



УДК 004.318

КОМП'ЮТЕРНІ СИСТЕМИ ЗНАНЬ: ОСНОВИ ПОДАННЯ ТА ОБРОБЛЕННЯ ПРЕДМЕТНО-ОРІЄНТОВАНИХ ЗНАНЬ

Петренко М.Г.¹, Бойко М.О.², Малахов К.С.³^{1,2,3}Інститут кібернетики імені В.М. Глушкова Національної Академії Наук України, м. Київ, УкраїнаORCID: ¹<http://orcid.org/0000-0001-6440-0706>, ²<http://orcid.org/0000-0003-1723-5765>,³<https://orcid.org/0000-0003-3223-9844>E-mail: ¹petrng@ukr.net, ²xeldag@ukr.net, ³malakhovks@nas.gov.ua

Copyright © 2024 by author and the journal “Automation of technological and business – processes”.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0>

DOI: 10.15673/atbp.v16i1.2771

Анотація. В статті досліджується еволюція онтолого-керованого оброблення природномовної інформації. Шляхом поглибленого дослідження запропоновано нову архітектуру лінгвістичного процесора, яка унікальним чином інтегрує лінгвістичну та онтологічну парадигми під час семантичного аналізу. Відходячи від традиційних методологій, такий підхід демонструє глибоке злиття методів видобування та представлення знань. Центральним моментом дослідження є розробка онтолого-керованої інформаційної системи, побудованої з вбудованим акцентом на саморозвиток і адаптацію. Особливістю системи є ефективне оперування елементарними знаннями у поєднанні з її динамічною здатністю сприяти розвитку інноваційних концептів і відношень. Особлива увага приділяється застосуванню системи в обробці наукової інформації, що свідчить про її потенціал у революційному перетворенні наукомістких додатків у наукових галузях. Основними зусиллями в статті є прагнення прокласти шлях до більш інтуїтивно зрозумілих, точних і розширених онтолого-орієнтованих інструментів у сфері видобування та представлення знань.

Abstract. This article delves into the evolving frontier of ontology-driven natural language information processing. Through an in-depth examination, we put forth a novel linguistic processor architecture, uniquely integrating linguistic and ontological paradigms during semantic analysis. Distancing from conventional methodologies, our approach showcases a profound merger of knowledge extraction and representation techniques. A central highlight of our research is the development of an ontology-driven information system, architected with an innate emphasis on self-enhancement and adaptability. The system's salient capability lies in its adept handling of elementary knowledge, combined with its dynamic aptitude to foster innovative concepts and relationships. A particular focus is accorded to the system's application in scientific information processing, signifying its potential in revolutionizing knowledge-based applications within scientific domains. Through our endeavors, we aim to pave the way for more intuitive, precise, and expansive ontology-driven tools in the realm of knowledge extraction and representation.

Ключові слова: Трансдисциплінарні дослідження, Онтологічний інжиніринг, Елементарні сенси, Подання знань, Знання здорового глузду, Глибокий штучний інтелект, Наукова модель світу

Keywords: Transdisciplinary research, Ontology engineering, Elementary Sense, Knowledge representation, Commonsense knowledge, Deep artificial intelligence, Scientific model of the World

Вступ

Розвиток динамічних комп'ютеризованих систем знань і систем глибокого штучного інтелекту (ГШІ) тісно пов'язані між собою, оскільки мають спільні фундаментальні корені. Основною метою цих систем є формулювання цілісної наукової моделі світу (НМС) та її ефективне використання. Перші системи спрямовані на створення глобальної мережі трансдисциплінарних знань, спрямованих на вирішення складних проблем людства, в той час як друга система націлена на наслідування людського здорового глузду і всеосяжного світогляду, притаманного людині.

Аналіз літературних даних і постановка проблеми

Прагнення досягти штучного інтелекту (ШІ) на рівні людини залишається головним завданням в дослідженнях ШІ. Таке досягнення може призвести до революційних інновацій з далекосяжними наслідками для людства. Наразі, однак, роль ШІ переважно полягає в тому, що він є практичним інструментом в конкретних, хоча й обмежених сферах застосування.



У міру розвитку систем штучного інтелекту, пристосованих до вирішення реальних завдань, вони, безперечно, отримуватимуть користь від сучасних наукових досягнень. Очікується, що в найближчому майбутньому потенціал штучного інтелекту буде продемонстрований переважно через сплеск спеціалізованих додатків. Ця траєкторія вже очевидна в різних секторах, від промисловості до економіки та соціальних аспектів. Широкі перспективи ШІ очевидні, оскільки він проникає в різні сфери, від наукових досліджень до сприяння великим технологічним інноваціям [1–3].

Тут доречно згадати погляд Девіда Ферруччі на ГШІ. Генеральний директор і засновник стартапу Elemental Cognition, що займається розробкою штучного інтелекту, постулює, що оволодіння розумінням природної мови є ключовим для досягнення універсального інтелекту. На відміну від підходів, подібних до DeepMind, які досліджують фізіологію мозку, позиція Ферруччі полягає у створенні системи, що перевершує людські можливості в розумінні природної мови та логічних міркуваннях. Він однозначно стверджує, що фундаментальні елементи для створення такого інтелекту вже доступні [4].

Щоб досягти цієї мети, Elemental Cognition розробляє гібридну систему. Ця модель об'єднує глибокі нейронні мережі та різноманітні стратегії машинного навчання з програмними модулями, відточеними для логіки та міркувань, які базуються на традиційних парадигмах програмування. Цікаво, що сучасний дослідницький клімат сприяє об'єднанню, а не розділенню символістських і коннекціоністських систем. Такий інтегрований напрям досліджень отримав назву "нейросимволічний ШІ" [3, 4].

Розуміння причинно-наслідкових зв'язків має вирішальне значення для розвитку креативності та пошуку альтернативних рішень. На відміну від алгоритмів навчання з підкріпленням у нейронних мережах, які вимагають повторних невдач для успішної розробки стратегій, люди за своєю природою практикують ментальні сценарії. Цей когнітивний акт дозволяє нам передбачати потенційні результати різних рішень, що ґрунтується на нашому внутрішньому розумінні причинно-наслідкових зв'язків. Це полегшує нам пошук відповідей на квінтесенційне "чому?". Опанування причинно-наслідкових зв'язків, особливо вміння ставити і вирішувати причинно-наслідкові питання, є невід'ємною частиною еволюції універсального машинного інтелекту [4].

Відмінною рисою людського інтелекту є здатність асимілювати інформацію з одного джерела і адаптувати її в різних сферах, що лежить в основі творчості та інновацій. Для того, щоб універсальний машинний інтелект був практично ефективним, він повинен виходити за рамки простого розуміння тексту. Його справжня перевага полягає в здатності використовувати свій запас знань для вирішення нових завдань. Вміння системи штучного інтелекту застосовувати знання в різноманітних і нових сценаріях цілком може стати визначальним критерієм для оцінки його інтелектуальної глибини [3].

Підтримуючи думку Д. Ферруччі, здатність екстраполювати інформацію з одного контексту і вміло застосовувати її в різних сценаріях є квінтесенцією для сприяння інноваціям [4].

З нашої точки зору, емуляція людських міркувань в ГШІ вимагає формулювання специфічної для домену НМС або, щонайменше, орієнтованої на дисципліну НМС. Побудова такої моделі має спиратися на онтологічну методологію, кульмінацією якої має стати Науково-онтологічна модель світу (НОМС). Подальший дискурс буде присвячений необхідності створення та використання НОМС, підкреслюючи ключову роль систематичного представлення онтологічних знань у відтворенні людських міркувань за допомогою ГШІ.

Опрацювання інформації та представлення знань на основі онтологій виникло в результаті пошуку стандартного протоколу для впорядкування знань у різноманітних сферах знань. Ця парадигма має на меті запропонувати уніфіковану схему та основні принципи для систематичного представлення, категоризації та взаємозв'язку знань, незалежно від галузі знань. Поява онтологічних стратегій уможливила ефективну побудову знаннево-орієнтованих систем і, що особливо важливо, заклала фундамент для трансдисциплінарної взаємодії та онтологічного інжинірингу у сфері сучасного ШІ [5–9].

Онтологічні компоненти та представлення знань

Онтологія, як формалізована структура для представлення знань, зазвичай визначається наступними чотирма компонентами:

- *Класи*. Символізують категорії або концепти в межах певної предметної області, пропонуючи засоби для групування об'єктів зі схожими атрибутами.
- *Властивості*. Окреслюють атрибути або асоціації класів та об'єктів, слугують для встановлення відношень за допомогою таких термінів, як "мати", "бути" або "бути частиною" та ін.
- *Індивіди*. Представляють собою матеріальні екземпляри класів, а також можуть розглядатися як окремі сутності, концепти або ситуації в межах предметної області.
- *Аксіоми або обмеження*. Це правила або логічні твердження, які диктують взаємозв'язки і поведінку в онтології, посилюючи її логічну узгодженість.

Ці компоненти лежать в основі онтологічного моделювання, пропонуючи надійну основу для структурованого представлення знань, специфічних для певної предметної області. Така структура не тільки допомагає людському розумінню, але й полегшує машинні міркування [10].

Онтологічні методології дають користувачам цілісний погляд на конкретні предмети або складні проекти. Використання онтологічних моделей дозволяє розмежовувати класи, сутності, функції та формальні теорії. Онтологічні інструменти підтримують створення аналітичних систем для дослідницьких і організаційних цілей,



які охоплюють функції від багатофакторного аналізу первинних даних до сприяння спільному прийняттю рішень. Крім того, онтології слугують як засіб маніпулювання, так і результат для технологій семантичного вебу.

Неоціненним інструментом онтологічного забезпечення є лінгвістично-онтологічна модель світу (ЛОМС). Задумана як лексикографічна система, ЛОМС є невід'ємною частиною загальної НМС і має ключове значення для систем, орієнтованих на розуміння об'єктів природної мови (ПМО) [10–12].

У цьому контексті ЛОМС діє як категоріальний каркас, забезпечуючи семантично збагачену базу для репозиторіїв знань у конкретних галузях. Вона також допомагає об'єднати різноманітні джерела знань. Об'єднуючи лінгвістичні та онтологічні компоненти, ЛОМС покращує розуміння, комунікацію та управління знаннями, сприяючи як галузевим, так і міждисциплінарним дослідженням.

Будь то людське мислення чи комп'ютерна система (КС), оброблення мовленнєвих чи текстових даних залежить від лінгвістичного процесора. У КС цей процесор має першорядне значення, він відповідає за розпізнавання і розуміння вхідних природномовних текстів, видобування ключових знань і представлення їх у логічному форматі.

Ці оброблені дані закладають основу для операцій, заснованих на знаннях, допомагаючи у вирішенні проблем, прийнятті рішень і виконанні різних пов'язаних з ними завдань. По суті, лінгвістичний процесор КС слугує каналом між людським мовленням і обчислювальними діями, видобуваючи і використовуючи знання для різноманітних застосувань.

Лінгвістичний процесор, апаратний чи програмний, розшифровує текстові дані, такі як документ, стаття, монографія або лінгвістичний корпус текстів (ЛКТ), за допомогою послідовних етапів лінгвістичної обробки. Зазвичай це включає граматичний, морфологічний, синтаксичний та поверхнево-семантичний аналіз, кожен з яких сприяє розумінню структури та семантики тексту [13–15].

Після обробки лінгвістичним процесором отримана інформаційна структура підготовлена для подальшого інтенсивного семантичного аналізу в екстралінгвістичній підсистемі. Тут першочерговим завданням є структурування концептів. По суті, це автоматизує вилучення знань з ПМО з метою прагматичної інтерпретації цих знань, що відображає людське розуміння та реакцію [10].

Архітектура системи аналізу і розуміння (САР), в основі якої лежить ЛОМС, показана на рис. 1 [10]. Первинним сховищем даних САР є ЛКТ, пов'язаний з конкретною областю знань або масивом наукових праць. Лінгвістичний корпус, скінченний набір текстів, представлений числом k , яке позначає загальну кількість текстів у ЛКТ. Ці тексти обробляються послідовно, спочатку проходячи через підсистему графематичного аналізу.

Коли текст обробляється алгоритмом лінгвістичного аналізу, він зазнає перетворень у графематичній, морфологічній, синтаксичній та семантичній структурах, кожна з яких має свої моделі та інструменти представлення. Вичерпний опис цієї процедури наведено в [10]. Відмінність цього методу від традиційного семантичного аналізу полягає в інтеграції мовно-онтологічної моделі світу в семантичний аналіз. Ця модель виходить за межі словникових семантичних даних, вбираючи в себе багаторівневі моделі як загальних, так і специфічних семантичних структур, що спостерігаються в простих реченнях.

Результатом обробки САР є таблиця шаблонів речень тексту. Це сховище охоплює інформаційну структуру всього тексту і слугує вихідними даними для підсистеми екстралінгвістичного опрацювання тексту. У межах цієї підсистеми виконується формально-логічне подання речень і загального тексту, зазвичай перетворюючи текст у відповідну формальну теорію першого порядку. Загальноприйнята методика включає проміжну фазу, на якій дані переробляються в модифіковані концептуальні граfi, після чого відбувається перехід до логіки предикатів першого порядку [7].

Формулювання ЛОМС вичерпно описано в [11], а в [12] подано детальну інформацію про розробку НОМС. Обидві роботи є незамінними джерелами, що проливають світло на методологію створення цих важливих компонентів системи представлення знань і семантичного аналізу.

Часова та просторова динаміка розвитку знань

Еволюція знань у будь-якій сфері, включаючи експансивну сферу науки, проявляється у вигляді параметрів, які моделюються у часі та просторі. Історично склалося так, що темпи накопичення знань на початкових етапах розвитку науки, як в цілому, так і в межах окремих спеціалізацій, меркнуть у порівнянні з нинішніми експоненціальними темпами зростання знань. Хоча кількість знань та їх інформаційне представлення взаємопов'язані, вони не є синонімами. Їхній зв'язок на будь-якому часовому відрізку значною мірою залежить від рівноваги між вербальною та формалізованою репрезентаціями. Останні, які включають аналітичні, табличні та графічні репрезентації, є більш стислими порівняно з вербальними і краще піддаються операційній обробці. З перебігом часу можна виділити три основні еволюційні траєкторії [12]:

1. Збільшення загального обсягу знань та інформації.
2. Розширення формалізованого сегменту знань.
3. Витіснення застарілих знань сучасним розумінням.

Характерною рисою еволюції загального знання є наявність дихотомічних тенденцій: диференціація наукових дисциплін, за якою слідує інтеграція. Історично першість належала диференціації, що призвело до генезису спеціалізованих дисциплін. Сьогодні диференціація співіснує з інтеграцією, що символізується появою мультидисциплінарних дослідницьких структур і зародженням трансдисциплінарних конвергентних кластерів. Ці кластери відіграють важливу роль, виступаючи в якості середовища для міждисциплінарної співпраці, тим



самим сприяючи узгодженим рішенням багатограних проблем і просуваючи людський епістемологічний прогрес [10].

Онтологічні знаннево-орієнтовані системи в інформатиці

Розвиток моделей і методологій для систем, заснованих на знаннях, залишається одним з основних напрямків діяльності в галузі комп'ютерних наук. Генезис онтолого-керованих систем знань (ОКСЗ), нерозривно пов'язаний з розвитком теоретичних засад і підходів до проектування. Невід'ємні елементи включають:

1. *Узагальнена архітектура та структура системи.* Це передбачає формулювання основоположних принципів, що регулюють архітектуру системи, забезпечуючи цілісну структуру для представлення та обробки знань.
2. *Формальні моделі представлення знань.* Цей аспект наголошує на структурованих моделях представлення знань, що сприяють ефективному управлінню та пошуку даних.
3. *Алгоритми оброблення знань.* Це стосується створення алгоритмів, що полегшують роботу зі структурованими знаннями, підтримуючи широкий спектр операцій, орієнтованих на знання.

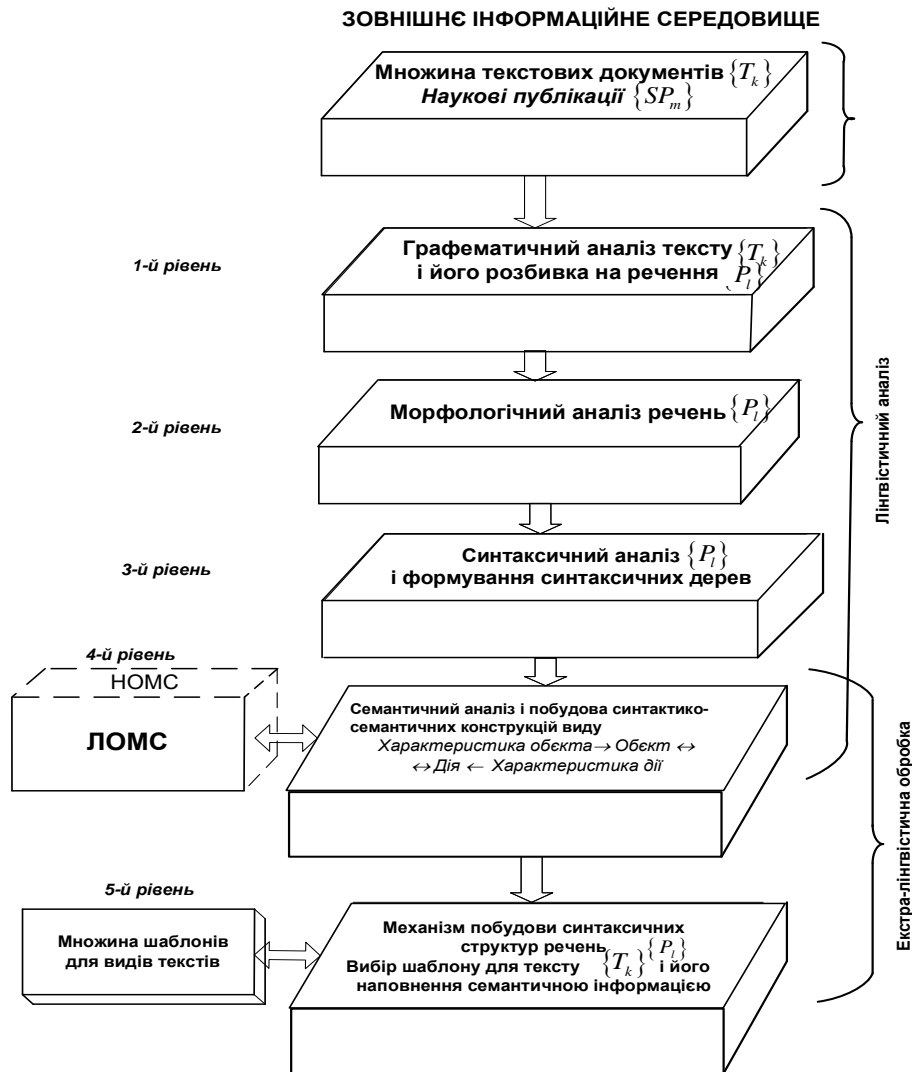


Рис. 1 Архітектура системи аналізу та розуміння
Fig. 1 The Analytical and Understanding System architecture

У сукупності ці розробки посилюють значення онтологічних знань у системах, заснованих на знаннях, особливо при вирішенні таких складних завдань, як розробка НОМС. Як підкреслювалося раніше, НОМС є невід'ємною частиною глибокого ШІ і зростаючої галузі "нейросимволічного" ШІ. Обидва напрямки прагнуть до об'єднання символічної та коннекціоністської парадигм ШІ, прагнучи до більш ємного втілення ШІ.

Розвиток архітектури інтелектуальної комп'ютерної системи, яка еволюціонує, можна розглядати через подвійну призму, що охоплює як зовнішні (орієнтовані на користувача), так і внутрішні виміри. Гармонійна координація між цими вимірами є квінтесенцією для максимізації ефективності ОКСЗ [12].

Архітектура ефективних онтолого-керованих комп'ютерних систем вимагає асиміляції сучасних областей комп'ютерних наук, включаючи штучний інтелект, оброблення знань і прагматичну модель лінгвістичної свідомості. Концептуалізована як продуктивна послідовність, їхня операційна синергія є дзеркальним відображенням: "Вхідний сигнал → Система знань → Реакція".



Розбудована ОКСЗ функціонує із заздалегідь визначеними цілями, що охоплюють як довгострокове бачення, так і безпосередні цілі. Таке оперативне узгодження моделюється взаємодією із зовнішнім інформаційним середовищем на основі зворотного зв'язку. Центральним елементом операційного принципу ОКСЗ є система знань, що розглядається як складна підсистема, яка синергічно взаємодіє із сукупністю підсистем знань у конкретних галузях [10, 16]. Ця багатогранна взаємодія дає змогу ОКСЗ ефективно накопичувати, обробляти і застосовувати знання з різних галузей, що відповідають її оперативним цілям.

За своєю суттю, ОКСЗ спирається на систему знань і спеціалізовані підсистеми знань, динамічно налаштовуючи свої реакції відповідно до загальних цілей і специфічних вимог зовнішньої інформаційної екосистеми. Ця адаптивність забезпечує постійну ефективність системи у виконанні її чітко окреслених завдань.

Архітектура ОКСЗ, що зображена на рис. 2, висвітлює ключовий компонент: механізм саморозвитку бази знань (БЗ), яка має безпосереднє відношення до визначеної предметної області. Ця функція саморозвитку забезпечує безперервну адаптацію та зростання БЗ, резонуючи з мінливими формами та розумінням предметної області.

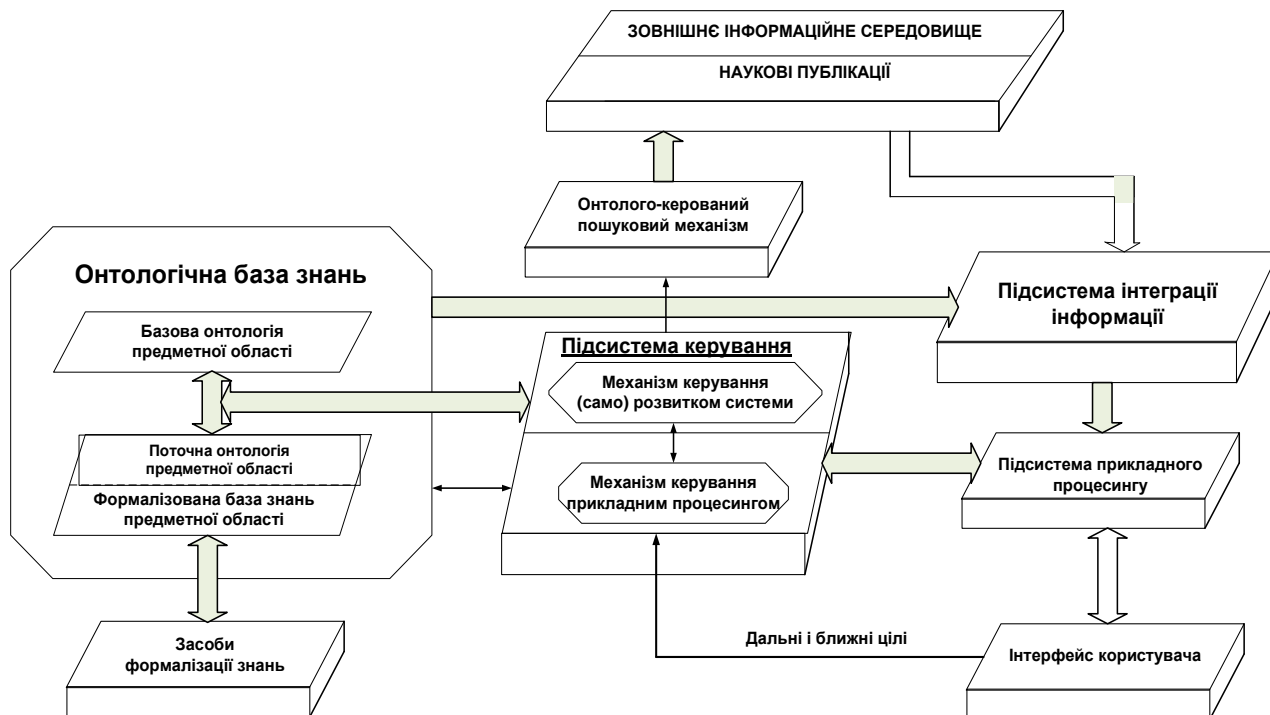


Рис. 2 Архітектура онтологічної системи, заснованої на знаннях

Fig. 2 The Ontological Knowledge-Based System architecture

Центральне місце в цьому механізмі саморозвитку займають онтологічні механізми управління двома фундаментальними процесами: дослідженням зовнішнього інформаційного простору та формуванням формалізованої бази знань. Еволюція БЗ реалізується двома основними каналами [11, 12].

1. Отримання даних із зовнішнього інформаційного середовища (переважно з Інтернету).
2. Генерація знань на основі умовиведень.

Обидва аспекти еволюції формалізованої БЗ в онтологічних системах, заснованих на знаннях, демонструють глибоку взаємодію. Генерація нових знань через канали виведення залежить від надходження сучасних даних, переважно з цифрового простору Інтернету. Цей безперервний потік даних забезпечує систему фундаментальними знаннями, що мають вирішальне значення для логічних виведень, інференційних процесів й активного формування нових сфер знань.

Крім того, онтологічне наповнення бази знань новими знаннями, що потенційно можуть вводити нові концепти, є невід'ємною частиною еволюції бази знань. Це передбачає ідентифікацію та асиміляцію нових концептів і динаміки із зовнішніх інформаційних сфер в онтології системи. Такий підхід динамічної інтеграції підкреслює важливість циклічної методології отримання знань, що забезпечує постійну адаптацію та врахування онтологічних знань, які еволюціонують [17].

Формалізація наукових публікацій для розвитку знань

Наукові публікації (НП) слугують важливим засобом синтезу нових знань, особливо з огляду на їхній унікальний науковий нарративний стиль і структурованість подання. За своєю суттю НП мають чітко визначену синтаксичну та семантичну структуру. Їхній шаблонний формат дозволяє автоматизувати формалізований опис змісту, що є ключовою характеристикою, висвітленою в [18].

У режимі розробки ОКСЗ обробляє запити користувачів, спрямовані до бази даних Natural Language Processing (NLP), як це детально описано в [18].



Онтологічна система (ОнС) для обробки НП представлена на рис. 3, де наведено схему основних компонентів та функцій.

Розробка ОнС характеризується двома різними операційними моделями:

1. Механізм логічного виведення. Цей режим використовує інструменти міркування, такі як логічний різонер Pellet, розширюючи знання за допомогою методів виведення. Він створює нові онтологічні зв'язки та розкриває нюанси відношень у вже існуючій базі знань.
2. Алгоритми обробки елементарних сенсів (ЕС). Цей режим використовує спеціалізовані алгоритми, пристосовані для аналізу елементарних сенсів. Робота "електронного колайдера" [18] є прикладом такого підходу. Шляхом розчленування основних лінгвістичних і семантичних одиниць знання видобуваються безпосередньо з аналізу елементарних сенсів.
3. Оброблення природномовної інформації на основі формул. Використання визнаних формул, таких як формула Брукса, підкреслює поведінково-орієнтовані стратегії для NL розуміння.

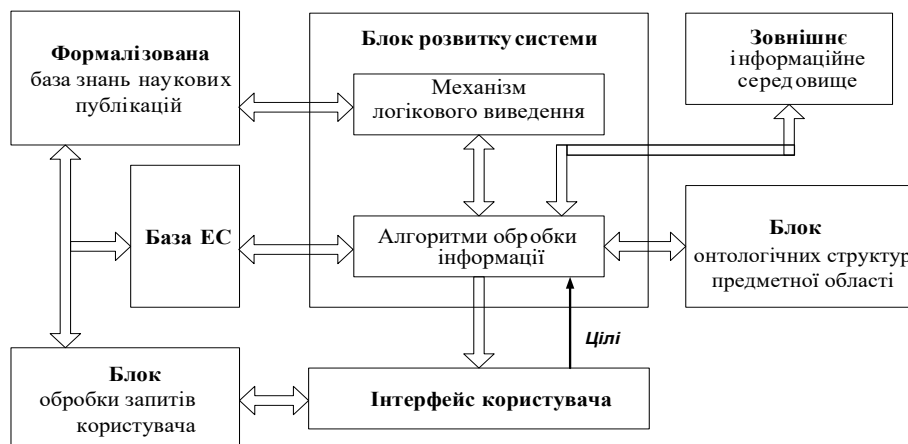


Рис. 3 Архітектура онтологічної системи обробки НП
Fig 3 The architecture of Ontology System for processing SP

Вибір підходу може залежати від стратегічних цілей системи, особливостей предметної області та конкретних цілей розвитку знань. Завдяки об'єднанню дедуктивних, алгоритмічних і формальних методів, ОнС демонструє універсальність у вирішенні завдань розвитку знань, поступово відточуючи свою ефективність.

Співвідношення між "Елементарний Сенс", "Здоровий Глузд" і "Міркування Здорового Глузду" в онтологерованих комп'ютерних системах і оброблення предметних знань

Визначення

1. Поняття елементарного сенсу, введене в [18]: просте двоскладове висловлювання, яке містить суб'єкт (підмет), предикат (присудок) і прямий об'єкт (пряме доповнення). Ці компоненти відповідають суб'єкту, предикату та об'єкту RDF-трійки Semantic Web.
2. Поняття "Знання здорового глузду", введене в [19]: універсальне розуміння світу, яким володіє більшість людей, що включає в себе очевидні умовиведення, а також охоплює широкий спектр галузей від природної мови до високорівневого передбачення. Воно формує фундаментальне ядро людських знань та інтелекту.
3. Поняття "міркування здорового глузду", введене в [19, 20]: акт виконання виведення на множині об'єктно-рівневої інформації з використанням на наборі інформації об'єктного рівня бази знань і її керуючого блоку. Воно передбачає отримання нових знань на основі наявних знань про загальні сценарії розвитку подій у світі.

Подібність.

- *Підстави для подання знань.* Як "Елементарні сенси", так і "Здоровий глузд" забезпечують фундаментальні структури для представлення інформації. Елементарний сенс пропонує стислий, структурований формат, подібний до RDF-трійок Semantic Web, тоді як "Здоровий глузд" надає всебічні знання про світ.
- *Виведення.* Як Знання здорового глузду, так і Міркування здорового глузду включають в себе процес формування умовиведень. У той час як перший компонент надає фундаментальні знання, другий компонент – це власне процес отримання нових знань на основі існуючих. Крім того, в елементарному сенсі виведення можна виконувати за допомогою SPARQL-запитів до RDF-трійок, таким чином отримуючи точні результати зі структурованого представлення даних.
- *Динамічне розуміння.* Всі три поняття підкреслюють динамічну природу знань. Елементарні сенси пропонують структуровані репрезентації, Знання здорового глузду охоплюють постійно зростаюче людське розуміння, а Міркування здорового глузду включає в себе безперервний процес отримання нових знань.

**Відмінності.**

Деталізація. Елементарні сенси фокусуються на точному представленні інформації у вигляді простих двоскладових висловлювань, узгоджених з RDF-триjkами, в той час як Здоровий глузд ширший і охоплює широкий спектр загальних областей знань.

Призначення. Основна мета Елементарного сенсу – запропонувати структуровані репрезентації для легкого опрацювання. На противагу цьому, Знання здорового глузду слугують фундаментальною базою для розуміння світу, а Міркування здорового глузду спрямовані на виведення нових ідей на основі цих фундаментальних знань.

Процес проти даних. Елементарний сенс стосується представлення та зберігання даних, що тісно пов'язано зі структурами семантичного вебу. З іншого боку, Знання здорового глузду стосуються самих даних, а Міркування здорового глузду орієнтовані на процес і зосереджені на акті виведення.

Видобування нових знань з існуючих знань.

Роль елементарного сенсу. У онтолого-керованих комп'ютерних системах Елементарні сенси відіграють вирішальну роль у спрощенні та структуруванні даних. Організуючи інформацію у форматі, аналогічному RDF-триjкам, вони спрощують видобування знань, прокладаючи шлях до більш глибокого семантичного та онтологічного аналізу.

Включення здорового глузду. Знання здорового глузду діють як фоновий резервуар під час процесу оброблення знань. Коли системи стикаються з неоднозначними або неповними даними, ці знання можуть бути використані для заповнення прогалів або визначення відсутніх компонентів. Широка сфера застосування Здорового глузду забезпечує всебічне розуміння, навіть за відсутності чіткої інформації.

Міркування та еволюція. Завдяки міркуванням на основі здорового глузду системи можуть виводити нову інформацію або зв'язки з наявних знань. Застосування виведення на основі бази знань, особливо в поєднанні зі структурованими даними з Елементарних сенсів і запитів SPARQL на RDF-триjках, збільшує потенціал для відкриття нових закономірностей, взаємозв'язків або знань. Таке динамічне міркування є ключовим для еволюції та самовдосконалення онтолого-керованих комп'ютерних систем.

Елементарні сенси, знання здорового глузду і міркування здорового глузду нерозривно пов'язані у сфері онтолого-керованих комп'ютерних систем, і оброблення предметно-орієнтованих знань. Разом вони сприяють точному представленню, всебічному розумінню і динамічному розширенню знань в системі.

ЕС-и є ключовими для аналізу НП. Розглянемо докладніше:

1. **Репозиторій наукових публікацій.** Це велика кількість структурованих наукових статей, повних назв, бібліографічних даних, текстів (структурованих у вигляді розділів, абзаців, речень, включаючи рисунки, формули і таблиці), анотацій, передмов, висновків і цитувань.
2. **Декомпозиція складних речень.** Складні структури речень розбиваються на простіші, в ідеалі прості двоскладні речення для спрощення аналізу.
3. **Представлення простих речень.** Прості двоскладні речення розкладаються на базову тріаду: підмет, присудок і пряме доповнення, аналогічно до Суб'єкт-Предикат-Об'єкт (SPO) або RDF-триjки, популярної у фреймворках семантичної павутини.

Сегментація складних конструкцій на елементарні сенси, виражені у вигляді RDF-триjок, забезпечує лаконічне представлення знань. Отже, це сприяє ефективному і точному аналізу та видобуванню нюансів знань з НП.

У середньому речення в тексті НП, зокрема українською мовою, може включати 3–4 ЕС. Маніпулювання цими елементарними сенсами в поєднанні з відповідними структурами онтологічних доменів знань відкриває шлях до потенційної генерації нових онтологічних концептів або відношень.

Для демонстрації процесу утворення ЕС:

“Теорія і практика створення та використання систем, заснованих на знаннях, – найбільш актуальний напрям Computer Science, що інтенсивно розвивається, який дозволяє підвищити ефективність створення і використання комп'ютерних технологій, прикладних систем та інструментальних засобів.”

Після декомпозиції на ЕС речення розпадається на основні компоненти: суб'єкт, предикат і об'єкт, а також контекстуальні коди предметної області знань (ПДО):

ЕС1 Computer Science має актуальний напрям, який інтенсивно розвивається

S – Computer Science P – мати O – актуальний напрям, який інтенсивно розвивається

Кнт ЕС1 – P7-5, P 9-5

ЕС2 Напрямок Computer Science є теорія і практика створення і використання систем

S – Напрямок Computer Science P – бути O – теорія і практика створення і використання систем

Кнт ЕС2 – P 7-5, P 9-5

ЕС3 Системи є системи обробки знань

S – Системи P – бути O – системи обробки знань

Кнт ЕС3 – P 7-5, P 9-3, P 9-5

ЕС4 Системи обробки знань дозволяють підвищити ефективність створення комп'ютерних технологій

S – Системи обробки знань P – дозволяють підвищити O – ефективність створення комп'ютерних технологій

Кнт ЕС4 – P7-5, P9-5

ЕС5 Системи обробки знань дозволяють підвищити ефективність створення прикладних систем



S – Системи обробки знань P – дозволяють підвищити O – ефективність створення прикладних систем
Кнт ЕС5 – P7-5, P10-6

ЕС6 Системи обробки знань дозволяють підвищити ефективність створення інструментальних засобів

S – Системи обробки знань P – дозволяють підвищити O – ефективність створення інструментальних засобів

Кнт ЕС6 – P7-5, P10-4

У прикладі прийнято наступні скорочення: S – суб’єкт ЭС, P – предикат ЭС, O – об’єкт ЭС; Кнт ЕС_m – контекст *m*-го елементарного сенсу; P*n-j-n* – номер рівня в онтографі контекстів, *j* – порядковий номер вершини концепту на рівні *n*.

Фрагмент онтографу контекстів наведено на рис. 4.

Слід відмітити, що в процесі обробки структурних компонентів ЕС активно використовується онтограф контекстів домена ПДО з метою однозначного їх співвіднесення до відповідної предметної галузі. При цьому сукупність ЕС з їх контекстами будемо називати *елементарними знаннями*.

Висновки

У цьому дослідженні зроблено спробу заглибитися у перетин онтологічних перспектив та обробки природної мови, щоб переглянути та доповнити спосіб видобування та представлення знань. Представляючи нову архітектуру лінгвістичного процесора, ми відійшли від традиційних реалізацій, об’єднавши лінгвістичні та онтологічні аспекти на етапі семантичного аналізу. Крім того, наше прагнення побудувати прогресивну онтолого-керовану інформаційну систему відзначається притаманним їй акцентом на безперервному самовдосконаленні. Особливістю нашого підходу є вдосконалення онтологічної системи, спеціально розробленої для обробки наукових даних. Перевага цієї системи полягає не лише в тому, що вона вмiло оперує елементарними знаннями, але й у її динамічній здатності породжувати нові концепти та створювати складні відношення. Такі досягнення відкривають величезні перспективи для підвищення ефективності та актуальності нашої системи в багатьох наукових галузях, що знаменує собою значний крок вперед у сфері представлення та аналізу знань.

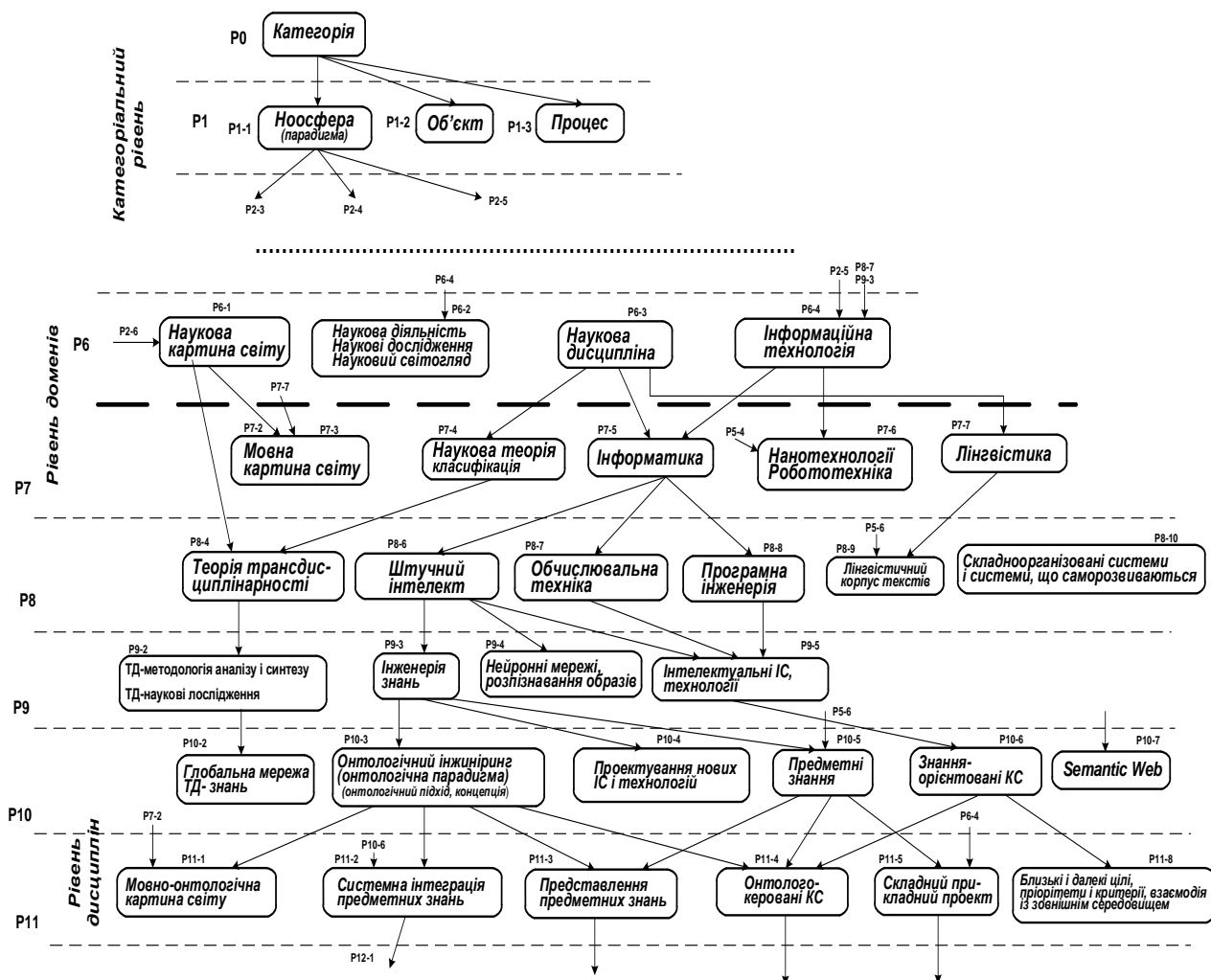


Рис. 4 Архітектура системи онтологій для обробки та вдосконалення НІІ
Fig. 4 The architecture of Ontology System for processing and enhancing SP

**Список використаних джерел**

1. Luger G. *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*. Pearson, Boston, Mass., 6th edition, 2008. ISBN 978-0-321-54589-3.
2. OpenAI. Gpt-4 technical report. Technical report, OpenAI, March 2023. URL <http://arxiv.org/abs/2303.08774>. arXiv:2303.08774 [cs]. (дата звернення: 15.10.2023).
3. Ford M. *Rule of the Robots: How Artificial Intelligence Will Transform Everything*. Basic Books, New York, first edition, 2021. ISBN 978-1-5416-7473-8.
4. Ford M. *Architects of Intelligence: The truth about AI from the people building it*. Pack Publishing, Birmingham, UK, 2018. ISBN 978-1-78913-151-2.
5. Gomez-Perez A., Fernandez-Lopez M., Corcho O. *Ontological Engineering*. Advanced Information and Knowledge Processing. Springer-Verlag, London, 1 edition, 2004. ISBN 978-1-85233-551-9. DOI: 10.1007/b97353.
6. Studer R. Staab S., editor. *Handbook on Ontologies*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2 editions, 2009. ISBN 978-3-540-70999-2. DOI: 10.1007/978-3-540-92673-3.
7. OntoChatGPT information system: Ontology-driven structured prompts for ChatGPT meta-learning / Palagin O., Kaverinskiy V., Litvin A., Malakhov K. *International Journal of Computing*, 22(2):170–183, July 2023. ISSN 2312-5381, 1727-6209. DOI:10.47839/ijc.22.2.3086.
8. Sowa J.F. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks. Cole, Boston, Mass, first edition, January 2000. ISBN 978-0-534-94965-5.
9. Guarino N. *Formal Ontology in Information Systems: Proceedings of the 1st International Conference June 6-8, 1998, Trento, Italy*. IOS Press, NLD, 1st edition, June 1998. ISBN 978-90-5199-399-8.
10. *Ontology-Driven Processing of Transdisciplinary Domain Knowledge* / O.V. Palagin and others. Iowa State University Digital Press, July 2023. ISBN 978-1-958291-06-1. DOI: 10.31274/isudp.2023.140.
11. Palagin O. V. An ontological conception of informatization of scientific investigations. *Cybernetics and Systems Analysis*, 52(1): 1–7, January 2016. ISSN 1060-0396, 1573-8337. DOI: 10.1007/s10559-016-9793-6.
12. Palagin O. V. Architecture of ontology-controlled computer systems. *Cybernetics and Systems Analysis*, 42(2):254–264, March 2006. ISSN 1060-0396, 1573-8337. DOI: 10.1007/s10559-006-0061-z.
13. Kurgaev A. F., Petrenko M. G. Processor structure design. *Cybernetics and Systems Analysis*, 31(4):618–625, July 1995. ISSN 1060-0396, 1573-8337. DOI: 10.1007/BF02366417.
14. Petrenko M.G., Sofiyuk A.A. On one approach to the transfer of an information structures interpreter to pld-implementation. *Upravlyayushchie Sistemy i Mashiny*, (6): 48–57, 2003. ISSN 0130-5395. URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0442276898&partnerID=40&md5=44974b40409363e5fe4378e240149c52>. (дата звернення: 15.10.2023).
15. Petrenko M.G., Kurgaev A.F. Distinguishing features of design of a modern circuitry type processor. *Upravlyayushchie Sistemy i Mashiny*, (5):16–19, 2003. ISSN 0130-5395. URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0347622333&partnerID=40&md5=7283307afdf891445ec9062c7b2ff80a>. (дата звернення: 15.10.2023).
16. Palagin O.V., Petrenko M.G., Velychko V.Y., Malakhov K.S. Development of formal models, algorithms, procedures, engineering and functioning of the software system “Instrumental complex for ontological engineering purpose”. In *CEUR Workshop Proceedings*, volume 1843, pages 221–232, Kyiv, Ukraine, May 2014. CEUR-WS. URL <http://ceurws.org/Vol-1843/221-232.pdf>. (дата звернення: 15.10.2023).
17. Kryvyi S.L. Formal ontological models in scientific researchers. *Upravlausie sistemy i masiny*, 263(3):04–15, June 2016. ISSN 01305395, 25181262. DOI: 10.15407/usim.2016.03.004.
18. Malakhov K.S., Petrenko M.G., Cohn E. Developing an ontology-based system for semantic processing of scientific digital libraries. *South African Computer Journal*, 35 (1):19–36, July 2023. ISSN 2313-7835, 1015-7999. DOI: 10.18489/sacj.v35i1.1219.
19. Davis E. Chapter 1 – Automating Common Sense, page 1–26. *The Morgan Kaufmann Series in Representation and Reasoning*. Morgan Kaufmann, January 1990. DOI:10.1016/B978-1-4832-0770-4.50009-5.
20. Mueller E. T. Chapter 1 – Introduction, page 1–16. *Morgan Kaufmann*, Boston, January 2015. ISBN 978-0-12-801416-5. DOI:10.1016/B978-0-12-801416-5.00001-2.

References

- [1] George Luger. *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*. Pearson, Boston, Mass., 6th edition edition, February 2008. ISBN 978-0-321-54589-3.
- [2] OpenAI. Gpt-4 technical report. Technical report, OpenAI, March 2023. URL <http://arxiv.org/abs/2303.08774>. arXiv:2303.08774 [cs].
- [3] Martin Ford. *Rule of the Robots: How Artificial Intelligence Will Transform Everything*. Basic Books, New York, first edition edition, September 2021. ISBN 978-1-5416-7473-8.
- [4] Martin Ford. *Architects of Intelligence: The truth about AI from the people building it*. Packt Publishing, Birmingham, UK, November 2018. ISBN 978-1-78913-151-2.
- [5] Asunci´on G´omez-P´erez, Mariano Fern´andez-L´opez, and Oscar Corcho. *Ontological Engineering*. Advanced Information and Knowledge Processing. Springer-Verlag, London, 1 edition, 2004. ISBN 978-1-85233-551-9. doi: 10.1007/b97353. URL <http://link.springer.com/10.1007/b97353>.



- [6] Studer R. Staab S., editor. Handbook on Ontologies. Springer Berlin Heidelberg, Berlin, Heidelberg, 2 edition, 2009. ISBN 978-3-540-70999-2. doi: 10.1007/978-3-540-92673-3. URL <http://link.springer.com/10.1007/978-3-540-92673-3>.
- [7] Oleksandr Palagin, Vladislav Kaverinskiy, Anna Litvin, and Kyrylo Malakhov. OntoChatGPT information system: Ontology-driven structured prompts for ChatGPT meta-learning. International Journal of Computing, 22(2):170–183, July 2023. ISSN 2312-5381, 1727-6209. doi:10.47839/ijc.22.2.3086. URL <https://computingonline.net/computing/article/view/3086>.
- [8] John F. Sowa. Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks / Cole, Boston, Mass, first edition edition, January 2000. ISBN 978-0-534-94965-5.
- [9] N. Guarino. Formal Ontology in Information Systems: Proceedings of the 1st International Conference June 6-8, 1998, Trento, Italy. IOS Press, NLD, 1st edition, June 1998. ISBN 978-90-5199-399-8.
- [10] Oleksandr Palagin, Mykola Petrenko, Sergii Kryvyi, Mykola Boyko, and Kyrylo Malakhov. Ontology-Driven Processing of Transdisciplinary Domain Knowledge. Iowa State University Digital Press, July 2023. ISBN 978-1-958291-06-1. doi: 10.31274/isudp.2023.140. URL <https://www.iastatedigitalpress.com/plugins/books/140/>.
- [11] O. V. Palagin. An ontological conception of informatization of scientific investigations. Cybernetics and Systems Analysis, 52(1):1–7, January 2016. ISSN 1060-0396, 1573-8337. doi: 10.1007/s10559-016-9793-6. URL <http://link.springer.com/10.1007/s10559-016-9793-6>.
- [12] O. V. Palagin. Architecture of ontology-controlled computer systems. Cybernetics and Systems Analysis, 42(2):254–264, March 2006. ISSN 1060-0396, 1573-8337. doi: 10.1007/s10559-006-0061-z. URL <http://link.springer.com/10.1007/s10559-006-0061-z>.
- [13] A. F. Kurgaev and M. G. Petrenko. Processor structure design. Cybernetics and Systems Analysis, 31(4):618–625, July 1995. ISSN 1060-0396, 1573-8337. doi: 10.1007/BF02366417. URL <http://link.springer.com/10.1007/BF02366417>.
- [14] M.G. Petrenko and A.A. Sofiyuk. On one approach to the transfer of an information structures interpreter to pld-implementation. Upravlyayushchie Sistemy i Mashyny, (6): 48–57, 2003. ISSN 0130-5395. URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0442276898&partnerID=40&md5=44974b40409363e5fe4378e240149c52>.
- [15] M.G. Petrenko and A.F. Kurgaev. Distinguishing features of design of a modern circuitry type processor. Upravlyayushchie Sistemy i Mashyny, (5):16–19, 2003. ISSN 0130-5395. URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0347622333&partnerID=40&md5=7283307afdf891445ec9062c7b2ff80a>.
- [16] Oleksandr Palagin, Mykola Petrenko, Vitalii Velychko, and Kyrylo Malakhov. Development of formal models, algorithms, procedures, engineering and functioning of the software system “Instrumental complex for ontological engineering purpose”. In CEUR Workshop Proceedings, volume 1843, pages 221–232, Kyiv, Ukraine, May 2014. CEUR-WS. URL <http://ceurws.org/Vol-1843/221-232.pdf>.
- [17] S.L. Kryvyi. Formal ontological models in scientific researchers. Upravl`a`u`sie sistemy i ma`siny, 263(3):04–15, June 2016. ISSN 01305395, 25181262. doi: 10.15407/usim.2016.03.004. URL <http://usim.org.ua/?pageid=4253&lang=en>.
- [18] Kyrylo Malakhov, Mykola Petrenko, and Ellen Cohn. Developing an ontology-based system for semantic processing of scientific digital libraries. South African Computer Journal, 35 (1):19–36, July 2023. ISSN 2313-7835, 1015-7999. doi: 10.18489/sacj.v35i1.1219. URL <https://sacj.cs.uct.ac.za/index.php/sacj/article/view/1219>.
- [19] Ernest Davis. Chapter 1 - Automating Common Sense, page 1–26. The Morgan Kaufmann Series in Representation and Reasoning. Morgan Kaufmann, January 1990. doi:10.1016/B978-1-4832-0770-4.50009-5. URL <https://www.sciencedirect.com/science/article/pii/B9781483207704500095>.
- [20] Erik T. Mueller. Chapter 1 - Introduction, page 1–16. Morgan Kaufmann, Boston, January 2015. ISBN 978-0-12-801416-5. doi: 10.1016/B978-0-12-801416-5.00001-2. URL <https://www.sciencedirect.com/science/article/pii/B9780128014165000012>.

Отримана в редакції 12.03.2024. Прийнята до друку 12.03.2024. Received 12 March 2024. Approved 12 March 2024. Available in Internet 10 April 2024