

Голові разової СБР із захисту
дисертації М'якенького А.В.
проф. Морозу Б.І.
опонента,
д.т.н., проф. Гончаренко Т. А.

ВІДГУК ОФІЦІЙНОГО ОПОНЕНТА

доктора технічних наук, професорки, завідувачки кафедри інформаційних технологій Київського національного університету будівництва і архітектури

ГОНЧАРЕНКО Тетяни Андріївни на дисертацію

М'ЯКЕНЬКОГО Арсенія Вячеславовича «Метод ідентифікації перифраз на основі графових репрезентацій для моделювання когнітивного процесу розуміння», подану до захисту на здобуття ступеня доктора філософії за спеціальністю 122 «Комп'ютерні науки»

Актуальність теми дисертаційного дослідження.

Сучасний етап розвитку інформаційних технологій характеризується інтенсивним використанням методів штучного інтелекту, машинного навчання, нейромережевого моделювання та обробки природної мови для аналізу великих масивів текстових даних. У цьому контексті особливої ваги набувають методи, здатні встановлювати не лише формальну подібність текстових фрагментів, а й їхню семантичну еквівалентність. Саме такою є задача ідентифікації перифраз, у межах якої різні за лексичною та синтаксичною формою висловлювання можуть передавати однаковий зміст.

Дисертаційна робота М'якенького А.В. є актуальною, оскільки спрямована на розроблення методу ідентифікації перифраз у текстах українською мовою на основі графових семантичних репрезентацій. Для українськомовного сегмента ця задача має особливе значення через обмеженість спеціалізованих ресурсів перифраз та AMR-розмітки, а також через потребу у створенні інструментів, здатних виконувати глибокий семантичний аналіз природномовної інформації.

Вибір автором AMR-графів, графових нейронних мереж і класифікаційних моделей є методично виправданим, оскільки графова форма подання дозволяє аналізувати текст як структурований інформаційний об'єкт. Такий підхід є близьким до сучасних тенденцій розвитку інформаційних технологій, у яких дедалі більшого значення набувають графові моделі, нейромережеві методи прогнозування та інтелектуальні системи підтримки прийняття рішень на основі структурованих даних.

З позиції розвитку інформаційних технологій важливо, що дисертація не обмежується використанням готових мовних моделей як «чорної

скриньки», а пропонує перехід до явного представлення смислової структури речення. Це підвищує інтерпретованість результатів, створює підстави для аналізу помилок і робить запропонований підхід перспективним для задач, де важливо не лише отримати класифікаційне рішення, а й розуміти, на яких семантичних ознаках воно ґрунтується.

Отже, тема дисертації відповідає актуальним напрямам розвитку комп'ютерних наук, зокрема обробки природної мови, семантичного аналізу, машинного навчання та побудови інтелектуальних інформаційних систем.

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

Дисертаційна робота виконана в Національному технічному університеті «Дніпровська політехніка» та пов'язана з такими науково-дослідними роботами: «Методи, моделі та технології обробки даних в комп'ютерних системах загального та спеціального призначення» (номер державної реєстрації 0124U004583) та «Високопродуктивні багатопроцесорні системи: особливості конструювання. Дослідження оцінок ефективності застосування до розв'язування прикладних задач» (номер державної реєстрації 0121U113718).

Отримані результати безпосередньо відповідають зазначеним науковим напрямам, оскільки стосуються розроблення методів, моделей, алгоритмів і програмних засобів для інтелектуальної обробки даних у комп'ютерних системах.

Ступінь обґрунтованості та достовірності наукових положень, висновків і рекомендацій, сформульованих у дисертації.

Сформульовані в дисертаційній роботі наукові положення, рекомендації та висновки повною мірою відповідають установленим вимогам до досліджень відповідного рівня. Високий ступінь обґрунтованості, достовірності та об'єктивності наукових результатів і висновків забезпечується здійсненням критичного аналізу сучасних джерел у межах досліджуваної предметної області, чітким визначенням мети, об'єкта, предмета і завдань дослідження, а також послідовним використанням сучасних методів комп'ютерних наук.

Обґрунтованість дисертації також підтверджується тим, що кожне з поставлених завдань має відображення у відповідних теоретичних, алгоритмічних або експериментальних результатах. Достовірність результатів підтверджується експериментальною перевіркою окремих етапів запропонованої інформаційної технології. У роботі досліджено збереження семантичної структури українськомовних текстів під час нейронного

машинного перекладу, якість побудови AMR-графів, ефективність графових нейронних мереж з механізмом уваги та результативність класифікаційного модуля. Така поетапна валідація підсилює обґрунтованість кінцевих висновків.

