

ВІДГУК

офіційного опонента про дисертаційну роботу Куроп'ятник Олени Сергіївни «Конструктивно-продукційні моделі природомовних текстів для виявлення запозичень у структурованих документах»,
подану на здобуття наукового ступеня кандидата технічних наук
за спеціальністю 01.05.02 – математичне моделювання
та обчислювальні методи

Актуальність теми дисертації.

У сучасних умовах особливої актуальності набуває питання неправомірних запозичень, перевірки навчальної та науково-технічної діяльності з приводу можливого плагіату. Розширення інформаційного простору породило практику інтелектуальних запозичень, що пов'язано з легким доступом та простою копіювання інформації. Перевірка кваліфікаційних робіт на плагіат є актуальною тенденцією в світовій освіті. Недоліком багатьох антиплагіатних сервісів є те, що вони не враховують структури документів, яка є характерною рисою для кваліфікаційних робіт. Таким чином, актуальною є науково-технічна задача розробки моделей природомовних текстів для виявлення запозичень у структурованих документах.

Про актуальність роботи також свідчить те, що тематика дисертаційного дослідження є частиною науково-дослідних робіт, виконаних на кафедрі комп'ютерних інформаційних технологій Дніпровського національного університету залізничного транспорту ім. академіка В. Лазаряна, зокрема «Прикладне конструктивне моделювання програмних сутностей» та «Підвищення конкурентоспроможності залізничного транспорту на основі уніфікованих інтелектуальних технологій процесів перевезень та експлуатації парків технічних систем», у яких авторка брала участь як відповідальний виконавець і виконавець відповідно.

Структура, задачі та короткий зміст роботи.

Робота складається зі вступу, чотирьох основних розділів, висновків, списку використаних джерел та додатків. Метою роботи є розробка моделей природомовних текстів, методу і засобів виявлення запозичень у структурованих документах з маскуванням.

Для досягнення поставленої мети в роботі були вирішені такі задачі: розроблено конструктивно-продукційну модель мови для формалізації семантичних аспектів порівняння тестів; розроблено конструктивно-продукційну та об'єктно-орієнтовану моделі мовних конструкцій та метод їх порівняння з ура-

хуванням структурних особливостей документів; побудовано модель процесів маскування запозичень та її комп'ютерну реалізацію; розроблено алгоритмічне та програмне забезпечення для виявлення запозичень у структурованих документах.

У першому розділі дисертації проведено детальний аналіз виконано існуючих підходів до моделювання різних аспектів мови, методів, алгоритмів та програмних засобів обробки та зіставлення текстів. На основі аналізу літературних джерел сформульовані завдання дисертаційних досліджень.

У другому розділі запропоновано об'єктно-орієнтоване моделювання смислових конструкцій та конструктивно-продукційне моделювання природної мови, графове представлення тексту, розроблені методи стиснення графа з текстовим навантаженням та порівняння структурованих документів. Побудована модель природної мови, яка дозволяє розглядати мову як конструктивний процес та формально представити класифікації мови. Побудована модель графового представлення тексту. Формальні граматичні та алгоритмічні структури використані як засіб формалізації тексту у вигляді конструкцій з атрибутами. Побудовано модель процесу зіставлення структурованих документів та запропоновано метод порівняння структурованих документів, що дозволяє автоматизувати відповідні процеси.

Третій розділ присвячено експериментальній частині дослідження. Розроблена конструктивно-продукційна модель процесу маскування запозичень. Утворено теоретичну основу формування бази тестів для перевірки здатності демаскування запозичень в програмах виявлення плагіату. На основі сформованої бази тестів виконано тестування комп'ютерної реалізації моделі графового представлення текстів. Виконано дослідження функціональної ефективності застосування моделі графового представлення текстів в задачах виявлення запозичень. Отримані результати порівняно з аналогом.

У четвертому розділі представлено результати розробки об'єктноорієнтованих моделей графового представлення текстів, документів та її програмної реалізації для виявлення запозичень. Встановлено зв'язок алгоритмічної складової моделі графового представлення тексту та логіки програми. Представлено результати розробки трьох додатків.

Наведені у висновках результати досить повно відображають суть дисертаційної роботи і мають практичне втілення.

Основні наукові результати досліджень і новизна дисертації.

