

А.В. М'ЯКЕНЬКИЙ, М.О. АЛЕКСЄЄВ
 Національний технічний університет «Дніпровська політехніка»

ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ СЕМАНТИЧНОЇ РЕПРЕЗЕНТАЦІЇ ТЕКСТІВ ПРИРОДНОЇ МОВИ У ЗАДАЧІ ІДЕНТИФІКАЦІЇ ПЕРИФРАЗ

У статті розглядається проблема ідентифікації семантичної еквівалентності текстів природної мови, що є ключовою для моделювання когнітивного процесу розуміння та розвитку інтелектуальних систем обробки мови. Розуміння тексту включає не лише лексичний та синтаксичний аналіз, а й глибоку когнітивну інтерпретацію смислових відносин і контекстуальних особливостей. Особливу складність становить варіативність мовних конструкцій, коли одна й та сама інформація передається різними способами, зберігаючи тотожність значень. У цьому контексті задача ідентифікації перифраз – висловлювань зі схожим змістом, але різною формою – є важливою для когнітивних моделей. Людський мозок здатен легко виявляти семантичну еквівалентність, проте створення комп'ютерних моделей із подібними властивостями є складним завданням через неоднозначність, контекстуальну залежність та мультимовність природної мови. Метою статті є обґрунтування методу пошуку семантичної еквівалентності текстів природної мови шляхом аналізу сучасних підходів до побудови семантичних репрезентацій у контексті розв'язання задачі ідентифікації перифраз. Обрана саме задача ідентифікації перифраз, оскільки вона є репрезентативною для вивчення семантичної еквівалентності та когнітивних механізмів розпізнавання смислу. У статті проведено аналіз трьох основних напрямів розв'язання задачі: статистичних методів на основі корпусного аналізу, методів заснованих на знаннях із використанням онтологій та семантичних мереж, а також методів глибокого навчання, що базуються на дистрибутивній семантиці і трансформерних архітектурах. Окремо розглядаються гібридні підходи, які поєднують графові семантичні репрезентації AMR із нейромережевими моделями, що дозволяють структурувати зміст тексту на концептуальному рівні та моделювати логіко-семантичні зв'язки між його складниками. Результати порівняльного аналізу сучасних досліджень підтверджують високу точність і здатність узагальнювати семантичну інформацію, властиву графовим підходам, у порівнянні з іншими методами. Це свідчить про доцільність їх застосування як основи для побудови інтелектуальних систем, здатних моделювати когнітивні процеси людського розуміння природномовного тексту. Практична цінність дослідження полягає у формуванні підґрунтя для подальшого розвитку когнітивних технологій у сферах машинного перекладу, автоматичного узагальнення, пошуку інформації та діалогових систем, що суттєво підвищує якість взаємодії людини і комп'ютера на основі природної мови.

Ключові слова: когнітивне моделювання, перифрази, розуміння, природна мова, AMR, семантичні графи, трансформери, штучний інтелект.

A. V. MIAKENKYI, M. O. ALEKSIEIEV
 Dnipro University of Technology

COMPARATIVE ANALYSIS OF SEMANTIC REPRESENTATION METHODS FOR NATURAL LANGUAGE TEXTS IN THE TASK OF PARAPHRASE IDENTIFICATION

The article addresses the problem of identifying semantic similarity of natural language texts, which is crucial for modeling the cognitive process of comprehension and developing intelligent language processing systems. Text comprehension involves not only lexical and syntactic analysis but also deep cognitive interpretation of semantic relations and contextual features. A particular challenge is the variability of linguistic constructions, where the same information is conveyed in different ways while preserving the identity of meanings. In this context, the task of paraphrase identification – expressions with similar meaning but different forms – is important for cognitive models. The human brain can easily detect semantic similarity; however, creating computer models with similar capabilities is a complex task due to ambiguity, context dependence, and multilingualism of natural language. The aim of the article is to substantiate a method for searching semantic similarity of natural language texts through analysis of modern approaches to building semantic representations in the context of solving the paraphrase identification task. The paraphrase identification task is chosen because it is representative for studying semantic similarity and cognitive mechanisms of meaning recognition. The article analyzes three main approaches to solving the task: statistical methods based on corpus analysis, knowledge-based methods using ontologies and semantic networks, and deep learning methods based on distributional semantics and transformer architectures. Hybrid approaches are considered separately, which combine graph-based semantic representations AMR with neural network models, allowing structuring of text content at the conceptual level and modeling logical-semantic connections between its components. Results of a comparative analysis of recent studies confirm the high accuracy and ability to generalize semantic information inherent in AMR-based approaches compared to other methods. This indicates the appropriateness of applying AMR as a foundation for building intelligent systems capable of modeling the cognitive processes of human understanding of natural language text. The practical value of the research lies in forming a basis for further development of cognitive technologies in machine translation, automatic summarization, information retrieval, and dialogue systems, significantly improving the quality of human-computer interaction based on natural language.

Key words: cognitive modeling, paraphrases, comprehension, natural language, AMR, semantic graphs, transformers, artificial intelligence.

Постановка проблеми

Одним із найважливіших викликів у сучасній когнітивній інформатиці та обробці природної мови є моделювання процесу розуміння тексту з урахуванням його когнітивних характеристик. Розуміння тексту має на увазі не лише аналіз лексичних або синтаксичних структур, а також глибокий когнітивний процес, який передбачає інтерпретацію смислових відносин, логічних залежностей і контекстуальних нюансів. Саме тому створення інформаційних моделей, здатних відтворити цей процес, є надзвичайно складним завданням. Особливо актуальним цей виклик стає у зв'язку з варіативністю мовних конструкцій, коли одна і та сама інформація може бути передана різними способами, що мають відмінності у лексичній та граматичній формі, але зберігають тотожність значень.

У цьому контексті задача ідентифікації перифраз набуває особливого значення. Перифрази – це мовні висловлювання, які передають одне і те саме значення, але відрізняються формою. Успішне розпізнавання таких смислово еквівалентних висловлювань може інтерпретуватися як здатність штучної системи до когнітивного міркування. Вона обумовлює те, що модель не обмежується лише поверхневим аналізом тексту, а здатна виявляти глибинні знання, оперуючи концептуальними і семантичними зв'язками, що властиві людському розумінню.

