

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

## ВІДГУК

офіційного опонента на кваліфікаційну наукову працю (дисертацію)  
М'якенького Арсенія Вячеславовича на тему «Метод ідентифікації перифраз на  
основі графових репрезентацій для моделювання когнітивного процесу  
розуміння», подану на здобуття наукового ступеня доктора філософії за  
спеціальністю 122 «Комп'ютерні науки»

### *1. Актуальність дисертації.*

Дисертаційна робота М'якенького А.В. присвячена актуальній науково-прикладній задачі комп'ютерних наук - розробленню методу ідентифікації перифраз на основі графових репрезентацій для моделювання когнітивного процесу розуміння природної мови. Актуальність теми зумовлена тим, що сучасні інтелектуальні інформаційні системи потребують механізмів не лише лексичного або статистичного зіставлення текстів, а й встановлення їхньої семантичної еквівалентності на рівні змісту, наміру та концептуальної структури висловлювання.

Особливого значення дослідження набуває в українськомовному контексті. На відміну від англійської мови, для якої сформовано значні корпуси, моделі та ресурси семантичного аналізу, українська мова все ще характеризується дефіцитом спеціалізованих ресурсів саме для задач ідентифікації перифраз та побудови AMR-репрезентацій. Тому запропонована автором технологія, яка поєднує підготовку українськомовного тексту, формування абстрактних семантичних графів, графове кодування та класифікацію семантичної еквівалентності, є своєчасною і відповідає сучасним потребам розвитку українськомовних NLP-систем.

Важливо, що здобувач розглядає перифразування не як поверхневу мовну трансформацію, а як збереження смислового інваріанта при зміні мовної форми. Саме такий підхід дозволяє пов'язати прикладну задачу автоматизованого виявлення перифраз із ширшою проблематикою моделювання когнітивного процесу розуміння. У цьому аспекті робота має не лише прикладне, а й теоретичне значення для розвитку методів подання знань, семантичного аналізу текстів та побудови інтелектуальних систем, здатних працювати з природномовною інформацією на змістовому рівні.

Тематика дисертації узгоджується із сучасними напрямками розвитку машинного навчання, графових нейронних мереж, трансформерних архітектур,

інформаційного пошуку, автоматичного реферування, виявлення дублювання змісту та підтримки прийняття рішень у системах, що опрацьовують великі обсяги текстових даних. З огляду на це обрана тема є актуальною та науково обґрунтованою.

## ***2. Зв'язок дисертаційної роботи з науковими програмами, планами і темами.***

Дисертаційну роботу виконано на кафедрі програмного забезпечення комп'ютерних систем Національного технічного університету «Дніпровська політехніка». Зміст дослідження відповідає науковій проблематиці кафедри та пов'язаний із розвитком методів, моделей і технологій обробки даних у комп'ютерних системах загального та спеціального призначення.

Зокрема, робота пов'язана з науково-дослідними темами «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 відповідають зазначеним науковим напрямом, оскільки стосуються алгоритмічного, програмного та експериментального забезпечення інтелектуальної обробки природномовних даних.

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

Сформульовані в дисертації наукові положення, висновки та рекомендації є достатньо обґрунтованими, логічно послідовними і такими, що відповідають поставленій меті та визначеним завданням. Автор коректно вибудував дослідницьку логіку: від аналізу когнітивного процесу розуміння та природи перифраза до формалізації семантичної еквівалентності, побудови AMR-графів, формування графових векторних репрезентацій і прийняття класифікаційного рішення.

Достовірність результатів забезпечується використанням сучасного апарату обробки природної мови, методів машинного навчання, нейронного машинного перекладу, трансформерних моделей, графових нейронних мереж з механізмом уваги та експериментального оцінювання якості інтелектуальних систем. Позитивним є те, що здобувач не обмежився демонстрацією кінцевого результату класифікації, а послідовно перевіряв окремі компоненти технологічного конвеєра: збереження семантичної структури під час перекладу, якість генерації AMR-графів, вплив графового кодування та результативність класифікації перифраз.

Одержані кількісні результати експериментальних досліджень підтверджують працездатність і результативність запропонованого підходу.

Наведені висновки узгоджуються з експериментальною частиною роботи та не суперечать поставленим завданням.

#### ***4. Наукова новизна та практичне значення одержаних результатів.***

Наукова новизна дисертаційної роботи М'якенького А.В. полягає у теоретичному обґрунтуванні та розробленні методу ідентифікації перифраз у текстах українською мовою на основі графових семантичних репрезентацій, що дозволяє здійснювати порівняння висловлювань на рівні концептуально-семантичних структур. Одержані результати мають цілісний характер, оскільки поєднують когнітивне трактування перифраза, технологію формування AMR-графів, графове кодування текстових структур і класифікаційне прийняття рішення щодо семантичної еквівалентності.

До основних положень наукової новизни слід віднести такі результати:

1. 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;

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

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

4. 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 семантичної структури висловлювання, що забезпечує підвищення точності у задачі оцінювання семантичної подібності текстів на 4% порівняно з наявними AMR орієнтованими підходами;

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

Зазначені положення наукової новизни є змістовно пов'язаними між собою та утворюють послідовну модельну й програмну основу для семантичного аналізу українськомовних текстів. На відміну від підходів, що переважно спираються на лексичну або статистичну подібність, у дисертації запропоновано орієнтацію на структуру змісту, що є принциповим для задачі ідентифікації перифраз.

Практичне значення одержаних результатів визначається можливістю їх використання для побудови програмних засобів автоматизованого аналізу українськомовних текстів, зокрема для встановлення семантичної близькості та виявлення перифраз між висловлюваннями. У межах дисертаційної роботи запропоновано інформаційну технологію, яка поєднує попередню обробку текстових даних, формування AMR-представлень, графове кодування та подальше прийняття рішення щодо змістової еквівалентності.

Реалізоване прикладне програмне забезпечення забезпечує можливість практичної перевірки запропонованих рішень і може застосовуватися як окремий дослідницько-прикладний інструмент або як складова більших інформаційних систем, орієнтованих на роботу з природномовним контентом. Практична спрямованість роботи посилюється тим, що запропоновані рішення адаптовано до особливостей українськомовних текстів, для яких наявність спеціалізованих корпусів перифраз і ресурсів семантичного графового подання є обмеженою. Це підвищує значущість отриманих результатів для подальшого розвитку вітчизняних систем обробки природної мови.

Прикладна цінність результатів підтверджується їх використанням у діяльності ТОВ «КМК «Арт Ерія» для обробки неструктурованої текстової інформації та автоматизації лінгвістичного аналізу, а також упровадженням в освітній процес Національного технічного університету «Дніпровська політехніка» під час підготовки здобувачів вищої освіти за спеціальністю 122 «Комп'ютерні науки».

#### ***5. Оцінка змісту дисертації, стилю викладення та повноти викладення положень, висновків та рекомендацій в опублікованих наукових працях.***

Дисертаційна робота М'якенького А.В. має логічну, завершену та послідовну структуру. Вона складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та додатків. Загальний обсяг дисертації становить 196 сторінок, з яких 142 сторінки становить основний текст. Робота містить 28 рисунків, 21 таблицю, список використаних джерел із 107 найменувань і 3 додатки. Такий обсяг і структура відповідають характеру поставленої наукової задачі та дають змогу достатньо повно висвітлити як теоретичну, так і прикладну складові дослідження.

У вступі обґрунтовано актуальність обраної теми, сформульовано мету, об'єкт, предмет і задачі дослідження, визначено методи, наукову новизну, практичне

значення та особистий внесок здобувача. Вступна частина загалом демонструє, що автор усвідомлює подвійну природу досліджуваної задачі: з одного боку, це прикладна задача автоматизованої класифікації текстових пар, а з іншого - задача моделювання процесу розуміння через зіставлення смислових структур.

