

УДК 004.032.26:37.09

## **РОЗГЛЯД ПИТАННЯ СТВОРЕННЯ КОНСПЕКТУ ЛЕКЦІЙ НА ОСНОВІ ВІДЕО ТА ПРЕЗЕНТАЦІЇ**

Максімов Г.Р.

email: hlib.maksimov@nure.ua

Науковий керівник – к.т.н., доц. Яковлева О.В.

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

Video content is a major source of information, yet time constraints often prevent users from fully engaging with it. This work presents an innovative application that automatically generates comprehensive lecture notes by analyzing video, audio, and presentation materials. The system utilizes advanced techniques such as OpenAI Whisper for transcription, Silero VAD for silence detection, FFmpeg for key frame extraction, and modern large language models for summarization. Unlike existing solutions, our approach effectively integrates visual cues from key frames to produce contextually rich summaries.

У сучасному інформаційному середовищі відеоконтент набуває все більшої популярності завдяки синтезу зорової та слухової інформації, що сприяє ефективнішому засвоєнню матеріалу. Проте, через обмеження часу користувачам часто буває складно переглянути повний відеоматеріал для отримання лише основних тез. Ця проблема особливо актуальна для студентів, які потребують швидкого ознайомлення з лекціями, та для нових співробітників, яким необхідно оперативно отримувати ключову інформацію зі зустрічей.

Ці випадки чітко висвітлюють задачу, що поставлена в рамках даної наукової роботи – створити можливість підсумувати відеоконтент в текстовому форматі, виділивши основні тези, розбивши на слайди або секції по декілька хвилин кожна. При цьому, бажано послатися на ключові зображення, слайди тощо. Це рішення може бути актуальним для прискорення процесу навчання та попереднього аналізу відеоконтенту.

На ринку вже існує ряд сервісів [1, 2], що надають подібний функціонал, проте кожен з них має свої суттєві недоліки. Такі застосунки, як Glasp, YouLearn, NoteGPT, kagi, Slider, мумар, Otio та getrecall, не забезпечують прив'язку до ключових кадрів у відео, що обмежує точність конспектування. Крім того, багато з них страждають від проблем якості, завищеної вартості або обмежень на тривалість відео. Часто користувач не може завантажити супровідну презентацію або попередньо визначити структуру майбутнього конспекту.

Всі ці фактори свідчать про те, що існуючі рішення залишаються недосконалими, а тому виникає потреба у створенні більш інтегрованого та точного інструменту.

Сучасні досягнення в галузях комп'ютерного зору, обробки аудіо даних і тексту значною мірою зумовлені розвитком нейромережових методів, які демонструють видатні результати в задачах класифікації, генерації та розуміння інформації. Проте на практиці широко застосовується комбінований підхід, при якому сучасні нейромережові моделі інтегруються з класичними методами комп'ютерного зору [3, 4] і обробки природної мови.

Для вирішення задачі створення конспектів лекцій на основі аналізу відео, аудіо контенту та презентації в роботі пропонується комплексний підхід на базі інтеграції сучасних інструментів обробки даних (табл. 1).

Таблиця 1 – Використані моделі ШІ та інструменти

Назва інструменту	Призначення
OpenAI Whisper [5]	транскрибування аудіо
Silero VAD	виявлення беззвучних фрагментів
OpenAI Whisper, pyannote	діаризація (дозволяє розподілити текст за окремими спікерами)
FFmpeg, scikit-learn	витяг ключових кадрів
PaddleOCR	розпізнавання тексту з візуальних елементів
GPT-4, Google Gemini	генерація стислого, сумаризованого тексту
Flask	розробка бекенду
HTML, CSS, JavaScript	побудова фронтенду
pydantic	структуризація отриманих даних

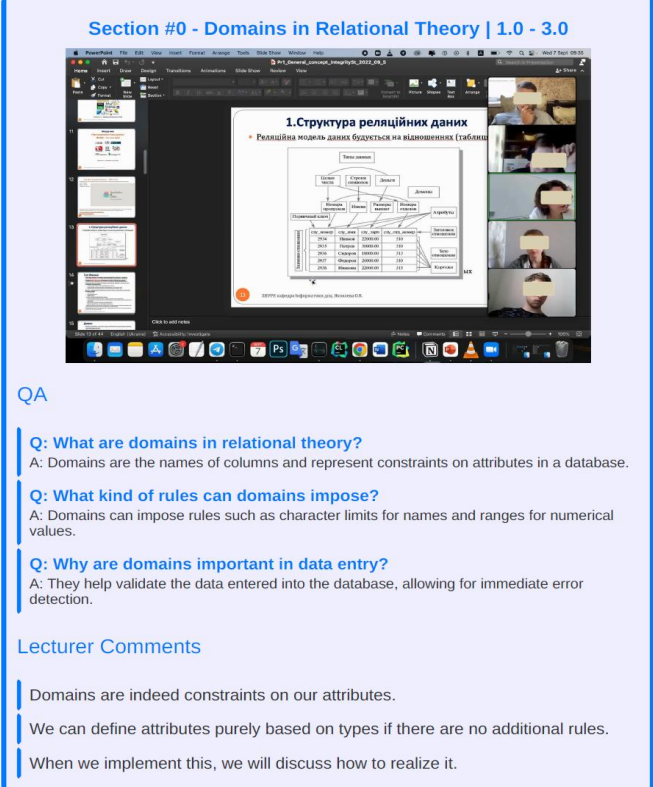
Процес роботи системи складається із таких етапів:

