

Автоматизация сбора данных для оценки выраженности личностных особенностей пользователей социальных сетей

Ляпин Н.Е., студент СПбГУ, st075646@student.spbu.ru
Корепанова А.А., м.н.с. СПб ФИЦ РАН, aak@dscs.pro

Аннотация

В данной статье описывается реализация инструмента для автоматизации сбора данных пользователей социальной сети ВКонтакте для последующего поиска взаимосвязей между личностными особенностями пользователей и публикуемым ими контентом, с целью применения найденных фактов в сфере информационной безопасности и других смежных областях.

Введение

Повседневность человека в 2022 году невозможно представить без взаимодействия с самыми разными информационными системами, переносящими в цифровую среду множество различных видов деятельности [1]. Общение также подвержено цифровизации, и это особенно заметно в последние годы [2]. Важной площадкой для данного процесса служат социальные сети, количество новых пользователей которых с каждым годом лишь увеличивается [3]. Вместе с ростом вовлечённости в них, растёт и объём данных, которые пользователи оставляют о себе в сети.

Личностные особенности и психологические черты миллионов людей отражены на их страницах в виде записей, фотографий, лайков и т.д. Это огромный массив информации, который может быть потенциально использован для различных исследований в областях, изучающих человеческое поведение или аспекты, связанные с ним [4-5].

Данный проект посвящён разработке веб-приложения для упрощения и ускорения сбора данных для оценки личностных особенностей пользователей социальных сетей. Найденные с его помощью зависимости могут быть применены во множестве областей, вроде информационной безопасности, психологии и др.

Концепция приложения

Для удобства сбора данных было решено реализовать мини-приложение с психологическими и иными тестами [6]. Оно легко интегрируется в социальную сеть ВКонтакте и позволяет получать информацию, представляющую собой результаты тестирований по любым специально разработанным методикам (в том числе, для исследования личностных особенностей респондентов).

В концепции приложения были реализованы следующие подходы:

- Каждый тест конструируется из нескольких блоков, взаимодействующих друг с другом. Если подходящих блоков среди ранее реализованных нет – их можно реализовать и внедрить. Благодаря данному подходу получается достичь большей гибкости за счёт того, что в приложение возможно добавить тестирования любой конфигурации и сложности.
- Документно-ориентированная база данных. Использование такой БД позволяет легко внедрять новые функциональности на уровне, связывающем серверную часть и БД, без необходимости постоянного исправления SQL-скриптов [7].

Использованные технологии

1. Интеграция:
 - 1.1. VK Mini Apps — платформа социальной сети ВКонтакте для мини-приложений;
2. Клиентская часть:
 - 2.1. React — JavaScript-библиотека для простой разработки пользовательских интерфейсов;
 - 2.2. VKUI, VK Bridge, VK Icons — библиотеки команды ВКонтакте, предназначенный для удобного создания React-приложений на платформе VK Mini Apps в общей для социальной сети стилистике;
3. Серверная часть:
 - 3.1. Node.js — программная платформа, выполняющая роль веб-сервера, предоставляющая возможность программирования на языке JavaScript;
 - 3.2. Express — фреймворк-обёртка для веб-приложений на

Node.js, позволяющий более простым образом использовать серверные функции;

3.3. Mongoose — библиотека для взаимодействия веб-приложения на Express с документно-ориентированной базой данных MongoDB;

4. База данных:

4.1. MongoDB — документно-ориентированная система управления базами данных, не требующая описания схемы таблиц, использующая JSON-подобные документы и схему базы данных;

4.2. MongoDB Compass — графический интерфейс для администрирования и просмотра данных MongoDB;

Архитектура приложения

Клиентская часть

Структура клиентской части веб-приложения была спроектирована через предварительную декомпозицию проекта на атомарные функциональные компоненты. Компоненты были объединены в следующие логические группы:

- Основные компоненты, которые описывают главные панели, содержащие все функциональности клиентской части.
- Панель с тестами, которая содержит компоненты с внешним оформлением тестов и функциональности, реализующие их логику.
- Панель с профилем пользователя, которая содержит компоненты с оформлением профиля и функциональности, отвечающие за показ результатов тестирований.
- Панель с настройками, которая содержит компоненты с внешним оформлением настроек и функциональности, отвечающие за их изменение внутри приложения.

