

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.75

Юдин
Василий Васильевич

Платформа управления данными для проведения целевых рекламных
кампаний

АВТОРЕФЕРАТ

на соискание академической степени
магистра информатики и вычислительной техники

по специальности 1-40 81 04 – Обработка больших объёмов информации

Научный руководитель
Н. А. Волорова
кандидат технических наук, доцент

Минск 2019

КРАТКОЕ ВВЕДЕНИЕ

Реклама играет важную роль как для бизнеса, так и для потребителя. Реклама формирует имидж брэнда и товара, стимулирует спрос. До нашего времени, когда интернет-реклама, как и сама интернет-сеть, не была настолько распространена, реклама была, по большому счету, неперсонифицирована: в основном, это были бумажные носители, телевидение, объявления. Данные виды рекламы позволяли только условно определить целевую аудиторию, что имело отрицательный эффект как для потребителя – множество навязчивой, но, при этом, не интересующей и, как следствие, раздражающей рекламы, – так и для бизнеса – соотношение количества действительно заинтересованных потребителей к количеству показов рекламы было сравнительно мало. Интернет-реклама – самый прогрессивный и эффективный вид рекламы в наше время. Правда, с появлением интернет-рекламы ситуация не сильно изменилась: реклама стала ещё более агрессивной и навязчивой, но не релевантной. Изначально данная проблема была условно решена для бизнеса: стоимость показа рекламного баннера либо в десятки раз меньше стоимости клика по баннеру, либо вовсе отсутствует. Но данный подход, так или иначе, обязывал бизнес платить за показы рекламы людям, которым она абсолютно не интересна, а также не решал проблему нерелевантной рекламы для потребителя. Со временем, интернет-компании, имеющие огромную аудиторию, такие как Google, Facebook, Apple, стали предоставлять наиболее релевантную рекламу на основе данных о пользователях для построения сегментов: использование истории поисковых запросов для определения сферы интересов и предпочтений, использование данных профиля, таких как пол, возраст, часто посещаемые страницы и сообщества. Подавая собственную рекламу через такие платформы, бизнес может достаточно точно установить критерии для определения сегмента, однако, запрос, конкретно определяющий целевую аудиторию, например «только пользователи, приобретавшие тёмно-синие джинсовые куртки брэнда X», до сих пор невозможен. Платформа управления данными для проведения целевых рекламных кампаний призвана решить описанную выше проблему, тем самым, не только предоставив наиболее релевантную рекламу пользователям интернет-сети, а также минимизировав расходы на рекламу для бизнеса, показывая ее значительно меньшему сегменту, также увеличивая конверсию, так как люди, увидевшие рекламу, будут с высокой долей вероятности заинтересованы в показанной именно им рекламе.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью диссертационной работы является разработка алгоритмов и программного обеспечения для решения задач мониторинга действий и поведения пользователей, составления их портретов, сегментации данных и последующего проведения целевых рекламных кампаний с целью минимизации расходов на их проведение и показа наиболее релевантной рекламы пользователям.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Разработать концепцию системы управления данными для проведения целевых рекламных кампаний.
2. Провести сравнительный анализ возможных технологий для разработки каждого компонента системы.
3. Разработать архитектуру системы.
4. Разработать алгоритмы хранения данных на каждом этапе работы программного комплекса.
5. Реализовать программное обеспечение, реализующее поставленную на каждом этапе задачу.
6. Проанализировать GDPR (General Data Protection Regulation) с целью принятия мер по соответствию продукта регламенту.

Объектом исследования являются системы для проведения целевых рекламных кампаний.

Предметом исследования является программное обеспечение для мониторинга действий пользователей в целях последующего составления их портретов и сегментации.

Основной *гипотезой*, положенной в основу диссертационной работы, является возможность использования вычислительных мощностей облачных сервисов для сбора и обработки больших объемов информации с целью вычисления портретов пользователей интернет-ресурсов для определения их предпочтений в целях проведения целевых рекламных кампаний, что позволит как максимально снизить стоимость таких кампаний, так и сделает рекламу менее навязчивой для пользователя благодаря показу только интересующей его рекламы.

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя Н.А. Волоровой заключается в формулировке целей и задач исследования.

Апробация результатов диссертации

Основные положения диссертационной работы докладывались и обсуждались на IV Международной научно-практической конференции «Big Data и анализ высокого уровня» (Минск, 2018).