Отримані результати є логічно пов'язаними між собою та безпосередньо впливають із поставленої мети і завдань дослідження. Окремі положення та результати дисертації пройшли апробацію, відображені у наукових публікаціях здобувача, що додатково підтверджує їхню обґрунтованість і відповідність тематиці проведеного дослідження.

Таким чином, наведені в дисертаційній роботі наукові положення, рекомендації та висновки можна вважати достатньо обґрунтованими й достовірними, а отримані результати такими, що відповідають меті дослідження та мають належне теоретичне і практичне значення.

Наукова новизна одержаних результатів полягає у теоретичному обґрунтуванні та розробленні методу ідентифікації перифраз для українськомовних текстів, що поєднує когнітивно орієнтований підхід, графові репрезентації AMR, нейронний машинний переклад, графові нейронні мережі та класифікаційні моделі. Основними науковими результатами дисертаційної роботи, що визначають її новизну, є такі:

- **вперше** розроблено когнітивно обґрунтовану модель перифраза як концептуально-еквівалентного смислового варіанта в межах семантичного простору абстрактних репрезентацій змісту (AMR), що дозволяє уніфікувати підхід до ідентифікації перифраз незалежно від їхньої лексичної та синтаксичної структури.

- **вперше** запропоновано метод формування AMR-графів для українськомовних текстів на основі нейронного машинного перекладу, який забезпечує збереження семантичної повноти текстів на рівні 94% за умов відсутності повноцінних нативних українськомовних AMR-корпусів.

- **удосконалено** метод репрезентації текстів природної мови за допомогою AMR-графів шляхом застосування графових нейронних мереж із механізмом уваги (GAT), що дає змогу формувати векторні представлення, які одночасно відображають локальні та глобальні властивості смислової структури висловлювання; це забезпечило підвищення точності оцінювання семантичної подібності текстів на 4% порівняно з існуючими AMR-орієнтованими підходами.

- **удосконалено** метод ідентифікації перифраз шляхом застосування класифікації на основі векторних представлень AMR-графів, що забезпечує прийняття рішень про семантичну еквівалентність на рівні концептуально-

сміслових структур та дозволяє досягати показників точності та F1-міри вищих на 11% та 4,2% відповідно порівняно з існуючими аналогами.

– **набув подальшого розвитку** когнітивно орієнтований підхід до ідентифікації перифраз у текстах українською мовою, який, на відміну від існуючих, використовує графові семантичні репрезентації для аналізу семантичної еквівалентності висловлювань на концептуальному рівні в межах моделювання когнітивного процесу розуміння.

Вважаю, що наведені положення наукової новизни повністю відповідають змісту роботи, є належно аргументованими та підтверджені результатами експериментальних досліджень. Вагомим здобутком автора є не лише розробка окремих алгоритмів класифікації, а й створення цілісного методичного та інформаційного забезпечення, адаптованого для автоматизованого опрацювання українськомовного контенту.

Практичне значення одержаних результатів полягає у розробці та програмній реалізації комплексної інформаційної технології автоматизованої ідентифікації перифраз та аналізу семантичної еквівалентності українськомовних текстів. Основні практичні здобутки роботи:

1) Розроблено програмний прототип у вигляді клієнт-серверного вебзастосунку з графічним інтерфейсом користувача, який реалізує повний конвеєр (pipeline) обробки даних: від попередньої підготовки текстів і побудови AMR-графів до формування графових вкладень та класифікації семантичної еквівалентності.

2) Створено модульну архітектуру, окремі компоненти якої (модулі препроцесингу українськомовних текстів, генерації AMR-структур, графового кодування та нейромережевої класифікації) можуть автономно інтегруватися у суміжні системи обробки природної мови (NLP), пошукові системи, системи виявлення плагіату та інтелектуального аналізу текстових масивів.

3) Забезпечено наочність та інтерпретованість прийняття рішень завдяки можливості візуалізації проміжних результатів графових репрезентацій та оцінок семантичної подібності на концептуальному рівні.