Для людини розуміння перифраз є природним і майже автоматичним процесом. Людський мозок володіє здатністю до абстрагування, узагальнення і логічного мислення, що дозволяє легко ідентифікувати ідентичність речень, навіть якщо вони виражені по-різному. При моделюванні процесу розуміння семантичної еквівалентності виникає низка викликів для створення ефективних обчислювальних моделей, здатних адекватно відображати когнітивні механізми обробки природної мови. Основні труднощі полягають у наступному:

1. **Висока варіативність природної мови.** Людська мова характеризується значною гнучкістю, що дозволяє передавати одне й те саме значення через безліч лексичних та граматичних форм. Варіанти можуть відрізнятися порядком слів, використанням синонімів, різними граматичними конструкціями тощо. Це означає, що одна й та сама ідея може бути сформульована багатьма способами, що створює серйозні виклики для автоматичних систем.

2. **Неоднозначність і контекстуальна залежність.** Значення окремих слів і конструкцій часто залежить від контексту, у якому вони вживаються. Слова можуть мати кілька значень, а їх інтерпретація визначається контекстом, у тому числі сусідніми словами, ситуацією, культурними чи тематичними факторами. Особливо це стосується мов із багатим словотвором і вільним порядком слів, таких як українська. Розпізнавання правильного значення в конкретному висловлюванні вимагає досить складного когнітивного аналізу.

3. **Мультимовна складність.** Семантичне моделювання ще більше ускладнюється у мультимовному контексті. Мови відрізняються не лише лексикою та граматику, але й концептуальними структурами, способом категоризації реальності, культурними референтами та прагматичними нормами. Відповідно, моделі, що ефективно працюють для однієї мови (наприклад, англійської), не завжди адекватно переносяться на інші мови без суттєвої адаптації. Це особливо актуально для мов із меншою кількістю цифрових ресурсів або тих, що мають складні морфологічні системи, як-от українська. Завдання полягає не лише у правильному відображенні лексичних та синтаксичних особливостей, але й у збереженні семантичної еквівалентності між мовами, що критично важливо для задач багатомовної ідентифікації перифраз.

У світлі цих викликів постає нагальна потреба у пошуку методів, що дозволяють точно визначати семантичну еквівалентність між текстовими фрагментами – незалежно від лексичних, синтаксичних або мовних варіацій. Визначення семантичної еквівалентності вимагає одночасно гнучкого захоплення смислу й здатності моделювати складні логічні та концептуальні взаємозв'язки.

Аналіз останніх досліджень і публікацій

Завдання визначення семантичної еквівалентності текстів є фундаментальним у сфері обробки природної мови, когнітивного моделювання й розвитку штучного інтелекту. Це завдання лежить у площині моделювання когнітивного процесу розуміння, який передбачає не лише обробку лексичних послідовностей, а й глибоке осмислення значення, семантичних відношень, контекстуальних залежностей.

Аналіз наукових джерел останнього десятиліття дозволяє виокремити три основні напрями підходів до задачі визначення семантичної еквівалентності: статистичні методи, засновані на аналізі корпусних даних; методи що засновані на знаннях, що використовують онтології та семантичні бази; методи комбінаційної дистрибутивної семантики на основі глибинного навчання.

Статистичні моделі стали першим етапом автоматизації семантичного аналізу тексту. До них належать такі методи, як Latent Semantic Analysis (LSA) [1], а також моделі на основі векторних представлень слів. Ці підходи ґрунтуються на гіпотезі дистрибутивної семантики, яка передбачає, що слова, що з'являються в подібних контекстах, мають подібне значення. Статистичні моделі дозволили значно підвищити ефективність систем тематичного аналізу, класифікації текстів та інформаційного пошуку, що обумовило їх використання для пошуку семантичної еквівалентності текстів у задачі ідентифікації перифраз.

Підходи засновані на знаннях спираються на експліцитне кодування семантичних відношень через використання онтологій, тезаурусів та семантичних мереж. Одним із найбільш впливових проєктів цього напрямку є WordNet, який забезпечив формалізацію лексико-семантичних відношень – синонімії, антонімії, гіпонімії, тощо. Перевагою таких підходів при обчисленні семантичної еквівалентності є їхня інтерпретованість і здатність моделювати відношення між поняттями на рівні концептів, а не лише статистичних залежностей.

Значного прогресу у сфері моделювання семантики було досягнуто завдяки застосуванню методів композиційної дистрибутивної семантики на основі глибинного навчання. Ці методи поєднують ідеї традиційної дистрибутивної семантики, яка базується на припущенні, що значення слова формується його контекстом, із можливостями сучасних нейронних мереж, що дозволяють враховувати складні залежності та контекстуальні варіації в тексті. Завдяки цьому формується більш гнучке та точне семантичне представлення лексичних одиниць, яке адаптується до конкретного контексту використання.

Особливістю таких підходів є використання трансформерних архітектур, зокрема моделей типу BERT та їх численних модифікацій, які навчаються на великих обсягах текстових даних і здатні створювати контекстно-залежні векторні представлення слів, фраз і цілих текстів. Це суттєво підвищує якість розпізнавання семантичних зв'язків і взаємозв'язків між елементами мови, що є критично важливим у задачах ідентифікації семантичної еквівалентності, таких як виявлення перифраз.

Ще одним важливим напрямом є розвиток гібридних моделей, які поєднують традиційні семантичні графові репрезентації з можливостями глибинного навчання, зокрема трансформерних архітектур та нейронних мереж. Серед них особливо можна виділити використання репрезентацій абстрактного значення AMR [2]. Перевага AMR полягає у здатності репрезентувати зміст висловлювання у формі графа, де вузли відповідають концептам, а ребра – семантичним відношенням між ними. Це дозволяє чітко фіксувати зв'язки між об'єктом, суб'єктом та дією, а також обставини дії та інші семантичні ролі. Сучасні дослідження демонструють ефективність використання AMR у задачах семантичного парсингу, генерації тексту та підсумовування [3]. Разом з тим застосування AMR для задачі визначення семантичної еквівалентності залишається недостатньо дослідженим.

У зв'язку з цим доцільним є проведення порівняльного аналізу традиційних підходів і методів на основі штучного інтелекту для визначення їх можливостей та обмежень у контексті розв'язання задачі ідентифікації семантичної еквівалентності. Це дозволить обґрунтувати

вибір оптимальної архітектури, здатної забезпечити баланс між точністю, інтерпретованістю та масштабованістю.

