

Получение, оценка и анализ данных из социальных сетей средствами приложений

М.Н.Дубовенко¹

Аннотация – This article shows the possibility of obtaining, processing and analysis of data from applications in social networks.

Ключевые слова – Социальные сети, Обработка, Получение, Информация, Анализ данных.

I. ВВЕДЕНИЕ

Социальные сети относительно недавно внедрились в человеческую жизнь и уже стали неотъемлемой частью современного общества. На данный момент крупные социальные сети хранят информацию о сотнях миллионов людей, что открывает новые возможности для работы с ней. Но есть ряд сложностей, связанных с получением и оценкой данных, которые являются предметом обсуждения в данной публикации.

II. ПОЛУЧЕНИЕ, ОЦЕНКА И АНАЛИЗ ДАННЫХ ИЗ СОЦИАЛЬНЫХ СЕТЕЙ СРЕДСТВАМИ ПРИЛОЖЕНИЙ

Единственный легальный способ получения информации из социальных сетей на автоматизированном уровне – *социальные приложения* (СП). Для разработки СП существуют документации по API, а также правила внедрения и использования информации. В документациях приводятся методы, параметры и способы передачи данных.

Ввиду тенденции развития Интернет-технологий, эти способы реализовываются на клиентской стороне технологиями JavaScript API и др., для уменьшения нагрузки на серверную часть. Эта задача особенно актуальна для СП, так как количество обращений к приложению может быть до сотен тысяч в сутки. Не остается в стороне вопрос безопасности, и каждая транзакция имеет hash, построенный по алгоритму md5, который сверяется на стороне социальной сети и СП.

Социальные сети раскрывают горизонты для проведения масштабных мониторингов, при условии неразглашения личной информации пользователей. Для задач такого рода становятся особо актуальными вопросы о качестве получаемой информации. Основными классами данных, для которых необходима оценка качества и достоверности являются *личные данные* и *данные мониторинга*. Оба класса связаны с объективностью, искренностью пользователей [1] и другими человеческими факторами. Данная проблема разрешима на уровне СП, где параметры пользователя

можно получать из сети и из СП, вычисляя коэффициент валидности (сходимости данных), который учитывается при анализе результатов. Оценка индивидуальна и разрабатывается для каждой задачи отдельно в виде математической модели. Второй класс данных оценивается исходя из поставленной задачи, и может использовать предыдущий способ, а также другие, например, коэффициент социального доверия.

Как и любую другую информационную систему, СП можно построить в виде пользовательской и управляющей подсистем. Функции пользовательской подсистемы закладываются в сборе информации, управленческой – в управлении пользователями и анализом данных. Так как СП предназначены для работы с большими массивами данных, их логично основывать на базе данных (БД). Проектирование БД для СП не имеет особых отличий от каких-либо других систем, и реализуется стандартными дизайнерскими пакетами.

СП, построенные на БД, удобны для Data (Web, Social) Mining [2]. У СП, априори, достаточно возможностей для получения данных именно в количественной форме, что позволяет применять традиционные алгоритмы кластеризации, регрессии, классификации и др. и зачастую не требует разработки специализированных средств.

III. ВЫВОД

Использование приложений в социальных сетях является эффективным способом получения данных, который исключает большие финансовые и временные затраты и дает возможность извлекать валидную информацию от заданной аудитории, основываясь на оценке входящих параметров.

СПИСОК ССЫЛОК

- [1] О.Н. Поляруш, А.В. Витько, "Применение байесовских вероятностных сетей для анализа социальных сетей" *Системы обработки информации*, Харьков: ХУПС, №6(87), с.208-211, 2010.
- [2] С. Царьков, "Web Mining: основные понятия" BaseGroup Labs, технологии анализа данных, http://www.basegroup.ru/library/web_mining/basic_conceptions/

¹ Международное научно-учебное учреждение информационных технологий и систем НАН и МОН Украины, пр. академика Глушкова, 40, Киев, 03680, УКРАИНА, E-mail: maksymkpm@rambler.ru