

# РАЗРАБОТКА РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ С ВИЗУАЛИЗАЦИЕЙ ДАННЫХ

**А. Г. Каркоцкий**

*Белорусский государственный университет, пр. Независимости, 4,  
220030, г. Минск, Беларусь, karkotskiy.alexander@yandex.by  
Научный руководитель — Е. В. Пазюра, старший преподаватель*

В статье рассматриваются методы фильтрации данных в рекомендательных системах, способы визуализации данных и разрабатывается рекомендательная система с визуализацией данных которая позволяет обрабатывать произвольный набор неструктурированных данных, предоставляет пользователю рекомендации, строит визуализацию полученных рекомендаций и учитывает предпочтения пользователя при построении рекомендаций и визуализаций.

**Ключевые слова:** рекомендательные системы; фильтрация; визуализация; язык программирования Python.

Для обработки данных в рекомендательных системах используются различные фильтрации. Основными подходами фильтрации являются такие подходы, как фильтрация на основе содержимого, совместная фильтрация и гибридная фильтрация [3].

Подход **фильтрации на основе содержимого** предоставляет рекомендации, находя сходства среди определенных атрибутов вариантов рекомендаций.

Подход **совместной фильтрации** предоставляет рекомендации, основываясь на взаимодействии различных пользователей с элементами [1, с. 215].

Подход **гибридной фильтрации** использует несколько моделей рекомендаций одновременно, для достижения большей точности [3].

В качестве подхода фильтрации в разработанной рекомендательной системе с визуализацией данных, используется подход фильтрации на основе содержимого.

С точки зрения реализации, этот подход можно описать следующим образом. Набор данных состоит из элементов, которые могут быть рекомендованы. Каждый из элементов описывается набором характеристик. Все характеристики элемента разбиваются на отдельные слова и ищется количество их вхождения по-отдельности в каждый элемент. Таким образом проводится операции векторизации, где каждому элементу рекомендации ставится в соответствие многомерный вектор, количество элементов которого соответствует количеству уникальных слов по всем элемен-

там рекомендаций, а значения элементов векторов равны количеству повторений уникальных слов, входящих в соответствующий элемент рекомендаций.

Затем происходит вычисление похожести между элементами рекомендаций, для чего используется косинусное сходство [4, с. 129]:

$$\text{cosine}(x, y) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}$$

где  $x$  и  $y$  – это элементы двух многомерных векторов и размерности, между которыми ищется сходство.

Таким образом получается матрица сходства, где на пересечении строк и столбцов находится число в диапазоне от 0 до 1, которое характеризует похоть соответствующих элементов. Чем больше число, тем более элементы рекомендаций похожи друг на друга.

После этого на основе предпочтений пользователя, выбирается элемент, который считается наиболее подходящим и на основе матрицы сходства получается набор элементов рекомендаций.

Визуализация является важнейшим этапом в рамках представления большого объема данных в удобном пользователю формате. В рамках разработанной рекомендательной системы используются такие типы визуализации, как гистограмма, столбчатая диаграмма, круговая-секторная диаграмма и диаграмма рассеивания (Рис. 1).

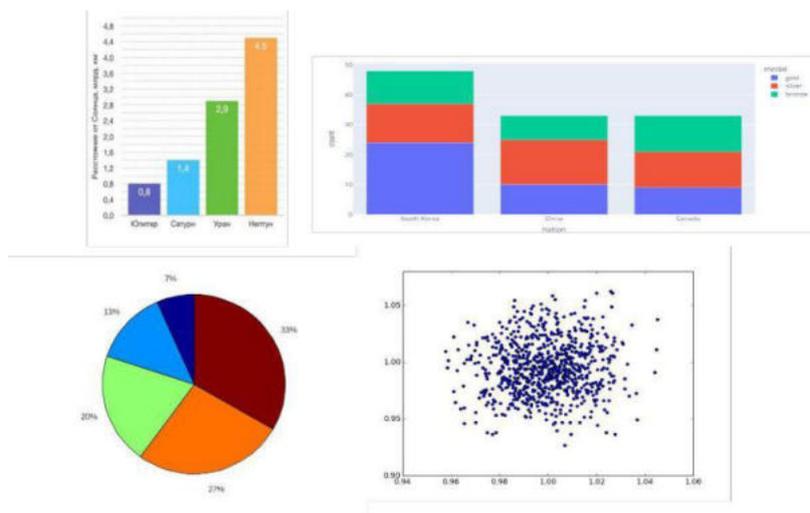


Рис. 1. Примеры типов визуализации

Разработанная рекомендательная система состоит из шести подсистем (Рис. 2). Этими подсистемами являются подсистемы авторизации пользователя, загрузки данных, предоставления рекомендаций, визуализации, сохранения и загрузки параметров и данных системы, и перехода состояний системы.

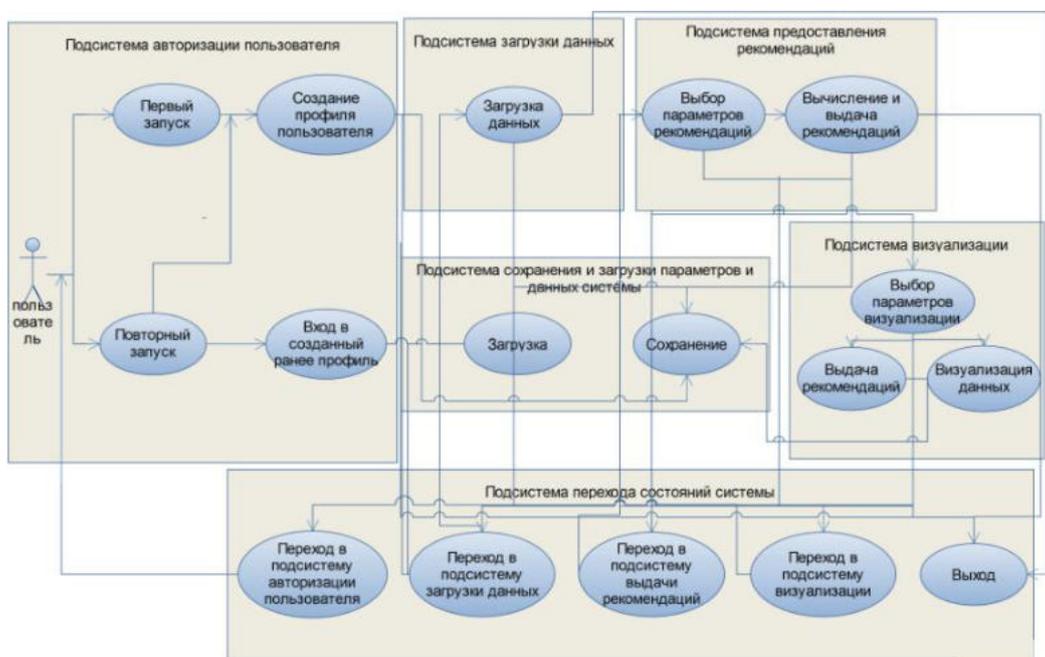


Рис. 2. Рекомендательная система с визуализацией данных

Разработка осуществлялась с использованием языка программирования Python и подключением таких библиотек, как Pandas для обработки данных, Scikit-learn для построения рекомендательной модели, Plotly для визуализации данных, и Tkinter для создания визуального интерфейса.

Подсистема авторизации пользователя позволяет осуществить регистрацию или авторизацию пользователя в разработанной системе, где в качестве параметров входа или регистрации выступают логин и пароль.

Подсистема загрузки данных позволяет осуществить загрузку данных. На входе в систему поступают данные в формате Excel. В системе данные хранятся в виде Pandas DataFrame библиотеки Pandas.

Подсистема предоставления рекомендаций позволяет осуществить выбор параметров рекомендаций, включающих количество рекомендаций

и приоритет атрибутов элементов рекомендаций, пользователем и получить персонализированные рекомендации на основе предпочтений.

Подсистема визуализации позволяет пользователю осуществить настройку параметров визуализации и, при необходимости, обратиться к

истории визуализаций. Помимо этого, если к данным применялась фильтрация, то незначимые элементы будут отнесены в группу Others.

Подсистема сохранения и загрузки параметров и данных системы позволяет осуществить сохранение и загрузку всех параметров или данных, которые образуются в результате работы системы в автоматическом режиме. Это позволяет пользователю вернуться к работе с системой в том положении, в котором она была завершена ранее.

Работа подсистемы перехода состояний системы заключается в завершении работы текущей запущенной подсистемы и запуска той, к которой осуществляется переход, где под состоянием системы понимается текущая запущенная подсистема.

Разработанная рекомендательная система с визуализацией данных реализует общий подход к построению рекомендаций с использованием фильтрации на основе содержимого и может применяться, без необходимости дополнительной настройки, к примеру, в таких сферах, как:

1. **Электронные торговые площадки.** Анализ оценок, отзывов, характеристик товаров, продаж с целью установить наиболее выгодные к закупкам и дальнейшей продаже изделий.

2. **Стриминговые платформы.** Анализ жанра, количество просмотров, оценок, отзывов фильмов или сериалов, для установки наиболее выгодного расширения репертуара.

3. **Финансовые услуги.** Анализ положения на рынке, с целью прогнозирования инвестиций.

Программный код разработанной рекомендательной системы с визуализацией данных представлен в открытом доступе в системе контроля версий GitHub [2].

Разработанная рекомендательная система с визуализацией данных, обладает следующими особенностями:

1. Использован подход фильтрации на основе содержимого.
2. Осуществляется обработка данных, подготовленных пользователем.
3. Осуществляется работа с произвольным набором неструктурированных данных.
4. Предоставляется возможность явной настройки получаемых пользователем рекомендаций.
5. Осуществляется визуализация построенных рекомендаций с возможностью настройки пользователем.

## Библиографические ссылки

1. Фальк К. Рекомендательные системы на практике / К. Фальк. – Москва: ДМК Пресс, 2020 – 448 с.
2. GraduateWork [Электронный ресурс]. URL: <https://github.com/Aleanderk/GraduateWork> (дата обращения 06.05.2024).
3. Recommender Systems: Behind the Scenes of Machine Learning [Электронный ресурс]. URL: <https://www.altexsoft.com/blog/recommender-system-personal-ization/> (дата обращения 01.05.2024).
4. Schrage M. Recommendation Engines (The MIT Press Essential knowledge series) / М. Schrage. Кембридж: MIT, 2020. 306 с.