

МЕТОДИ АВТОМАТИЗОВАНОЇ КОНВЕРСІЇ ВІДЕОКОНТЕНТУ В СТРУКТУРОВАНІЙ МЕДІАМАТЕРІАЛ

Левикін І.В.

д.т.н., доцент, кафедра «Медіасистеми та технології»,
Харківський національний університет радіоелектроніки

Шимко Д.І.

аспірант, кафедра «Медіасистеми та технології»,
Харківський національний університет радіоелектроніки

***Анотація.** У роботі досліджено процес автоматизованої конверсії відео в структурований текстовий контент із контекстно синхронізованими зображеннями. Розглянуто сучасні інструменти на базі штучного інтелекту для транскрипції мовлення, семантичної обробки тексту та генерації візуального супроводу. Проведено експериментальне тестування на комплексному мультимедійному матеріалі.*

***Ключові слова:** АВТОМАТИЗАЦІЯ, МЕДІАКОНВЕРСІЯ, ТРАНСКРИПЦІЯ, РОЗПІЗНАВАННЯ МОВЛЕННЯ, СТРУКТУРОВАНІЙ ТЕКСТ, ШТУЧНИЙ ІНТЕЛЕКТ.*

Вступ

Відео є одним із найбільш комплексних і насичених форматів передачі інформації. Його унікальність полягає в одночасному залученні кількох каналів сприйняття, що дозволяє глибше зануритися в контент і краще запам'ятовувати подану інформацію. Така багаторівнева взаємодія з користувачем значно підвищує ефективність навчального чи інформаційного матеріалу. Відео забезпечує більш інтегрований підхід до подання змісту, дозволяючи краще візуалізувати об'єкти, процеси чи події, демонструвати приклади, експерименти чи інтерв'ю в динаміці.

Незважаючи на переваги використання відеоформату, він також має низку важливих недоліків. Зокрема, зберігання відеоматеріалів потребує значного обсягу пам'яті через високу вагу файлів. Часто якість відео чи звуку може бути незадовільною через технічні обмеження запису або обробки, що ускладнює сприйняття. Крім того, відео може містити багато несуттєвих елементів, таких як заставки, переходи або рекламні вставки, які не несуть змістового навантаження і лише відволікають глядача. Також відео практично позбавлене можливості швидкої навігації по змісту, оскільки пошук або перехід по навігації у відео неможливо реалізувати з такою ж швидкістю та зручністю, як це зроблено в текстових матеріалах. До того ж ефективність відтворення відеофайлів часто залежить від технічних характеристик пристроїв, на яких вони переглядаються.

На цьому тлі текстові формати демонструють цілий ряд переваг, що робить їх незамінними в багатьох сферах. Текст дозволяє швидко орієнтуватися в

матеріалі завдяки чіткій структурі, змісту, заголовкам і підзаголовкам. За допомогою функцій пошуку можна миттєво знаходити потрібні фрагменти, що є особливо корисним під час аналізу великих обсягів даних або навчального контенту. Також важливо, що текстові файли займають мінімум місця, легко редагуються, копіюються, зберігаються й відтворюються практично на будь-якому пристрої, що значно підвищує доступність контенту для широкої аудиторії.

Сучасні інструменти з використанням моделей штучного інтелекту дозволяють автоматизувати створення ключових деталей та змісту обговорення, що значно спрощує підсумовування офісних зустрічей та зідзвонів через інтернет [1]. У деяких випадках відеоформат матеріалу обирається не через огляд та вибір між перевагами та недоліками цього формату, а через неможливість збереження оригінального матеріалу у інших форматах.

Наприклад, запис онлайн-зустрічей, які відбуваються у багатьох компаніях у таких месенджерах, як Zoom, Slack, Google Meet та інших, що дає можливість зберегти увесь обговорюваний матеріал. Подальше переглядання цього матеріалу, який може складатися з годин відеозапису, у намаганнях виписати важливі ключові деталі та зміст усіх обговорень є складним та нераціональним методом обробки.

У сучасному світі, особливо з урахуванням реалій життя у пост-пандемічний час та через вплив інших факторів нестабільності, навчання у 71.6% освітніх закладах України відбувається дистанційно [2]. Можливість збереження та подальшої обробки лекційних матеріалів може допомогти викладачам створювати більш детальні та ефективні навчальні матеріали. Це особливо буде корисним для тих студентів, які не могли бути присутніми на лекціях, або віддають перевагу використанню текстових та графічних матеріалів, а не лекцій у живому форматі.

Незважаючи на популярність текстових форматів, велика частина сучасного мультимедійного контенту, зокрема ЗМІ, освітні платформи, подкасти та відеоблоги, створюються насамперед у форматі відео. Хоча ці матеріали можуть бути надзвичайно цінними, їх подальше використання у текстовому вигляді потребує мануальної роботи. Перегляд відео, запис тексту, підбір зображень, що відповідають змісту, та додаткової структуризації. Такий підхід потребує багато часу та зусиль, що обмежує масштабування контенту в різних форматах.

Враховуючи недоліки відеоматеріалів та складнощі їх перетворення у текстовий формат, виникає потреба в автоматизованих рішеннях, які здатні трансформувати відео у текстовий формат із збереженням якостей важливих візуальних і аудіоелементів. Це повинно забезпечити точність та ефективність такого перетворення.

Замість того щоб ставити текст і відео в опозицію одне до одного, сучасні підходи публікації та просування медіаматеріалів пропонують інтеграцію цих форматів у рамках єдиного мультимедійного підходу. Створення

кросплатформеного контенту, адаптованого до кількох форматів, як відео, тексту та зображень, дозволяє значно розширити охоплення аудиторії. Завдяки публікації одного й того ж матеріалу на різних платформах у відповідних форматах досягається більша універсальність. Такий кросплатформенний підхід стає все більш актуальним у сучасному інформаційному середовищі.

Через особливості та обмеження кожної популярної медіаплатформи виникають складнощі з адаптацією матеріалів або зовсім неможливістю їх публікації. Наприклад, відеохостинг Youtube призначений для публікації змістовних відео, як у інших популярних соцмережах, як Twitter або Facebook, переважає текстова інформація, яка доповнюється зображеннями.

Разом із тим, під час простої транскрипції відео у текст, як правило, втрачається велика частина його смислового наповнення. Якщо відео значною мірою спирається на візуальні елементи, простий текст не передає цього змісту повноцінно. У випадку інтерв'ю або подкастів ситуація може бути простішою, однак виникають складнощі з визначенням розповідачів та їх голосів. Доповнення графічних матеріалів можуть мати суттєве значення для вивчення та засвоєння матеріалу, а їх втрата може знизити інформативність транскрипції.

Тому ключовим викликом стає збереження візуального й контекстного наповнення відео під час його трансформації у текстову форму. Для цього необхідно не лише точно відображати мовленнєвий зміст, а й доповнювати його зображеннями, які відповідають контексту кожної частини. Важливо, щоб ці зображення були точно синхронізовані з текстом, оскільки навіть кілька секунд затримка чи неправильна асоціація можуть втратити доречність, що негативно впливає на сприйняття матеріалу.

Такий підхід дозволяє отримати гібридну форму представлення матеріалу, яка поєднує переваги тексту, візуалізації й аудіо. Одним із найперспективніших напрямів у цій галузі є застосування штучного інтелекту. Наприклад, модель "Whisper" від OpenAI може автоматично транскрибувати аудіо з відео, визначаючи часові мітки для кожного фрагмента. Це дає змогу прив'язати зображення або скріншоти до відповідного тексту, значно покращуючи сприйняття матеріалу [3].

