



**ЕЛЕКТРОННІ РЕСУРСИ  
ТА ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНІ ТЕХНОЛОГІЇ**  
**ELECTRONIC RESOURCES AND INFORMATION  
AND COMMUNICATION TECHNOLOGIES**

---

УДК 004.421:[006(083.7):81](=161.2)

DOI: <https://doi.org/10.31866/2617-796X.9.1.2026.362659>

**Костянтин Ткаченко,**

*кандидат економічних наук, доцент,*

*доцент кафедри інженерії програмного забезпечення,*

*Національний університет «Київський авіаційний інститут»,*

*Київ, Україна*

*tkachenko.kostyantyn@gmail.com*

*<https://orcid.org/0000-0003-05493396>*

**ВИКОРИСТАННЯ ОНТОЛОГІЧНОГО МОДЕЛЮВАННЯ  
ТА ЛІНГВІСТИЧНОГО АНАЛІЗУ ПІД ЧАС РОБОТИ З ТЕКСТАМИ  
НОРМАТИВНИХ ДОКУМЕНТІВ**

На сьогодні розробка та перевірка відповідних нормативних документів є копітким процесом, який часто виконують без використання сучасних інформаційних систем і технологій. У роботі розглянуто проблему, актуальність якої не викликає сумнівів щодо розробки підходу до формування текстів нормативної документації на основі використання онтологій та методів лінгвістичного аналізу текстів національних стандартів.

**Метою статті** є аналіз і дослідження проблем моделювання процесів формування документації для продукції різних галузей на основі поєднання онтологічного моделювання та лінгвістичного аналізу українських текстів відповідних стандартів.

**Методами дослідження** є методи систематизації підходів до процесів обробки текстів та формалізації цих процесів у вигляді відповідної онтологічної моделі, порівняльного аналізу основних алгоритмів і методів лінгвістичного аналізу українськомовних текстів.

**Новизною проведеного дослідження** є аналіз проблем автоматизованого розроблення та перевірки українськомовних текстів нормативних документів; розробка підходу, що ґрунтується на поєднанні онтологічного моделювання відповідної предметної області та лінгвістичного аналізу текстів і розроблення відповідної інформаційної системи, що базується на запропонованому підході.

**Висновки.** У роботі досліджено та визначено основні проблеми розробки та перевірки українськомовних текстів нормативних документів. Крім того, проведено аналіз і систематизацію сучасних підходів, що підтримують окремі етапи обробки текстів нормативних документів. Запропоновано підхід до обробки українськомовних текстів нормативних документів, який поєднує онтологічне моделювання відповідних предметних областей та методи лінгвістичного аналізу текстів, що надані природною мовою. Запропоновано алгоритми та методи обробки нормативних документів ДСТУ й ТУ, які ґрунту-

ються на адаптації наявних алгоритмів і методів до предметної області, що розглядається. Описано інформаційну систему щодо обробки (розробки та перевірки) нормативних документів ДСТУ й ТУ, яка ґрунтується на використанні онтологічних моделей відповідних предметних областей та методів лінгвістичного аналізу українськомовних текстів.

**Ключові слова:** українськомовний текст; нормативний документ; стандарти та технічні умови на продукцію; онтологічне моделювання; лінгвістичний аналіз; інформаційна система.

**Вступ.** На сьогодні розробка нормативних документів (НД) згідно зі стандартами (національними, світовими) з виготовлення будь-якої продукції (промислових чи продовольчих товарів, програмних продуктів в ІТ-галузі, результатів будівництва, наукових досліджень тощо) є складним процесом, що обумовлено, зокрема, недостатнім рівнем автоматизованого формування текстів документного супроводження продукції, використання інформаційних (інтелектуальних) технологій для генерації відповідних текстів документації, використання комп'ютерно-лінгвістичних технологій аналізу українськомовних текстів наявних стандартів для подальшої генерації відповідних текстів нормативної (стандартизованої – такої, що розробляється відповідно до стандартів) документації, яка супроводжує процес виробництва продукції.

Чинні в Україні вимоги до основної технологічної та технічної документації регламентують детальний опис численних атрибутів (характеристик, параметрів оцінювання, властивостей) продукту виробництва, у тому числі й ІТ-продуктів, у тексті відповідного НД (Всеукраїнський державний науково-виробничий центр стандартизації, метрології, сертифікації і захисту прав споживачів, 2016). Це стосується й документа «Технічні умови» (ТУ), який зазвичай розробляють на основі документа «Державний стандарт України» (ДСТУ).

Структура документів ДСТУ і ТУ на споріднений вид продукції має бути однаковою і в основному не залежати від окремих відмінностей у продукті, що описується в ТУ (порівняно з описом згідно з ДСТУ). До складу такого документа входить певна частина стандартних розділів, які можна перевіряти відповідними уніфікованими програмними засобами. Зазначені документи містять такі стандартні розділи: титульний аркуш; основну частину; обов'язкові рекомендовані та довідкові додатки (за необхідності); аркуш реєстрації змін.

Необхідність регламентації правил використання продукту в ТУ визначає розробник, додаючи до ТУ додаткові розділи. Наприклад, розділ «Правила застосування», який розміщують після розділу «Правила транспортування та зберігання». Крім того, ТУ можуть бути доповнені обов'язковими, рекомендованими й довідковими додатками. Довідкові та рекомендовані додатки розміщують після обов'язкових, а в разі їх відсутності – після основної частини в порядку, що відповідає порядку посилань на них у тексті ТУ. Останнім розміщують довідковий додаток з переліком посилальних документів.

Розробка і перевірка відповідних ДСТУ і ТУ – це копіткий і тривалий процес, у якому беруть участь як спеціалісти конкретної предметної галузі виробництва,

у тому числі і в ІТ-галузі, так і фахівці в галузі стандартизації. На сьогодні, на жаль, ця робота поки що виконують в основному без залучення сучасних інформаційних систем і технологій. Тому актуальність розробки підходу до формування текстів відповідної технічної та технологічної документації на основі використання онтологій та методів лінгвістичного аналізу текстів національних стандартів (наприклад, для різних галузей виробництва, ІТ-сфери, науки тощо) не викликає сумнівів.

На сьогодні основну частину роботи під час формування документації, що розробляють згідно з відповідними галузевими стандартами, проводять без використання інформаційних (інтелектуальних) технологій і систем. В ІТ-галузі є спроби автоматизувати деякі роботи під час формування документації. Наприклад, (Верховна Рада України, 2014) роблять спробу визначення та формалізованого опису структури відповідного документа з переліком його основних складників. Але процес автоматизації розробки документації в цій роботі не розглядається.

Побудову якісної документації до програмного забезпечення розглянуто в роботах багатьох вітчизняних і зарубіжних учених (Марковець та Синько, 2021; Isay, 2021). У вищенаведеній праці акцентовано на проведенні аналізу коректності та якості документації, що формується, та визначенні переліку вимог щодо оцінки її якості. Серед цих вимог автори виділяють зручність у використанні та максимальну інформативність. Здійснено спробу описати процес формування документації, але інформатизацію цього процесу не розглянуто.

У дослідженні українських учених (Пасічник та Яромиш, 2025) проаналізовано переваги й обмеження автоматизованого підходу щодо розробки нормативної документації до програмних продуктів. Використання великих мовних моделей у процесі генерації відповідної документації стало підґрунтям розробки відповідного програмного забезпечення, що реалізує ці моделі та використовує механізми GPT-4 для генерації технічних описів та інструкцій користувача. Практичне застосування запропонованого підходу визначило й низку проблем. Зокрема для повної відповідності вимогам стандартів може знадобитися додаткове редагування текстів документів і включення більш детальних описів параметрів.

У роботі, присвяченій проблемам автоматизації формування документів (Ярка та ін., 2016), запропоновано для формування документів використовувати розроблені шаблони. Такий підхід, крім переваг, що втілюють збільшені можливості автоматизації, є не гнучким і не динамічним. Проведений аналіз останніх досліджень і публікацій ще раз свідчить про необхідність розробки нового (наприклад, онтологічного) підходу до генерації відповідної документації та обґрунтовує актуальність запропонованого в статті підходу, який поєднує використання онтологічного моделювання предметної області (ПрО), що розглядається, та методів і технологій лінгвістичного аналізу українськомовних текстів.

*Мета статті.* Метою є аналіз і дослідження проблем та процесів формування документації для продукції різних галузей на основі поєднання онтологічного моделювання та лінгвістичного аналізу українських текстів відповідних стандартів. Досягнення мети передбачає виконання таких завдань:

– визначення основних проблем формування документації для різних категорій продуктів;

- проведення аналізу та систематизації наявних підходів до формування нормативної документації;
- проведення аналізу та систематизації деяких сучасних методів та алгоритмів лінгвістичного аналізу українськомовних текстів нормативної документації;
- опис авторського підходу до формування нормативної документації, що ґрунтується на поєднанні онтологічного моделювання відповідної ПрО та лінгвістичного аналізу українськомовних текстів нормативної документації.

