ПРОГРАМНО-ТЕХНІЧНІ КОМПЛЕКСИ

УДК 004.89

О.В. ПАЛАГІН
Інститут кібернетики ім. В.М. Глушкова НАН України, Київ, Україна,
e-mail: *palagin_a@ukr.net*.

В.В. КАВЕРИНСЬКИЙ
Інститут проблем матеріалознавства ім. І.М. Францевича НАН України, Київ, Україна,
e-mail: *insamhlaithe@gmail.com.*

К.С. МАЛАХОВ
Інститут кібернетики ім. В.М. Глушкова НАН України, Київ, Україна,
e-mail: *k.malakhov@outlook.com.*

М.Г. ПЕТРЕНКО
Інститут кібернетики ім. В.М. Глушкова НАН України, Київ, Україна,
email: *petrng@ukr.net.*

ОСНОВИ МЕТОДУ КОМПЛЕКСНОГО ВИКОРИСТАННЯ НЕЙРОМЕРЕЖЕВОЇ ТА ОНТОЛІНГВІСТИЧНОЇ ПАРАДИГМ: КОМПЛЕКСНИЙ ПІДХІД[1]

Анотація. Описано комплексний підхід, який передбачає використання нейромережевої та онтолінгвістичної парадигм. Розроблений метод включає методологічні засади, інформаційну технологію та систему MedRehabBot, які сукупно реалізують основоположні принципи метанавчання, структурованих підказок та підвищують ефективність взаємодії інформаційної системи з чат-ботами і пошуку інформації на основі онтологій. Метод забезпечує можливість адаптування системи MedRehabBot до використання в різних LLM-системах.

Ключові слова: трансдисциплінарні наукові дослідження, онтологія, онтологічний інжиніринг, онтолого-керована інформаційна система, LLM-система, ChatGPT, MedRehabBot, інжиніринг підказок, чат-бот.

ВСТУП

Сучасний етап розвитку науки та її застосунків має явно трансдисциплінарний (ТД) характер. «Цей факт зумовив потребу у розробленні строгої методології ТД наукових досліджень, розширення мережі ТД міжнародних центрів та шкіл, нарешті, визначення місця та ролі інформатики у системно-технологічній підтримці ТД-досліджень та використання їхніх результатів під час розв'язання глобальних проблем розвитку сучасної цивілізації. ТД-парадигма передбачає побудову в найближчому майбутньому загальнонаукової картини світу або, що те саме, єдиної ТД-системи знань, яка забезпечує формалізовану постановку та розв'язання конкретних задач під час виконання комплексних проєктів високої складності, соціальної значущості, конфліктності та конкурентності» [1]. Іншими словами, йдеться про метанауку, здатну пояснювати процеси і явища навколишнього світу, а також розв'язувати прикладні проблеми природи й суспільства, які стають дедалі складнішими.

[1] Дослідження виконано за підтримки Національного фонду досліджень України (грант № 2021.01/0136 «Розробка хмарної платформи пацієнт-центричної телереабілітації онкологічних хворих на основі математичного моделювання») на базі Інституту кібернетики ім. В.М. Глушкова НАН України, Київ, Україна.

ТД-дослідження охоплюють зони прикордонних (демаркаційних) ареалів наукових дисциплін, інтегрують сутнісні основи останніх, утворюючи кластери конвергенції, в яких відбувається потужна синергетична взаємодія за рахунок взаємопроникнення парадигм і конкретних поточних результатів кожної дисципліни у той чи інший кластер. Ця взаємодія відображає цілісність реального світу і є стимулом (і водночас гарантією успішності) до проведення ТД-досліджень, реалізації пов'язаних із нею практичних проєктів, а також нетривіальності і значимості їхніх результатів [1, 2].

Важливим кроком у напрямку трансдисциплінарності є формування перспективних самодостатніх кластерів трансдисциплінарних досліджень, які забезпечують врахування наслідків (ризиків) та взаємний вплив основних факторів та зворотних зв'язків у процесі теоретичного аналізу, цілеспрямованих фізичних експериментів та реалізації глобальних системних проєктів сталого розвитку людського суспільства, збереження довкілля, розвитку науки загалом тощо [3]. Яскравим прикладом може слугувати кластер NBIC-конвергенції (N — «нано», B — «біо», I — «інфо», C — «когніто»), який визначив проривні напрями розвитку новітніх технологій стирання граней між живими та неживими системами, наноробототехніки з її численними застосунками, глобальних суперкомп'ютерних агломерацій з високим рівнем штучного інтелекту тощо. До них слід додати єдину розподілену ТД-систему знань яка є глобально-комунікаційним варіантом загальнонаукової картини світу та знаменує наступний етап розвитку вже наявних мереж «Інтернет» та Semantic Web.

Вважають, що ТД-дослідження мають забезпечити: ефективну ТД-взаємодію на всіх етапах життєвого циклу розв'язання фундаментальних, та прикладних наукових проблем; методологічний супровід та забезпечення процесів інтеграції, конвергенції та уніфікованого формалізованого подання ТД-знань, та операцій над ними; загальну основу взаєморозуміння представників різних наукових дисциплін; включення людини у внутрішньонауковий контекст, загальне зближення природничих та гуманітарних знань тощо.

З огляду на трансдисциплінарний характер сучасних наукових програм і проєктів, актуальними видаються методи аналізу сутнісних процесів конвергенції предметних галузей та пов'язаних з нею процесів когнітивної взаємодії на рівні понятійних структур. Це безпосередньо стосується інформаційної системи MedRehabBot, яка може стати високопродуктивним помічником саме у розробленні нових кластерів конвергенції предметних галузей із застосуванням засобів штучного інтелекту (ШІ).

Використання природної мови під час взаємодії людини і комп'ютера завжди було актуальним, особливо через його зручність порівняно з навчанням спеціальних мов або інструкцій. Замість того, щоб вивчати складні інтерфейси, користувачам було б зручно виражати свої наміри природною мовою. Упровадження природної мови в інтерактивні системи покращить їхню інтуїтивність та ефективність. Віртуальні помічники, як-от: Apple Siri, Amazon Alexa та Google Assistant, вже використовують зазначені можливості. Система діалогових сервісів ChatGPT (далі ChatGPT) від OpenAI, ґрунтується на GPT (Generative Pre-trained Transformer) LLM (Large Language Model) моделях GPT-3, GPT-3.5, GPT-3.5-turbo та GPT-4 [4, 5], доопрацьованих для розмовних застосунків (natural language applications) з використанням методів навчання під контролем і з підкріпленням [4, 6]. Її поява ознаменувала собою значний прорив у галузі ШІ, зокрема у галузях оброблення (NLP, Natural Language Processing) і розуміння (NLU, Natural Language Understanding) природної мови і є суттєвим кроком вперед.

Виконані авторами цієї статті дослідження є продовженням, розвитком та вдосконаленням їхньої попередньої праці [7]. Вони дали можливість розробити та апробувати новітню інтерактивну інформаційну систему підтримки лікаря фізичної і (теле) реабілітаційної медицини (ФРМ), студентів спеціальності ФРМ та пацієнтів, названу MedRehabBot. Вона дає змогу надавати інформацію та виконувати логікові виведення (logic inference) на основі певного набору контекстів, функціонуючи як діалогова система. Слід зазначити, що роботу виконано в межах грантового проєкту Національного фонду досліджень України «Розробка хмарної платформи пацієнт-центричної телереабілітації онкологічних хворих на основі математичного моделювання» [8–10]. Авторський колектив входить до складу Міжвідомчої робочої групи з питань розроблення Концепції впровадження телемедицини в Україні, а саме підгрупи «Технічні питання та архітектура телемедицини».

Реалізацію методу підвищення ефективності взаємодії онтолінгвістичних та нейромережевих засобів продемонстровано з використанням набору наукових робіт, написаних українською та англійською мовами у галузі ФРМ, зокрема е-реабілітації [11].

Платформа OpenAI [12] має обмежений API (Application Program Interface), основною особливістю якого є відправлення цільових повідомлень-підказок (prompts) природною мовою без чітко визначеної структури. Для більш зручного користування API та вказування параметрів і налаштувань створено пакет-обгортку LangChain [13–15]. Основною мовою для GPT-моделей від OpenAI та ChatGPT є англійська [4]. Через обмеження на кількість токенів, які за одним запитом може обробити ChatGPT, інструкції повинні бути стислими, але інформативними. Експериментальні дані роботи з ChatGPT [7, 16, 17] показали, що одним з ефективних підходів для надання стислих, але вичерпних команд та інструкцій є їхнє представлення у форматі JSON. У такий спосіб можна гарантувати впорядкованість, легкість, інтерпретовність та максимальну ефективність інформації для ChatGPT. Такий підхід дає змогу чітко подати системі бажані дії та очікування і забезпечує оптимізацію взаємодії між користувачами та ChatGPT [18–20].

Використання онтологій, як сховища правил поведінки, розглянуто в [21, 22], хоча без прямих посилань на ChatGPT. Онтологія може сприяти прийняттю рішень та представленню даних в інтерфейсі. Онтолого-керовану архітектуру розглянуто у роботі [23].

МОДЕЛІ ІНТЕРАКТИВНОЇ ДОВІДКОВО-ІНФОРМАЦІЙНОЇ СИСТЕМИ MedRehabBot ДЛЯ ПІДТРИМКИ РЕАБІЛІТОЛОГІВ, СТУДЕНТІВ ТА ПАЦІЄНТІВ У ГАЛУЗІ ФІЗИЧНОЇ РЕАБІЛІТАЦІЇ ТА ТЕЛЕРЕАБІЛІТАЦІЇ

Інформаційна модель та інструментарій розроблення системи MedRehabBot. Система MedRehabBot використовує інформаційну модель, ґрунтовану на композитному сервісі (S) з трикомпонентного кортежу [7, 24, 25]. Ця модель дає змогу інтегрувати різноманітні сервіси у межах такої системи:

$$S = \langle D, F, E \rangle$$

де $D = \{ws_w, as_d \mid w = \overline{1, k}, d = \overline{1, l}\}_{k, l = \overline{1, N}}$ — набір вебсервісів (ws_w) та прикладного програмного забезпечення (as_d), доступний для розробників, який дає змогу розробляти різноманітні застосунки та сервіси всередині системи, тобто набір для розроблення MedRehabBot, N — множина цілих невід'ємних чисел; $F = D : \{C_j \mid j = \overline{1, n}\}_{n = \overline{1, N}}$ — набір функцій, які охоплюють функ-

ціональні аспекти інформаційної технології, яку реалізує система MedRehabBot. Кожна функція відповідає певному конвеєру або процесу керування знаннями, який реалізується в результаті інтеграції та взаємодії елементів у межах D; $C_j \subseteq D, C_j = \{ws_o, as_p \mid o, p \geq 0, o \leq k, p \leq l\}_{o,\, p=\overline{1,N}}$ — підмножина вебсервісів та прикладного програмного забезпечення, які потрібні для успішної реалізації j-ї функції в межах D; $E = \{prl, os, floss\}$ — набір елементів, які об'єднуються разом у межах рівневої структури та формують інтегроване середовище розвитку знань (K-IDE) [24].

Елемент *prl* являє собою фізичне обладнання та ресурси об'єкта, як визначено у [24]. Елемент *os* належить рівню операційної системи (ОС) та представляє гостьову ОС у K-IDE. Рівень ОС призначено для використання Unix-подібних ОС, таких як Ubuntu Server для систем x86 та DietPi — легка ОС на основі Debian для одноплатних комп'ютерів (single-board computer, SBC) на базі архітектури ARM. Цей рівень забезпечує основу для роботи фреймворку K-IDE та сумісність з вибраними ОС і середовищами десктопних комп'ютерів. Рівень FLOSS, позначений як *floss*, представляє компонент вільного та відкритого програмного забезпечення (free and open-source software, FLOSS) у складі K-IDE. Цей рівень охоплює як внутрішні, так і зовнішні програмні компоненти. Формалізоване представлення вебсервісів *ws* та *as*, а також їхні складові детально описані в [7].

Функціональна модель системи MedRehabBot. Функціональне розширення системи MedRehabBot представлено таким набором F функцій, синтезованих з D:

$$F = \{C_1, C_2, C_3\},$$

де C_1 — автоматична генерація OWL-онтології (мультимножина RDF-трійок), яка формально описує природномовний текст; C_2 — функція онтолого-керованого діалогу, яка інтегрує ChatGPT та структуровані підказки; C_3 — структуровані підказки для функції метанавчання ChatGPT.

МЕТОДОЛОГІЧНІ ЗАСАДИ ПОБУДОВИ ТА ВИКОРИСТАННЯ СИСТЕМИ MedRehabBot

Методологію можна розділити на дві ключові складові, кожна з яких слугує окремій меті в розробленні системи MedRehabBot.

У першій складовій зосереджено увагу на техніці метанавчання на основі створення структурованих підказок для ChatGPT. Цей підхід забезпечує керування процесом метанавчання ChatGPT, що дає змогу генерувати більш контекстуально релевантні та точні відповіді. Детально розглянуто процедури розроблення та впровадження цих підказок, наголошено на їхній важливості для покращення розмовних можливостей ChatGPT.

У другій складовій основну увагу приділено розробленню автоматичної онтолого-керованої діалогової системи, яка інтегрує ChatGPT зі структурованими підказками. Основною ідеєю розроблення цієї системи є ефективне маніпулювання контекстами певних предметних галузей, які можуть містити специфічну для домену інформацію, не повністю охоплену базою знань ChatGPT. Ці контексти зберігаються в базі даних, наприклад у документо-орієнтованій системі керування базами даних (СКБД) MongoDB або об'єктно-реляційній СКБД PostgreSQL, і пов'язані з наборами іменованих сутностей зі своєю власною структурою, подібною до онтології. До того ж для категоризації контекстів можна використовувати аналіз наративів. Прив'язка іменованих сутностей до відповідних контекстів включає семантичні компоненти, що пояснюють роль сутності в контексті. За допомогою цих додаткових функцій підвищують релевантність і чіткість вибраного контексту для подальшого оброблення. Для автоматизації цих процесів використовують раніше розроблений інструментарій [2, 7, 26].

Для семантичного аналізу та виокремлювання іменованих сутностей із фраз, наданих користувачем, ChatGPT виявляється корисним ресурсом. Саме

для цього створено спеціалізовані підказки. Крім того, ChatGPT використовують для аналізу інтенцій користувацьких фраз. Визначені інтенції разом з виокремленими іменованими сутностями, анотованими їхніми семантичними ролями, і вибраним переліком контекстів подають як вхідні дані для ChatGPT. Ці вхідні дані супроводжують відповідними структурованими підказками, які пояснюють інформацію, що має бути виокремлена, і бажаний формат її представлення.

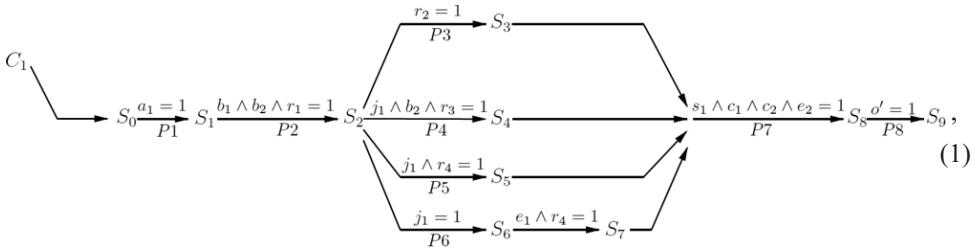
Загальну схему діаграми моделі C4 [27] (зокрема, контексту та контейнера) наведено в [7].

Однією з головних особливостей системи є гнучкість структурованих підказок для ChatGPT. Їх динамічно генерують замість фіксованих підказок залежно від відповідної ситуації за допомогою інструкцій, наданих у вигляді метаонтології. Остання описує поля, які мають бути включені у структуру формату JSON (або XML), і відповідні фрази підказок, які потрібно вставити. Кожна інструкція або структурована підказка для ChatGPT має власний набір полів і попередньо визначених значень, які можуть бути включені. Крім того, підказка містить шаблонну структуру для відповіді, що забезпечує узгодженість і спрощує подальше оброблення. Під час створення фраз підказок використовують перевірені методики з [18, 28, 29] для досягнення ефективних і послідовних підказок.

У структурі системи MedRehabBot є декілька компонентів, станів, процедур і змінних, що сприяють функціональності та роботі діалогової системи. Нижче наведено огляд ключових елементів.

Роботу системи MedRehabBot можна представити за допомогою схеми, подібної до мережі Петрі [30], а саме у вигляді модифікованого графу маркування (marking graph) системної мережі [31]. Таке подання забезпечує структуроване представлення функціонування системи з відтворенням потоку та взаємодії між різними компонентами. Більш детальний розгляд функціонування складових графу проілюстровано наведеними нижче виразами.

До автоматичного створення онтології контекстів можна застосувати різні підходи залежно від природи вихідних даних і особливостей роботи системи в цілому. Нижче наведено один з можливих прикладів автоматичного створення такої онтології на базі набору вихідних файлів у форматі PDF, що мають регулярну, наперед відому структуру. Базова структура онтології у цьому випадку також є відомою. Зазначений вище компонент C_1 можна описати за допомогою такої модифікованої мережі Петрі:



(1)

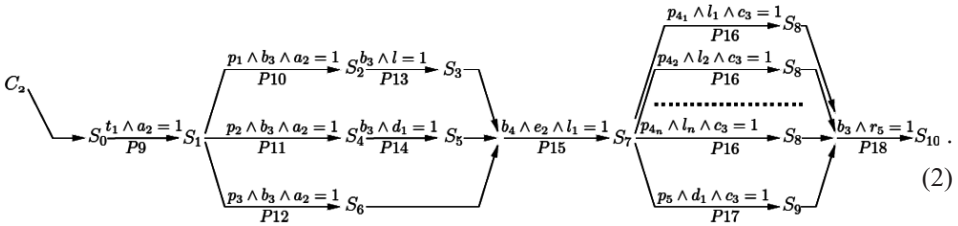
де стани для компонента C_1 (від S_0 до S_9) є такими: S_0 — вихідний стан (набір файлів у форматі PDF); S_1 — список текстових представлень PDF-документів; S_2 — набір документів у структурованому JSON представленні; S_3 — програмне подання базових компонентів онтології, що не залежить від вихідних даних. Сюди входить опис декларацій і предикатів, а також класів і властивостей верхнього рівня, що є базовими елементами. У цьому випадку, наприклад, до цих властивостей віднесено ті, що відповідають за прив'язку контекстів до певної статті (документа) та іменованих сутностей до

відповідних контекстів; S_4 — програмне представлення класів онтології; S_5 — програмне представлення екземплярів (Named Individual) онтології; S_6 — виділені з відповідних контекстів іменовані сутності; S_7 — програмне представлення іменованих сутностей як екземплярів (Named Individual) онтології; S_8 — сукупне програмне представлення створюваної онтології; S_9 — файл, що містить OWL опис онтології контекстів.

Процедури у виразі (1) є такими: $P1$ — отримання текстових даних з вихідних PDF- файлів. Для кожного файлу текст отримують у вигляді списку рядків; $P2$ — отримання на основі текстових даних JSON-структур заданого вигляду із заповненням відповідними контекстами; $P3$ — побудова програмного представлення базової онтологічної структури заданого вигляду; $P4$ — створення програмного представлення набору класів онтології. Класи переважно створюють на базі ключів JSON-словників представлення документів, беручи до уваги заданий шаблон структури; $P5$ — створення програмного представлення набору екземплярів (Named Individual) онтології на базі контекстів чи іменованих сутностей; $P6$ — розпізнавання іменованих сутностей (named entity recognition) із контекстів; $P7$ — складання програмного представлення онтологічної структури з отриманих класів, властивостей та екземплярів; $P8$ — серіалізація представлення створеної онтологічної структури у вигляді OWL файлу.

Вираз (1) містить такі змінні: a_1 — вихідний набір PDF-документів; b_1 — список текстових представлень PDF-документів; b_2 — задана структура створюваних JSON-представлень; r_1 — набір правил і інструкції щодо розбору текстових представлень документів; r_2 — набір правил і інструкції побудови загальної структури онтології; j_1 — сукупність JSON-представлень вихідних документів; r_3 — набір правил і інструкції побудови програмних представлень класів онтології; r_4 — набір правил і інструкції побудови програмних представлень екземплярів онтології; e_1 – сукупність виділених із контекстів іменованих сутностей; s_1 — програмне представлення загальної структури онтології; c_1 — програмне представлення класів онтології; c_2 — програмне представлення екземплярів контекстів онтології; e_2 — програмне представлення екземплярів іменованих сутностей онтології; o' — програмне представлення онтології контекстів.

Компонент C_2 у виразі наведеному нижче — це онтолого-керована діалогова функція, яка інтегрує ChatGPT та структуровані підказки.



(2)

Вираз (2) відображає процес діалогового акту між користувачем та системою MedRehabBot, де стани для компонента C_2 (від S_0 до S_{10}) є такими: S_0 — початковий стан отримання текстового повідомлення від користувача; S_1 — стан попередньо обробленого тексту, підготовленого до подальшого виконання операцій; S_2 — стан списку визначених інтенцій; S_3 — стан сформованого шаблону запиту на виокремлення інформації; S_4 – стан логікового виведення, який необхідно виконати; S_5 — стан сформованого шаблону підказки виведення висновків; S_6 — стан видобутого списку іменованих сутностей; S_7 —

стан списку вибраних контекстів; S_8 — стан видобутої інформації з контекстів; S_9 — стан висновків, отриманих з контекстів; S_{10} — стан отриманих результатів у зручній для користувача формі.

Процедури у виразі (2) є такими: $P9$ — первинне оброблення тексту; $P10$ — визначення інтенцій, виражених у вхідному тексті; $P11$ — визначення того, чи передбачає інтенція отримання висновків, спонукаючи ChatGPT надавати відповідну інформацію з переданих контекстів або пов'язаних з ними знань; $P12$ — розпізнання іменованих сутностей з вхідного тексту; $P13$ — формування підказок для визначення інтенцій; $P14$ — формування підказки для генерації висновків; $P15$ — вибір релевантних контекстів на основі ідентифікованих іменованих сутностей та інтенцій; $P16$ — виокремлення інформації з вибраних контекстів відповідно до інтенції; $P17$ — виведення висновків з вибраних контекстів; $P18$ — форматування та представлення результатів користувачеві.

Змінні у виразі (2) є такими: t_1 — набір правил та операцій, які будуть застосовані під час попереднього оброблення вхідного тексту; a_2 — початковий текст, наданий користувачем; a_3 — попередньо оброблений і підготовлений текст для подальших операцій; b_3 — метаонтологія, що включає правила роботи та інструкції з підготовки оперативних повідомлень; b_4 — онтологія контекстів, пов'язаних із системою; p_1 — підказка для визначення інтенцій; p_2 — підказка аналізу, чи необхідно виконати якусь дію; p_3 — підказка на виокремлення іменованої сутності; $p_{4_{n,\, n=\overline{1,N}}}$ — підказки для виокремлення інформації на основі конкретної інтенції n, $n=\overline{1,N}$; p_5 — підказка для ChatGPT генерувати висновки із заданих контекстів; l — набір інтенцій, визначених у вхідному тексті, що активується відповідно до певних сутностей; l_n — конкретна єдина інтенція, $n=\overline{1,N}$; d_1 — вказує на те, чи будуть зроблені висновки і якщо так, то на які саме висновки очікують; e_2 — множина іменованих сутностей, знайдених у вхідному тексті; c_3 — вибрані контексти; r_5 — структури даних, що представляють отримані результати від ChatGPT.

Процедуру діалогового акту можна описати у такий спосіб: після отримання текстової інформації (запиту) від користувача вона проходить аналіз інтенцій з використанням ChatGPT за допомогою спеціалізованої підказки. Результатом є список словників, що містять такі ключі:

- «назва» ("name") — назва інтенції зі списку, наданого у запиті;
- «тип» ("type") — більш загальна класифікація інтенції, наприклад, «розповідь», «опитування» або «імператив»;
- «ймовірність» ("probability") — значення з плаваючою комою в діапазоні від 0 до 1, що вказує на ймовірність наявності інтенції в тексті користувача;
- «суб'єкт» ("subject") — суб'єкт, пов'язаний з інтенцією, якщо цей зв'язок можливий;
- «об'єкт» ("object") — об'єкт, пов'язаний з інтенціями, якщо цей зв'язок можливий.

Запит може містити різні можливі інтенції, як-от: «кількість», «спосіб виконання», «об'єкт», «суб'єкт», «дія», «місце», «напрямок», «місце дії», «умови», «інструмент», «співучасник», «зв'язок», «причина», «послідовність», «походження» тощо. Ці інтенції представляють семантичні категорії та забезпечують основу для розуміння запиту користувача. До того ж структурований запит містить поля для надання інформації, мови та інших технічних деталей, пов'язаних із введенням і виведенням даних.

Одночасно за допомогою ChatGPT та іншої підказки виконують виокремлення іменованих сутностей. Результат має містити лематизовані слова, згрупо-

вані за сутностями, тип кожної групи (іменникова чи дієслівна), а також головне слово в кожній групі.

На основі виокремлених з інтенцій іменованих сутностей та їхніх семантичних ролей система підбирає відповідні контексти з онтології контекстів.

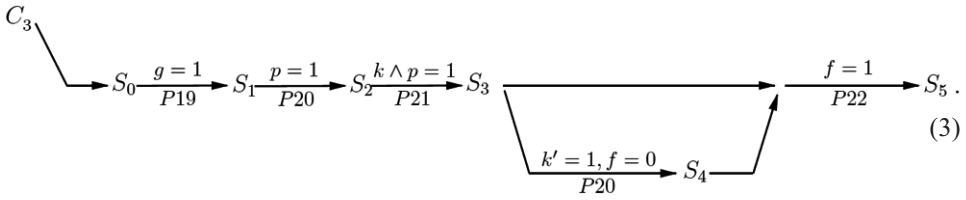
Далі виконують запит до ChatGPT, який включає попередньо вибрані контексти та списки інтенцій, а також відповідні сутності (суб'єкти та об'єкти), актуалізовані у підказках. Форма цього запиту може суттєво відрізнятися і залежить від певної інтенції або списку інтенцій. Правила його побудови включно з обов'язковими полями та фразами підказок визначені в метаонтології.

Метаонтологія також охоплює правила для остаточного представлення результатів, залежно від типів задіяних полів. Формою представлення результату може бути звичайний текст, числа, дати, списки, таблиці або посилання на зовнішні ресурси. Є й інші можливі варіанти представлення результату.

Створення метаонтології передбачає метанавчання, яке охоплює розроблення та доопрацювання відповідних підказок. У цьому ітеративному процесі беруть участь інженер зі знань (окремий актор) та ігровий майданчик (playground) платформи OpenAI. Спочатку інженер зі знань формулює структуровану підказку у форматі JSON (XML, YAML тощо), що складається з ключів і фраз підказок, розроблених з певною метою, спираючись на попередній досвід розроблення підказок і загальні знання. Потім цю підказку надають ChatGPT, і відповідь аналізують, щоб визначити, чи відповідає вона поставленій меті.

Якщо відповідь ChatGPT є повністю неправильною або містить недоліки, до початкової підказки вносять зміни. Цими змінами може бути додавання додаткових полів, видалення зайвих або неефективних аспектів підказки, а також редагування фраз підказки для покращення зрозумілості завдання. Покращену підказку передають знову до ChatGPT для подальших ітерацій. Цей ітеративний процес триває доти, доки отримана відповідь не буде повністю відповідати бажаному та очікуваному результату.

Після розроблення відповідного набору підказок і вивчення поведінки ChatGPT на цих підказках їх об'єднують у формат метаонтології. Створення метаонтології є ручним процесом, де структуру підказок і властивості їхніх полів описують на основі різних передбачуваних ситуацій і цілей розв'язування задачі. Формалізоване графічне представлення компонента C_3 процесу метанавчання показано у такому виразі:



(3)

Тут стани для компонента C_3 (від S_0 до S_5) є такими: S_0 — мета створення підказки; S_1 — початкова підказка, створена на основі здогадок та життєвої мудрості; S_2 — відповідь ChatGPT на початкову підказку; S_3 — виправлена та відредагована підказка; S_4 — відповідь ChatGPT на виправлену підказку; S_5 — метаонтологія з інструкціями та правилами для нової підказки.

Процедури у виразі (3) є такими: $P19$ — висловлення обґрунтованих здогадок і міркувань щодо нової підказки; $P20$ — ініціалізація ChatGPT; $P21$ — виправлення та редагування підказки; $P22$ — інтеграція нової підказки в метаонтологію.

У цьому виразі вжито такі змінні: g — вихідна ідея для створення нової підказки; p — початкова підказка; p' — доопрацьований варіант запиту; f — фінальна версія запиту; k — відповідь ChatGPT на початковий запит; k' — відповідь ChatGPT на змінений запит.

ОТРИМАНІ РЕЗУЛЬТАТИ

Розроблено прототип запропонованої системи MedRehabBot, який включає всі основні компоненти. Докладні робочі матеріали можна знайти в публічному репозиторії GitHub [32], пов'язаному з цією статтею. Підказки в системі формулюються у вигляді JSON-структур. Наведена нижче JSON-схема, яка представляє приклад підказки, орієнтована на визначення інтенцій та їхнього зв'язку з відповідними сутностями (якщо це визначення є можливим):

```
{
    "information to provide": [
    "define intents",
    "find subjects",
    "find objects"
    ],
    "text": "<A text to be analyzed>",
    "language": "Ukrainian",
    "input information field": "text",
    possible intents": [
        "quantity", "place", "way of doing", "object", "subject", "action", "location",
        "direction", "scene of action", "conditions", "instrument", "collaborator",
        "relation", "cause", "sequence", "origin"
  ],
  "several intents": true,
  "intents probability": true,
  "show intent subject": true,
  "max intents number": 4,
  "intents arrange": "by probability",
  "output format": "JSON",
  "output representation template": {
    "result": [
      {
        "intent": "intent name - string",
        "type": "narration, interrogation or imperative",
        "probability": "float value",
        "subject": "subject of the intent as a name group - string",
        "object": "object of the intent as a name or verb group - string"
      }
    ]
  }
}
```

Суттєвою частиною є вказівка «інформація для надання» ("information to provide"), яка формулює основні цілі цієї підказки–завдання. Текст вхідного повідомлення, яке підлягає аналізу, зазначено в полі "text". Для поліпшення продуктивності визначено мову вхідного тексту. У цьому прикладі це українська ("Ukrainian"). Список "possible intents" («можливі інтенції») використано для точного встановлення множини інтенцій, які можна виявити

у вихідному тексті. Для надання технічної інформації про виведення передбачено кілька полів булевого типу. У цьому випадку вони представлені як “several intents” («кілька інтенцій»), “intents probability” («ймовірність інтенцій») та “show intent subject” («показати суб’єкти інтенції»).

Перше поле дає змогу ідентифікувати кілька інтенцій у наданому тексті, друге поле допомагає ChatGPT оцінити ймовірність кожної інтенції, а останнє поле вказує на сутності, які конкретизують інтенції. Для уникнення зайвого розпізнавання інтенцій їхню кількість можна обмежити (у цьому дослідженні взято чотири інтенції) та впорядкувати їх за ймовірністю. Для спрощення подальшого оброблення результатів рекомендовано визначити шаблон структури вихідних даних, зазначивши формат, в якому інформація повертається. Це регулюють полем “output representation template" («шаблон представлення вихідних даних»).

Запропонований метод протестовано на прикладі системи MedRehabBot, побудованої на основі набору інформаційних джерел у галузі фізичної та реабілітаційної медицини. Для створення бази знань використано датасет “EBSCO articles dataset (domain knowledge: rehabilitation medicine) + JSON of every article” [33] та побудовано на його основі онтологію контекстів [32]. Класифікацію отриманих відповідей здійснено у такий спосіб:

• «дійсно (істинно) позитивні» (TP): відповідь була згенерована ChatGPT і виявилася правильною;

• «дійсно (істинно) негативні» (TN): ChatGPT визнав свою неосвіченість або вказав на недостатність інформації. В цю категорію також включено випадки, коли ChatGPT не зміг надати відповідь, і відповідна інформація відсутня у контекстах;

• «хибно позитивні» (FP): система намагалася надати відповідь, але вона виявилася неправильною;

• «хибно негативні» (FN): ChatGPT не надав відповідь, хоча правильна відповідь була наявна у контекстах.

Слід зауважити, що на етапі тестування були виключені запитання та фрази з не пов’язаних між собою предметних галузей, що містили назви сутностей, відсутніх у контекстній онтології. У таких випадках отримують негативний результат, оскільки не буде вибрано жодного релевантного контексту, і подальше оброблення не відбудеться. Отже, всі тести були сформульовані у такий спосіб, щоб вони відповідали запропонованому підходу.

Також потрібно зазначити, що запропонований підхід допускає можливість декількох відповідей на одне запитання, насамперед завдяки наявності кількох визначених інтенцій. Під час оцінювання всі надані відповіді були враховані, незалежно від того, чи були вони надані на одне й те саме запитання, чи на різні.

Маніпулювання експериментальними даними дає змогу отримати комплексну оцінку продуктивності системи та оцінку її здатності обробляти різноманітні запити, враховуючи визначені інтенції та виокремлюючи релевантну інформацію з вибраних контекстів.

Результати тестування наведено у табл. 1.

У результаті тестування запропонованого методу отримано такі значення для стандартних метрик оцінювання:

• точність: 0.8684;
• достовірність: 0.9310;
• відтворюваність: 0.9000;
• оцінка F1: 0.9152.

Таблиця 1. Результати тестування пропонованого методу

Відповідь	Істинно	Хибно
Позитивна	27	2
Негативна	6	3

Ці метрики забезпечують кількісну оцінку роботи системи з погляду її точності, достовірності, відтворюваності та загальної ефективності. Метрика точності відображає частку правильних відповідей, наданих системою, порівняно із загальною кількістю запитів. Показник точності вимірює здатність системи надавати точні відповіді з-поміж тих, що вона генерує. Показник достовірності вказує на здатність системи знаходити всі релевантні відповіді з наявних контекстів. Показник F1 поєднує в собі точність і відтворюваність, щоб забезпечити збалансовану оцінку загальної продуктивності.

Отримані значення метрик демонструють потенційну придатність запропонованого методу, хоча його ще не можна назвати досконалим.

Автори цієї статті усунули недолік, виявлений в їхньому попередньому дослідженні [7], що полягав у відносно низькому значенні точності внаслідок високої ймовірності хибно позитивних відповідей. Це поліпшення зумовлене досконалішою структурою онтології контекстів та запитів до неї, що виокремлюють більш релевантні контексти, які дійсно містять відповіді. Завдяки цьому ChatGPT не має потреби виконувати пошук наближеної інформації у переданому йому великому, але не завжди релевантному контексті.

Також варто зазначити, що загалом у межах розглядуваної задачі хибно позитивні відповіді часто супроводжувалися істинно позитивними і самі собою не становлять суттєвої проблеми. Отже, їх можна розглядати як додаткову інформацію, що може мати дотичне відношення до основної відповіді. Це спостереження навіть підкреслює потенційну цінність врахування хибно позитивних відповідей у практичному контексті.

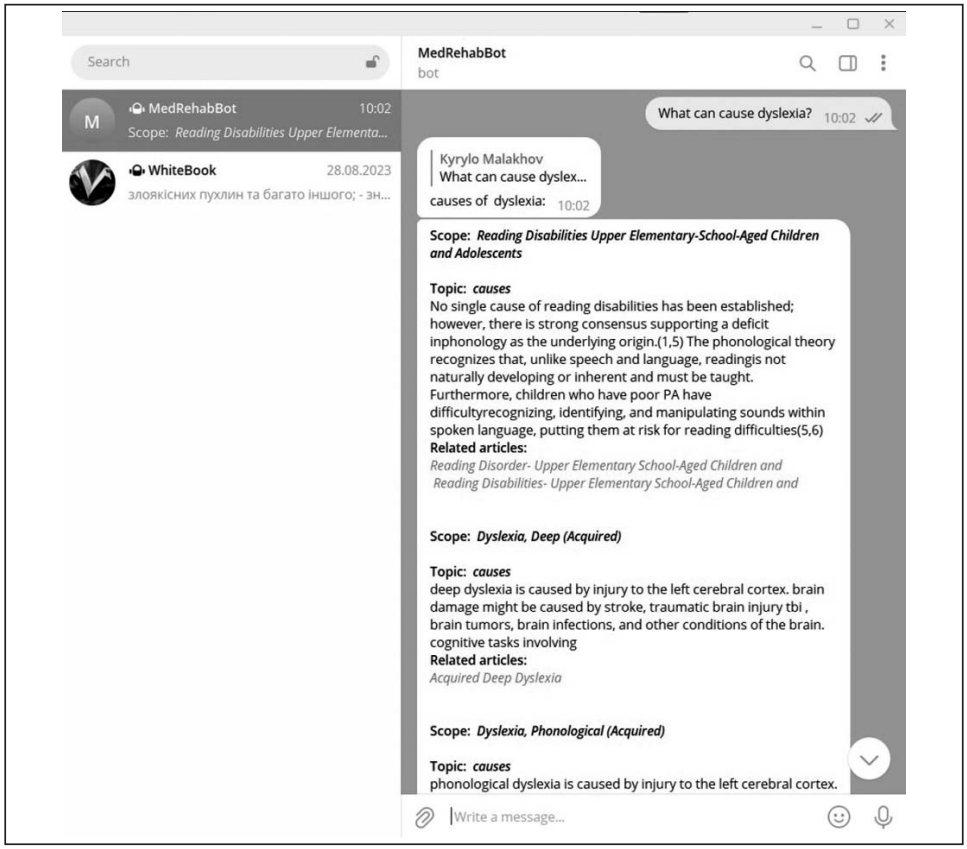


Рис. 1. Графічний інтерфейс користувача програмної реалізації системи MedRehabBot (у вигляді додатка для системи обміну повідомленнями «Телеграм»)

Розглянемо такий приклад. На рис. 1 наведено графічний інтерфейс користувача програмної реалізації системи MedRehabBot (у вигляді застосунку для системи обміну повідомленнями «Телеграм»). Початкова фраза є такою: “What can cause dyslexia?” («Що може спричинити дислексію?»). Система розпізнає такі інтенції:


```
[
    {
        “intent”: “cause”,
        “type”: “interrogation”,
        “probability”: 0.8,
        “subject”: “”,
        “object”: “dyslexia”
    },
    {
        “intent”: “origin”,
        “type”: “interrogation”,
        “probability”: 0.7,
        “subject”: “dyslexia”,
        “object”: “”
    },
    {
        “intent”: “way of doing”,
        “type”: “interrogation”,
        “probability”: 0.6,
        “subject”: “cause dyslexia”,
        “object”: “”
    },
    {
        “intent”: “action”,
        “type”: “interrogation”,
        “probability”: 0.5,
        “subject”: “can cause dyslexia”,
        “object”: “”
    }
]
```


Для вибору контекстів знайдено такі іменовані сутності:


```
[
    {
        “words”: [“cause”, “dyslexia”],
        “type”: “noun”,
        “main word”: “cause”
    },
    {
        “words”: [“dyslexia”],
        “type”: “noun”,
        “main word”: “dyslexia”
    }
]
```


Автори визначили такі інтенції: “cause” («причина»), “origin” («походження») та “way of doing” («спосіб дії»), “action” («дія»). У всіх випадках речення визначене як пи-

тальне (”interrogation”). Усі ці інтенції мають відносно високу ймовірність і мають бути враховані для отримання остаточного набору відповідей. Отримано такі відповіді:

— “No single cause of reading disabilities has been established; however, there is strong consensus supporting a deficit in phonology as the underlying origin. The phonological theory recognizes that, unlike speech and language, reading is not naturally developing or inherent and must be taught. Furthermore, children who have poor PA have difficulty recognizing, identifying, and manipulating sounds within spoken language, putting them at risk for reading difficulties” (“cause”).

— “Deep dyslexia is caused by injury to the left cerebral cortex. Brain damage might be caused by stroke, traumatic brain injury TBI, brain tumors, brain infections, and other conditions of the brain. cognitive tasks involving”. (“origin”, “way of doing”, “action”).

Фактично було надано два варіанти відповіді. Друга версія, отримана згідно з інтенцією (“origin”, “way of doing” та “action”), є більш прямою і вказує на фізіологічні причини глибокої дислексії. Перший варіант теж можна вважати релевантним та таким, що змістовно доповнює другий.

Загалом довідкова система може обмежуватися наданням лише набору фрагментів готових текстів з онтології контекстів. Проте пропускання їх через велику мовну LLM-модель (як-от GPT-3.5-turbo) із зазначеним набором інтенцій дає змогу сформувати більш конкретну і чітку відповідь саме на поставлене запитання, для якої контексти слугують джерелом інформації.

Як уже зазначено, запропонований підхід надає ChatGPT додаткову інформацію і розширює його можливості для надання відповідей у порівнянні з використанням лише наявної бази знань. Нижче розглянуто приклад, що демонструє таке порівняння роботи просто ChatGPT і ChatGPT з передачею підказки, яка містить інтенції і контексти. Текст запитання є таким: “What is ICD-10 code for breast cancer?”

ChatGPT надає відповідь “The ICD-10 code for breast cancer is C50”. Відповідь правильна. Запропонована модель, що використовує передачу контекстів зі створеної заздалегідь онтології, також надає значення коду C50. Проте вона, до того ж, надає і пояснення до нього: “personal history of malignant neoplasm of breast (ICD codes are provided for the reader’s reference, not for billing purposes)”. Також ця модель надає набір посилань на відповідні статті і ширшу інформацію про ICD-10 коди, пов’язані з названою предметною галуззю, наприклад, Z90.1 — acquired absence of breast; I97.2 — postmastectomy lymphoedema syndrome; Z85.3 — personal history of malignant neoplasm of breast; Z80.3 — family history of malignant neoplasm of breast; R92 — abnormal findings on diagnostic imaging of breast.

Розглянемо інший приклад на базі запиту ICD-9 коду: “What is ICD-9 code for breast cancer?”. У контекстній онтології немає коду ICD-9 для цього випадку. Тому отримано стандартну відповідь “Sorry, but I don’t know about ICD-9 code for breast cancer”. Своєю чергою ChatGPT повертає повідомлення: “The ICD-9 code for breast cancer is 174.9. Please note that ICD-9 codes have been largely replaced by ICD-10 codes for medical coding and classification”. Отже, особливість запропонованого підходу полягає в тому, що ChatGPT формує відповіді, спираючись саме на передані контексти й інформацію, представлену (або відсутню) саме в них.

ВИСНОВКИ

У результаті виконаного дослідження розроблено продуктивну тріаду, яка охоплює три ключові компоненти: методологічні засади використання онтолого-керованих структурованих підказок у метанавчанні ChatGPT, нову інформаційну технологію та комплексний сервіс MedRehabBot. Останній є новітньою інтерактивною інформаційною системою підтримки лікаря фізичної і (теле) реабілітаційної медицини, студентів спеціальності ФРМ та пацієнтів, створеною у результаті розвинення та вдосконаленням попередніх досліджень авторів [7].

Розроблено метод підвищення ефективності взаємодії онтолінгвістичних та нейромережевих засобів, який включає метанавчання для створення та налашту-

вання структурованих підказок і контекстної онтології, а також послідовність операцій для виявлення інтенцій, ідентифікації іменованих сутностей, вибору контекстів і формування остаточної підказки та відповіді на основі інформації та висновків, які будуть надані в контекстах. Ключовою особливістю цього методу є його здатність надавати ChatGPT специфічну інформацію за допомогою вибраних контекстів, що може розширити його знання в конкретних предметних галузях і поліпшити його роботу з даними на різних мовах.

Зауважимо, що запропоновані методологічні засади можна застосовувати не тільки до системи діалогових сервісів ChatGPT та GPT-моделей від OpenAI, використаних у дослідженні, але й до інших систем чат-ботів, що ґрунтуються на LLM-моделях (зокрема Google Bard). Основоположні принципи та методи метанавчання, формування структурованих підказок та пошуку інформації на основі онтології можна адаптувати та використати в поєднанні з різними LLM-системами. Ця можливість увиразнює потенційну універсальність і масштабованість запропонованого підходу на різних платформах чат-ботів, що дає змогу ширше застосовувати його в галузі оброблення природної мови та діалогових систем.

Онтологія в інформаційній системі MedRehabBot забезпечує такі ефекти:

- структуровані знання;
- семантичне розуміння;
- підвищена точність;
- розширення та адаптивність знань до конкретного домену;
- інтероперабельність;
- знання-керовані підказки;
- підтримка метанавчання.

Запропонований підхід слугує прототипом для розроблення більш досконалого діалогу користувача та системи в цілому. Для того, щоб підвищити продуктивність системи, заплановано здійснити кілька удосконалень для структурованих підказок у форматі JSON. Додаткові вхідні дані, а саме інтенції, виявлені у початковому повідомленні та вибраних контекстах, будуть додані до підказок. За очікуваннями авторів ці вдосконалення дадуть змогу покращити значення критерію точності системи.

Напрямом майбутніх досліджень є інтеграція запропонованого комплексного методу підвищення ефективності взаємодії онтолінгвістичних та нейромережевих засобів з медичними інформаційними системами реабілітаційного профілю та апаратної підтримки в діалогових системах з використанням для її реалізації логічних апаратних технологій [34].

Програмна реалізація системи MedRehabBot (у вигляді застосунку для системи обміну повідомленнями «Телеграм»), метаонтологія, онтологія контекстів, SPARQL запити до метаонтології, зразки структурованих JSON підказок для ChatGPT, тестові запитання та результати, є у відкритому доступі в репозиторії GitHub за таким посиланням: https://github.com/knowledge-ukraine/MedRehabBot.

Датасет “EBSCO articles dataset (domain knowledge: rehabilitation medicine) + JSON of every article” доступний на платформі Zenodo за посиланням [33].

Демо-версія програмної реалізації MedRehabBot доступна для вивчення та оцінювання у вигляді застосунку для системи обміну повідомленнями «Телеграм»: https://t.me/MedicalRehabBot.

Детальну інформацію про використання та доступ до системи MedRehabBot можна отримати за запитом до авторського колективу.

СПИСОК ЛІТЕРАТУРИ

1. Palagin A.V. Transdisciplinarity problems and the role of informatics. *Cybernetics and Systems Analysis*. 2013. Vol. 49, N 5. P. 643–651 (2013). https://doi.org/10.1007/s10559-013-9551-y.
2. Palagin O., Petrenko M., Kryvyi S., Boyko M., Malakhov K. Ontology-Driven Processing of Transdisciplinary Domain Knowledge. Iowa State University Digital Press, 2023. 189 p. https://doi.org/10.31274/isudp.2023.140.

3. Palagin A.V., Petrenko N.G. Methodological foundations for development, formation and IT-support of trandisciplinary research. *J. Automat. Inf. Sci.* 2018. Vol. 50, Iss.10. P. 1–17. https://doi.org/10.1615/JAutomatInfScien.v50.i10.10.
4. OpenAI: GPT-4 Technical Report. 2023. https://doi.org/10.48550/arXiv.2303.08774.
5. Kublik S., Saboo S. GPT-3: The Ultimate Guide To Building NLP Products With OpenAI API. Packt Publishing, 2023. 150 p.
6. Rothman D., Gulli A. Transformers for Natural Language Processing: Build, train, and fine-tune deep neural network architectures for NLP with Python, Hugging Face, and OpenAI's GPT-3, ChatGPT, and GPT-4. Birmingham Mumbai: Packt Publishing, 2022. 602 p.
7. Palagin O., Kaverinskiy V., Litvin A., Malakhov K. OntoChatGPT information system: ontology-driven structured prompts for ChatGPT meta-learning. *International Journal of Computing.* 2023. Vol. 22, Iss. 2. P. 170–183. https://doi.org/10.47839/ijc.22.2.3086.
8. Malakhov K.S. Letter to the editor – update from Ukraine: Development of the cloud-based platform for patient-centered telerehabilitation of oncology patients with mathematical-related modeling. *Int. J. Telerehab.* 2023. Vol. 15, N 1. https://doi.org/10.5195/ijt.2023.6562.
9. Romaniv S.V., Palaniza Yu.B., Vakulenko D.V., Galaychuk I.Y. The method of using fractal analysis for metastatic nodules diagnostics on computer tomographic images of lungs. In: Horizons in Cancer Research. Watanabe J.S. (Ed.) Vol. 85. P. 231–247. Nova Science Publishers, Inc., 2023. 265 p.
10. Vakulenko D., Vakulenko L., Zaspa H., Lupenko S., Stetsyuk P., Stovba V. Components of Oranta-AO software expert system for innovative application of blood pressure monitors. *Journal of Reliable Intelligent Environments.* 2023. Vol. 9, Iss. 1. P. 41–56. https://doi.org/10.1007/s40860-022-00191-4.
11. Palagin O.V., Malakhov K.S., Velychko V.Yu., Semykopna T.V. Hybrid e-rehabilitation services: SMART-system for remote support of rehabilitation activities and services. *Int. J. Telerehab.* 2022. https://doi.org/10.5195/ijt.2022.6480.
12. OpenAI: OpenAI API Reference. URL: https://platform.openai.com/docs/api-reference (Last accessed: 01.06.2023).
13. Kondrashchenko I. First steps in LangChain: The ultimate guide for beginners (part 1). URL: https://medium.com/@iryna230520/first-steps-in-langchain-the-ultimate-guide-for-beginners-part-1-2baf5a4e1b81 (Last accessed: 07/09/2023).
14. Kondrashchenko I. First steps in LangChain: The ultimate guide for beginners (part 2). URL: https://medium.com/@iryna230520/first-steps-in-langchain-the-ultimate-guide-for-beginners-part-2-d17a2f057f43 (Last accessed: 07/09/2023).
15. Amri A.E. OpenAI GPT for Python developers: The art and science of developing intelligent apps with OpenAI GPT-3, DALL·E 2, CLIP, and Whisper - Suitable for learners of all levels. FAUN, 2023. 378 p.
16. GPT 4 is Smarter than You Think: Introducing SmartGPT. 2023. URL: https://www.youtube.com/watch?v=wVzuvf9D9BU.
17. Gill S.S., Xu M., Patros P., Wu H., Kaur R., Kaur K., Fuller S., Singh M., Arora P., Parlikad A.K., Stankovski V., Abraham A., Ghosh S.K., Lutfiyya H., Kanhere S.S., Bahsoon R., Rana O., Dustdar S., Sakellariou R., Uhlig S., Buyya R. Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots. *Internet of Things and Cyber-Physical Systems.* 2024. Vol. 4. P. 19–23. https://doi.org/10.1016/j.iotcps.2023.06.002.
18. Hebenstreit K., Praas R., Kiesewetter L.P., Samwald M. An automatically discovered chain-of-thought prompt generalizes to novel models and datasets. 2023. https://doi.org/10.48550/arXiv.2305.02897.
19. Wei J., Wang X., Schuurmans D., Bosma M., Ichter B., Xia F., Chi E., Le Q., Zhou D. Chain-of-thought prompting elicits reasoning in large language models. 2023. https://doi.org/10.48550/arXiv.2201.11903.
20. Bhatt P., Sethi A., Tasgaonkar V., Shroff J., Pendharkar I., Desai A., Sinha P., Deshpande A., Joshi G., Rahate A., Jain P., Walambe R., Kotecha K., Jain N.K. Machine learning for cognitive behavioral analysis: datasets, methods, paradigms, and research directions. *Brain Informatics.* 2023. Vol. 10. Article number 18. https://doi.org/10.1186/s40708-023-00196-6.
21. Zhou Y., Muresanu A.I., Han Z., Paster K., Pitis S., Chan H., Ba J. Large language models are human-level prompt engineers. 2023. https://doi.org/10.48550/arXiv.2211.01910.

22. Khan A. Knowledge graphs querying. *ACM SIGMOD Record.* 2023. Vol. 52, N 2. P. 18–29. https://doi.org/10.1145/3615952.3615956.
23. Palagin O.V., Petrenko M.G., Velychko V.Yu., Malakhov K.S. Development of formal models, algorithms, procedures, engineering and functioning of the software system "Instrumental complex for ontological engineering purpose." *Proc. 9th International Conference of Programming UkrPROG* (20-22 May 2014, Kyiv, Ukraine). Kyiv, 2014. *CEUR Workshop Proceedings.* 2018. Vol. 1843. P. 221–232.
24. Palagin O.V., Velychko V.Yu., Malakhov K.S., Shchurov O.S. Research and development workstation environment: The new class of current research information systems. *Proc. 11th International Conference of Programming UkrPROG 2018* (22–24 May 2018, Kyiv, Ukraine). Kyiv, 2018. *CEUR Workshop Proceedings.* 2018. Vol. 2139. P. 225–269. URL: https://ceur-ws.org/Vol-2139/255-269.pdf.
25. Petrie C.J.: Formalization of Web Service Composition. In: Web Service Composition. Petrie C.J. (Ed.). Cham: Springer International Publishing, 2016. P. 41–53. https://doi.org/10.1007/978-3-319-32833-1_3.
26. Markov K., Vanhoof K., Mitov I., Depaire B., Ivanova K., Velychko V., Gladun V. Intelligent data processing based on multi-dimensional numbered memory structures. In: Diagnostic Test Approaches to Machine Learning and Commonsense Reasoning Systems. Naidenova X., Ignatov D.I. (Eds.). 2012. P. 156–184. https://doi.org/10.4018/978-1-4666-1900-5.ch007.
27. Richards M., Ford N. Fundamentals of Software Architecture: An Engineering Approach. Sebastopol, CA: O'Reilly Media, Inc, 2020. 410 p.
28. Moghaddam S.R., Honey C.J. Boosting Theory-of-Mind performance in large language models via prompting. 2023. http://arxiv.org/abs/2304.11490, https://doi.org/10.48550/arXiv.2304.11490.
29. Ni S., Kao H.-Y. KPT++: Refined knowledgeable prompt tuning for few-shot text classification. *Knowledge-Based Systems.* 2023. Vol. 274. Article number 110647. https://doi.org/10.1016/j.knosys.2023.110647.
30. Reisig W. Understanding Petri Nets. Heidelberg; Berlin: Springer Verlag, 2013. XXVII, 230 p. https://doi.org/10.1007/978-3-642-33278-4.
31. Reisig W. The basic concepts. In: Understanding Petri Nets: Modeling Techniques, Analysis Methods, Case Studies. Reisig W. (Ed.). Berlin; Heidelberg: Springer, 2013. P. 13–24. https://doi.org/10.1007/978-3-642-33278-4_2.
32. MedRehabBot. 2023.URL: https://github.com/knowledge-ukraine/MedRehabBot.
33. Malakhov K., Vakulenko D., Kaverinsky V. EBSCO articles dataset (domain knowledge: rehabilitation medicine) + JSON of every article. 2023. URL: https://zenodo.org/record/8308214, https://doi.org/10.5281/ZENODO.8308214.
34. Kryvyi S., Grinenko O., Opanasenko V. Logical approach to the research of properties of software engineering ecosystem. *Proc. 2020 IEEE 11th International Conference on Dependable Systems, Services and Technologies (DESSERT)* (14–18 May 2020, Kyiv, Ukraine). Kyiv, 2020. https://doi.org/10.1109/DESSERT50317.2020.9125033.

O. Palagin, V. Kaverinskiy, K. Malakhov, M. Petrenko

FUNDAMENTALS OF THE INTEGRATED USE OF NEURAL NETWORK AND ONTOLINGUISTIC PARADIGMS: A COMPREHENSIVE APPROACH

Abstract. This article presents an integrated approach that combines neural-network and ontolinguistic paradigms. The method encompasses methodological underpinnings, information technology, and the MedRehabBot system. Collectively, they embody the core principles of meta-learning and structured prompts, ultimately enhancing the efficiency of information system interaction with Chatbots and information retrieval rooted in ontologies. The method also offers the flexibility to adapt the MedRehabBot system for utilization within different Large Language Model (LLM) systems.

Keywords: transdisciplinary scientific research, ontology, ontological engineering, ontology-driven information system, LLM-system, ChatGPT, MedRehabBot, prompt engineering, Chatbot.

Надійшла до редакції 12.09.2023