МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
УКРАЇНИ "КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ
ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО"

НАВЧАЛЬНО-НАУКОВИЙ
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ

THEORETICAL AND APPLIED CYBERSECURITY

**Матеріали другої Всеукраїнської
науково-практичної конференції**

Випуск 2

Київ – 2024

Рекомендовано до друку Вченою радою
КПІ ім. Ігоря Сікорського
(протокол № 14 від 12 червня 2024 р.)

Theoretical and Applied Cybersecurity. Матеріали другої всеукраїнської науково-практичної конференції (TACS-2024). – Київ: Інжиніринг. – 190 с. ISBN 978-966-2344-98-1

До збірника увійшли матеріали доповідей, представлених на другій всеукраїнській науково-практичній конференції Theoretical and Applied Cybersecurity (TACS-2024, 30 травня 2024 року, м. Київ, Україна).

У збірнику представлені матеріали, присвячені питанням кібернетичної безпеки критичних інфраструктур, моделювання та протидії інформаційним операціям, технологій інформаційно-аналітичних досліджень на основі відкритих джерел інформації. Наведені матеріали з актуальних проблем інформаційної та кібернетичної безпеки, можливості застосування штучного інтелекту, системного аналізу при забезпеченні підтримки прийняття рішень, комп'ютерному моделюванні процесів і систем, актуальні завдання забезпечення інформаційної та кібербезпеки.

Для фахівців в області інформаційних технологій, інформаційної і кібернетичної безпеки, а також для аспірантів і студентів старших курсів вищої школи відповідних спеціальностей.

ISBN 978-966-2344-98-1

ВИВЕДЕННЯ ТА АНАЛІЗ НАРАТИВІВ У ІНФОРМАЦІЙНИХ ПОТОКАХ ЗА ДОПОМОГОЮ ШТУЧНОГО ІНТЕЛЕКТУ

О.О. Гуменюк, А.В. Комар, І.М. Свобода, Д.В. Ланде

Навчально-науковий Фізико-технічний інститут, НТУУ «КПІ ім. Ігоря Сікорського», м.Київ, Україна
olehhumeniukba@gmail.com

Це дослідження пропонує інноваційний підхід до виведення та аналізу наративів у великих інформаційних потоках за допомогою штучного інтелекту. Використання алгоритмів машинного навчання та обробки природної мови (NLP) дозволяє автоматизувати процеси токенізації, екстракції подій та наративів. Інструменти штучного інтелекту дозволяють виявляти приховані зв'язки та структури в текстових даних, значно підвищуючи ефективність та точність аналізу. Результати дослідження демонструють потенціал ШІ у покращенні розуміння та управління інформаційними потоками в різних сферах життя.

Ключові слова: штучний інтелект, GenAI, обробка природної мови, машинне навчання, токенізація, подія, наратив, мережі подій, кластеризація, аналіз наративів, інформаційні потоки.

Вступ

Аналіз наративів в інформаційних потоках стає все більш актуальним в сучасному світі, де обсяги даних постійно зростають. Наративи, як форма представлення інформації, допомагають структурувати та інтерпретувати події, що відбуваються навколо нас [1, с.15]. Штучний інтелект (ШІ) відіграє ключову роль в обробці великих обсягів текстових даних та виявленні наративів, автоматизуючи процеси аналізу та дозволяючи виявляти приховані структури в текстах [2, с.34]. Застосування алгоритмів машинного навчання та NLP) відкриває нові

можливості для ефективного і точного аналізу наративів [3, с.50].

Мета цієї роботи – дослідити методи аналізу наративів у великих інформаційних потоках за допомогою ШІ. Основна увага зосереджена на створенні мереж подій і кластеризації для розуміння структури наративів та їх впливу на інформаційний контекст.[4, с.75]

Токенізація тексту

Токенізація — це основний етап обробки тексту, коли він розбивається на елементи, такі як слова, фрази або символи, для подальшої структуризації інформації. Вибір інструментів для токенізації, таких як бібліотеки NLTK та SpaCy, залежить від завдань аналізу. [1, с. 15].

$$F_{tokenize\_weighted}(t_i, w_i) = \{(t_i, w_i)\},$$

де t_i є токенами в тексті $T, w_i = f(t_i, T)$, w_i є вагою токена, яка може бути визначена через частоту, TF-IDF, або іншу метрику, що визначає важливість слова в контексті тексту.

