

СПОСОБЫ СБОРА ДАННЫХ ПОЛЬЗОВАТЕЛЕЙ СОЦСЕТЕЙ

Рассматриваются основные подходы и способы сбора данных пользователей социальных сетей, а также их особенности.

ВВЕДЕНИЕ

Понятие социальной сети использовалось социологами еще в 20-х годах прошлого века для изучения взаимосвязей между участниками различных сообществ. Одним из вариантов стало использование социограмм, на которых отдельные индивиды представлялись в виде точек, а взаимосвязи между ними — в виде линий. Идею использования аппарата теории графов для изучения взаимоотношений и взаимосвязей между людьми подхватили специалисты в области социологии, психологии, антропологии, политологии, экономики.

I. АКТУАЛЬНОСТЬ ПРОБЛЕМЫ СБОРА ДАННЫХ

Современные социальные сети существенно изменили постановку вопроса — сегодня у исследователей имеются ресурсы для изысканий, а стремительное распространение социальных онлайн-сервисов и развитие технологий Больших Данных инициировали интерес к использованию сведений из социальных сетей в различных отраслях. Совместное использование структурных и контентных данных потенциально позволяет применять социальные сети для решения широкого круга бизнес-задач: борьбы с мошенничеством, управления брендом, рекламы товаров и услуг, формирования новых каналов сбыта и др.

II. СРАВНЕНИЕ СПОСОБОВ СБОРА ДАННЫХ ПОЛЬЗОВАТЕЛЕЙ СОЦСЕТЕЙ

Самый простой способ сбора данных — воспользоваться услугами специализированных компаний, собирающих и постоянно обновляющих данные из множества источников. Главное преимущество здесь — быстрота получения информации, что существенно при больших объемах клиентской базы и использовании различных социальных сетей. Недостаток — платная подписка на обновления данных.

Следующий способ — использовать программные интерфейсы, предоставляемые почти всеми популярными социальными сетями. Для различных сетей API отличаются набором до-

ступных данных, ограничениями на количество запросов и стоимостью доступа к интерфейсам. Например, если с помощью программного интерфейса сети «В контакте» можно получить полную информацию о пользователе, то Facebook предоставляет API, возвращающий практически «нулевые» сведения о пользователе. К недостаткам этого метода относится ограничение на количество одновременных запросов и на количество обращений, которые приложение может делать в единицу времени. Кроме того, необходимо постоянно отслеживать изменения в API и обновлять приложение по сбору данных, причем некоторые социальные сети предоставляют важные данные только на платной основе. Преимуществами метода являются возможность получения данных об одном пользователе в структурированной форме (JSON или XML), а также простота интеграции вызовов API в собственное приложение.

Еще один способ — ручной разбор веб-страниц социальных сетей, а также использование готовых краулеров для сбора данных с последующим разбором. В этом случае имеется доступ ко всем открытым данным и отсутствуют ограничения на скорость их сбора. К недостаткам следует отнести сложность реализации — веб-страница каждой социальной сети уникальна, поэтому каждый раз придется разрабатывать свои правила разбора. Недостатками являются также сложность поддержки и необходимость больших вычислительных ресурсов, правда, этот процесс хорошо распараллеливается.

III. Выводы

Каждый способ сбора данных пользователей соцсетей имеет свои плюсы и минусы. Таким образом, наиболее быстрым по времени реализации будет метод, использующий открытый API соцсети, однако данный метод может иметь ограничения по количеству обращений и типу сведений о пользователе.

1. Горчинская, О. Ю. Анализ данных социальных сетей / О. Ю. Горчинская // Открытые системы. СУБД. — 2015. — №3.

Стельмачёнок Максим Олегович, магистрант кафедры интеллектуальных информационных технологий БГУИР, stelmachenokm@gmail.com.

Научный руководитель: Бойко Игорь Михайлович, в. н. с. УП ГИС НАНБ, кандидат технических наук, igor_m_boyko@hotmail.com.