

# ИНФОРМАЦИОННАЯ СИСТЕМА АНАЛИЗА СОЦИАЛЬНЫХ СЕТЕЙ

Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь

Белов А.В.

Лукашевич М.М. – к.т.н., доцент

Статья посвящена разработке информационной системы анализа социальных сетей. Перечислены основные вопросы и проблемы, приводится обоснование использования для разработки определенного перечня библиотек. Рассмотрены результаты работы информационной системы, а также перспективы использования.

Целью настоящего исследования является реализация информационной системы анализа данных социальных сетей, пригодную для изучения публично доступной информации о студентах специальности ВМСиС. Объектом исследования являются социальные сети, предметом – алгоритмы сбора, анализа и визуализации данных социальных сетей.

Рассматриваются существующие социальные сети, их особенности, доступные инструменты разработчика для работы с данными каждой, а также основные проблемы в использовании. Описываются основные подходы к анализу данных социальных сетей, а также производится обзор существующих аналогов и выбор на основе их технологических стеков языков программирования, основных библиотек и алгоритмов для использования.

С целью использования в процессе анализа данных рассматриваются основные теоретические аспекты рекомендательных систем. Рассмотрены вопросы пригодности рекомендательных систем для анализа данных социальных сетей, имеющиеся методы их построения и основные алгоритмы, пригодные для реализации. Описан процесс выбора конкретного типа рекомендательных систем и алгоритмов.

Рассматриваются вопросы проектирования и программной реализации информационной системы анализа данных социальных сетей. Также подробно описаны алгоритмы получения и анализа данных, построения рекомендательных систем, а также алгоритмы визуализации данных и типы используемых интерактивных графиков. Описываются основные сущности модели данных, а также классы, используемые в программной реализации и их взаимосвязи.

Описан процесс тестирования реализованной информационной системы, описываются характеристики ее работы и выявляются точки, пригодные для дальнейшей оптимизации. Оценивается точность работы рекомендательной системы с данными пользователей Github и наглядность используемых средств для визуализации данных [1].

Для реализации алгоритмов анализа данных наиболее подходящим признан язык программирования Python, а также библиотеки алгоритмов машинного обучения scikit-learn, быстрых численных вычислений numru, научных расчетов scipy, а также библиотека для построения рекомендательных систем implicit.

Реализация сбора данных реализована с применением библиотек Octokit для работы с API сервиса Github, Koala для использования небольшого числа доступных для использования методов работы с данными социальной сети Facebook, Linkedin-scraper, позволяющий произвести импорт данных пользователя из публично доступного профиля в профессиональной социальной сети LinkedIn, а также библиотеки Mechanize, служащей для получения данных, недоступных с помощью иных способов через эмуляцию действий пользователя в процессе работы с веб-интерфейсами социальных сетей.

Часть данных, необходимых, в частности, для построения рекомендательной системы, получена с использованием облачного сервиса Google BigQuery и SQL запросов особого вида [2].

Для реализации пользовательского интерфейса используется веб-фреймворк Ruby on Rails для языка программирования Ruby, а также Javascript библиотека построения интерактивных графиков d3.js.

В процессе работы над данной диссертацией была спроектирована и разработана информационная система анализа данных социальных сетей. В качестве социальных сетей, пригодных для сбора релевантных данных, были использованы социальные сети Facebook и LinkedIn, а также сервис совместной работы Github. Для анализа репозитория программных проектов системы Github была построена рекомендательная система коллаборативной фильтрации по методу соседства с использованием алгоритма чередующихся наименьших квадратов [3].

В качестве объектов для тестирования получившейся системы были выбраны студенты специальности ВМСиС и их профили в названных социальных сетях и сервисах. Среди информации, представляющей большой интерес, следует отметить данные о распределении студентов и мест работы по странам мира, декларируемые в профессиональной социальной сети LinkedIn навыки, а также стек используемых в работе над проектами с открытым исходным кодом технологий, полученный с помощью Github. Для пользователей, имеющих аккаунты в системе Github и достаточное количество отмеченных программных проектов, были сформированы рекомендации по наиболее интересным для дальнейшего изучения программным проектам и направлениям.

Список использованных источников:

1. Building recommender system for Github [Электронный ресурс]. – Режим доступа: [https://medium.com/@andrey\\_lisin/building-recommender-system-for-github-a8108f0cb2bd](https://medium.com/@andrey_lisin/building-recommender-system-for-github-a8108f0cb2bd).
2. Recommending Github Repositories with Google BigQuery and the implicit library [Электронный ресурс]. – Режим доступа: <https://towardsdatascience.com/recommending-github-repositories-with-google-bigquery-and-the-implicit-library-e6c6e666c77>
3. Чередующиеся наименьшие квадраты [Электронный ресурс]. – Режим доступа: <http://rsc.chemometrics.ru/Tutorials/mcr.htm#Ch4.2>.