**Результати дослідження.** *Онтологічні моделі нормативних документів.* У процесі створення, збереження та перевірки НД запропоновано використовувати семантичну модель (базу знань), що містить засоби обробки таких документів і, зокрема, ТУ для виробництва та споживання продуктів відповідно до стандартів України та міжнародних стандартів.

Під час створення онтологічної моделі можуть використовувати різні методи: на початковому етапі – методи ручного проектування онтології; на деяких наступних етапах – методи автоматизації створення онтології, наприклад вилучення термінів з тексту, класифікацію термінів, вилучення об'єктів, їх відносин, властивостей тощо.

У роботі запропоновано використовувати базову онтологію НД і онтологію конкретної ПрО. Під час створення зазначених моделей та їх наповнення визначають для конкретної ПрО основні об'єкти (терміни, поняття) та зв'язки між ними, знаходять посилання на відповідні розділи наявних нормативних і регулювальних документів (на цьому етапі проводять аналіз різних вихідних документів для створення основних класів об'єктів та відношень між ними).

Розроблена модель НД проектується та відіграє роль інтелектуального ядра під час надання засобів для автоматизованої перевірки наявного чи розробки нового НД. Під час створення та наповнення моделі використовують розроблений алгоритм ідентифікації термінів НД, що сприяє автоматичному формуванню нових сутностей у процесі парсингу (Галка, 2025) та перевірки цих документів.

Онтологічне моделювання передбачає визначення класів, об'єктних властивостей, властивостей даних та аксіом. Згідно з результатами проведеного аналізу НД, розроблених відповідно до наявних стандартів (Всеукраїнський державний науково-виробничий центр стандартизації, метрології, сертифікації і захисту прав споживачів, 2016; ДСТУ 4163:2020, 2021), виділено такі структурні компоненти, що становитимуть основу онтологічного моделювання відповідних НД: структура документа; поняття (терміни), які використовуються в документі; типи даних, які використовуються в документі; лінгвістичні характеристики тексту НД.

Онтологія «Нормативний документ» (рис. 1) вміщує такі основні класи: документ з підкласами «Стандарт» (галузевий, державний міжнародний), «Закон», «Технічні умови», «Автор» (орган влади, технічний комітет, спеціаліст), «Класифікатор» (предметні галузі), «Кодекс усталеної практики», «Регламент».

Клас «Документ» має властивості даних (назва, ідентифікатор, розробник, дата створення, дата поновлення, номер редакції) та об'єктні властивості (Має\_Галузь\_застосування, Має\_Нормативні посилання, Має\_Розробника та інші). Онтологія має 28 класів і підкласів, 9 об'єктних властивостей і 13 властивостей даних.

Граф моделі «Нормативний документ», побудований за допомогою інструменту OntoGraf, відображає основні класи та зв'язки між ними (рис. 2).

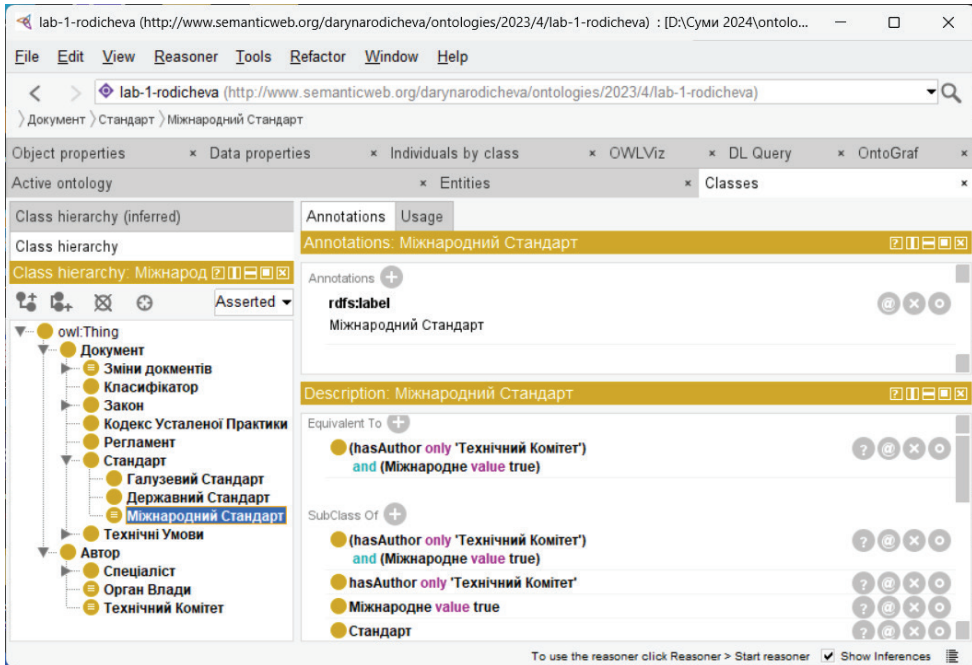


Рис. 1. Приклад опису класу «Міжнародний стандарт» онтології «Нормативний документ» в редакторі Protege.  
Джерело: авторська розробка

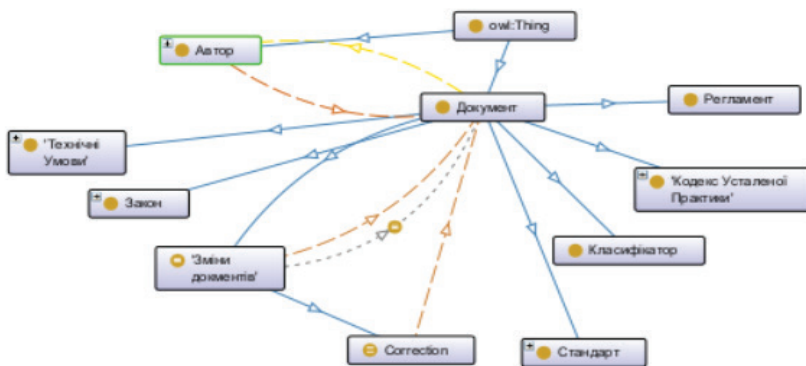


Рис. 2. Граф моделі «Нормативний документ».  
Джерело: авторська розробка

Відмінними рисами онтології НД, зокрема, є: гнучкість та можливість налаштування під час описування структурних компонентів документів; можливість створення нових сутностей, що входять як терміни (поняття) до складу класифікаторів і словників відповідних ПрО; можливість поповнення онтології новими сутностями в процесі її перевірки.

Така онтологія НД в конкретній ПрО (галузі знань або виробництва) є онтологією верхнього рівня (базовою онтологією), яка охоплює основні поняття та відношення, що використовуються для опису й формалізації знань конкретної ПрО, об'єднаних загальним призначенням. У дослідженні таку базову онтологію використано під час створення та поповнення онтології харчових продуктів у галузі сиро-молочного виробництва.

Онтологія містить пов'язані поняття, властивості й анотації знань про харчові молочні продукти, поживні речовини, харчові добавки й іншу інформацію (калорійність, вміст жиру, наявність радіонуклідів, металів), яку вдалося знайти в різних джерелах. Поточна версія онтології містить 78 класів і підкласів, 9 об'єктних властивостей, 50 властивостей даних, 194 індивідуали, 2452 семантичних правила та аксиоми. Сформовано ієрархію із шести основних класів, які мають дочірні класи:

- Instruments – обладнання та механізми, які використовують для обробки сировини та матеріалів;
- Package – упаковка;
- Production – готова продукція (молочні вироби), які виробляють на заводі та доставляють до магазинів;
- Sources – матеріали та сировина, які використовують для виготовлення молочних продуктів;
- Territories – цехи та ділянки заводів з обробки ресурсів та виготовлення готової продукції;
- Workers – персонал заводів.

У процесі роботи з онтологією використано (періодично) різні машини логічного виведення (Reasoner) для виявлення нових знань і перевірки онтології на невідповідність.

На рис. 3 показано результат парсингу сайту (Інформаційне забезпечення у сфері технічного регулювання ДП «УКРМЕТРТЕСТСТАНДАРТ», б.д.) для перевірки актуальності ДСТУ щодо сиру, а також доповнення описаної вище онтології відповідною анотацією індивідуалу.

Одним зі способів виділення початкового набору понять (термінів) є забезпечення процесу вилучення назв сутностей ПрО в результаті парсингу текстів за допомогою пошуку ключових термінів та їх класифікації щодо класів і властивостей онтології. Це є завданням розпізнавання іменованих сутностей (*Named Entity Recognition (NER)*, n.d.), коли в тексті слід виділяти безперервні фрагменти тексту як сутності із задалегідь заданого набору категорій.

