

УДК 004.8

DOI <https://doi.org/10.32782/IT/2024-4-17>

Марина МОЛЧАНОВА

викладач кафедри комп'ютерних наук, Хмельницький національний університет, вул. Інститутська, 11, Хмельницький, Україна, 29016

ORCID: 0000-0001-9810-936X

Scopus Author ID: 57456998700

Бібліографічний опис статті: Молчанова, М. (2024). Нейромережеве визначення цілей пропаганди у текстовому контенті з візуальною аналітикою результатів. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, 144–150, doi: <https://doi.org/10.32782/IT/2024-4-17>

НЕЙРОМЕРЕЖЕВЕ ВИЗНАЧЕННЯ ЦІЛЕЙ ПРОПАГАНДИ У ТЕКСТОВОМУ КОНТЕНТІ З ВІЗУАЛЬНОЮ АНАЛІТИКОЮ РЕЗУЛЬТАТІВ

Мета роботи полягає у створенні методу нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою результатів, який повинен включати розширення множини цілей пропаганди та використання контекстних вікон для аналізу взаємозв'язків між пропагандистськими прийомами та цілями.

Методологія. Розроблений метод передбачає аналіз текстового контенту за допомогою нейромереж, що дозволяє виявляти цілі та прийоми пропаганди. Використовуються контекстні вікна для оцінки взаємозв'язків між прийомами та цілями пропаганди. Результати представляються у вигляді візуалізації семантичної моделі пропаганди, що включає оцінку семантичної значущості та важливості взаємозв'язків.

Наукова новизна. Основною відмінністю запропонованого методу є розширення множини цілей пропаганди шляхом доповнення множини їх словесних варіантів. Впровадження контекстних вікон для аналізу взаємозв'язків між пропагандистськими прийомами та цілями дозволяє підвищити точність виявлення пропаганди та забезпечити наочне відображення результатів.

Висновки. Експериментальні результати підтвердили ефективність розробленого методу, який демонструє покращення виявлення цілей пропаганди у порівнянні з існуючими підходами. Висока кореляція з експертними оцінками підтверджує надійність і точність підходу. Візуальна аналітика забезпечує прозорість та пояснюваність результатів, що сприяє глибшому розумінню механізмів пропагандистського впливу. Метод має значний потенціал для використання в автоматичному аналізі текстового контенту та може бути адаптований для різних текстових джерел.

Ключові слова: пропаганда, цілі пропаганди, текстовий контент, візуальна аналітика.

Maryna MOLCHANOVA

Teacher at the Department of Computer Science, Khmelnytskyi National University, 11 Instytutska Str., Khmelnytskyi, Ukraine, 29016, m.o.molchanova@gmail.com

ORCID: 0000-0001-9810-936X

Scopus Author ID: 57456998700

To cite this article: Molchanova, M. (2024). Neiomerezheve vyznachennia tsilei propahandy u tekstovomu kontenti z vizualnoiu analytkoiu rezultativ [Neural network identification of propaganda goals in text content with visual analytics of results]. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, 144–150, doi: <https://doi.org/10.32782/IT/2024-4-17>

NEURAL NETWORK IDENTIFICATION OF PROPAGANDA GOALS IN TEXT CONTENT WITH VISUAL ANALYTICS OF RESULTS

The work aim is to create a method for neural network identification of propaganda goals in text content with visual analytics of the results, which should include expanding the set of propaganda goals and using context windows to analyze the relationships between propaganda techniques and goals.

Methodology. The developed method involves analyzing text content using neural networks, which allows identifying propaganda goals and techniques. Context windows are used to assess the relationships between propaganda techniques and goals. The results are presented in the form of visualization of the semantic model of propaganda, which includes assessing the semantic significance and importance of relationships.

Scientific novelty. The main difference of the proposed method is the expansion of the set of propaganda goals by supplementing the set of their verbal variants. The introduction of context windows to analyze the relationships between propaganda techniques and goals allows to increase the accuracy of propaganda detection and provide a visual display of the results.

Conclusions. Experimental results confirmed the effectiveness of the developed method, which demonstrates improved detection of propaganda targets compared to existing approaches. High correlation with expert assessments confirms the reliability and accuracy of the approach. Visual analytics provides transparency and explainability of the results, which contributes to a deeper understanding of the mechanisms of propaganda influence. The method has significant potential for use in automatic analysis of text content and can be adapted for various text sources.

