УДК 004.65

Малініч І. П., асистент
Козловський О. А., аспірант
Іванчук Я. В., докт. техн. наук, професор
Вінницький національний технічний університет, malinich@vntu.edu.ua

ЗАСТОСУВАННЯ СЕМАНТИЧНИХ MCP-ШЛЮЗІВ ДЛЯ ЗВ'ЯЗКУ АГЕНТІВ ШТУЧНОГО ІНТЕЛЕКТУ З ОНТОЛОГІЧНИМИ БАЗАМИ ЗНАНЬ

Для дослідження таких предметних областей як хмарні технології, операційні системи та комп'ютерні мережі онтологічні бази знань є дуже потужними інструментами для встановлення взаємозв'язків між концептами та автоматичного логічного виведення (різонінгу). Нові IT-технології постійно продовжують з'являтися, маючи при цьому новий функціонал та можливо нові методи взаємодії з іншими інформаційними системами, тому можливостей традиційних баз даних недостатньо для дослідницьких цілей. Іншою стороною питання всіх сучасних досліджень є використання великих мовних моделей (LLM), які значним чином збільшують можливості дослідників. Саме тому у онтологічному моделюванні дедалі більшою стає роль генеративного штучного інтелекту, який використовує онтологічні бази знань як контекстну базу [1]. До прикладу технологія цифрових двійників також значним чином покладається на онтологічні бази знань [2]. Тому для застосування можливостей агентів штучного інтелекту у онтологічних базах знань застосовуються семантичні MCP-шлюзи (рис. 1), які виступають у ролі інтерфейсу, що дає доступ LLM-моделям до елементів таких баз знань [3].

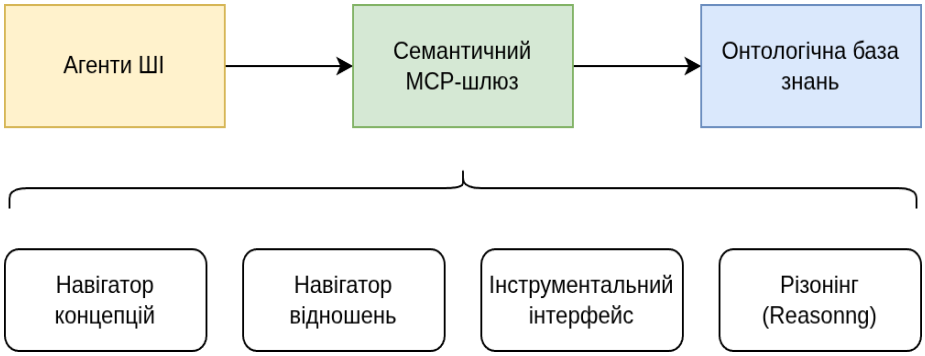

Рис. 1 – Концепція роботи семантичного MCP-шлюзу

Протокол моделі контексту (*англ.* Model Context Protocol, MCP) – це стандарт, який дозволяє LLM-моделям легко взаємодіяти з різноманітними зовнішніми джерелами даних і сторонніми інструментами за допомогою уніфікованої архітектури. Замість написання унікальних інтеграцій для доступу до різних API, розробники використовують MCP як спільний інтерфейс, де хост (наприклад, ChatGPT чи Gemini) підключається до серверу MCP, який надає доступ до змісту файлів, баз даних або до API [4]. Цей механізм працює за принципом клієнт-серверної взаємодії, де ШІ-модель надсилає запит на виконання певної дії або отримання контексту, а MCP-сервер повертає структуровані дані, що значно розширює можливості моделі за межами її тренувальних даних, забезпечуючи актуальність та точність відповідей у реальному часі [5]. Прикладом застосування MCP можна вважати випадок коли користувач простить LLM-сервіс повідомити прогноз погоди. LLM-модель звертається до MCP-серверу метеорологічного сервісу та отримує необхідні для користувача дані.

У випадку з онтологічними базами знань MCP-сервер виступає посередником, тому його найбільш доцільно називати MCP-шлюзом. Він перекладає абстрактні запити моделі у формальні запити до онтологічної бази [4]. Найпростіший сценарій застосування подібного шлюзу передбачає що LLM-модель отримує конкретні сутності та зв'язки, визначені в

онтології. Онтології можуть також доповнювати функціонал LLM-моделей, виступаючи у ролі верифікатора їх рішень. Наприклад коли модель пропонує дію, яка суперечить логіці онтології, MCP-сервер може повернути помилку валідації. Завдяки цьому можливо збільшити надійність та безпеку рішень ШІ, зокрема у таких сферах як інженерія та медицина. Однак інтеграція з онтологічними бази знань застосовується переважно у LLM-сервісах, пов'язаних з наукою та медициною, таких як Scite, Consensus та Elicit.

