***УДК 004.652***

**ИЗВЛЕЧЕНИЕ И СТРУКТУРИРОВАНИЕ ДАННЫХ ПОЛЬЗОВАТЕЛЕЙ ИЗ СОЦИАЛЬНОЙ СЕТИ ВКОНТАКТЕ**

**А.А. Михайличенко, Д.А. Вербицкий**

Кубанский государственный университет

ул. Ставропольская 149, 350040, Краснодар, Россия

**Ключевые слова:** СУБД, SQLite, социальные сети, Python, ВКонтакте, структурирование данных, извлечение данных, vk.com, API

**Аннотация**

Статья посвящена разработке программы на языке Python для извлечения и структурирования информации о пользователях из социальной сети ВКонтакте. Программа гибко и безопасно извлекает данные, учитывая ограничения API, и сохраняет их в СУБД SQLite.

**Введение.** Одним из крупнейших источников информации о пользователях являются социальные сети. Они предоставляют данные о подписках, географическом распределении, фотографиях, интересах и поведении пользователей. Автоматический сбор такой информации существенно улучшает ее точность и количество сведений для будущего анализа. Компании активно используют такие данные для повышения качества своих продуктов и разработки новых информационных систем.

**Извлечение данных пользователей с помощью API.** Данные пользователей – это информация, которая собирается социальными сетями о своих зарегистрированных пользователях. Они отражают характеристики, интересы и прочую активность человека. В дальнейшем эта информация может быть проанализирована и использована в маркетинговых либо в исследовательских целях. Данные пользователей используются для прогнозирования активности пользователей, запуска таргетированной рекламы, предсказания трендов и других целях.

Для взаимодействия с данными социальной сети vk.com используется ее API. Он предоставляет разработчикам обращаться к различным методам и функциям для интеграции своих приложений с социальной сетью. Для запросов к API vk.com используются HTTP-запросы, а ответы приходят в формате JSON, что обеспечивает их удобную обработку и интеграцию с системой. Разработчику для получения доступа к данным пользователей vk.com сначала потребуется получить токен доступа (access token), выдаваемый OAuth 2.0.

В целях обеспечения безопасности пользователей было придумано разделение прав доступа. Для обеспечения конфиденциальности большинство чувствительных данных могут быть получены только, если пользователь явно согласится их предоставить приложению. Таким образом обеспечивается прозрачность и законное использование этих данных. У каждого приложения, которое обращается к API, есть лимит на частоту отправки запросов, превышение которого может привести к блокировке доступа.

Для извлечения данных пользователей используются методы API users.get, users.getSubscriptions и friends.get. Затем происходят валидация и фильтрация этих данных. Эта обработка полученной информации является важным этапом и подготавливает данные для последующего анализа.

Для структурированного хранения данных используется реляционная база данных. Для взаимодействия с ней применяются ORM-модели, реализованные с помозью библиотеки SQLAlchecmy. Это позволяет описать базу данных на уровне классов Python, что существенно упрощает работу с данными.

Модель ER диаграммы показана на рисунке 1.

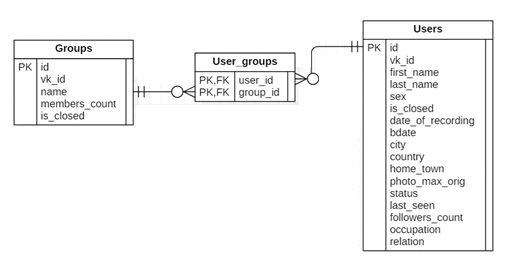


Рисунок 1 – ER – диаграмма

Основные данные пользователей извлекаются и помещаются в соответствующие модели. Также извлекается информация о подписках пользователя. Таблица, в которой хранятся данные пользователей, показана на рисунке 2.

