РЕФЕРАТ

Дипломну роботу виконано на 56 аркушах, вона містить 2 додатки та перелік посилань на використані джерела з 19 найменувань. У роботі наведено 19 рисунків та 1 таблиця.

**Актуальність теми.** На сьогоднішній день сучасним медіакомпаніям є необхідним мати розмічений архів новинних відеосюжетів для подальшої ефективної його обробки, отримання аналітичних даних, успішного спілкування з замовниками. Сама така задача класифікації новинних репортажів виникла на телеканалі «Україна». Новинні канали виконують класифікацію та обробку новинних сюжетів вручну або за допомогою підрядників, проте цю задачу можна виконати завдяки штучному інтелекту. Проблема є актуальною, адже такої системи немає у використанні ні в однієї медіа компанії України.

**Мета і задачі дослідження.** Метою даної дипломної роботи є створення математичного та програмного забезпечення для обробки й класифікації архіву новинних відеорепортажів.

При розробленні відповідного забезпечення потрібно розв’язати наступні завдання:

* проведення порівняльного аналізу існуючих методів nlp та cv;
* вибір та адаптація існуючого методу для вирішення задачі nlp та cv;
* розробка програмного забезпечення на базі вибраного математичного методу;
* тестування розробленої автоматизованої системи.

Реалізована система має задовольняти такі вимоги:

* визначати день тижня та номер тижня з назви відеофайлу;
* визначати start-end time кожного окремо сюжету;
* визначати хронометраж сюжету;
* визначати назву сюжету;
* визначати текст сюжету;
* визначати тип (новини, анонс або реклама);
* визначати журналістів та гостей;
* визначати ведучих;
* визначати теги;
* класифікувати сюжет на позитивний чи негативний;
* додавати до результуючої таблиці канал та назву новинної програми;
* мати високі показники ефективності розпізнавання start-end time сюжету (точність повинна бути до 1 секунди);
* представляти таблицю обробки даних новинного випуску у вигляді файлу формату Microsoft Excel;

**Методи дослідження.** У даній роботі проаналізовано існуючі підходи до розв'язання поставленої задачі. Методи розв'язання поставленої задачі відносяться до таких сучасних напрямків роботи з інформацією, як computer vision та natural language processing.

**Характеристика розробленої системи.** У даній роботі проаналізовано існуючі підходи та створено автоматизовану систему для виявлення спроектовано систему та реалізовано програмний засіб, що дозволяє виконувати обробку і класифікацію архіву новинних відеорепортажів.

**Результати випробувань системи.** За допомогою створеного ПЗ було оброблено більше 80 годин новинних випусків, для кожної окремої групи результатів проаналізовано їх точність шляхом порівняння з данними, отриманими вручну, встановлено точність отриманих результатів на рівні 85-100% для кожної категорії.

**Практичне застосування.** Розроблене програмне забезпечення впроваджене у роботу для телеканала «Україна». Наступним кроком є підлаштування агоритму під інші новинні канали, такі як: «Україна24», «Інтер», «СТБ», «ICTV», тощо.

**Ключові слова:** медіа сюжети, обробка відеопотоку, новини, natural language processing, computer vision.