

Цифрова трансформація

УДК 004.912

ЛАНДЕ Д.В., доктор технічних наук, професор, керівник Наукового центру інформатики і права ДНУ ПБП НАПрН України, завідувач кафедри НН ФТІ КПІ ім. Ігоря Сікорського.
ORCID: <https://orcid.org/0000-0003-3945-1178>.

ФОРМУВАННЯ СЕМАНТИЧНОЇ МАПИ ПОНЯТЬ В ГАЛУЗІ ПАРЛАМЕНТСЬКОГО КОНТРОЛЮ

Анотація. У статті викладено методiku формування моделі предметної області парламентського контролю у вигляді семантичної карти. Ця методика базується на використанні генеративних систем штучного інтелекту і програми аналізу і візуалізації мережевої інформації LegalGraph. Показано можливість формування семантичної карти предметної області, задачі, яка раніше потребувала залучення великих часових та людських ресурсів. Показано, як інтегруються засоби інтелектуальної текстової аналітики та аналізу і візуалізації мереж. Методика може застосовуватися для мережного аналізу інформації з будь-яких предметних областей, побудови відповідних семантичних мап.

Ключові слова: семантична мережа, генеративний штучний інтелект, правова інформація, парламентський контроль, квазієрархічна мережа, візуалізація.

Summary. The article presents a methodology for constructing a model of the subject area of parliamentary control in the form of a semantic map. This methodology is based on the use of generative artificial intelligence systems and the LegalGraph network information analysis and visualization program. It demonstrates the possibility of creating a semantic map of the subject area, a task that previously required significant time and human resources. It also shows how tools of intelligent text analysis and network analysis and visualization are integrated. This methodology can be applied to network information analysis in various subject areas and the creation of corresponding semantic maps.

Keywords: semantic network, generative artificial intelligence, legal information, parliamentary control, quasi-hierarchical network, visualization.

Постановка проблеми. Останнім часом генеративні системи штучного інтелекту (далі – ГШІ), такі як ChatGPT (<https://chat.openai.com>), стають дедалі більш поширеними в різних галузях. Найпоширеніші застосування включають машинний переклад, резюмування текстів, створення різних рівнів обговорення, таких як формулювання питань до навчальних матеріалів. Великі можливості у виділенні основних концепцій та іменних сутностей роблять ChatGPT корисним у фактографічних системах, зокрема в медицині та економіці [1]. Звісно, інтелектуальні чати інтегруються з зовнішніми системами, такими як геоінформаційні [2], системи аналізу та візуалізації графіків і мереж [3]. Зокрема, в роботі [3] показано, як можна створювати мережі зв'язків між персонажами літературних творів і мережі в областях з відношеннями “загальне-часткове”.

Такі системи ГШІ, як ChatGPT, зазвичай називають генеративними мовними моделями (GLM), великими мовними моделями (LLM) або просто мовними моделями (LM). Ці терміни широко використовуються в наукових і технічних колах для позначення систем, здатних генерувати текст на основі введення та контексту.

Основи та принципи таких систем, як ChatGPT, базуються на передових методах ШІ. Ці системи використовують глибоке навчання та моделі генерації мови для створення зрозумілих людині відповідей на текстові запити. Вони тренуються на великих обсягах даних, щоб зрозуміти контекст і створити відповідні та інформативні відповіді. Принципи роботи таких систем включають розуміння тексту, генерацію відповідей та покращення якості моделі за допомогою зворотного зв'язку та додаткового навчання.

Метою роботи є представлення нових можливостей обробки правової інформації із застосуванням систем генеративного штучного інтелекту, зокрема, ChatGPT, для вирішення задач побудови семантичних мап, семантичного індексування, аналізу та візуалізації, що дозволяє розглядати такі системи як корисний аналітичний інструмент в галузі права, зокрема, для здійснення парламентського контролю.

Виклад основного матеріалу. Ця стаття присвячена опису методики створення причинних мереж, отриманих шляхом багаторівневої декомпозиції поняття “Парламентський контроль” за допомогою системи ChatGPT, а також візуалізації та аналізу цих мереж за допомогою програми LegalGraph (<https://bigsearch.space/legal.html>). Створену мережу можна розглядати як семантичну карту, де кожен її вузол або зв'язок є гіперпосиланням, що веде до юридичної пошукової системи.