Автоматизація побудови термінологічного ядра онтології під час розробки та перевірки НД з використанням базової онтології (зі стандартизації), онтології конкретної ПрО та корпусу текстів у цій ПрО сприяють суттєвому підвищенню ефективності якості та швидкості виконання такої роботи.

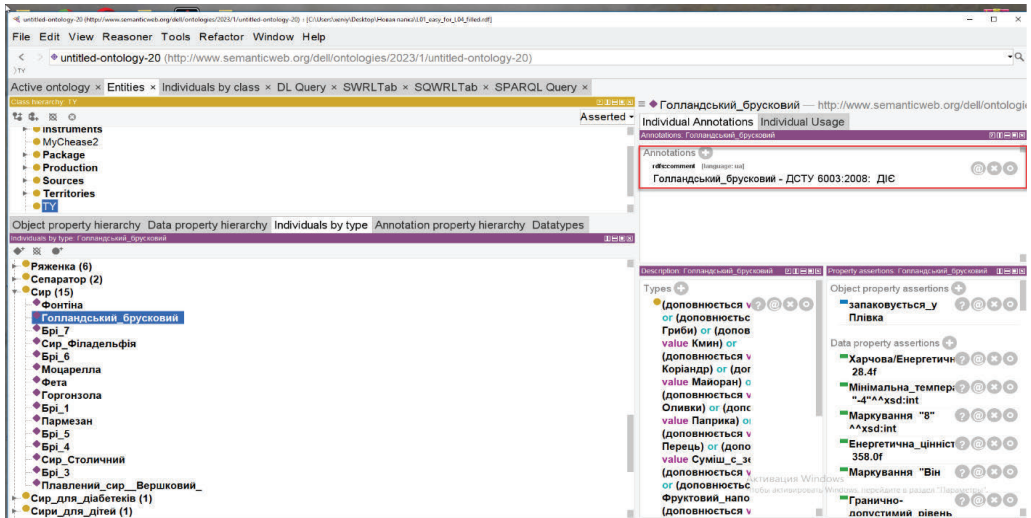


Рис. 3. Доповнення онтології інформацією про актуальний ДСТУ для індивідуалу «Сир Голландський брусковий».

*Джерело:* авторська розробка

Розроблені алгоритми й відповідні програмні засоби для вирішення цієї задачі сприяють виявленню відсутніх понять та уточненню основних відношень, визначенню помилок і невідповідностей моделі ще на етапі її формування.

Варто описати декілька методів вилучення термінів і генерації сутностей онтологічної моделі з використанням розроблених алгоритмів, частина яких була реалізована програмно. У процесі експериментів зібрано й оброблено корпус українських НД з вебсайтів (Інформаційне забезпечення у сфері технічного регулювання ДП «УКРМЕТРТЕСТСТАНДАРТ», б.д.; Бізнес-портал «Леонорм», б.д.), а також з інших постачальників контенту.

В алгоритмі вилучення термінів та їх обробки спочатку виконується парсинг вихідного НД, потім до українськомовного тексту застосовуються методи комп'ютерної лінгвістики: токенизація, вилучення СТОП-слів та символів пунктуації, лематизація (приведення словоформ до нормальної (словникової) форми), стемінг (приведення до основної словоформи), підрахунок сформованих векторних представлень tokenів і словосполучень. Оброблений у такий спосіб вихідний документ готується до подальшого пошуку та перевірки термінів відповідно до сутностей об'єктної моделі. Застосування таких методів обробки текстів НД дає змогу під час пошуку термінів оцінювати ступінь їхньої схожості, порівнювати їх з уже наявними в онтології, знаходити нові.

Під час обробки тексту й пошуку термінів, що є релевантними сутностям онтології, використовуються тезауруси, словники української мови (Словник.ua., б.д.; Про сайт, б.д.; Граматичний словник української мови, б.д.; Lang-uk projects, n.d.) та класифікатори зі стандартизації (Державний класифікатор продукції та послуг ДК 016-2010, б.д.; Науково-дослідний інститут метрології вимірювальних і управляючих систем, 2020; Пошуковий інструмент кодів УКТЗЕД за обраними товарами,

б.д.). Водночас у процесі обробки можуть утворюватися (генеруватися) необхідні n-грами, словосполучення типу «іменник-прикметник», «прикметник-іменник» або певні фрази, якщо того вимагає вбудований алгоритм пошуку з набору слів.

Пошук у тексті термінів та підрядків виконується на основі повнотекстового пошуку за ключовим словом з використанням регулярних виразів. Результат пошуку не залежить від обов'язкової підтримки певного формату, а ґрунтується тільки на відповідності до зафіксованого заданого шаблону. Також пошук виконується за розміром і місцем розташування терміна / словосполучення в тексті документа. Точність такого пошуку залежить від якості створеного документа в сенсі його оформлення відповідно до заданого шаблону. Може бути використано метод підрахунку кількості слів певного типу у фрагменті документа, який вміщує необхідні дані.

У процесі обробки деяких документів доцільно використовувати попередню їх розмітку спеціальними символами, вибір яких виконують експерти в ПрО, що розглядається. Аналіз відбувається за допомогою спеціальної програми вилучення термінів для подальшого їх перетворення в триплети. Інша програма може обробляти вихідні документи без розмітки, і результати роботи двох програм порівнюються з метою оцінки точності запропонованих алгоритмів.

*Алгоритми перевірки нормативного документа.* Структурованість НД типу ТУ (надалі мова буде йти саме про цей тип документа), яка відображається у створеній онтологічній моделі, дає змогу автоматизувати процес перевірки розробленого документа. Процес перевірки може бути суміщений з процедурою формування простору (множини) сутностей або поповнення моделі релевантною інформацією індивідуалів.

Більша частина розділів ТУ має відповідати розділам ДСТУ, на базі якого створено ТУ. Тому перевірка на відповідність ДСТУ є необхідним етапом. Загалом кожен документ ТУ може мати стандартний набір розділів, тому їх можливу наявність і вміст слід перевіряти. У разі відсутності розділу в ТУ, що є у відповідному ДСТУ, слід надати відповідне повідомлення. Під час перевірки кожного розділу ТУ виконують певні дії щодо поповнення / перевірки онтології.

*Перевірка титульної сторінки.* Назва ТУ має розташовуватися в середній частині титульної сторінки під блоками тексту «Затверджую». Назва документа має починатися з іменника, який повинен відповідати назві продукту згідно з кодом з класифікаторів ДКПП (Державний класифікатор продукції та послуг ДК 016-2010, б.д.), УКНД (Науково-дослідний інститут метрології вимірювальних і управляючих систем, 2020), УКТЗЕД (Пошуковий інструмент кодів УКТЗЕД за обраними товарами, б.д.). Під час перевірки виконується приведення першого слова назви ТУ (іменника) до однини й пошук його в словнику української мови. Якщо це слово знайдено, то аналізується друге слово назви, якщо воно є, що зазвичай є прикметником (може бути в однині чи множині) і доповнює перше слово. Правильність його напису також перевіряється за словником. Пара <іменник–прикметник> є базовою для перевірки назви документа за класифікаторами. Наступні доповнювальні слова назви перевіряються на правильність правопису тільки за словником без залучення класифікатора. Іменник у назві ТУ використовується і в наступних перевірках кодів класифікаторів.

Код продукту, який відповідає назві ТУ, має бути у верхньому лівому куті титульної сторінки та наявності в класифікаторі ДКПП. Назва продукту за цим кодом має відповідати назві продукту (перше слово – іменник у назві ДСТУ/ТУ) і збігатися з назвою в класифікаторі ДКПП. Наприклад, код ДКПП 10.20 повинен бути в ТУ рибних морепродуктів.

Код продукту за класифікатором УКНД має бути у верхньому правому куті титульної сторінки. Цей код гармонізований ICS:2005 MOD, де MOD – один зі способів гармонізації, що описані у відповідних національних та міждержавних стандартах (УкрНДІССІ Держстандарту України, 2001; Україна переходить із міждержавних стандартів (ГОСТ) на міжнародні та європейські, 2024). Код за класифікатором УКТЗЕД зазначають додатково для перевірки на митниці (коди деяких видів продуктів з цього довідника надаються й у довіднику ДКПП). Для деяких продуктів код УКТЗЕД обов'язково вставляється в текст ТУ. Перевірка по УКНД та УКТЗЕД виконується аналогічно перевірці за класифікатором ДКПП.

Ідентифікатор ТУ розташовується по центру титульної сторінки під назвою ТУ і ключовими словами «Технічні умови»; містить таку інформацію: код продукту за ДКПП, ідентифікатор власника ТУ, порядковий номер (кількість) різних ТУ на підприємстві, рік створення.

