

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 І АНАЛІЗ МЕРЕЖ ПОДІЙ У СФЕРІ ПАРЛАМЕНТСЬКОГО КОНТРОЛЮ НА ОСНОВІ ЗАСТОСУВАННЯ СИСТЕМ ШТУЧНОГО ІНТЕЛЕКТУ

Анотація. У статті наведено методологію формування і аналізу мережі подій в новинних повідомленнях у сфері парламентського контролю на основі застосування Генеративного штучного інтелекту і приклад її застосування. Революція в галузі штучного інтелекту дозволяє вирішувати завдання не тільки виявлення, але й формування Каузальних мереж подій, в якій у явному вигляді наведені події-причини і події-наслідки. Завдяки використанню великих лінгвістичних моделей отримані зручні методи екстрагування подій із текстів, їх фільтрації, кластеризації. Виявлення причинно-наслідкових зв'язків здійснюється із застосуванням штучного інтелекту, що значно спрощує роботи з природною мовою. Візуалізація і кластерний аналіз сформованих мереж може здійснюватись із застосуванням традиційних інструментів аналізу мереж.

Ключові слова: Генеративний штучний інтелект, виявлення подій, екстрагування зв'язків, Каузальна мережа, візуалізація мережі, кластерний аналіз.

Summary. The article introduces a methodology for forming and analyzing the network of events in news reports related to parliamentary control. This methodology relies on the application of generative artificial intelligence, and the article provides examples of its practical implementation. The revolution in artificial intelligence enables the solution of tasks not only related to identification but also to the formation of causal networks of events, where the causes and consequences are clearly presented. The utilization of large linguistic models has yielded convenient methods for extracting events from texts, filtering them, and clustering. Artificial intelligence is employed for identifying cause-and-effect relationships, significantly simplifying the processing of natural language. Visualization and cluster analysis of the formed networks can be performed using traditional tools for network analysis.

Keywords: Generative artificial intelligence, event detection, connection extraction, Causal network, network visualization, cluster analysis.

Постановка проблеми. Виявлення нових подій із текстових новинних повідомлень – це традиційна задача з області інтелектуальної обробки природної мови, розв'язанню якої присвячено багато наукових робіт [1; 2]. У сфері обробки правових документів задачі виявлення подій також досліджувались на перетині двох наук – права і комп'ютерної лінгвістики [3; 4].

Разом з цим, революція в галузі штучного інтелекту (GenAI), поява великих лінгвістичних моделей (LLM) дозволяє вирішувати завдання не тільки виявлення, але й формування Каузальної мережі подій, в якій у явному вигляді наведені події – причини і події-наслідки. Для вирішення цієї задачі на цей час можуть використовуватись великі

лінгвістичні моделі, зокрема, ChatGPT (<https://chat.openai.com>), Llama-2 (<https://www.llama2.ai>), Gemini (<https://gemini.google.com/app>), Deep Seek (<https://deepseek.com>).

Подальший аналіз такої мережі дозволяє виявляти кластери і ланцюжки подій, аналогічні тим, що виявляються в сценарному аналізі [5]. Відмінність нового підходу полягає у тому, що вузлами традиційних Каузальних мереж виступають концепти, тобто поняття, а методологія що пропонується дозволяє розглядати як вузли деякі події. Сформовані мережі можуть виступати як навігатори у світі подій, бути інструментами для проведення досліджень і розслідувань, зокрема, у галузі парламентського контролю. Аналіз і візуалізація сформованих Каузальних мереж подій можуть здійснюватись звичайними інструментами аналізу мереж, зокрема, такими як Gephi [6] або GraphViz [7]. Як і у випадку традиційних Каузальних мереж при цьому можуть виявлятися найбільші за вихідним ступенем вузли (події, що є причиною багатьох подій-наслідків), визначатися різної довжини ланцюжки, ланками яких виступають події, що є наслідками багатьох причин, події добрі “посередники”, що поєднують багато причин і наслідків, тобто є ланками зв’язку для одночасно багатьох причин і наслідків.

Метою роботи є представлення нових можливостей обробки правової документації із застосуванням систем Генеративного штучного інтелекту для здійснення парламентського контролю.

Виклад основного матеріалу.

