

## ДОСЛІДЖЕННЯ ТА РЕАЛІЗАЦІЯ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ НА ОСНОВІ МЕТОДУ КОЛАБОРАТИВНОЇ ФІЛЬТРАЦІЇ

*Запорізька державна інженерна академія*

Велика кількість нині існуючих веб-ресурсів використовують рекомендаційні системи для своїх користувачів. Вони генерують різні пропозиції, наприклад, супутні товари користувачам зі схожими інтересами, зареєстрованим на цьому сайті. Такі рекомендаційні механізми обробляють величезні обсяги інформації для визначення потенційних вподобань користувачів. Рекомендаційні системи - це програми, які намагаються визначити що хочуть знайти користувачі, що може їх зацікавити, і рекомендують їм це. Дані механізми вдосконалили взаємодію між користувачем і сайтом. Замість статичної інформації вони надають динамічно змінювану: рекомендації генеруються окремо для кожного користувача, ґрунтуючись на його попередній активності на даному веб-ресурсі, до того ж може враховуватися інформація, що надходить від інших відвідувачів ресурсу.

У більшості випадків розвиток рекомендаційних систем полягає в поліпшенні алгоритмів рекомендації. Мета цього прогресу - дати відвідувачам сайтів найбільш точні рекомендації, що задовольняють їхні запити. Для досягнення цього математичні алгоритми, що лежать в основі рекомендаційних сервісів, повинні постійно навчатися. Тут вступає в силу машинне навчання та інтелектуальний аналіз даних *Data mining*.

Об'єктом дослідження даної роботи є рекомендаційні системи, предметом дослідження – методи розробки (колаборативна фільтрація та метод найменших квадратів) рекомендаційних систем.

Мета роботи – створення та дослідження тестової рекомендаційної системи використання медіа-ресурсів та дослідження методів розробки рекомендаційних систем.

В роботі наведені вимоги до розроблюваної рекомендаційної системи.

Далі в роботі наведено огляд відомих рекомендаційних систем та технологій розробки рекомендаційних систем.

*Pandora* - одна з найпопулярніших музичних рекомендаційних систем - базує свої рекомендації на даних, отриманих з проекту *Music Genome*. Він приписує до 400 атрибутів кожній пісні. Це робиться спеціальними музичними працівниками. Процес такого призначення може займати до півгодини на пісню. Незважаючи на вражаючі результати роботи *Pandora*, і те, що нам зазвичай рекомендується цікава і нова музика, цей підхід слабо масштабується, і бібліотека *Pandora* часом буває обмежена.

*The Hype Machine* - відрізняється від інших сервісів на корені. Замість того, щоб збирати дані про музику, яка подобається користувачеві, з існуючих плейлистів або ґрунтуючись на проставлених ним рейтингах і потім намагаючись запропонувати нову музику, яка може йому сподобатися, *The Hype Machine* шукає в мережі Інтернет найбільш обговорюваних артистів і виконавців, їх альбоми. Він відстежує не те, що люди слухають, а то, про що вони пишуть в блогах. Тут особливо важливо, що відслідковуються не всі блоги, а тільки кращі з кращих. Команда *Hype Machine* відбирає свої джерела дуже ретельно, і в їх базу дуже важко потрапити блогам музичних стартапів або звукозаписних лейблів [1].

*Колаборативна фільтрація* - метод побудови рекомендаційних систем, заснований на припущенні про те, що користувачам зі схожими оцінками предметів, переглянутих раніше, буде подобатися одне і те ж в майбутньому. Група користувачів, найбільш схожа за інтересами з активним, називається «сусідами».

*Неявний збір даних* - коли відвідувач відмовляється надавати інформацію про себе, актуальною стає наступна методологія - неявний збір даних. Тут присутнє свого роду стеження за людиною, в ході якого його дії фіксуються спеціальною програмою, яка

щороку збирає необхідні відомості для подальшого аналізу і застосування. Сюди включаються дані про здійснені покупки, про проставлені оцінки будь-яких продуктів або послуг, дані про перегляди і коментарі [3].

Проблеми рекомендаційних систем:

1) Методи збору інформації, що надаються Інтернетом, значно спростили використання громадської думки за допомогою колаборативної фільтрації. Але, з іншого боку, великий обсяг інформації ускладнює втілення цієї можливості. Наприклад, поведінка одних людей досить ясно піддається моделюванню, в той час як інші поведуться абсолютно непередбачувано. І саме другі впливають на зміщення результатів рекомендаційної системи і зниження її ефективності. Інший приклад: користувачі можуть використовувати рекомендаційну систему для навмисного зростання популярності одного продукту щодо іншого. Вони можуть залишати не просто позитивні відгуки про вподобаний товар, але і писати негативні про конкурентів.

2) Інша проблема, притаманна великим рекомендаційним системам, це масштабованість. Традиційні алгоритми непогано справляються з невеликою кількістю даних. Але чим більше інформації має оброблятися, тим важче отримати точні результати.

3) Рекомендаційні системи, що працюють на основі неявних користувальницьких оцінок, тягнуть за собою проблему дотримання конфіденційності користувачів, захисту персональних даних.

#### **Висновки :**

- Проведені дослідження методів розробки рекомендаційних систем і принципів їх побудови.
- Проаналізовані засоби створення рекомендаційних систем, такі як: фреймворк Apache Spark, мова Scala; фреймворк Django, мова Python [4-5].
- В ході дослідження була створена тестова рекомендаційна система для користувачів медіа-ресурсів, основана на колаборативній фільтрації, досліджено можливості створеної рекомендаційної системи.
- Також в роботі були розглянуті проблеми рекомендаційних систем та можливі їх вирішення [6-7].

#### **Література**

1. Hype Machine — саундтрек блогосфери [Електронний документ] (<http://www.ferra.ru/ru/techlife/news/2013/02/11/HypeMachine-sitesoftheday/>).
2. Авхадеев Б., Воронова Л., Охупкина Е. Разработка рекомендательной системы на основе данных из профиля социальной сети «ВКонтакте» [Електронний документ] (<http://cyberleninka.ru/article/n/razrabotka-rekomendatelnoy-sistemy-na-osnove-dannyh-iz-profilya-sotsialnoy-seti-vkontakte>).
3. Лагун Д. Анализ неявных предпочтений пользователей [Електронний документ] (<https://habrahabr.ru/company/yandex/blog/223121/>).
4. Парошина И. Как музыкальный сервис Pandora узнаёт и предсказывает предпочтения слушателей [Електронний документ] (<http://www.computerra.ru/87479/pandora-bigdata/>).
5. Издательский дом «Питер» Знакомство с Apache Spark [Електронний документ] (<https://habrahabr.ru/company/piter/blog/276675/>).
6. Гончаров М. Data Mining: Рекомендательные системы [Електронний документ] ([http://kek.ksu.ru/EOS/WM/50\\_132-670.pdf](http://kek.ksu.ru/EOS/WM/50_132-670.pdf)).
7. Сэнди Риза, Ури Лезерсон, Шон Оуэн, Джош Уиллс Spark для профессионалов: современные паттерны обработки больших данных. — СПб.: Питер, 2017.

