

Правова інформатика

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>.

СТРАШНОЙ Л.Л., старший фахівець з архітектури даних (Senior Data Architect), Університет Каліфорнії (UCLA), Лос-Анджелес, США.

ФОРМУВАННЯ МЕРЕЖ ПОНЯТЬ В ГАЛУЗІ ПРАВА ЗА ДОПОМОГОЮ СИСТЕМИ ШТУЧНОГО ІНТЕЛЕКТУ

DOI: ...

***Анотація.** Описується методика формування мереж понять, пов'язаних із правовими документами і поняттями предметної галузі “Інформаційне право”. Методика базується на використанні системи штучного інтелекту ChatGPT і програми аналізу і візуалізації мереж Gephi. Показано можливість побудови мереж концептів на основі використання ChatGPT, вирішуються завдання, які раніше вирішувалися із залученням великих часових та людських ресурсів. Показано, як інтегруються засоби інтелектуальної текстової аналітики та аналізу мереж, а також їхня візуалізація. Запропоновано емуляцію багатьох експертів шляхом багаторазової побудови запиту до системи ChatGPT. Методика може застосовуватися для мережного аналізу документальних джерел, побудови моделей предметних галузей.*

***Ключові слова:** ChatGPT, мережі понять, візуалізація графів, мережевий аналіз документів, модель домену, штучні експерти, інформаційне право.*

***Summary.** The article describes a methodology for constructing networks of concepts related to legal documents and the subject area of “Information Law”. The methodology is based on the use of the artificial intelligence system ChatGPT and the Gephi network analysis and visualization program. The article demonstrates the possibility of constructing concept networks using ChatGPT, solving tasks that previously required large time and human resources. It shows how tools of intelligent text analytics and network analysis are integrated, as well as their visualization. The article proposes emulating multiple experts by repeatedly querying the ChatGPT system. The methodology can be applied to network analysis of documentary sources and the construction of subject area models.*

***Keywords:** ChatGPT, Concept networks, Graph visualization, Book network analysis, Domain model, Artificial experts, Information law.*

Постановка проблеми. У сучасному світі здатність обробляти дані стала одним з найважливіших пріоритетів. Великі дані (Big Data) революціонізують наше розуміння світу, але екстрагування значущої інформації з такої величезної кількості даних є нелегкою задачею. Тут на допомогу приходять системи штучного інтелекту, зокрема, ChatGPT, що дозволяє користувачам ефективно обробляти та аналізувати масиви даних великих обсягів.

Метою статті є представлення нових можливостей обробки правової інформації із застосуванням систем штучного інтелекту, зокрема, ChatGPT, для вирішення задач семантичного аналізу та візуалізації, що дозволяє розглядати такі системи як корисний інструмент для аналізу правових документів, формування моделей предметних галузей.

Виклад основного матеріалу.

Технологічне підґрунтя.

Мережевий аналіз та візуалізація є важливими засобами для розуміння зв'язків між елементами даних та виявлення прихованих закономірностей, що може допомогти дослідникам оптимізувати свій робочий процес.

Для мережевого аналізу та візуалізації в поєднанні з системою ChatGPT можуть використовуватися сучасні інструменти, такі як Neo4j та Gephi. Neo4j є потужною графовою базою даних на сервері, яка дозволяє ефективно зберігати та обробляти великі обсяги пов'язаних даних. Більш орієнтований на користувача Gephi є інтерактивною персональною платформою, яка дозволяє легко досліджувати та візуалізувати складні мережі. Саме цим інструментом будуть демонструватися приклади в цій статті.