Опис методології. Методологія формування мережі подій, що містяться в новинних повідомленнях, передбачає реалізацію таких етапів:

1. Отримання новинних повідомлень, що відповідають деякій цільовій тематиці, за допомогою наявних систем пошуку новин (вільних або пропріетарних).
2. За допомогою засобів GenAI і застосування LLM здійснюється виявлення подій, що містяться у вибраних новинних повідомленнях, створення масиву позначень цих подій.
3. Виявлення оригінальних подій серед відібраного масиву.
4. Зв’язування оригінальних подій причинно-наслідковими зв’язками.
5. Формування і візуалізація семантичної мапи за допомогою графічного інструменту на основі бібліотеки GraphViz.
6. Виявлення концептів (ключових слів), що відповідають вибраним поняттям.
7. Формування і візуалізація об’єднаної мережі понять і концептів.
8. Кластеризація подій на основі аналізу об’єднаної мережі.

Слід зазначити, що у той час, як система Gephi дозволяє візуалізувати, виявляти кластери, проводити підрахунок і аналіз параметрів сформованих Каузальних мереж, програми побудовані на основі GraphViz дозволяють створювати інтерактивні семантичні мапи за рахунок застосування графічного формату SVG. У цьому випадку кожний вузол або ребро мережі може містити гіперпосилання на ресурси мережі Інтернет, зокрема на пошукові системи Google, Google News, Bing, Bing News тощо, або корпоративні новинні системи.

Приклад застосування.

Розглянемо приклад, що стосується подій щодо забезпечення прозорості діяльності комітетів Верховної Ради України і їх інтерпретацій.

Для цього на першому етапі формується масив цільових документів, для чого у новинній базі даних виконується запит вигляду “комітет Верховної Ради прозорість”. Серед отриманих документів розглядається документ Укрінформу з таким вмістом:

Рух ЧЕСНО: Рейтинг прозорості очолили одразу сім комітетів ВР.

Громадська організація “Рух ЧЕСНО” презентувала рейтинг прозорості комітетів Верховної Ради за друге півріччя 2023 року, його очолили одразу сім парламентських комітетів. Про це повідомили учасники прес-конференції, що відбулась в Укрінформі.

“Більшість комітетів Верховної Ради суттєво покращила відкритість своєї роботи та інформування громадян. Наразі лише один комітет має найменшу кількість балів у нашому рейтингу. Проте ми бачимо, що найвищий бал мають одразу 7 парламентських комітетів”, - поінформував головний редактор Руху ЧЕСНО Олександр Саліженко ...

На другому етапі для знайдених документів застосовується запит (промпт):

В тексті описані деякі події. Детектуй їх, назви мені у вигляді нумерованого списку. Кожна подія - 4-5 слів
Ось текст: Рух ЧЕСНО: Рейтинг прозорості очолили одразу сім комітетів ВР. Громадська організація “Рух ЧЕСНО” презентувала рейтинг прозорості комітетів Верховної Ради за друге півріччя 2023 року, його очолили одразу сім парламентських комітетів. Про це повідомили учасники прес-конференції, що відбулась в Укрінформі ...

У результаті множинного виконання цього промпту у різних системах штучного інтелекту (Gemini, ChatGPT, DeepSeec) формується масив позначень подій, не всі серед яких є оригінальними. Тобто можливі повтори, перекази різними словами:

1. “Презентація рейтингу прозорості комітетів Верховної Ради за друге півріччя 2023 року”.
2. “Сім парламентських комітетів очолили рейтинг прозорості”.
3. “Головний редактор Руху ЧЕСНО відзначив найбільше зростання рейтингу у Комітеті з питань інтеграції України до ЄС”.
4. “Голова Комітету ВР з питань інтеграції України, Іванна Климпуш-Цинцадзе, наголосила на необхідності повернення до прозорої роботи парламенту та публікації порядку денних”.
5. “Зазначено важливість відновлення трансляцій засідань комітетів, Погоджувальної ради та самого парламенту” ...

На третьому етапі здійснюється фільтрація подій, тобто обираються оригінальні, для чого весь масив відібраних позначень подій надається системі GenAI та виконується промпт:

Із названих подій обері оригінальні і найважливіші. Дублікати вилучай. Ось події:

1. Презентація рейтингу прозорості комітетів Верховної Ради за друге півріччя 2023 року.
2. Сім парламентських комітетів очолили рейтинг прозорості.
3. Головний редактор Руху ЧЕСНО відзначив найбільше зростання рейтингу у Комітеті з питань інтеграції України до ЄС ...

