

ВІДГУК

офіційного опонента, завідувача кафедри комп'ютерних інформаційних технологій Дніпровського національного університету залізничного транспорту імені академіка В. Лазаряна, д-ра технічних наук, професора Шинкаренка Віктора Івановича, на дисертацію Ковиліна Єгора Романовича «Модель генерації відповідей в пошукових системах на основі неструктурованої бази знань», подану на здобуття наукового ступеня кандидата технічних наук за спеціальністю 01.05.02 – математичне моделювання та обчислювальні методи

Актуальність теми дисертації. Перманентне зростання обсягів та різноманіття інформації у сучасних інформаційних системах потребує наявності методів її обробки та постійного підвищення їх ефективності – в принципі, можна просто додати слово «методи». У першу чергу це стосується пошуку ревалентної інформації. Задача значно ускладнюється коли мова йде про природномовні тексти, та й ще на таких складних як українська.

Звісно, є досить ефективні пошукові системи у мережі Internet, але вони потребують подальших досліджень та вдосконалення. Нерідко пошукова система надає у якості результату мільйони документів, при цьому необхідні можуть бути з невисоким рангом.

Широке поле для досліджень залишається щодо задачі пошуку інформації у локальних сховищах природно мовних документів. Наприклад, сховища лабораторних, курсових та дипломних робіт.

Беручи до уваги на постулат: чим більше ми знаємо про об'єкт чи процес, тим ефективніше ми можемо з ним працювати та керувати, пошукові системи мають спиратись на синтаксичні та семантичні властивості текстів.

На вирішення науково-практичної задачі генерації ревалентних відповідей пошукових систем з урахуванням деяких семантичних властивостей і спрямована дана дисертаційна робота. При цьому автор намагається уникати зовнішніх засобів таких як словники, глобальні бази знань, тощо.

Автором роботи пропонується модель пошукової системи, яка не тільки спроможна генерувати нові знання, чим спрощує процес пошуку інформації для кінцевого користувача, а і дозволяє оброблювати семантично неструктуровані документи без залучення великих об'ємів лінгвістичних знань та ручного складання семантичних розміток, завдяки чому значно спрощується створення семантично-орієнтованих моделей та систем.

Дисертаційна робота виконана у відповідності з планом науково-дослідних робіт кафедри комп'ютерних наук та інформаційних технологій Дніпровського національного університету імені Олеся Гончара. Робота дисертанта пов'язана з НДР «Методи та інформаційні технології цифрової обробки багатоканальних даних» (реєстраційний номер 0116U001297).

Достовірність та обґрунтованість основних висновків і результатів роботи. У дисертації приділено достатньо уваги обґрунтуванню наукових положень, висновків і рекомендацій.

Обґрунтованість і достовірність наукових результатів, висновків і рекомендацій, викладених в дисертаційній роботі, досягаються ретельним системним аналізом процесу розроблення математичних та алгоритмічних моделей, використанням апробованих та відомих методів

Для отриманих моделей вперше розроблено і виконано систему оцінок адекватності, яка складається з:

- перевірки залежності розробленої семантичної моделі тексту від семантичних характеристик тексту, яка вказує на універсальність моделі і можливість її застосування для розробок великої кількості семантично-орієнтованих систем;
- оцінки висунутої гіпотези про семантичні мітки тексту, яка полягає в обчисленні значень семантичної місткості тематично важливих термінів у кластерах розробленої семантичної моделі;
- перевірки точності автоматичної класифікації моделі фільтрації вхідних даних за критерієм семантичної зв'язності;
- оцінки згенерованих відповідей за бальним методом, вирахованої на основі оцінок, виставлених операторами-експертами;
- оцінки якості формування моделлю множини текстів-кандидатів до включення у результуючу відповідь, з точки зору її тематичної відповідності до вхідного запиту.

Достовірність результатів роботи також підтверджується численними доповідями на наукових конференціях, комп'ютерним моделюванням, а також результатами, які відображені у документах впровадження.

Наукова новизна результатів дисертації. Загалом погоджуюсь з викладеним автором що в роботі вперше отримано такі результати:

- розроблено семантичну модель текстових даних, яка на відміну від існуючих аналогів, дозволяє отримувати кількісні показники семантичних властивостей і сенсові зв'язки між компонентами тексту без необхідності будь-якої попередньої семантичної розмітки, впровадження словників або залучення лінгвістичних знань;
- створено модель автоматичної класифікації знань за ступенем їх семантичної зв'язності, що використовує числові дані, отримані із розробленої семантичної моделі тексту, яка дозволила збільшити надійність використання моделі генерації відповідей шляхом верифікації первинної інформації;
- побудовано модель автоматичної генерації відповідей у пошуковій системі із неструктурованої бази текстових знань на основі створених моделей, яка дозволила автоматизувати роботу користувача із пошуковими системами;

- розроблено систему оцінок адекватності створеної моделі генерації відповіді на основі неструктурованої бази знань та отримали подальший розвиток:
- семантичні моделі текстових даних для флективно багатих мов із вільним порядком слів: розроблена семантична модель текстових даних дозволяє уникнути процесу ручного опису семантичної структури документа;
- методи організації пошуку інформації: створені моделі дозволяють генерувати релевантні до запиту користувача знання на основі неструктурованої бази знань, чим спрощують роботу користувача із пошуковими системами.