Для титульної сторінки здійснюється перевірка таких параметрів: наявність ключових літер ТУ У (за регулярним виразом); код продукту (має збігатися з кодом ДКПП); ідентифікатор власника (за довідником); порядковий номер (кількість) різних ТУ на певному підприємстві; рік – за регулярним виразом (має бути в певних межах). Наприклад, код ТУ У 10.2-00447853-014-2022 вміщує таку інформацію: 10.2 – код продукту по ДКПП, 00447853 – ідентифікатор власника ТУ, 014 – порядковий номер (кількість) різних ТУ на вказаному підприємстві, 2022 – рік.

Перевірка наявності затвердження документа з прізвищем директора, дати затвердження (під кодом УКНД – у правій частині титульної сторінки) реалізована за допомогою перевірки місця розташування цих атрибутів та ключових слів, наприклад: «Затверджую», «Директор». Дата перевіряється за шаблоном з використанням обмеження на рік. Подання імені та прізвища керівника й виконавця: ім'я – повністю, прізвище – великими літерами.

Наявність розробника документа перевіряється за місцем розташування цього атрибута та за ключовим словом «Розроблено». Дата перевіряється за шаблоном з використанням обмежень на рік. Назва підприємства перевіряється за довідником (з використанням інтернету). Наявність ключового слова «Вперше» (якщо документ надається вперше), дати надання чинності, обмеження терміну дії перевіряються за ключовим словом і місцем розташування. Дати перевіряються аналогічно до попередніх атрибутів титульної сторінки.

ТУ слід перевіряти регулярно, але не рідше одного разу на п'ять років після введення в дію чи останньої перевірки, якщо не було потреби перевірити їх раніше. Перевіряється це обмеження і в процесі автоматизованої перевірки документа й у разі закінчення терміна дії видається повідомлення про необхідність нової перевірки, що надсилається поштою власникові ТУ з пропозицією виконати чергову перевірку. Після перевірки, яку виконала програма, ті поля на титульній

сторінці, що були не на своєму місці, неправильно оформлені чи відсутні, відмічаються в тексті червоним кольором.

Під час обробки цього розділу НД вирішують завдання створення / перевірки існування сутностей онтології: класів – на основі іменників (наприклад, сир, консерви); властивостей даних (наприклад, код УКНД, розробник); суперкласів до класів у процесі використання словників і тезаурусів (наприклад, харчові продукти до класу «Сир»). Під час перевірки титульної сторінки виконуються такі дії:

- пошук словосполучення «технічні умови» (за ключовими словами та за розташуванням – приблизно по центру сторінки та по центру рядка);
- пошук назви ТУ (над словосполученням «технічні умови»);
- обробка текстового рядка назви:
  - видалення СТОП-слів (прийменників типу з, та, і, для, а також знаків пунктуації);
  - перевірка всіх слів у словнику української мови;
  - приведення до однини називного відмінка (наприклад, назва «Пресерви з риби та морепродуктів в олії» буде перетворена в «Пресерва риба морепродукти олія»; якщо друге слово – прикметник, то замість назви «Сири м'які» отримуємо «Сир м'який»);
  - перевірка наявності першого слова в довіднику: воно є основним словом у повній назві і має бути іменником (перевірка частини мови здійснюється за словником); значення наступних слів зменшується прямо пропорційно до відстані від першого слова;
  - якщо перше слово має загальний характер, наприклад продукт, виріб, консерви, то може траплятися в класифікаторі кілька разів або зараховуватися до іншої сутності; тоді слід до пошуку додати друге (а іноді й третє) слово з назви ТУ; одним з варіантів подальшого пошуку в разі загального першого слова є перетворення прикметника в іменник і пошук по двох іменниках;
  - аналіз другого слова; якщо воно є прикметником (перевірка частини мови за словником), то для подальшого пошуку треба використовувати пару <іменник – прикметник>; створення у зворотному порядку з цих слів пари, яка знадобиться й для подальшої перевірки;
  - перевірка першого слова назви (або декількох слів) відповідно до кодів у класифікаторах ДКПП і УКНД реалізується знаходженням цього слова в класифікаторі та порівнянням відповідного коду з кодом на титульній сторінці. Якщо перше слово – іменник назви ТУ є у ДКПП, то за ним і за уточнювальними словами визначають код. Наприклад назва «Вироби сиркові» має бути присутня в ДКПП – 10.51 та УКНД – 67.100;
  - якщо першого слова – іменника назви ТУ немає у ДКПП, то пошук слова проводять за Тлумачним словником української мови, тезаурусом, вікіпедією тощо; за отриманим визначенням слова з перших слів визначення (наприклад, для слова «пресерви», якого немає в класифікаторах, це будуть слова «нестерилізовані харчові продукти») вибираємо основне слово – іменник (продукт) і додаткове слово (харчовий) (після приведен-

ня їх до нормальної форми); ця комбінація слів «продукт харчовий» чи її інверсна форма «харчовий продукт» використовується для подальшої перевірки;

- аналіз назви ТУ відбувається за основними словами його визначення, що отримані раніше (у прикладі – це харчовий продукт); потрібно знайти словосполучення (наприклад, продукт харчовий – підрозділ 10 «Продукти харчові» у ДКПП); у цьому розділі класифікатора ДКПП слово «рибна» (прикметник від «риба» – друге слово у назві ТУ) знаходиться з кодом 10.20 – «Продукція рибна», тобто слово «пресерви» може відповідати визначенню – «харчовий продукт, в якому є риба» і мати код 10.10;
- альтернативний спосіб перевірки – це попередній пошук коду в класифікаторі і, якщо він є, то його назва аналізується на відповідність з назвою ТУ (наприклад, код 67.120.30 з ТУ знаходимо за допомогою повнотекстового пошуку з назвою «Рибні морепродукти»);
- обробка всіх слів знайденого словосполучення в назві (попередньо перетворивши їх у нормальні форми) і перевірка на збіг з основними словами в назві ТУ (у наведеному прикладі до слів «риба морепродукти» є збіг з двома словами в назві ТУ, код 67 – Технологія виробництва харчових продуктів вміщує слова «харчовий продукт»; це можна вважати достатнім для твердження, що код на титульній сторінці ТУ відповідає назві);
- можливість додаткової перевірки слова, якого немає в класифікаторах, за тезаурусом (наприклад, слово «пресерви» має бути в підрозділі «Рибні продукти» розділу «Харчові продукти»; це означає належність цього слова до харчових продуктів з риби);
- перевірка назви ТУ (наприклад, «Вироби сиркові») має відбуватися із залученням другого важливого слова – прикметника, перетвореного в іменник (сир), і за словосполученням двох іменників (Вироби сир) можна знайти їх відповідно до ДКПП 10.51 та УКНД 67.100.

На рис. 4 показано результат роботи програми з перевірки титульного аркуша ТУ. Алгоритм перевірки основної частини документа (загальної структури) передбачає перевірку таких параметрів, як:

– наявність змісту ТУ й перевірка його відповідно до списку розділів, який відповідає сутностям онтологічної моделі НД, тому дає змогу виконувати додаткові перевірки об'єктних властивостей і властивостей даних моделі; у разі відсутності розділу з певною назвою і номером під час подальшої перевірки номер наступного розділу зменшувати на одиницю;

– відповідність назви розділу в змісті до його назви в повному тексті документа; у разі наявності розбіжностей у назвах або відсутності розділу – надавати відповідне повідомлення; назва розділу може частково відрізнятись, але основні ключові слова мають бути (наприклад, назва розділу «ВИМОГИ ЩОДО БЕЗПЕКИ» може бути змінена на «ВИМОГИ БЕЗПЕКИ»);

– порядок розташування розділів, який має відповідати шаблону;

– наявність вмісту в кожному розділі (за наявності підрозділів або наявності довільного тексту довжиною не менше зазначеної);

- послідовність нумерації розділів та підрозділів;
- якщо назва спорідненого розділу та його номер відрізняються від зразка, то поруч зі зразком слід вивести споріднену назву, яку запропонував автор;
- таблиці: наявність підписів і відповідність за назвою (ключовими словами) зі зразком (наявність та послідовність нумерації не перевіряється, бо вона може не відповідати зразку).

ДКПП 10.89.19: ОК
УКНД 67.100.99: ОК
ЗАТВЕРДЖУЮ: ОК
Директор: ОК
Ім'я ПРИЗВИЩЕ: НЕ ЗНАЙДЕНО
РОЗРОБЛЕНО: НЕ ЗНАЙДЕНО
Розробник: НЕ ЗНАЙДЕНО
ТУ У: ОК
Дата надання чинності: ОК
Чинні до: ОК
Рік перевірки: 2025

Рис. 4. Протокол перевірки титульного аркуша ТУ.