Ця робота присвячена опису методів формування семантичних мереж, онтологічних структур, пов'язаних з різними поняттями конкретного юридичного документа (Конституції США) та цілої предметної області “Інформаційне право”. Про значення таких мереж для науки та практики написано багато робіт, створені концепції, спеціальні програми та мови опису даних [1 – 2]. Методика дозволяє формувати мережі, графові структури, що мають різні топологічні ознаки: прості ненаправлені, зважені, направлені зважені, направлені з підписаними ребрами. І звичайно, методика базується на використанні системи ChatGPT, навченої лінгвістичної мережі, тому в статті не буде використано жодної формули чи явно прописаного розрахунку. Це саме те, про що писав колись С. Вольфрам, коли йшлося про клітинні автомати [3]. До речі, сьогодні він також багато пише про можливості ChatGPT [4]. Методика включає візуалізацію сформованих мереж. Для цього використовується програма Gephi (gephi.org) – найпопулярніша на сьогодні програма візуалізації графових структур з вільною ліцензією [5]. Для завантаження даних в середовище Gephi досить підходить формат CSV, тому всі запити до ChatGPT будуть супроводжуватися вимогою до формату.

Формування мережі понять із юридичного документу.

Отже, в наших планах передбачено описання процедур формування мереж з різними властивостями. Перейдемо безпосередньо до опису задач та їх розв'язань. Ми покажемо, як формуються мережі понять на прикладі аналізу Конституції США. Слід зауважити, що не кожен документ був охоплений системою ChatGPT під час її навчання. Очевидно, що система “знає” документ такого масштабу. Для побудови мережі достатньо отримати файл у форматі CSV та завантажити його в програму Gephi.

Створення ненаправленої мережі.

Для початку треба знати декілька, наприклад, 20 основних пар понять у тексті. Система ChatGPT може допомогти отримати вміст такого файлу у потрібному форматі (поля, які відповідають назвам понять, розділені крапкою з комою). Для цього можна застосувати, наприклад, такий запит до системи ChatGPT:

Prompt 1 (з англ. “підкажіть”) – Надайте мені 20 пар найважливіших зв'язаних понять із Конституції США. Кожна пара понять нехай видається у форматі “поняття 1; поняття 2”. Кожна така пара понять виводиться з нового рядка.

У результаті ChatGPT видає таку відповідь:

1. Федералізм; Державний суверенітет.
2. Конгрес; Сенат.
3. Представницький дім; Обрання президента.
4. Кабінет міністрів; Президент.
5. Верховний суд; Конституційні права.
6. Конституційні права; Законодавча влада.

7. Виконавча влада; Право вето.
8. Право вето; Конгрес.
9. Конституційні права; Судді.
10. Державний суверенітет; Законодавча влада.
- ...

Завантаживши отримані дані в систему Gephi, вибираємо розмір вузлів, пропорційний ступеню (кількості прилеглих зв'язків) і, розділивши мережу на кластери за критерієм модулярності, отримуємо наочний графік (Рис. 1).

Отже, перша мережа побудована і по ній видно, що центральну частину Конституції США займають “Законодавча влада” та “Конституційні права”, що загалом підтверджує правильність вибору понять системою ChatGPT.

Очевидно, сформована мережа слабо зв'язана і неповна. Вважатимемо, що це мережа, отримана в результаті опитування одного штучного експерта.

Емуляція багатьох експертів шляхом багаторазового виконання запиту.

Система ChatGPT може видавати різні варіанти відповідей на текст у різні моменти обробки, причому всі вони можуть бути правильними з точки зору людської логіки. Кожен з таких варіантів можна сприймати як відповідь від окремого штучного експерта [6 – 7].



Рис. 1. Ненаправлена мережа 20 головних понять Конституції США за оцінкою програми ChatGPT

Можна припустити, що, поєднуючи відповіді багатьох подібних експертів, можна отримати більш повну та точну відповідь. Щоб забезпечити повноту набору понять (або їх пар), а також точність (незначні зв'язки матимуть малу вагу), можна скористатися можливістю запуску “рою” штучних агентів. Для цього можна запустити попередній запит декілька разів та об'єднати відповіді в агрегований CSV-файл. Мережу можна доповнювати до тих пір, поки вона не стане достатньо повною згідно з оцінками експерта-людини. Центральний фрагмент отриманої мережі після первинної обробки представлено на Рис. 2.