Мета дослідження

Метою статті є обґрунтування методу пошуку семантичної еквівалентності текстів природної мови шляхом аналізу сучасних підходів до побудови семантичних репрезентацій природної мови у контексті розв'язання задачі ідентифікації перифраз. Дослідження спрямоване на виявлення переваг і обмежень традиційних корпусно-статистичних методів, моделей на основі знань, методів комбінаційної дистрибутивної семантики, а також сучасних гібридних моделей, що поєднують нейромережеві технології з формальними семантичними структурами.

Виклад основного матеріалу дослідження

Процес розуміння природної мови є ключовим когнітивним механізмом, що забезпечує здатність інтерпретувати зміст висловлювань, встановлювати зв'язки між ними, робити висновки та будувати знання про навколишній світ. У контексті когнітивної науки розуміння полягає у співвіднесенні мовних форм із концептуальними структурами, які відображають значення [4]. У цьому сенсі процес розуміння включає здатність співвідносити різні мовні вирази, що передають однаковий або подібний зміст, попри їхню формальну різницю. Ця властивість безпосередньо є об'єктом моделювання в задачі ідентифікації перифраз. Пошук перифраз, або встановлення семантичної еквівалентності між текстами, є важливим як для когнітивного моделювання, так і для широкого спектра прикладних задач обробки природної мови – від машинного перекладу до інформаційного пошуку та побудови діалогових систем.

Основним поняттям у цій задачі є поняття семантичної еквівалентності. Під ним зазвичай розуміють відношення між двома мовними виразами, які передають однаковий або дуже близький зміст у межах певного комунікативного контексту. Водночас це відношення не є абсолютно строгим і може варіювати залежно від умов використання, прагматичних факторів та цільової задачі. У деяких підходах семантична еквівалентність визначається як симетричне логічне включення смислів, в інших – як взаємна заміненість у контекстах, а ще в інших – як подібність розподілу лексичних одиниць у схожих ситуаціях.

Відповідно до цього, пошук семантичної еквівалентності може ґрунтуватися на різних методологічних підходах. Аналіз сучасних досліджень показав, що основні підходи до побудови семантичних представлень тексту можна умовно поділити на три ключові напрямки:

1. Методи, засновані на статистичних характеристиках текстових корпусів.
2. Методи, які спираються на структуровані лінгвістичні ресурси, тобто на бази знань про мову, створені вручну або напівавтоматично.
3. Методи композиційної дистрибутивної семантики, що базуються на машинному навчанні та глибоких нейронних мережах.

Методи засновані на корпусах спираються на аналіз великих обсягів тексту з метою виявлення статистичних закономірностей у вживанні слів та фраз. Вони використовують інформацію про частоту спів вживання лексичних одиниць, що дозволяє моделювати семантичні зв'язки без потреби у ручному кодуванні знань. Одним із класичних методів цього напрямку є латентний семантичний аналіз (LSA), який використовує матричну факторизацію для виявлення латентної структури в текстових даних.

Латентний семантичний аналіз (LSA) – це метод математичного моделювання, який застосовується для виявлення прихованих семантичних структур у великому корпусі текстів. Основна ідея LSA полягає у представленні тексту у вигляді матриці термінів та документів, де кожен рядок відповідає терміну, а кожен стовпець – документу або іншій текстовій одиниці. Значення в матриці відображають частоту появи відповідного терміну у відповідному документі, часто із застосуванням вагових коефіцієнтів, які підсилюють значущість рідкісних, але

інформативних слів. Ця матриця фактично кодує інформацію про те, які терміни та в якій кількості присутні у різних текстах, створюючи числове уявлення текстових даних [1].

Для обчислення ваги терміну часто використовується метрика TF-IDF, яка поєднує у собі дві метрики: частоту терміна (tf) та обернену частоту документів (idf) і визначається як:

$$tfidf(t, d) = tf(t, d) \times idf(t), \quad (1)$$

$$tf(t, d) = \frac{f_{t,d}}{\sum_k f_{k,d}}, \quad (2)$$

$$idf(t) = \log \frac{N}{|\{d: t \in d\}|}, \quad (3)$$

де $tf(t, d)$ – частота частота терміну t у документі d , що обчислюється як відношення кількості появ $f_{t,d}$ до загальної кількості термінів у документі; $idf(t)$ – обернена частота документа; N – загальна кількість документів у корпусі, а $|\{d: t \in d\}|$ – кількість документів, у яких зустрічається термін t .

Отримана в результаті обчислень матриця, як правило, дуже велика і розріджена, а також містить багато надлишкової інформації. Для виявлення латентних (прихованих) семантичних структур застосовують матричну факторизацію – метод зменшення розмірності даних, який полягає у поділі великої матриці на добуток двох або більше менших матриць, що зберігають приховану семантичну інформацію. Завдяки факторизації можна виявити латентні ознаки, які відображають семантичні зв'язки між словами чи фразами. Одним із найпоширеніших методів є сингулярне розкладання матриці (SVD), яке дозволяє знизити розмірність простору та представити кожне слово у вигляді вектора в латентному просторі. SVD розкладає матрицю A розміром $m \times n$ на добуток трьох матриць:

$$A = U \times \Sigma \times V^T, \quad (4)$$

де U – матриця лівих сингулярних векторів (терміни та латентні концепти), Σ – діагональна матриця сингулярних значень, які визначають вагу кожного латентного концепту, V^T – матриця правих сингулярних векторів (латентні концепти та документи).

Для зменшення розмірності залишаються лише перші k найбільших сингулярних значень з матриці Σ_k , відповідні їм стовпці в U_k та V_k , що дає апроксимацію матриці A_k . Ця апроксимація відображає основні латентні семантичні структури, дозволяючи порівнювати значення термінів і документів у зменшеному просторі. Через зменшення розмірності LSA дозволяє виявити глибинні семантичні взаємозв'язки між словами та текстами, які неяви у початкових частотних даних. Так, синонімічні чи близькі за змістом терміни наближаються у латентному просторі, навіть якщо вони не збігаються лексично. Це дозволяє використовувати отримані вектори при розв'язанні задачі ідентифікації перифраз шляхом оцінки схожості пари векторів з використанням метрик подібності.

