

УДК 004.738.5:339

ДОСЛІДЖЕННЯ МЕТОДІВ ПОШУКУ ТЕКСТОВОЇ ІНФОРМАЦІЇ ТА ЇХ ВИКОРИСТАННЯ В ІТ-ПРОЄКТАХ ЕЛЕКТРОННОЇ КОМЕРЦІЇ

Семенець К. Є.

Науковий керівник – д.т.н., проф. Петров К. Е.

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

тел.: +38(066)407-02-22, e-mail: kateryna.semenets@nure.ua

In the modern digital world, the success of IT companies often depends on the effectiveness of search on their resources. This paper examines the problem of relevant search in e-commerce IT projects, as well as existing methods to address specific issues related to this topic, their advantages and disadvantages. The study justifies the relevance of the problem, determines the purpose of the research, and outlines ways to solve these problems.

Системи електронної комерції стають все більш популярними в сучасному світі, але однією з найважливіших складових їх успішності є реалізація якісного пошуку товару, що потрібен споживачу. Від того, наскільки якісно та швидко користувачі зможуть знайти потрібний їм товар, залежить частка продажів та задоволеність клієнтів [1].

Реалізація релевантного пошуку по товарних позиціях є складною задачею, оскільки користувачі можуть шукати товари різними способами та використовувати різні формулювання, робити помилки у запитах або здійснювати пошук, використовуючи трансліт.

Тому актуальним завданням є розробка нових, більш досконалих методів, які дозволили б розробити ефективний інструмент пошуку релевантної інформації.

Найпростішим з методів є пошук за фразою, коли товари, назва яких містить дану фразу, виводяться у результатах пошуку. Однак цей метод є не ефективним, оскільки при пошуку можуть виникнути труднощі через різне написання тих самих слів, різні закінчення, порядок слів у фразі або через те, що одне зі слів записане із помилкою.

Для вирішення цих проблем можна використати метрику подібності Левенштейна. Відстань Левенштейна дозволяє знайти відповідності між двома словами, порівнюючи їх на відстані редагування, тобто кількості операцій (вставка, видалення, заміна), необхідних для перетворення одного слова в інше. Використання цієї метрики дозволяє механізму пошуку видавати користувачу товари, що містять інформацію, яка є близькою до його запиту.

Але і в цьому випадку ще залишаються такі проблеми: використання синонімічних, але різних за написанням слів; транслітні записи іншомовних слів; випадкове використання іншої розкладки клавіатури.

Для вирішення проблем, пов'язаних з морфологією, можна використовувати так званий тезаурус – словник, який містить синоніми та близькозначні слова. Використання тезаурусів допомагає системі знаходити інші слова, які можуть бути пов'язані з запитом користувача, і тим самим збільшує кількість можливих релевантних результатів.

Для більш глибокого вирішення контекстної проблеми пошуку також застосовують алгоритми машинного навчання. Одним із напрямків штучного інтелекту, що динамічно розвивається, є розуміння природної мови (Natural Language Understanding, NLU) [2].

Використовуючи підходи NLU пошукова система “розуміє” контекст запиту користувача та здатна зв'язати його з контекстом документів, що зберігаються в базі даних. Але навіть цей підхід може мати низьку якість пошуку через те, що у словниковому наборі не буде знайдено нового для нього слова або система зіткнеться із проблемою наявності різних значень одного й того ж слова.

Отже, деякі з перерахованих методів пошуку є простими, але неефективними, тоді як реалізація інших підходів може бути громіздкою та складною. Проте, комбінування кращих методів може допомогти створити більш досконалий метод для реалізації релевантного пошуку інформації.

Для досягнення цієї мети, було детально проаналізовано та проведене порівняння ефективності існуючих методів, визначені їх переваги й недоліки та запропоновані підходи для розробки вдосконаленого методу пошуку на основі комбінації найбільш ефективних з них.

Також варто зазначити, що розроблений комбінаторний метод може бути застосований не тільки в системах електронної комерції, а й в інших проєктах, в яких важливим є релевантний пошук. Враховуючи те, що споживачі все більше звертають увагу на зручність і швидкість пошуку в ІТ-продуктах, реалізація такого методу може значно поліпшити користувацький досвід та збільшити конверсію проєкту.

Список використаних джерел:

1. Погоріла, Ю. (2023, 10 квітня). Де ви втрачаєте покупців та «зливаєте» продажі: реальні ситуації з інтернет-магазинів. <https://mc.today/blogs/sezon-rasprodazh-vyderzhit-li-poisk-v-vashem-internet-magazine/>.

2. Patel, A. A., & Arasanipalai, A. U. (2021). Applied Natural Language Processing in the Enterprise. " O'Reilly Media, Inc."