Формування мережі понять – моделі предметної області.

Мережу понять предметної області представимо у вигляді спрямованої зваженої мережі. У попередніх прикладах ми формували мережу понять, пов'язану із конкретним документом – Конституцією США, який, звичайно ж, охоплювався системою ChatGPT

при її навчанні. Далі розглянемо завдання формування мережі понять, пов'язаних з цілою предметною областю. Очевидно, ChatGPT знає основні поняття з найважливіших предметних областей. Нас цікавить область інформаційного права, тому запитаємо у системи знайти 20 пар найпопулярніших понять з даної предметної галузі, пов'язаних між собою. При цьому визначимо відношення порядку – нехай перше поняття у парі буде загальнішим, ніж друге.

Запропонуємо системі ChatGPT виконати такий запит:

Prompt 2 – Назвіть мені 20 пар важливих понять із галузі Інформаційного права. Перше слово нехай буде більш загальне, ніж друге. Кожна пара понять нехай видається у форматі “поняття 1; поняття 2”. Кожна така пара понять виводиться з нового рядка.

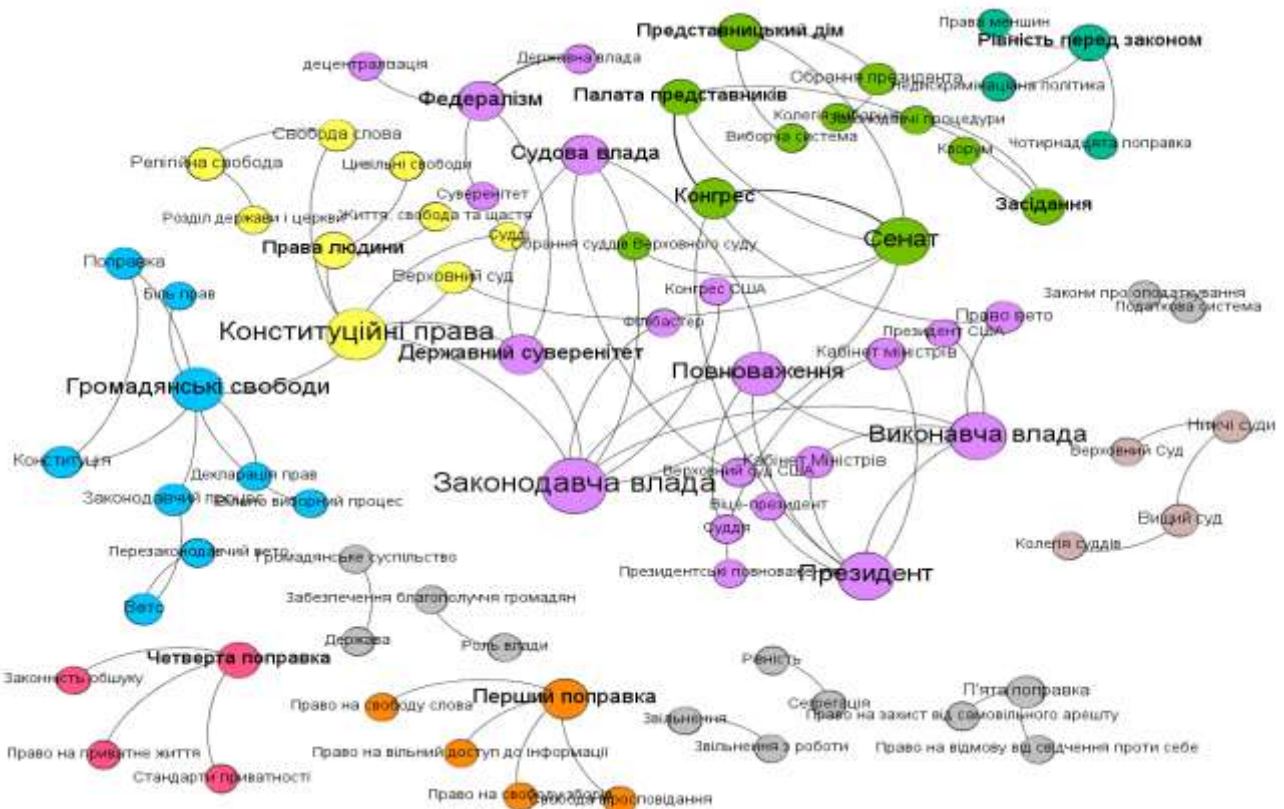


Рис. 2. Центральний фрагмент агрегованої мережі понять Конституції США

Зважаючи на широкий спектр галузей, що стосуються “Інформаційного права”, надамо список пар понять з різних підгалузей:

1. Конфіденційність інформації; захист персональних даних.
2. Авторське право; інтелектуальна власність.
3. Комп’ютерний злочин; кібербезпека.
4. Свобода слова; регулювання контенту в Інтернеті.
5. Право на доступ до інформації; прозорість влади.
6. Цифровий підпис; електронний документообіг.
7. Інформаційна безпека; захист від кібератак.
8. Електронна комерція; правила дистанційної торгівлі.
9. Антимонопольне законодавство; боротьба з монополіями в галузі ІТ.
10. Інтернет-права; регулювання використання Інтернету.

...

Емуляція багатьох експертів шляхом багаторазового повторення запиту.

Як було показано раніше, при побудові мережі понять за Конституцією США, для забезпечення повноти та точності множини понять можна запустити “рій” штучних агентів. Для цього досить запустити наведений вище запит кілька разів і об’єднати відповіді. На Рис. 3 показано центральний фрагмент мережі понять, що відповідає предметній області “Інформаційне право” після завантаження агрегованих даних, багато разів отриманих від системи ChatGPT.