Інші сервіси, як AssemblyAI, забезпечують розпізнавання окремих голосів у діалогах, що дозволяє автоматично розділити висловлювання різних розповідачів і створити покращену структуру тексту. Це особливо корисно в подкастах чи інтерв'ю, де важливо розуміти, хто саме говорить у конкретний момент часу. Відсутність відокремлення різних розповідачів може створити матеріал повністю спотвореним та неприйнятним, коли репліки декількох людей будуть йти єдиним текстом без жодного відокремлення. Завдяки таким технологіям автоматизована обробка відео може досягти нових рівнів точності та зручності.

Втім, використання таких інструментів пов'язане з низкою етичних викликів. Наразі відсутні чітко визначені стандарти для автоматичної обробки медіа з використанням ШІ, а також не завжди можна гарантувати достовірність

створених текстів чи зображень. Особливо гострою є проблема використання генеративного ШІ для створення фейкових матеріалів, що загрожує поширенням дезінформації.

Однак у випадку перетворення відео в текст із використанням початкового оригінального контенту, а не генерації нових матеріалів, ризики спотворення матеріалів значно менші. Такий підхід дозволяє ефективно використовувати ШІ без притаманних йому загроз, водночас розширюючи можливості для журналістики, освіти та інших сфер.

Проведення системного аналізу існуючих інструментів та їх тестування на прикладах реальних відеоматеріалів дозволить виявити переваги та обмеження кожного з підходів. На основі цього можна розробити універсальні методи для створення автоматизованих систем трансформації медіаконтенту з дотриманням балансу між точністю, доступністю та етичністю.

Мета та задачі дослідження

Метою дослідження є створення, апробація та оцінювання ефективності автоматизованої інформаційної технології, яка забезпечує конверсію відеоматеріалів у текстові документи зі збереженням контекстуальних особливостей і візуального супроводу.

Дослідження зосереджено на вивченні можливостей сучасних інструментів автоматичної транскрипції, мовного аналізу та генерації тексту для побудови цілісного алгоритму. Ці інструменти дозволяють здійснювати трансформацію відео у структурований медіаматеріал, придатний для подальшої публікації або інформаційної адаптації.

Задачі дослідження включають наступні.

1. Розробка методики конверсії відео у структурований текстовий матеріал із релевантними зображеннями, яка буде дозволяти повністю автоматизувати даний процес у подальшому. Важливим аспектом є вибір таких інструментів, які можуть виконуватися програмно та бути універсальними при обробці матеріалів різної тематики та складової.

2. Вибір інструментів для отримання та обробки відео, які також будуть дозволяти подальшу автоматизацію процесів.

3. Аналіз можливостей сучасних популярних інструментів створення текстової транскрипції з аудіозапису.

4. Визначення кількості помилок та обмежень кожного з розглядаємих інструментів на конкретному прикладі відеоматеріалу. Відео повинно мати комплексний семантичний матеріал, який повинен ускладнити транскрипцію для кращого аналізу успішності кожного з оглядаємих інструментів. Важливо, щоб обраний матеріал мав аудіозапис зі словами, які включають визначення, імена та окремі назви, що повинно ускладнити правильне визначення тексту з аудіо.

5. Форматування визначеного тексту зі збереженням часових відміток, що дозволить у подальшому додавати зображення, які будуть синхронізовані та точно відповідати тексту, поруч з яким вони будуть розміщені.

6. Виконання додаткових операцій над текстом, як переклад та перевірка на помилки, які могли бути допущені під час виконання операцій над текстом.

7. Визначення універсального запиту до моделей ШІ, який дозволить структурувати текст шляхом визначення ключових тем та відокремлення частин тексту одна від одної.

8. Перетворення структурованого тексту у формат HTML, що дозволить одразу розмістити його на веб-сайті без необхідності додаткової обробки та редагування.

9. Формування кінцевого результату з зображеннями, які відповідають визначеним часовим відміткам у попередньо отриманому структурованому тексті, який часові мітки на зображення.

Комплексне завдання дослідження полягає не лише в механічному перетворенні аудіовізуального контенту на текстовий, а й у логічному структуруванні отриманого результату відповідно до норм і вимог сучасної подачі медіаматеріалів.

Це включає обробку мовлення, поділ на смислові блоки, створення абзаців і заголовків, ідентифікацію мовців, вставлення релевантних зображень у відповідні місця, а також оформлення матеріалу у форматі, придатному для цифрового поширення.

У межах дослідження об'єктом розгляду є процеси автоматизації конверсії медіаресурсів з одного виду контенту в інший. Зокрема, розглядається перетворення відео у текст із подальшим формуванням структурованого матеріалу з графічним супроводом.

Предметом дослідження є методи, алгоритми, моделі та інформаційні технології, що забезпечують реалізацію такого перетворення.

Особливу увагу приділено інструментам автоматичного розпізнавання мовлення та сучасним моделям генеративного штучного інтелекту, які забезпечують обробку тексту, підвищують якість конверсії та сприяють його структуризації для кращого сприйняття. Додатково розглянуто методи інтеграції зображень у фінальний текст, що відповідають контексту висловлювання в конкретні моменти часу та дозволяють зберегти сутність оригінального відеоматеріалу, компенсуючи ті аспекти, які неможливо повноцінно передати лише текстовими засобами.

Важливим завданням цього дослідження є не лише реалізація окремих етапів і процедур, а й вибір таких інструментів, які забезпечують можливість їх подальшої автоматизації. Проведено практичне тестування окремих компонентів системи, що у перспективі можуть бути повністю автоматизовані, мінімізуючи потребу у втручанні людини в процес перетворення відео в структурований текстовий матеріал.

Важливою особливістю шляхом її застосування до відео, які містять складні елементи: змінну кількість мовців, фонові візуальні ряди, використання термінології з окремих галузей знань, а також закадровий супровід. Такий вибір дозволяє оцінити гнучкість, універсальність та якість роботи запропонованого підходу в умовах, максимально наближених до реальних сценаріїв інформаційного середовища.

Протягом дослідження експериментально реалізується повний цикл конверсії. З відеофайлу витягується звукова доріжка, далі вона трансформується в текст за допомогою систем автоматичного розпізнавання мови, після чого текст обробляється з метою очищення, редагування та логічного структурування. На фінальному етапі здійснюється перетворення тексту в структурований формат матеріалу, який доповнений релевантними зображеннями.

Особливий інтерес становить перевірка здатності системи формувати змістовно обґрунтовану структуру матеріалу. А саме виділення ключових смислових моментів, формулювання заголовків, визначення тем матеріалу, а також генерацію скріншотів, що ілюструють змістовні етапи мовлення, відповідно до кожної частини тексту. Значення мають не лише технічні показники швидкодії та точності, але й відповідність естетичним і комунікативним нормам сучасного контенту, придатного для поширення в цифрових медіа.

Експериментальне підтвердження життєздатності такого підходу може відкрити нові перспективи в автоматизації медіапроцесів, зокрема у створенні адаптивних контентних рішень для ЗМІ, соціальних платформ, цифрової освіти або професійних баз знань.

Отже, дослідження має на меті не лише практичну реалізацію процесу трансформації відео у текстовий формат, а й формування підходу, що міг би бути покладений в основу нових інтелектуальних медіаплатформ. Ці платформи можуть надавати користувачу можливість отримувати адаптований, багатформатний контент без втрати змісту, структури та візуальної репрезентації оригіналу.

