

АВТОМАТИЗАЦИЯ ПОЛУЧЕНИЯ И АНАЛИЗА СТАТИСТИЧЕСКИХ ДАННЫХ РЕКЛАМНЫХ ПЛАТФОРМ

Пузикова А.С., Мизеев В.Д., студенты группы гр.753504

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Ермакова Е.В. – канд. экон. наук., доцент

Аннотация. В данной работе рассматривается механизм загрузки, обработки и хранения данных об эффективности интернет-рекламы на различных порталах. Выделяются и описываются структурные особенности процесса, лежащего в основе создаваемой для автоматизации данного процесса ETL-системы. Дается обоснование использования ETL-процессов для автоматизации данного механизма. Продемонстрированы другие модули системы (управляющий веб-портал), связанные с данными процессами.

Ключевые слова. Big data, ETL-процесс, реклама, маркетинг, статистика, аналитика, веб-портал.

Работа с большим объемом данных – одна из ключевых особенностей современного маркетинга. Стек технологий «Big data» применяется для сбора и анализа этих данных со следующими целями:

Проведение идентификационных операций (концепция «Know your customer»).

Бизнес-аналитика брендов: сбор информации, необходимой для улучшения существующих продуктов или увеличения дохода; проверка узнаваемости бренда.

Привлечение новых покупателей на основании результатов клиентской аналитики.

Масштабность таких данных не предполагает их ручную обработку: к примеру, объемы показов рекламы на всех типах устройств могут достигать триллионов. В таких ситуациях применяются ETL-процессы, представляющие собой автоматизированную работу по очистке и консолидации разрозненных наборов данных. ETL расшифровывается как «Извлечение, преобразование, загрузка», что позволяет выделить следующие этапы данного процесса:

Извлечение данных из источника, к примеру, API рекламного портала.

Преобразование данных и их абстрагирование для соответствия применяемым для анализа бизнес-моделям.

Загрузка преобразованных данных в хранилище.

В рамках данной работы мы представляем приложение, позволяющее автоматизировать сбор и анализ рекламной информации с различных порталов, таких как Amazon (в т.ч. Amazon DSP, Amazon Vendor Central), Bing, Apple, Facebook (в т.ч. Instagram, Facebook Messenger), Adwords (Google) и т.д. Данная система позволяет пользователю собрать воедино (с помощью веб-портала) статистические данные всех рекламных процессов, которые проводятся им на вышеупомянутых платформах.

Рассмотрим работу приложения в соответствии с выделенными нами этапами классического ETL-процесса:

Извлечение данных

Данный этап предполагает получение данных от рекламных порталов. В зависимости от того, к какому portalу мы обращаемся, извлечение выполняется одним из двух способов: через REST запросы к API портала или использование Selenium WebDriver в том случае, если применение первого метода невозможно (например, отсутствует клиентское API, как в случае с платформой Amazon VCD). Приложение получает от платформы данные о рекламе продуктов на каждом из используемых мастер-аккаунтов за определённый срок (чаще всего – последние 30 или 60 дней) в виде потока данных или csv-таблиц. Данные таблицы могут содержать такие поля как «дата»,

«рекламодатель», «стоимость», «количество показов», «количество кликов», «количество заказов» и т.д.

Преобразование данных

На втором этапе программа фильтрует данные и приводит их к виду, который в дальнейшем будет записан в хранилище. Данный шаг один из самых сложных: каждая платформа предполагает работу над отдельным набором данных с большим числом допустимых форматов. В рамках текущего этапа потоки данных или файлы проходят через синтаксический анализ (например, с помощью парсера CSV), при необходимости агрегируются (к примеру, некоторые платформы предоставляют ежедневные данные, которые нужно суммировать в ежедневный отчёт) или преобразуются (например, для дальнейшего применения некоторых плоских аналитических таблиц необходимо соединить несколько наборов данных и извлечь из полученного массива фиксированный набор полей). Процесс преобразования уровня данных Creative на уровне кода приведён на рисунке 1.

```
public List<CreativeReportRow> GetReportCreatives(string reportCsvText)
{
    using (TextReader sr = new StringReader(reportCsvText))
    {
        var csvHelper = new CsvReader(sr);
        csvHelper.Configuration.SkipEmptyRecords = true;
        csvHelper.Configuration.RegisterClassMap<CreativeReportEntityRowMap>();
        var creatives = csvHelper.GetRecords<CreativeReportRow>().ToList();
        return creatives;
    }
}
```