На четвертому етапі система Генеративного штучного інтелекту формує причинно-наслідкові зв'язки серед виявлених оригінальних подій, для чого виконується промпт:

Пари взаємопов'язаних подій за принципом причина-наслідок. Ось події:

- презентація рейтингу прозорості комітетів Верховної Ради;
- більшість комітетів покращили відкритість своєї роботи;
- оновлення сайтів та відновлення онлайн-трансляцій;
- пропозиції щодо покращення прозорості роботи парламенту;
- рух ЧЕСНО презентував рейтинг прозорості комітетів Верховної Ради ...

Відповідь, зокрема, системи Gemini на цей промпт видає результат типу:

Рух ЧЕСНО презентував рейтинг прозорості комітетів Верховної Ради;

- презентація рейтингу прозорості комітетів Верховної Ради;
- покращення прозорості роботи комітетів;
- оновлення сайтів та відновлення онлайн-трансляцій;
- сім парламентських комітетів очолили рейтинг прозорості;
- більшість комітетів покращили відкритість своєї роботи ...

На п'ятому етапі відібрані події групуються у CSV-файл, записи якого мають формат "подія-причина; подія-наслідок". Цей файл може завантажуватись у програми аналізу і візуалізації мереж.

На Рис. 1 наведено інтерфейс введення даних у програмі CSV2Graph (<https://bigsearch.space/uli.html>), розробленій на базі бібліотеки GraphViz. У результаті виконання цієї програми формується відображення графу подій (Рис. 2) Кожне ребро і вузол цього графа є гіперпосиланням із відповідними запитами до систем Google або Bing.

CSV => Graph

Insert text - pairs of concepts separated by a semicolon:

```

Рух ЧЕСНО презентував рейтинг прозорості комітетів Верховної
Ради; Презентація рейтингу прозорості комітетів Верховної
Ради
Презентація рейтингу прозорості комітетів Верховної
Ради;Покращення прозорості роботи комітетів
Оновлення сайтів та відновлення онлайн-трансляцій;
Покращення прозорості роботи комітетів
Презентація рейтингу прозорості комітетів Верховної Ради;
Сім парламентських комітетів очолили рейтинг прозорості
Більшість комітетів покращили відкритість своєї роботи;
Презентація рейтингу прозорості комітетів Верховної Ради
Вимога відновлення доступу журналістів до Верховної Ради;
Наголошено на важливості відновлення трансляцій засідань
Зростання прозорості Комітету з питань інтеграції до ЄС;
Найвищий бал у рейтингу
Пропозиції щодо покращення прозорості роботи парламенту;
Дискусія про забезпечення більш якісної комунікації з
українським парламентом
Покращення прозорості роботи комітетів; 16 комітетів оновили
сайти
Покращення прозорості роботи комітетів; Більшість комітетів
відновили онлайн-трансляції засідань
Деякі комітети публікують матеріали англійською мовою;
Покращення прозорості роботи комітетів
Презентація рейтингу прозорості комітетів Верховної Ради;
Рейтинг очолили сім комітетів
Більшість комітетів покращили відкритість своєї роботи;16
комітетів оновили сайти.
Більшість комітетів покращили відкритість своєї роботи;
Деякі комітети публікують матеріали англійською мовою

```

Links: Google Google News Bing Bing News

Directed

Рис. 1. Інтерфейс введення отриманих даних.

Задача кластеризації подій потребує формування простору параметрів, у якості якого запропоновано застосовувати множину ключових слів, що відповідають подіям. Для виявлення ключових слів для кожної визначеної події було відпрацьовано відповідний промпт, що включав запит і назву події, наприклад:

Назви 10 ключових слів, пов'язаних з подією: Презентація рейтингу прозорості комітетів Верховної Ради

Після отримання множини ключових слів, кожна подія і кожне відповідне ключове слово розглядалась як вузли мережі, а зв'язками між ними виступало відношення приналежності ключового слова події (одне ключове слово при цьому може відноситись до декількох подій).

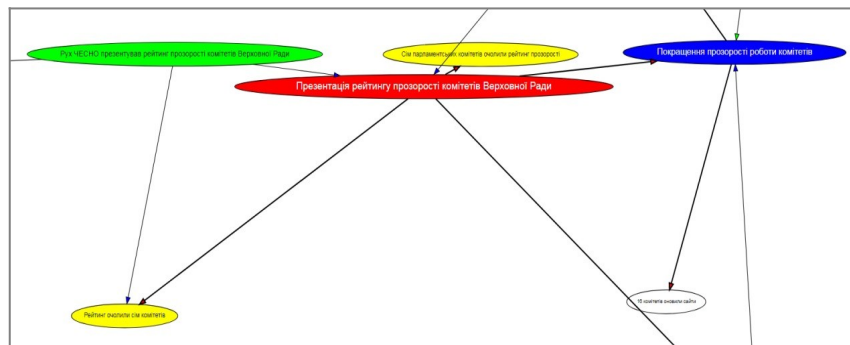


Рис. 2. Фрагмент графа, що відповідає мережі подій, сформований системою CSV2Graph, розроблений на базі бібліотеки GraphViz.