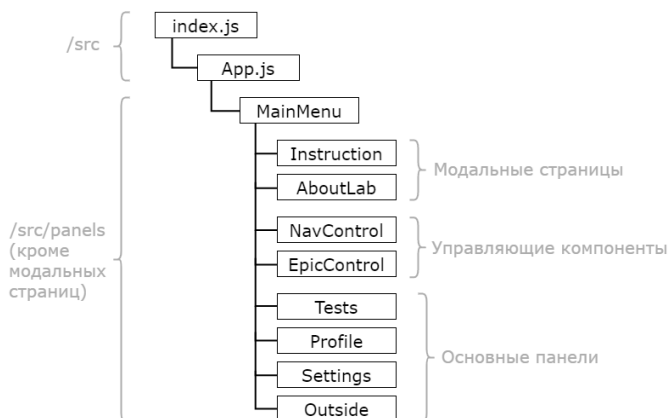


Рисунок 1: Иерархия основных компонентов

Серверная часть

Архитектура серверной части представлена в виде API, позволяющего взаимодействовать на клиенте с базой данных. Функции API были объединены в следующие логические группы:

- tests – функции, отвечающие за взаимодействие с тестами (например, получение конкретного теста для его прохождения).
- users – функции, отвечающие за взаимодействие с данными пользователя (например, инициализация стандартных пользовательских настроек при первом входе в приложение).
- passings – функции, отвечающие за прохождения тестов (например, сохранения ответов пользователя).
- results – функции, отвечающие за формирование и сохранение результатов тестирований (например, сохранение результатов пользователя по завершению прохождения).

База данных

Для хранения и управления данными в приложении используется документно-ориентированная NoSQL СУБД MongoDB. Для приложения были спроектированы следующие три коллекции, содержащиеся в БД:

- Tests – содержит информацию обо всех имеющихся тестах, каждый документ в ней представляет собой описание

конкретного теста.

- Users – содержит информацию обо всех пользователях, каждый документ представляет собой описание конкретного пользователя.
- Passings – содержит информацию о прохождении тестов, каждый документ представляет собой описание конкретного прохождения.

Заключение

В статье было представлено описание веб-приложения, являющегося полноценным, масштабируемым инструментом для автоматизации сбора данных пользователей социальной сети ВКонтакте с целью последующей оценки связи между публикуемым пользователями контентом и их личностными особенностями. Факты, установленные с помощью этих данных, можно будет применить во множестве областей. Например, для более полного изучения и моделирования распространения социоинженерных атак и иных аспектов информационной безопасности, связанных с человеческим фактором, а также в социологии, психологии, маркетинговых исследованиях и т.д. [8–11].

Литература

1. Forbes: «Record 3.8 Trillion Hours Spent On Mobile Apps During 2021 In Another Blockbuster Year For Digital Economy». <https://www.forbes.com/sites/roberthart/2022/01/12/record-38-trillion-hours-spent-on-mobile-apps-during-2021-in-another-blockbuster-year-for-digital-economy/?sh=76ca663b2a42> [дата просмотра: 18.04.2022].
2. Nguyen, Minh Hao; Gruber, Jonathan; Fuchs, Jaelle (2020). Changes in digital communication during the COVID-19 global pandemic: implications for digital inequality and future research. Social Media and Society.
3. We Are Social: «DIGITAL 2022: ANOTHER YEAR OF BUMPER GROWTH». <https://wearesocial.com/uk/blog/2022/01/digital-2022-another-year-of-bumper-growth-2/> [дата просмотра: 18.04.2022].
4. Open Access Government «Social media data for social and behavioural research». <https://www.openaccessgovernment.org/social-media-data/113292/> [дата просмотра: 18.04.2022].
5. Cerruto, F., Cirillo, S., Desiato, D. et al. Social network data analysis to highlight privacy threats in sharing data. J Big Data 9, 19 (2022).
6. Мини-приложение ВКонтакте «Психологические тесты». <https://vk.com/app7794698> [дата просмотра: 19.04.2022].
7. Эрик Редмонд, Джим. Р. Уилсон. Семь баз данных за семь недель.

Введение в современные базы данных и идеологию NoSQL = MongoDB in Action. — ДМК Пресс, 2013. — 384 с.

8. Тулупьева Т.В., Абрамов М.В., Тулупьев А.Л. Социоинженерные атаки: социальные сети и оценки защищенности пользователей // Монография.: Санкт-Петербург, 2018.
9. Корепанова А.А., Олисеенко В.Д., Абрамов М.В. Применимость коэффициентов сходства в задаче сравнения социального окружения // Международная конференция по мягким вычислениям и измерениям. 2020. Т. 1. С. 39-42.
10. Корепанова А.А., Тулупьева Т.В. Идентификация аккаунтов пользователя в различных социальных сетях по социальному окружению // В сборнике: ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ РЕГИОНОВ РОССИИ (ИБРР-2019). материалы конференции. 2019. С. 442-443.
11. Арамов М.В., Азаров А.А., Фильченков А.А. Распространение социоинженерной атаки злоумышленника на пользователей информационной системы, представленных в виде графа социальных связей // Международная конференция по мягким вычислениям и измерениям. 2015. Т. 1. С. 329-331.