УДК 004

АВТОМАТИЗАЦИЯ АНАЛИЗА СОДЕРЖИМОГО ГРУПП И КАНАЛОВ МЕССЕНДЖЕРА TELEGRAM

Е. М. БОРИСОВ, К. В. БОГДАШЕВА
Научный руководитель Д. А. ДЕНИСЕВИЧ
Белорусско-Российский университет
Могилев, Беларусь

Ручной мониторинг содержимого каналов мессенджера Telegram становится неэффективным вследствие лавинообразного роста информационного шума, риска безвозвратной потери данных при редактировании или удалении публикаций авторами, а также неприемлемо высоких трудозатрат на постоянный контроль десятков источников. Автоматизация процессов сбора (парсинга) призвана решить задачи формирования «вечного архива», исключающего утрату информации, и построения структурированных баз знаний, в которых хаотичный поток сообщений преобразуется в чистые массивы данных, пригодные для углубленной аналитики.

Современные программные системы анализируют не только текст, но и все медиа-вложения, а также динамические показатели – реакции, количество репостов и полные ветки комментариев, что позволяет количественно оценить реальную вовлечённость аудитории. Технологический стек таких решений опирается на языки программирования C#, Python, Node.js и библиотеки парсинга. Для надёжного хранения и структуризации собираемых архивов применяются реляционные системы управления базами данных, в частности PostgreSQL и Microsoft SQL Server.

При разработке инструментов неизбежно столкновение с защитными механизмами платформы, в первую очередь с Flood Control. Для предотвращения блокировок реализуются адаптивные задержки и инкрементальная стратегия сбора: система фиксирует идентификатор последнего обработанного сообщения и докачивает только новые данные, что снижает сетевую нагрузку в десятки раз. Собранные архивы служат фундаментом для применения методов машинного обучения: автоматического анализа тональности комментариев и построения графов информационных связей, позволяющих установить первоисточник новостного повода.

Таким образом, парсинг Telegram сегодня представляет собой не единичные скрипты, а полноценную промышленную аналитику. Корректно спроектированная база данных даёт возможность создать интерфейс, с помощью которого аналитик без навыков программирования может в один клик экспортировать отчёты в форматах Excel или JSON. Подобные системы экономят сотни часов рутинного труда и превращают мессенджер из развлекательной платформы в надёжный структурированный источник бизнес-знаний.