- Вперше виконано формалізацію процесів формування образів людини засобами об'єктно-орієнтованого моделювання та розроблено конструктивно-

продукційну модель природної мови;

- вперше розроблено конструктивно-продукційну модель графового представлення текстів та метод їх порівняння для виявлення запозичень із урахуванням структурних особливостей документів;

- вперше побудовано конструктивно-продукційну модель процесів маскування запозичень.

Отримали подальший розвиток:

- методи конструктивно-продукційного моделювання, які дозволили розробити комп'ютерні реалізації моделі графового представлення тексту та моделі процесів маскування;

- методи виявлення семантичних запозичень у текстах, які дозволили зменшити вплив зміни лексичної та синтаксичної структури;

- методи обробки та стиснення графів.

Практичне значення та рекомендації щодо використання результатів дисертації.

Теоретичні розробки дисертації доведені автором до конкретних методів, розроблено алгоритмічні та програмні засоби виявлення запозичень у структурованих документах на основі моделі графового представлення тексту. Вони були використані для перевірки наявності запозичень у структурованих документах. Розроблена модель процесів маскування дозволила отримати інструментарій для отримання бази тестів для програм-антиплагіатів.

Також практичне значення результатів роботи підтверджується актом впровадження та прийняття у дослідну експлуатацію програмного засобу «Система виявлення запозичень у цифровому представленні структурованих документів» в Дніпровському національному університеті ім. академіка В. Лазаряна; актом впровадження в філії «Проектноконструкторське бюро інформаційних технологій» АТ «Українська залізниця» при проектуванні та розробці алгоритмів та програмних засобів порівняння та видалення тотожних частин документів; актом впровадження результатів дисертаційної роботи в ТОВ «СОВЛАНУТ» при виконанні проекту «Розробка системи виявлення запозичень»; та двома свідоцтвами про реєстрацію авторського права на твір.

Результати дисертаційної роботи доцільно застосовувати в системах перекладу і антиплагіату, експертних системах, а також у різноманітних додатках, де є обробка інформації, представлена природною мовою.

Обґрунтованість та достовірність одержаних результатів.

Отримані результати мають практичне втілення і піддаються чіткій інтерпретації, що свідчить про обґрунтованість наукових положень, висновків та ре-

комендацій дисертаційної роботи. Наукові положення дисертації мають належне теоретичне обґрунтування. Слід зазначити коректне застосування автором методів теорії графів, теорії множин, формальних граматики, статистичних методів. Також достовірність одержаних результатів підтверджується зіставленням сучасних наукових і технічних досягнень в області моделювання мов і розробки алгоритмів та систем виявлення запозичень; апробацією основних теоретичних і експериментальних результатів роботи в друкованих працях та доповідях на конференціях.

Оформлення дисертації та автореферату. Дисертаційна робота написана зрозуміло і грамотно, науково-технічна термінологія використовується коректно. Структура та обсяг дисертації відповідають встановленим вимогам на здобуття наукового ступеню кандидата технічних наук. Дисертацію написано з використанням сучасної бібліографії та наукової термінології. Зміст та результати досліджень викладено лаконічно та аргументовано. Стиль викладення матеріалів дисертації та автореферату логічний. Зміст автореферату повністю відповідає основним положенням та висновкам, зробленим у дисертації. Дисертаційна робота відповідає паспорту спеціальності 01.05.02 – математичне моделювання та обчислювальні методи.

Повнота викладення наукових положень, висновків і рекомендацій в опублікованих працях.

За результатами дисертаційного дослідження опубліковано 24 наукових праці, в тому числі 7 статей у журналах, рекомендованих МОН України для публікації результатів дисертацій, та закордонних виданнях; у тезах доповідей та трудах міжнародних та всеукраїнських конференцій – 15. Отримано 2 свідоцтва про реєстрацію авторського права на твір.

Зауваження.

Серед недоліків дисертації слід зазначити такі.

1. Основний текст дисертації погано структурований всередині розділів, містить велику кількість формул і визначень, поданих у порядку перерахування без обговорення, що дуже ускладнює сприйняття і розуміння роботи. Не виділено основні положення і пункти, які зв'язують подання дослідження воедино. Висновки за кожним розділом також є наскрізним текстом. В тексті роботи не виділено, які положення вже були, а які запропоновано автором. Все перераховане робить текст дисертації погано читабельним.

2. Розділ 2 в дисертаційній роботі і відповідне його подання в авторефераті є перевантаженим, в порівнянні з іншими розділами. У другому розділі зіб-

рано більшу частину запропонованої автором наукової новизни. З урахуванням того, що четвертий розділ переважно містить опис програмної реалізації, який можна винести у додатки, було б доцільним збалансовано перерозподілити за розділами основний текст.