Але зауважу, що деякі пункти пов'язані між собою та стосуються результатів тільки цієї роботи.

Наукове і практичне значення роботи. Роботу відрізняє інноваційність розроблених моделей та нестандартне використання відомих підходів математичного моделювання для побудови нових моделей та методів автоматичної обробки текстів.

На основі теоретичних напрацювань побудовано комп'ютерну систему, набір програмних інструментів якої дозволяє створювати та обробляти тематичні повнотекстові бази знань без попередньої семантичної розмітки і будувати програмну семантичну модель текстових знань, що дозволяє вирішувати задачу пошуку релевантних документів та інші завдання автоматичної обробки текстів без необхідності залучати будь-які лінгвістичні знання.

Розроблений програмний додаток впроваджено у міському комунальному закладі культури «Централізована система бібліотек для дітей» м. Дніпро як пошуковий інструмент обробки електронних текстів, у ТОВ «Сітал Україна» як засіб автоматичної генерації текстових інструкцій та у АТ «ДніпроАзот» як інструмент покращення процесів пошуку в системах електронного документообігу.

Оцінка змісту дисертації, її завершеності в цілому. Дисертація складається із анотацій, змісту, вступу, трьох розділів, висновків, списку використаних джерел і дев'ятнадцяти додатків. Робота в цілому має завершений вигляд. Мета та завдання дослідження досягнуті. Зміст та об'єм досліджень дисертації повною мірою розкривають її сутність. Побудова дисертації відповідає прийнятним для наукового дослідження рекомендаціям.

Аналіз змісту дисертації, її завершеності й оформлення. Побудова дисертації відповідає прийнятним для наукового дослідження рекомендаціям. Дисертація складається із анотації двома мовами, змісту, вступу, трьох розділів, висновків, списку використаних джерел і дев'ятнадцяти додатків.

У вступі обґрунтовано актуальність теми дослідження та наукових завдань; наведено інформацію про зв'язок роботи з науковими темами; сформульовано мету й завдання дослідження; розкрито наукову новизну, практичне значення отриманих результатів; вказані методи дослідження, на-

ведено відомості про апробацію, публікації та впровадження результатів дослідження.

У *першому розділі* проаналізовано існуючі математичні моделі текстів та існуючі аналоги моделі генерації текстів на основі неструктурованої бази знань у пошуковій системі. Здійснено і обґрунтовано вибір базової концепції моделі та зазначені основні складнощі та наукові задачі, які пов'язані із її використанням у пошукових системах. Сформульована модель генерації відповідей із неструктурованої бази знань та основні кроки її роботи.

Особливу увагу автор приділив теорії «Сенс ↔ Текст». Розглянуті основні переваги та недоліки обраних теорій з точки зору їх застосування до задачі автоматичної генерації відповідей у пошуковій системі.

Слід відмітити, що розділ дещо переобтяжено, можливо було б зменшити обсяг, залишившись в межах вимог до обсягу кандидатської дисертації.

Другий розділ є основою дослідження. У ньому викладені основні надбання роботи: семантична модель текстового документу і метод її формування. Наведені результати автоматичної побудови семантичної моделі документа та проведена перевірка адекватності розробленого підходу.

Оригінальний метод формування латентно-семантичної моделі текстового документу включає: на основі відстані Левенштейна приведення однокореневих слів до деякої загальної форми – стеми; формується матриця, рядки якої відображають речення, стовпці – стеми, а значення – число входжень стеми в речення; виконується операція сингулярного розкладання матриці; видаляються значення лівої і правої матриці, які відповідають низьким сингулярним значенням, залишаються тільки два найбільших, що представляють собою координати, які проектуються на двомірну площину як масив точок для стем і для речень; проводиться кластеризація отриманих точок-координат для стем і для речень за алгоритмом k-means; формуються семантичні контури у двомірній площині; кластер речення приєднується в моделі до кластеру-стеми, де вага зв'язку – це кількість точок, що потрапили в контур кластера-стеми.