Перший розділ присвячено інформаційно-аналітичному огляду стану проблеми моделювання когнітивних процесів та ідентифікації перифраз. Автор розглядає теоретико-методологічні засади когнітивного пізнання, процес розуміння природномовних висловлювань, поняття семантичної еквівалентності та сучасні підходи до автоматизованого виявлення перифраз. Важливо, що огляд не зводиться до простого переліку методів, а підводить до висновку про недостатність суто лексичних, статистичних або поверхневих embedding-підходів у випадках, коли необхідно встановити тотожність змісту при різній мовній формі.

У цьому розділі обґрунтовано доцільність звернення до AMR як до засобу подання абстрактної семантичної структури речення. Такий перехід є методично виправданим, оскільки саме графова репрезентація дозволяє моделювати відношення між поняттями, ролями та подіями у висловлюванні. Для дисертації це має принципове значення: автор показує, що перифрази доцільно порівнювати не лише як послідовності токенів, а як структуровані інформаційні об'єкти.

У другому розділі сформовано теоретико-методичну основу дослідження. Здобувач розглядає процес розуміння як зіставлення смислових структур, формалізує задачу ідентифікації перифраз та описує метод формування абстрактних семантичних репрезентацій. Зміст цього розділу є важливим, оскільки саме тут здійснено перехід від загальної когнітивної постановки задачі до конкретної обчислювальної процедури, яка може бути реалізована в програмній системі.

Окремо слід відзначити, що автор достатньо послідовно обґрунтовує використання попереднього нейронного машинного перекладу для побудови AMR-графів українськомовних текстів. За умов відсутності повноцінних українськомовних AMR-корпусів таке рішення є прагматичним і науково вмотивованим. Водночас здобувач не залишає цей етап як технічне припущення, а перевіряє збереження семантичної структури, що підсилює обґрунтованість усієї технології.

Третій розділ містить основне інженерно-алгоритмічне ядро дисертації. У ньому розроблено архітектуру інтелектуальної системи ідентифікації перифраз, визначено призначення її функціональних модулів, описано підготовку текстових даних, генерацію AMR-графів, формування графових вкладень та класифікацію семантичної еквівалентності. З позиції офіційного опонента позитивно оцінюю те, що автор доводить запропоновані методи до рівня програмної реалізації, а не обмежується концептуальним описом моделі.

Практична реалізація у вигляді інтерактивного клієнт-серверного вебзастосунку підтверджує прикладну завершеність дослідження. Архітектура системи демонструє послідовність технологічного конвеєра: користувацький запит, попередня обробка текстів, переклад, AMR-парсинг, побудова графових репрезентацій, векторизація та класифікація. Така структура є зрозумілою, відтвореною і придатною для подальшої модифікації або інтеграції в інші інформаційні системи.

Четвертий розділ присвячено експериментальному дослідженню запропонованої технології. Автор послідовно перевіряє якість підготовки українськомовних текстів до AMR-парсингу, оцінює побудову графових репрезентацій, досліджує вплив механізму уваги у графових нейронних мережах та аналізує результативність класифікації перифраз. Така логіка експериментальної частини є методично коректною, оскільки дозволяє оцінити не лише кінцевий результат, але й внесок окремих компонентів технології.

Загалом зміст дисертації свідчить про належну теоретичну підготовку здобувача, володіння сучасними методами комп'ютерних наук та здатність виконувати повний цикл наукового дослідження: від постановки проблеми й аналізу предметної області до розроблення методу, програмної реалізації, експериментальної перевірки та формулювання узагальнених висновків. Стиль викладення є науковим, термінологічно насиченим, у цілому послідовним і зрозумілим для фахівців у галузі інформаційних технологій.

Основні положення дисертаційної роботи достатньо повно висвітлені у 10 наукових публікаціях здобувача. Серед них 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 апробаційного характеру. Тематика публікацій відповідає змісту дисертації та відображає її ключові наукові результати.

#### ***6. Відсутність/наявність порушення академічної доброчесності.***

Аналіз дисертаційної роботи М'якенького А.В. дає підстави зробити висновок про дотримання здобувачем принципів академічної доброчесності. Дисертація є самостійним науковим дослідженням, у якому використані ідеї, результати та матеріали інших авторів супроводжуються належними посиланнями на першоджерела. Ознак фабрикації, фальсифікації результатів або некоректного використання текстових запозичень не встановлено.

