

# Облікова картка дисертації

## I. Загальні відомості

**Державний обліковий номер:** 0826U000469

**Особливі позначки:** відкрита

**Дата реєстрації:** 05-03-2026

**Статус:** Запланована

**Реквізити наказу МОН / наказу закладу:**



## II. Відомості про здобувача

**Власне Прізвище Ім'я По-батькові:**

1. Нарушинська Ольга Олександрівна

2. Olga O. Narushynska

**Кваліфікація:**

**Ідентифікатор ORCID ID:** 0009-0000-0628-8218

**Вид дисертації:** доктор філософії

**Аспірантура/Докторантура:** так

**Шифр наукової спеціальності:** 122

**Назва наукової спеціальності:** Комп'ютерні науки

**Галузь / галузі знань:** інформаційні технології

**Освітньо-наукова програма зі спеціальності:** Комп'ютерні науки

**Дата захисту:**

**Спеціальність за освітою:** 122 Комп'ютерні науки

**Місце роботи здобувача:** Національний університет "Львівська політехніка"

**Код за ЄДРПОУ:** 02071010

**Місцезнаходження:** вул. Степана Бандери, Львів, 79013, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

### **III. Відомості про організацію, де відбувся захист**

**Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради):** PhD 12012

**Повне найменування юридичної особи:** Національний університет "Львівська політехніка"

**Код за ЄДРПОУ:** 02071010

**Місцезнаходження:** вул. Степана Бандери, Львів, 79013, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

### **IV. Відомості про підприємство, установу, організацію, в якій було виконано дисертацію**

**Повне найменування юридичної особи:** Національний університет "Львівська політехніка"

**Код за ЄДРПОУ:** 02071010

**Місцезнаходження:** вул. Степана Бандери, Львів, 79013, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

### **V. Відомості про дисертацію**

**Мова дисертації:** Українська

**Коди тематичних рубрик:** 20.54.03

**Тема дисертації:**

1. Інформаційна технологія інтелектуальної ієрархічної класифікації коротких описів товарів у сфері електронної комерції
2. Information Technology for Automation of Short Product Descriptions Hierarchical Classification in E-Commerce

**Реферат:**

1. У дисертації розглянуто підвищення ефективності процесів ієрархічної класифікації коротких текстів у сфері електронної комерції шляхом інтеграції ручного маркування даних, методів машинного навчання та сучасних підходів до побудови багаторівневих класифікаційних моделей і стандартизації інформаційних потоків. У вступі обґрунтовано актуальність теми дисертаційної роботи, сформульовано мету та основні завдання досліджень, визначено наукову новизну роботи і практичне значення отриманих результатів, показано зв'язок роботи з науковими темами. Подано відомості про апробацію результатів роботи та особистий внесок автора та його публікації. У першому розділі дисертаційного дослідження здійснено ґрунтовний аналіз сучасних підходів до ієрархічної класифікації коротких текстів, зокрема в контексті електронної комерції, де класифікація товарів потребує особливої точності та узгодженості. Детально розглянуто наукові досягнення у сфері розроблення стандартів, що забезпечують уніфікацію інформаційних