- завантаження користувачем відеоматеріалу та супровідних презентацій на хмарні сховища, наприклад, Google Drive;
- після цього контент автоматично завантажується сервером, де проводиться його попередня обробка: відео очищується від беззвучних фрагментів, а презентація розбивається на окремі сторінки у форматі PDF;
- відео ділиться на фрагменти тривалістю близько 30 секунд, для кожного з яких визначаються ключові параметри: витягнутий текст, часові мітки та ключові кадри;
- отримані дані групуються відповідно до тематичних блоків, що дозволяє не тільки розподілити матеріал за слайдами, але й виділити «відповіді на запитання» та «коментарі лектора»;
- нарешті, всі результати інтегруються у фінальний конспект у форматі PDF, який поєднує текстову та візуальну інформацію.

Приклади результату роботи застосунку наведено на рисунку 1.

Розробка даного застосунку є кроком уперед у напрямку оптимізації роботи з відеоматеріалами. Він особливо актуальний для студентів заочної форми навчання та співробітників, яким необхідно оперативно аналізувати великі обсяги інформації. Комплексний підхід, що включає аналіз як аудіо-та відеоконтенту, так і супровідних презентацій, дозволяє значно прискорити процес вилучення основних тез, забезпечуючи користувача не лише

стислим викладом, а й візуальними елементами, що допомагають краще зрозуміти матеріал.



The screenshot shows a Zoom meeting interface. The main window displays a presentation slide with the title "1. Структура реляційних даних" (Structure of Relational Data) and the subtitle "Реляційна модель даних базується на відносних таблицях" (Relational data model is based on relative tables). The slide contains a diagram of a relational database structure and a table with columns for "Підприємство" (Company), "Відділ" (Department), "Ім'я" (Name), "Вік" (Age), "Стать" (Gender), "Посада" (Position), and "Зарплата" (Salary). Below the slide, there is a Q&A section with three questions and answers, and a "Lecturer Comments" section with three comments.

**Section #0 - Domains in Relational Theory | 1.0 - 3.0**

**1. Структура реляційних даних**  
Реляційна модель даних базується на відносних таблицях

**QA**

**Q: What are domains in relational theory?**  
A: Domains are the names of columns and represent constraints on attributes in a database.

**Q: What kind of rules can domains impose?**  
A: Domains can impose rules such as character limits for names and ranges for numerical values.

**Q: Why are domains important in data entry?**  
A: They help validate the data entered into the database, allowing for immediate error detection.

**Lecturer Comments**

- Domains are indeed constraints on our attributes.
- We can define attributes purely based on types if there are no additional rules.
- When we implement this, we will discuss how to realize it.

Рисунок 1 – Приклад роботи застосунку

Таким чином, запропонований інструмент має всі передумови для того, щоб стати унікальним рішенням у сфері мультимодального конспектування, сприяючи підвищенню якості освіти та ефективності робочих процесів.

Список використаних джерел:

1. YouTube Summary with ChatGPT. URL: <https://glasp.co/youtube-summary> (дата звернення: 04.03.2025).
2. NoteGPT – AI Summarizer & Generator for Enhanced Learning. URL: <https://glasp.co/youtube-summary> (дата звернення: 04.03.2025) (дата звернення: 04.03.2025).
3. Application a Committee of Kohonen Neural Networks to Training of Image Classifier Based on Description of Descriptors Set / V. Gorokhovatskyi et al. *IEEE Access*. 2024. P. 1. URL: <https://doi.org/10.1109/access.2024.3404371>.
4. Gorokhovatskyi O., Yakovleva O. Medoids as a packing of ORB image descriptors. *Advanced Information Systems*. 2024. Vol. 8, no. 2. P. 5–11. URL: <https://doi.org/10.20998/2522-9052.2024.2.01>.
5. Robust Speech Recognition via Large-Scale Weak Supervision. *arXiv.org*. URL: <https://doi.org/10.48550/arXiv.2212.04356> (дата звернення: 04.03.2025).