

РОЗРОБКА ІНТЕРАКТИВНОГО ВЕБ-ДОДАТКУ «ІНТЕЛЕКТУАЛЬНИЙ ЕКЗАМЕНАТОР» НА ОСНОВІ LLM ТА API GROQ

Вінницький національний технічний університет

Анотація

У роботі представлено розробку інформаційної технології для автоматизації процесу створення перевірочних тестів на основі довільного навчального матеріалу. Запропоновано архітектуру веб-додатка, що забезпечує вилучення тексту з різних джерел (PDF, HTML-сторінки) та його обробку за допомогою великих мовних моделей родини LLaMa через платформу API Groq. Описано механізм динамічного парсингу відповідей ШІ за допомогою регулярних виразів для формування інтерактивного інтерфейсу на базі фреймворку Streamlit. Проведене тестування підтвердило високу швидкість генерації завдяки використанню LPU-процесорів та виявило специфіку токенизації україномовного контенту в умовах обмежень безкоштовних тарифів API.

Ключові слова: інтелектуальний екзаменатор; великі мовні моделі; LLaMa; API Groq; Streamlit; автоматизація тестування; обробка природної мови.

Abstract

The paper presents the development of an information technology for automating the process of creating check tests based on arbitrary educational material. A web application architecture is proposed that ensures text extraction from various sources (PDF, HTML pages) and its processing using large language models of the LLaMa family via the Groq API platform. A mechanism for dynamic parsing of AI responses using regular expressions to form an interactive interface based on the Streamlit framework is described. The testing confirmed the high speed of generation due to the use of LPU processors and revealed the specifics of tokenization of Ukrainian-language content under the conditions of free API rate limits.

Keywords: intelligent examiner; large language models; LLaMa; Groq API; Streamlit; testing automation; natural language processing.

Вступ

Процес підготовки контрольних матеріалів є важливою складовою сучасної освіти, проте традиційний ручний аналіз тексту та формування запитань вимагає значних витрат часу [1]. Поява високоточних великих мовних моделей, зокрема родини LLaMa відкрила можливість створення систем, що здатні глибоко аналізувати природномовний контекст та автоматично генерувати структуровані тестові завдання [2]. У таких розробках ключову роль відіграє зручний інтерактивний інтерфейс, який дозволяє швидко перетворювати неструктуровані дані на візуальні блоки для взаємодії з користувачем [3].

Метою роботи є створення програмних основ інформаційної технології, що забезпечує автоматизовану генерацію перевірочних тестів на основі веб-додатка, здатного вилучати навчальний матеріал з різних джерел, формувати запитання за допомогою API Groq та виводити готовий для проходження інтерактивний інтерфейс.

Результати дослідження

Для реалізації інтерактивного веб-додатку «Інтелектуальний екзаменатор» розроблено архітектуру, що базується на трьох основних модулях: збору даних, промпт-інжинірингу та візуалізації результатів.

Перший етап роботи системи передбачає багатоканальне вилучення навчального матеріалу. Для обробки PDF-документів застосовано бібліотеку PyPDF2, а для автоматизованого збору даних із веб-ресурсів (веб-скрапінгу) — бібліотеку BeautifulSoup, що дозволяє отримувати текст та очищувати HTML-код від технічних тегів [4]. Інтерфейс вибору джерела матеріалу та завантаження файлів

зображено на рис. 1.

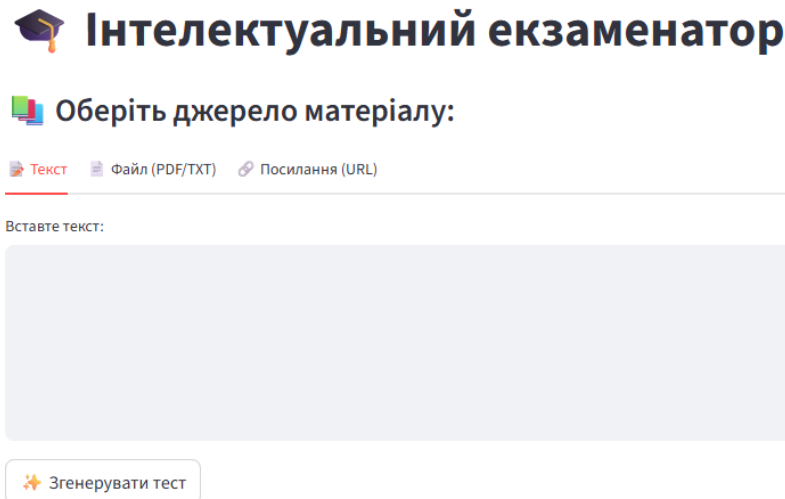
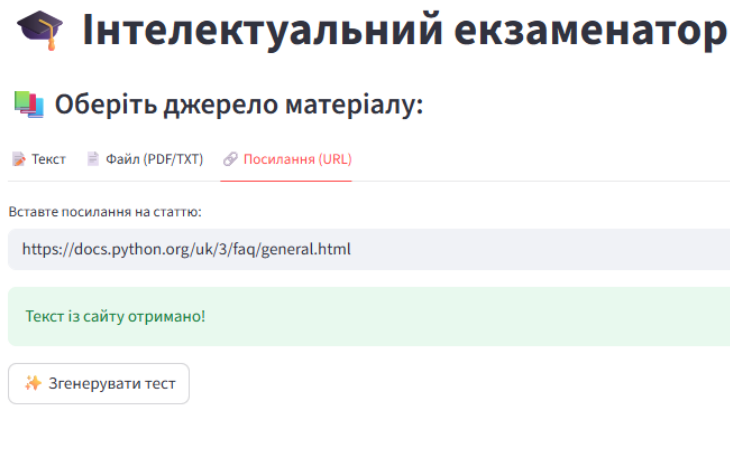


Рис. 1. Інтерфейс вибору джерела матеріалу та завантаження файлів

Центральним елементом системи є взаємодія з хмарною платформою Groq через API. Використання спеціалізованих процесорів LPU (Language Processing Units) забезпечує надвисоку швидкість генерації тестів у режимі реального часу [5]. У ході дослідження було проведено порівняльне тестування моделей Llama-3.1-8b-instant та Llama-3.3-70b-versatile. Встановлено, що модель 70b генерує складніші логічні запитання, проте модель 8b працює значно швидше.

