УДК 004.89:004.93

ВИКОРИСТАННЯ ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ СТВОРЕННЯ ВІДЕО НА ОСНОВІ ТЕКСТУ

Шатило І.Ю., Чала Л.Е.
e-mail: ihor.shatylo@nure.ua, larysa.chala@nure.ua
Харківський національний університет радіоелектроніки, каф. ШІ
м. Харків, Україна

This paper introduces a technology designed for automated video generation from web-based articles. The system follows a structured process that includes content parsing, translating it, summarising key points, extracting keywords, and analysing sentiment. It selects relevant video clips and background music, synthesises voiceovers using text-to-speech technology, and assembles the final video. The result is a fully automated pipeline that converts textual content into dynamic video presentations, with applications in news automation, educational media, and digital content production.

Зі зростанням обсягів цифрової інформації особливо актуальним є питання швидкого й ефективного сприйняття текстових матеріалів, які часто потребують значних часових затрат. Автоматизація їхнього перетворення у відеоформат може суттєво підвищити доступність контенту в журналістиці, освіті та маркетингу. Основна ідея полягає в тому, щоб автоматично отримувати текстову інформацію з різних джерел, виділяти з неї ключові моменти, ілюструвати її відповідним візуальним та аудіоконтентом і створювати відео без участі людини. Створення відео для текстового контенту веб-сайтів можна розділити на 4 важливих етапи: парсинг контенту, аналіз контенту, пошук або відео- та аудіоконтенту для кожної сцени відео, і етап створення фінального відео.

На рисунку 1 наведено узагальнену схему запропонованої технології створення відео за текстовим контентом.

На першому етапі відбувається отримання тексту з веб-сторінок, що може ускладнюватись їхньою динамічною структурою та великою кількістю допоміжних елементів. Для розв’язання цієї проблеми застосовується метод динамічного рендерингу сторінок, коли контент завантажується у браузероподібному середовищі, що дає змогу обробляти веб-сайти з активним JavaScript. Реалізація цього підходу здійснюється за допомогою бібліотеки Playwright. Для виокремлення основного текстового блоку використовується бібліотека newspaper4k для Python, яка застосовує евристичні методи та статистичний аналіз структури HTML-документів. В основі її роботи лежать алгоритми визначення найбільш інформаційно насичених блоків на основі кількості текстового вмісту, співвідношення тексту до коду та аналізу DOM-структури. Також застосовується механізм фільтрації нерелевантного контенту на основі регулярних виразів та

попередньо визначених шаблонів, що дозволяє ефективно відсікати рекламу, коментарі та навігаційні панелі [1].

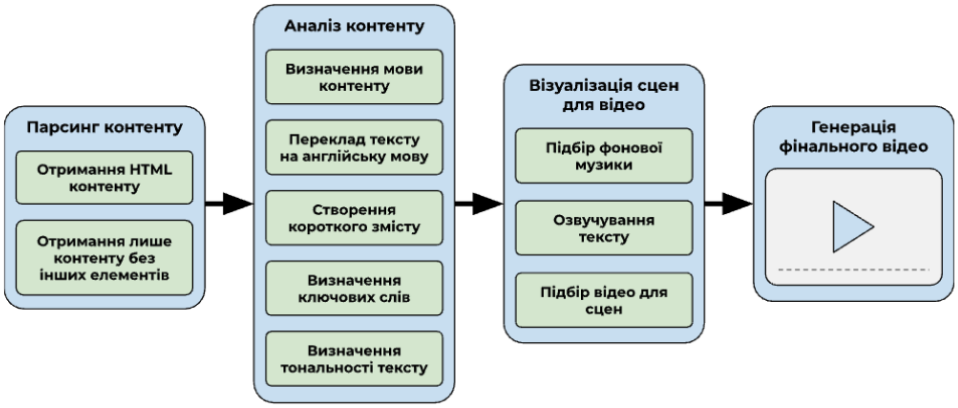

Рисунок 1 – Узагальнена схема технології створення відео з текстового контенту

Після отримання контенту відбувається його обробка за допомогою методів NLP. Визначення мови тексту є важливим кроком, оскільки більшість мовних моделей працюють переважно з англомовними даними. Для цього застосовуються статистичні методи, реалізовані у бібліотеках langdetect та textblob. Переклад здійснюється за допомогою sequence-to-sequence моделей, що використовують механізм уваги для коректного збереження контексту під час трансформації тексту. Для цього у даній роботі використовуються моделі Helsinki NLP (MarianMT), які демонструють високу ефективність завдяки попередньому тренуванню на багатомовних корпусах. Для узагальнення змісту тексту застосовуються методи реферування, які поділяються на екстрактивні (виділення ключових речень) та абстрактивні (генерація нового тексту зі збереженням змісту). Реалізація цих методів можлива на основі трансформерних моделей BART, T5 та GPT [2, 3]. Вони використовують глибоке контекстуальне представлення тексту, що дозволяє будувати якісні скорочені виклади. У даній роботі використовується модель BART для генерації короткого змісту тексту.

Важливою задачею є також аналіз тональності тексту, який здійснюється за допомогою нейромережевих моделей. Для цього використовується модель DistilBERT, яка є адаптованим варіантом BERT [4], та яка навчається на розмічених корпусах для розпізнавання позитивних, негативних чи нейтральних висловлювань. Визначення ключових слів, необхідних для подальшого підбору аудіо- та відеоконтенту, реалізується за допомогою використання моделі KeyBERT, яка враховують контекстне значення слів.

Після аналізу тексту відбувається пошук або генерація відповідного мультимедійного контенту. Для підбору зображень і відео використовується метод пошуку за ключовими словами, реалізований у

відкритому API від Pexels. Фоновий музичний супровід отримується шляхом генерації музики на основі глибоких нейромережевих підходів. Для цього у даній роботі використовується модель Facebook MusicGen, яка використовує рекурентні й трансформерні архітектури для автоматичного створення мелодійних послідовностей, враховуючи заданий стиль і темп.

Додатково для озвучення тексту застосовується метод синтезу мовлення (text-to-speech), що передбачає послідовне перетворення текстових даних у фонеми, а потім у звукові хвилі. Для цього у даній роботі використовуються спеціалізовані моделі StyleTTS2, які дозволяють досягти більш природного звучання голосу завдяки механізму гнучкого контролю інтонації та стилю мовлення. На відміну від традиційних TTS-систем, що використовують попередньо визначені мовні шаблони, StyleTTS2 базується на генеративних підходах, що забезпечують високу варіативність та адаптивність синтезованого мовлення.

На завершальному етапі відбувається об'єднання всіх елементів у єдиний відеофайл. Для цього використовується метод послідовного монтажу, що дозволяє програмно керувати накладанням відеофрагментів, аудіодоріжок і текстових субтитрів. У цій роботі це реалізується за допомогою бібліотеки MoviePy для Python, яка забезпечує широкий набір інструментів для редагування відео.

Автоматизація створення відео з тексту спрощує підготовку контенту, робить його більш доступним і усуває потребу в професійних відеоредакторах. Сучасні моделі NLP забезпечують високу якість, але залишаються певні проблеми, зокрема візуальна узгодженість, точність згенерованого тексту та природнє озвучення. Подальший розвиток штучного інтелекту та мультимедійних технологій допоможе вдосконалити цей підхід і розширити його застосування.

Список використаних джерел:

1. Rajat Thakur. What Are The Different Types Of Web Scraping Approaches? DEV Community. URL: https://dev.to/digitallyrajat/what-are-the-different-types-of-web-scraping-approaches-1e4g (дата звернення: 08.02.2025).

2. Abstractive Text Summarisation Using Keywords with Transformers Model / P. Shanmugam S. та ін. 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), м. Ченнай, 6-7 квіт. 2023 р. 2023. С. 1-9. URL: https://ieeexplore.ieee.org/document/10142867 (дата звернення: 09.02.2025).

3. Дудник М.П., Удовенко С.Г., Чала Л.Е., Соколовська М.М. Нейромережева технологія багатомовної класифікації електронних текстів // Біоніка інтелекту. – 2021. – Вип. 2 (97). – С. 3-12.

4. Bharath K. BERT Transformers for Natural Language Processing. Paperspace by DigitalOcean Blog. URL: https://blog.paperspace.com/bert-natural-language-processing/ (дата звернення: 11.02.2025).