Прикладом одного з найбільш відомих сервісів структурування тексту з відеозапису можна виділити відеохостинг Youtube. Він дозволяє автоматично створити субтитри до відео, що дозволяє зрозуміти матеріали навіть якщо вони були створені на іншій мові.

Автоматично згенеровані субтитри є лише переліком слів, які вдалося розпізнати та розбити на короткі фрагменти по кілька секунд для відображення користувачу.

Якість розпізнавання не враховує тематику чи контекст відео, а базується переважно на визначенні слів з найвищою імовірністю. Також відсутня пунктуація, що перетворює субтитри на звичайний набір слів, який може довільно розривати речення, ускладнюючи правильне розуміння й сприйняття матеріалу.

Основна частина

У сучасному інформаційному просторі відеоконтент відіграє дедалі важливішу роль як джерело даних, що може бути ефективно трансформоване у текстовий формат для подальшого використання в наукових, освітніх або публіцистичних цілях. Процес конверсії відео у структуровану статтю потребує чітко визначеного алгоритму дій, починаючи від вибору відповідного матеріалу до обробки аудіо- та текстових даних.

У цьому дослідженні розглянуто методи, інструменти та підходи, що забезпечують трансформацію відео в повноцінну публікацію наукового або інформаційно-освітнього характеру. Для повноцінного тестування запропонованої методики необхідно обрати комплексний відеоматеріал, який дозволить всебічно перевірити кожен етап перетворення та виявити як його сильні сторони, так і потенційні недоліки.

На першому етапі реалізації було визначено ключові критерії, яким має відповідати відео для подальшої адаптації у формат структурованого тексту. По-перше, матеріал повинен мати чітку тематичну спрямованість, що дозволить логічно структурувати текст за змістовими розділами.

По-друге, стиль оповіді у відео має бути наративним або науковим, що сприятиме збереженню відповідної стилістики у текстовій версії. Такий підхід дозволяє отримати змістовний і водночас інформативний матеріал, придатний для формування структурованої тематичної доповіді.

Для реалізації експерименту було обрано відео, розміщене на офіційному YouTube-каналі Харківського Національного університету радіоелектроніки (ХНУРЕ). Зокрема, вибрано відеоматеріал, присвячений історії Кафедри медіасистем та технологій (МСТ), тривалістю понад 10 хвилин, що містить значний обсяг аудіоінформації.

Важливо зазначити, що відео було записане приблизно п'ять років тому переважно російською мовою. Це створює певні труднощі під час адаптації, проте водночас відкриває додаткові можливості для перекладу та трансформації контенту з урахуванням мовних потреб цільової аудиторії.

Оскільки YouTube як основна платформа розміщення відео не передбачає можливості безпосереднього завантаження відеофайлів для офлайн-обробки, контент транслюється у форматі потокового мовлення. Це значно ускладнює процес прямої конвертації відео у формат, придатний для подальшої транскрипції та аналізу.

З огляду на це, виникає потреба у використанні сторонніх рішень, як онлайн-сервіси та програмних бібліотек, які дозволяють завантажувати відеофайли з відкритих джерел. Одним із таких рішень є сервіс *savefrom.net*, що забезпечує зручне завантаження відео у форматі, придатному для обробки, з базовим збереженням якості зображення та звуку.

У межах поточного дослідження відеофайл було завантажено у форматі .mp4 з роздільною здатністю 360p, що забезпечило компроміс між розміром

файлу (65,3 МБ) та якістю аудіо. Оскільки основною метою є подальше вилучення та обробка аудіодоріжки, висока якість відеозображення не є критичною. Зменшення роздільної здатності сприяє пришвидшенню завантаження та спрощує подальшу обробку.

Однією з ключових цілей дослідження є забезпечення можливості подальшої автоматизації процесу конверсії, що дозволить мінімізувати або повністю усунути необхідність ручного втручання користувача на кожному етапі.

Серед доступних програмних засобів одним з найпоширеніших та найефективніших рішень є кросплатформна бібліотека FFmpeg, яка вважається індустріальним стандартом для обробки медіафайлів. Цей інструмент підтримує широкий спектр форматів і забезпечує високу гнучкість у налаштуванні процесів конвертації, кодування, декодування та фільтрації.

Вилучення аудіозапису з відеофайлу здійснювалося за допомогою простого консольного запиту, у результаті чого було отримано окремий аудіофайл у форматі .mp3 розміром 14 МБ.

Обраний формат є зручним для подальшої транскрипції, оскільки поєднує високу якість звуку з помірним обсягом файлу. За потреби можливе перетворення у формати з вищою якістю, наприклад .wav, проте в цьому дослідженні вирішено не застосовувати додаткове стиснення, щоб уникнути втрати якості, яка може вплинути на точність подальшого розпізнавання мовлення.

Окрім роботи з аудіо, FFmpeg також дозволяє здійснювати вилучення окремих кадрів із відео, що може бути корисним для створення ілюстративного матеріалу. Скріншоти з оригінального відео мають потенціал значно підвищити інформативність підготовленого контенту, доповнивши текст візуальними елементами, які полегшують його сприйняття. Такий підхід сприяє кращому передаванню атмосфери, хронології та ключових моментів подій, що висвітлюються у відео.

Після отримання якісного аудіофайлу відкривається можливість переходу до етапу транскрипції, тобто автоматизованого або напівавтоматичного перетворення мовлення у текст. Для цього можуть бути використані спеціалізовані сервіси, що підтримують обробку української та російської мов і забезпечують можливість подальшого перекладу.

Отриманий у результаті транскрипції текст стане основою для формування повноцінного медіаматеріалу, який відповідатиме вимогам публіцистичного стилю й буде доповнений тематичними зображеннями та структурованими підрозділами.

Оригінальний текст із вибраного відео містить приблизно 1600 слів і 12800 символів.

Вибір та аналіз сучасних сервісів транскрипції тексту з відео

Для роботи з сервісами транскрипції були обрані провайдери популярних AI-технологій, як OpenAI та Google Cloud. Додатково було протестовано сервіс AssemblyAI, який хоча й має вищу вартість, проте надає розширений функціонал, зокрема можливість визначення окремих мовців під час транскрипції, а також оцінку ймовірності кожного розпізнаного слова.

Тестування кількох сервісів є важливим, адже кожен із них має свої особливості, налаштування та рівень підтримки різних мов. Більшість таких платформ орієнтовані насамперед на англomовний контент, що може створювати труднощі під час обробки матеріалів українською мовою.

У сучасних умовах стрімкого розвитку інформаційних технологій та, зокрема, штучного інтелекту, критично важливо залишатися гнучким і готовим до переходу на нові, більш ефективні моделі й сервіси. Залежність від одного провайдера створює ризик втрати якості або актуальності у зв'язку з технологічним прогресом конкурентів.

Можливість надсилання запитів до кількох постачальників AI-моделей забезпечує гнучкий вибір найбільш сучасного й точного рішення на кожен конкретний момент часу. Оцінка ефективності кожної моделі потребує попереднього тестування та порівняння з вже відомими результатами. Для автоматизації такого тестування можна використати кілька відео із заздалегідь відомими текстами, які дозволять порівняти оригінал з транскрибованим результатом. Порівняння результатів дає змогу визначити відхилення за кількістю символів або слів, що дозволить обчислити коефіцієнт помилки. Цей показник, у свою чергу, стане базою для подальшого порівняння якості транскрипції між різними моделями.