Отриманий від нейромережі суцільний текст обробляється модулем парсингу на основі регулярних виразів (re), який автоматично розділяє його на окремі візуальні блоки. Кожне запитання супроводжується радіокнопками для вибору відповідей та системою самоперевірки. Візуалізацію згенерованого інтерактивного тесту подано на рис. 2.



Пройдіть тест:

Запитання №1 ⇄

Що таке Python?

Варіант 1:

- А) Мова програмування для веб-розробки
- Б) Мова програмування для операційних систем
- В) Інтерпретована, інтерактивна, об'єктно-орієнтована мова програмування
- Г) Мова програмування для баз даних

Рис. 2. Візуалізація згенерованого інтерактивного тесту

Особливу увагу в процесі тестування було приділено дослідженню обмежень інфраструктури при роботі з україномовним контентом. Виявлено, що через менш ефективну токенизацію кирилиці ліміт у 6000 токенів за хвилину вичерпується дуже швидко. Для забезпечення надійності та відмовостійкості в коді реалізовано алгоритм перехоплення виключень. У разі перевищення ліміту (Error 413: rate_limit_exceeded) система не зупиняє роботу аварійно, а за допомогою блоку try...except виводить користувачеві інформативне повідомлення з описом проблеми.

Висновки

Розроблене програмне забезпечення інформаційної технології «Інтелектуальний екзаменатор» успішно вирішує актуальне завдання автоматизації процесу створення перевірочних тестів. Використання хмарної платформи Groq та спеціалізованих LPU-процесорів дозволило забезпечити надвисоку швидкість генерації завдань на основі сучасних великих мовних моделей (LLaMa 8b та 70b), усунувши проблему тривалої затримки відповідей.

Програмна реалізація на базі фреймворку Streamlit підтвердила ефективність застосування динамічного парсингу за допомогою регулярних виразів для перетворення неструктурованого тексту на зручні інтерактивні візуальні блоки.

Проведене тестування довело надійність та відмовостійкість системи, зокрема завдяки успішному перехопленню виключень (Error 413) при вичерпанні лімітів API. Виявлена під час роботи специфіка неефективної токенизації кириличного контенту визначає напрямок подальших досліджень – впровадження алгоритмів автоматичного розбиття великих текстів (чанкінг) перед їх відправкою до мовної моделі, яку варто удосконалювати на основі ідей та формалізмів теорії асоціативного образного мислення людини [6, 7].

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Jurafsky D. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition / D. Jurafsky, J. H. Martin. – 3rd ed. draft. – Stanford University, 2023. – 600 p.
2. Touvron H. LLaMA: Open and Efficient Foundation Language Models / H. Touvron, T. Lavril, G. Izacard et al. // arXiv preprint arXiv:2302.13971. – 2023. – 14 p.
3. Streamlit Documentation — Official Website. URL: <https://docs.streamlit.io/>
4. Mitchell R. Web Scraping with Python: Collecting More Data from the Modern Web / R. Mitchell. – 2nd ed. – O'Reilly Media, 2018. – 308 p.
5. Groq Cloud API Documentation — Official Website. URL: <https://console.groq.com/docs/>
6. Дадиверін В. В., Бісікало О. В. Аналіз та імплементація ідеї навчання великих мовних моделей за аналогією з дитячим когнітивним розвитком. – Вчені записки Таврійського національного університету імені В.І. Вернадського, Серія: Технічні науки, Том 36 (75) № 4 2025, Частина 2. – СС.103-109. – ISSN 2663-5941.
7. Омельченко, В.; Бісікало, О. Удосконалення великих мовних моделей формальними засобами оцінки сенсу текстової інформації. ВІТКП ВНТУ. Факультет інтелектуальних інформаційних технологій та автоматизації, Ukraine, mar. 2026. Available at: <<https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2026/paper/view/27435/22812>>. Date accessed: 10 Mar. 2026.

Бісікало Олег Володимирович — д-р. техн. наук, професор, завідувач кафедри АІТ, Вінницький національний технічний університет, м. Вінниця, email: obisikalo@vntu.edu.ua

Апанасенко Каріна Євгенівна — студентка групи ІІСТ-23б, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця, e-mail: karinaapanasenko2006@gmail.com

Кириченко Катерина Дмитрівна — студентка групи ІІСТ-24б, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця, e-mail: kk9955458@gmail.com

Bisikalo Oleg V. — Dr. Sc. (Eng.), Professor, Head of the AIT Department, Vinnytsia National Technical University, Vinnytsia, email: obisikalo@vntu.edu.ua

Apanasenko Karina Y. — student of the IIST-23b group, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: karinaapanasenko2006@gmail.com

Kyrychenko Kateryna D. — student of the IIST-24b group, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: kk9955458@gmail.com