

ВІДГУК

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” на здобуття наукового ступеня доктора філософії за спеціальністю 122 “Комп’ютерні науки”,

доктора технічних наук Погорілого Сергія Дем’яновича

1. Актуальність теми дисертаційної роботи.

Представлені у роботі дослідження зосереджені на актуальному напрямку розвитку комп’ютерних наук, яким є автоматизована обробка природномовних текстів, зокрема їх семантичний аналіз і синтез. На даний час використання автоматизованої обробки природномовних текстів набуває все більшого поширення у різних сферах діяльності. Відповідно, розглянуті у роботі питання автоматизованої побудови баз знань, формальних запитів до них та синтезу відповідей на базі результатів запитів є ключовими для ефективного її застосування.

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

Результати, отримані у рамках виконання роботи Літвін А. А. використовувались у науково-дослідних роботах «Трансдисциплінарна інтелектуальна інформаційно-аналітична система супроводження процесів реабілітації при пандемії (TISP)» (№ 159/01/0245), що виконувалася у 2020 – 2021 роках, ДКР «Сфера» (Державний контракт № 181/18/3.), що виконувалася у 2019 році, «Створення методів та технологічних засад формування інтерактивних баз знань» (№ 0117U000005), що виконувалася у 2021 році.

3. Обґрунтованість та достовірність отриманих результатів.

Результати дисертації ґрунтуються на коректних постановках виділених типів задач автоматизованого семантичного аналізу і синтезу природномовного тексту. Достовірність результатів підтверджується результатами обчислювальних експериментів, проведених з метою дослідження ефективності роботи розроблених програмних діалогових систем, що використовують запропоновані методи аналізу природномовного тексту, а також екскрементами з синтезу природномовних висловлювань за онтологічною моделлю. Результати експерименту супроводжуються визначенням числових формальних критеріїв оцінки.

4. Основні результати та їх наукова новизна.

Наукова новизна роботи полягає в наступному:

- отримала подальший розвиток теорія систем семантичних відношень, зокрема, розроблено підхід для виведення різноманітних семантичних категорій з аксіоматичного набору понять верхнього рівня. Метод дозволяє створити обґрунтовану основу для онтологічного представлення семантичних категорій при автоматизованій побудові графових баз знань;

- вперше створено метод отримання формальних запитів мовами SPARQL і Cypher за аналізом природномовних текстів, адаптований до мов флективного типу, зокрема, української. Запити будуються автоматично на основі гнучких шаблонів, які автоматично налаштовуються за визначеними у вхідному тексті семантичними категоріями, та сутностями, що їх конкретизують;

- отримали подальший розвиток методи автоматичної побудови онтологій на основі природномовного тексту, на базі глибокого семантичного аналізу ефективні як для мов флективного типу, зокрема, української, так і для англійської мови;

- отримали подальший розвиток методи для генерації відповідей за допомогою шаблонів гнучкої структури на основі набору понять, отриманих шляхом виконання формального запиту до бази знань;

- вперше розроблений підхід метанавчання великої мовної моделі, який базується на застосуванні структурованих інструкцій-підказок для аналізу вхідних даних і контекстів, що вибираються із онтологічної бази знань. При цьому формування зазначених інструкцій відбувається автоматично із застосуванням мета-онтології.

5. Практичне значення отриманих результатів.

Розроблені методи створення онтологій на основі структурованих та неструктурованих природномовних текстів, дозволили створити бази знань для різних областей: листування, реабілітаційна медицина, а також та тестові онтології із різних галузей. Онтології відповідають стандартам, можуть використовуватися з графовими СУБД (Apache Jena Fuseki, Neo4J, RDFlib) та стати основою для діалогових та довідкових систем.

На основі семантичного аналізу та синтезу природномовних текстів розроблено онтологокеровані діалогові системи у сферах медицини, економіки, екології, листування, педагогіки та інших.

Створено комбіновану систему OntoChatGPT, що використовує великі мовні моделі (API ChatGPT) та онтологічну базу знань для мета-навчання. Мета-онтологія дозволяє створювати структуровані інструкції-підказки для великої мовної моделі, розширюючи її можливості і збільшуючи релевантність відповідей.