Методи засновані на знаннях – це підходи, що спираються на використання структурованих лінгвістичних і семантичних ресурсів, які відображають знання про мову та світ. Такі ресурси включають онтології, тезауруси, семантичні мережі, такі як WordNet, словники синонімів, глосарії та бази концептуальних зв'язків. Ці методи оперують інформацією про лексичні та семантичні відношення між словами, зокрема синонімією, гіперонімією, гіпонімією, антонімією, а також семантичними ролями та концептуальними зв'язками.

У контексті ідентифікації перифраз підходи, засновані на знаннях, дозволяють виявляти семантичну еквівалентність між виразами внаслідок зіставлення їхніх концептуальних представлень. Наприклад, якщо два висловлювання містять різні лексичні одиниці, але ці одиниці є синонімами у відповідній базі знань, система може зробити висновок про можливу семантичну еквівалентність.

Ступінь семантичної подібності у методах, заснованих на знаннях, визначається шляхом обчислення значень спеціалізованих метрик, які дозволяють кількісно оцінити смислову близькість між поняттями або текстовими фрагментами. Як показано на рисунку 1, ці метрики поділяються на дві основні групи: метрики семантичної подібності та метрики семантичної спорідненості [5].



Рис. 1. Таксономія метрик семантичної еквівалентності у методах заснованих на знаннях

Метрики семантичної подібності спрямовані на оцінку схожості між поняттями, що мають спільну семантичну природу. До цієї групи належать такі підходи, як обчислення довжини найкоротшого шляху між поняттями в онтології, врахування глибини спільного предка, або комбіновані моделі, які враховують як топологічні, так і статистичні характеристики. Загалом дані метрики розділяють на дві групи.

Перша група метрик, до якої належать Resnik (res), Lin (lin) та Jiang & Conrath (jcn), ґрунтується на концепції інформаційного наповнення. Ці метрики використовують статистичні характеристики корпусу для оцінки ймовірності зустрічі понять, припускаючи, що чим менш ймовірне поняття, тим більшою є його інформаційна цінність. Відповідно, семантична подібність між концептами визначається на основі їхньої спільної інформації [5].

Метрика res визначає семантичну подібність через інформаційне наповнення найближчого спільного предка LCS у ієрархії WordNet:

$$sim_{res}(A, B) = IC(LCS(A, B)), \quad (5)$$

$$IC(c) = -\log P(c), \quad (6)$$

де $IC(c)$ – інформаційне наповнення концепту c , $P(c)$ – ймовірність зустрічі концепту c у корпусі текстів, $LCS(A, B)$ – це найближчий спільний предок для концептів A та B у таксономічній ієрархії.

Обчислене значення може приймати значення 0 та бути більше. Верхня межа значення, як правило, досить велика і змінюється залежно від розміру корпусу, який використовується для визначення значень інформаційного вмісту.

Метрики lin та jcn також базуються на інформаційному вмісті, але враховують не лише спільне, а й відмінне між поняттями. Обидві метрики доповнюють інформаційний вміст найближчого спільного предка сумою інформаційного вмісту самих концептів A і B . При цьому lin оцінює семантичну подібність як відношення подвоєного інформаційного вмісту спільного предка до суми інформаційного вмісту обох концептів, тобто фокусується на співвідношенні спільної та унікальної інформації:

$$sim_{in}(A, B) = \frac{2 \times IC(LCS(A, B))}{IC(A) + IC(B)}, \quad (7)$$

де $IC(A)$, $IC(B)$ – інформаційне наповнення концептів A і B відповідно, $LCS(A, B)$ – найближчий спільний предок для концептів A та B .

Це значення наближається до 1, якщо поняття мають спільного предка з високим інформаційним вмістом і мінімальними відмінностями.

Натомість метрика jcn розраховує семантичну відстань між концептами як суму їхнього власного інформаційного вмісту мінус подвоєний інформаційний вміст спільного предка:

$$dist_{jcn}(A, B) = IC(A) + IC(B) - 2 \times IC(LCS(A, B)), \quad (8)$$

де $IC(A)$, $IC(B)$ – інформаційне наповнення концептів A і B відповідно, $LCS(A, B)$ – найближчий спільний предок для концептів A та B .

Чим менше це значення, тим ближчі поняття. Відповідно, для перетворення відстані у подібність використовують обернене значення або інші нормалізації. Таким чином, обидві метрики є більш чутливими до структурної та статистичної інформації, оскільки вони одночасно враховують кількість спільної інформації (через спільного предка) та кількість унікальної інформації кожного з концептів.

Друга група метрик базується на топологічних характеристиках лексичної мережі, зокрема на довжині шляху між поняттями в ієрархії WordNet. До цієї групи входять метрики Leacock & Chodorow (lch), Wu & Palmer (wup) та Path Length ($path$). Вони оперують глибиною ієрархії та відстанню між вузлами: чим коротший шлях або вища спільна ієрархічна вершина, тим більшою вважається семантична подібність між поняттями [5].

Метрика $path$ обчислює семантичну подібність як довжину найкоротшого шляху між концептами A і B в онтологічній ієрархії:

$$sim_{path}(A, B) = \frac{1}{length(A, B)}, \quad (9)$$

де $length(A, B)$ – довжина найкоротшого шляху між концептами A та B .

Показник lch повертає оцінку, яка вказує на те, наскільки схожі значення двох слів, на основі найкоротшого шляху між цими двома значеннями та максимальної глибини таксономії, у якій зустрічаються значення та обчислюється за формулою:

$$sim_{lch}(A, B) = -\log \frac{length(A, B)}{2 \times D}, \quad (10)$$

де $length(A, B)$ – найкоротший шлях між поняттями A та B , D – максимальна глибина ієрархії в WordNet.

Остання метрика wup оцінює подібність на основі глибини найближчого спільного предка відносно глибини самих понять:

$$sim_{wup}(A, B) = \frac{2 \times depth(LCS(A, B))}{depth(A) + depth(B)}, \quad (11)$$

де $depth(A)$, $depth(B)$ – глибина понять A та B у ієрархії, $LCS(A, B)$ – найближчий спільний предок понять A та B .