Впровадження результатів дослідження. Практичну цінність результатів дисертаційної роботи підтверджено відповідними актами та довідками про впровадження:

1) У промисловий сектор: у діяльність ТОВ «КМК «Арт Ерія» — для автоматизації процесів семантичного аналізу та обробки великих масивів неструктурованої текстової інформації.

2) В освітній процес: у Національному технічному університеті «Дніпровська політехніка» на кафедрі комп'ютерних наук при викладанні навчальних дисциплін з обробки природної мови та інтелектуального аналізу даних для здобувачів вищої освіти за спеціальністю 122 «Комп'ютерні науки».

Оцінка змісту дисертації, її завершеності та відповідності встановленим вимогам.

Дисертаційна робота має логічну та послідовну структуру, відповідає поставленій меті та завданням дослідження. Робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел і додатків. Загальний обсяг дисертації становить 196 сторінок, основна частина - 142 сторінки. Робота містить 28 рисунків, 21 таблицю, 107 літературних джерел і 3 додатки.

У вступі обґрунтовано актуальність теми, сформульовано мету, основні завдання, об'єкт, предмет і методи дослідження, наведено положення наукової новизни та практичного значення. Вступна частина демонструє чітке розуміння автором місця дослідження в сучасній проблематиці обробки природної мови та інтелектуального аналізу текстових даних.

Перший розділ присвячено аналізу сучасного стану розробок і досліджень у галузі моделювання когнітивних процесів, семантичного аналізу текстів та ідентифікації перифраз. У ньому автор узагальнює теоретичні засади когнітивного пізнання, розглядає наявні підходи до встановлення семантичної еквівалентності та обґрунтовує доцільність переходу від поверхневого зіставлення текстів до графового подання змісту. Слід відзначити, що у першому розділі достатньо чітко показано зв'язок між лінгвістичною природою перифраза та комп'ютерною задачею його автоматизованого розпізнавання. Це важливо, оскільки для задач такого типу некоректна постановка на рівні предметної області часто призводить до формального порівняння текстів без належного урахування змістової еквівалентності.

Другий розділ містить методичну основу дослідження. У ньому сформовано підхід до ідентифікації перифраз на основі графових семантичних репрезентацій, описано логіку формування AMR-графів та визначено принципи порівняння смислових структур. З погляду інформаційних технологій важливо, що автор переводить природномовне висловлювання у формалізовану структуру, придатну для подальшого машинного аналізу. Для оцінки цієї роботи особливо показовим є використання графових структур як проміжної форми між природномовним

текстом і нейромережевою моделлю. Такий підхід узгоджується із сучасними тенденціями застосування графових представлень у задачах прогнозування, моніторингу та аналізу складних інформаційних об'єктів.

Третій розділ присвячено розробленню інформаційної технології ідентифікації перифраз, її архітектури, функціональних модулів та програмної реалізації. У цьому розділі найбільш чітко проявляється прикладна спрямованість роботи: здобувач описує технологічний конвеєр обробки тексту, формування AMR-репрезентацій, графового кодування та класифікації.

Четвертий розділ містить експериментальне дослідження розробленої технології. Автор послідовно оцінює якість окремих компонентів і кінцевого класифікаційного рішення. Такий підхід є методично коректним, оскільки дозволяє оцінити не лише загальну ефективність системи, а й внесок окремих модулів у кінцевий результат. Позитивним є те, що експериментальна частина побудована не як одноразове порівняння кінцевої точності, а як перевірка послідовного технологічного ланцюга. Це дозволяє оцінити вразливі місця системи, простежити вплив проміжних репрезентацій на класифікаційне рішення та обґрунтувати доцільність використання саме графового подання семантики.

За своїм змістом дисертація є завершеним науковим дослідженням, у якому теоретичні положення, методичні рішення, програмна реалізація та експериментальна перевірка узгоджені між собою. Мова і стиль викладення відповідають вимогам до наукових праць такого рівня.

Повнота викладу основних положень дисертації в опублікованих наукових працях. Основні результати дисертаційної роботи достатньо повно відображені у 10 наукових публікаціях здобувача, 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 дисертації. Тематика опублікованих праць відповідає змісту дисертаційної роботи та відображає основні положення, що виносяться на захист.

Дотримання принципів академічної доброчесності. За результатами ознайомлення з дисертаційною роботою можна зробити висновок, що дисертація М'якенького А.В. є самостійним та оригінальним дослідженням. Використані в роботі ідеї, матеріали та результати інших авторів супроводжуються належними посиланнями. Ознак фабрикації, фальсифікації або некоректних текстових запозичень не виявлено.

