

Рішення
разової спеціалізованої вченої ради
про присудження ступеня доктора філософії
від 18.08.2026 р.

Здобувач ступеня доктора філософії Олександр Ніколенко, 1972 року народження, громадянин України, освіта вища: закінчив у 1994 році ДВНЗ «Ужгородський національний університет» за спеціальністю «Прикладна математика», працює директором в компанії «Карбук Софт», м. Київ, виконав акредитовану освітньо-наукову програму на здобуття наукового ступеня доктора філософії зі спеціальності 122 «Комп'ютерні науки».

Разова спеціалізована вчена рада утворена наказом Державного вищого навчального закладу «Ужгородський національний університет» Міністерства освіти і науки України, м. Ужгород від «03» червня 2026 року №342/01-04 у складі:

Голови разової спеціалізованої вченої ради –	Олександра Міца, доктора технічних наук, професора, професора кафедри інформаційних управляючих систем та технологій факультету інформаційних технологій ДВНЗ «Ужгородський національний університет» МОН України;
Рецензентів -	Ігоря Ляха, доктора технічних наук, професора, професора інформатики та фізико-математичних дисциплін факультету інформаційних технологій ДВНЗ «Ужгородський національний університет» МОН України; Юрія Андрашка, кандидата технічних наук, доцента, доцента кафедри системного аналізу та теорії оптимізації факультету математики та цифрових технологій ДВНЗ «Ужгородський національний університет» МОН України;
Офіційних опонентів -	Олександра Гожого, доктора технічних наук, професора, професора кафедри інтелектуальних інформаційних систем факультету комп'ютерних наук Чорноморського національного університету імені Петра Могили МОН України; Альони Кудряшової, доктора технічних наук, доцента, доцента кафедри систем віртуальної реальності Інституту комп'ютерних наук та інформаційних технологій Національного університету «Львівська політехніка» МОН України.

на засіданні «18» серпня 2026 року прийняла рішення про присудження ступеня доктора філософії з галузі знань 12 Інформаційні технології за спеціальністю 122 «Комп'ютерні науки» Олександрю Ніколенку на підставі публічного

захисту дисертації на тему: «Методи класифікації та кластеризації для адаптивної таксономії україномовних технічних текстів».

Дисертацію виконано в Державному вищому навчальному закладі «Ужгородський національний університет» Міністерства освіти і науки України, м. Ужгород.

Науковий керівник Сергій Машталір, доктор технічних наук, професор, професор кафедри інформаційних управляючих систем та технологій факультету інформаційних технологій ДВНЗ «Ужгородський національний університет» МОН України.

Дисертацію подано у вигляді спеціально підготовленого рукопису українською мовою, який повністю відповідає вимогам до оформлення дисертації, затвердженим Наказом МОН України від 12.01.2017 р., № 40.

Представлені в дисертації результати характеризуються належним рівнем наукової новизни та розширюють наявні підходи до розв'язання досліджуваної науково-прикладної задачі. Серед основних результатів, одержаних уперше, слід відзначити створення інтегрованого фреймворку для формування та адаптивного оновлення таксономії ремонтних робіт на основі даних GMS Carbook. Запропоноване рішення об'єднує доменно-орієнтоване попереднє оброблення даних, Тор-К ієрархічну класифікацію, семантичну кластеризацію дій, засоби індукції та еволюції таксономії, а також механізми її керування і версіонування. Крім того, вперше розроблено спеціалізований конвеєр оброблення україномовних і змішаних ремонтних записів, який передбачає нормалізацію аббревіатур, лематизацію та виокремлення ключових компонентів. Його застосування дає змогу зменшити варіативність словника без втрати суттєвої для подальшого аналізу інформації.

Удосконалено підхід до Тор-К ієрархічної класифікації завдяки використанню уніфікованого показника впевненості, призначеного для оцінювання та впорядкування ієрархічних шляхів під час мультишляхового Тор-К інференсу. Це забезпечує обґрунтоване застосування порогових значень і маршрутизацію рішень у прикладних сценаріях. Також удосконалено методи формування множини дій та їх семантичної кластеризації. Для цього запропоновано гібридний конвеєр розпізнавання «дія/не дія» у просторі ембеддингів із наступною кластеризацією та оцінюванням когерентності, що дає можливість визначати кластери, придатні для формування вузлів рівня «дія».