Метрики семантичної спорідненості є ширшим поняттям і дозволяють враховувати не лише таксономічні відношення, а й асоціативні зв'язки між поняттями – такі як функціональна сумісність, тематична спільність або причинно-наслідкові відношення. Серед метрик семантичної спорідненості виокремлюють три основні підходи: Hirst and St-Onge Measure (HSO),

Lesk та векторні представлення [5]. Метод HSO ґрунтується на топології лексичної бази WordNet і враховує не лише довжину найкоротшого шляху між поняттями, але й кількість змін типів семантичних зв'язків на цьому шляху. Чим менше таких змін і коротший шлях, тим вищий ступінь спорідненості між поняттями. Метрика Lesk базується на принципі перетину словникових визначень. Оцінювання спорідненості здійснюється шляхом аналізу кількості спільних слів у визначеннях двох концептів: чим більший цей перетин, тим вищою є семантична близькість. Підхід vector спирається на статистичне моделювання спів вживання термінів у текстовому корпусі. Для кожного слова, що зустрічається в глосах WordNet, формується вектор контексту, який відображає частотні співвідношення. Відповідно, кожна концепція репрезентується як усереднений вектор термінів, що входять до її глоси.

Варто зазначити, що метрики семантичної спорідненості моделюють радше загальну асоціативну або тематичну близькість понять, а не точну еквівалентність змісту. Саме тому ці підходи мають обмежене застосування у задачах ідентифікації перифраз, де критично важливою є точна семантична відповідність, на відміну від метрик семантичної подібності, що демонструють кращу ефективність.

У роботі [6] автори демонструють використання цих підходів для оцінки семантичної близькості між словами у двох текстах. Перетворення семантичної подібності між концептами в подібність між словами здійснюється шляхом зіставлення їхніх відповідників у лексичній мережі WordNet.

Попри широку застосовність, як методи, засновані на корпусах, так і методи, що спираються на бази знань, мають суттєві обмеження у контексті задачі ідентифікації перифраз. Головна їхня спільна характеристика полягає у тому, що обчислення семантичної близькості виконується переважно на рівні окремих лексичних одиниць або пар слів. У задачі ідентифікації перифраз така обмеженість виявляється критичною. Перифразування за своєю природою передбачає не лише заміну окремих слів синонімами, а й складні трансформації, пов'язані з перебудовою синтаксичної структури, використанням метафор, фразеологізмів, зміною акцентів або способів концептуалізації одного і того ж змісту.

Ці обмеження визначають необхідність застосування складніших моделей семантичного представлення, здатних інтегрувати не лише лексичні, але й синтаксичні та концептуальні аспекти змісту. Важливою властивістю таких моделей є здатність відображати зв'язки між словами у межах речення, включно з ролями агентів, пацієнтів, інструментів, а також каузальних, темпоральних та інших відношень.

З розвитком глибинного навчання парадигма обробки природної мови поступово змістилася від традиційних методів, що базувалися на ручному конструюванні ознак і використанні фіксованих лінгвістичних ресурсів, до моделей розподіленого представлення, які формують семантичні подання мовних одиниць у вигляді векторів у безперервному багатовимірному просторі. На відміну від попередніх підходів, де знання про значення слів і фраз явно кодувалися за допомогою правил або таксономічних структур, сучасні моделі нейронних мереж дозволяють автоматично навчатися узагальнених семантичних залежностей із великих обсягів тексту.

Одним із ключових напрямів у сучасній парадигмі обробки природної мови є композиційна дистрибутивна семантика (Compositional Distributional Semantics). Цей підхід базується на гіпотезі, що значення складних мовних одиниць – таких як фрази, речення або цілі тексти – може бути отримано шляхом композиції векторних подань їхніх складових елементів. Основною ідеєю є поєднання принципів дистрибутивної семантики, яка розглядає значення слова як похідне від контекстів його вживання, із принципами формальної семантики, що описує правила композиції значень.

Сучасні методи композиційної дистрибутивної семантики класифікуються залежно від рівня мовного представлення, на якому формуються векторні репрезентації: методи на рівні слів, методи на рівні фраз, методи на рівні речень.

Методи на рівні слів ґрунтуються на ідеї побудови векторів окремих слів на основі їхнього контексту у великих корпусах тексту. Найвідоміші моделі цієї групи, такі як Word2Vec або GloVe, формують статичні векторні представлення, де кожному слову відповідає один вектор незалежно від контексту та використовуються у групі моделей під назвою SWEM (*Simple Word-Embedding-Based Models*) [7]. Основна ідея цих моделей полягає у використанні векторних представлень слів, зазвичай попередньо навчених на великих корпусах, що відображають латентну семантичну схожість між словами у багатовимірному континуальному просторі. У межах SWEM-підходу кожне речення подається як послідовність векторів, відповідних словам у реченні. Далі застосовуються прості операції агрегації, що дозволяють отримати єдине фіксованої довжини представлення всього речення. Найпоширенішими типами агрегації є:

- агрегування по середньому – обчислення середнього значення по кожній координаті всіх векторів речення;
- агрегування по максимуму – вибір максимального значення по кожній координаті;
- конкатенація середнього та максимуму – конкатенація результатів агрегації по середньому та максимуму для збереження більшої кількості семантичної інформації;
- сумарне агрегування – обчислення сумарного вектора всіх слів у реченні.

Отримані агреговані вектори подаються на вхід класифікаційній моделі, яка навчається визначати ступінь семантичної відповідності між парами речень, класифікуючи їх як перифрази або не перифрази. Головним недоліком описаної групи моделей є нечутливість до порядку слів, що особливо характерно для варіантів із середнім та максимальним згортанням. Через втрату інформації про порядок, такі моделі не здатні розрізняти речення, які мають однаковий лексичний склад, але різну синтаксичну структуру. Це обмеження суттєво знижує застосовність простих SWEM-моделей у задачах, де семантична еквівалентність вимагає точного урахування синтаксису або логіко-семантичних відношень.

Методи на рівні фраз становлять важливу роль в моделюванні локальної семантики, що виникає внаслідок взаємодії між словами в межах фразових конструкцій. На відміну від моделей, які формують значення тексту шляхом простого агрегування векторних представлень слів, фразові методи орієнтовані на врахування синтаксичної структури та граматичних зв'язків між елементами, що дозволяє моделювати складні мовні вирази. Для моделювання семантичного значення на рівні фраз широко застосовуються нейронні моделі, орієнтовані на обробку послідовностей, зокрема рекурентні нейронні мережі.