Розроблено засіб для конвертації формату онтологій між Graph Editor та OWL у вигляді RDF/XML. Це дозволяє ефективно використовувати створені вручну у графічному інтерфейсі онтології у графових СУБД, таких як Apache Jena Fuseki та Neo4J.

6. Повнота викладення наукових положень та висновків в опублікованих працях.

Наукові положення та висновки викладені в повному обсязі як а опублікованих Літвін А. А. працях у фахових наукових виданнях, так і власне матеріалах дисертації.

7. Недоліки дисертації щодо її змісту і оформлення.

1. Структура дисертації, на погляд опонента, є надто деталізованою, про що свідчить наявність великої кількості підрозділів, пунктів і підпунктів, в тому числі нумерованих, багато з яких мають об'єм одного або декількох абзаців.

2. Як один з основних результатів роботи виносяться положення про розвиток теорії семантичних відношень та створення методу виведення системи семантичних категорій із аксіоматичного набору. Проте у тексті дисертації відповідний матеріал представлено досить стисло, в тому числі в аспекті конкретизації подальшого практичного використання розробленого підходу.

3. Має місце повторення і переказування викладеного матеріалу. Зокрема у розділі 4, де розповідається про приклади реалізації онтологокерованих діалогових систем, надається повторний, можливо, надто детальний опис структури створених онтологій, методів аналізу тексту і автоматизованої побудови формальних запитів, раніше розкритих у розділі 2.

4. Написи на багатьох ілюстраціях, зокрема 2.3, 2.5, 2.6, 2.7, 2.8, 2.9, 4.1, 4.4, 4.5, виконані англійською мовою.

5. У розділі 4 «Практична реалізація аналізу і синтезу природномовних текстів при розробці діалогових і довідкових систем» безпосередньо опису конкретних прикладів програмної реалізації таких систем і їх аналізу присвячено лише підрозділ 4.4. Підрозділи 4.1, 4.2 та 4.3 відносяться скоріш

до методології та підходів до їх розробки. Можливо, краще було б розмежувати викладення цього матеріалу.

6. Додаток Д відрізняється від інших додатків великим об'ємом представленої текстової інформації детального опису прикладів. В принципі, цей матеріал можна було б і не виносити у додаток, а викласти у рамках основного тексту.

Зазначені недоліки не є принциповими і не впливають на загальну позитивну оцінку роботи.

8. Рекомендації щодо впровадження результатів дисертаційного дослідження в практику.

Розроблені алгоритми та програмні засоби можна використовувати для автоматизованого створення систематизованих баз знань на основі текстової інформації і відповідних довідкових і діалогових систем з природномовним інтерфейсом у багатьох галузях діяльності. Результати дисертації доцільно також впровадити у навчальний процес, де вони можуть бути покладені в основу створення інтерактивних посібників при підготовці спеціалістів відповідно профілю бази знань. Окрім того, самі розроблені моделі і алгоритми можуть скласти базис для спеціалізованих навчальних курсів чи модулів у рамках існуючих дисциплін з інженерії знань, які викладаються у закладах освіти країни.

9. Висновок про відповідність дисертації встановленим вимогам.

Дисертація Літвін А. А. "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 завершеною науковою роботою, в якій отримано нові науково обґрунтовані результати у сфері

автоматизованої обробки природномовних текстів і онтологічної інженерії. Робота повністю відповідає вимогам “Порядку присудження ступеня доктора філософії”, затвердженого постановою Кабінету Міністрів України №44 від 12 січня 2022 року, а її авторка заслуговує на присудження ступеня доктора філософії за спеціальністю 122 “Комп’ютерні науки”.

Офіційний опонент:

Професор кафедри комп’ютерної інженерії Київського національного університету імені Тараса Шевченка, доктор технічних наук, професор


Сергій ПОГОРІЛИЙ
(підпис)

Підпис Сергія ПОГОРІЛОГО засвідчую.

Декан факультету радіофізики, електроніки та комп’ютерних систем



Андрій НЕТРЕБА
(підпис)