

УДК 004.62

И.А. Губанов¹

¹Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация

ПОДХОД К ИМИТАЦИИ ДЕЙСТВИЙ ПОЛЬЗОВАТЕЛЯ ДЛЯ СБОРА ДАННЫХ СИСТЕМЫ АНАЛИТИКИ

Аннотация. В статье рассматривается решение проблемы потери полезных данных систем веб-аналитики электронной коммерции, базирующееся на имитации пользовательских действий. Имитация действий заключается в применении headless-браузеров в соответствии с разработанными автором алгоритмами.

Ключевые слова: информационная система, веб-аналитика, имитация, электронная коммерция.

I.A. Gubanov¹

¹Irkutsk State Transport University, Irkutsk, Russian Federation

AN APPROACH TO SIMULATION OF USER ACTIONS FOR DATA COLLECTION OF ANALYTICS SYSTEM

Abstract. The article discusses the solution to the problem of the loss of useful data from e-commerce web analytics systems, based on imitation of user actions. The simulation of actions involves the use of headless browsers in accordance with the algorithms developed by the author.

Keywords: information system, analytics, simulation, e-commerce.

Введение

В наши дни наблюдается тенденция роста количества представителей в сфере электронной коммерции (ЭК), которая представляет собой коммерческую активность, целью которой считается приобретение выгоды, и которая базируется на комплексной автоматизации платного цикла за счет внедрения компьютерных сетей [1]. Это подтверждается статистикой: ЭК на сегодняшний день является одним из наиболее перспективных секторов экономики, ее оборот в России по данным [2] в 2019 году составил 1,66 трлн руб.

А, как известно, с ростом представителей в определенной сфере растет и конкуренция. Поэтому цель каждого из них - предлагая наиболее интересные и выгодные предложения, заполучить как можно больше клиентов, тем самым увеличив свой доход. Однако, для привлечения потенциальных клиентов необходимо знать их потребности. Среди инструментов для изучения потребностей и поведения клиентов наиболее мощными являются инструменты веб-аналитики.

Веб-аналитика предоставляет возможность анализировать рекламные кампании, оценивать затраты на определенный канал рекламы для того, чтобы расходовать рекламный бюджет наиболее эффективно, анализировать структуру и контент сайта, получать исчерпывающую информацию о посетителях и многое другое [3]. При этом веб-аналитика сайта позволяет понять:

- количество просмотренных веб-страниц;
- ключевые слова и фразы, по которым посетители находят сайт в поисковых системах;
- географию посетителей;
- время, проведенное на веб-странице посетителем;
- переходы между веб-страницами;
- аудиторию сайта (случайные, постоянные посетители и т. д.);
- удобство навигации сайта для посетителей и т. д.

Например, география посетителей сайта может показать информацию о местоположении

клиентов, иными словами, люди из каких стран и областей являются вашими действующими клиентами. Ориентируясь на географические группы данных можно понять, клиентам из какой страны стоит уделить большее внимание для извлечения наибольшей прибыли. Опираясь на полученные сведения, опытному маркетологу не составит труда оценить эффективность веб-сайта на предмет выполнения поставленных задач, а также разработать план по его улучшению, тем самым увеличив эффективность бизнеса. Опираясь на полученные данные можно определять слабые стороны бизнес-процессов и искать пути их устранения.

Имитация пользовательских действий с применением headless-браузеров

Схема работы систем веб-аналитики представлена на рисунке 1. Как можно заметить, данные в систему аналитики могут поступать из различных источников. Браузер пользователя передает в систему аналитики информацию о действиях, выполняемых им на сайте. Сайт в свою очередь, также может передавать данные как напрямую в систему аналитики, так и в интегрированную CRM-систему. CRM, используя полученные от сайта данные, может передавать их сразу в систему аналитики или же после их предварительной обработки. Несмотря на множество источников, из которых данные могут поступать в систему аналитики, наиболее исчерпывающими являются те данные, которые передаются в систему непосредственно браузером, поскольку они наиболее релевантны и не были подвержены модификации.

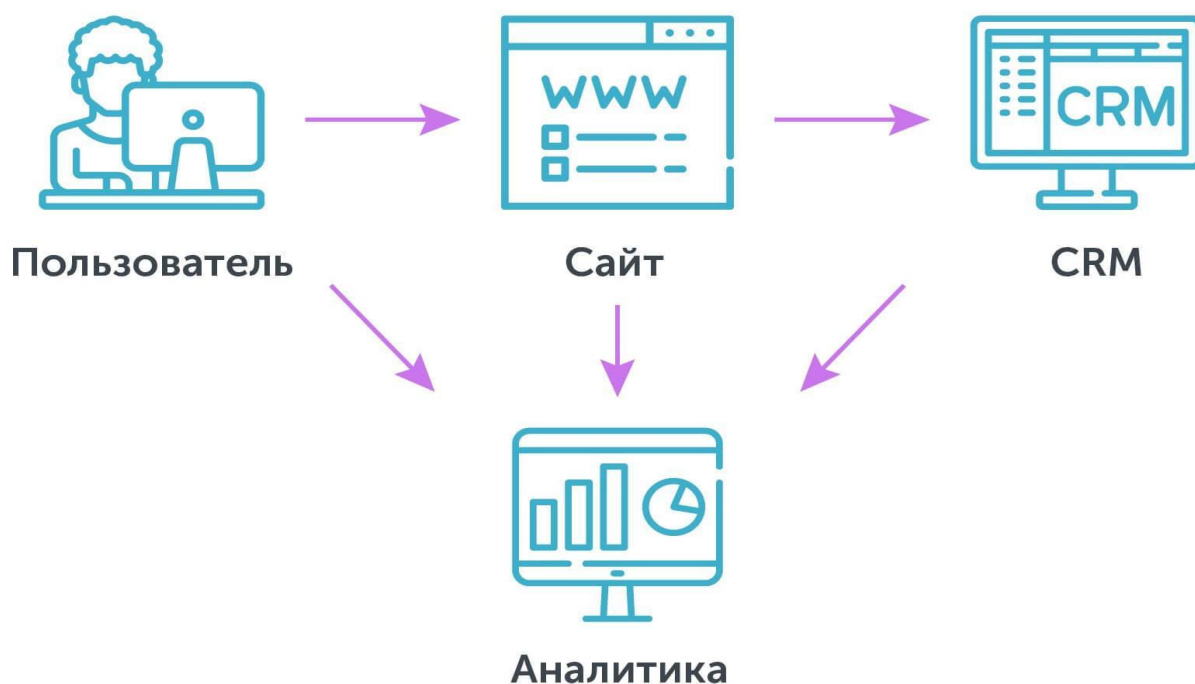


Рисунок 1. Схема работы систем аналитики.

Для сбора пользовательских данных при открытии сайта на страницу встраивается код, который непосредственно и выполняет сбор данных. С помощью этого можно отслеживать передвижения клиента по вашему сайту и строить граф переходов, определять наиболее покупаемые товары или услуги, отслеживать успешные и неуспешные транзакции и выяснять, как те или иные факторы влияют на конверсию. Конверсия в интернет-маркетинге - это отношение числа посетителей сайта, выполнивших на нём какие-либо конверсионные действия (покупку, регистрацию, подписку, посещение определенной страницы сайта, переход по рекламной ссылке), к общему числу посетителей сайта, выраженное в процентном соотношении [4].

Однако серьезной проблемой являются случаи, когда данные о совершенном клиентом

действию из-за различных факторов не доходят до системы аналитики. Такими факторами могут быть программные, технические или другие трудности. Например, клиентское приложение (зачастую таким приложением выступает браузер) не может выполнить передачу данных в систему аналитики. Причины отказа приложения передавать данные в аналитику могут быть различными: установлены программы, блокирующие сбор данных; отсутствие программной возможности исполнить JS-код, на языке которого описываются инструкции для сбора данных.

Другим фактором может быть асинхронная работа приложения - результат действий становится известен намного позже самого действия. Например, если платеж выполняется банковским переводом, то между созданием счета на оплату и зачислением платежа могут пройти несколько дней. Когда платеж все-таки зачислится браузер клиента может быть выключен или вкладка сайта может быть закрыта, соответственно и возможности передать информацию об успешном платеже не будет.

Для данных о переходах по сайту подобные потери не особо критичны, однако они могут негативно сказаться на совершенных транзакциях, которые не попали в систему аналитики. Потеря части полезных сведений может привести к неточно выдвинутым гипотезам и, соответственно, финансово-убыточным решениям. Чтобы такого избежать, нужно каким-то образом восстановить отсутствующие данные.

Решением этой проблемы может быть имитация действий пользователя. Имитация представляет собой переход на определенную страницу сайта и исполнение некоторого набора последовательных действий от лица пользователя. Такой подход позволит передать в систему аналитики все недостающие данные, причем система аналитики не способна будет различить, где действия выполнялось реальным пользователем, а где программно посредством вышеописанного метода. Однако для запуска имитации необходимы некоторые сведения:

- нужна ли имитация;
- окружение пользователя в момент совершения действия;
- результат действия пользователя.

Необходимость имитации можно определить по различным параметрам в момент совершения действия: клиентское приложение, метод оплаты и прочие. Если, проанализировав параметры, мы пришли к выводу, что имитация нужна, то нам нужно восстановить окружение пользователя, в котором совершалось действие. Окружение пользователя представляет собой совокупность различных параметров: уникальные идентификаторы в системах аналитики, источники перехода на сайт, клиентское приложение, идентификатор транзакции и прочие сведения. Затем, когда система получит результат совершенного пользователем действия, можно запускать имитацию.

Для имитации можно использовать готовые инструменты, позволяющие открывать сайты с сервера, не имеющего графической оболочки, а соответственно и открывать оконные приложения. Наиболее популярными инструментами для этого можно считать:

- Headless Chrome;
- PhantomJS.

Все из перечисленных выше инструментов представляют собой браузеры, не имеющие графического интерфейса, также их еще называют “безголовыми” браузерами [5]. Компоненты обеспечивающие работоспособность обычного и “безголового” браузеров для сравнения изображены на рисунках 2 и 3.

Одной из основных задач, решаемых “безголовыми” браузерами является скрейпинг веб-страниц, который представляет собой извлечение данных, расположенных на веб-страницах [6]. Однако, помимо этого они имеют встроенный интерпретатор для исполнения JS-кода, который в настоящее время широко применяется в интернет-сфере. JS используется для манипуляции данными и динамического изменения веб-содержимого страницы.

Обычный браузер

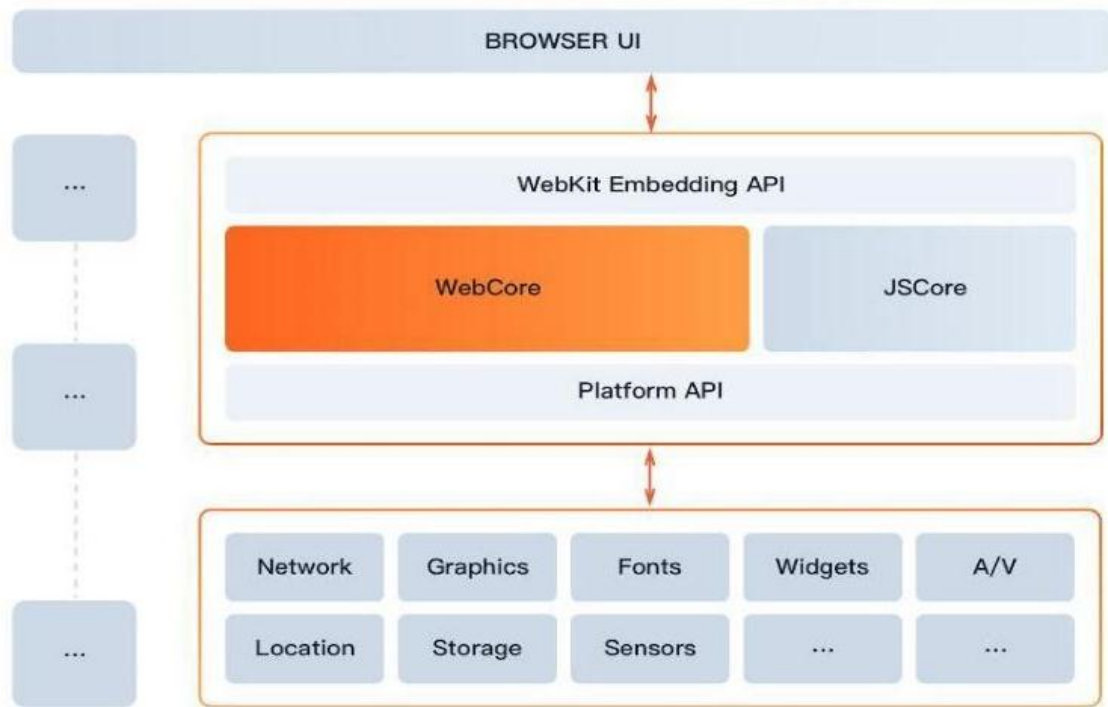


Рисунок 2. Устройство обычного браузера

Безголовый браузер

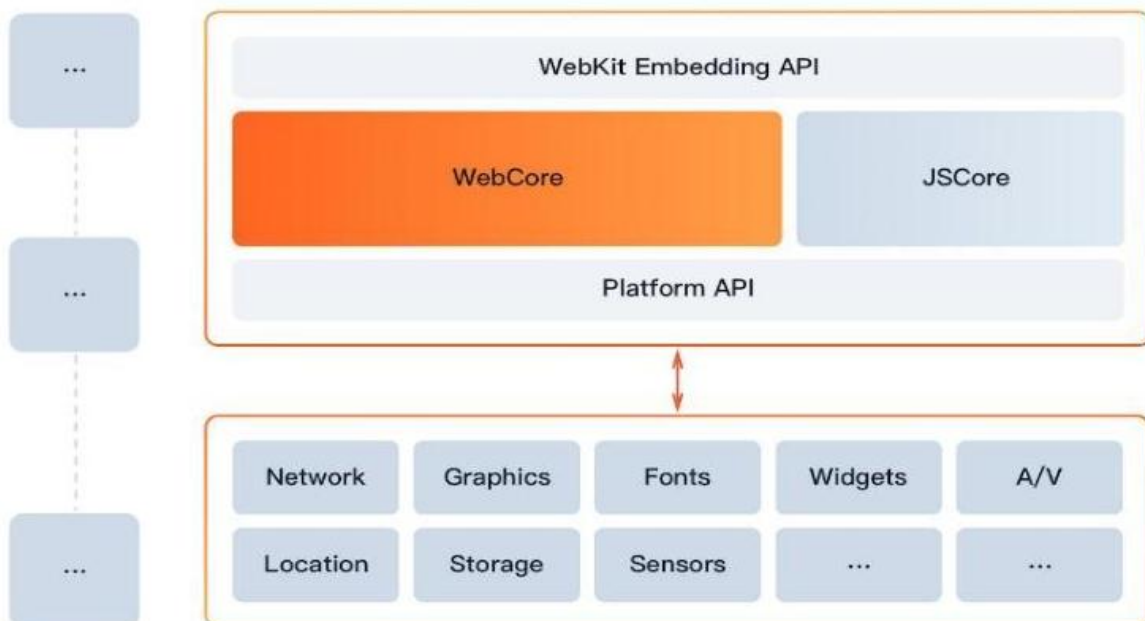


Рисунок 3. Устройство “безголового” браузера

Все из представленных инструментов оснащены механизмами для выполнения рендера страниц и интерпретатором для исполнения JS-кода, расположенного на ней. Однако наличие подобного набора возможностей у “безголовых” браузеров для успешной имитации недостаточно - необходимо также сгенерировать инструкции, которые осуществляют передачу данных; подготовить данные, которые должны быть переданы в систему аналитики; а также восстановить окружение пользователя, чтобы связать результаты выполнения с действиями, осуществленными реальным пользователем ранее.

На рисунке 4 изображена диаграмма последовательности имитации действий пользователя при зачислении банковского платежа. В соответствии с разработанной последовательностью, имитацию запускает серверное приложение, которое выполняет подготовку и затем запускает процесс имитации. Подготовив данные и запустив процесс, управление передается одному из возможных инструментов описанных выше. При этом выполняется переход на нужную страницу сайта и исполняется весь JS-код, расположенный на ней. Исполняя код, инструмент передает данные о платеже, такие как: сумма, валюта, оплаченный товар или услугу для последующего анализа в систему аналитики, которая связывает эти данные с предыдущими действиями пользователя.

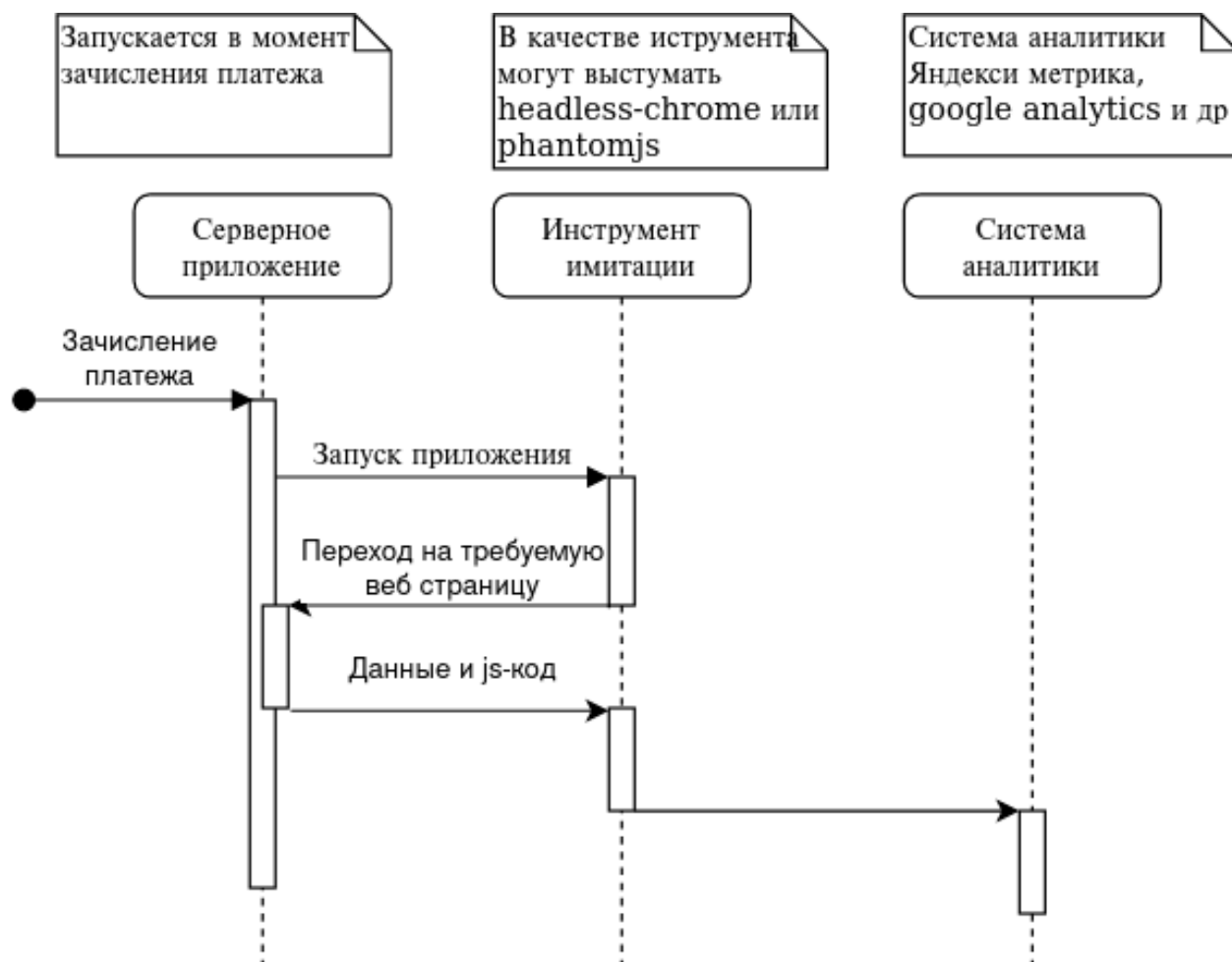


Рисунок 4. Диаграмма последовательности имитации действий пользователя

Заключение

В результате применения предложенного подхода к имитации действий пользователя возможно передать в систему аналитики не зафиксированные в ней ранее данные о совершенных действиях клиента и результаты этих действий. Благодаря восстановлению пропущенных данных увеличивается репрезентативность выборки данных, что в свою очередь позволяет проводить более детальный анализ.

Имея больше информации, мы можем выдвигать наиболее точные гипотезы, относительно постоянно меняющегося рынка интернет торговли и потребностей клиентов. Это позволит нам поддерживать конкурентоспособность и увеличивать извлекаемую прибыль, своевременно подстраиваясь под меняющиеся требования общества.

Подход к имитации описанный в статье был реализован в биллинговой панели Billmanager и успешно применяется для сбора информации о платежах, которые не были переданы в систему аналитики. На текущий момент необходимость имитации определяется в Billmanager на основании метода оплаты выбранного при создании платежа.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Власова А.С. Электронная коммерция [Электронный ресурс] - URL: <https://cyberleninka.ru/article/n/elektronnaya-kommertsiya-3>. (Дата обращения: 14.05.2020)
2. Электронная торговля - 2019: время возможностей : [Электронный ресурс] - URL: <https://oborot.ru/articles/elektronnaya-torgovlya-2019-vremya-vozmozhnostej-i108296.html>. (Дата обращения: 14.05.2020)
3. Яковлев А.Я. Веб-аналитика: основы, секреты, трюки. А. Яковлев, А. Довжиков. - СПб.:БХВ-Петербург, 2010. 272с
4. Романенко Е.В. Влияние конверсии на эффективность интернет-магазина // Международный научный журнал «Инновационная наука». 2016. N 6.
5. Как работает Headless Chrome. [Электронный ресурс]. URL: <https://habr.com/ru/company/oleg-bunin/blog/421137/> (Дата обращения: 15.05.2020)
6. Редькин О.К. Подходы к представлению текста для определения типа источника информационного сообщения // Интернет-журнал «НАУКОВЕДЕНИЕ». 2016. Том 8, N 6. <http://naukovedenie.ru/PDF/73TVN616.pdf> (доступ свободный).
7. Оладько В.С. Аудит информационной безопасности в электронной коммерции // Международный научно-исследовательский журнал. 2019. N 11 (42).
8. Олевинский М. А. Веб аналитика. сравнение систем веб-аналитики // Научное периодическое издание «In situ». 2015. N4.

REFERENCES

1. Romanenko E.V. The effect of conversion on the effectiveness of the online store // Innovative science. 2016. № 6.
2. E-commerce-2019: time of possibilities: [Electronic resource] - URL: <https://oborot.ru/articles/elektronnaya-torgovlya-2019-vremya-vozmozhnostej-i108296.html>. (Date accessed: 05/14/2020)
3. Redkin O.K. Approaches to the presentation of text to determine the type of source of an informational message // Internet journal "SCIENCE". 2016. Volume 8, N 6. <http://naukovedenie.ru/PDF/73TVN616.pdf>.
4. Oladko V.S. Information Security Audit In E-Commerce // International Research Journal.2019. N 11 (42).
5. Vlasova A.S. E-commerce: [Electronic resource] - URL: <https://cyberleninka.ru/article/n/elektronnaya-kommertsiya-3>. (Date accessed: 05/14/2020).
6. Olevinsky M. A. Web analytics. comparison of web analytics systems // Scientific periodical edition "In situ". 2015. N 4.
7. Yakovlev, A.Y47 Web analytics: basics, secrets, tricks I A. Yakovlev, A. Dovzhikov. - SPb.: BHV-Petersburg, 2010. P. 272.
8. How Headless Chrome works: [Electronic resource] - URL: <https://habr.com/en/company/oleg-bunin/blog/421137/> (Date of access: 05/15/2020).

Информация об авторе

Губанов Илья Алексеевич – студент группы ИСм.1-18-1, Иркутский государственный университет путей сообщений, г. