

РОЗРОБКА МЕТОДУ ГРУПУВАННЯ ОБ'ЄКТІВ НА ОСНОВІ
ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ТЕКСТОВОГО ОПИСУ

Вступ. Сучасні платформи електронної комерції вимушені постійно підвищувати ефективність алгоритмів рекомендацій. Одним із кроків у цьому процесі є групування товарів на основі порівняння їх споживчих характеристик. Дослідження питань схожості товарних пропозицій та їх групування базується на попередніх роботах, зокрема [1]. Як показує аналіз, питанням порівняння товарів, які наявні у веб, присвячено багато досліджень. Так, у роботі [2] пропонується процес класифікації побудований на регулярних виразах. Обробка зображень також використовується для визначення відповідності елементів, наприклад у роботі [3]. Однак, існуючі підходи досить не забезпечують загальної схеми вирішення задачі групування.

Постановка задачі. Завдання порівняння товарів стосується виявлення продуктивних пропозицій, що надходять з різних веб-ресурсів, та які відносяться до одного і того ж реального продукту. Товарні пропозиції публікуються разом з текстовими описами. Синтаксична, структурна і семантична неоднорідність пропозицій робить порівняння товарів складним завданням.

Для **вирішення цієї задачі** пропонується технологія краудсорсінгу для збору оцінок подібності товарів на основі зображень, а також алгоритм побудови ядра тегів на основі обробки текстових даних з опису товарів. Загальну схему запропонованого підходу зображено на рисунку 1.

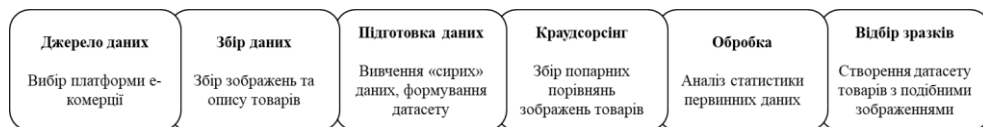


Рис. 1. Схема методу групування товарів

Підхід реалізовано на прикладі товарів з платформи eBay (<https://www.ebay.com/>) для товарної групи «кросівки». Для формування тестового набору даних були зібрані дані, відокремлені зображення та текстові описи товарів. Для реалізації технології краудсорсінгу розроблено мобільний додаток Item Matcher, який дозволяє на основі реалізації парних порівнянь зображень визначити схожість товарів. Отримані статистичні дані використані для визначення групи схожих зі споживчої точки зору товарів. Для сформованої групи товарів було побудовано ядро тегів. Відповідна процедура обробки текстових описів товарів наведена на рис. 2.



Рис. 2. Процедура обробки текстових описів товарів

Висновки. На основі аналізу тестових даних визначено оцінки Accuracy = 0.74, Precision = 0.88, Recall = 0.89, що свідчить про задовільні результати експерименту. Запропонований підхід оцінки подібності товарів показав достатньо високий результат, тому подальші дослідження будуть направлені на визначення обмежень запропонованого підходу до інших груп товарів, а також на порівняння товарів з різних торговельних майданчиків.

Література

- 1 Cherednichenko, O., Vovk, M., Kanishcheva, O., Godlevskiy M.: Studying Items Similarity for Dependable Buying on Electronic Marketplaces. Proc. 2nd Int. Conf. On Computational Linguistics and Intelligent Systems (COLINS), Volume I. Vol. 2136. pp.78-89 CEUR-WS (2018).
- 2 Hoffmann, U., Silva, A., Carvalho, M.: Finding Similar Products in E-commerce Sites Based on Attributes. Published in AMW (2015).
- 3 Yashima, T., Okazaki, N., Inui, K., Yamaguchi, K., Okatani, T.: Learning to Describe E-Commerce Images from Noisy Online Data. pp. 85-100 (2017). https://doi.org/10.1007/978-3-319-54193-8_6.