потоків, таких як Global Product Classification (GPC). Увагу приділено класифікаційним моделям, що враховують багаторівневу структуру даних, зокрема глобальним, локальним і плоским моделям, а також їхнім перевагам і недолікам. Особливий акцент зроблено на специфіці коротких текстів, які є обмеженими в контексті, часто містять скорочення, числові позначення та унікальні терміни. Аналіз показав, що традиційні моделі, хоча й демонструють певні успіхи, потребують адаптації до умов складних ієрархій і високого рівня дисбалансу даних. На основі аналізу сформульовано ключові вимоги до сучасних моделей ієрархічної класифікації: точність результатів навіть у випадку рідкісних категорій, ефективність роботи з великими обсягами даних, масштабованість і здатність до автоматизованого розширення функціоналу. У другому розділі вдосконалено процес синтезу навчальних наборів даних для задач ієрархічної класифікації коротких текстів. Побудовано метод оптимізації процесу ручного маркування даних, який ґрунтується на впорядкуванні текстів за допомогою методу упорядкування листків у ієрархічній кластеризації. Цей метод дає змогу розташовувати схожі описи товарів поруч, що значно знижує когнітивне навантаження на анотаторів і підвищує ефективність процесу маркування. Підхід інтегровано в інструменти для маркування даних, такі як Label Studio, де налаштування відповідали стандарту GPC і забезпечували точність анотацій на всіх рівнях ієрархії. Впровадження сортування даних забезпечило скорочення часу маркування та забезпечити узгодженість між експертами. Крім того, вдосконалено метод опрацювання коротких текстів, який базується на інтеграції методу кодування пар байтів (Byte Pair Encoding, BPE) із контекстуальними вкладеннями слів (англ. word embedding) BERT. Поєднання цього методу з BERT моделлю забезпечило синтез контекстуальних репрезентативних векторних представлень тексту, які враховують як локальні семантичні зв'язки між підсловами, так і глобальний контекст. Така інтеграція дала змогу підвищити точність класифікації, особливо на нижніх рівнях ієрархії, де деталізація даних є найбільш критичною. Розроблений метод опрацювання коротких текстів також сприяв зменшенню розмірності ознакового простору, що знижує обчислювальні витрати та мінімізує ризик перенавчання моделі. Такий метод є універсальним і може бути інтегрований у широкий спектр задач, пов'язаних із ієрархічною класифікацією текстів. У третьому розділі розроблено модель ієрархічної класифікації текстів, засновану на показнику штрафного інформаційного приросту (Penalized Information Gain, PIG) та його динамічній модифікації, що дає змогу адаптивно враховувати ієрархічні зв'язки між класами, мінімізувати каскадні помилки та забезпечувати вищу точність передбачення. Проведені експерименти продемонстрували ефективність моделей у задачах класифікації, що підтверджується підвищенням точності на різних рівнях ієрархії. Додатково, інтеграція цих моделей у бізнес-процеси електронної комерції показала практичну цінність у реальних умовах, де точність класифікації безпосередньо впливає на ефективність логістичних і комерційних процесів. У четвертому розділі розроблено розподілену інформаційну технологію для автоматизації ієрархічної класифікації коротких текстів. Розроблена структура ґрунтується на модульному підході, що забезпечує масштабованість і гнучкість системи. У технології використано сучасні методи інтеграції даних і стандарти GPC, що уможливорює ефективну роботу з багатомовними й багатокультурними ринками. Особливу увагу приділено забезпеченню сумісності та відкритості програмно-апаратного забезпечення, що забезпечує умови для широкого використання технології.

2. The dissertation addresses improving the efficiency of hierarchical classification processes for short texts in the field of e-commerce through the integration of manual data annotation, machine learning methods, and modern approaches to building multi-level classification models and standardizing information flows. The introduction substantiates the relevance of the dissertation topic, formulates the aim and main objectives of the research, defines the scientific novelty and practical significance of the obtained results, and demonstrates the connection of the work with ongoing research themes. Information on the approbation of the results, as well as the author's personal contribution and publications, is provided. The first chapter presents a comprehensive analysis of contemporary approaches to hierarchical classification of short texts, particularly in the context of e-commerce, where product classification requires high precision and consistency. Scientific achievements in the development of standards that ensure the unification of information flows, such as the Global Product Classification (GPC), are examined in detail. Attention is given to classification models that account for multi-level data structures,

including global, local, and flat models, along with their advantages and limitations. Special emphasis is placed on the specificity of short texts, which are context-limited and often contain abbreviations, numerical designations, and domain-specific terminology. The analysis demonstrates that traditional models, while achieving certain success, require adaptation to complex hierarchies and significant data imbalance. Based on the analysis, key requirements for modern hierarchical classification models are formulated: high accuracy even for rare categories, efficiency in processing large-scale datasets, scalability, and the ability to support automated functional expansion. The second chapter improves the process of constructing training datasets for hierarchical classification tasks involving short texts. A method for optimizing manual data annotation is developed, based on text ordering using the leaf ordering technique in hierarchical clustering. This approach allows similar product descriptions to be positioned adjacently, significantly reducing annotators' cognitive load and increasing labeling efficiency. The method is integrated into data annotation tools such as Label Studio, where configurations were aligned with the GPC standard to ensure annotation accuracy at all hierarchy levels. The implementation of data sorting reduced labeling time and ensured inter-expert consistency. Additionally, a method for short text processing is enhanced by integrating Byte Pair Encoding (BPE) with contextual word embeddings generated by BERT. The combination of BPE and BERT enables the synthesis of contextual vector representations that capture both local semantic relationships between subwords and the global textual context. This integration improves classification accuracy, particularly at the lower levels of the hierarchy, where fine-grained distinctions are most critical. The developed short text processing method also contributes to reducing the dimensionality of the feature space, thereby decreasing computational costs and minimizing the risk of model overfitting. The proposed approach is universal and can be integrated into a wide range of hierarchical text classification tasks. The third chapter introduces a hierarchical text classification model based on the Penalized Information Gain (PIG) metric and its dynamic modification, which adaptively accounts for hierarchical relationships between classes, minimizes cascading errors, and ensures higher prediction accuracy. Experimental results demonstrate the effectiveness of the proposed models in classification tasks, confirmed by improved accuracy across different hierarchy levels. Furthermore, integrating these models into e-commerce business processes demonstrates their practical value in real-world conditions, where classification accuracy directly affects the efficiency of logistics and commercial operations. The fourth chapter presents a distributed information technology framework for automating the hierarchical classification of short texts. The developed architecture is based on a modular approach, ensuring system scalability and flexibility. The technology incorporates modern data integration methods and GPC standards, enabling effective operation in multilingual and multicultural markets. Particular attention is given to ensuring software and hardware compatibility and openness, which creates conditions for the broad adoption of the technology.