Зауваження, коментарі та дискусійні положення щодо змісту дисертації. Під час детального ознайомлення з результатами досліджень, викладеними в дисертаційній роботі М'якенького А.В., було виявлено такі зауваження, коментарі та дискусійні положення:

1. **Стосовно формалізації математичної моделі.** У другому розділі дисертації смислове ядро висловлення розглянуто як сукупність концептів і відношень між ними за формулою (2.1), а AMR-граф використано як форму подання концептуальної структури речення за формулою (2.11). Для більшої наочності та цілісності викладу бажано було б чіткіше та детальніше пояснити математичну відповідність (відображення) елементів смислового ядра висловлення вузлам, ребрам та міткам AMR-графа.

2. **Стосовно конкретизації семантичних критеріїв.** При формалізації задачі ідентифікації перифраз смислове ядро висловлення подано як сукупність концептів і семантичних відношень між ними. Водночас у межах формальної постановки задачі варто було б конкретизувати, які саме елементи цього ядра (наприклад, кореневі вузли, підграфи суб'єктно-об'єктних зв'язків чи конкретні семантичні ролі) є визначальними для встановлення семантичної еквівалентності двох висловлень.

3. **Стосовно оцінки внеску етапів навчання (Ablation Study).** У роботі запропоновано двоетапну схему навчання графового кодувальника, що поєднує попереднє навчання на реконструкції структурних зв'язків AMR-графа та подальше тонке налаштування (fine-tuning) на задачі семантичної текстової подібності. Для повнішої експериментальної характеристики методу доречно було б навести результати дослідження усунення (ablation study), щоб показати, наскільки змінюється точність підсумкової класифікації за відсутності кожного з цих етапів окремо.

4. **Стосовно часової складності та продуктивності.** У третьому розділі описано програмну реалізацію інтелектуальної системи ідентифікації перифраз, яка передбачає послідовне виконання кількох етапів конвеєра: нейронного перекладу, побудови AMR-графів, формування графових представлень (embeddings) та класифікації. Бажано було б навести детальнішу часову оцінку (профіль виконання) кожного з цих етапів, що дозволило б визначити найбільш ресурсомісткі компоненти реалізованої технології та оцінити її придатність для роботи у режимі реального часу.

5. **Стосовно термінологічної уніфікації.** У тексті дисертації місцями паралельно використовуються українськомовні та англomовні терміни, зокрема назви функцій втрат (loss functions), способів агрегування

(pooling) та компонентів нейромережевої архітектури. Для досягнення строгої академічної уніфікації було б доцільним при першому згадуванні подавати англomовний термін у дужках, а надалі послідовно дотримуватися обраного варіанта українomовної адаптації.

Вважаю, що вищезазначені зауваження, коментарі та дискусійні положення жодним чином не применшують високої наукової та практичної цінності дисертаційної роботи М'якенького А.В., а лише висвітлюють напрями для подальшого розвитку дослідження та слугуватимуть предметом конструктивної наукової дискусії під час публічного захисту.

Загальний висновок щодо дисертаційної роботи. З урахуванням актуальності обраного напрямку, наукової новизни, практичної цінності, обсягу публікаційного доробку, а також високого ступеня достовірності та обґрунтованості одержаних результатів і висновків, вважаю, що дисертаційна робота М'якенького Арсенія В'ячеславовича на тему: «Метод ідентифікації перифраз на основі графових репрезентацій для моделювання когнітивного процесу розуміння» є завершеним, самостійним та кваліфікованим науковим дослідженням, у якому розв'язано актуальне науково-прикладне завдання у галузі комп'ютерних наук.

Дисертація *повністю відповідає* «Вимогам до оформлення дисертації», затвердженим наказом Міністерства освіти і науки України від 12.01.2017 р. № 40, та «Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», затвердженому постановою Кабінету Міністрів України від 12.01.2022 р. № 44 (зі змінами).

Її автор, М'якенький Арсеній В'ячеславович, *заслуговує* на присудження ступеня доктора філософії з галузі знань 12 «Інформаційні технології» за спеціальністю 122 «Комп'ютерні науки».

Офіційний опонент

доктор технічних наук, професорка,
завідувачка кафедри інформаційних технологій
Київського національного університету
будівництва і архітектури

Тетяна ГОНЧАРЕНКО