

**РОЗРОБКА QA-СИСТЕМИ
ПІДТРИМКИ ДИСТАНЦІЙНОГО НАВЧАННЯ
В УМОВАХ НЕСТАБІЛЬНОЇ ІНФРАСТРУКТУРИ**

Щербина П.О., Гриньова О.Є.

e-mail: polina.shcherbyna@nure.ua, olena.hrynova@nure.ua

Харківський національний університет радіоелектроніки, каф.ІІІ
м. Харків, Україна

The article develops a retrieval-based QA system to support distance learning amidst Ukraine's unstable infrastructure. The approach utilizes TF-IDF vectorization and cosine similarity to automate responses to student queries, ensuring reliability without generative hallucinations. A specialized Ukrainian-language corpus of 800 entries across six categories was created to handle linguistic variability. The system demonstrates high performance on low-end hardware, maintaining educational continuity during power outages. Testing via Accuracy and F1-score metrics confirms the model's effectiveness as a primary support line. This tool reduces teacher workload while preserving their central role in the pedagogical process.

Сьогодні дистанційне навчання в Україні часто переривається через проблеми зі світлом чи інтернетом. У такі моменти студентам важко оперативно отримати відповіді на прості питання: коли здавати роботу, як її оформити чи де знайти матеріали. Викладачі ж отримують сотні однакових повідомлень у месенджерах, що забирає багато часу. Запропоноване рішення – розробка QA-системи, яка автоматично відповідає на такі організаційні питання. Вона працює як «розумний» помічник і допомагає студентам миттєво, навіть якщо викладач у цей момент перебуває поза мережею.

Роль системи в навчанні. Важливо розуміти, що ця система не намагається замінити викладача. Навчання – це живий процес спілкування, і жоден алгоритм не замінить пояснення складної теми. Система потрібна лише для того, щоб взяти на себе рутину – відповіді на типові запитання про дедлайни та регламенти. Це звільняє час викладача для творчої роботи зі студентами, а самим студентам дає змогу швидко знайти потрібну інформацію 24/7.

Ціль роботи – розробка інтелектуальної системи підтримки студентів для написання стандартних запитів, коли викладач фізично не може вийти на зв'язок через відключення світла. Таким чином, час викладача для творчої педагогічної роботи та вирішення складних навчальних кейсів звільнюється. А система стає компонентом, що покращує взаємодію в системі «викладач-студент», а не заміщує її.

Методи та алгоритми. В основі розробленої системи покладено принцип retrieval-based QA-Model, що забезпечує високу достовірність

відповідей за рахунок виключення генеративних «галюцинацій», оскільки відповідь вилучається з верифікованої бази знань [1]. Процес інтелектуального аналізу запиту поділяється на три ключові етапи.

Етап 1. Багатоступенева попередня обробка тексту. Для забезпечення стійкості до варіативності природної мови в умовах нестабільної інфраструктури, кожен запит проходить через модуль препроцесингу:

- нормалізація та очищення. Приведення до нижнього регістру та видалення спеціальних символів, що не несуть смислового навантаження;
- токенизація та фільтрація. Розбиття тексту на окремі лексеми з подальшим видаленням стоп-слів (прийменників, часток), які мають низьку інформативність.

Етап 2. Векторизація на основі методу TF-IDF. Для перетворення текстової інформації у математичну форму використано статистичну міру TF-IDF (Term Frequency-Inverse Document Frequency). Це дозволяє системі оцінювати важливість слова в контексті конкретного запиту та всієї бази знань. Математично вага терміну t у документі d обчислюється як:

$$w_{t,d} = \text{tf}_{t,d} \cdot \log\left(\frac{N}{\text{df}_t}\right),$$

де $\text{tf}_{t,d}$ – частота терміна в документі, N – загальна кількість документів у корпусі, а df_t – кількість документів, що містять цей термін. Використання матричних обчислень у пакеті Scikit-learn забезпечує миттєву обробку навіть на застарілому апаратному забезпеченні.

Етап 3. Обчислення схожості. Коли студент вводить запит, система порівнює його «цифровий відбиток» із тими, що вже є в базі, за допомогою косинусної схожості. Вибирається та відповідь, де збіг найбільший. Якщо запит зовсім незрозумілий – система повідомить, що не знає відповіді, і не буде вводити в оману [2].

Структура бази знань та генерація даних. Якість роботи QA-системи безпосередньо залежить від репрезентативності та збалансованості вхідних даних. Для підвищення точності було сформовано спеціалізований корпус запитів українською мовою, що охоплює 200 базових питань та 600 синтетичних перефразувань. Синтетичне розширення дозволило охопити варіативність лексичних форм та синтаксичних конструкцій, що притаманні природній мові студентів. Корпус охоплює шість тематичних категорій: організаційні питання, оцінювання, оформлення робіт, терміни подачі, доступ до матеріалів та нормативні положення кафедри:

- запити загального характеру щодо розкладу консультацій, форматів проведення занять та контактних даних викладачів;
- питання щодо критеріїв нарахування балів, ваги окремих видів робіт у підсумковій оцінці та умов отримання автоматичних заліків;
- методичні запити стосовно стандартів ДСТУ для списків літератури, вимог до структури пояснювальних записок та оформлення титульних аркушів;

– найбільш критична категорія в умовах нестабільної інфраструктури, що стосується термінів виконання лабораторних та практичних робіт, проходження тестів [3] та deadlines ліквідації академічної заборгованості;

– запити щодо посилань на навчальні курси в системі Moodle, паролів до архівів з лекціями та доступу до віртуальних лабораторних стендів;

– роз'яснення щодо регламентів перескладань, правил вибору дисциплін та інших внутрішніх положень кафедри ІІІ.

Для кожного питання прописано мінімум три варіанти того, як студент може його сформулювати. Наприклад, на питання про дедлайн додано варіанти: «Доки треба здати?», «Коли дедлайн по лабі?», «Чи можна ще завантажити звіт сьогодні?». Такий підхід робить систему надійною.

Використання саме retrieval-based моделі дозволяє підтримувати ці дані у актуальному стані без необхідності перенавчання нейронних мереж, що є критично важливим для оперативного оновлення інформації викладачем.

Програмний прототип побудовано за модульним принципом (Python, Scikit-learn). Завдяки матричним обчисленням система демонструє високу швидкодію на малопотужному обладнанні. Тестування за метриками Accuracy та F1-score підтвердило ефективність розпізнавання варіативних запитів.

Впровадження системи відповідає викликам цифровізації освіти в умовах форс-мажорних обставин. Технологічна автономність та швидкість доступу до інформації є критичними для збереження якості навчання під час війни. Запропонована QA-система оптимізує рутинну роботу кафедри та створює фундамент стійкої комунікації в умовах нестабільної інфраструктури.

Список використаних джерел:

1. Jurafsky D., Martin J.H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 3rd Edition. – Pearson, 2021. URL: https://web.stanford.edu/~jurafsky/slp3/ed3book_jan26.pdf (дата звернення: 23.02.2026).

2. Vaswani A. et al. Attention is All You Need. Advances in Neural Information Processing Systems. 2017. Vol. 30. P. 5998–6008. URL: <https://arxiv.org/abs/1706.03762> (дата звернення: 23.02.2026).

3. Hurzhy T., Hrynova O. Leveraging LLMs for automated test and glossary generation in education. Proceedings of the International Workshop “Chatbot, Uzhhorod, 2025. 28–32p. URL: <https://dspace.uzhnu.edu.ua/handle/lib/74262> (дата звернення: 23.02.2026).