3. У тексті Висновків знаходиться рисунок В1, де представлено взаємозв'язок на різних рівнях запропонованих автором моделей та їх реалізацій. Якщо даний рисунок із відповідними поясненнями помістити в початок роботи, то це значно допомогло би розумінню суті дисертаційного дослідження, його результатів і практичної реалізації.

4. В роботі досить часто зустрічаються рисунки та формули, компоненти яких були описані значно раніше, але посилань на цей опис не наводиться. Так, наприклад, рисунок 2.8, знаходиться на стор. 70, а розшифровку часті позначень наведено на стор. 61; або компоненти формули 3.5, поданої на стор. 113 було приведено на стор. 82 без жодного посилання.

5. На стор. 11 автореферату подано 20 алгоритмів опису операцій у вигляді формульних позначень, математичний сенс яких складно зрозуміти, оскільки частина позначень не приведена.

6. Підрозділи 3.2 та 3.3 зовсім не пов'язані між собою: в 3.2 розглядається структуровані документи на основі структури бакалаврської роботи, а в 3.3 проведено аналіз текстових фрагментів англійською мовою. Було б доцільно і доречно, якби дослідження були проведені на однотипному матеріалі.

7. При побудові конструкції образів речень та слів в таблицях 3.4-3.7 основними величинами є значення «фокус» і «момент часу». Сенс цих понять в тексті не розкритий та їх значення не пояснені.

8. В експериментальних дослідженнях ефективності методу стиснення графа (експерименти 3.1-3.5, 3.7) була використана експериментальна база - 64 текстових файлів за тематикою «Розробка ПЗ», а для експериментів 3.6 і 3.8 - 25 текстових файлів, які є документацією до дипломних проектів ОКР «Бакалавр» за напрямом «програмна інженерія». Було б доцільним і статистично коректним провести експерименти на більш різноманітній за тематикою і стилю написання вибірці даних. Це зауваження також стосується всіх експериментів, результати яких представлені у 3-му розділі.

9. Було б доцільно розширити експерименти щодо дослідження функціональної ефективності комп'ютерної реалізації запропонованих методів та дослідження до впливу маскування запозичень. А саме: провести порівняння не з одним, а з декількома аналогами на різноманітних вибірках даних, а також при дослідженні різних сценаріїв маскування зіставити результати роботи розробленого програмного продукту з результатами інших аналогів.

Наведені зауваження мають окремий характер, не знижують високий науковий рівень дисертаційної роботи і не впливають на її загальну позитивну оцінку.

Висновок. Розглянувши дисертаційну роботу Куроп'ятник Олени Сергіївни «Конструктивно-продукційні моделі природомовних текстів для виявлення запозичень у структурованих документах», автореферат, опубліковані наукові праці та додаткові матеріали, можна зробити такі висновки:

- дисертація відповідає паспорту спеціальності 01.05.02 – Математичне моделювання та обчислювальні методи;
- тематична спрямованість роботи є актуальною, суспільно корисною й перспективною у плані продовження розпочатих досліджень;
- зміст автореферату відповідає основним положенням дисертаційної роботи;
- дисертація є цілісною, завершеною, оригінальною, самостійною кваліфікаційною науковою працею.

Вважаю, що дисертаційна робота є завершеною науково-дослідницькою роботою, в якій отримано нові науково обґрунтовані результати, що в сукупності вирішують актуальну наукову-технічну задачу виявлення запозичень у структурованих документах кваліфікаційного характеру, та відповідає вимогам п. п. 9, 11, 12 «Порядку присудження наукових ступенів», які висуваються до кандидатських дисертацій, а її автор, Куроп'ятник О. С., заслуговує на присудження наукового ступеня кандидата технічних наук за спеціальністю 01.05.02 – математичне моделювання та обчислювальні методи.

Офіційний опонент
професор кафедри
прикладної математики
Харківського національного
університету радіоелектроніки,
доктор технічних наук, професор

Кіріченко Л. О.

Підпис проф. Кіріченко Л. О. засвідчую.

Учений Секретар
Харківського національного
університету радіоелектроніки



I. В. Магдаліна

«27» 05 2020 р.

Відгук надійшов
у Раду: 04.06.2020
Вчений секретар: