

## НЕКОТОРЫЕ АНАЛИТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЕЙ САЙТА



**С.Е. Кушмар**

*Разработчик в компании IBA Group, магистр технических наук, Республика Беларусь*



**И.И. Пилецкий**

*Научный руководитель совместной лаборатории БГУИР-IBA и АЦКТ ИВМ, главный архитектор отделения по программному обеспечению IBA IT Park, кандидат физико-математических наук*

*Белорусский государственный университет информатики и радиоэлектроники, [skushmar@gmail.com](mailto:skushmar@gmail.com), [ianmenski@gmail.com](mailto:ianmenski@gmail.com)*

This paper describes how Big Data technologies can be used to analyze web user behavior. At present such techniques as click heat map and session replay are used in studying web site usability and understanding user behavior. However analysis is made manually and therefore not all available data can be processed. The goal is to create real-time processing applications for web analytics that allow making decision instantly.

Активно развивающиеся технологии Big Data позволяют анализировать новые источники данных, которые совсем недавно были недоступны ввиду их значительного объема, большой скорости поступления и невозможности традиционной структуризации. Применение новых источников данных позволяет перейти на качественно новый уровень решения задач, связанных с применением аналитических методов анализа данных.

В настоящий момент все чаще используются методы, которые позволяют анализировать поведение пользователей сайта. К таким методам относятся тепловые карты сайта (heat map), а также воспроизведение сеанса пользователя (session replay) [1]. Тепловые карты сайта позволяют визуализировать активность взаимодействия пользователей с различными элементами веб-страницы. К тепловым картам сайта относятся: карта кликов, карта ссылок, карта скроллинга.

Карта кликов отображает статистику по кликам на сайте. Клики на карте подсвечиваются разными цветами в зависимости от их частоты. Такая карта помогает оценить удобство использования сайта, определить, какие элементы дизайна кажутся посетителям сайта ссылкой, но таковой не являются.

Тепловая карта ссылок – это инструмент для измерения статистики переходов по ссылкам на сайте. Ссылки в карте подсвечиваются разными цветами в зависи-

мости от их популярности: теплые и яркие цвета соответствуют популярным ссылкам, а тусклые и холодные — непопулярным.

Карта скроллинга предназначена для анализа того, как распределяется внимание пользователей сайта на определенных областях страницы. С помощью такой карты можно определить, сколько времени пользователи просматривают различные элементы страницы. Использование такой карты позволяет подобрать оптимальную длину страницы и правильно разместить важную информацию.

В качестве примера метода анализа поведения с помощью воспроизведения сеанса можно привести технологию Вебвизор, интегрированную в сервис "Яндекс.Метрика". Вебвизор записывает действия посетителей сайта и позволяет затем просматривать действия посетителя в режиме видео. Данная технология позволяет увидеть точное повторение всех действий посетителя: движения и клики мыши, прокрутка страницы, нажатия на клавиши и заполнение форм, выделение и копирование текста [2]. Анализ данных, полученных указанными выше методами, позволяет узнать, как ведут себя посетители сайта, оценить удобства использования веб-ресурса, найти ошибки в структуре и дизайне сайта. Однако в настоящее время на использование таких методов анализа накладывается ряд ограничений. Во-первых, не все возможные данные доступны для анализа. Так, например, Вебвизор записывает не более 1000 посещений в день, а данные хранятся лишь в течение 14 дней [3]. Схожий по возможностям сервис Mouseflow предоставляет до 100 тысяч записей в месяц, а время хранения записей ограничено тремя месяцами [4]. Во-вторых, формат предоставляемых данных нацелен на проведение анализа человеком: для тепловых карт это будут изображения, а записи действий пользователей будут представлены в формате видео. Использование таких форматов значительно затрудняет проведение анализа в автоматическом режиме. В-третьих, отсутствует возможность анализа данных в реальном времени, что, соответственно, не позволяет принимать решения в текущий момент времени. Для устранения указанных ограничений необходимо применение технологий Big Data.

Данные можно считать "большими" не только с точки зрения объема, но и с точки зрения разнообразия, скорости передачи и сложности [5]. Данные, полученные путем записи действий пользователей сайта, обладают всеми указанными аспектами. Такие данные характеризуются как большим объемом, так и высокой скоростью поступления, так как записываются все действия всех пользователей веб-ресурса в каждый момент времени. Данные являются разнообразными ввиду различия записываемых действий. Также полученные данные отличаются сложностью, так как такие данные описывают связь между разнообразными действиями пользователя и многочисленными элементами сайта.

В настоящее время все чаще создаются приложения, в которых применяется потоковая аналитика (streaming analytics). Такие приложения позволяют анализи-

ровать потоки данных большой скорости и в режиме реального времени принимать решения на основе этих данных. Среди платформ для разработки приложений потоковой аналитики можно выделить IBM InfoSphere Streams, Informatica Platform for streaming analytics, SAP Event Stream Processor, IBO: Arama and Terracotta, Tibco StreamBase, Apache Storm [6]. Применение указанных выше технологий для анализа поведения пользователей позволит использовать все возможные данные (поступающие и накопленные данные, данные из других источников), а также принимать решения в режиме реального времени. Рассмотрим возможные варианты применения анализа поведения пользователя и принятия решений в режиме реального времени. Исходя из анализа данных за какой-либо период времени выясняется, что один и тот же пользователь интернет-магазина вновь и вновь возвращается к странице с определенным товаром, изучает отзывы на данный товар, но не покупает его. Для такого пользователя в режиме реального времени может быть сделана персональная скидка на этот конкретный товар. Другой вариант применения – динамическое создание сайта, персонализированного для конкретного пользователя, используя накопленные данные о взаимодействии с сайтом. Еще один вариант применения – интерактивная помощь пользователю в случае возникновения каких-либо ошибок либо затруднения пользования сайтом.

Таким образом, применение технологий Big Data при анализе поведений пользователей сайта позволяет повысить качество решения уже имеющихся задач, а также позволит перейти на качественно новый уровень решения задач, ранее недоступный с использованием традиционных информационных технологий.

#### *Литература:*

1. Baesens, B. Analytics in a Big Data World: The Essential Guide to Data Science and its Applications / B. Baesens. — Hoboken : Wiley, 2014. — 256 с.
2. Яндекс.Метрика внедрила Вебвизор // Компания Яндекс [Электронный ресурс]. — 2015. Режим доступа : [https://company.yandex.ru/press\\_releases/2011/0421/index.xml](https://company.yandex.ru/press_releases/2011/0421/index.xml). — Дата доступа : 11.05.2015.
3. Вебвизор // Яндекс.Помощь.Метрика [Электронный ресурс]. — 2015. Режим доступа : <https://help.yandex.ru/metrika/qanda/webvisor-qanda.xml#number-of-records>. — Дата доступа : 11.05.2015.
4. Mouseflow – Pricing // Mouse Tracking, Website Heatmaps – Mouseflow [Электронный ресурс]. — 2015. Режим доступа : <http://mouseflow.com/pricing/>. — Дата доступа : 11.05.2015.
5. Фрэнкс, Б. Укрошение больших данных. Как извлекать знания из массивов информации с помощью глубокой аналитики / Б. Фрэнкс; пер. с англ. Андрея Баранова. — М. : Манн, Иванов и Фербер, 2014. — 352 с.
6. The Forrester Wave: Big Data Streaming Analytics Platforms // Forrester Research [Электронный ресурс]. — 2015. Режим доступа : <https://www.forrester.com/The+Forrester+Wave+Big+Data+Streaming+Analytics+Platforms+Q3+2014/fulltext/-/E-RES113442?objectid=RES113442>. — Дата доступа : 11.05.2015.