Парламентський контроль – це система, яка дозволяє парламентарям контролювати діяльність уряду. Ця система є важливою для забезпечення прозорості та підзвітності уряду перед народом [4]. Наведемо декілька причин актуальності застосування ГШІ при здійсненні парламентського контролю:

– *Ефективність.* ГШІ може автоматизувати багато завдань, які в даний час виконуються людьми, таких як моніторинг новин, аналіз даних та підготовка звітів. Це може звільнити час парламентарям для більш важливих завдань, таких як розробка законів та контроль за діяльністю уряду.

– *Прозорість.* ГШІ може використовуватися для створення нових інструментів, які підвищують прозорість діяльності уряду. Наприклад, ГШІ можна використовувати для створення чат-ботів, які можуть відповідати на запитання громадян про діяльність уряду.

– *Об'єктивність.* ГШІ може використовуватися для аналізу даних об'єктивно. Це може допомогти парламентарям приймати більш обґрунтовані рішення.

Звичайно, використання ГШІ у сфері парламентського контролю також пов'язане з деякими ризиками, такими як:

- *Зловживання ГШІ для поширення дезінформації.* ГШІ можна використовувати для створення фейкових новин або інших видів дезінформації, які можуть використовуватися для впливу на громадську думку або підриву довіри до уряду.

- *Непрозорість використання ГШІ.* Може бути важко зрозуміти, як ГШІ використовується для прийняття рішень, що може призвести до зниження довіри до парламенту.

Для того, щоб мінімізувати ці ризики, необхідно розробити чіткі правила та стандарти щодо використання ГШІ у сфері парламентського контролю. Ці правила повинні враховувати такі аспекти, як прозорість, відповідальність та справедливість.

Загалом, парламентський контроль на базі ГШІ має потенціал для революціонізації сфери парламентського контролю. Ця технологія може допомогти парламентарям бути більш ефективними та прозорими, а також може допомогти їм приймати більш обґрунтовані рішення.

Основна проблема, яка виникає при створенні семантичних карт, полягає в тому, що зазвичай для цього потрібні великі ресурси та залучення експертів. Звісно, існують

віддалені спроби автоматизованого створення каузальних мереж; однак запропонований підхід створення рою віртуальних експертів [5] може суттєво спростити і прискорити цей процес.

Формування семантичної мапи.

Прикладом моделі предметної області “Парламентський контроль”, в якості якої можна представити мережевий масив даних, який буде зручним для обробки комп’ютером, є направлена зважена мережа термінів. Направлена зважена мережа термінів (Directed Weighted Network of Terms – DWNT, або просто мережа термінів) – це семантична модель предметної області, де мережі є ключові терміни (слова та словосполучення), що відповідають сутностям, назвам концептів предметної галузі, а ребра – семантичні зв’язки між цими термінами. Порівняння DWNT, отриманих для різних текстів, дає змогу визначити семантичну близькість відповідних текстів.

У даному випадку нас, наприклад, цікавить проблематика парламентського контролю, тому запросимо у сервісу ChatGPT декомпозувати це поняття і видати відомі їй часткові поняття, що складають суть парламентського контролю. Тобто. центральним вузлом майбутньої мережі має стати поняття “Парламентський контроль”. Успішне відпрацювання такого запиту визначить другий рівень ієрархії – поняття пов’язані з парламентським контролем, відповідні часткові поняття. Після цього для кожного такого часткового поняття також здійснюється декомпозиція. Такий процес може тривати нескінченно, але в роботі зупинимось лише на трьох рівнях. Незважаючи на ієрархічне формування такої мережі, отримана мережа загалом не буде строго ієрархічною структурою, а скоріше близькою до неї квазіієрархічною. Деякі часткові поняття при декомпозиції різних понять можуть співпадати.

