

ВИЯВЛЕННЯ ПРОМО-ЕЛЕМЕНТІВ У ВІДЕОРЕКЛАМІ СОЦІАЛЬНИХ МЕРЕЖ НА ОСНОВІ OCR ТА LLM

Вінницький національний технічний університет;

Анотація

Рекламні відео в соціальних мережах часто містять промо-елементи у вигляді накладеного тексту (знижки, ціни, дедлайни акцій, умови пропозиції), які є критичними для аналітики кампаній та коректного маркування контенту. Автоматичне вилучення таких даних ускладнюється короткочасністю появи тексту, різноманіттям шрифтів і мов, а також артефактами стиснення відео. Запропоновано підхід до структурованого маркування промо-атрибутів на основі поєднання OCR для вилучення текстових фрагментів з кадрів [1] та великих мовних моделей (LLM) для семантичної інтерпретації й нормалізації значень у контрольовану схему з інструкційним форматом відповіді [3], [4]. Описано компоненти конвеєра обробки, модель даних промо-сутностей і механізми валідації, що підвищують відтворюваність результатів.

Ключові слова: відеореклама, соціальні мережі, маркування, промо-атрибути, OCR, великі мовні моделі, нормалізація даних.

Abstract

Social media video ads frequently include promotional overlays (discounts, prices, deadlines, and offer conditions) that are essential for campaign analytics and reliable content labeling. Automated extraction is challenged by short on-screen durations, multilingual typography, and compression artifacts. This paper proposes a structured labeling approach that combines OCR for frame-level text extraction [1] with LLM-based semantic interpretation and normalization into a controlled schema via instruction-following inference [3], [4]. The processing pipeline, promo-entity data model, and validation mechanisms are described to improve reproducibility.

Keywords: video advertising, social networks, labeling, promotional attributes, OCR, large language models, data normalization.

Вступ

Відеореклама у соціальних мережах часто містить промо-елементи (наприклад, «-30%», «до 31.03», «2=1», «безкоштовна доставка»), які визначають зміст пропозиції й мають бути відображені у структурованих атрибутах маркування. Підхід, заснований лише на OCR та правилах, є ефективним для простих числових патернів, але недостатній у випадках контекстної неоднозначності (наприклад, відрізнити «-50%» як знижку від «50% більше», інтерпретувати «до понеділка» як часову умову, нормалізувати «від 299 грн» як мінімальну ціну) [1]. Це мотивує застосування LLM як семантичного інтерпретатора поверх OCR-виходу із заданим форматом результату (контрольована схема), що узгоджується з принципами few-shot/in-context підходів [3] та інструкційного вирівнювання моделей із наміром користувача [4]. У контексті сучасних моделей родини GPT-5 також актуальним є представлення можливостей системи у вигляді офіційних описів/карток безпеки для практичного впровадження [5], [6].

Постановка задачі та модель промо-сутностей

Нехай задано відеофрагмент реклами VVV з часовою віссю $t \in [0, T] \forall t \in [0, T] \exists t \in [0, T]$. Потрібно сформувати множину промо-сутностей EEE , де кожна сутність має тип і нормалізовані поля. Запропоновано мінімальне ядро типів: Discount (значення та одиниця — %/сума), Price (значення,

валюта, тип ціни), PromoDeadline (дата/час або відносний дедлайн), Condition (формалізована умова). Для практичного використання кожній сутності додається поле evidence: фрагмент OCR-тексту та таймкод(и) кадрів, що забезпечує аудит і модерацію на основі першоджерела [1].

Метод виявлення промо-елементів (OCR + LLM + валідація)

Конвеєр складається з послідовних етапів. Спочатку виконується вибір кадрів із адаптивним семплуванням, щоб зменшити обсяг обробки без втрати промо-інформації. Далі застосовується детекція текстових областей і OCR: для «scene text» доцільні детектори, що локалізують текстові блоки в кадрі, після чого виконується розпізнавання [2], [1].

Наступний крок — темпоральна агрегація: рядки OCR об'єднуються у «треки» за схожістю та перекриттям областей між сусідніми кадрами, що знижує шум і відновлює частково втрачені фрагменти. Після агрегації застосовується LLM-парсинг: модель отримує промо-текст із таймкодами та повертає структуру у строгой схемі (наприклад, JSON з дозволеними полями/типами). Стабільність формату й семантична інтерпретація підсилюються інструкційними техніками [3], [4].

Фінальний етап — детермінована валідація та нормалізація: перевірка допустимих меж (наприклад, 0–100% для знижки), форматів дат, узгодження валюти з локаллю, очищення OCR-артефактів. Некоректні сутності відхиляються або позначаються як такі, що потребують перевірки. За потреби підхід може розширюватись на мультимодальні сценарії аналізу відео/зображень у системах рівня GPT-5 [5], [6] та на споріднені підходи візуального instruction tuning для задач “vision-language” [7].

Результати досліджень

Основний результат полягає у формалізації промо-атрибутів у контрольованій схемі з доказовими полями (таймкоди + OCR-фрагменти), що підвищує відтворюваність маркування й спрощує аудит [1]. Розподіл ролей зменшує кількість неоднозначностей: OCR забезпечує прив'язку до кадру та текстову основу [1], [2], LLM — контекстну інтерпретацію і приведення до схеми [3], [4], а валідаційний шар гарантує мінімальні інваріанти коректності. Практично це створює сумісний модуль для системи маркування аудіо- та відеорядів реклами в соціальних мережах із можливістю масштабування на великі обсяги креативів.

Висновки

Запропоновано підхід до виявлення та структурованого маркування промо-елементів відеореклами, що поєднує OCR, LLM-інтерпретацію та детерміновану валідацію [1]–[4]. Визначено модель промо-сутностей і описано конвеєр нормалізації знижок, цін, дедлайнів та умов у контрольовану схему з доказовими полями. Підхід узгоджується з практиками застосування сучасних LLM-систем родини GPT-5 [5], [6] і може бути розширений на мультимодальні сценарії [7].

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. An Overview of the Tesseract OCR Engine [Електронний ресурс] - Режим доступу до ресурсу: <https://research.google.com/pubs/archive/33418.pdf>
2. EAST: An Efficient and Accurate Scene Text Detector [Електронний ресурс] - Режим доступу до ресурсу: https://openaccess.thecvf.com/content_cvpr_2017/papers/Zhou_EAST_An_Efficient_CVPR_2017_paper.pdf
3. Language Models are Few-Shot Learners [Електронний ресурс] - Режим доступу до ресурсу: <https://arxiv.org/abs/2005.14165>
4. Training language models to follow instructions with human feedback [Електронний ресурс] - Режим доступу до ресурсу: <https://arxiv.org/abs/2203.02155>

5. GPT-5 System Card [Електронний ресурс] - Режим доступу до ресурсу: <https://cdn.openai.com/gpt-5-system-card.pdf>
6. Introducing GPT-5 [Електронний ресурс] - Режим доступу до ресурсу: <https://openai.com/index/introducing-gpt-5/>
7. Visual Instruction Tuning [Електронний ресурс] - Режим доступу до ресурсу: <https://arxiv.org/abs/2304.08485>

Куліш Сергій Павлович – аспірант групи 121-24а, ФІТКІ, Вінницький національний технічний університет, м. Вінниця, e-mail: kulish@vntu.edu.ua

Ткаченко Олександр Миколайович – к.т.н., доцент кафедри програмного забезпечення, Вінницький національний технічний університет, м. Вінниця, e-mail: alextk1960@gmail.com.

Serhii Pavlovych Kulish – PhD student, group 121-24a, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: kulish@vntu.edu.ua

Tkachenko Oleksandr Mykolayovych – Ph.D., Associate Professor of the Department of Software, Vinnytsia National Technical University, Vinnytsia, e-mail: alextk1960@gmail.co