*Джерело:* авторська розробка

Приклад роботи програми з перевірки основної частини умовного документа ТУ «Продукти молоковмісні» наведено на рис. 5. Як зразок базового документа, з яким порівнювалися ТУ, використано структуру, що відображається в розробленій онтологічній моделі НД.

*Перевірка розділу «Нормативні посилання»* передбачає перевірку на:

- відповідність номерів і назв ДСТУ й інших НД, наданих у файлах ТУ з тими номерами та назвами, що є на офіційному сайті ДП «Укрметрестандарт» та інших споріднених сайтах;
- повноту (перевіряються два документи спорідненого змісту): перелік НД ТУ має охоплювати принаймні всі посилання з відповідного базового ДСТУ, на основі якого був створений документ ТУ;
- актуальність зазначеного НД (програма перевірки актуальності документа, результат роботи якої наведено в табл. 1, у колонці «Стан» виводить слово «Діє», якщо документ є актуальним).

У табл. 1 порівнювалися два НД: Док1 – ДСТУ «Сири тверді» і умовний документ Док2 – ТУ «Продукти молоковмісні». В онтологію НД записуються (або перевіряються на існування): клас ДСТУ і його властивості даних, зокрема номер, назва, розробник, ...; клас ТУ і його властивості даних, наприклад: номер, назва, розробник, посилання на НД, ...; об'єктні властивості: посилання ТУ на ДСТУ, ДСТУ на ТУ тощо.

**Перевірка наявності розділів та вмісту в них:**

- 1 СФЕРА ЗАСТОСУВАННЯ – відповідає змісту, має вміст  
2 НОРМАТИВНІ ПОСИЛАННЯ – відповідає змісту, порожній розділ  
Терміни та визначення понять – відсутній розділ  
Класифікація – відсутній розділ  
3 ТЕХНІЧНІ ВИМОГИ – відповідає змісту, має ВМІСТ  
4 ВИМОГИ БЕЗПЕКИ – відповідає змісту, має вміст (споріднена назва до «Вимоги щодо безпеки»)  
5 ВИМОГИ ОХОРОНИ ДОВКІЛЛЯ, УТИЛІЗАЦІЯ – відповідає змісту, (споріднена назва до «Вимоги до охорони довкілля»  
Маркування – відсутній розділ  
Пакування – відсутній розділ (знайдено у розділі 3.4)  
6 ПРАВИЛА ПРИЙМАННЯ – відповідає змісту, має вміст (порушено порядок слідування) (порушено нумерацію)  
7 МЕТОДИ КОНТРОЛЮВАННЯ – відповідає змісту, має вміст (порушено порядок слідування)  
8 ТРАНСПОРТУВАННЯ ТА ЗБЕРІГАННЯ – відповідає змісту, має вміст (споріднена назва до Правила транспортування та зберігання) (порушено порядок слідування)  
9 ГАРАНТІЇ ВИРОБНИКА – порожній розділ  
ДОДАТОК А – відповідає змісту, має вміст  
ДОДАТОК Б – відповідає змісту, має вміст  
ДОДАТОК В – відсутній розділ

**Перевірка наявності таблиць, пустих таблиць (або тих, що мають менше 2 рядків), таблиць без підписів:**

- Таблиця 1 – Органолептичні показники – таблиця ВІДСУТНЯ  
Таблиця 2 – Органолептичні показники – таблиця ВІДСУТНЯ  
Таблиця 3 – Фізико- хімічні показники (рядків: 5)  
Таблиця – Мікробіологічні показники – таблиця ВІДСУТНЯ  
Таблиця – Маса нетто – таблиця ВІДСУТНЯ  
Таблиця – Строк придатності головок сиру – таблиця ВІДСУТНЯ  
Таблиця – Строк придатності в спожитковому пакуванні – таблиця ВІДСУТНЯ  
Таблиця 6 – Показники безпеки (рядків: 6)  
Таблиця 8 (рядків: 10) – не має вмісту  
Таблиця (рядків: 31) – не має вмісту, без підпису

**Перевірка нумерації розділів і підрозділів:**

- Нумерація підрозділів для рівня [3]: 3.1, 3.2, 3.3, 3.4  
Нумерація підрозділів для рівня [4]: 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 4.10, 4.11, 4.12, 4.13, 4.14, 4.15  
Нумерація підрозділів для рівня [5]: 5.1, 5.2, 5.3, 5.4  
Нумерація підрозділів для рівня [6]: 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9  
Нумерація підрозділів для рівня [7]: 7.1, 7.2, 7.3, 7.4, 7.5, 7.6, 7.7  
Нумерація підрозділів для рівня [8]: 8.1, 8.2, 8.3, 8.4

*Рис. 5. Протокол перевірки основної частини документа ТУ.*

*Джерело: авторська розробка*

Таблиця 1

## Результати перевірки актуальності ДСТУ

Номер ДСТУ	Назва ДСТУ	Док1	Док2	Стан
016-97	Державний класифікатор продукції та послуг	+		
4834:2007	Молоко та молочні продукти. Правила приймання, методи відбирання та готування проб до контролювання	+	+	Діє
4457:2005	Ферментні препарати. Загальні технічні умови	+	-	Діє
...	...	...	...	...

Перевірка інших розділів НД виконується аналогічно до попередніх розділів.

*Алгоритми порівняння розділів документів на схожість.* Більшість розроблених ТУ спираються на базовий ДСТУ, тому важливим способом перевірки є порівняння певних розділів документів на відповідність. Також можна порівнювати два споріднені ТУ між собою або дві версії одного ТУ (початкову і поновлену). Перевірка за однаковими номерами розділів у двох документах виявляється неефективною, бо не всі розділи та підрозділи ДСТУ обов'язково мають бути присутніми в ТУ. Порядок слідування розділів може відрізнитися. Тому важливим є порівняння вмісту назв заголовків підрозділів (вони можуть дещо відрізнитися) та вмісту самого підрозділу. Для цього використовуються відомі та розроблені авторські алгоритми порівняння, адаптовані для певної ПрО. Під час порівняння окремих слів і речень у назвах розділів та в їх вмісті використовуються такі наявні алгоритми:

1. *Відстань Левенштейна (2025).* Порівнюються два схожих слова за допомогою відстані Левенштейна, коли визначається кількість операцій, необхідних для перетворення одного слова в інше. Алгоритм Вагнера-Фішера, розраховує відстань Левенштейна за допомогою спеціально побудованої матриці. Такий алгоритм доцільно застосовувати під час перевірки невеликих фрагментів тексту (наприклад, заголовків розділів документа) або окремих слів, які можуть уміщувати орфографічні помилки або мати специфічний напис у певній ПрО.

2. *Наближеність текстів по Жаккару.* Якщо фрагменти тексту не дуже великі, то можна визначити індекс близькості між ними – коефіцієнт Жаккара (Клебанова та ін., 2024). Щоб обчислити його для речень А і В документів, слід знайти кількість однакових токенів у двох реченнях і розділити їх на загальну кількість токенів у двох реченнях, застосувавши формулу:  $Jaccard = \frac{\text{intersection}(A,B)}{\text{union}(A,B)}$ .

3. *Алгоритм шинглів.* Якісно покращити пошук дублікатів у тексті можна за допомогою шинглів (Перевірити текст по шинглам, б.д) (відомих також як n-грами (Rose-Collins, 2025)). Ідея полягає в тому, щоб представити текст як послідовності з n елементів, що йдуть поспіль у тексті (посимвольно або послівно). Якщо порівнювати з попереднім підходом, то метод з n-грамами більш чутливий до перестановки слів у тексті та друкарських помилок. На рис. 6 показано використан-

ня шинглів (розбиття тексту на n-грами по три символи) та обчислений коефіцієнт Жаккара для порівняння аналогічних фрагментів з документів ТУ (Документ 1) та ДСТУ (Документ 2). Скорочений текст таких документів наведено нижче.

Документ 1. Контроль мікробіологічних показників (бактерії групи кишкових паличок), наявність патогенних мікроорганізмів досліджують у порядку державного санітарного нагляду санітарно-епідеміологічною службою з періодичністю, затвердженою у встановленому порядку.

Документ 2. Наявність патогенних мікроорганізмів досліджують у порядку державного санітарного нагляду санітарно-епідеміологічними станціями з періодичністю, затвердженою у встановленому порядку.