У *третьому розділі* представлена модель генерації відповідей у пошукових системах на основі неструктурованої бази знань. Розкриті структура і алгоритм роботи створеної моделі пошукової системи із використанням генерації текстів, показані архітектури програмного додатку і створеної неструктурованої бази знань, а також проведено тестування моделі генерації відповідей на основі неструктурованої бази знань в пошукових системах і перевірена її адекватність. Наведено приклади отриманих за допомогою системи відповідей та проведені їх аналіз і оцінка.

У *висновках* стисло сформульовані основні наукові і практичні результати дисертаційної роботи з якими цілком можна погодитись.

Відповідність змісту автореферату основним положенням дисертації. Зміст автореферату відповідає змісту дисертації. Автореферат не містить надмірних подробиць та не містить інформації, якої немає в дисертації. Автореферат оформлено згідно до вимог і рекомендацій МОН України.

Повнота викладених результатів дисертації в опублікованих працях. Положення дисертації у повній мірі викладені в 14 наукових працях, в тому числі 8 статей у журналах, рекомендованих МОН України для публікації результатів дисертацій, та закордонних виданнях: «Математичне моделювання» – 1 (індексована у НМБ Index Copernicus), «Системні технології. Регіональний збірник міжвузівських наукових праць» – 5 (НМБ Index Copernicus), «Вісник Херсонського національного технічного університету» (НМБ Google Scholar) – 1, «Modern engineering and innovative technologies» (Німеччина, НМБ Index Copernicus) – 1; у тезах доповідей та трудах міжнародних та всеукраїнських конференцій – 6 (у тому числі 1 – у НМБ Scopus).

Слід відзначити якісне та повне висвітлення матеріалів дисертаційного дослідження у зазначених публікаціях та апробацію результатів на міжнародних конференціях. Зміст публікацій відповідає темі дисертаційної роботи та в достатньому обсязі відображає її основні положення та наукові результати.

При загальному позитивному ставленні до роботи, слід відмітити, що вона не вільна від зауважень.

Зауваження до дисертації.

1. У дисертації є декілька недостатньо обґрунтованих рішень щодо:
 - кількості кластерів-стем для побудови основи семантичної моделі документа обирається за деякою формулою, а не визначається за якістю результатів;
 - розмірності простору кластеризації – обрана 2, а чому не 3 і більше;
 - формування відповідей із відібраних документів.
2. На рисунках 2.21–2.24 дисертації надані графіки з не упорядкованими даними по осі абсцис. Не зрозуміло як їх інтерпретувати.
3. Не зрозуміло яким чином отримані дані у табл. 2.3, 3.3, 3.4 з точністю до 14 знака після коми і яка їх дійсна точність.
4. У авторефераті недостатньо висвітлене питання яким чином формується відповідь на запит за визначеними документами.
5. Не зрозуміло яким чином використовується нейронна мережа для класифікації документів, якщо вона має лише один вихід.
6. Не визначені показники часової ефективності або обчислювальної складності розробленої програми, що не знімає питання її працездатності при значному збільшенні кількості документів та їх обсягів.
7. Розрахунок визначеного у роботі коефіцієнту семантичної значущості моделі залежить від порядку обробки документів. Не вказано як враховується цей факт.

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

Висновок.

Вважаю, що дисертаційна робота Ковиліна Єгора Романовича «Модель генерації відповідей в пошукових системах на основі неструктурованої бази знань» є завершеним науковим дослідженням, містить важливі наукові та

практичні результати. У дисертаційній роботі вирішено актуальну науково-практичну задачу ефективної генерації відповідей в пошукових системах.

Робота відповідає напрямам досліджень паспорту спеціальності 01.05.02 – математичне моделювання та обчислювальні методи (технічні науки).

Дисертаційна робота за актуальністю, науковою новизною та практичною цінністю відповідає вимогам пп. 9, 11, 12 «Порядку присудження наукових ступенів», затвердженого постановою Кабінету Міністрів України від 24 липня 2013 р. № 567 (зі змінами, внесеними згідно з Постановами № 656 від 19.08.2015, № 1159 від 20.12.2015, № 567 від 27.07.2016 та № 943 від 20.11.2019), які висуваються до кандидатських дисертаційних робіт, а її автор, Ковилін Єгор Романович, заслуговує присудження наукового ступеня кандидата технічних наук за спеціальністю 01.05.02 – математичне моделювання та обчислювальні методи.

Офіційний опонент

завідувач кафедри комп'ютерних інформаційних технологій
Дніпровського національного університету
залізничного транспорту ім. ак. В. Лазаряна,
доктор технічних наук, професор

Шинкаренко В.І.

Підпис проф. Шинкаренка В.І. засвідчую.

Учений Секретар

Дніпровського національного університету
залізничного транспорту ім. ак. В. Лазаряна

Радкевич Т.О.

«27» травня 2020 р.