Крім того, у цій мережі враховувались визначені раніше причинно-наслідкові зв'язки між подіями. Цю мережа також спочатку було представлено як файл у форматі CSV, після чого її було завантажено у систему Gephi (<https://gephi.org>). У середовищі цієї системи проведено фільтрацію створеної мережі (вилучені всі вузли-листя із ступенем 1), а також кластеризацію за критерієм модулярності. Отриману мережа, що містить 39 вузлів і 149 зв'язків, наведено на Рис. 3, на якому кожному кластеру відповідає свій колір вузлів.



Рис. 3. Кластери в мережі подій у середовищі системи Gephi.

Після аналізу наведених результатів кластеризації експертним шляхом вибрано п'ять подій, що відповідають визначеним кластерам, а саме:

1. Рух ЧЕСНО презентував рейтинг прозорості комітетів Верховної Ради
2. Зростання прозорості Комітету з питань інтеграції до ЄС
3. Вимога відновлення доступу журналістів до Верховної Ради
4. Деякі комітети публікують матеріали англійською мовою
5. Оновлення сайтів та відновлення онлайн-трансляцій

Висновки.

У статті наведено методологію побудови і кластеризації мереж подій в новинних повідомленнях та приклад застосування цієї методології на основі Генеративного штучного інтелекту.

Завдяки використанню Генеративного штучного інтелекту отримані зручні методи екстрагування подій із текстів правової спрямованості, їх фільтрації, кластеризації. Виявлення причинно-наслідкових зв'язків також здійснюється із застосуванням штучного інтелекту, що значно спрощує роботу з природною мовою, алгоритми якої вбудовано у великі лінгвістичні моделі. Візуалізація і аналіз сформованих мереж здійснюється із застосуванням традиційних графічних засобів.

При цьому запропоновано формування інтерактивної функціональної семантичної мапи на основі застосування формату SVG і формування гіперпосилань до глобальних пошукових систем. На прикладі показано можливість і ефективність застосування наведеної методики для аналізу текстів, які мають пряме відношення до завдань парламентського контролю. Визначені головні події (центри кластерів) у розглянутому прикладі дійсно відповідають реалізації задач відкритості, прозорості діяльності комітетів Верховної Ради.

Використана література

1. [Samaneh Karimi](#), [Azadeh Shakeri](#) and [Rakesh M. Verma](#). Enhancement of Twitter event detection using news streams. *Natural Language Engineering*, Volume 29, Issue 2, March 2023. Pp. 181-200. DOI: <https://doi.org/10.1017/S1351324921000462>.
2. Lande D.V., Prishchepa S.V. The automatic detection of the information operations event basis. Preprint arxiv:1807.03360. DOI: <https://doi.org/10.48550/arxiv.1807.03360>.
3. P. Quaresma, T. Gonçalves. Using linguistic information and machine learning techniques to identify entities from juridical documents. In: Francesconi, E., Montemagni, S., Peters, W., Tiscornia, D. (eds). *Semantic Processing of Legal Texts. Lecture Notes in Computer Science*, vol 6036. Springer, Berlin, Heidelberg. URL: https://doi.org/10.1007/978-3-642-12837-0_3
4. Navas Loro, María (2021). *Processing, Identification and Representation of Temporal Expressions and Events in Legal Documents*. Thesis (Doctoral), E.T.S. de Ingenieros Informáticos (UPM). URL: <https://doi.org/10.20868/UPM.thesis.69772>.
5. Oleh Dmytrenko, Dmitry Lande, Oleh Andriichuk. Method for Searching of an Optimal Scenario of Impact in Cognitive Maps during Information Operations Recognition. Preprint arxiv: 1904.13308. DOI: <https://doi.org/10.48550/arXiv.1904.13308>.
6. Ken Cherven. *Mastering Gephi Network Visualization*. Packt Publishing, 2015. 378 p.
7. Lambert M. Surhone, Mariam T. Tennoe, Susan F. Henssonow. *Graphviz*. VDM Publishing, 2010. 108 p.

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