Обчислені значення коефіцієнта Жаккара (рис. 6, 7) показують більшу схожість цих документів під час використання символічних шинглів. Це характерно для невеликих фрагментів тексту (повідомлень), де добре працюють посимвольні n-грами зі значенням n від трьох до п'яти. З великими текстами (пошук дублікатів між повними нормативними документами, статтями, вебсторінками) можна використовувати послівні n-грами з n від семи до десяти.

```
Контроль мікробіологічних показників (бактерії групи кишкових паличок), наявність патогенних мікроорганізмів досліджують у порядку державного санітарного
Шингли: Кон, онт, нтр, тро, рол, оль, ль, ь м, ні, мік, ікр, кро, роб, обі, біо, іол, оло, лог, огі, гіч, ічн, чни, них, их, х п, по, пок, ока, каз
Наявність патогенних мікроорганізмів досліджують у порядку державного санітарного нагляду санітарно-епідеміологічними станціями з періодичністю, затвер
Шингли: Ная, аяв, яяв, яні, ніс, іст, сть, ть, ь п, па, пат, ато, тог, оге, ген, енн, нни, них, их, х м, ні, мік, ікр, кро, роо, оор, оор, рга, ген

Речення для порівняння:
Контроль мікробіологічних показників (бактерії групи кишкових паличок), наявність патогенних
Наявність патогенних мікроорганізмів досліджують у порядку державного санітарного нагляду са
Введіть n (розмір шинглу): 3
Кількість шинглів у першому реченні: 203
Кількість шинглів у другому реченні: 150
Кількість спільних шинглів: 135
Кількість усіх шинглів: 218
Коефіцієнт Жаккара (jac): 0.6192660550458715
```

Рис. 6. Розбиття тексту на символічні n-грами (n=3) для обчислення коефіцієнта Жаккара.

*Джерело:* авторська розробка

На рис. 7 показано шингли, що збігаються в процесі розбиття тексту на n-грами по три слова для порівняння тих самих фрагментів у документах, і обчислений коефіцієнт Жаккара.

Спільні шингли: [затвердженою у встановленому, з періодичністю затвердженою, у встановленому порядку, у порядку державного, патогенних мікроорганізмів досліджують, періодичністю затвердженою у, державного санітарного нагляду, наявність патогенних мікроорганізмів, досліджують у порядку, мікроорганізмів досліджують у, порядку державного санітарного]

Коефіцієнт Жаккара (n=3): 0,4231

Рис. 7. Обчислення коефіцієнта Жаккара з використанням шинглів-слів.

*Джерело:* авторська розробка

*Використання пріоритетів та вагових коефіцієнтів слів.* Під час порівняння речень в ТУ є слова, які мають особливу вагу для розуміння змісту речення. У запропонованому алгоритмі підраховується кількість однакових слів у двох реченнях з різних документів, але окремим словам надається подвійна вага. Такими словами є посилання на НД (ДСТУ) з однаковими номерами, не зважаючи на їх розташування, тому що вони відображають змістовий складник речення. Ураховуючи специфіку текстових виразів в ТУ, слова на початку речення є більш значущими, ніж інші. Наприклад, нехай перевіряються на схожість два однакових за змістом речення з різних документів:

Методи відбирання проб і готування їх до випробовування – згідно з ДСТУ.

Відбирання та готування проб здійснюється згідно з ДСТУ.

Слова з пріоритетами мають такий вигляд: ДСТУ (7,6) – у тексті номер ДСТУ не зазначено, тому слову спеціальний пріоритет не надається, метод (1,0), відбирання (2,1), проба (3,3), готування (4,2), випробування (5,0), здійснення (0,4), згідно (6,5). Після перемноження пріоритетів виведено слова, які повторюються (з не нульовим пріоритетом після перемноження) відповідно до їх пріоритету, починаючи з найменшого: відбирання (2), готування (8), проба (9), згідно (30), ДСТУ (42). Результат роботи програми показано на рис. 8.

К-сть співпадаючих слів (без стоп-слів): 5  
Загальна к-сть слів у двох реченнях (без стоп-слів): 13  
К-сть різних слів (union) у двох реченнях: 8  
% збігів до загальної к-сті слів (без стоп-слів): 38,46 %  
% збігів до к-сті різних слів (union): 62,50 %

*Рис. 8.* Статистика після обробки тексту з використанням пріоритету слів.

*Джерело:* авторська розробка

Надаючи додаткову вагу ДСТУ за умови збігу його номерів і враховуючи наближеність до початку речення трьох слів, які збігаються, можна вважати, що збіг дорівнює приблизно 80 %, тобто речення є еквівалентними за змістом.

*Векторизація тексту.* Для визначення подібності текстів використовується векторизація тексту з показниками Term Frequency (TF) та Inverse Document Frequency (IDF) (Данильченко та ін., 2025). Компонент TF обчислює кількість (частоту) терміна в документі й ділить її на загальну кількість термінів цього документа. Значення  $IDF = \log(N/N1)$  ( $N1$  – кількість зустрічальності певного слова в  $N$  документах) стосується всіх документів; вимірює частоту слова у всіх документах і тому обчислюється лише один раз. Значення TF-IDF отримується множенням компонентів TF та IDF. Матриця, яку в результаті отримують, є основним показником подібності документів. Потім створюється матриця косинусної подібності (Клебанова та ін., 2024), яка сприятиме порівнянню між собою наданих речень щодо схожості слів, які в них входять.

Функція Окарі BM25 ранжує набір документів на основі термів запиту, що трапляються в кожному документі, незалежно від їхньої близькості один до одного.

Це покращений, більш збалансований варіант оцінки TF-IDF. Вираз, який доповнює оцінку TF-IDF, використовується для нормалізації довжини документа. Швидкість роботи функції BM25 може бути суттєво покращена завдяки використанню індексації. Створюється так званий інвертований індекс, де в таблиці зберігаються для кожного терміна значення TF та довжина документа, у якому він є.

**Висновки.** У роботі досліджено та визначено основні проблеми розробки та перевірки українськомовних текстів нормативних документів, крім того:

- проведено аналіз і систематизацію сучасних підходів, що підтримують окремі етапи обробки текстів нормативних документів;
- запропоновано підхід до обробки українськомовних текстів нормативних документів, який поєднує онтологічне моделювання відповідних предметних областей та методи лінгвістичного аналізу текстів, які надані природною мовою;
- запропоновано алгоритми та методи обробки нормативних документів ДСТУ та ТУ, що ґрунтуються на адаптації наявних алгоритмів і методів до Про, що розглядається;
- описано інформаційну систему щодо обробки (розробки та перевірки) нормативних документів ДСТУ й ТУ, яка ґрунтується на використанні онтологічних моделей відповідних предметних областей і методів лінгвістичного аналізу українськомовних текстів.

## СПИСОК ПОСИЛАНЬ

- Бізнес-портал «Леонорм», б.д. *Леонорм*. [online] Доступно: <<http://www.leonorm.lviv.ua/Default.php?rescode=0510&Page=stcatalog>> [Дата звернення 20 лютого 2026].
- Верховна Рада України, 2014. *Перелік форматів даних електронних документів постійного і тривалого (понад 10 років) зберігання*. Наказ Міністерства юстиції України [online] 11 листопада, № 1886/5. Доступно: <<https://zakon.rada.gov.ua/laws/show/z1422-14#Text>> [Дата звернення 21 лютого 2026].
- Відстань Левенштейна, 2025. *Вікіпедія*, [online] 28 серпня. Доступно: <[https://uk.wikipedia.org/wiki/Відстань\\_Левенштейна](https://uk.wikipedia.org/wiki/Відстань_Левенштейна)> [Дата звернення 20 лютого 2026].
- Всеукраїнський державний науково-виробничий центр стандартизації, метрології, сертифікації і захисту прав споживачів, 2016. *Технічні умови України. Настанови щодо типової побудови, викладення, оформлення, позначення, прийняття та надання чинності. СОУ КЗПС 74.9-02568182-003:2016*. Наказ [online] 13 січня, № 003. Київ: Укрметрестстандарт. Доступно: <<https://zakon.isu.net.ua/sites/default/files/normdocs/sou.pdf>> [Дата звернення 20 лютого 2026].
- Галка, Л., 2025. Парсинг сайтів: що це, як працює та навіщо парсити дані. *Promodo*, [online] 18 березня. Доступно: <<https://www.promodo.ua/blog/parsing-saytiv-shcho-ce-yak-pracuyue-ta-navishcho-parsiti-dani>> [Дата звернення 15 лютого 2026].
- Грамотичний словник української мови, б.д. *Mova.info*. [online] Доступно: <<http://www.mova.info/grmasl.aspx>> [Дата звернення 10 лютого 2026].
- Данильченко, В.М., Отрох, С.І., Шалигін, М.О. та Донець, А.Г., 2025. Багатокритеріальне розпізнавання відповідності текстів темі на основі алгоритму TF-IDF. *Зв'язок*, [e-journal] 3, с.36-40. <https://doi.org/10.31673/2412-9070.2025.027728>

