

**РАЗРАБОТКА РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ ПУБЛИКАЦИИ
В НАУЧНОЕ ИЗДАНИЕ**

Чепрасова Е. В. (Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»),
Радченко И. А. (Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

Ввиду большого количества научных изданий возникает проблема выбора наиболее подходящих изданий для публикации научных статей. Существующие рекомендательные системы обладают ограниченной функциональностью. В рамках данной работы рассматриваются подходы к построению улучшенной системы рекомендации в научное издание.

Введение.

В настоящее время существует большое количество научных изданий. Как следствие, ручной выбор журнала для публикации является затруднительным, и для ученых остро встает проблема выбора наиболее подходящего из них.

Существует ограниченное количество систем, рекомендуемых научное издание. Большинство из них анализируют только название статьи и ее аннотацию, что может приводить к неточности результатов.

В работе рассматриваются часто используемые подходы к построению рекомендательных систем – коллаборативная фильтрация и фильтрация на основе содержимого. Системы коллаборативной фильтрации строятся на основе исторических данных о пользователях (user-based) или контенте (item-based).

Основная часть.

Было принято решение разрабатывать систему с использованием открытого набора данных, предоставляемого системой arXiv — известным сервисом публикации препринтов. Набор содержит данные о почти 1.8 млн научных трудов таких областей как математика, физика, экономика, электротехника; представлена информация об авторах статьи, название журнала, аннотация статьи, область знаний. Непосредственно текст статьи может быть получен посредством API сервиса arXiv. Для получения сведений о новых статьях используется отдельный компонент системы.

Ядром разрабатываемой системы является сервис рекомендаций. В его основе лежит идея о возможности выявления общих признаков для статей, опубликованных в одном журнале. Таким образом первым этапом является анализ содержимого статьи и подбор наиболее похожих по смыслу статей. При сравнении документов может использоваться алгоритм TF-IDF, идея которого заключается в выявлении ключевых слов в документах путем анализа частотности. Для выявления похожих статей может также применяться гипотеза о возможности определения близких документов путем анализа литературных источников, на которые ссылаются статьи.

Еще одним подходом является анализ данных об авторах статей для построения ассоциаций между научными изданиями на основе истории их публикаций или взаимосвязей между учеными (рассматривается гипотеза о близкой тематике работ ученых-соавторов). Стоит отметить, что данный подход будет работать корректно в случае наличия уникального идентификатора для каждого ученого.

Для взаимодействия пользователя с рекомендательной системой спроектирован отдельный компонент.

Выводы.

В работе представлены различные подходы к построению рекомендательных систем, предложены гипотезы для построения ядра системы — сервиса рекомендаций, спроектирована архитектура системы.

Чепрасова Е. В. (автор)