Моделі OpenAI. Процес роботи з OpenAI API є дуже простим та досить зручним, оскільки вимагає лише надання особистого ключа до API, який можна згенерувати у панелі керування. Текст, який був отриманий в результаті виконання транскрипції, має пунктуацію та орфографічні виправлення.

У результаті транскрипції за допомогою моделей OpenAI були допущені певні помилки та є декілька пропусків тексту, але загалом можна відзначити чудову якість визначення ПБ, дат та назв, які були згадані у тексті. Пропущений текст не мав значну важливість, а помилки у словах можна виправити у подальшому за допомогою моделей ChatGPT від OpenAI.

При аналізі тексту з оригіналу відео та результату транскрипції було виявлено 9 помилки при визначенні слів та 11 помилок у пропусках слів або помірних помилок у синтаксисі слів, що має дуже маленький відсоток помилок – 0.56% та 0.08%, відповідно до кількості слів та символів.

Додатково у моделях транскрипції від OpenAI існують параметри передачі додаткового контексту до перекладу, які дозволяють значно покращити якість визначення слів та їх відносність до якоїсь конкретної сфери діяльності або об'єктів. Підказки можуть допомогти виправити певні слова або акроніми, які модель неправильно розпізнає в аудіо [4].

При автоматизації процесу конверсії відео у текстові матеріали може бути відсутня можливість надання достатнього контексту матеріалу до запиту на транскрипцію через відсутність спеціально підготовлених даних для цього. Головним джерелом корисних даних про відео може бути його назва та опис з таких сервісів, як Youtube, але дуже часто опис спотворений рекламою та різними деталями, які не відносяться до відео, а назва відео може бути створена лише з метою збільшити кількість переходів на нього без відображення сутності матеріалу.

Оскільки наявність такого контексту може допомогти правильно розпізнати певні слова та терміни, видобуток таких даних може значно підвищити якість транскрипції. Одним із варіантів є підхід до автоматизованої конверсії відео та аудіо матеріалів у текст, де той самий матеріал передається повторно до обробки, але вже з текстовим описом про те, до чого належить матеріал, в чому його сутність та які особливості можуть бути значущими для його визначення [5].

Отримання таких даних можливо з першого варіанту транскрипції аудіозапису, який додатково обробляється спеціальним запитом до чат моделей генеративного штучного інтелекту, які надають опис матеріалу та визначають його ключові особливості, які після цього відправляються на повторну обробку аудіозапису.

Особливості назв, аббревіатур, імен та назв об'єктів також можливо отримати за допомогою додаткової обробки веб-сайтів, які можуть відноситися до відео, над яким ведеться обробка. Наприклад, відео у поточному дослідженні відноситься до кафедри МСТ ХНУРЕ, що дозволяє автоматично обробити декілька перших запитів з пошуку у Google та аналогічних пошуковиках для того, щоб відкрити найбільш релевантні сторінки, додатково обробити їх для визначення ключових даних для покращення якості транскрипції і відправити на новий етап обробки з вже знайденими даними.

Сучасні моделі від OpenAI дозволяють відправити простий текстовий запит, який дозволить системі самостійно зробити пошук в інтернеті та виділити ключові слова та особливості, які відносяться до конкретно розглядаємої теми, або предмету.

Сервіс Google Cloud speech-to-text. Робота з Google Cloud speech-to-text є більш складною з точки зору взаємодії з API. Якщо для OpenAI достатньо лише надати ключ до API, то в Google Cloud ця опція доступна лише для файлів, довжина запису яких не більше однієї хвилини. Якщо ж довжина більша, то необхідно створити внутрішнє сховище та попередньо завантажити туди файл. Тільки після попереднього завантаження з'являється можливість надати на нього внутрішнє посилання та відправити запит на транскрипцію. Також замість ключа до API потрібно створювати спеціальний акаунт з доступами до зміни сховища, у яке ми завантажуюмо файл та необхідно згенерувати ключі доступів, які можна зберегти у файл ("key.txt" у нашому випадку) та постійно надавати до авторизації.

Після завантаження файлу аудіозапису у Cloud Storage, який ми отримали із відео, ми отримуємо можливість відправлення завантаженого файлу до сервісу speech-to-text для транскрипції. Важливо зазначити, що сервіс надає можливість розпізнавання різних голосів у запису, але ця можливість працює лишень як бета-версія. Спробуємо задати додатково кількість голосів більше двох, щоб сервіс спробував визначити унікальні голоси на записі.

У отриманому результаті не було визначено унікальні голоси. Найбільш вірогідно, що дана функція працює тільки для англійської мови. Головним недоліком можна відзначити відсутність будь-якої пунктуації та перевірки орфографії. Також контекст зовсім не перевіряється і частини тексту можуть виглядати дуже непов'язаними одна з одною.

Схоже, що даний сервіс виконує транскрипцію тим самим чином, як і транскрипція, яка автоматично виконується у відео на відеохостингу Youtube.

У даному випадку можна відзначити набагато більшу кількість помилок, неправильні визначення дат, технічних назв та ПІБ. Також існує набагато більша кількість пропущень тексту, ніж у попередньому випадку.

При аналізі тексту з оригіналу відео та результату транскрипції було виявлено 31 помилки при визначенні слів та 40 помилок у пропусках слів або помірних помилок у синтаксисі слів, що має набагато більший відсоток помилок, ніж попередній результат – 2.6% та 0.31%, відповідно до кількості слів та символів.

Сервіс Assemblyai. Даний сервіс окрім зручного використання (у порівнянні з Google Cloud) також пропонує можливість визначення голосів різних людей у запису, додавши до запиту параметр “speaker_labels: true”. Також необхідно вказати мову оригіналу для якісної обробки.

Сервіс зміг визначити різні голоси на запису, але лише частково. Один з голосів був визначений неправильно, бо та сама людина продовжувала говорити, а також ще один голос було визначено правильно, хоча людина на відео змінилася. Крім цього, інші голоси визначені правильно та це можна використовувати для формування структури статті. У подальшому така можливість буде невід'ємно важливою для транскрипції відео з інтерв'ю, бо саме у таких відео критично важливо розділити запитання та відповіді. У поточному варіанті це допоможе краще структурувати текст, але досконало визначити усіх унікальних розповідачів не вийшло.

У даному випадку можна відзначити високу якість визначення тексту та мінімальні помилки. Майже усі імена, назви та технічні терміни були визначені правильно. Існують одиничні помилки, які дійсно можна почути на записі, що може свідчити про погану якість запису, або погане вимовляння розповідача.

Окрім високої якості визначення тексту також були визначені унікальні розповідачі. Хоча визначення і мало певні неточності (один пропущений та один визначений неправильно), це може допомогти в написанні статті.

При аналізі тексту з оригіналу відео та результату транскрипції було виявлено 7 помилки при визначенні слів та 10 помилок у пропусках слів або

помірних помилок у синтаксисі слів, що має набагато більший відсоток помилок, ніж попередній результат – 0.58% та 0.08%, відповідно до кількості слів та символів.

Серед оглядаємих сервісів саме Google Cloud text-to-speech мав найгіршу якість як з точки зору цільності, пунктуації, синтаксисі та правильному визначенні імен, назв та дат. Формування текстового матеріалу з такої транскрипції мало би дуже погану якість та помилки, що могло би призводити до вводи в оману читачів та мати погані наслідки, якщо на сформованому тексті хтось буде робити певні висновки, або вести свою наукову роботу.