- Державний класифікатор продукції та послуг ДК 016-2010, б.д. *Dkpp.rv.ua*. [online] Доступно: <<https://dkpp.rv.ua/>> [Дата звернення 11 лютого 2026].
- ДСТУ 4163:2020 «Уніфікована система організаційно-розпорядчої документації. Вимоги до оформлення документів», 2021. *Document*, [online] 1 вересня. Доступно: <<https://document.vobu.ua/doc/6310>> [Дата звернення 11 лютого 2026].
- Інформаційне забезпечення у сфері технічного регулювання ДП «УКРМЕТРТЕСТСТАНДАРТ», б.д. *Csm.kiev.ua*. [online] Доступно: <<https://csm.kiev.ua/>> [Дата звернення 13 лютого 2026].
- Клебанова, Т.С., Гур'янова, Л.С., Чаговець, Л.О., Панасенко, О.В., Сергієнко, О.А. та Яценко, Р.М., 2024. *Бізнес-аналітика багатомірних процесів*. [e-book] Харків: Харківський національний економічний університет імені Семена Кузнеця. Доступно: <<http://ebooks.git-elt.hneu.edu.ua/babar/index.html>> [Дата звернення 10 лютого 2026].
- Марковець, О.В. та Синько, А.І., 2021. Формування якісної технічної документації до програмного забезпечення. *Вісник Вінницького політехнічного інституту*, [e-journal] 2, с.98-106. <https://doi.org/10.31649/1997-9266-2021-155-2-98-106>
- Науково-дослідний інститут метрології вимірювальних і управляючих систем, 2020. *Український класифікатор нормативних документів НК 004:2020*. Наказ Міністерства України [online] 18 травня, № 905. Київ: Український науково-дослідний і навчальний центр проблем стандартизації, сертифікації та якості. Доступно: <[https://zakon.isu.net.ua/sites/default/files/normdocs/nk\\_004\\_2020\\_0.pdf](https://zakon.isu.net.ua/sites/default/files/normdocs/nk_004_2020_0.pdf)> [Дата звернення 11 лютого 2026].
- Пасічник, В. та Яромич, М., 2025. Автоматизоване формування технічної документації в галузі іт із використанням великих мовних моделей. *Studia methodologica*, [e-journal] 59, с.250-272. <https://doi.org/10.32782/2307-1222.2025-59-22>
- Перевірити текст по шинглам, б.д. *Ve.gayn.cx.ua*. [online] Доступно: <<https://key.mirab.v.ua/articles/pereviriti-tekst-po-shinglam.html>> [Дата звернення 10 лютого 2026].
- Пошуковий інструмент кодів УКТЗЕД за обраними товарами, б.д. *Дія*. [online] Доступно: <<https://business.diiia.gov.ua/service/uktzed>> [Дата звернення 11 лютого 2026].
- Про сайт, б.д. *Горох*. [online] Доступно: <<https://goroh.pp.ua/>> [Дата звернення 10 лютого 2026].
- Словник.ua*, б.д. [online] Доступно: <<https://slovnuk.ua/index.php>> [Дата звернення 20 лютого 2026].
- Україна переходить із міждержавних стандартів (ГОСТ) на міжнародні та європейські, 2024. *Uteka.ua*. [online] Доступно: <<https://uteka.ua/ua/publication/news-14-delovye-novosti-36-ukraina-perexodit-s-mezhgosudarstvennyh-standartov-gost-na-mezhdunarodnye-i-evropejskie>> [Дата звернення 12 лютого 2026].
- УкрНДІССІ Держстандарту України, 2001. ДСТУ 1.1:2001 *Національна стандартизація. Стандартизація та суміжні види діяльності. Терміни та визначення основних понять. З Поправкою (ІПС № 4-2003)*. Наказ [online] 29 травня, № 247. Доступно: <<https://studfile.net/preview/11639149/page:4/>> [Дата звернення 12 лютого 2026].
- Ярка, У.Б., Мина, Ж.В., Пелещишин, О.П. та Білушак, Т.М., 2016. Оптимальні методи створення документації згідно стандартів управління якістю на основі ISO 9000. *Управління розвитком складних систем*, [online] 27, с.162-169. Доступно: <[https://www.researchgate.net/publication/311307326\\_Optimalni\\_metodi\\_stvorennia\\_dokumentacii\\_zgidno\\_standartiv\\_upravlinna\\_akistu\\_na\\_osnovi\\_ISO\\_9000](https://www.researchgate.net/publication/311307326_Optimalni_metodi_stvorennia_dokumentacii_zgidno_standartiv_upravlinna_akistu_na_osnovi_ISO_9000)> [Дата звернення 15 лютого 2026].
- Isay, I., 2021. Як створювати та оформлювати технічну документацію в ІТ: рекомендації для початківців і підказки для досвідчених. *DOU.ua*, [online] 13 травня. Доступно: <<https://dou.ua/lenta/columns/creating-technical-documentation/>> [Дата звернення 15 лютого 2026].

Lang-uk projects, n.d. *Lang-uk*. [online] Available at: <<https://lang.org.ua>> [Accessed 11 February 2026].

*Named Entity Recognition (NER)*, n.d. [online] Available at: <<https://www.ultralytics.com/glossary/named-entity-recognition-ner>> [Accessed 13 February 2026].

Rose-Collins, F., 2025. N-грами в НЛП: як вони працюють та їхня роль в аналізі тексту. *Ranktracker*, [online] 18 березня. Доступно: <<https://www.ranktracker.com/uk/blog/n-grams-in-nlp-how-they-work-their-role-in-text-analysis/>> [Дата звернення 10 лютого 2026].

---

## REFERENCES

---

Biznes-portal "Leonorm" [Business portal "Leonorm"], n.d. *Leonorm*. [online] Available at: <<http://www.leonorm.lviv.ua/Default.php?rescode=0510&Page=stcatalog>> [Accessed 20 February 2026].

Danylchenko, V.M., Otrokh, S.I., Shalyhin, M.O. and Donets, A.H., 2025. Bahatokryterialne rozpoznavannya vidpovidnosti tekstiv temi na osnovi alhorytmu TF-IDF [Multi-criteria recognition of text-to-topic correspondence based on the TF-IDF algorithm]. *Zviatok*, [e-journal] 3, pp.36-40. <https://doi.org/10.31673/2412-9070.2025.027728>

Derzhavnyi klasyfikator produktii ta posluh DK 016-2010 [State Classifier of Products and Services DK 016-2010], n.d. *Dkpp.rv.ua*. [online] Available at: <<https://dkpp.rv.ua/>> [Accessed 11 February 2026].

DSTU 4163:2020 "Unifikovana systema orhanizatsiino-rozporiadchoi dokumentatsii. Vymohy do oformlennia dokumentiv" [DSTU 4163:2020 "Unified system of organizational and administrative documentation. Requirements for document execution"], 2021. *Document*, [online] September 1. Available at: <<https://document.vobu.ua/doc/6310>> [Accessed 11 February 2026].

Halka, L., 2025. Parsynh saitiv: shcho tse, yak pratsiuie ta navishcho parsyty dani [Parsing sites: what it is, how it works and why to parse data]. *Promodo*, [online] March 18. Available at: <<https://www.promodo.ua/blog/parsing-saytiv-shcho-ce-yak-pracyuie-ta-navishcho-parsiti-dani>> [Accessed 15 February 2026].

Hramatychnyi slovnyk ukrainskoi movy [Grammatical Dictionary of the Ukrainian Language], n.d. *Mova.info*. [online] Available at: <<http://www.mova.info/grmasl.aspx>> [Accessed 10 February 2026].

Informatsiine zabezpechennia u sferi tekhnichnoho rehuliuвання DP "UKRMETRTSTANDART" [Information support in the field of technical regulation of the State Enterprise "UKRMETRTSTANDART"], n.d. *Csm.kiev.ua*. [online] Available at: <<https://csm.kiev.ua/>> [Accessed 13 February 2026].

Klebanova, T.S., Hurianova, L.S., Chahovets, L.O., Panasenko, O.V., Serhiienko, O.A. and Yatsenko, R.M., 2024. *Biznes-analytika bahatovymirnykh protsesiv* [Business Analytics of Multidimensional Processes]. [e-book] Kharkiv: Simon Kuznets Kharkiv National University of Economics. Available at: <<http://ebooks.git-elt.hneu.edu.ua/babap/index.html>> [Accessed 10 February 2026].

Lang-uk projects, n.d. *Lang-uk*. [online] Available at: <<https://lang.org.ua>> [Accessed 11 February 2026].