Key words: propaganda, propaganda goals, text content, visual analytics.

Актуальність проблеми. У сучасному світі пропаганда є одним із найефективніших засобів маніпулювання масовою свідомістю, використовуючи текстовий контент для формування потрібних нарративів. Її цілі можуть варіюватися від політичного впливу до соціальних чи економічних трансформацій. Вона спрямована на різні цілі, серед яких – окремі особи, соціальні групи, організації (Nazarov V., 2023; Молчанова М.О., 2023).

Пропаганда діє через широкий спектр прийомів, що впливають на емоційний стан, когнітивні процеси та поведінкові реакції аудиторії. Важливість вивчення цього явища обумовлена не лише необхідністю ідентифікації цих прийомів, але й розумінням того, на кого вони спрямовані та які мають наслідки, тобто виникає потреба визначати цілі пропаганди (Молчанова М.О., 2024, С. 101-116).

Аналіз останніх досліджень. На сучасному етапі завдання ідентифікації пропаганди в текстовому контенті, як і інші аспекти семантичного аналізу, зокрема визначення токсичного змісту або аналізу емоційної складової текстів (Krak I., 2024, С. 16-28), вирішуються за допомогою сучасних методів обробки природної мови (Zalutskaya O., 2023, С. 344–356). У контексті виявлення пропагандистського впливу важливо не лише розпізнати самі пропагандистські прийоми, але й встановити, на які цілі або аудиторії вони спрямовані. Ідентифікація цілей пропаганди має вагомe значення для розкриття механізмів маніпуляції свідомістю, що робить це питання особливо актуальним для наукових досліджень.

У статті (Ahmad P. N., 2024, С. 1–15.) розглядається проблема зростання кількості пропагандистських текстів та активностей у цифровому просторі, що стало наслідком стрімкого розвитку соціальних медіа. Для вирішення цього питання запропоновано автоматизований підхід до визначення семантичних характеристик пропаганди у текстах новин. Метод ґрунтується на використанні мета-навчання для класифікації текстів на рівні речень і включає поєднання багатозадачного навчання, моделі CRF, BiLSTM

та попередньо натренованих мовних моделей. Ефективність підходу підтверджено на основі аналізу корпусу новинних текстів, де модель досягла F1-оцінки 0.61 для багатомовних даних і 0.688 для одномовних.

У статті (Rodríguez-García R., 2024) представлено дослідження визначення пропаганди в текстовому контенті. У двох із п'яти сценаріїв MTL перевершило інші методи, досягнувши F1-оцінки 0.78, що значно перевищує базовий результат трансформерів (0.68). Автори також рекомендують досліджувати вплив моральних і логічних аспектів для подальшого вдосконалення моделей.

У статті (Szwach J., 2024, С. 4330) досліджується застосування LLM, таких як GPT-4, для автоматичного визначення та класифікації пропагандистських прийомів у текстах електронної преси. Результати демонструють, що моделі здатні досягати високої точності в окремих аспектах, наприклад, 74% точності у бінарному виявленні пропаганди в польських текстах і 69% у класифікації прийомів. Однак загальні показники F1-оцінки були нижчими за базові рівні, коливаючись у межах 20-50%. Експерименти також показали, що результати моделей є нестабільними та важко відтворюваними.

Описані підходи до визначення пропаганди із застосуванням великих мовних моделей мають низку обмежень, які впливають на їх ефективність і практичну придатність. Однією з ключових проблем є нестабільність результатів, яка ускладнює відтворення отриманих даних. Багато експериментів показали, що F1-оцінка моделей суттєво поступається базовим підходам, що свідчить про їх обмежену точність у виявленні та класифікації пропаганди. Також описані підходи не містять візуальної аналітики результатів, що не викликає високий рівень довіри до отриманих результатів

Таким чином виникає потреба в застосуванні не тільки підходу на основі класифікації текстів, а й на основі визначення цілей пропаганди у текстовому контенті з візуальною аналітикою результатів.

Метою роботи є розробка методу нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою результатів, що дозволить визначати, на кого та на що орієнтовані конкретні пропагандистські прийоми, а також надавати візуальну інтерпретацію результатів.