Найчастіше при використанні ГШІ сутності або пари сутностей (а саме це потребується при побудові мережевих структур) можуть витягувати безпосередньо з навченої моделі системи. Запропонувавши ChatGPT відпрацювати деякий запит (промпт), отримаємо множину часткових понять первинного поняття. Промпт для виявлення сутностей – результатів декомпозиції основного поняття може мати такий вигляд:



Декомпозуйте поняття "парламентський контроль". Виведіть 10 сутностей, які разом утворюють це поняття. Інформацію (позначення сутностей) прошу виводити у вигляді переліку окремих коротких записів без додаткових пояснень

У результаті ГШІ видає відповідь такого вигляду:



1. Парламент
2. Контроль
3. Нагляд
4. Виконавча влада
5. Звітність
6. Вплив
7. Механізми
8. Комітети
9. Голосування
10. Співпраця.

Система ChatGPT може надати різні відповіді під час обробки тексту в різний час, і всі вони розумні з точки зору людської логіки. Кожну з цих відповідей можна сприйняти як відповідь штучного експерта. Узагальнивши відповіді багатьох таких експертів, можна отримати більш повну і точну відповідь. Тому наведений промпт можна задати системі декілька разів, запустивши тим самим “рой віртуальних агентів”. Багаторазове виконання дозволяє отримати різні відповіді, багато з яких співпадуть. Із статистичного погляду, – це самі достовірні результати. У разі розгляду первинного поняття “Парламентський контроль”, отримуємо такі поняття:

ПАРЛАМЕНТ
 КОНТРОЛЬ
 НАГЛЯД
 ВИКОНАВЧА ВЛАДА
 ЗВІТНІСТЬ
 ВПЛИВ
 МЕХАНІЗМИ
 КОМІТЕТИ
 ГОЛОСУВАННЯ
 СПІВПРАЦЯ

Після виконання першого етапу – отримання результату декомпозиції, для кожного із отриманих часткових понять здійснюється подальша декомпозиція за допомогою промптів типу:



Декомпозуйте поняття "ВИКОНАВЧА ВЛАДА" виходячи із його функціональності як частини парламентського контролю. Виведіть 10 сутностей, які разом утворюють це поняття з погляду парламентського контролю. Інформацію прошу виводити у вигляді переліку окремих коротких записів у форматі "ВИКОНАВЧА ВЛАДА; сутність" без додаткових пояснень

При цьому система ChatGPT може допомогти отримати зміст CSV-файлу (поля, розділені точкою з комою), а саме:

ВИКОНАВЧА ВЛАДА; ВИКОНАННЯ ЗАКОНІВ
 ВИКОНАВЧА ВЛАДА; РЕГУЛЮВАННЯ ГАЛУЗЕЙ ГОСПОДАРСТВА
 ВИКОНАВЧА ВЛАДА; РОЗРОБКА ТА ВИКОНАННЯ БЮДЖЕТУ
 ВИКОНАВЧА ВЛАДА; ЗОВНІШНІ СПРАВИ
 ВИКОНАВЧА ВЛАДА; НАЦІОНАЛЬНА БЕЗПЕКА
 ВИКОНАВЧА ВЛАДА; ПРИЗНАЧЕННЯ ГРОМАДСЬКИХ СЛУЖБОВЦІВ
 ВИКОНАВЧА ВЛАДА; ГУБЕРНАТОРИ (У ФЕДЕРАЛЬНИХ СИСТЕМАХ)
 ВИКОНАВЧА ВЛАДА; АДМІНІСТРАТИВНІ РІШЕННЯ
 ВИКОНАВЧА ВЛАДА; КОНТРОЛЬ ЗА БЮДЖЕТНИМИ ВИДАТКАМИ
 ВИКОНАВЧА ВЛАДА; ВІДПОВІДНІСТЬ КОНСТИТУЦІЇ

Такого типу відповіді можуть слугувати інформаційною основою для побудови графових структур – семантичних мап, які одночасно є джерелом посилань, тому що кожний вузол і кожний зв’язок цієї мережі містить гіперпосилання на відповідні запити до пошукової системи порталу Верховної Ради України (<https://zakon.rada.gov.ua/laws>). Кожне екстраговане поняття може розглядатись як вузол мережі. Крім того, з подібним промптом можна звернутися до декількох систем ШІ, серед яких ChatGPT, Bard

(<https://bard.google.com/chat>), GPT Free (<https://gptfree.co>) тощо. Після отримання відповідей від різних систем їх можна поєднати, при цьому реалізується концепція “віртуальних експертів”, якими виступають такі системи.