Обидва сервіси від OpenAI та AssemblyAI дуже якісно виконали транскрипцію тексту з мінімальними помилками. Якість визначення навіть викликає певний подив з точки зору точності ПШБ, які були згадані у відео-записі. Навіть самостійне прослуховування у тих, хто не знайомий зі згаданими особистостями, може викликати запитання до озвучених прізвищ, хоча згадані сервіси чудово справилися з цим завданням.

На відміну від OpenAI, сервіс AssemblyAI додатково визначив унікальних розповідачів, що додає певних додаткових переваг для структурування статті, тому для подальшої роботи ми зупинимося на даному варіанті.

Обробка та структуризація тексту

Для отримання якісного текстового матеріалу та збереження графічної складової відео потрібно зберегти важливі часові мітки в тексті, щоб потім замінити їх на зображення, які точно відповідатимуть конкретному моменту у відео. Це допоможе передати візуальну частину й доповнити сприйняття тексту відповідно до початкового матеріалу.

Оскільки кожен із сервісів надає часові мітки для кожної визначеної частини тексту з аудіозапису, ми можемо обрати частоту вставки зображень відповідно до часу або кількості символів.

Також є можливість вставляти зображення лише під час зміни теми або зміни розповідача. Типові чат-моделі генеративного штучного інтелекту дають змогу визначити такі місця відповідно до запиту користувача.

Кількість міток повинна відповідати бажаній кількості зображень, яке планується використовувати у кінцевому медіаматеріалі, виходячи із довжини тексту та розміру зображень.

Потенційно можливо залишити часові мітки для кожного визначеного речення та подальшого визначення того, які з них будуть залишені у кінці процесу обробки, але надмірна кількість зайвих символів та числових міток може негативно вплинути на інші етапи обробки тексту. На рис. 1 відображено приклад тексту з часовими мітками.

Перед передачею тексту для його зміни ці часові мітки важливо визначити у тексті, відокремивши їх спеціальними символами (як-от квадратні дужки). Також необхідно повідомити в інструкції до ШІ про те, щоб він не видаляв, не змінював їх та не звертав на них уваги при зміні тексту. На жаль, навіть при

наданні чіткої та всеосяжної інструкції ІІІ все одно може змінювати текст та ігнорувати певні частини запитів до себе, що викликає потребу додатково аналізувати результат його роботи перед тим, як виконувати над ним подальші кроки. У випадку з визначенням часових міток можна досить легко перевірити, скільки саме їх було перед опрацюванням та після цього.

[10:33]

Паперова книга на сьогоднішній день зазнає досить серйозної конкуренції з боку інших джерел. Читачі віддають перевагу дайджестам багатосторінкових електронних видань, відеоконтенту, а також мультимедіа-контенту. Саме такі мультимедіа-електронні видання створюють наші студенти на кафедрі медіа-системи та технологій, які сьогодні дуже затребувані на ринку праці. І в повному обсязі на сучасних платформах вони представлені курсами нашої кафедри.

[11:01]

Новою для нашої кафедри спеціальністю є спеціальність технологій електронних мультимедійних видань, яка зараз дуже популярна, дуже популярна серед наших студентів. та основними напрямками цієї спеціальності є створення мультимедійних видань, розробка веб-сайтів та створення мобільних додатків. Щодо продуктів, які зможуть наші випускники створювати після закінчення цієї спеціальності, це веб-сайти, насамперед статичні, динамічні, **[11:31]**це мобільні програми. Ну і, звісно, мультимедійні видання, це вміння створювати аудіо, відео, 3D, анімацію. все для того, щоб створювати повноцінні мультимедійні видання. Один із напрямків роботи нашої кафедри – підготовка фахівців, дизайнерів інтерфейсу електронних та мобільних додатків. Ця область дуже сильно зараз розвивається, як через створення нових технологій, застосування нових технологій доповненої віртуальної реальності, голосовий інтерфейс, через більш поглиблене опрацювання окремих деталей, створення прототипів, алгоритмів автоматизованого створення інтерфейсів, так і через розширення сфери діяльності дизайнера на рівень стратегії продукту та бізнесу в цілому.

[12:16]

Основна робота колективу «Кафедру з нового набору» проводиться протягом усього навчального року. Тут різноманітні заходи університету, дні відкритих дверей, що відбуваються протягом усього року, участь у виставках, конференціях, олімпіадах, пікніках, науках та багатьох інших заходах, в яких можна розкритися та показати таланти наших студентів. У навчальному плані підготовки бакалаврів маємо дуже багато творчих спеціальностей. Основи графічного дизайну, художні основи проектування видань **[12:47]** обробка відео, обробка аудіоінформації та багато інших дисциплін, які саме пов'язані з творчістю. Не випадково ми часто говоримо, що спеціальність видавництва та поліграфія поєднують творчі можливості з гідністю точних наук. Наша кафедра є провідною у ВНЗ. за кількістю перемог на олімпіадах та студентських конкурсах, саме з поліграфії, це видавництво та поліграфія у Києві, молотиполіграфія у Львові, студентські олімпіади з поліграфії, дизайну, упаковки. Більшість наших студентів самі зацікавлені участю у всіх цих конкурсах. Завданням викладача в основному є просто направити їх, допомогти розвинути свій талант, любов до цієї справи і добре виступити на заходах, що проводяться.

[13:36]

Наші випускники завжди знайдуть собі роботу. Це підтверджує специфіку працевлаштування наших випускників. Зараз вони розкидані не лише територією всієї України, у нас є хлопці, які влаштувалися в Мексиці, Австралії, Німеччині, Англії, Польщі. Вони працюють за фахом, роботодавці задоволені їхніми професійними навичками і ми сподіваємось на те, що міжнародне **[14:07]**трудове співробітництво у нас тільки розширюватиметься і надалі. Хочеться побажати нашій кафедрі добрих та талановитих студентів, успіхами та досягненнями яких ми могли б пишатися. А вони, наші майбутні випускники могли гордо сказати «Я випускник ФНОРЕ» у спеціальності видавництва ТАП поліграфія. Насамкінець, аналізуючи в історичному аспекті шлях нашої кафедри на освітній ниві, хотілося б сказати таке. Може не завжди ми, і наші попередники, на сміливість сказати їхнє ім'я, перебували на висоті тих вимог, які до нас прийшли. пред'являв час у плані підготовки кадрів, наукових розробок, навчальної роботи, але динаміка розвитку кафедри у

Рисунок 1 – Текст з виділеними часовими мітками

Якщо кількість сильно змінилася або часові мітки всередині було видозмінено, можна відправити новий запит з тими самими параметрами (бо ІІІ може надавати різні результати на той самий запит залежно від моделі), або відправити наступне повідомлення разом із попередніми та надіслати інструкцію до виправлення помилок.

Для створення зображень відповідно до збережених важливих часових міток можна також скористатися бібліотекою FFmpeg, яка відповідно до наданих запитів створить зображення необхідного формату. На рис. 2 відображено список збережених кадрів з обраного відео.