Визначення понять і подій

Після токенізації текст аналізується на поняття та події, що включає використання методів машинного навчання та GenAI. Цей процес допомагає виявити ключові елементи, які формують основу наративів. [2, с. 34].

$$F_{concepts\_events\_context}(T, Context) = \{(e_j, C_j)\},$$

де e_j - події у тексті $T, C_j = g(e_j, Context)$, C_j представляє контекст події, що може включати залежності від інших подій, історичний контекст, або реляційні зв’язки.

Створення мережі подій

Методика побудови мережі подій полягає у з'єднанні подій, понять, слів та токенів для створення структур, які дозволяють візуалізувати та аналізувати взаємозв’язки між елементами тексту. Для цього був використаний інструмент Gephi. [4, с. 75].

$$F_{event\_netwok\_extended}(E) = \{(N, s(e_k, C_l))\}$$

де N є мережею $s(e_k, C_l)$ є силою звязку між e_k та e_l, $s(e_k, C_l)$ може визначатися через кількість спільних контекстів, спільних згадок, або інші відносини.

Для отримання наративів застосовувався GPT-4. Було проаналізовано новину «Контроль над IT-інфраструктурою допомагає США реалізувати кольорові революції в країнах – Радбез». З новини отримано 20 наративів, 20 подій та 20 концептів для кожного з них. Використовуючи ChatGPT-4, встановлено можливі причинно-наслідкові зв'язки між наративами та подіями.

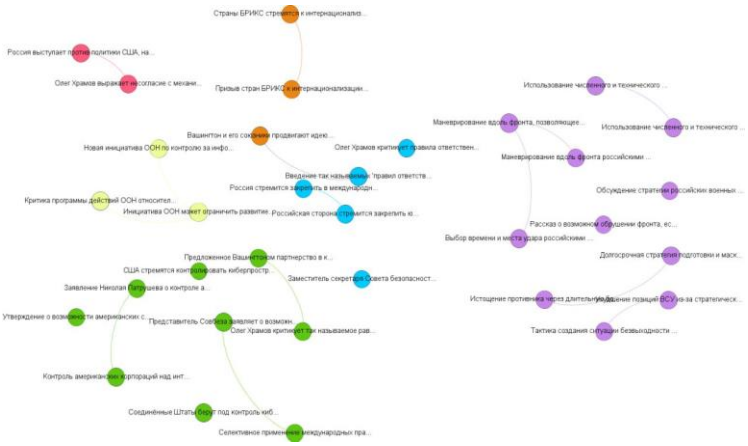

Рисунок 1 - Візуалізація кластеризованої мережі наративів і подій

Висновки

Дослідження підтвердило, що методи аналізу наративів на основі ШІ є потужним інструментом для обробки великих обсягів текстових даних. Це може бути корисним для наукової спільноти та практичного застосування у різних галузях, таких як журналістика, соціологія, політика та інші. Методологічний підхід, запропонований у цьому

дослідженні, відкриває нові перспективи для автоматизації та вдосконалення процесів аналізу інформаційних потоків, сприяючи кращому розумінню та управлінню наративами у сучасному світі.

Перелік використаних джерел

1. Salgado A., Zeng J., Abnar S. A Survey on Event-based News Narrative Extraction. — 2023-07-17. — P. 15.
2. Smith R., Doe J., Lee K. A Survey on Narrative Extraction from Textual Data. — 2023-01-06. — P. 34, 60.
3. Williams B., Chen Y., Patel R. Testing the Quantitative Spacetime Hypothesis using Artificial Narrative Comprehension. — 2020-09-23. — P. 50.
4. Martin J., Brown T., Doe P. Using Narrative Function to Extract Qualitative Information. — 2020-12-28. — P. 75.

ЗМІСТ