Аналіз і візуалізація семантичної мапи.

Аналіз мереж є важливим методом для розуміння взаємозв'язків між об'єктами та виявлення зв'язків між ними, що може значно полегшити дослідження конкретної сфери. Для проведення аналізу та візуалізації мереж, спільно з системами, подібними до ChatGPT, можна використовувати сучасні графові інструменти, такі як Neo4j та Gephi [7]. Однак при використанні таких програмних продуктів аналітики стикаються з двома труднощами: потребою встановлення програм, що не завжди можливо, особливо при обмеженнях на встановлення стороннього програмного забезпечення, та потребою досконально вивчити особливості функціонування цих систем, розібратися в численних параметрах, режимах розміщення графів, кластеризації тощо.

Для вирішення проблеми на основі бібліотеки системи GrahViz розроблено програму LegalGraph, доступного наразі в Інтернет за адресою <https://bigsearch.space/legal.html>. Ця програма забезпечує первинний аналіз і відображення графів, інформація про які відповідає формату CSV (від англ. Comma-Separated Values – значення, розділені комою), кожен запис в якому відповідає парі сутностей. При цьому кожному вузлу і ребру графа у відповідність ставиться гіперпосилання на законодавчу базу даних, розміщену в Інтернеті.

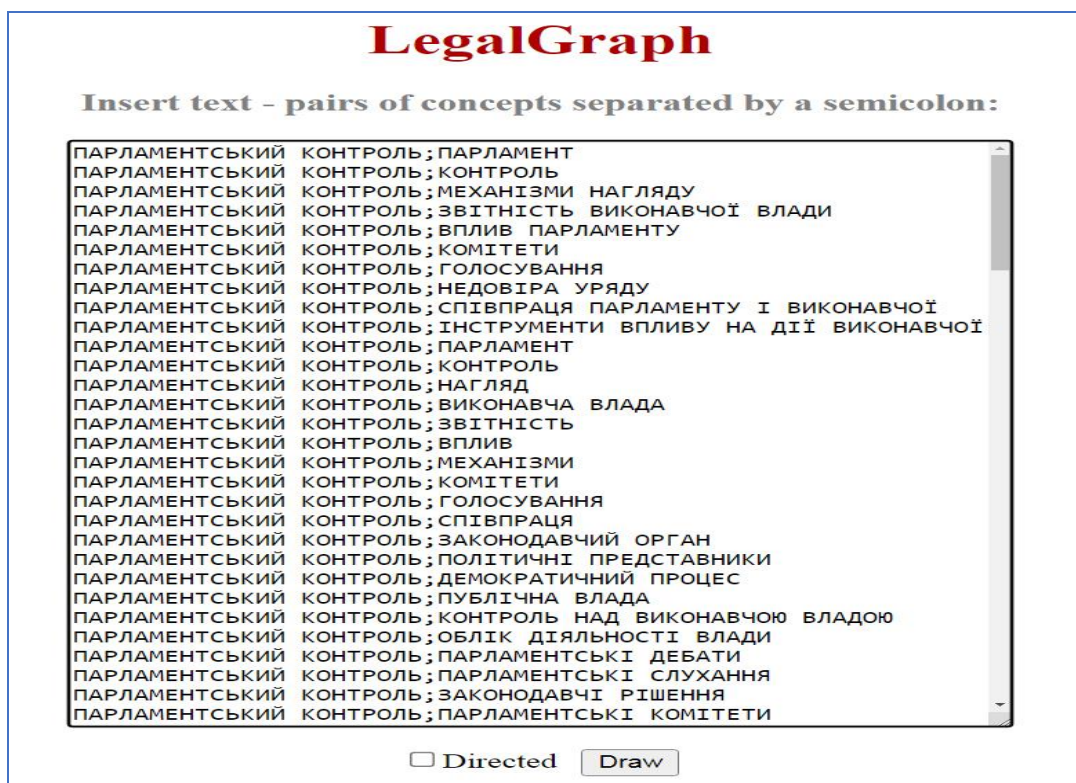


Рис. 1. Інтерфейс введення інформації в системі LegalGraph.

На Рис. 2 показано результат – фрагмент семантичної мапи предметної області “Парламентський контроль”, який було сформовано на ресурсі сервісу ChatGPT. Як бачимо, ця мережа квазіієрархічна, до вузлів, що поєднують різні часткові поняття, тобто є транзитними, відіграють особливу роль в процесі парламентського контролю, віднесено такі:

КОНТРОЛЬ ЗА ВИКОНАВЧОЮ ВЛАДОЮ;
 ЗВІТНІСТЬ ВИКОНАВЧОЇ ВЛАДИ;
 ПАРЛАМЕНТСЬКІ СЛУХАННЯ;
 АУДИТ БЮДЖЕТНИХ ВИТРАТ;
 ЗВІТУВАННЯ ПЕРЕД ПАРЛАМЕНТОМ;
 ІНСПЕКЦІЇ ТА РЕВІЗІЇ;
 ЗАКОНОДАВЧИЙ ОРГАН.

У наведеному графі – семантичній мапі предметної області кожний вузол і кожне ребро виступають гіперпосиланнями на пошукову систему. При активізації такого гіперпосилання здійснюється перехід на сторінку цього веб-ресурсу з результатами пошуку (Рис. 3).

Програмою формується відображення орієнтованих графів, вузли ранжуються за ступенем і розфарбовуються, визначається товщина і напрямок ребер. Покладання графа на площину виконується методами, вбудованими в GraphViz. При цьому формується зображення графа у форматі SVG, що дає можливість формувати гіперпосилання, що ведуть до пошукових форм системи Верховної Ради України. з вузлів і ребер графа. Масштабована векторна графіка (SVG) – це формат для визначення двовимірної графіки за допомогою XML. Він підтримує інтерактивність. Специфікація SVG, відкритий стандарт, розроблений World Wide Web Consortium з 1999 року, дозволяє масштабувати зображення без втрати якості. Ці зображення зберігаються в текстових XML-файлах, що робить їх доступними для пошуку, індексування, сценаріїв і стискування. Усі основні настільні браузери почали підтримувати SVG. Дані у файлі SVG є текстом, а не зображенням, тому можна вбудовувати інтерактивні функції, такі як гіперпосилання на веб-ресурси, які використовуються в програмі LegalGraph. Реалізація гіперпосилань надає можливість перегляду нормативних документів, що з одного боку надає інформаційне розкриття поняття або зв'язку, а з іншого – підтверджує коректність побудови семантичної мапи.