Рисунок 2 – Кадри з обраного відео, відповідно до часових міток

Оскільки обраний відеозапис був записаний близько 5 років тому, на ньому розмовники використовують переважно російську мову. Для покращення якості матеріалу та актуалізації до поточної мови наукового, освітнього та суспільного соціуму ми перекладемо статтю українською мовою. Переклад текстового матеріалу додасть додатковий шар обробки, що може призвести до зайвих труднощів, але це дозволить підтвердити можливість конвертації матеріалів різними мовами, що повинно підвищити охоплення та сприйняття створених матеріалів.

При використанні моделі GPT-4o текст було якісно перекладено зі збереженням часових міток, які не вплинули на точність перекладу. При запиті до моделей генеративного штучного інтелекту важливо вказати подробиці про формат тексту, наявність часових міток та потребу в збереженні початкової сутності без зміни тексту.

Оскільки при автоматизації конверсії відеоматеріалів та подальшій обробці тексту може виконуватися велика кількість операцій перетворення, це може спричинити виникнення додаткових помилок.

Транскрипція, переклад, створення часових міток та подальша зміна тексту потребують додаткової перевірки результату. Для цього можна відправити початковий та кінцевий текст на перевірку до тих самих моделей ChatGPT із запитом перевірки, порівняння та виявлення помилок і виправлень.

Порівняння початкового та кінцевого станів тексту може дозволити визначити та виправити помилки або небажані зміни, які відбулися з текстом. Цей крок можливо виконувати при кожному перетворенні тексту, відправляючи додатковий запит до моделей ШІ для порівняння текстів та негайного виправлення помилок.

Такий підхід може значно підвищити якість конверсії текстових матеріалів, але водночас відбувається підвищення використання відповідних сервісів, що варто враховувати при розрахунку вартості конверсії.

Після отримання відредагованого, переведеного та виправленого тексту, необхідно приступити до формування статті. Частіше за все, статті формуються по підрозділах із зображеннями.

Для переформатування тексту у формат статті з заголовками потрібно визначити основні теми та частини, які були виділені у ньому. Для адаптації і зручності подальшого використання текст зі структуруванням необхідно оформити у форматі HTML, що дозволить повністю автоматизувати використання на веб-ресурсах.

HTML-формат також дозволить створювати спеціальні графічні матеріали шляхом створення скріншотів та переформатування у спеціальні формати для публікації у соцмережах, які не дозволяють розміщувати довгі та структуровані текстові матеріали, що дозволить підвищити охоплення.

Для структуризації тексту та перетворення у формат HTML необхідно відправити запит до моделей генеративного штучного інтелекту з інструкціями у такому форматі.

1. Рівномірно розділи текст на 3-5 частин, а потім надай назву кожній частині.
2. Збережи кожне речення з початкового тексту, нічого не видаляй!
3. Не видаляй значення у дужках "[...]", збережи їх, вони дуже потрібні для подальшої вставки зображення.
4. Надай результат у форматі html.

У результаті було отримано добре структурований матеріал з заголовками та рівномірно розподіленим текстом.

Останнім етапом потрібно замінити часові мітки на зображення відповідно до створених скріншотів з відео. Додатково, оскільки при форматуванні тексту часові мітки у деяких випадках перемістилися у середину тексту, необхідно перенести їх на початок речення – це допоможе правильно сприймати матеріал при його перегляді.

Результати дослідження

У результаті проведеного дослідження була створена “методика автоматизації конверсії відеоматеріалів у структурований текст із релевантними зображеннями”. Ця методика включає дев’ять послідовних етапів, кожен з яких спрямований на покращення якості кінцевого результату.

1. На початку відбувається завантаження та попередня обробка обраного відеофайлу. Із нього відокремлюється аудіодоріжка, а також зберігаються кадри відео у форматі зображень. Ці зображення можуть бути вибрані за фіксованими інтервалами або за ключовими моментами вмісту вже після транскрипції та аналізу тексту.

2. Виконується транскрипція відео за допомогою обраного сервісу розпізнавання мовлення. Сервіс підбирається з урахуванням таких вимог, як підтримка обраної мови, наявність спеціальних функцій та параметрів, які можуть покращити точність розпізнавання або надавати такі можливості, як визначення унікальних голосів та розподіл тексту відповідно до них.

3. Після отримання первинного тексту здійснюється визначення контексту вихідного відеоматеріалу. Аналізуються тематика, характерна термінологія та можливі особливості мовлення, що дозволяє краще підготувати транскрипт до подальшої обробки, враховуючи ймовірні неточності чи помилки.

4. Для покращення точності визначення слів у тексті виконується повторне відправлення аудіозапису на транскрипцію, але вже з урахуванням раніше визначених особливостей. Додатковий контекст передається у вигляді інструкції до сервісу, що дозволяє точніше розпізнати рідковживані або специфічні слова, включаючи власні назви, технічні терміни тощо.

5. Після повторної транскрипції виконується виправлення орфографічних та пунктуаційних помилок у тексті. Це підвищує загальну якість тексту та забезпечує його відповідність мовним нормам.

6. На проміжному етапі проводиться додаткова обробка тексту залежно від потреб. Це може бути переклад іншою мовою або зміна стилістики (наприклад, адаптація під науковий, офіційний чи публіцистичний стиль). Для цього використовуються сервіси генеративного штучного інтелекту або спеціалізовані інструменти перекладу та редагування.

7. Визначаються основні частини та теми тексту з метою створення заголовків і підзаголовків. Це допомагає структурувати текст для кращого сприйняття та подальшої адаптації під різні платформи, включаючи блоги, соцмережі чи освітні сайти.

8. Після структуризації текст перетворюється у HTML-формат. У ньому застосовуються відповідні теги для оформлення заголовків, списків, гіперпосилань та інших елементів, що дозволяє легко інтегрувати підготовлений матеріал на інтернет-ресурсах.

9. Завершальний етап полягає в інтеграції зображень у текст. У відповідних місцях тексту, де були часові мітки або описані ключові моменти,

вставляються зображення, отримані раніше з відео. Це забезпечує візуальну підтримку інформації та покращує сприйняття матеріалу кінцевим користувачем.

Розроблена методика була протестована на складному відеоматеріалі, що містив велику кількість визначень, імен, назв, а також декілька різних розповідачів із різними манерами мовлення.

Такий формат відео дозволив перевірити, наскільки ефективно система справляється із транскрипцією, обробкою мовного контенту, розпізнаванням рідковживаних термінів та адаптацією тексту під структурований формат.

У результаті тестування було отримано повноцінний структурований медіаматеріал у форматі HTML, що включає кілька ключових складових. Зокрема, структура готового документа охоплює:

- заголовки, що відповідають окремим темам або логічним частинам відео, завдяки чому матеріал легко сприймається, а навігація стає зручною як для читача, так і для технічної інтеграції в системи пошуку;

- перекладений текст, у якому було попередньо виправлено пунктуаційні та орфографічні помилки. Текст рівномірно розподілений між заголовками та зображеннями, що забезпечує логічну послідовність і полегшує читання;

- зображення, які були автоматично отримані шляхом фіксації ключових кадрів із відео. Кожне зображення прив'язане до відповідного фрагмента тексту відповідно до моменту часу, коли розповідається певна частина змісту. Це дозволяє візуально підсилити сприйняття інформації;

- додаткове форматування в HTML, яке включає такі елементи, як марковані та нумеровані списки, гіперпосилання, стилістичні теги для виділення тексту та інші засоби оформлення, що забезпечують сучасний вигляд і технічну сумісність з більшістю онлайн-платформ.

Результати тестування підтвердили практичну придатність методики для обробки складного відеоконтенту та створення на його основі якісного, структурованого й мультимедійно підтриманого текстового матеріалу.