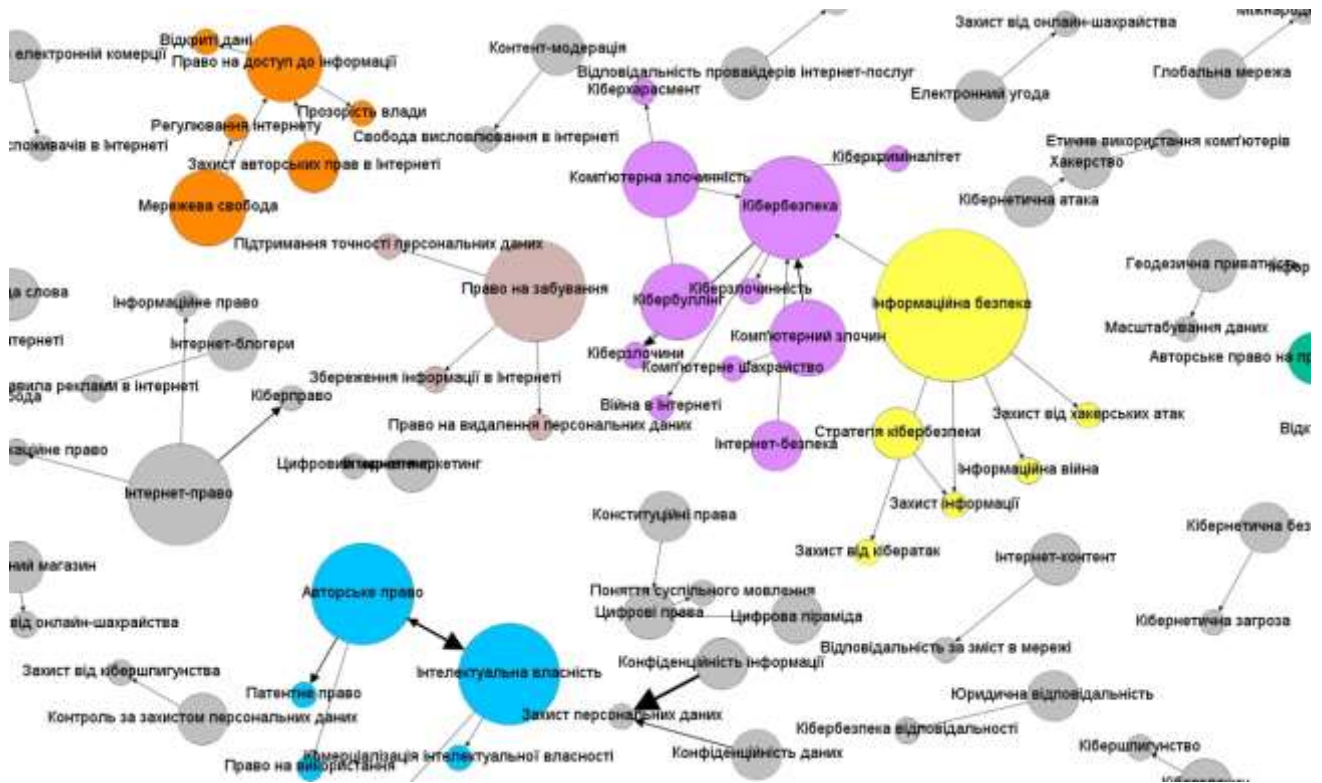


Рис. 3. Мережа понять з предметної галузі “Інформаційне право”

Висновки.

У цій роботі:

1. Показано можливість використання ChatGPT для побудови концептуальних мереж, що має потенціал для значного скорочення ресурсів, які зазвичай потрібні для виконання цих завдань вручну. Крім того, ми показали, що інтеграція інструментів текстової аналітики та мережевого аналізу може виявитися дуже корисною для отримання інформації з великих обсягів неструктурованих даних.

2. Вперше запропоновано емуляцію “рою” експертів шляхом багаторазового виконання запиту до системи штучного інтелекту. Цей підхід може значно підвищити ефективність екстрагування знань і забезпечити глибше розуміння структури та значення окремих документів і цілих предметних областей.

3. Запропоновано методику, яка може застосовуватися для мережевого аналізу документів, побудови моделей різних предметних областей.

Ця робота демонструє практичне застосування передової технології штучного інтелекту для аналізу тексту в правовій галузі. На прикладах показано, як використання алгоритмів машинного навчання може допомогти розкрити приховані ідеї та закономірності в текстових даних.

Важливо зазначити, що цей підхід не позбавлений обмежень. Інтерпретація результатів вимагає досвіду в досліджуваній області, зокрема в галузі права, і все ще існує потреба в нагляді людини для забезпечення повноти і точності результатів.