**Державний реєстраційний номер ДіР:**

**Пріоритетний напрям розвитку науки і техніки:** Інформаційні та комунікаційні технології

**Стратегічний пріоритетний напрям інноваційної діяльності:** Розвиток сучасних інформаційних, комунікаційних технологій, робототехніки

**Підсумки дослідження:** Нове вирішення актуального наукового завдання

**Публікації:**

- 1. Нарушинська О. О., Мотрунич В. І., Арзубов М. В., Теслюк В. М. Інформаційна технологія для аналізу пунктів продажу мобільного оператора на основі методів кластеризації. Український журнал інформаційних технологій. 2023. Т. 5, № 2. С. 105–113. <https://doi.org/10.23939/ujit2023.02.105>
- 2. Лозинський А. Я., Машевська М. В., Нарушинська О. О., Теслюк В. М. Застосування мультиагентного підходу для покращення параметрів синхронізації в ігрових системах. Моделювання та інформаційні технології. 2017. Вип. 79. С. 183–190. <http://jnas.nbuv.gov.ua/uk/article/UJRN-0000762182>

- 3. Нарушинська О. О., Теслюк В. М., Денисюк П. Ю. Застосування методів машинного навчання для мінімізації повернень товару в системах електронної комерції. Актуальні проблеми економіки. 2017. № 3 (189). С. 342–347.  
[https://www.researchgate.net/publication/328998595\\_MACHINE\\_LEARNING\\_METHODS\\_APPLICATION\\_FOR\\_COMMERCE](https://www.researchgate.net/publication/328998595_MACHINE_LEARNING_METHODS_APPLICATION_FOR_COMMERCE)
- 4. Narushynska O., Arzubov M., Teslyuk V. Construction of hierarchical classification model for product management: Penalized Information Gain considering dynamic weight coefficients. Східно-Європейський журнал передових технологій. 2025. №1/3 (133). С. 17–27. <https://doi.org/10.15587/1729-4061.2025.321273>
- 5. Narushynska O., Teslyuk V., Doroshenko A., Arzubov M. Data sorting influence on short text manual labeling quality for hierarchical classification. Big Data and Cognitive Computing. 2024. Vol. 8, iss. 4. 41. <https://doi.org/10.3390/bdcc8040041>
- 6. Narushynska O., Doroshenko A., Teslyuk V., Antoniv V., Arzubov M. Enhancing hierarchical classification in tree-based models using level-wise entropy adjustment. Big Data and Cognitive Computing. 2025. Vol. 9, iss. 3. 65. <https://doi.org/10.3390/bdcc9030065>
- 7. Arzubov M., Narushynska O., Batyuk A., Cherkas N. Concept of server-side clusterization of semi-static big geodata for web maps. International Scientific and Technical Conference on Computer Sciences and Information Technologies. 2023. 18th IEEE Intern. conf. on computer science and information technologies, CSIT 2023, Lviv, Ukraine, 19–21 Oct. 2023. <https://doi.org/10.1109/CSIT61576.2023.10324155>
- 8. Teslyuk V., Doroshenko A., Narushynska O. Preprocessing product descriptions with byte pair encoding: a solution for abbreviation-heavy texts. CEUR Workshop Proceedings. 2024. Vol. 3861 : Proc. of the computational intelligence application workshop (CIAW 2024), Lviv, Ukraine, Oct. 10–12, 2024. P. 57–76. <https://ceur-ws.org/Vol-3861/paper5.pdf>
- 9. Narushynska O., Teslyuk V., Arzubov M., Martyniuk T., Trachuk T. Module for the analysis of digital skin images aimed at early diagnosis of dermatological conditions based on deep learning methods. CEUR Workshop Proceedings. 2025. Vol. 4013 : Proc. of the 2nd Intern. workshop on intelligent & cyberphysical systems ICyberPhyS 2025, Khmelnytskyi, Ukraine. July 4, 2025. P. 97–110. <https://ceur-ws.org/Vol-4013/paper8.pdf>
- 10. Teslyuk V., Narushynska O., Arzubov M., Prots D. End-to-end development of a retrieval-augmented large language model for cloud-based healthcare applications. CEUR Workshop Proceedings. 2025. Vol. 4005 : Proc. of the Modern data science technologies doctoral consortium MoDaST 2025, Lviv, Ukraine, June 15–16, 2025. P. 141–162. <https://ceur-ws.org/Vol-4005/paper11.pdf>