Для дослідження хмарних розгортань у подальших дослідженнях передбачається застосування онтологічної бази знань у вигляді графу хмарних сервісів. В такому графі вершинами виступатимуть хмарні сервіси, а ребрами – зв'язки між ними. Зв'язками вважатимуться як залежності, так і пряма мережева комунікація між сервісами. Така структура є досить зурчною для LLM-моделей, оскільки для роботи із сервісами не доведеться завантажувати всю базу знань у вікно контексту, а лише дозавантажувати потрібні гілки онтологічного дерева за запитом LLM-моделі. Завдяки використанню подібної структури на рівні MCP-шлюзу можна заощаджувати кількість токенів та працювати з доволі великими базами знань у реальному часі. Семантична точність у онтологічному моделюванні обчислювальних хмар є критичною, оскільки вона забезпечує однозначність і несуперечність опису понять, відносин та властивостей у хмарних системах. Її дотримання важливе для правильної роботи MCP-серверу, який комунікуватиме з LLM-моделями. Це дозволяє легше поєднувати різні сервіси, передбачати забезпечення обчислювальними ресурсами та знаходити проблемні аспекти їх застосування. Компанії, які залучають до виконання завдань інтелектуальних агентів, зазвичай заощадливо використовують токени, віддаючи перевагу найбільш простим LLM-моделями всюди де це можливо [6].

Висновок. MCP-шлюзи можуть перетворити онтології з пасивних сховищ у динамічну пам'ять для ШІ-агентів, надаючи їм крім безпосереднього доступу до даних також розуміння структури цих даних. MCP-шлюзи, маючи великий функціонал роботи з графовими структурами, можуть бути використані для досліджень хмарних розгортань. Для написання висновків використано LLM-модель Gemini 3.

Список посилань

1. Галушка О. В. Комплексні онтологічні та нейромережеві моделі фотографічних образів / О. В. Галушка, В. І. Шинкаренко // Інформаційні технології в металургії та машинобудуванні : матеріали міжнар. наук.-техн. конф. – Дніпро : НМетАУ, 2024. – С. 440–444. – DOI: 10.34185/1991-7848.itmm.2024.01.085.

2. Атаманюк О. В. Огляд моделей для проєктування цифрових двійників [Електронний ресурс] / Атаманюк О. В. // Сучасні інформаційні технології та системи в управлінні : зб. матеріалів VI Міжнар. наук.-практ. конф. молодих вчених, аспірантів і студентів, 10–11 квіт. 2025 р. / М-во освіти і науки України, М-во з питань стратег. галузей промисловості України, Київ. нац. екон. ун-т ім. Вадима Гетьмана [та ін.] ; [редкол.: О. М. Помазун (голова) та ін.]. – Електрон. текст. дані. – Київ : КНЕУ, 2025. – С. 125–126. – Назва з титул. екрану.

3. An Ontology-driven MCP Agent Framework for Ensuring Integrity in Architectural Structural Design / I. Paik, D. Kim, J. Lim, Y. Roh, S. Na // IEEE Access. – 2026. – PP. 1–1. – DOI: 10.1109/ACCESS.2026.3681621.

4. Ayyagari V. Model Context Protocol for Agentic AI: Enabling Contextual Interoperability Across Systems / V. Ayyagari // International Journal of Computational and Experimental Science and Engineering. – 2025. – Vol. 11. – DOI: 10.22399/ijcesen.3678.

5. Kumar P. Beyond the Universal Connector: Why MCP Needs Ontology to Scale in Enterprise [Електронний ресурс]. – Режим доступу: https://medium.com/@cloudpankaj/beyond-the-universal-connector-why-mcp-needs-ontology-to-scale-in-enterprise-3c476cbd23e0

6. Model context protocol (MCP): Landscape, security threats, and future research directions [Електронний ресурс] / X. Hou, Y. Zhao, S. Wang, H. Wang // arXiv preprint. – 2025. – arXiv:2503.23278. – Режим доступу: https://arxiv.org/abs/2503.23278.