У результаті виконання дослідження було вирішено низку задач, спрямованих на розробку, перевірку та вдосконалення методики автоматизованої конверсії відеоматеріалів у структурований текстовий формат з використанням сучасних цифрових інструментів.

1. Визначено та протестовано інструменти, необхідні для отримання і попередньої обробки відеофайлів. Це включає програмні засоби для відокремлення аудіодоріжки та збереження окремих кадрів у форматі зображень.

2. Обрано кілька популярних сервісів транскрипції аудіо та проведено їхнє порівняння на основі комплексного тестового відео, яке містить рідковживані слова, власні імена, географічні та комерційні назви, а також терміни з вузькопрофільних сфер. Це дозволило оцінити ефективність кожного інструменту в умовах підвищеної складності.

3. Детально проаналізовано переваги, недоліки та специфічні особливості роботи розглянутих транскрипційних сервісів, включаючи їхню точність, швидкість та обсяг підтримуваних мов, гнучкість налаштувань.

4. Запропоновано методику автоматизованого тестування нових моделей транскрипції, що дозволяє швидко й об'єктивно оцінювати їхню ефективність у порівнянні з попередніми рішеннями. Зокрема, було розроблено підхід до розрахунку коефіцієнтів помилок у поєднанні з вартісною ефективністю, що дозволяє приймати обґрунтовані рішення щодо доцільності переходу на нову модель.

5. Проведено повний цикл транскрипції аудіодоріжки відео. Отриманий текст було відредаговано для усунення граматичних, орфографічних та пунктуаційних помилок, після чого виконано переклад, що дозволило забезпечити мультимовну доступність контенту.

6. Сформульовано універсальний запит до моделей штучного інтелекту для автоматичної структуризації тексту. В результаті текст було поділено на тематичні блоки з відповідними заголовками, що значно покращило зручність сприйняття та подальшу адаптацію до різних форматів.

7. Здійснено перетворення структурованого тексту у формат HTML із використанням відповідних тегів для заголовків, списків, посилань та інших елементів веб-оформлення.

8. На основі оброблених даних сформовано фінальний варіант текстового матеріалу. До нього були додані релевантні зображення, синхронізовані з відповідними фрагментами тексту, що забезпечило візуальну підтримку та посилення змістовної складової.

9. Проведено аналіз розробленої методики з точки зору її ефективності та точності. Виявлено переваги та недоліки обраного підходу, а також запропоновано шляхи його подальшого вдосконалення та розширення функціональності відповідно до потреб різних категорій користувачів.

В результаті було отримано готовий матеріал у форматі HTML, придатний для негайного розміщення на вебсайтах та інших інтернет-платформах. Такий підхід дозволяє автоматизувати процес публікації без потреби у додатковому редагуванні. За необхідності, стиль і розмітку матеріалу можна легко змінити шляхом коригування CSS-стилів або HTML-тегів. На рис. 3 представлено приклад кінцевого медіаматеріалу, згенерованого на основі обраного відеозапису.

Серед можливих напрямів подальшого вдосконалення розробленої методики можна виокремити кілька перспективних рішень, що сприяють розширенню її функціональності та адаптації до сучасних форматів цифрового контенту. Це особливо актуально для платформ, на яких не рекомендується або неможливо розміщувати великі обсяги текстової інформації чи повноцінні відео.

HTML-формат, у якому формується кінцевий структурований матеріал, відкриває широкі можливості для автоматизованої адаптації цього контенту у графічний та інші візуальні формати. Завдяки наявності чітко окреслених заголовків, тематичних блоків і зображень, з'являється змога автоматично генерувати інфографіку, графічні картки та компактні візуальні публікації.

Нові напрямки та досягнення

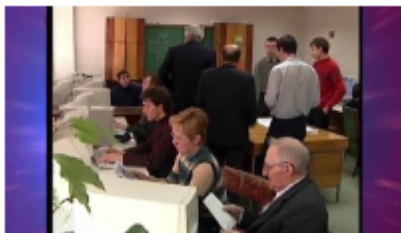
У 1997 році абсолютно новий етап у нашій роботі. Після консультації з викладачами, з ректором Української академії друку, видавничо-поліграфічного інституту КПІ, тоді він ще був факультетом. За їх підтримки відкриваємо нову спеціальність. Кафедра стає випусковою технологій автоматизованої обробки інформації. Таким чином відновлюємо підготовку фахівців у місті Харкові для видавничо-поліграфічної галузі. У постанові Кабінету Міністрів України на початку 2000-х років затверджено новий напрямок підготовки фахівців — видавничо-поліграфічна справа. Кафедра поступово ліцензує спеціальності, які входили до складу цього напрямку.



2002 рік — спеціальність комп'ютерні технології систем видавничо-поліграфічних виробництв. 2005 рік — технології друкованих видань спеціальність і 2011 рік — сучасна спеціальність технології електронних мультимедійних видань.

Напружено працюють у цей час викладачі кафедри над розробкою нового методичного забезпечення. Допомагає нам у цьому і

спираємося на досвід Київського політехнічного інституту, видавничо-поліграфічного факультету та Української академії друку Львів. На сьогодні видано вже спільно з ними в тому числі 15 навчальних посібників, кілька монографій. У цей час розвиваються нові наукові напрямки на кафедрі. Деякі з них — проблеми обробки цифрових зображень та кольоровідтворення в поліграфії, комп'ютерна графіка очолює напрямком професор Кулішова Н.Е. та інформаційних систем і технологій Г.Н.І.



Технології розробки електронних мультимедійних видань Єгорової Ірини Миколаївни, системи автоматизації управління поліграфічним виробництвом — професор Левикін Ігор Вікторович.

Успішні проекти та міжнародна співпраця

У 2002 році наказом ректора заснована науково-дослідна лабораторія геоінформаційної системи і комп'ютерна графіка, яка виконала цілий ряд проектів. До успішних проектів, виконаних у лабораторії, я б відніс наступний. 2004 рік — Харкову 350 років. Харківський госпкомітет замовляє нам комплексний проект, який ми реалізували спільно з міським інформаційним центром, картографічним інститутом «Схід Геоінформ» та регіональним центром «Сприяння в бізнесі». У складі проекту була реалізована електронна векторна карта міста Харкова масштабу 1:10 тисяч, поліграфічний атлас міста Харкова. На базі електронної карти була реалізована геоінформаційна довідкова система про об'єкти міста Харкова, випущена на CD-диску.



І крім того, ще один CD-диск, випущена була інформаційна система підприємств міста Харкова, яка з такою претензією назвою вийшла «У Харкові все є». Ми за цю роботу отримали подяку від міського виконавчого комітету міста Харкова та його голови.

Окрім цього, був реалізований ще цілий ряд ГІС-проектів. У цей час, крім ХОС-договірних, виконувалися також держбюджетні теми.

Наприклад, технології автоматизованої обробки зображень і кольоровідтворення у відкритих поліграфічних системах. Відповідальний виконавець — професор Кулішова Нонна Євгенівна. За результатами цієї роботи була видана монографія. Зараз кафедра продовжує виконувати замовлені госпдоговірні роботи. Серед них — це система відеоконтролю якості зображень і також на замовлення книжкової фабрики Юнісук виконуємо роботу з розробки системи управління якістю продукції.

