

РОЗРОБКА РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ ЕЛЕКТРОННОЇ КОМЕРЦІЇ З ПРОДАЖУ КНИЖОК

Синьова В. О.

Науковий керівник – к.т.н., с.н.с. Коваленко А. І.

Харківський національний університет радіоелектроніки, каф. СТ
м. Харків, Україна

e-mail: viktoriiia.synova@nure.ua

The report discusses the stages of designing and developing a recommendation system for e-commerce selling books, implementing a business function that allows recommending users books with the highest average rating among registered users, and books similar to the chosen one, using TF-IDF vectorization, cosine similarity, and Data Mining methods.

У доповіді розглядаються етапи проектування та розробки рекомендаційної системи електронної комерції (е-системи) з продажу книжок із реалізацією бізнес-функції, що дозволяє рекомендувати користувачам книжки з найвищим середнім рейтингом серед зареєстрованих користувачів, і книги, що схожі на обрану, з використанням векторизації TF-IDF (Term Frequency – Inverse Document Frequency) і косинусної подібності. Для розробленої рекомендаційної е-системи визначені такі ролі її користувачів.

«Незареєстрований користувач» має можливість: зареєструватися та авторизуватися в е-системі; переглядати каталоги, використовуючи фільтрацію та сортування; користуватися кошиком для оформлення замовлення; ознайомитися з детальною інформацією про будь-яку книгу, зокрема, її назву, авторів, видавництво, мову написання, рік видавництва, короткий опис змісту, зображення обкладинки, вартість, жанри, теги, середній рейтинг, кількість сторінок, ISBN, відгуки, рекомендації подібних книг з оцінкою схожості.

«Зареєстрований користувач» додатково може користуватися функціями особистого кабінету, оформлювати замовлення, писати відгуки до книг і оцінювати схожість запропонованих рекомендацій до книжки.

«Адміністратор» має доступ до функцій обробки замовлень, редагування системної інформації і формування звітів з продажів книг.

«Зареєстрований користувач» може оформити тільки один відгук на кожну книгу. До цього відгуку входить обов'язкова оцінка від одиниці до п'яти, а також необов'язковий текст відгуку. З використанням усіх оцінок користувачів за книгою розраховується їх середнє арифметичне значення, яке є її рейтингом. За замовчанням книги без відгуків мають рейтинг, що дорівнює нулю. На основі рейтингу книжок їх можна відповідно відсортувати у каталозі за зростанням (спаданням) оцінки або за їх діапазоном.

Під час перегляду детальної інформації про книгу також відображається список до п'яти найбільш подібних до неї інших книжок, відсортованих за оцінкою схожості, яку виставляють зареєстровані користувачі. Тобто користувачі можуть впливати на оцінку схожості, збільшуючи або зменшуючи її значення. За замовчанням оцінка схожості дорівнює нулю, та нижче нього опуститися не може, а верхньої межі немає.

Для реалізації функції е-системи з визначення рейтингу використовується значення статистичної міри «TF-IDF», яка застосовується для тексту книг [1]. Коефіцієнт TF визначає оцінку важливості тегу (ключового слова) і знаходиться як відношення числа його входжень до загального числа слів книги. Коефіцієнт IDF – це інверсія частоти, з якою тег зустрічається у книзі. Статистична міра TF-IDF є добутком двох співмножників: TF-IDF. Велику вагу за TF-IDF отримують теги з високою частотою в межах тексту книги. Кожна книжка представляється у вигляді вектора коефіцієнтів TF, що дозволяє визначити їх схожість за мірою косинусної подібності.

Дані книжок отримуються з датасет-файлу у форматі CSV, після чого їх теги оброблюються та аналізуються завдяки векторизації TF-IDF і косинусній подібності. Окрема книга має свій список тегів. Під час аналізу коефіцієнтів TF кожному присутньому тегу книги присвоюється значення «1», а відсутньому тегу – значення «0». Далі з використанням коефіцієнтів IDF аналізується вага (важливість) тегу відносно всіх тегів інших книг. Чим більше книжок мають певний тег, тим менший вплив він має під час оцінки схожості. На основі отриманих значень TF-IDF розраховується ступінь схожості між усіма парами книг із використанням міри косинусної подібності. За останнім кроком, для книжки визначаються до п'яти найбільш схожих книжок з найбільшим значенням подібності.

Рекомендаційна е-система книжкової крамниці побудована за архітектурою «клієнт–сервер» для глобальної мережі Інтернет. Клієнтська частина містить HTML-сторінки та JavaScript компоненти, які розроблені на платформі React.JS. Для взаємодії користувача з е-системою та забезпеченню отримання даних від СУБД розроблені серверні процедури мовою Java, які обробляють події, що відбуваються під час використання інтерфейсу. База даних е-системи розроблена на платформі СУБД MySQL. Рекомендаційна функція визначення схожості книг та обробки датасету розроблена мовою Python.

Список використаних джерел:

1. Understanding TF-IDF for Machine Learning: вебсайт. URL: <https://www.capitalone.com/tech/machine-learning/understanding-tf-idf/> (дата звернення 29.02.2024).