Виклад основного матеріалу дослідження. Використання прийомів пропаганди в текстовому контенті може бути виявлене за допомогою сучасних нейромережевих підходів (Kraak I., 2024, С. 158–170), що водночас дозволяє створювати моделі, спеціалізовані на детекції конкретних пропагандистських прийомів. Запропонований метод забезпечує цілісний аналіз взаємозв'язків між прийомами пропаганди та їхніми цілями, а також спроможний узагальнювати альтернативні згадування цілей у текстах. Його унікальність полягає у поєднанні двох завдань: ідентифікації цілей пропаганди та кількісного співвіднесення цих цілей із виявленими пропагандистськими прийомами.

Запропонований метод передбачає використання 17 попередньо навчених нейромережевих моделей архітектури трансформер, кожна з яких спеціалізується на ідентифікації одного з основних пропагандистських прийомів, таких як «Appeal to fear-prejudice», «Causal Oversimplification», «Doubt», «Exaggeration», «Flag-Waving», «Labeling», «Loaded Language», «Minimisation», «Name Calling», «Repetition», «Appeal to Authority», «Black and White Fallacy», «Reductio ad Hitlerum», «Red Herring», «Slogans», «Thought Terminating Clichés» та «Whataboutism» (Yoosuf S., 2019, С. 87–91).

Навчання моделей здійснювалося на основі розмічених даних, підготовлених командою «Analysis Project» (Propaganda Analysis Project, 2024), яка провела детальний аналіз текстів, виявивши фрагменти, що містять зазначені прийоми, та анотувавши їх відповідно до типу. Для цієї мети створено корпус із 550 новинних статей (Zenodo, 2024), розмічених вручну на рівні кожного фрагмента для точного визначення застосованих прийомів пропаганди.

Метод нейромережевого визначення цілей пропаганди з візуальною аналітикою результатів базується на використанні моделей глибокого навчання та складається з кількох ключових етапів. Спочатку здійснюється розпізнавання іменованих сутностей для формування первинної множини цілей пропаганди. Далі проводиться попередня обробка тексту та розширення цієї множини шляхом додавання альтернативних словесних подань цілей. Наступним кроком є створення контекстних вікон для

кожного словесного подання, з урахуванням мінімального порогу розміру вікна. На основі цих контекстних вікон виконується оцінка рівня застосування пропагандистських прийомів із використанням нейромережевих моделей. Завершальним етапом є визначення значущих зв'язків між прийомами та цілями пропаганди, враховуючи мінімальний рівень прояву пропаганди, та подальше представлення результатів у візуальній формі. Схема та кроки методу нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою результатів наведено на рисунку 1.

Запропонований метод нейромережевого визначення цілей пропаганди базується на аналізі тестового тексту, множини прийомів пропаганди та набору спеціалізованих моделей, натренованих на визначення цих прийомів.

Перший крок передбачає пошук іменованих сутностей (NER) із видаленням повторів на рівні лем, що дозволяє отримати множину унікальних цілей пропаганди.

На другому кроці здійснюється розширення цієї множини за допомогою моделі «FastText», яка виявляє семантично схожі цілі, розширюючи аналіз за межі класичних NER. Використання «FastText», донавченої на пропагандистських текстах. У результаті цього кроку формується ширша множина цілей пропаганди завдяки включенню різних варіантів їх словесних подань. Параметр мінімальної семантичної близькості встановлюється залежно від конкретної задачі та визначається експериментально. У рамках цього дослідження порогове значення не використовувалося.

Третій крок зосереджений на формуванні контекстних вікон, які представляють речення, що містять цілі пропаганди. Якщо у межах одного контекстного вікна трапляється кілька різних цілей пропаганди, то таке вікно не дублюється. Однак, якщо в контекстному вікні присутні різні словесні варіанти однієї й тої ж самої цілі пропаганди, то таке вікно буде дублюватися.

На четвертому кроці виконується аналіз контекстних вікон нейромережевими моделями для визначення рівня застосування пропагандистських прийомів. Контент вікон векторизується, а моделі оцінюють його відповідність конкретним прийомам, виявленим у тексті.

Завершальним, п'ятим кроком, є побудова множини значущих зв'язків між прийомами та цілями пропаганди. При цьому враховується порогове значення, нижче якого прийом не вважається застосованим до відповідної групи цілей.