Загалом очевидно, що штучний інтелект має величезний потенціал для трансформації сфери аналізу тексту та візуалізації мереж. Хоча існує ще багато проблем, які потрібно вирішити, це дослідження є певним кроком у розкритті можливостей технологій штучного інтелекту для глибшого розуміння знань, закладених в правові документи.

Використана література

1. Berners-Lee, T., Hendler, J., Lassila, O. The semantic web. *Scientific American*, 2001. Vol. 284. No. 5. Pp. 34-43.
2. Ланде Д.В., Дмитренко О.О. Побудова семантичних мереж та визначення ступеня розбіжності текстів. *Інформація і право*. № 2(41)/2022. С. 44-51.
3. Wolfram S. *A New Kind of Science*. Wolfram Media Inc. 2002. 1264 p.
4. Wolfram S. *What Is ChatGPT Doing and Why Does it Work?* Wolfram Media Inc. 2023. 112 p.
5. Cherven K. *Mastering Gephi. Network Visualization*. Packt Publishing. 2015. 378 p.
6. Mnih, V., Badia, A.P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., Kavukcuoglu, K. Asynchronous methods for deep reinforcement learning. *International Conference on Machine Learning*. 2016. Pp. 1928-1937.
7. Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M. Mastering the game of go with deep neural networks and tree search. *Nature*, 529 (2016). Pp. 484-503.

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