

ПРОЄКТУВАННЯ ЛОКАЛЬНОГО ЗАСТОСУНКУ ТРАСКРИБУВАННЯ ТА СУМАРИЗАЦІЇ АУДІО КОНТЕНТУ

Бізюк А.В., професор, кафедра МСТ, ХНУРЕ

Лозін Д.Д., студент, кафедра МСТ, ХНУРЕ

Анотація. У роботі досліджено актуальні тенденції розробки застосунків для транскрибування та сумаризації аудіо- та відеоматеріалів з акцентом на мінімалістичному дизайні та високому рівні безпеки даних. Обґрунтовано переваги локальних моделей, та запропоновано підходи до оптимізації користувацького інтерфейсу. Розглянуто перспективні функції та інтеграцію локальних мовних моделей для створення стислих підсумків і відповідей на запитання.

Ключові слова: МУЛЬТИМЕДІА, UX/UI ДИЗАЙН, КОНФІДЕНЦІЙНІСТЬ, ІНФОРМАЦІЙНА БЕЗПЕКА, КРОСПЛАТФОРМНІСТЬ.

Попит на інструменти транскрипції та сумаризації аудіо та відео зростає через дві ключові причини. По-перше, щороку збільшується обсяг аудіо інформації, що потребує швидкого й ефективного оброблення: від освітніх лекцій до бізнес-зустрічей [1]. По-друге, користувачі дедалі більше дбають про безпеку та конфіденційність своїх даних. Це стимулює розробників шукати офлайн-рішення, здатні гарантувати, що записи не залишають пристрою й не потрапляють у хмарні сервіси. Саме в цій площині актуальність набуває поєднання мінімалістичного дизайну, який оптимізує увагу користувача, із сучасними технологіями безпеки, що забезпечують локальну обробку та надійне шифрування.

Мінімалістичність інтерфейсу дозволяє спростити взаємодію з програмою. Коли користувач відкриває додаток для транскрибування й бачить лише одну кнопку «Почати запис» або зону «Перетягнути файл», це значно зменшує когнітивне навантаження. Усі додаткові інструменти, як-от налаштування чутливості, вибір розміру моделі чи параметри експорту, можна сховати в окремий розділ «Опції» чи «Налаштування». Також бажано надати можливість вибору розміру чи точності мовної моделі, щоб задовольнити потреби як користувачів із потужними комп'ютерами, так і тих, хто користується базовими ноутбуками. Дослідження Nielsen Norman Group [2] свідчать, що мінімалістичні рішення покращують загальний користувацький досвід, оскільки люди швидше концентруються на суті завдання.

Ще один важливий аспект – це безпека даних. Використання локальних моделей на основі проєктів типу OpenAI Whisper чи GPT4All уможлиблює цілковиту автономність застосунку: аудіо- та відеоматеріали обробляються без під'єднання до зовнішніх серверів. З погляду конфіденційності це майже ідеальне рішення, оскільки будь-який ризик витоку інформації мінімізується. Однак, аби справді перевершити чинні офлайн-пропозиції на ринку, варто додати можливість генерувати короткі підсумки або Q&A без виходу з програми, а також запропонувати вбудоване шифрування результатів.

Крім того, перспективним кроком є реалізація автоматичного визначення спікерів, що дасть змогу розрізняти голоси різних співрозмовників. Це особливо затребувано у сфері корпоративних онлайн-зустрічей, подкастів та інтерв'ю, де важливо чітко

розуміти, хто і коли говорить. Наступним етапом розвитку може стати інтеграція локальних мовних моделей: від покоління тез і ключових ідей до можливості швидких запитань «Що сказала доповідачка про дедлайни?» з миттєвою відповіддю в додатку. Такі можливості дають змогу повноцінно реалізувати концепцію «Все в одному», коли користувач отримує повний цикл оброблення аудіо — від транскрипції до підсумку або навіть стислого FAQ.

Важливо зазначити, що не можна нехтувати питанням адаптації під різні платформи. Обравши підхід до інтерфейсу в якому буде однакова структура меню й розкладка елементів, навіть якщо в основу будуть взяті Apple Human Guidelines [3] то ці концепції точно не встануть на заваді версії застосунка для Windows чи навіть Linux.

Слід також зазначити, що студентами ХНУРЕ також проводяться дослідження в цьому напрямку. Так, зокрема, в кваліфікаційній роботі Морковського К.Д. [4] проаналізовано існуючі методи та підходи до обробки природної мови, за допомогою яких може бути здійснена сумаризація наукового тексту (публікації, статті, тощо) досить великого розміру. Проведено їх порівняння в контексті теперішнього часу.

Отже, якщо стоїть на меті створити офлайн-застосунок для транскрипції й сумаризації, який перевершить існуючі аналоги, потрібно зосередитися на таких головних напрямках як мінімалістичний дизайн, який виносить на перший план лише основні функції. І також реалізувати повноцінну локальну обробку даних, підкріплену шифруванням та інтегрованими інструментами безпеки операційної системи з можливістю вибору мовної моделі для аналізу даних. Таке поєднання допоможе не тільки задовольнити запит на конфіденційність та простоту, а й створить можливість для розвитку цієї галузі у майбутньому.

Література.

1. Murel, J. What is text summarization? <https://www.ibm.com/think/topics/text-summarization>.
2. Fessenden, T. Aesthetic and Minimalist Design (Usability Heuristic #8). <https://www.nngroup.com/articles/aesthetic-minimalist-design>.
3. Human Interface Guidelines. <https://developer.apple.com/design/human-interface-guidelines>.
4. Морковський, К.Д. (2023). Дослідження методів обробки природної мови для сумаризації тексту наукових публікацій : пояснювальна записка до кваліфікаційної роботи здобувача вищої освіти на другому (магістерському) рівні, спеціальність 122 Комп'ютерні науки / М-во освіти і науки України, Харків. нац. ун-т радіоелектроніки. Харків.