Рекурентна нейронна мережа (RNN) – це клас штучних нейронних мереж, що дозволяють зберігати та передавати інформацію в часовому порядку. Така архітектура забезпечує обробку послідовних даних, зокрема природної мови, де значення поточного елемента залежить від попереднього контексту. На кожному часовому кроці RNN приймає на вхід поточний елемент послідовності та попередній прихований стан, обчислюючи новий стан, який несе узагальнену інформацію про вже оброблену частину вхідної послідовності (рис. 2).

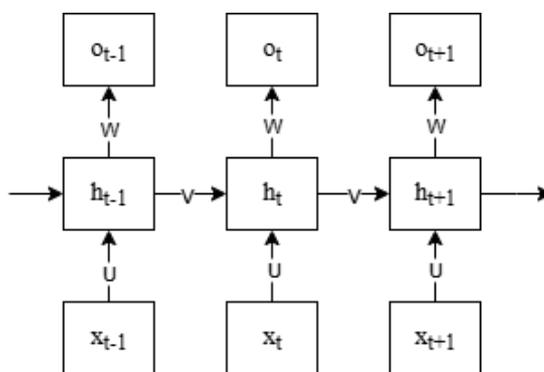


Рис. 2. Архітектура рекурентної нейронної мережі

В контексті задачі ідентифікації перифраз рекурентні мережі використовуються для імплементації архітектури енкодер-декодер. На першому етапі, **енкодер** рекурентної мережі приймає на вхід послідовність слів вхідної фрази та послідовно кодує її в компактне векторне подання, яке відображає семантичний зміст фрази в багатовимірному просторі. Таким чином формується контекстний вектор, що акумулює інформацію про всі елементи речення, враховуючи їхню послідовність та взаємозв'язки, після чого декодер, отримуючи на вхід другу фразу, використовує механізм уваги, щоб зіставити частини обох речень у процесі класифікації.

Такий підхід, наприклад, застосовано у роботі [8], де автори використали архітектуру енкодер-декодер на базі рекурентних нейронних мереж. Запропонований метод спочатку обчислює схожість між парами фраз у вихідному та цільовому реченнях, використовуючи приховані стани RNN-енкодера. Для кожної пари фраз у системі застосовується спеціальний механізм оцінювання, який автоматично визначає, наскільки семантично сумісні ці фрази. Завдяки такому вирівнюванню з використанням RNN-моделей вийшло отримати гнучке представлення фраз, що враховує не лише локальний контекст, а й загальну семантичну сумісність у межах всього речення.

Методи композиційної дистрибутивної семантики на рівні речень є природним продовженням та розвитком підходів, що оперують фразовими структурами. Основна мета таких моделей полягає у побудові цілісного семантичного представлення речення, яке інтегрує як локальні взаємозв'язки між словами, так і глобальний контекст. У порівнянні з методами на рівні фраз, де моделювання зазвичай обмежується короткими синтаксичними структурами (наприклад, іменними чи дієслівними групами), реченнєві моделі здатні охоплювати ширші зв'язки між елементами тексту, включно з довготривалими залежностями та складною семантичною композицією.

Альтернативно до рекурентних мереж, методи на рівні речень також використовують згорткові нейронні мережі – клас нейронних мереж, який спеціалізується на виявленні локальних патернів у структурованих даних, таких як зображення або послідовності. У класичному застосуванні CNN обробляє дані за допомогою згорткових фільтрів, які «ковзають» по вхідному масиву та виявляють характерні ознаки, що є релевантними для поставленої задачі. У контексті обчислення семантичного представлення речення CNN обробляє послідовність векторів слів за допомогою згорткових фільтрів фіксованого розміру, які паралельно обробляють підмножини речення, дозволяючи не зважати на порядок слів у реченні, що є перевагою даного класу моделей перед рекурентними архітектурами, які обробляють дані послідовно. Ці фільтри автоматично вивчають локальні семантичні шаблони, які є релевантними для розв'язання задачі ідентифікації перифраз.

Однією з класичних моделей, що реалізують підхід на основі згорткових нейронних мереж для порівняння речень, є ARC-I (Architectural Convolutional model I) [9]. Дана модель реалізує підхід незалежної обробки вхідних речень, у межах якого кожне речення кодується окремо за допомогою ідентичних згорткових блоків, а обчислення подібності здійснюється лише на завершальному етапі класифікації.

На першому етапі кожне речення представляється як послідовність векторів слів. На отриману послідовність накладаються згорткові фільтри фіксованої ширини k , що дозволяє виявляти локальні контексти у тексті. Далі застосовується операція агрегування по максимуму над усіма виходами згорткового шару, яка забезпечує інваріантність до довжини речення та відбір найбільш релевантних ознак. Таким чином, кожне речення кодується у фіксовану векторну репрезентацію. На наступному етапі ці вектори комбінуються та подаються на вхід багаточаровому перцептроні, який виконує класифікацію пари речень як перифраза або не перифраза. Модель ARC-I продемонструвала конкурентну ефективність на задачах семантичного порівняння, зокрема у випадках, де релевантність текстів залежить від локальних шаблонів або ключових словосполучень. Разом із тим, модель має певні обмеження. Зокрема, незалежне

кодування кожного речення до порівняння не дозволяє явно враховувати взаємозалежності між елементами обох речень, що може призводити до втрати контекстуально важливої інформації.

Попри ефективність методів комбінованої дистрибутивної семантики, зокрема векторних моделей, що відображають контекстуальне значення слів і речень у багатовимірному просторі, ці підходи мають низку обмежень, які суттєво ускладнюють їх використання в задачах глибокого семантичного аналізу. Зокрема, одним із ключових недоліків є спосіб агрегування векторного представлення текстових одиниць: у більшості моделей, семантична репрезентація речення формується шляхом застосування операцій середнього або максимального агрегування над векторами слів. Такий підхід ігнорує порядок слів та синтаксичну структуру речення, що може призводити до некоректної інтерпретації його змісту.

У зв'язку з вищезгаданими обмеженнями, обґрунтованим видається перехід до гібридного підходу, який поєднує переваги глибоких контекстуальних представлень із формалізованими структурними описами значення. Такий підхід дозволяє моделювати не лише поверхневу семантичну схожість між текстами, а й глибинні семантичні зв'язки, що критично важливо для задачі ідентифікації перифраз.