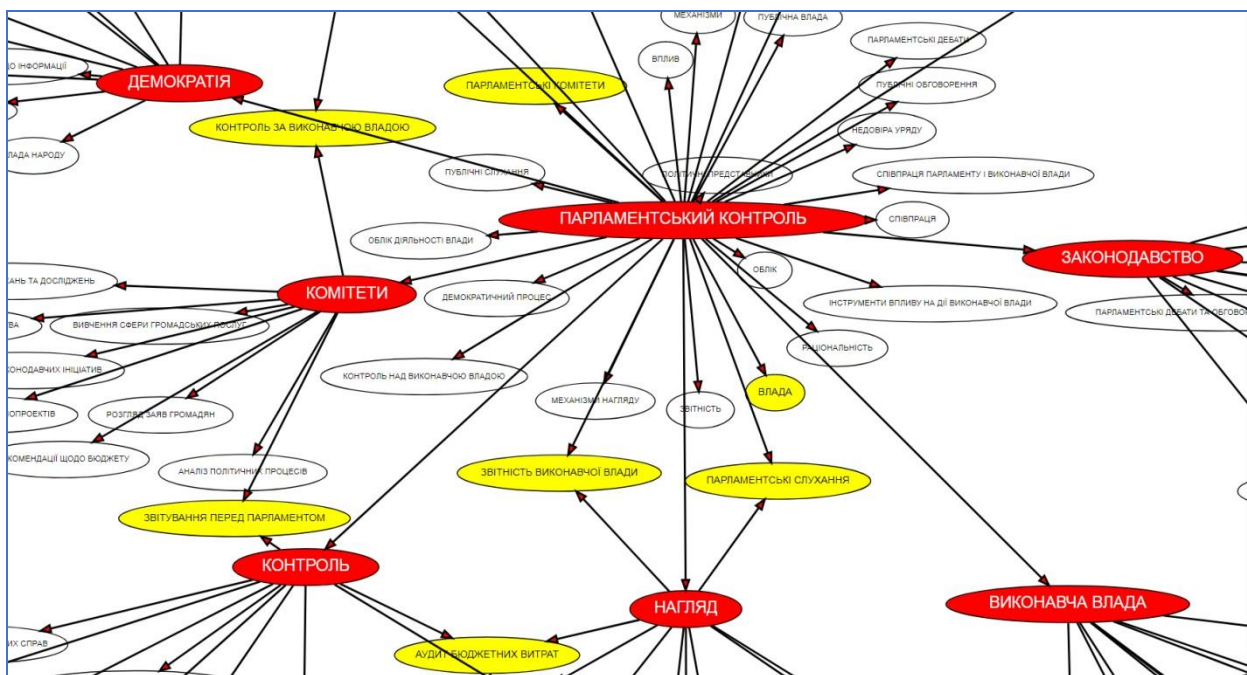


Рис. 2. Фрагмент мережі понять, що відповідають тематиці
 “Парламентський контроль”.

Висновки.

Створення семантичної мапи на основі технології ГШІ і системи LegalGraph може допомогти юристам при використанні правових знань при здійсненні парламентського контролю завдяки:

- зручному доступу до інформації – мапа посилань робить навігацію по поняттях щодо парламентського контролю більш інтуїтивною, що дозволяє користувачам швидше і ефективніше знаходити потрібну інформацію;
- семантична мапа може допомогти в структуруванні та візуалізації знань, роблячи їх більш зрозумілими;
- семантична мапа дозволяє встановлювати зв'язки між різними сутностями і документами, допомагаючи користувачам легше розуміти, як різні інформаційні елементи взаємодіють між собою;
- семантична мапа може бути важливим методичним інструментом, дозволяючи здобувачам освіти швидше знаходити відповіді на свої питання, виходити на документи-першоджерела;
- семантична мапа може стати важливим інструментом для поширення знань серед широкої аудиторії, надаючи зручний доступ до правової інформації та сприяючи її поширенню.

