

Голові разової спеціалізованої вченої ради
Національного технічного університету
«Дніпровська політехніка»
д.т.н., проф. Морозу Б.І.

РЕЦЕНЗІЯ

*завідувача кафедри системного аналізу та управління, доктора технічних наук, доцента **Желдака Тімура Анатолійовича** на дисертаційну роботу **М'якенького Арсенія Вячеславовича** «Метод ідентифікації перифраз на основі графових репрезентацій для моделювання когнітивного процесу розуміння», подану на здобуття наукового ступеня доктора філософії в галузі знань 12 «Інформаційні технології» за спеціальністю 122 «Комп'ютерні науки»*

Рецензію підготовлено на основі аналізу тексту дисертації, наукових публікацій здобувача, в яких висвітлено результати дисертаційного дослідження, а також документальних матеріалів, що засвідчують упровадження отриманих результатів у практику.

Актуальність роботи. У сучасних умовах зростання обсягів українськомовних текстових даних актуальним є розвиток методів обробки природної мови, здатних встановлювати семантичну еквівалентність висловлювань за істотної відмінності їхньої лексичної та синтаксичної форми. Автоматизоване розпізнавання перифраз є важливою прикладною задачею для інформаційного пошуку, інтелектуального аналізу текстів, машинного перекладу та інших напрямів штучного інтелекту.

Наявні підходи до ідентифікації перифраз, що переважно спираються на лексичні, статистичні або поверхневі синтаксичні ознаки, не завжди забезпечують достатню якість розпізнавання у випадках, коли текстові фрагменти мають різну мовну форму, але зберігають спільний зміст. Для української мови ця проблема

ускладнюється обмеженістю спеціалізованих корпусів перифраз і ресурсів для побудови формалізованих семантичних представлень.

У цьому контексті використання графових репрезентацій, нейронного машинного перекладу, графових нейронних мереж і методів машинного навчання є обґрунтованим напрямом удосконалення комп'ютерних технологій розпізнавання перифраз. Запропонований у дисертації підхід спрямований на формалізоване моделювання семантичної структури висловлювань і підвищення точності автоматизованої обробки українськомовних текстових даних. Таким чином, обрана М'якеньким А.В. тема дисертаційної роботи є актуальною, науково обґрунтованою та відповідає сучасним напрямам розвитку галузі комп'ютерних наук, зокрема штучного інтелекту, обробки природної мови, машинного навчання та інтелектуального аналізу текстових даних.

Зв'язок роботи з науковими програмами, планами, темами. Актуальність дисертаційного дослідження також визначається його зв'язком із науково-дослідними роботами кафедри програмного забезпечення комп'ютерних систем НТУ «Дніпровська політехніка»: «Методи, моделі та технології обробки даних в комп'ютерних системах загального та спеціального призначення» (державний реєстраційний номер 0124U004583) та «Високопродуктивні багатопроесорні системи: особливості конструювання. Дослідження оцінок ефективності застосування до розв'язування прикладних задач» (державний реєстраційний номер 0121U113718).

Наукова новизна одержаних результатів. На підставі критичного аналізу дисертації М'якенького А.В., а також наукових публікацій, у яких висвітлено основні результати досліджень, встановлено, що здобувачем досягнуто основну мету та розв'язано поставлені дослідні завдання. Вважаю, що до найбільш вагомих положень наукової новизни відносяться такі:

- запропоновано когнітивно орієнтовану методику до ідентифікації перифраз у текстах українською мовою, що ґрунтується на використанні семантичних графових репрезентацій та дозволяє здійснювати аналіз семантичної

еквівалентності висловлювань на концептуальному рівні в межах моделювання когнітивного процесу розуміння;

- вперше розроблено когнітивно обґрунтовану модель перифраза як концептуально-еквівалентного смислового варіанту в межах семантичного простору абстрактних репрезентацій змісту, що дозволяє уніфікувати підхід до ідентифікації перифраз незалежно від їхньої лексичної та синтаксичної форми;