Зокрема, перспективним є представлення семантичної структури речень у вигляді графів, які експліцитно кодують концепти, ролі учасників ситуації, типи подій, логічні зв'язки та інші елементи глибинного змісту. Графові структури такого типу, як абстрактна репрезентація значення або AMR (Abstract Meaning Representation), дозволяють абстрагуватися від поверхневої форми та виявляти еквівалентні висловлювання навіть за суттєвих лексико-синтаксичних відмінностей.

Водночас, сучасні трансформерні архітектури, зокрема моделі типу BERT демонструють високу ефективність у задачах побудови таких графів з природної мови, а також у задачах класифікації чи порівняння графових представлень [4]. Це створює передумови для формування гібридної моделі, в якій семантика тексту спочатку інтерпретується у вигляді AMR-графів, а потім аналізується за допомогою нейронних підходів, що враховують як структурну, так і дистрибутивну інформацію.

Формально абстрактне семантичне представлення (AMR) моделюється як орієнтований мічений граф $G=(V, E)$, де вершини V відповідають поняттєвим одиницям, а дуги $E \subseteq V \times L \times V$ мають мітки з множини L , що позначають семантичні відношення між концептами. Побудова AMR-графа передбачає існування функції відповідності $f: S \rightarrow G$, яка здійснює перехід від простору речень природної мови S до простору графових репрезентацій G , що відображають їх глибинну семантику [2].

На рівні змістового представлення AMR дозволяє явно кодувати:

1. Семантичні ролі: такі як агент ARG0, пацієнт ARG1, інструмент ARG2 та інші, які вказують на різних учасники події або дії, що виражена в реченні.
2. Лексичні концепти: кожен змістовий компонент речення, такі як іменник, дієслово, прикметник, позначається уніфікованим поняттям із лексичного ресурсу.
3. Квантори, модальність, час: AMR дозволяє позначати логічні конструкції, часові характеристики подій, а також модальність.
4. Структуру співвіднесених подій: граф дозволяє представити причинно-наслідкові, умовні й інші зв'язки, що вкрай важливо для цілісного моделювання тексту.

AMR-графи таким чином забезпечують багаторівневу семантичну структуру, яка робить можливим глибший аналіз значення речення у порівнянні з дистрибутивними моделями, які зазвичай не фіксують внутрішньої логіки подієвих чи рольових зв'язків.

Наприклад, робота Fuad Issa та співавторів [10] досліджує використання AMR-репрезентацій разом із методами латентного семантичного аналізу для розв'язання задачі ідентифікації перифраз. У цьому дослідженні автори пропонують комбіновану модель, яка поєднує структурну інформацію з AMR-графів та дистрибутивні ознаки з простору латентних тем, отриманих

за допомогою LSA. Такий підхід дозволяє враховувати як формальні семантичні зв'язки між концептами, так і латентну тематичну близькість між реченнями.

У моделі реалізовано зіставлення пар AMR-графів шляхом обчислення схожості на основі частково вирівняного зіставлення підграфів, після чого ці структурні ознаки комбінуються з векторами, отриманими через LSA, у єдиному класифікаторі на основі SVM. Дослідження охоплювало два основні підходи: побудова розріджених ознак на основі графових компонентів для SVM-класифікатора; формування семантичного представлення речення за допомогою латентно-семантичного аналізу (LSA), де ваги вершин AMR-графів задавалися через PageRank або TF-IDF. Результати досліджень показали, що використання комбінації AMR-графів із техніками латентного аналізу дозволяє досягти вищої продуктивності та більшої семантичної точності при ідентифікації перифраз.

Разом із цим варто відзначити, що ефективність AMR-репрезентацій значною мірою залежить від якості AMR-парсингу. Останні досягнення у сфері трансформерних архітектур, зокрема моделей типу Seq2Seq зі спеціалізованими енкодерами для графових структур, відкривають нові можливості для поліпшення якості автоматичної побудови AMR-графів, так і їхнього подальшого використання в задачах глибокого семантичного аналізу. Інтеграція AMR-структур у трансформери дозволяє:

- кодувати граф як послідовність або мультиреляційне представлення для вхідного шару моделі;
- враховувати глобальний контекст семантичних ролей і структурних залежностей;
- навчати модель на завданнях класифікації перифраз без необхідності ручного створення ознак.

Таким чином, гібридні архітектури, що поєднують AMR-графи як структуровану семантичну репрезентацію з потужністю трансформерних моделей, виглядають як обґрунтований напрям для побудови високоточних систем виявлення перифраз. Такий підхід дозволяє поєднати переваги знанневих репрезентацій та розподілених векторних подань, забезпечуючи як семантичну точність, так і гнучкість у навчанні моделей на великих корпусах.

Висновки

У статті було проведено систематичний аналіз сучасних підходів до обчислення семантичної еквівалентності та побудови семантичних репрезентацій в контексті задачі ідентифікації перифраз. Розглянуто традиційні методи дистрибутивної семантики на рівнях слів, фраз і речень, включно з моделями типу SWEM, ARC-I та нейронними мережами, що враховують синтаксичну і семантичну структуру. Показано, що хоча ці методи демонструють хорошу продуктивність у випадках лексичної або локально-семантичної подібності, вони мають обмеження щодо розуміння глибоких семантичних залежностей, зокрема причинно-наслідкових зв'язків, логічних операторів та узагальнень.

На цьому тлі було обґрунтовано доцільність використання графових семантичних репрезентацій, зокрема Abstract Meaning Representation, які дозволяють формалізувати зміст речення у вигляді орієнтованого графа з концептами та семантичними ролями. Розглянуті дослідження, які застосовували AMR у задачі класифікації перифраз, показали зростання точності при інтеграції AMR-структур у класифікаційний підхід.

У підсумку сформульовано перспективу переходу до гібридних архітектур, що поєднують AMR-графи з трансформерними моделями. Такий підхід дозволяє подолати недоліки класичних дистрибутивних методів, підвищити семантичну глибину репрезентацій та побудувати більш інтерпретовану, структурно узгоджену систему ідентифікації перифраз. Надалі розвиток цієї гібридної моделі дозволить досягти вищої точності, зокрема в мовах із гнучким порядком слів, таких як українська, де поверхневі ознаки є менш інформативними.