**Наукова (науково-технічна) продукція:** методи, теорії, гіпотези; програмні продукти, програмно-технологічна документація

**Соціально-економічна спрямованість:** забезпечення промисловості чи населення новим видом інформаційно-комунікаційних послуг

**Охоронні документи на ОПВ:**

**Впровадження результатів дисертації:** Впроваджено

**Зв'язок з науковими темами:** 0124U000822; 0122U000891

## **VI. Відомості про наукового керівника/керівників (консультанта)**

**Власне Прізвище Ім'я По-батькові:**

1. Теслюк Василь Миколайович
2. Vasyl M. Teslyuk

**Кваліфікація:** д.т.н., професор, 05.13.06

**Ідентифікатор ORCID ID:** 0000-0002-5974-9310

**Додаткова інформація:** завідувач кафедри автоматизованих систем управління Інститут комп'ютерних наук та інформаційних технологій

**Повне найменування юридичної особи:** Національний університет "Львівська політехніка"

**Код за ЄДРПОУ:** 02071010

**Місцезнаходження:** вул. Степана Бандери, Львів, 79013, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

## **VII. Відомості про офіційних опонентів та рецензентів**

### **Офіційні опоненти**

#### **Власне Прізвище Ім'я По-батькові:**

1. Бісікало Олег Володимирович
2. Oleg V. Bisikalo

**Кваліфікація:** д.т.н., професор, 05.13.06

**Ідентифікатор ORCID ID:** 0000-0002-7607-1943

**Додаткова інформація:** Профіль у Scopus:

<https://www.scopus.com/authid/detail.uri?authorId=57105837600>; Профіль у WoS:

<https://www.webofscience.com/wos/author/record/J-9715-2015>; Профіль у Google Scholar:

[https://scholar.google.com.ua/citations?user=YoHjJ\\_wAAAAJ&hl=en](https://scholar.google.com.ua/citations?user=YoHjJ_wAAAAJ&hl=en)

**Повне найменування юридичної особи:** Вінницький національний технічний університет

**Код за ЄДРПОУ:** 02070693

**Місцезнаходження:** вул. Хмельницьке шосе, Вінниця, Вінницький р-н., 21021, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

#### **Власне Прізвище Ім'я По-батькові:**

1. Мулеса Оксана Юріївна
2. Oksana Y. Mulesa

**Кваліфікація:** д.т.н., професор, 05.13.06

**Ідентифікатор ORCID ID:** 0000-0002-6117-5846

**Додаткова інформація:**

**Повне найменування юридичної особи:** Державний вищий навчальний заклад "Ужгородський національний університет"

**Код за ЄДРПОУ:** 02070832

**Місцезнаходження:** вул. Підгірна, Ужгород, Ужгородський р-н., 88000, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

## Рецензенти

**Власне Прізвище Ім'я По-батькові:**

1. Піх Ірина Всеволодівна

2. Iryna V. Pikh

**Кваліфікація:** д. т. н., професор, 05.13.06

**Ідентифікатор ORCID ID:** 0000-0002-9909-8444

**Додаткова інформація:**

**Повне найменування юридичної особи:** Національний університет "Львівська політехніка"

**Код за ЄДРПОУ:** 02071010

**Місцезнаходження:** вул. Степана Бандери, Львів, 79013, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

**Власне Прізвище Ім'я По-батькові:**

1. Литвин Василь Володимирович

2. Vasyl V. Lytvyn

**Кваліфікація:** д. т. н., професор, 01.05.03

**Ідентифікатор ORCID ID:** 0000-0002-9676-0180

**Додаткова інформація:** Scopus Author ID: 56446930100; Web of Science Researcher ID: P-8557-2016;  
<https://scholar.google.com.ua/citations?user=bleBe9kAAAAJ>

**Повне найменування юридичної особи:** Національний університет "Львівська політехніка"

**Код за ЄДРПОУ:** 02071010

**Місцезнаходження:** вул. Степана Бандери, Львів, 79013, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

## VIII. **Заключні відомості**

**Власне Прізвище Ім'я По-батькові  
голови ради**

Цмоць Іван Григорович

**Власне Прізвище Ім'я По-батькові  
головуючого на засіданні**

Цмоць Іван Григорович

**Відповідальний за підготовку  
облікових документів**

Віктор Михайлович Хавалко

**Реєстратор**

Юрченко Тетяна Анатоліївна

**Керівник відділу УкрІНТЕІ, що є  
відповідальним за реєстрацію наукової  
діяльності**



Юрченко Тетяна Анатоліївна