- вперше розроблено інформаційну технологію формування AMR графів для українськомовних текстів на основі їхнього попереднього машинного перекладу. Така технологія гарантує збереження семантичної повноти текстів українською мовою при перекладі на рівні 94% за відсутності повноцінних українськомовних AMR корпусів;

- удосконалено метод репрезентації текстів природної мови за допомогою AMR шляхом використання графових нейронних мереж з механізмом уваги, які дозволяють формувати векторні представлення, що відображають як локальні, так і глобальні властивості смислової структури висловлювання, що забезпечує підвищення точності у задачі оцінювання семантичної подібності текстів на 4% порівняно з наявними AMR орієнтованими підходами;

- удосконалено метод ідентифікації перифраз шляхом застосування класифікації на основі векторних представлень AMR графів, що забезпечує прийняття рішень про семантичну еквівалентність висловлювань на рівні концептуально-смислових структур та дозволяє досягати показників точності та F1-міри на 11% та на 4,2% відповідно вищими за наявні підходи, що використовують AMR графи для ідентифікації перифраз.

Важливим є те, що обробка текстових даних у дисертації базується не лише на поверхневому зіставленні окремих слів чи статистичній близькості речень, а на аналізі структурованих семантичних представлень. Використання AMR-графів дозволяє розглядати висловлювання як складні інформаційні об'єкти, у яких суттєвими є не тільки окремі лексеми, але й зв'язки між концептами, предикатами та аргументами. Вважаю, що всі положення наукової новизни відповідають змісту

роботи та є повністю обґрунтованими і доведеними в результаті проведеного дисертаційного дослідження.

Практична цінність, використання та впровадження отриманих результатів. Здобувачем М'якеньким А.В. отримано низку практично орієнтованих результатів, які націлено на підвищення ефективності автоматизованої обробки українськомовних текстів, а саме: когнітивно орієнтовану методику ідентифікації перифраз на основі графових семантичних репрезентацій; інформаційну технологію формування AMR-графів для українськомовних висловлювань; програмні компоненти побудови векторних представлень AMR-графів із використанням графових нейронних мереж; класифікаційний модуль для визначення семантичної еквівалентності текстових пар.

Практичну значущість роботи підсилює те, що здобувач не обмежився теоретичним описом методу, а реалізував програмний комплекс, який забезпечує повний цикл обробки текстових даних: підготовку вхідних висловлювань, формування семантичних графових репрезентацій, побудову векторних ознак і класифікацію перифраз. Результати проведених обчислювальних експериментів свідчать про готовність запропонованої технології до подальшої інтеграції в реальні аналітичні платформи, орієнтовані на обробку українськомовного контенту.

Практична цінність результатів дисертаційної роботи підтверджується їх упровадженням у діяльність ТОВ «КМК «Арт Ерія», де запропоновані рішення використано під час обробки неструктурованої текстової інформації, аналізу масивів даних та автоматизації лінгвістичного аналізу. Окремі теоретичні й практичні здобутки дисертації також упроваджено в освітній процес Національного технічного університету «Дніпровська політехніка» під час підготовки здобувачів вищої освіти за спеціальністю 122 «Комп'ютерні науки».

Достовірність і обґрунтованість наукових положень, висновків і рекомендацій, які сформульовано та висвітлено в дисертації. Наукові положення, висновки та рекомендації, що сформульовані у дисертаційній роботі, є достатньо обґрунтованими, логічно послідовними та підтвердженими результатами теоретичних і експериментальних досліджень. Достовірність отриманих

результатів забезпечується використанням сучасного математичного апарату, методів когнітивного моделювання, алгоритмів інтелектуальної обробки природномовних даних, графових репрезентацій AMR, трансформерних моделей, графових нейронних мереж та методів машинного навчання.

Важливим результатом роботи є інтеграція методів машинного навчання у процес порівняння складних інформаційних об'єктів, представлених у вигляді AMR-графів. Це дозволило перейти від формального порівняння текстових послідовностей до машинного аналізу структурованих семантичних репрезентацій, що є методично виправданим для задачі ідентифікації перифраз.