Список використаної літератури

1. Landauer T. K., Foltz P. W., Laham D. An introduction to latent semantic analysis. *Discourse Processes*. 1998. Vol. 25, no. 2–3. P. 259–284. <https://doi.org/10.1080/01638539809545028>.
2. Banarescu L., Bonial C., Cai S., Georgescu M., Griffitt K., Hermjakob U., Knight K., Koehn P., Palmer M., Schneider N. Abstract Meaning Representation for Sembanking. In: *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*. Sofia, Bulgaria; 2013. P. 178–186. Association for Computational Linguistics. URL: <https://aclanthology.org/W13-2102.pdf>
3. Dohare S., Karnick H., Gupta V. Text Summarization using Abstract Meaning Representation. *arXiv preprint*. 2017. <https://doi.org/10.48550/arXiv.1706.01678>
4. Yingxu Wang et al. A layered reference model of the brain (LRMB). *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*. 2006. Vol. 36, no. 2. P. 124–133. <https://doi.org/10.1109/tsmcc.2006.871126>.
5. Corley C., Mihalcea R. Measuring the Semantic Similarity of Texts. In: *Proceedings of the ACL Workshop on Empirical Modeling of Semantic Equivalence and Entailment*. Ann Arbor, Michigan; 2005. P. 13–18. Association for Computational Linguistics. URL: <https://aclanthology.org/W05-1203.pdf>.
6. Mihalcea R., Corley C., Strapparava C. Corpus-based and Knowledge-based Measures of Text Semantic Similarity. In: *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*. Boston, MA; 2006. P. 775–780. Association for the Advancement of Artificial Intelligence. URL: <https://aaai.org/Papers/AAAI/2006/AAAI06-123.pdf>.
7. Shen D., Wang G., Wang W., Min M. R., Su Q., Zhang Y., Li C., Henao R., Carin L. Baseline Needs More Love: On Simple Word-Embedding-Based Models and Associated Pooling Mechanisms. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia; 2018. P. 440–450. Association for Computational Linguistics. <https://doi.org/10.18653/v1/P18-1041>.
8. Cho K., van Merriënboer B., Gulcehre Ç., Bougares F., Schwenk H., Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *CoRR*. 2014. URL: <http://arxiv.org/abs/1406.1078>.
9. Hu B., Lu Z., Li H., Chen Q. Convolutional Neural Network Architectures for Matching Natural Language Sentences. *arXiv preprint*. 2015. URL: <https://arxiv.org/pdf/1503.03244>.
10. Issa F., Damonte M., Cohen S. B., Yan X., Chang Y. Abstract Meaning Representation for Paraphrase Detection. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. New Orleans, Louisiana; 2018. P. 442–452. Association for Computational Linguistics. URL: <https://aclanthology.org/N18-1041.pdf>.

References

1. Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25(2–3), 259–284. <https://doi.org/10.1080/01638539809545028> [in English].
2. Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., & Schneider, N. (2013). Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse* (pp. 178–186). Association for Computational Linguistics. <https://aclanthology.org/W13-2102.pdf> [in English].
3. Dohare, S., Karnick, H., & Gupta, V. (2017). Text summarization using abstract meaning representation. *arXiv preprint*. <https://doi.org/10.48550/arXiv.1706.01678> [in English].

4. Wang, Y., S. Chen, H. Yin, et al. (2006). A layered reference model of the brain (LRMB). *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 36(2), 124–133. <https://doi.org/10.1109/tsmcc.2006.871126> [in English].
5. Corley, C., & Mihalcea, R. (2005). Measuring the semantic similarity of texts. In *Proceedings of the ACL Workshop on Empirical Modeling of Semantic Equivalence and Entailment* (pp. 13–18). Association for Computational Linguistics. <https://aclanthology.org/W05-1203.pdf> [in English].
6. Mihalcea, R., Corley, C., & Strapparava, C. (2006). Corpus-based and knowledge-based measures of text semantic similarity. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)* (pp. 775–780). Association for the Advancement of Artificial Intelligence. <https://aaai.org/Papers/AAAI/2006/AAAI06-123.pdf> [in English].
7. Shen, D., Wang, G., Wang, W., Min, M. R., Su, Q., Zhang, Y., Li, C., Henao, R., & Carin, L. (2018). Baseline needs more love: On simple word-embedding-based models and associated pooling mechanisms. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 440–450). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P18-1041> [in English].
8. Cho, K., van Merriënboer, B., Gulcehre, Ç., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *CoRR*. <http://arxiv.org/abs/1406.1078> [in English].
9. Hu, B., Lu, Z., Li, H., & Chen, Q. (2015). Convolutional neural network architectures for matching natural language sentences. *arXiv preprint*. <https://arxiv.org/pdf/1503.03244> [in English].
10. Issa, F., Damonte, M., Cohen, S. B., Yan, X., & Chang, Y. (2018). Abstract meaning representation for paraphrase detection. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 442–452). Association for Computational Linguistics. <https://aclanthology.org/N18-1041.pdf> [in English].

М'якенький Арсеній Вячеславович – аспірант кафедри програмного забезпечення комп'ютерних систем Національного технічного університету «Дніпровська політехніка». E-mail: miakenkyi.ar.v@nmu.one, ORCID: 0000-0002-4141-001X.

Алексєєв Михайло Олександрович – д.т.н., професор, завідувач кафедри програмного забезпечення комп'ютерних систем Національного технічного університету «Дніпровська політехніка». E-mail: alekseev.m.o@nmu.one, ORCID: 0000-0001-8726-7469.

Miakenkyi Arsenii Viacheslavovich – Postgraduate Student at the Department of Software Engineering (DoSE) of the Dnipro University of Technology. E-mail: miakenkyi.ar.v@nmu.one, ORCID: 0000-0002-4141-001X.

Aleksieiev Mykhailo Oleksandrovich – Doctor of Technical Sciences, Professor, Head of the Department of Software Engineering (DoSE) of the Dnipro University of Technology. E-mail: alekseev.m.o@nmu.one, ORCID: 0000-0001-8726-7469.



Отримано: 28.10.2025
Рекомендовано: 27.11.2025
Опубліковано: 30.12.2025