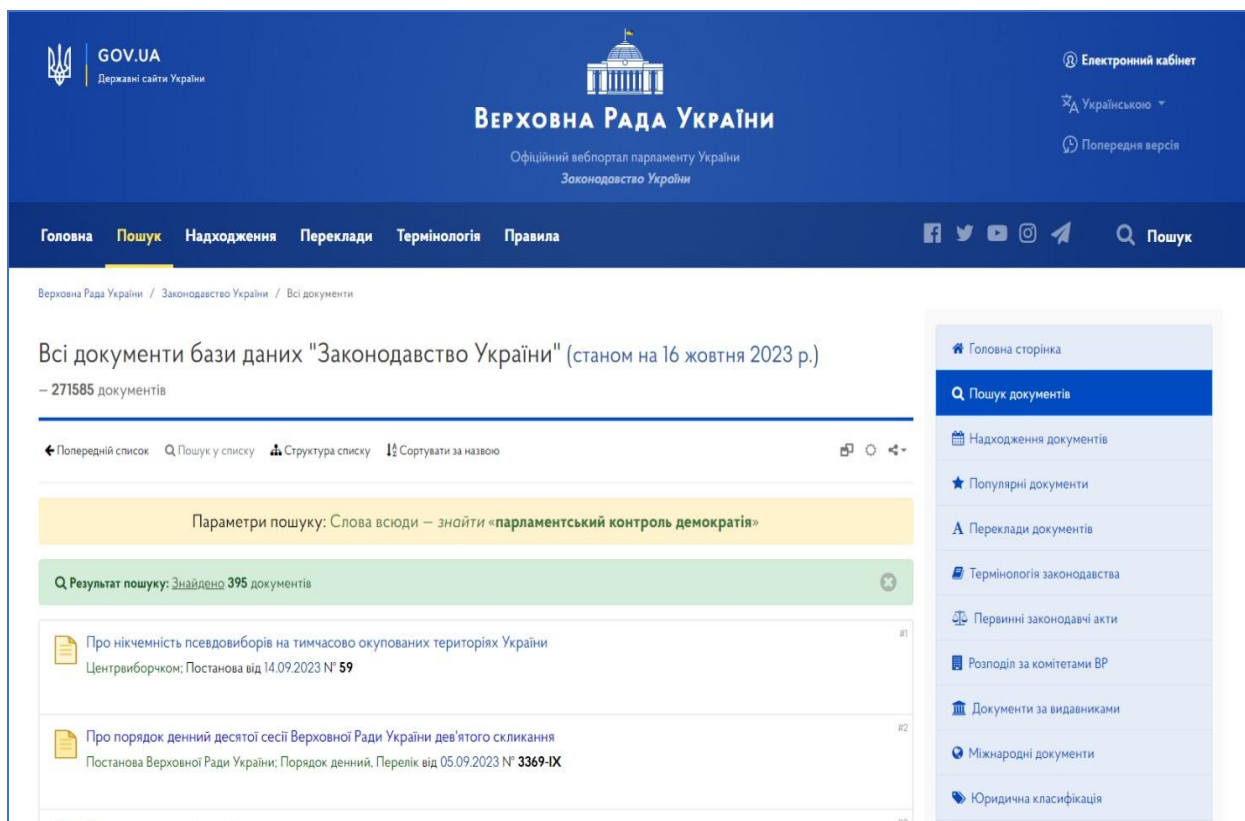


Рис. 3. Результат переходу за гіперпосиланням на ребрі між вузлами “Парламентський контроль” і “Демократія”.

Отже, використання методики формування семантичних мап може сприяти покращенню рівня парламентського контролю завдяки зручному охопленню всієї понятійної бази і швидкому доступу до відповідної нормативно-правової інформації.

Загалом продемонстровано практичне застосування передової технології ШІ в галузі аналізу тексту та візуалізації мережі.

Показано, як використання алгоритмів машинного навчання може допомогти розблокувати раніше приховані ідеї та закономірності в великих обсягах текстових даних, а також отримати глибше розуміння складних явищ у різних областях, зокрема, в правовій науці.

Використана література

1. Lande, Dmitry and Strashnoy, Leonard, Concept Networking Methods Based on ChatGPT & Gephi (April 17, 2023). SSRN. URL: <http://dx.doi.org/10.2139/ssrn.4420452>
2. Tamilla Triantoro. Graph Viz: Exploring, Analyzing, and Visualizing Graphs and Networks with Gephi and ChatGPT (March 30, 2023). ODSC Community. URL: <https://open.datascience.com/graph-viz-exploring-analyzing-and-visualizing-graphs-and-networks-with-gephi-and-chatgpt>
3. Solat J. Sheikh, Sajjad Haider, Alexander H. Levis, On semi-automated extraction of causal networks from raw text, Engineering Applications of Artificial Intelligence, Volume 123, Part A, 2023, 106189, ISSN 0952-1976. URL: <https://doi.org/10.1016/j.engappai.2023.106189>
4. Д.В. Ланде, В.М. Фурашев, С.М. Брайчевський. Інформатика парламентського контролю: посібник. Київ, 2022. 256 с. ISBN 978-966-2344-80-6.
5. Dmytro Lande, Leonard Strashnoy. GPT Semantic Networking: A Dream of the Semantic Web – The Time is Now. Kyiv: Engineering, 2023. 168 p. ISBN 978-966-2344-94-3.
6. Ken Cherven. Mastering Gephi Network Visualization. Packt Publishing, 2015. ISBN 78-1-78398-734-4.

~~~~~ \* \* \* ~~~~~