

РЕЦЕНЗІЯ

на дисертаційну роботу Літвін Анни Андріївни
**«Методи та засоби аналізу і синтезу природномовних текстів на основі
онтологічної моделі предметної області
в аспекті створення діалогових систем»**,
подану на здобуття наукового ступеня доктора філософії за спеціальністю
122 «Комп'ютерні науки»

1. Актуальність обраної теми дисертації

Враховуючи стрімкий розвиток технологій автоматизованої обробки природномовних текстів і широке їх застосування у різноманітних сферах діяльності, актуальність теми дослідження Літвін А. А. не викликає жодного сумніву. Напрямок, який обрала авторка дослідження, має суттєве значення для підвищення ефективності роботи діалогових і довідкових інформаційних систем із природномовним інтерфейсом – віртуальних помічників, чат-ботів, рекомендаційних підсистем.

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

Дисертаційна робота виконана у відповідності з планами науково-дослідних робіт Інституту кібернетики імені В. М. Глушкова НАН України:

ДКР «Сфера» (2019 р.);

«Трансдисциплінарна інтелектуальна інформаційно-аналітична система супроводження процесів реабілітації при пандемії (TISP)», 0221U105840 (2020–2021 рр.);

«Створення методів та технологічних засад формування інтерактивних баз знань», 0117U000005 (2021 р.).

3. Основні результати та їх наукова новизна

Робота розглядає три взаємопов'язані класи задач: семантичний аналіз природномовних текстів, автоматизована побудова онтологічних графових баз знань і синтез осмисленого тексту на основі онтологічної моделі. Наукова новизна полягає в розробці оригінальних моделей семантичного аналізу мови

флексивного типу, автоматизованого створення онтологічних баз знань та синтезу природномовних відповідей. Здобутки включають методи поєднання великих мовних моделей і онтологічного підходу для метанавчання, що підвищує їх ефективність у спеціалізованих областях.

Теорія семантичних відношень розвинута, включаючи виведення семантичних категорій з аксіоматичного набору понять верхнього рівня, надаючи основу для онтологічного представлення семантичних категорій.

Розроблено метод отримання формальних запитів з використанням гнучких шаблонів на основі аналізу природномовних повідомлень користувачів. Далі розвинено методи автоматичної побудови онтологій та генерації природномовних відповідей на основі результатів формальних запитів. Розроблено підхід метанавчання великої мовної моделі на основі структурованих інструкцій-підказок з використанням мета-онтології.

Результати можуть застосовуватися для створення діалогових інформаційних систем в різних областях, забезпечуючи їх ефективність і релевантність. Розроблене програмне забезпечення розгорнуте на сервері і має інтероперабельні можливості.

4. Практичне значення отриманих результатів

Робота має важливе практичне значення у розробці методів побудови онтологій на основі природномовних текстів. Створені онтології відповідають стандартам і використовуються як бази знань для діалогових та довідкових систем в різних галузях. Розроблені методи семантичного аналізу і синтезу природномовних текстів використовуються в системі OntoChatGPT, яка використовує великі мовні моделі, такі як API ChatGPT, і онтологічну базу для мета-навчання та вдосконалення результатів у спеціалізованих областях.

Розроблені в рамках дисертації методи аналізу і синтезу природномовних текстів дозволяють автоматизовано побудувати графові бази знань і автоматизовано створювати запити до них. Результати є

суттєвим внеском у методологію семантичного аналізу та синтезу текстів, зокрема у розробку діалогових і довідкових систем із онтологічними базами знань. Робота дозволяє створювати ефективні та релевантні системи в різних сферах, включаючи реабілітаційну медицину.

5. Ступінь обґрунтованості основних положень та висновків дисертації

Авторка чітко сформулювала та детально розглянула задачі, що постали перед сучасними програмними засобами для семантичного аналізу і синтезу природномовного тексту, а також онтологічної інженерії.

Основні наукові положення, висновки, рекомендації дисертаційної роботи мають належне теоретичне, методологічне та практичне обґрунтування. Обґрунтованість положень дисертації базується на системному дослідженні здобувачем великої кількості актуальних вітчизняних і зарубіжних джерел, які відповідають темі дисертації, а також на проведених обчислювальних експериментах.

Структура дисертаційного дослідження логічна, послідовна і у повному обсязі розкриває тему.

6. Повнота викладення наукових положень та висновків в опублікованих працях

Сутність основних отриманих результатів виконаного дослідження ті їх наукова новизна достатньо відображені у 12 наукових публікаціях, п'ять з яких входять до науко метричних баз Scopus та Web of Science.

Результати проведеного дослідження доповідалися на 10 конференціях, в тому числі міжнародних.

Повнота викладення наукових положень і висновків в опублікованих працях дозволяє науковій спільноті отримати повне уявлення про результати наукової роботи.

7. Недоліки дисертації щодо її змісту та оформлення

Суттєвих недоліків щодо змісту і оформлення дисертації не вбачаю, але варто навести деякі зауваження.