Результатом методу є візуальна аналітика семантичної моделі пропаганди для тексту,

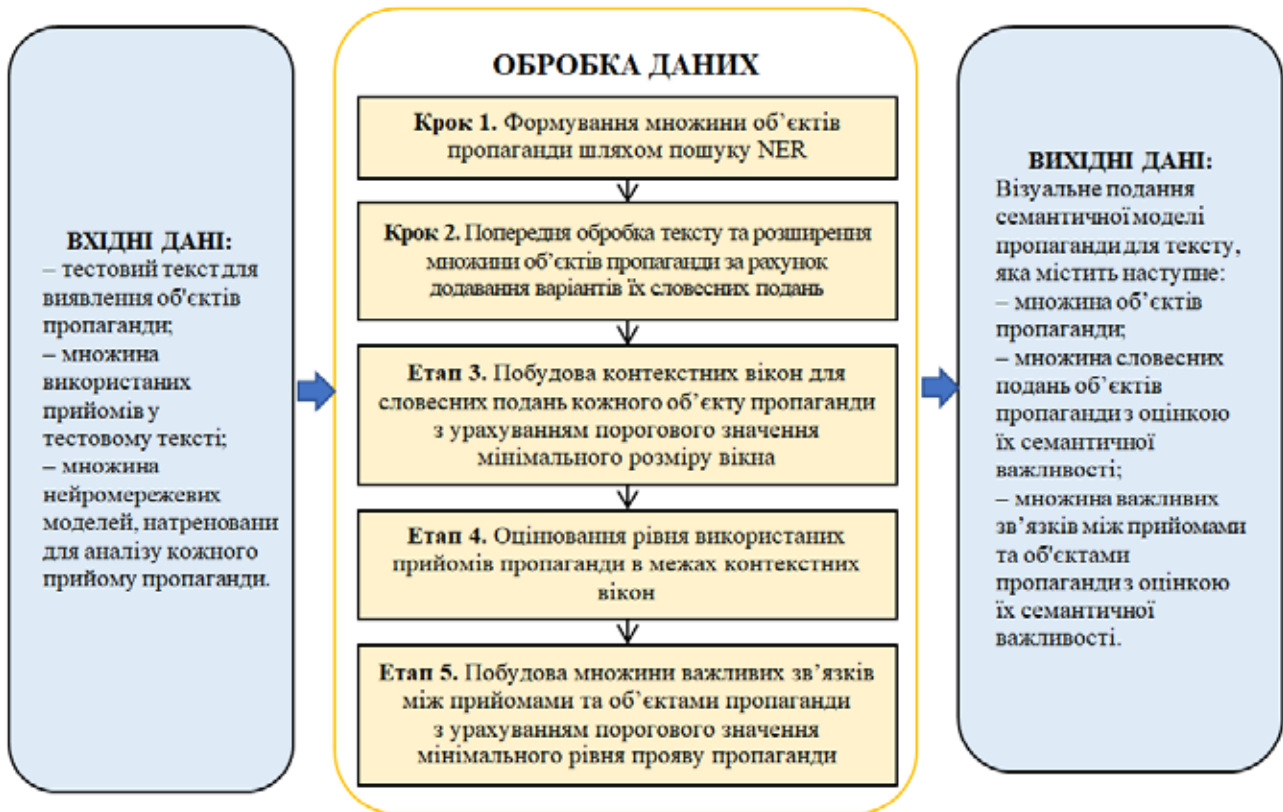


Рис. 1. Схема та кроки методу нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою

яка демонструє сукупність цілей пропаганди, їхні словесні варіанти з відповідними оцінками семантичної значущості, а також ключові зв'язки між пропагандистськими прийомами і цілями пропаганди із зазначенням рівня їхньої семантичної важливості.

У результаті було запропоновано метод, який забезпечує нейромережеве визначення цілей пропаганди у текстовому контенті, визначення їхньої належності до застосованих у тексті пропагандистських прийомів, а також розширення множини цілей пропаганди за рахунок додавання їхніх словесних варіантів. Використання контекстних вікон дозволило встановити взаємозв'язки між пропагандистськими прийомами та цілями, що сприяло підвищенню ефективності ідентифікації цілей пропаганди та забезпечило можливість візуальної аналітики результатів.

Експеримент, результати та дискусія.

