів, що використовуються у визначенні авторства природньомовних текстів для різних мов. Наведено їх недоліки, переваги та результативність.

У другому розділі «Методи та моделі встановлення авторства природньомовних текстів» на 34 сторінках подано опис методів та моделей, що були використані для вирішення задачі встановлення авторству україномовних текстів.

У третьому розділі описані отримані результати роботи кожного з методів та моделей, що були подані вище. Досліджено та проаналізовано їх ефективність.

Четвертий розділ присвячено інструментарію для автоматичного аналізу тексту та порівняння робіт, включаючи конструктори для автоматичної генерації правил та їх порівняння.

У загальних висновках наведені отримані результати дисертаційної роботи, що відповідають поставленій меті та завданням. Висновки узагальнюють результати, одержані дисертанткою при проведенні досліджень, їх опрацюванні та аналізу.

#### **6. Зауваження до змісту та оформленню дисертації:**

1. В результаті проведених експериментів у розділі 3 наведена оцінка ефективності запропонованих моделей та методів, однак, отримані результати не зовсім чітко відображають на скільки запропонований автором підхід краще за раніше існуючі.

2. Автор декларує, що використовує генетичний алгоритм для покращення результатів та визначення значущих параметрів тексту, проте не приводить його виведення у повному обсязі.

3. Доцільно було б більше уваги приділити роботі з технічними текстами, оскільки наразі є актуальною проблема наявності плагіату в академічній сфері.

Виявлені зауваження не стосуються загальної сутності роботи і не знижують її практичного та теоретичного значення.

#### **7. Висновок щодо відповідності дисертації чинним вимогам**

Дисертація Демидович І.М. «Розвиток методів та засобів для визначення авторства україномовних текстів на основі конструктивно-продукційного моделювання» повністю відповідає вимогам наказу МОН України № 40 від 12.01.2017 р. “Про затвердження вимог до оформлення дисертації” та КМ від 12 січня 2022 р. №44 “Порядку проведення експерименту з присудження ступеня доктора філософії та скасування рішення разової спеціалізованої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії” з останніми змінами внесеними постановою КМ №341 від 21 березня 2022.

За актуальністю теми, обсягом дослідженого матеріалу, діапазоном методик і методологією дослідження, повнотою викладу принципів наукових положень, достатньо високим рівнем наукової новизни отриманих результатів, практичним і теоретичним значенням робота відповідає вимогам до дисертації на здобуття ступеня доктора філософії і здобувач Демидович І. М. заслуговує присудження ступеня доктора філософії зі спеціальності 122 Комп'ютерні науки.

Доцент кафедри Комп'ютерні  
інформаційні технології  
Українського державного  
університету науки та технологій,  
к.т.н., доцент