Рисунок 1 – Преобразование уровня данных Creative для платформы Amazon DSP из CSV-файла

Загрузка данных

Преобразованные данные сохраняются в базе данных Azure SQL Database. Пользователь имеет доступ как к таблицам данной базы, что позволяет ему проверить полноту данных с помощью набора запросов, так и к управляющему веб-порталу, функционал которого предполагает возможность добавления новых аккаунтов для каждой из анализируемых платформ и проверку того, готовы ли на текущий момент данные по той или иной платформе. Кроме того, данный портал используется для просмотра логов каждой из «работ» (запуск выгрузки данных для конкретной платформы) в целях контроля их выполнения и исправления возможных ошибок. Пример работы таблиц портала приведён на рисунке 2.

splid	id	Search Account Name (X = disabled)	Channel	DaySums	ConvSums	CallSums	VideoSums	By Campaign
25	284		Google	8/13/2019 - 12/20/2019	8/26/2019 - 12/20/2019	-	-	view
25	279		Google	9/17/2019 - 3/13/2020	9/17/2019 - 3/11/2020	-	-	view
25	234		Google	11/2/2018 - 2/21/2021	11/5/2018 - 3/11/2020	-	10/1/2020 - 2/21/2021	view
25	251		Google	3/7/2019 - 3/13/2020	3/7/2019 - 3/8/2020	-	-	view
25	277		Google	6/20/2019 - 12/20/2019	6/20/2019 - 12/20/2019	-	-	view
25	280		Google	9/24/2019 - 12/12/2019	9/24/2019 - 12/12/2019	-	-	view
25	282		Google	8/8/2019 - 2/21/2021	8/8/2019 - 9/29/2020	-	3/27/2020 - 2/21/2021	view

Рисунок 2 – Пример работы таблиц управляющего портала, необходимых для проверки загрузки данных

Данная система запускается с помощью планировщика задач Windows: в качестве действия (Action) назначается запуск основной программы с параметрами, соответствующими названию необходимой платформы и данными, которые необходимо получить. Так, например, команда `Extractor.exe SynchFacebookStats -d=31` запустит загрузку данных для Facebook-аккаунтов за последний месяц.

Созданная нами система позволяет автоматизировать процесс получения, обработки и сохранения данных из различных рекламных источников. Она легко масштабируется, позволяет добавлять новые платформы и аккаунты с минимальными время- и ресурсозатратами и использовать гибкие настройки (например, указать время выгрузки данных для каждой из платформ, соответствующее обновлению этих данных на рекламном портале). Это помогает оптимизировать действия команды, использующей результаты работы данной программы для разработки дальнейшей маркетинговой стратегии компании.

Список использованных источников:

57-я научная конференция аспирантов, магистрантов и студентов БГУИР, 2021 г

1. Что такое ETL: как справиться с анализом big data [Электронный ресурс.] – Режим доступа: <https://mcs.mail.ru/blog/что-такое-etl-ili-kak-spravitsya-s-analizom-big-data>. – Дата доступа: 23.02.2021.
2. Kimball, R. The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data / R. Kimball, J. Caserta – New Jersey: Wiley, 2011. – 825 p.
3. Беляцкая, Т. Н. Автоматизированные системы управления электронным рынком рекламы / Т. Н. Беляцкая, О. М. Маклакова, А. С. Рыбакова // Экономика. Управление. Инновации. – 2017. – № 3. – С. 63–67.
4. Беляцкая, Т. Н. Диффузия цифровых технологий / Т. Н. Беляцкая // Цифровая трансформация экономики и промышленности: проблемы и перспективы / А. А. Алетдинова [и др.] ; под ред. А. В. Бабкина. – СПб., 2017. – С. 158–178.
5. Беляцкая, Т. Н. Маркетинг: анализ данных : учеб.-метод. пособие / Т. Н. Беляцкая, О. М. Маклакова. – Минск : Белорус. гос. ун-т информатики и радиоэлектроники, 2018. – 128 с.