З метою оцінки ефективності запропонованого методу нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою було створено спеціалізоване програмне забезпечення. Розроблене програмне забезпечення дозволяло ідентифікувати цілі пропаганди та аналізувати їх у контексті

застосованих прийомів пропаганди. Результати, представлені у вигляді візуальної аналітики, порівнювалися з висновками авторитетних джерел та експертів у сфері визначення і протидії пропаганді.

Для аналізу було використано пости із соціальних мереж, які попередньо опрацював «Центр стратегічних комунікацій» (<https://spravdi.gov.ua/>). Дані матеріали вже містили експертні оцінки та мітки, що дозволило зіставити їх із даними, отриманими завдяки запропонованому методу. Створене програмне забезпечення реалізує запропонований метод нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою. На рисунку 2 наведено покроковий опис запропонованого методу.

З огляду на вміст тестового тексту, прийоми пропаганди майже рівномірно розподілені по контекстним вікнам. Проте, для першого контекстного вікна прийом «Loaded Language» має менш виражений характер, тоді як прийом «Appeal to fear/prejudice» виявляється більш вираженою порівняно з другим контекстним вікном.

При навчанні нейромережевих моделей архітектури трансформер, заснованих на перенавчених моделях BERT, було досягнуто

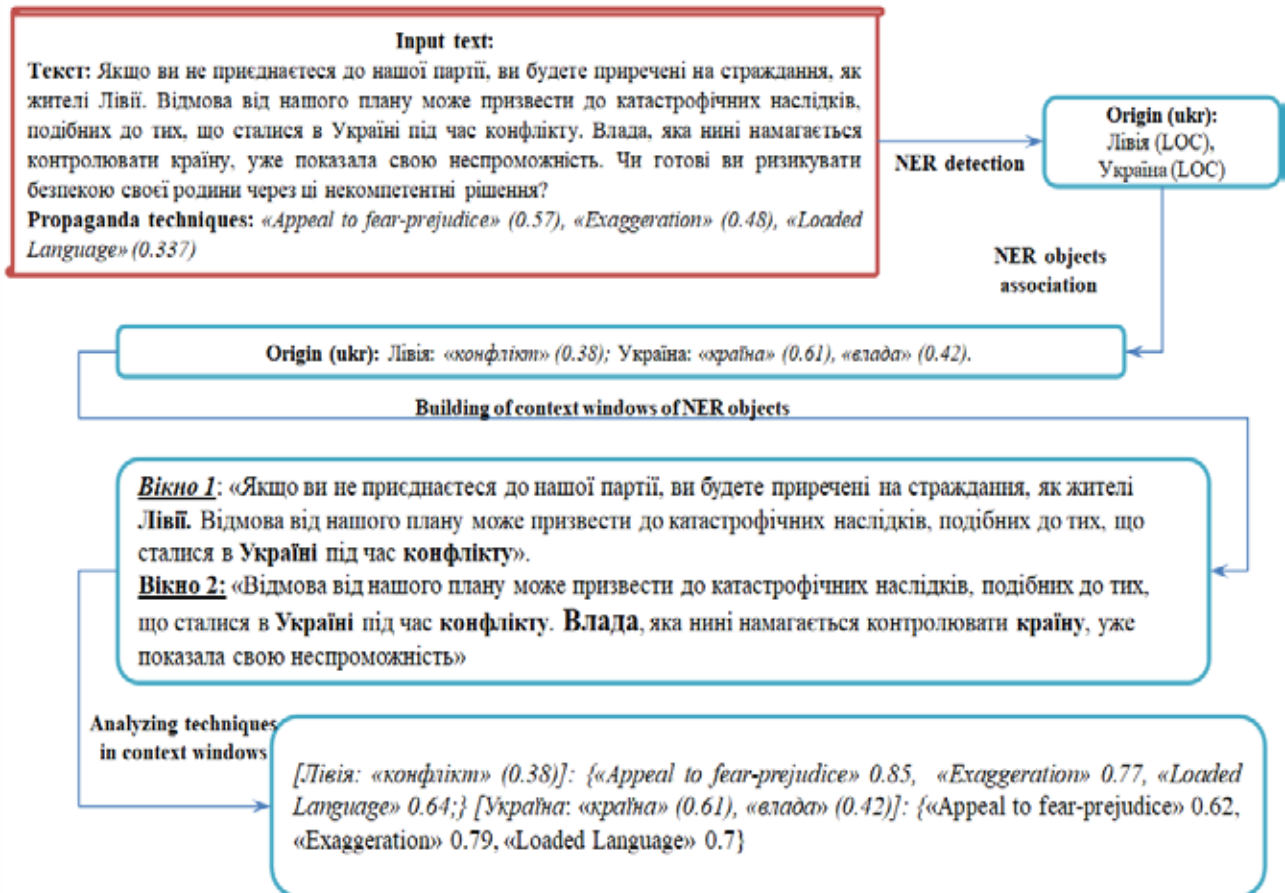


Рис. 2. Приклад перетворення даних цілей пропаганди за методом нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою

точності понад 80% для визначення прийомів пропаганди. Результати точності моделей BERT представлені в таблиці 1.

Таблиця 1

Точність визначення прийомів пропаганди моделями нейронних мереж

Прийом пропаганди	Accurasy
«Appeal to fear-prejudice»	0.87
«Causal Oversimplification»	0.82
«Doubt»	0.93
«Exaggeration»	0.8
«Flag-Waving»	0.92
«Labeling»	0.96
«Loaded Language»	0.97
«Minimisation»	0.9
«Name Calling»	0.92
«Repetition»	0.94
«Appeal to Authority»	0.89
«Black and White Fallacy»	0.91
«Reductio ad hitlerum»	0.87
«Red Herring»	0.8
«Slogans»	0.86
«Thought terminating Cliches»	0.8
«Whataboutism»	0.79

Приклад виконаного аналізу за допомогою розробленого програмного забезпечення показано на рисунку 3.

Як видно з рисунку 3, було виявлено не тільки цілі пропаганди та використані прийоми, а й визначено зв'язок між знайденими цілями та прийомами, до яких вони належать.

Отже, в результаті дослідження ефективності запропонованого методу було встановлено, що метод нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою забезпечує результати, які мають високу кореляцію з висновками експертів. Крім того, завдяки розгляду пропаганди як цілісної моделі та використанню візуальної аналітики отриманих результатів, вдалося здійснити комплексний аналіз взаємозв'язків між прийомами та цілями пропаганди, а також забезпечити узагальнення для цілей пропаганди та їх альтернативних згадок у текстах.

Висновки. Розроблено метод нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою результатів. Основною відмінністю цього методу є розширення множини цілей пропаганди через