Експериментальна перевірка розроблених рішень проведена поетапно: досліджено ефективність нейронного машинного перекладу для збереження семантичної структури українськомовних текстів, якість побудови AMR-графів за допомогою AMRBART, вплив механізму уваги у графових нейронних мережах на формування векторних репрезентацій, а також результативність класифікації перифраз. Отримані автором показники точності оцінювання семантичної подібності підтверджують достовірність і практичну цінність запропонованого підходу.

Оцінка змісту роботи та повноти викладу результатів дисертації в опублікованих працях. Дисертаційна робота має логічно вибудовану структуру та охоплює повний цикл наукового дослідження: від аналізу сучасного стану проблеми моделювання когнітивного процесу розуміння до розробки, тестування й програмної реалізації інформаційної технології ідентифікації перифраз на основі графових репрезентацій.

У вступі чітко сформульовано мету, завдання, об'єкт і предмет дослідження, визначено наукову новизну та практичну значущість результатів, наведено відомості про апробацію та впровадження. Вступна частина засвідчує, що здобувач розглядає ідентифікацію перифраз не як ізольовану задачу зіставлення текстів, а як складову ширшої проблеми моделювання когнітивного процесу розуміння природної мови.

Перший розділ присвячено аналізу сучасного стану проблеми моделювання когнітивних процесів та автоматизованої обробки природної мови. Автор послідовно розглядає теоретичні засади когнітивного пізнання, підходи до комп'ютерного подання змісту висловлювань, а також наявні методи виявлення семантичної еквівалентності. Особливо важливим є те, що в розділі показано обмеженість підходів, зорієнтованих переважно на поверхнєве зіставлення слів, і обґрунтовано потребу переходу до структурного аналізу смислу.

У другому розділі сформовано методичну основу дисертаційної роботи. Здобувач обґрунтовує використання абстрактних семантичних репрезентацій AMR як інструменту подання концептуальної структури речення, формалізує задачу ідентифікації перифраз та визначає логіку переходу від природномовного тексту до графового представлення. Для рецензованої роботи це є принциповим моментом, оскільки саме графова форма дозволяє порівнювати складні інформаційні об'єкти за їхнім змістом, а не лише за зовнішньою синтаксичною подібністю.

Третій розділ має прикладний та системотехнічний характер. У ньому розроблено архітектуру інтелектуальної системи ідентифікації перифраз, описано функціональні модулі, алгоритм роботи та програмну реалізацію технології. До позитивних аспектів слід віднести те, що автор подає систему як послідовний технологічний ланцюг: від підготовки текстових даних і формування AMR-графів до побудови графових векторних репрезентацій та прийняття класифікаційного рішення щодо семантичної еквівалентності.

У четвертому розділі наведено результати експериментальної перевірки запропонованої інформаційної технології. Автор аналізує якість підготовки українськомовних текстів до AMR-парсингу, ефективність побудови графових репрезентацій, вплив механізму уваги у графових нейронних мережах та результативність класифікації перифраз. Така послідовність експериментів дає змогу оцінити не лише кінцеву точність системи, а й роботу її окремих компонентів, що є важливим для подальшої інтеграції технології в реальні аналітичні платформи.

Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та додатків. Загальний обсяг дисертації становить 196

сторінок, основна частина - 142 сторінки, робота містить 28 рисунків, 21 таблицю, 107 літературних джерел і 3 додатки. Структура та обсяг роботи відповідають чинним вимогам до дисертацій на здобуття ступеня доктора філософії.

Основні результати дисертації достатньо повно відображені у 10 наукових публікаціях, з яких 5 статей опубліковано у фахових періодичних виданнях України з технічних наук, у тому числі 1 стаття категорії А, що індексується в Scopus, 4 статті категорії Б, а також 5 праць, що засвідчують апробацію матеріалів дисертаційного дослідження. Тематика публікацій відповідає змісту дисертації та відображає основні положення, що виносяться на захист.

Оформлення дисертації та дотримання вимог академічної доброчесності. Дисертація М'якенького А.В. написана українською мовою. Стиль викладення матеріалів досліджень є логічним, лаконічним і структурованим. За результатами аналізу тексту дисертації, наукових публікацій здобувача та матеріалів, що характеризують виконане дослідження, не встановлено порушень принципів академічної доброчесності. Усі результати, подані на захист, є оригінальними та отримані здобувачем особисто з дотриманням вимог чинного законодавства України та нормативних документів закладу вищої освіти. Використані положення інших авторів супроводжуються посиланнями на відповідні джерела, а особистий внесок здобувача у публікаціях, виконаних у співавторстві, визначено коректно.

Дискусійні положення та зауваження щодо змісту та оформлення дисертації. На підставі детального вивчення дисертаційної роботи були сформульовані такі зауваження та дискусійні положення:

1. У роботі обґрунтовано використання нейронного машинного перекладу як проміжного етапу перед побудовою AMR-графів для українськомовних текстів. Водночас ширше окреслення перспектив прямого генерування AMR-графів з української мови без проміжного перекладу дозволило б повніше показати можливі напрями підвищення швидкодії та загальної ефективності системи.

2. У другому розділі обґрунтовано використання AMR-графів як засобу подання концептуальної структури висловлювань. Разом із тим корисним було б

ширше прокоментувати межі застосування такого підходу для українськомовних текстів у випадках семантичної варіативності вузлів, контекстної залежності ролей та структурної неоднорідності AMR-графів.

3. У четвертому розділі наведено результати експериментальної перевірки запропонованої технології на українськомовному наборі даних, сформованому на основі матеріалів СУМ-20. Додатковий аналіз її роботи на вузькоспеціалізованих українськомовних текстах, зокрема інженерної чи ІТ-сфери, дав би змогу точніше оцінити універсальність запропонованого підходу з урахуванням термінологічної й семантичної специфіки таких текстів.

4. У роботі для оцінювання якості побудови AMR-графів використано метрику Smatch. Водночас вважаю, що недостатньо уваги приділено її обмеженням щодо різних типів помилок у графах, зокрема втрати або некоректного визначення типів вузлів чи ребер. Таке уточнення дозволило б повніше оцінити вплив якості AMR-графів на подальшу ідентифікацію перифраз.

Зазначені зауваження не мають принципового характеру та не знижують загальної позитивної оцінки дисертаційної роботи.

Висновок. Дисертаційна робота М'якенького Арсенія Вячеславовича є завершеною кваліфікаційною науковою працею, виконаною на належному науковому та практичному рівнях. У роботі розв'язано актуальну науково-прикладну задачу підвищення точності ідентифікації перифраз у текстах українською мовою в контексті моделювання когнітивного процесу розуміння шляхом розробки інтелектуальної системи, яка використовує семантичні графові репрезентації для виявлення семантичної еквівалентності між висловлюваннями на концептуальному рівні.

За актуальністю, науковою новизною, теоретичною та практичною значущістю отриманих результатів дисертація відповідає вимогам чинного законодавства України щодо дисертацій на здобуття ступеня доктора філософії, зокрема вимогам наказу МОН України № 40 від 12.01.2017 р. «Про затвердження вимог до оформлення дисертації» (зі змінами) та «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 від 12.01.2022 р. № 44.

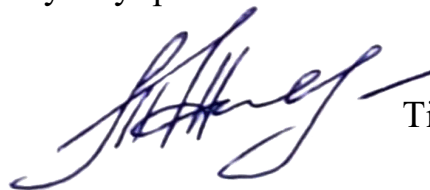
Здобувач М'якенький Арсеній Вячеславович заслуговує на присудження наукового ступеня доктора філософії в галузі знань 12 «Інформаційні технології» за спеціальністю 122 «Комп'ютерні науки».

Рецензент:

завідувач кафедри системного аналізу та управління

НТУ «Дніпровська політехніка»,

доктор технічних наук, доцент



Тімур ЖЕЛДАК