Рисунок 3 — приклад формату результату створення структурованого текстового матеріалу з доповненням графічних матеріалів

Розміщення ключових зображень, тематичних заголовків і коротких анотацій разом із гіперпосиланнями на повний матеріал дозволяє не лише розширити охоплення аудиторії, а й перенаправляти користувачів на платформи, більш зручні для споживання повного змісту.

З урахуванням тенденцій розвитку сучасних месенджерів та соціальних мереж, зростає популярність коротких, лаконічних публікацій, які складаються лише з кількох ключових тез та супроводжуються зображеннями. Такий формат дозволяє користувачеві швидко ознайомитися з основною ідеєю чи новиною без необхідності прочитання повного тексту. За бажанням, користувач може перейти за посиланням до повної версії матеріалу, що зберігає баланс між стислою подачею та глибиною змісту.

Форматування інформаційних карток у форматі HTML із подальшим зберіганням у вигляді графічних зображень дає можливість публікувати їх майже на всіх популярних платформах, включаючи Facebook, Twitter, Instagram, TikTok, Telegram та інші. Це розширює можливості крос-платформеного розповсюдження та підвищує впізнаваність контенту.

Крім того, після розділення оригінального відеоматеріалу на окремі зображення та структурований текст з нього можна автоматично створювати короткі відеоролики, що містять ключові теми та відповідні кадри з оригінального відео. Такий формат відеоісторій (stories) є надзвичайно популярним у соціальних мережах, зокрема в Instagram, YouTube та TikTok.

Висновки

У процесі виконання дослідження було здійснено розробку та апробацію інформаційної технології, що дозволяє автоматизовано трансформувати відеоконтент у структурований текстовий формат із контекстуальними візуальними вставками. Реалізований підхід довів свою ефективність при застосуванні до складних і неоднорідних відеоматеріалів, які включають декількох мовців, наявність фонових зображень, закадровий супровід, а також термінологічну насиченість, характерну для професійних чи тематичних відео.

Отримані результати підтверджують доцільність і практичну корисність такого підходу для вирішення широкого спектру задач у галузях медіатехнологій, освіти, цифрової журналістики та інформаційного забезпечення.

Однією з ключових переваг розробленої методики є її гнучкість на фінальних етапах формування статті. Зокрема, кількість, розміри та позиціонування зображень не є жорстко фіксованими й можуть адаптуватися залежно від потреб редактора, вимог видавничого середовища або особливостей платформи публікації. Цей рівень адаптивності дозволяє налаштовувати візуальне оформлення матеріалу без втрати цілісності змісту та без необхідності втручання у попередні етапи обробки.

Особливу увагу заслуговує ефективне використання часових міток, згенерованих під час транскрипції, які дозволяють точно синхронізувати

фрагменти тексту з відповідними візуальними елементами відео. Це забезпечує логічну та семантичну відповідність між текстовим і візуальним рядами, що суттєво підвищує рівень сприйняття матеріалу.

Експериментально підтверджено, що завдяки фіксації моментів переходу між мовцями можливе автоматичне формування маркерів прямої мови з додатковими атрибутами, наприклад, ініціалами мовця (якщо вони присутні в відеоматеріалі або аудіозаписі), що є надзвичайно важливим для збереження структури.

Важливою перевагою досліджуваного підходу є його технічна реалізованість через поєднання інструментів штучного інтелекту та утиліт для обробки медіаданих. Зокрема, кожен із ключових етапів, як транскрипція з розпізнаванням пунктуації, переклад, редагування тексту, формування HTML-структури і вставка зображень може бути реалізований за допомогою простих запитів до відповідних API або програмних засобів на кшталт ffmpeg. Це дозволяє у подальшому досягти повної автоматизації процесу без потреби у ручному втручанні.

Водночас проведене дослідження дозволило виявити і низку обмежень, які мають бути враховані при практичному впровадженні технології. Зокрема, незважаючи на загалом високу якість перетворення, можуть мати місце втрати окремих смислових акцентів, особливо у випадках, коли йдеться про складні терміни, фахову лексику або специфічні мовні звороти.

Для вирішення цієї проблеми запропоновано багатокрокову транскрипцію з уточненням даних, які можуть покращити якість визначення слів, відповідно до тематики та особливостей відео.

Також було виявлено, що при автоматичному доборі зображень іноді фіксуються кадри, які не є репрезентативними, зокрема, коли мовці закривають очі або вираз обличчя є непридатним для використання у статті. Це свідчить про необхідність реалізації механізму попереднього фільтрування скріншотів або створення декількох варіантів на вибір, що дозволило б покращити візуальне оформлення кінцевого документа.

Також для вирішення цієї проблеми можливо також обробляти та оцінювати зображення за допомогою моделей штучного інтелекту. Необхідно створити декілька скріншотів з одного часового проміжку та оцінити їх по відповідності до текстового матеріалу, який був викладений у відео. Кадри з найкращими оцінками потрібно додати до кінцевого матеріалу, а інші видалити.

Крім того, важливим аспектом є економічна складова реалізації. Попри низьку вартість окремих запитів до зовнішніх сервісів ШІ, масштабне використання або інтеграція в промислові інформаційні системи може вимагати фінансових витрат. Також варто враховувати обмеження, які пов'язані з підтримкою певних мов або діалектів, що наразі не охоплюються окремими мовними моделями і можуть призводити до підвищеного рівня помилок у транскрипції або перекладі.

У результаті дослідження було перевірено, що створена методика трансформації відео у текстовий формат з візуальним супроводом є ефективним інструментом, який може бути застосований у широкому колі сфер. Використання методики відкриває нові можливості для адаптації контенту до різних типів пристроїв та соціальних платформ.

Методика дозволяє забезпечувати зручність для користувачів із різними потребами, включно з людьми з вадами слуху або зоровими особливостями, а також дає змогу повторно використовувати зміст у нових інформаційних форматах, включно зі скороченим викладом, інтерактивною навігацією та швидким пошуком.

Проведене дослідження не лише підтверджує життєздатність і потенціал застосування інформаційної технології автоматизованої конвертації відео у текст із графічною адаптацією, але й окреслює напрями для її подальшого вдосконалення. Це стосується як підвищення якості конверсії матеріалів, так і розробки механізмів інтелектуального добору візуального контексту, що в сукупності забезпечить краще збереження смислової структури первинного відеоматеріалу у трансформованому цифровому форматі.

Список літератури.

1. Akshatha, P.S., Akash, B., Harshith, V., Sumukh, C., Reddy, V.G., & Shetty, S. (2024). NLP-Driven Video Transcription: A Comprehensive Survey and Innovative Office Meeting Automation. In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT). (p. 1-7).
2. Про.Світ. (2020, 1 березня). Дослідження стану реалізації дистанційного навчання в Україні (березень–квітень 2020). <https://prosvitcenter.org/product/doslidzhennia-stanu-realizatsii-dystantsiynoho-navchannia-v-ukraini-berezen-kviten-2020/>.
3. Greenberg, A.K., Gross, M., & Wright, M.C. (2014). Effects of image-based and text-based activities on student learning outcomes. In Proceedings of the Fourth International Conference on Learning Analytics And Knowledge. (p. 275-276).
4. OpenAI. Prompting. OpenAI Documentation. <https://platform.openai.com/docs/guides/speech-to-text#prompting>.
5. Левикін, І.В., & Шимко, Д.І., (2025). Підхід до автоматизації конверсії відео та аудіо матеріалів у текст. Радіоелектроніка та молодь у XXI столітті. (с. 624-626).