#### ***7. Зауваження та дискусійні положення.***

Дисертаційна робота М'якенького А.В. характеризується належним науковим рівнем, завершеністю та практичною спрямованістю. Водночас, як офіційний опонент, вважаю доцільним висловити такі зауваження, коментарі та дискусійні положення:

1. У підрозділі 2.5 метод генерації AMR-графів подано як послідовність перетворень тексту, однак недостатньо явно описано, що відбувається з реченнями, для яких AMR-парсер формує неповний, некоректний або синтаксично пошкоджений граф. Передбачення формального критерію відбракування таких графів або механізму їх корекції дозволило б підвищити надійність роботи запропонованої технології у випадках некоректного чи неповного AMR-парсингу.

2. У підрозділі 3.5 описано модуль формування графових вкладень на основі AMR-графів. Водночас варто було б чіткіше пояснити, яким чином різні типи AMR-відношень між вузлами враховуються параметрами моделі та як ця інформація використовується під час графового кодування й подальшого агрегування інформації.

3. У розділі 3 описано функціональні модулі інтелектуальної системи та алгоритм її роботи. Водночас розмежування етапів, що виконуються під час навчання моделей, та етапів практичного застосування вже навченої системи дозволило б повніше охарактеризувати логіку функціонування запропонованої технології та уточнити обчислювальну роль її окремих компонентів.

4. У третьому розділі наведено фрагменти програмної реалізації окремих модулів системи. Водночас доцільно було б детальніше подати порядок відтворення програмного середовища: версії основних бібліотек, спосіб підключення навчених моделей і послідовність запуску серверної та клієнтської частин.

5. У четвертому розділі під час експериментального дослідження графового кодувальника AMR-графів і методу класифікації перифраз наведено прирости якості за основними метриками. Для повнішої характеристики цих результатів корисним було б додатково навести оцінку їх варіативності, зокрема за повторними запусками моделей або різними розбиттями даних. Це дозволило б ширше охарактеризувати стійкість отриманих експериментальних показників.

6. У таблиці 4.11 наведено розмірність вихідного вектора AMR-графа. Водночас бажано було б коротко обґрунтувати вибір саме цієї розмірності або порівняти результати для кількох варіантів, оскільки цей параметр впливає як на якість векторного представлення, так і на обчислювальні витрати моделі.

Зазначені зауваження мають дискусійний характер, окреслюють напрями подальшого розвитку дослідження та не ставлять під сумнів наукову новизну, достовірність і практичну цінність отриманих результатів.

### ***8. Загальний висновок щодо дисертаційної роботи.***

Дисертація М'якенького Арсенія Вячеславовича на тему «Метод ідентифікації перифраз на основі графових репрезентацій для моделювання когнітивного процесу розуміння» є завершеною кваліфікаційною науковою працею, що містить нові науково обґрунтовані результати у галузі комп'ютерних наук. У роботі вирішено важливе науково-прикладне завдання підвищення точності

ідентифікації перифраз у текстах українською мовою шляхом використання AMR-графів, графових нейронних мереж і класифікаційних моделей.

За актуальністю, змістом, науковою новизною, ступенем обґрунтованості, практичною значущістю, повнотою опублікування результатів і рівнем виконання дисертаційна робота відповідає вимогам наказу Міністерства освіти і науки України від 12.01.2017 р. № 40 «Про затвердження вимог до оформлення дисертації» та постанови Кабінету Міністрів України від 12 січня 2022 р. № 44 «Про затвердження порядку присудження ступеня доктора філософії» (пункти 5, 6, 8).

Вважаю, що автор дисертації, М'якенький Арсеній Вячеславович, заслуговує на присудження ступеня доктора філософії за спеціальністю 122 «Комп'ютерні науки».

***Офіційний опонент:***

професор кафедри інженерії програмного забезпечення  
Івано-Франківського національного технічного  
університету нафти і газу,  
доктор технічних наук, професор

Василь ШЕКЕТА