

РЕФЕРАТ

Актуальність теми. Актуальність теми полягає в необхідності пришвидшення процесу підсумовування тексту без втрат по точності. У сучасній інформаційній епосі, коли потоки текстової інформації зростають експоненційно, важливо мати засоби для швидкого оброблення текстових документів. Розробка системи, яка поєднує метод TopicRank та Text-to-Text трансформерну нейронну мережу для швидкого та точного підсумовування тексту, є актуальною і важливою з наукового та практичного погляду. Такий підхід може покращити продуктивність та зручність роботи з текстовою інформацією у різних галузях.

Об'єкт дослідження: методи генерування узагальнень з об'ємних текстових даних.

Предмет дослідження: оптимізація процесу швидкого створення текстових узагальнень зі збереженням точності семантичного змісту.

Мета роботи: дослідження існуючих методів аналізу текстів; аналіз особливостей алгоритмічних методів та методів глибокого навчання для підсумовування; створення системи підсумовування великих текстів.

Наукова новизна:

- Запропоновано новий метод для створення підсумків з об'ємних текстів.
- Розроблено систему, яка за запропонованим методом швидше генерує текстові підсумки, при цьому зберігаючи високу точність.

Практична цінність отриманих результатів презентує те, що запропонований метод дає змогу підвищити швидкість процесу підсумовування тексту, при цьому зберігаючи певну семантичну точність. Ця система може бути застосована в різних областях, включаючи медицину, наукові дослідження, новинарство та багато інших. Результати дослідження

готові до використання, і рекомендації щодо їх практичного застосування можуть бути корисними для організацій та індивідуальних користувачів, які шукають швидкий та ефективний спосіб автоматичного підсумовування тексту.

Апробації роботи:

1. XVI наукова конференція магістрантів та аспірантів «Прикладна математика та комп'ютинг» ПМК-2023 (Київ, 28-30 квітня 2023 р.)
2. XII International Scientific and Practical Conference "SCIENTIFIC RESEARCH IN THE MODERN WORLD" (Toronto, 21-23 September 2023)
3. I International Scientific and Practical Conference "MODERN RESEARCH IN SCIENCE AND EDUCATION" (14-16 September 2023)

Публікації:

1. Суханюк І. С., Івасенко Д. В., Потапова К. Р. Використання нейронних мереж для аналізу текстів // Міжнародний науковий журнал "Інтернаука". — 2023. — №13. <https://doi.org/10.25313/2520-2057-2023-13-9036>
2. Sukhaniuk I. S., Potapova K. R. USAGE OF CONVOLUTIONAL NEURAL NETWORKS IN NATURAL LANGUAGE PROCESSING. p. 233. URL: <https://sci-conf.com.ua/wp-content/uploads/2023/09/SCIENTIFIC-RESEARCH-IN-THE-MODERN-WORLD-21-23.09.23.pdf>
3. Sukhaniuk I. S., Potapova K. R. USAGE OF RECURRENT NEURAL NETWORKS IN NATURAL LANGUAGE PROCESSING. p. 67. URL: <https://sci-conf.com.ua/wp-content/uploads/2023/09/MODERN-RESEARCH-IN-CIENCE-AND-EDUCATION-14-16.09.23.pdf>
4. Sukhaniuk I.S., Potapova K. R., Nalyvaichuk M. V., Vovk L. B. TEXT SUMMARIZATION BASED ON TOPICRANK METHOD AND TEXT-TO-TEXT TRANSFORMER NEURAL NETWORK // Науковий журнал

«Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія: Технічні науки», Томі 34 (73) № 6, 2023

5. Суханюк І. С., Потапова К. Р. СИСТЕМА УЗАГАЛЬНЕННЯ ТЕКСТУ НА БАЗІ МЕТОДУ TOPICRANK ТА ТРАНСФОРМЕРНОЇ НЕЙРОННОЇ МЕРЕЖІ ТЕХТ-ТО-ТЕХТ // XVI наукова конференція магістрантів та аспірантів «Прикладна математика та комп'ютинг» ПМК-2023 (Київ, 28-30 листопада 2023 р.)
6. Свідоцтво №с202306573 авторського права на твір наукового характеру, «ВИКОРИСТАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ АНАЛІЗУ ТЕКСТІВ» (Київ, 06 вересня 2023 р.)
7. Свідоцтво №с202307642 авторського права на твір наукового характеру, «Програмний модуль «Text Topic Transformer»» (Київ, 30 жовтня 2023 р.)

Структура та обсяг роботи. Магістерська дисертація складається з чотирьох розділів та висновків.

У першому розділі надано загальну характеристику, проведено оцінку актуальності проблеми та сучасного стану досліджень. Визначено мету та поставлені завдання досліджень. Проаналізовано існуючі методи підсумовування.

У другому розділі представлено детальний огляд технологій та методів, застосованих для реалізації модифікацій у проекті. Увага приділена перевагам кожного із використаних інструментів та методів, а також тому, як вони сприяють поліпшенню загальної функціональності та ефективності роботи системи.

У третьому розділі наведено запропонований метод підсумовування великих текстів. Описана програмна реалізація методу.

У четвертому розділі приводено результати експериментальних досліджень.

У висновках представлено результати проведеної роботи.

Робота представлена на 86 аркушах, містить посилання на список використаних літературних джерел.

Ключові слова: обробка природної мови (NLP), аналіз текстів за допомогою нейронних мереж, трансформерні нейронні мережі, алгоритм TopicRank, згортковий нейронні мережі, рекурентні нейронні мережі.

ABSTRACT

Actuality of theme. The relevance of the topic lies in the need to expedite the text summarization process without sacrificing accuracy. In today's information era, where textual information streams are growing exponentially, it is crucial to have tools for rapid processing of textual documents. The development of a system that combines the TopicRank method and Text-to-Text transformer neural network for fast and accurate text summarization is relevant and important from both a scientific and practical perspective. Such an approach can enhance productivity and convenience when working with textual information in various fields.

The object of the research: Methods of generating summaries from extensive textual data.

The subject of the research: Optimization of the process of creating fast text summaries while preserving the accuracy of semantic content.

The goal of the work: Investigation of existing text analysis methods; analysis of the features of algorithmic and deep learning methods for summarization; creation of a system for summarizing large texts.

Scientific novelty:

- A new method has been proposed for creating summaries from extensive texts.
- A system has been developed that generates text summaries faster using the proposed method while maintaining high accuracy.

The practical value: The presented method allows for speeding up the text summarization process while preserving semantic accuracy. This system can be applied in various fields, including medicine, scientific research, journalism, and many others. The research results are ready for use, and recommendations for

their practical application can be beneficial to organizations and individual users seeking a fast and efficient way of automatic text summarization.

Work approbation:

1. XVI Scientific Conference of Master's and Ph.D. Students "Applied Mathematics and Computing" PMK-2023 (Kyiv, April 28-30, 2023)
2. XII International Scientific and Practical Conference "SCIENTIFIC RESEARCH IN THE MODERN WORLD" (Toronto, September 21-23, 2023)
3. I International Scientific and Practical Conference "MODERN RESEARCH IN SCIENCE AND EDUCATION" (September 14-16, 2023)

Publications:

1. Суханюк І. С., Івасенко Д. В., Потапова К. Р. Використання нейронних мереж для аналізу текстів // Міжнародний науковий журнал "Інтернаука". — 2023. — №13. <https://doi.org/10.25313/2520-2057-2023-13-9036>
2. Sukhaniuk I. S., Potapova K. R. USAGE OF CONVOLUTIONAL NEURAL NETWORKS IN NATURAL LANGUAGE PROCESSING. p. 233. URL: <https://sci-conf.com.ua/wp-content/uploads/2023/09/SCIENTIFIC-RESEARCH-IN-THE-MODERN-WORLD-21-23.09.23.pdf>
3. Sukhaniuk I. S., Potapova K. R. USAGE OF RECURRENT NEURAL NETWORKS IN NATURAL LANGUAGE PROCESSING. p. 67. URL: <https://sci-conf.com.ua/wp-content/uploads/2023/09/MODERN-RESEARCH-IN-CIENCE-AND-EDUCATION-14-16.09.23.pdf>
4. [Sukhaniuk I.S., Potapova K. R., Nalyvaichuk M. V., Vovk L. B. TEXT SUMMARIZATION BASED ON TOPICRANK METHOD AND TEXT-TO-TEXT TRANSFORMER NEURAL NETWORK // Науковий журнал](#)

[«Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія: Технічні науки», Томі 34 \(73\) № 6, 2023](#)

5. [Sukhaniuk I. S., Potapova K. R. TEXT SUMMARIZATION SYSTEM BASED ON TOPICRANK METHOD AND TEXT-TO-TEXT TRANSFORMER NEURAL NETWORK // XVI Scientific Conference of Master's and Ph.D. Students "Applied Mathematics and Computing" PMK-2023 \(Kyiv, November 28-30, 2023\)](#)
6. Свідоцтво №с202306573 авторського права на твір наукового характеру, «ВИКОРИСТАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ АНАЛІЗУ ТЕКСТІВ» (Київ, 06 вересня 2023 р.)
7. Свідоцтво №с202307642 авторського права на твір наукового характеру, «Програмний модуль «Text Topic Transformer»» (Київ, 30 жовтня 2023 р.)

Structure and scope of the work. The master's thesis consists of four chapters and conclusions.

The first section provides a general description, assesses the relevance of the problem and the current state of research. The purpose and objectives of the research are defined. The scientific novelty and practical value of the results obtained are noted, as well as information on the testing and the possibility of implementing the data obtained.

In *the second section*, a detailed overview of the technologies and methods applied for implementing modifications in the project is presented. Attention is given to the advantages of each of the tools and methods used, as well as how they contribute to improving the overall functionality and efficiency of the system. *The third section* presents the proposed method of summarizing large texts.

The fourth section presents the results of experimental studies on different sets of texts.

The conclusions present the results of the work carried out.

The work is presented on 86 pages and contains references to the list of used literary sources.

Keywords: natural language processing (NLP), text analysis using neural networks, transformer neural networks, TopicRank algorithm.