1. Має місце певне повторення матеріалу, викладеного у підрозділах 2.4.2.3 і 3.2.2.1, що, можливо, зручно для читання, коли необхідні дані і формули розміщуються поруч, але повторення було б кориснішим у дещо інших термінах і стислішому вигляді.

2. Нумерація формул (2.2)–(2.7) на с. 94, в принципі, необов'язкова, бо в тексті відсутні посилення на конкретні формули з цього набору, але є посилання на всю групу цих формул.

3. Нумерація посилань у списку літератури не зовсім відповідає порядку їх першого згадування у тексті.

4. Система для конвертації форматів зберігання опису онтології згадується лише у підрозділі 4.5. Немає заперечень щодо практичної важливості даного інструменту, але слід докладніше вказати його місце у роботі в цілому.

5. Відсутній додаток із списком публікацій здобувача за темою дисертації та відомостями про апробацію результатів дисертації (азначаються назви конференції, конгресу, симпозіуму, семінару, школи, місце та дата проведення, форма участі), але ця інформація в основному міститься на с. 4–5.

6. Відсутній перелік скорочень.

7. Термінологічні зауваження стосуються виразів «видобуток тексту» (с. 20), «біаси» (с. 36, 39, 41), українського перекладу виразів «baseline» (с. 39), «separatibility of an action» (с. 82), вживання терміну «метрика» (с. 122). Подібні зауваження стосуються більшості документів щодо сучасного ПЗ.

8. Трапляються пропуски букв (с. 17, 30), неточності букв (с. 59, 212), описки (с. 197), неточності формул (с. 206).

9. Нумерація титульної сторінки є зайвою. Також бажано вказувати на титульній сторінці галузь знань.

10. Бажано вказати, що велика мовна модель ChatGPT підтримує в тому числі й українську мову. Варто уточнити відмінності векторної бази даних від графової.

11. Один із сучасних напрямків досліджень методів обробки природномовних текстів орієнтований не на виокремлення онтологій та семантичних взаємозв'язків, виокремлення інших атрибутів, проміжних етапів моделювання, а на побудову моделей нейронних мереж, що були б навчені на великих обсягах текстових даних. Такі великі моделі, маючи трансформерну архітектуру, відразу навчаються на оригінальному вході та дають фінальний вихід. Альтернативна імплементація діалогової системи могла б бути реалізована на основі бібліотеки LangChain та технології RAG (Retrieval Augmented Generation), що дозволило б кінцевому користувачу завантажувати файли pdf чи html з його фахової спеціальності та отримувати на їх основі відповіді на питання від LLM (великої мовної моделі) або машинне узагальнення текстової інформації без попередньої онтологічної чи семантичної обробки.

12. Метрики Accuracy, Recall та Precision орієнтовані на задачі класифікації, були використані в підрозділі 5.1 для природномовних текстів, по суті оцінюючи релевантність/нерелевантність відповіді. На практиці відповіді систем на основі природномовних текстів оцінюють також за критеріями змістовності, повноти, корисності тощо.

Наведені зауваження в жодному разі не впливають на загальну позитивну оцінку дисертаційної роботи.

8. Рекомендації щодо впровадження результатів дисертаційного дослідження в практику

Отримані результати можуть бути застосовані для розробки та вдосконалення спеціалізованих діалогових і довідкових інформаційних програмних систем як у приватних компаніях, так і у державних установах.

Рекомендації для впровадження результатів дисертації можуть включати створення додаткового інструментарію для побудови онтологічних баз знань з огляду на особливості предметної області, розробку додаткових програмних інтерфейсів взаємодії для ширшої інтеграції з різними сервісами обміну повідомленнями та інформаційно-аналітичними системами, адаптованих під специфічні потреби різних галузей.

9. Відповідність дисертації встановленим вимогам

Дисертаційна робота Літвін Анни Андріївни «Методи та засоби аналізу і синтезу природномовних текстів на основі онтологічної моделі предметної області в аспекті створення діалогових систем» на здобуття наукового ступеня доктора філософії за спеціальністю 122 «Комп'ютерні науки» є вагомим внеском у розвиток обраної тематики. Ефективність запропонованих моделей і методів перевірена на основі аналізу обчислювальних експериментів.

Таким чином, дисертація Літвін А. А. за актуальністю, обсягом і рівнем 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 №4 4 від 12 січня 2022 року, а її авторка Літвін Анна Андріївна заслуговує присудження ступеня доктора філософії за спеціальністю 122 «Комп'ютерні науки».

Рецензент

завідувач відділу інтелектуальних інформаційних технологій
Інституту кібернетики імені В. М. Глушкова
Національної академії наук України
доктор фізико-математичних наук,
старший науковий співробітник



Підпис *В. Горбачук*
ЗАСВІДОК
Зав. канц. *Степан Степаненко*
ІК НАН України

Василь ГОРБАЧУК