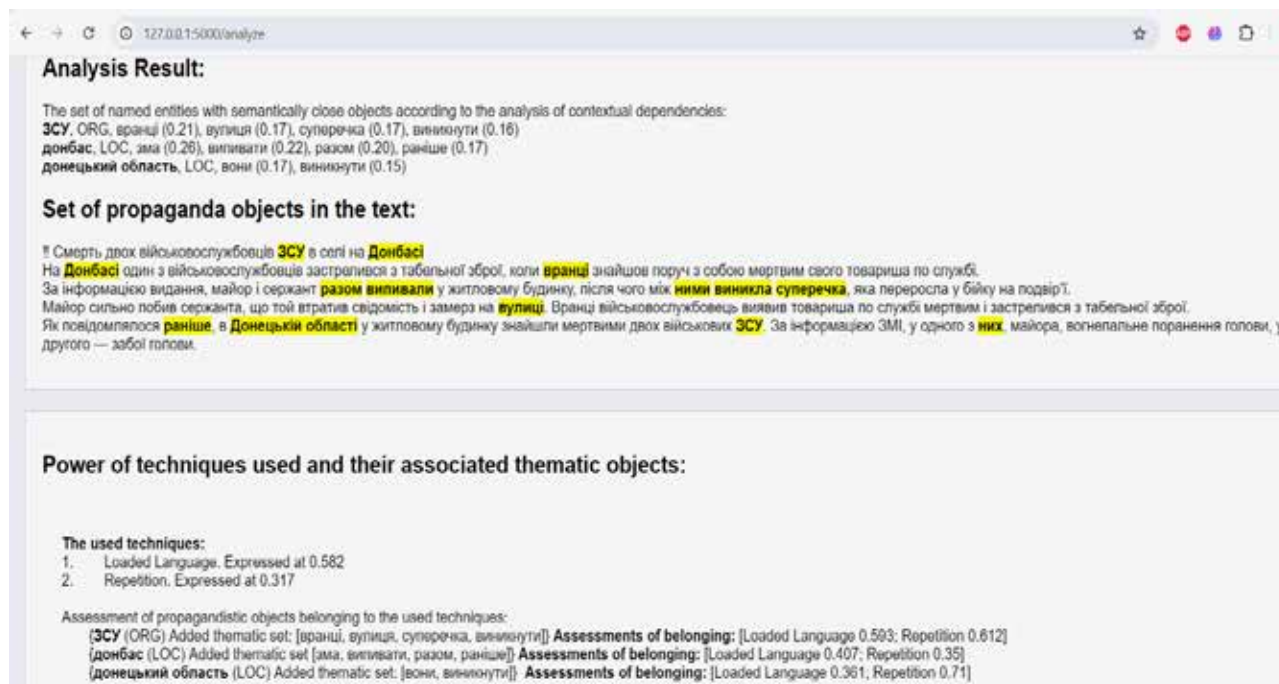


Рис. 3. Аналіз тестового тексту з пропагандою розробленим програмним забезпеченням

додавання варіантів їх словесного вираження та використання контекстних вікон для аналізу зв'язків між прийомами й цілями пропаганди, що дозволило значно покращити якість визначення та забезпечити наочне представлення результатів.

У межах основної мети підвищення точності та якості визначення прийомів і цілей пропаганди за семантичними маркерами у текстовому контенті з використанням нейронних мереж та візуальною аналітикою результатів було розроблено метод, що дозволяє визначити цілі пропаганди та встановлювати, на кого або на що спрямовані пропагандистські прийоми. Метод вирішує ключові проблеми, зокрема відсутність комплексного аналізу взаємозв'язків між прийомами та цілями пропаганди, а також

нестачу узагальнень і альтернативних згадок цілей у текстах.

Експериментально підтверджено ефективність запропонованого методу, який, на відміну від існуючих аналогів, поєднує пошук іменованих сутностей за допомогою бібліотеки «STANZA» із розширенням переліку цілей пропаганди завдяки бібліотеці машинного навчання «FastText». Крім того, метод оцінює, як знайдені цілі пов'язані з пропагандистськими прийомами. Дослідження показали, що результати, отримані за допомогою цього методу, узгоджуються з експертними оцінками. Для наочного представлення результатів надається візуальна аналітика, яка дозволяє визначити цілі впливу у контексті відповідних пропагандистських прийомів.

ЛІТЕРАТУРА:

1. Nazarov V., Molchanova M. Information System for Detecting Abusive Speech in Audio Content by Means of Natural Language. *Proceedings of V International Scientific and Practical Conference «Modern strategies of global scientific solutions»*. Stockholm, Sweden: International Scientific Unity, 2023. С. 132–135
2. Молчанова М. О., Мазурець О. В., Собко О. В. Віт Р. В., Назаров, В. В. Алгоритм виявлення аб'юзивного вмісту в україномовному аудіоконтенті для імплементації в об'єктно-орієнтовану інформаційну систему. *Вісник Хмельницького національного університету. Технічні науки*. 2024. № 1 (331). С. 101–106.
3. Молчанова М. О., Залуцька О. О., Бармак О. В. Метод інтелектуального аналізу тональності текстів. *Матеріали XII Всеукраїнської науково-практичної конференції «Глушковські читання»*. Київ, 2023. С. 113–116
4. Krak I., Zalutka O., Molchanova M., Mazurets O., Bahrii R., Sobko O., Barmak O. Abusive Speech Detection Method for Ukrainian Language Used Recurrent Neural Network. *CEUR Workshop Proceedings*. 2024. Vol. 3688. С. 16–28.