Markovets, O.V. and Synko, A.I., 2021. Formuvannia yakisnoi tekhnichnoi dokumentatsii do prohramnoho zabezpechennia [Formation of High-Quality Technical Documentation for

- Software]. *Visnyk of Vinnytsia Politechnical Institute*, [e-journal] 2, pp.98-106. <https://doi.org/10.31649/1997-9266-2021-155-2-98-106>
- Named Entity Recognition (NER)*, n.d. [online] Available at: <<https://www.ultralytics.com/glossary/named-entity-recognition-ner>> [Accessed 13 February 2026].
- Naukovo-doslidnyi instytut metrolohii vymiriuvalnykh i upravliaiuchykh system [Research Institute of Metrology of Measuring and Control Systems], 2020. *Ukrainskyi klasyfikator normatyvnykh dokumentiv NK 004:2020* [Ukrainian Classifier of Regulatory Documents NK 004:2020]. Order of the Ministry of Economy of Ukraine [online] May 18, No. 905. Kyiv: Ukrainskyi naukovo-doslidnyi i navchalnyi tsentr problem standartyzatsii, sertyfikatsii ta yakosti. Available at: <[https://zakon.isu.net.ua/sites/default/files/normdocs/nk\\_004\\_2020\\_0.pdf](https://zakon.isu.net.ua/sites/default/files/normdocs/nk_004_2020_0.pdf)> [Accessed 11 February 2026].
- Pasichnyk, V. and Yaromych, M., 2025. Avtomatyzovane formuvannia tekhnichnoi dokumentatsii v haluzi it iz vykorystanniam velykykh movnykh modelei [Automated generation of technical documentation in the it industry using large language models]. *Studia methodologica*, [e-journal] 59, pp.250-272. <https://doi.org/10.32782/2307-1222.2025-59-22>
- Pereviryty tekst po shynhnam [Check text for shingles], n.d. *Be.gayn.cx.ua*. [online] Available at: <<https://key.mirab.v.ua/articles/perevirity-tekst-po-shinglam.html>> [Accessed 10 February 2026].
- Poshukovy instrument kodiv UKTZED za obranymy tovaramy [UKTZED codes search tool for selected goods], n.d. *Diia*. [online] Available at: <<https://business.diia.gov.ua/service/uktzed>> [Accessed 11 February 2026].
- Pro sait [About the site], n.d. *Horokh*. [online] Available at: <<https://goroh.pp.ua/>> [Accessed 10 February 2026].
- Rose-Collins, F., 2025. N-hramy v NLP: yak vony pratsiuut ta yikhnia rol v analizi tekstu [N-grams in NLP: how they work and their role in text analysis]. *Ranktracker*, [online] March 18. Available at: <<https://www.ranktracker.com/uk/blog/n-grams-in-nlp-how-they-work-their-role-in-text-analysis/>> [Accessed 10 February 2026].
- Slovnnyk.ua*. [Dictionary.ua.], n.d. [online] Available at: <<https://slovnnyk.ua/index.php>> [Accessed 20 February 2026].
- Ukraina perekhodyt iz mizhderzhavnykh standartiv (HOST) na mizhnarodni ta yevropeiski [Ukraine is transitioning from interstate standards (GOST) to international and European ones], 2024. *Uteka.ua*. [online] Available at: <<https://uteka.ua/ua/publication/news-14-delovye-novosti-36-ukraina-perexodit-s-mezhgosudarstvennyx-standartov-gost-na-mezhdunarodnye-i-evropejskie>> [Accessed 12 February 2026].
- UkrNDISSI Derzhstandartu Ukrainy [UkrNDISSI Derzhstandartu Ukrainy], 2001. *DSTU 1.1:2001 Natsionalna standartyzatsiia. Standartyzatsiia ta sumizhni vydy diialnosti. Terminy ta vyznachennia osnovnykh poniat. Z Popravkoiu (IPS № 4-2003)* [DSTU 1.1:2001 National standardization. Standardization and related activities. Terms and definitions of basic concepts. With Amendment (IPS No. 4-2003)]. Order [online] May 29, No. 247. Available at: <<https://studfile.net/preview/11639149/page:4/>> [Accessed 12 February 2026].
- Verkhovna Rada of Ukraine, 2014. *Perelik formativ danykh elektronnykh dokumentiv postiinoho i tryvaloho (ponad 10 rokiv) zberihannia* [List of data formats of electronic documents of permanent and long-term (over 10 years) storage]. Order of the Ministry of Justice of Ukraine [online] November 11, No. 1886/5. Available at: <<https://zakon.rada.gov.ua/laws/show/z1422-14#Text>> [Accessed 21 February 2026].
- Vidstan Levenshteina [Levenshteyn distance], 2025. *Wikimedia*, [online] August 28. Available at: <[https://uk.wikipedia.org/wiki/Vidstan\\_Levenshteina](https://uk.wikipedia.org/wiki/Vidstan_Levenshteina)> [Accessed 20 February 2026].

Vseukrainskyi derzhavnyi naukovo-vyrobnychy tsestr standartyzatsii, metrolohii, sertyfikatsii i zakhystu prav spozhyvachiv [All-Ukrainian State Scientific and Production Center for Standardization, Metrology, Certification and Consumer Rights Protection], 2016. *Tekhnichni umovy Ukrainy. Nastanovyshchodo typovoi pobudovy, vykladennia, oformlennia, poznachennia, pryiniattia ta nadannia chynnosti. SOU KZPS 74.9-02568182-003:2016* [Technical conditions of Ukraine. Guidelines for typical construction, presentation, design, designation, acceptance and entry into force. SOU KZPS 74.9-02568182-003:2016]. Order [online] January 13, No. 003. Kyiv: Ukrmetrteststandart. Available at: <<https://zakon.isu.net.ua/sites/default/files/normdocs/sou.pdf>> [Accessed 20 February 2026].

Yarka, U.B., Myna, Zh.V., Peleshchyshyn, O.P. and Bilushchak, T.M., 2016. Optymalni metody stvorennia dokumentatsii zghidno standartiv upravlinnia yakistiu na osnovi ISO 9000 [Optimal practices for creating documentation according to the standards of quality management based on ISO 9000]. *Upravlinnia rozvytkom skladnykh system*, [online] 27, pp.162-169. Available at: <[https://www.researchgate.net/publication/311307326\\_Optimalni\\_metodi\\_stvorennia\\_dokumentacii\\_zghidno\\_standartiv\\_upravlinna\\_akistu\\_na\\_osnovi\\_ISO\\_9000](https://www.researchgate.net/publication/311307326_Optimalni_metodi_stvorennia_dokumentacii_zghidno_standartiv_upravlinna_akistu_na_osnovi_ISO_9000)> [Accessed 15 February 2026].

**UDK 004.421:[006(083.7):81(=161.2)**

***Kostiantyn Tkachenko,***

*PhD in Economics, Associate Professor,  
Associate Professor at the Department  
of Software Engineering,  
National University "Kyiv Aviation Institute",  
Kyiv, Ukraine  
e-mail: [tkachenko.kostyantyn@gmail.com](mailto:tkachenko.kostyantyn@gmail.com)  
<https://orcid.org/0000-0003-0549-3396>*

## **USE OF ONTOLOGICAL MODELLING AND LINGUISTIC ANALYSIS WHEN WORKING WITH REGULATORY DOCUMENTATION TEXTS**

Nowadays, the development and verification of relevant regulatory documents is a laborious process that is often performed without the use of modern information systems and technologies. The research considers the problem, the relevance of which is beyond doubt, of developing an approach to the formation of regulatory documentation texts based on the use of ontologies and methods of linguistic analysis of texts of national standards.

**The purpose of this research** is to analyse and study the problems of modelling the processes of documentation formation for products of various industries based on the combination of ontological modeling and linguistic analysis of Ukrainian texts of relevant standards.

**The research methodology** form methods of systematising approaches to text processing processes and formalising these processes in the form of a corresponding ontological model, comparative analysis of the main algorithms and methods of linguistic analysis of Ukrainian-language texts.

**The scientific novelty** of the conducted research is the analysis of the problems of automated development and verification of Ukrainian-language texts of regulatory documents; development of an approach based on a combination of ontological modeling of the relevant

subject area and linguistic analysis of texts and development of an appropriate information system based on the proposed approach.

**Coinclusions.** This article studies and identifies the main problems of development and verification of Ukrainian-language texts of regulatory documents. In addition, the work analyses and systematises modern approaches that support individual stages of processing texts of regulatory documents; offers an approach to processing Ukrainian-language texts of regulatory documents, which combines ontological modeling of the relevant subject areas and methods of linguistic analysis of texts provided in natural language; offers algorithms and methods for processing regulatory documents of DSTU and TU, which are based on the adaptation of existing algorithms and methods to the subject area under consideration; an information system for processing (development and verification) of regulatory documents of DSTU and TU is described, which grounds on the use of ontological models of the relevant subject fields and methods of linguistic analysis of Ukrainian-language texts.

**Keywords:** Ukrainian-language text; regulatory document; standards and technical conditions for products; ontological modelling; linguistic analysis; information system.

Надійшла 02.03.2026

Прийнята 16.04.2026

Стаття була вперше опублікована онлайн 29.05.2026



This is an open access journal, and all published articles are licensed under a Creative Commons Attribution 4.0.