Подальшого розвитку набули методи побудови, модифікації та супроводу таксономій. Зокрема, формалізовано процес еволюції структури таксономії, який охоплює операції додавання, об'єднання, декомпозиції та перейменування її елементів. Визначено правила виконання таких структурних перетворень з урахуванням результатів класифікації й кластеризації, а також запропоновано механізми керування версіями таксономії та її інтеграції у промислові

інформаційні системи.

Практична цінність отриманих у дисертаційній роботі результатів визначається можливістю їх використання для вдосконалення формування та супроводу довідників ремонтних робіт, підвищення їх повноти й узгодженості та скорочення частки операцій, що потребують ручного кодування. Застосування Тор-К рекомендацій і механізмів порогової маршрутизації дає змогу прискорити опрацювання ремонтних записів, підвищити якість аналітичного опрацювання даних за видами робіт і системами та створює передумови для подальшої автоматизації бізнес-процесів підприємств автосервісу на основі даних.

Розроблені підходи пройшли практичну перевірку в діяльності ТОВ «Карбук Софт» та на сервісній станції «Автоцентр Зеніт» (ТОВ «Тринік Авто»), а також були використані в освітньому процесі Ужгородського національного університету. Результати такого впровадження засвідчили придатність запропонованого фреймворку до практичного застосування, зокрема для підтримки прийняття рішень, автоматизованого опрацювання даних та використання під час підготовки фахівців у галузі інформаційних технологій.

Здобувач має 7 наукових публікацій за темою дисертації, у тому числі:

- 3 статті у наукових фахових виданнях України (з них 2 включено до наукометричної бази Scopus),
- 1 розділ монографії;
- 3 публікації у матеріалах науково-практичних конференцій (з них 2 матеріали конференції включено до наукометричної бази Scopus).

1. Mashtalir S. V., Nikolenko O. V. Data preprocessing and tokenization techniques for technical Ukrainian texts // *Applied Aspects of Information Technology*. 2023. Vol. 6, No. 3. P. 318–326. DOI: 10.15276/aait.06.2023.22. ISSN 2617-4316 (Print), 2663-7723 (Online).

2. Mashtalir S. V., Nikolenko O. V. Optimizing hierarchical classifiers with parameter tuning and confidence scoring // *Herald of Advanced Information Technology*. 2024. Vol. 7, No. 3. P. 231–242. DOI: 10.15276/hait.07.2024.15. ISSN 2663-0176 (Print), 2663-7731 (Online).

3. Mashtalir S. V., Nikolenko O. V. From classification to taxonomy: automated structuring of vehicle repair names in multilingual corpora // *Herald of Advanced Information Technology*. 2025. Vol. 8, No. 2. P. 151–163. DOI: 10.15276/hait.08.2025.9. ISSN 2663-0176 (Print), 2663-7731 (Online).

4. Mashtalir S. V., Nikolenko O. V. Hierarchical Classification of Ukrainian Technical Texts Using Tree-Based Models: Applications in the Automotive Industry // *Розвитки інформаційно-керуючих систем та технологій : монографія / під наук. ред. В. Вичужаніна. Львів–Торунь : LihaPres, 2024. P. 288–314. DOI: 10.36059/978-966-397-422-4. ISBN 978-966-397-422-4.*

5. Mashtalir S. V., Nikolenko O. V. Tree-Based Classification of the Technical Ukrainian Texts // Advances in Computer Science for Engineering and Education VII: Proceedings of ICCSEEA 2024. Lecture Notes on Data Engineering and Communications Technologies. 2025. Vol. 242. P. 339–348. DOI: 10.1007/978-3-031-84228-3_29. ISBN 978-3-031-84227-6 (Print), 978-3-031-84228-3 (Online).

6. Mashtalir S. V., Nikolenko O. V. Top-K Hierarchical Classification for Precision in Automotive Technical Data Analysis // 2024 IEEE 19th International Conference on Computer Science and Information Technologies (CSIT). Lviv, 2024. P. 1–4. DOI: 10.1109/CSIT65290.2024.10982679. ISBN 979-8-3315-4263-4.

7. Машталір С. В., Ніколенко О. В. Комплексне навчання моделі: налаштування параметрів та оцінка довіри в ієрархічних класифікаторах // Інформаційні управляючі системи і технології (ІУСТ-Одеса-2024) : матеріали XII Міжнародної науково-практичної конференції (23–25 вересня 2024 р., Одеса). Одеса, 2024. С. 93–96. ISBN 978-617-554-315-3.

У дискусії взяли участь голова і члени разової спеціалізованої вченої ради та присутні на захисті фахівці:

Ігор Лях – доктор технічних наук, професор, професор кафедри інформатики та фізико-математичних дисциплін факультету інформаційних технологій ДВНЗ «Ужгородський національний університет» МОН України (рецензент).

Зауваження Ігоря Ляха:

1. Для побудови семантичних подань у роботі використано модель ембеддингів multilingual-e5-large. Вибір моделі аргументовано її багатомовністю та якісними характеристиками, проте експериментального порівняння з альтернативними багатомовними або орієнтованими на українську мову моделями подань у роботі не наведено. Таке порівняння дозволило б кількісно підтвердити оптимальність зробленого вибору для цільового домену.

2. Обчислювальні характеристики конвеєра (підрозділ 4.3.6) оцінено на портативному комп'ютері, що є достатнім для підтвердження принципової придатності рішень. Водночас у роботі не описано архітектуру серверного розгортання класифікаційного модуля в промисловому контурі та не досліджено поведінку конвеєра за зростання навантаження, зокрема можливості горизонтального масштабування, що для системи, що обслуговує понад 500 станцій, має практичне значення.

3. У роботі показано, що механізм індукції вузлів системно працює з довгим хвостом рідкісних формулювань, а за період апробації індуковано 12 вузлів рівня «дія». Проте динаміку ефекту між версіями таксономії не оцінено кількісно: не наведено, як змінюється покриття хвоста корпусу та якість класифікації рідкісних класів після кожного циклу оновлення таксономії, хоча саме ця динаміка характеризує адаптивність запропонованого підходу в часі.

4. У роботі присутні граматичні та термінологічні помилки. Зазначені зауваження мають рекомендаційний характер, стосуються переважно напрямів розширення дослідження і не впливають на загальну позитивну оцінку дисертації.

Юрій Андрашко – кандидат фізико-математичних наук, доцент, доцент кафедри системного аналізу та теорії оптимізації факультету математики та цифрових технологій ДВНЗ «Ужгородський національний університет» МОН України (рецензент).

Зауваження *Юрія Андрашка*:

Незважаючи на загальну високу позитивну оцінку рецензованої дисертаційної роботи, можна сформулювати деякі зауваження:

1. У першому розділі дисертації автором виконано аналіз сучасних підходів до ієрархічної класифікації, семантичної кластеризації та побудови адаптивних таксономій, однак недостатньо повно висвітлено сучасний стан розвитку україномовних технологій обробки природної мови. Доцільно було б розширити огляд систематизованим порівнянням доступних україномовних корпусів, лексичних ресурсів, засобів лематизації (окрім використаного словника ВЕСУМ), а також сучасних мовних моделей і моделей векторних подань тексту, зокрема mBERT, XLM-RoBERTa, SlavicBERT, multilingual-E5, BGE-M3 та інших моделей, що підтримують українську мову. Такий аналіз дозволив би більш аргументовано обґрунтувати вибір використаних інструментальних засобів, чіткіше визначити місце запропонованого підходу серед сучасних методів обробки природної мови та посилити практичну цінність роботи в контексті розвитку українських мовних технологій.

2. У дисертації заявлено розроблення методів для україномовних технічних текстів як класу, проте вся експериментальна база обмежена єдиним пропріетарним корпусом однієї вузької предметної області – автомобільного сервісу. Узагальнювальна здатність запропонованих методів за межами цього домену залишається емпірично непідтвердженою. У зв'язку з цим висновки роботи щодо ефективності методів коректно поширювати лише на розглянуту предметну область, що варто було б явно застережити в тексті дисертації.

Підтвердження загальності підходу потребує додаткової перевірки щонайменше на одному відкритому корпусі іншої галузі технічних текстів.

3. Якість семантичної кластеризації дій у дисертації оцінено за допомогою експертного показника когерентності h , а також показників стабільності кластеризації (ARI, NMI), що загалом є обґрунтованим підходом.

Разом з тим оцінювання могло б бути доповнене використанням внутрішніх метрик якості кластеризації, зокрема силуетного коефіцієнта, індексу Девіса–Болдіна або інших показників, придатних для аналізу щільнісних методів кластеризації. Це дало б змогу більш формалізовано

порівнювати різні конфігурації параметрів алгоритму HDBSCAN (`min_cluster_size`, `min_samples`), оцінити їх вплив на структуру сформованих кластерів та додатково обґрунтувати вибір робочої конфігурації алгоритму.

4. Запропонований механізм порогової маршрутизації рішень є логічним і практично обґрунтованим. Зокрема, у табл. 4.12 наведено залежність між значеннями показника впевненості $c(x)$ та фактичною точністю класифікації, що підтверджує можливість використання цього показника для автоматичного прийняття рішень. Водночас вибір порогового значення здійснено емпірично. На мою думку, цей етап доцільно було б формалізувати як задачу оптимізації з урахуванням компромісу між часткою автоматично прийнятих рішень, ризиком помилкової класифікації та витратами на експертну перевірку. Це дозволило б більш обґрунтовано визначати оптимальне значення порога залежно від умов практичного використання системи.

5. Запропонований у роботі агрегований показник впевненості продемонстрував високу практичну ефективність і підтверджений результатами експериментальних досліджень. Разом із тим аналіз його властивостей доцільно було б доповнити оцінюванням калібрування отриманих імовірнісних оцінок із використанням сучасних кількісних показників, зокрема Expected Calibration Error або Brier Score. Проведення такого аналізу дозволило б кількісно оцінити відповідність прогнозованих рівнів впевненості фактичній імовірності правильної класифікації та додатково посилити теоретичне обґрунтування запропонованого підходу.

Зазначені зауваження стосуються окремих аспектів обґрунтування запропонованих рішень і можливостей подальшого розвитку дослідження. Вони мають рекомендаційний характер, не ставлять під сумнів достовірність і наукову новизну отриманих результатів.

Олександр Гожий – доктор технічних наук, професор, професор кафедри інтелектуальних інформаційних систем факультету комп'ютерних наук Чорноморського національного університету імені Петра Могили МОН України (офіційний опонент).

Зауваження Олександра Гожого:

1. У другому розділі ієрархічну класифікацію реалізовано на основі локальних класифікаторів з інтерпретованими розрідженими поданнями, проте в роботі відсутнє пряме експериментальне порівняння з сучасними неймережевими підходами до класифікації текстів, зокрема з донавчанням трансформерних мовних моделей. Таке порівняння дозволило б точніше позиціонувати запропонований метод на шкалі «якість – обчислювальна вартість» і кількісно обґрунтувати вибір архітектури.

2. У другому розділі в підрозділі 2.5.3 (стор.106), для навчання моделі уніфікованої впевненості використовується ансамбль бустингових дерев рішень, але використання цього типу ансамблевої моделі не обґрунтовано достатньо. В роботі бажано б навести порівняння різних типів ансамблевих моделей. Та визначити їх особливості та переваги для вирішення завдань навчання.

3. Звертає на себе увагу суттєвий розрив між загальною точністю класифікації ($Acc_1 = 0,9238$) та макроусередненою F1-мірою (близько 0,64), що зумовлено вираженим дисбалансом класів і довгим хвостом рідкісних формулювань. У роботі цей ефект констатовано, однак спеціальні прийоми підвищення якості для малочисельних класів (аугментація даних, переважування, ієрархічне згладжування) не досліджувалися, що варто було б принаймні обговорити як напрям удосконалення.

4. Механізм еволюції таксономії повністю реалізовано та експериментально перевірено лише для рівня дій, тоді як його узагальнення на вищі рівні ієрархії (підрозділ 3.4) подано на концептуальному рівні. Проведення хоча б пілотного експерименту для рівня компонентів посилило б обґрунтованість тверджень про багаторівневу застосовність запропонованого підходу.

5. Порогові параметри двозонної політики відбору кластерів ($\tau_{low} = 0,6$; $\tau_{high} = 0,8$) та поріг автоматичного прийняття рішень τ обрано емпірично; аналіз чутливості прикладних показників конвеєра до варіації цих порогів у роботі не наведено, хоча для промислового використання така інформація має істотне значення.

6. У четвертому розділі в підрозділі 4.2.1 (стор.162) для узагальненої оцінки ефекту нормалізації використовується коефіцієнт скорочення словника r_v , який представлено приблизною оцінкою. Бажано було описати алгоритм процесу оцінки ефекту нормалізації з врахуванням особливостей всіх елементів процесу (оцінки хвостової метрики, врахування орфографічних, мовних і морфологічних розбіжностей).

Висловлені зауваження та рекомендації мають здебільшого уточнюючий та дискусійний характер, вони не впливають на загальну високу оцінку дисертаційної роботи, не ставлять під сумнів її наукову новизну, обґрунтованість і практичну цінність.

Альона Кудряшова – доктор технічних наук, доцент, доцента кафедри систем віртуальної реальності Інституту комп'ютерних наук та інформаційних технологій Національного університету «Львівська політехніка» МОН України. (офіційний опонент).

Зауваження *Альони Кудряшової*:

1. Конвеєр препроцесингу включає гармонізацію мови та лематизацію на основі словника ВЕСУМ, однак якість самих цих кроків на доменній лексиці

кількісно не оцінено: у роботі не наведено частки помилкових лематизацій чи некоректних мовних зведень для жаргонних і скорочених форм, хоча такі помилки безпосередньо впливають на подальші етапи класифікації та кластеризації.

2. Показник когерентності кластерів q_{coh} , який відіграє ключову роль у відборі кандидатів на вузли таксономії, визначається через експертне оцінювання релевантності елементів. При цьому в роботі не описано кількість залучених експертів і не наведено оцінок міжекспертної узгодженості (наприклад, каппа Коена), що залишає відкритим питання відтворюваності цього показника.

3. Експериментальну перевірку здійснено на корпусі однієї інформаційної системи одного домену. Твердження про можливість адаптації запропонованих методів до інших технічних галузей є правдоподібним, але в межах роботи його не підкріплено навіть обмеженим експериментом на сторонньому корпусі, що варто розглядати як перспективний напрям подальших досліджень.

4. У тексті трапляються поодинокі редакційні огріхи, зокрема некоректний запис інтервалу впевненості в табл. 4.12 та окремі неуніфіковані позначення діапазонів значень, що не впливають на зміст роботи в цілому.

Висловлені зауваження мають рекомендаційний і частково дискусійний характер, не ставлять під сумнів наукову новизну, обґрунтованість і практичну цінність отриманих результатів та не змінюють загальної позитивної оцінки дисертації.

Результати відкритого голосування:

«За» п'ять членів ради,

«Проти» - немає.

На підставі результатів відкритого голосування разова спеціалізована вчена рада присуджує Олександрю Ніколенку ступінь доктора філософії з галузі знань 12 Інформаційні технології за спеціальністю 122 «Комп'ютерні науки».

Відеозапис трансляції захисту дисертації додається.

Голова разової спеціалізованої
вченої ради



Олександр МІЦА