5. Zalutska O., Molchanova M., Sobko O., Mazurets O., Pasichnyk O., Barmak O., Krak I. Method for Sentiment Analysis of Ukrainian-Language Reviews in E-Commerce Using RoBERTa Neural Network. *CEUR Workshop Proceedings*. 2023. Vol. 3387. C. 344–356
6. Ahmad P. N., Yuanhao L., Aurangzeb K., Anwar M. S., Haq Q. M. U. Semantic web-based propaganda text detection from social media using meta-learning. *Service Oriented Computing and Applications*. 2024. C. 1–15.
7. Rodríguez-García R., Centeno R., Rodrigo Á. Together we can do it! A roadmap to effectively tackle propaganda-related tasks. *Internet Research*. 2024
8. Szwoch J., Staszko M., Rzepka R., Araki K. Limitations of Large Language Models in Propaganda Detection Task. *Applied Sciences*. 2024. T. 14, № 10. C. 4330
9. Krak I., Molchanova M., Mazurets O., Sobko O., Zalutska O., Barmak O. Method for Neural Network Detecting Propaganda Techniques by Markers With Visual Analytic. *CEUR Workshop Proceedings*. 2024. Vol. 3790. C. 158–170.
10. Yoosuf S., Yang Y. Fine-Grained Propaganda Detection with Fine-Tuned BERT. Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda. Hong Kong, China, 2019. C. 87–91.
11. Propaganda Analysis Project. URL: <https://propaganda.math.unipd.it/index.html>. Дата звернення (20.11.2024)/
12. Zenodo. Propaganda. Propy Corpus 1.0. URL: <https://zenodo.org/records/3271522#.XS6qRUUzau4>. Дата звернення (20.11.2024)

REFERENCES:

1. Nazarov, V., & Molchanova, M. (2023). Information system for detecting abusive speech in audio content by means of natural language. *Proceedings of the V International Scientific and Practical Conference «Modern strategies of global scientific solutions»*, Stockholm, Sweden: International Scientific Unity, 132–135.
2. Molchanova, M. O., Mazurets, O. V., Sobko, O. V., Vit, R. V., & Nazarov, V. V. (2024). Alhorytm vyjavlennia abuzyvnoho vmistu v ukrainomovnomu audiokontenti dlia implementatsii v obiektno-orientovanu informatsiinu systemu [Algorithm for detecting abusive content in Ukrainian-language audio content for implementation in an object-oriented information system]. *Visnyk Khmelnytskoho natsionalnoho universytetu. Tekhnichni nauky*, 1(331), 101–106 [in Ukrainian].
3. Molchanova, M. O., Zalutska, O. O., & Barmak, O. V. (2023). Metod intelektualnoho analizu tonalnosti tekstiv [Method of intellectual analysis of text tone]. *Materialy XII Vseukrainskoi naukovo-praktychnoi konferentsii «Hlushkovski chytannia»*, 113–116 [in Ukrainian].
4. Krak, I., Zalutska, O., Molchanova, M., Mazurets, O., Bahrii, R., Sobko, O., & Barmak, O. (2024). Abusive speech detection method for Ukrainian language using recurrent neural network. *CEUR Workshop Proceedings*, 3688, 16–28.
5. Zalutska, O., Molchanova, M., Sobko, O., Mazurets, O., Pasichnyk, O., Barmak, O., & Krak, I. (2023). Method for sentiment analysis of Ukrainian-language reviews in e-commerce using RoBERTa neural network. *CEUR Workshop Proceedings*, 3387, 344–356.
6. Ahmad, P. N., Yuanhao, L., Aurangzeb, K., Anwar, M. S., & Haq, Q. M. U. (2024). Semantic web-based propaganda text detection from social media using meta-learning. *Service Oriented Computing and Applications*, 1–15.
7. Rodríguez-García, R., Centeno, R., & Rodrigo, Á. (2024). Together we can do it! A roadmap to effectively tackle propaganda-related tasks. *Internet Research*.
8. Szwoch, J., Staszko, M., Rzepka, R., & Araki, K. (2024). Limitations of large language models in propaganda detection task. *Applied Sciences*, 14(10), 4330.
9. Krak, I., Molchanova, M., Mazurets, O., Sobko, O., Zalutska, O., & Barmak, O. (2024). Method for neural network detecting propaganda techniques by markers with visual analytic. *CEUR Workshop Proceedings*, 3790, 158–170.
10. Yoosuf, S., & Yang, Y. (2019). Fine-grained propaganda detection with fine-tuned BERT. Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda, Hong Kong, China, 87–91.
11. Propaganda Analysis Project. (2024, November 20). Retrieved from: <https://propaganda.math.unipd.it/index.html>.
12. Zenodo. (2019, November 20). Propaganda. Propy Corpus 1.0. Retrieved from: <https://zenodo.org/records/3271522#.XS6qRUUzau4>.