Опубликованность результатов диссертации

По теме диссертации опубликована 1 печатная работа в сборнике трудов и материалов международных конференций.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, восьми глав, заключения, списка использованных источников и приложений. В первой главе представлен анализ предметной области, описана общая концепция и архитектура программного продукта. Вторая глава посвящена вопросам импорта и сбора данных, необходимых для составления портрета пользователя. В третьей главе подробно рассмотрены вопросы обработки несвязанных данных пользователей с целью получения полного портрета каждого пользователя в отдельности. Четвертая глава повествует о вопросах сегментации полученных профилей. В пятой главе описан процесс экспорта данных на удаленные платформы с целью последующего проведения рекламных кампаний. Шестая глава полностью посвящена вопросам GDPR (General Data Protection Regulation), анализу регламента и принятым мерам для соответствия законодательству. В седьмой главе описаны возможности интерфейса приложения, а также процесс создания сегмента для последующего проведения рекламной кампании. Восьмая глава посвящена вопросам автоматизации тестирования программного продукта.

Общий объем работы составляет 60 страниц, из которых основного текста – 38 страниц, 21 рисунок на 16 страницах и список использованных источников из 10 наименований на 1 странице.

ОСНОВНОЕ СОДЕРЖАНИЕ

Во **введении** показана актуальность темы диссертационной работы, описана предметная область и практическая значимость разработки программного продукта.

В **первой главе** подробно описана архитектура и концепция разработанного программного комплекса. Большое внимание в данной главе уделено сбору информации и пользователям, как основе, на которой базируется работа платформы.

Исходными данными являются анонимизированные данные о пользователях с привязкой к уникальному идентификатору, а также набор сегментов и их правил. Следующими шагами являются:

- Унификация данных.
- Импорт данных.
- Генерация массива профилей пользователей.
- Сегментация.
- Экспорт идентификаторов на рекламную платформу.

Для работы Платформы используется AWS (Amazon Web Services). Унификация данных представляет собой приведение данных, поступающих из разных источников, к единому формату (в нашем случае, к Apache ORC), а также разделение данных по категориям (события, статические данные, внешние сегменты и пары идентификаторов). Затем все данные импортируются в базу данных (Apache HBase – нереляционная распределенная база данных на EMR-кластере AWS).

Во **второй главе** объяснен процесс импорта данных о поведении пользователей из внешних источников с целью последующего создания профилей пользователей для их сегментации.

Третья глава посвящена процессу генерации массива профиля пользователей на основе импортированных данных. Для упрощения процесса сегментации данных (определения принадлежности профиля к определенным сегментам), в процессе обработки выполняется объединение всех данных в профили. В процессе объединения идентификатор, избранный основной вершиной графа профиля, идентифицирует весь профиль. С данным идентификатором связываются все остальные идентификаторы пользователя, а также все данные, принадлежащие идентификаторам.

Четвертая глава повествует о процессе сегментации профилей. Под сегментацией подразумевается обработка данных пользователей для определения принадлежности предустановленным сегментам. Список сегментов

предустанавливается в интерфейсе, доступном пользователям DMP. При настройке сегмента пользователю доступен следующий набор инструментов:

- Фильтр постоянных значений (констант).
- Фильтр динамических значений (событий).
- Фильтр внешних сегментов (маркеров).
- Фильтр функций агрегации.
- Объединение фильтров в группы с применением логического оператора.

Основная задача этапа сегментации – ответить на вопрос принадлежности каждого профиля каждому сегменту либо положительно, либо отрицательно, без вычисления дополнительных метрик.

В пятой главе описан процесс экспорта данных на удаленную рекламную площадку или систему управления данными. Экспортированные данные могут быть использованы в рамках проведения определенных рекламных кампаний.

Шестая глава полностью посвящена GDPR – регламенту о защите персональных данных пользователей. В данной главе проведен анализ требований регламента, а также описаны меры, принятые для полного соответствия программного продукта законодательству.

В седьмой главе продемонстрирован пользовательский интерфейс системы управления данными и наглядно описан процесс создания сегмента.

Восьмая глава повествует об автоматизации тестирования и развертывания программного продукта.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

В рамках научной работы над диссертацией был разработан программный комплекс для проведения целевых рекламных кампаний с целью снижения их стоимости и повышения производительности.

Рекомендации по практическому использованию результатов

Разработанный программный комплекс имеет возможность гибкой интеграции с любой системой сбора информации о пользователях. Продукт также имеет широкую область применения, так как может быть использован для проведения любой рекламной кампании в сети. Таким образом, использова-

ние платформы при проведении рекламных кампаний положительно влияет на их производительность и понижает их стоимость.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

[1] Юдин, В.В. Целевая интернет-реклама / В.В. Юдин, Н.А. Волорова // Сборник докладов IV Международной научно-практической конференции «Big Data и анализ высокого уровня» – Минск, 2018