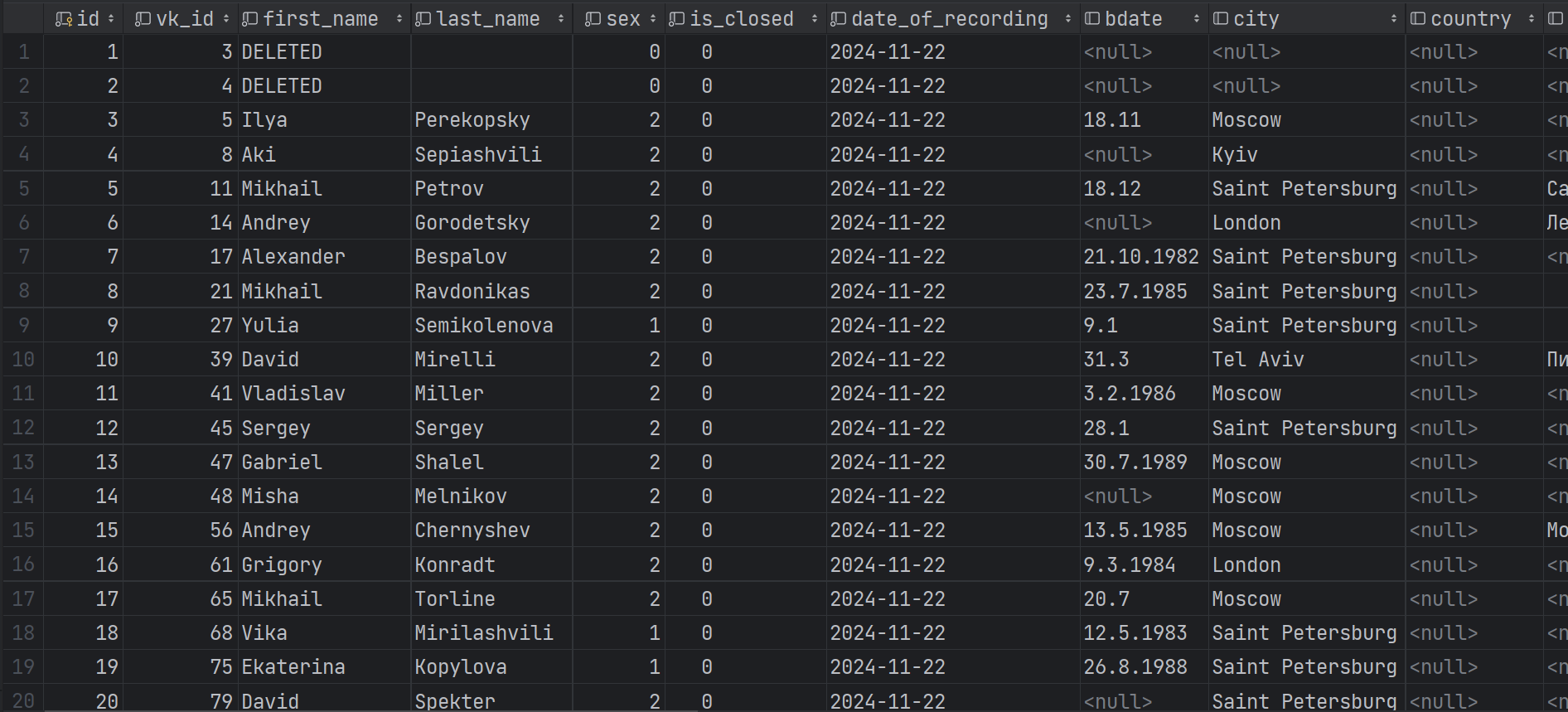


Рисунок 2 – Таблица users

**Заключение.** Была разработана программа для извлечения и структурирования информации о пользователях и сообществах из социальной сети ВКонтакте.

***Библиографический список***

1. VK API Documentation // Официальный сайт разработчиков VK: [сайт]. – 2024. – URL: https://dev.vk.com/ (дата обращения: 26.11.2024).
2. SQLAlchemy Documentation // Документация библиотеки SQLAlchemy: [сайт]. – 2024. – URL: https://docs.sqlalchemy.org/ (дата обращения: 26.11.2024).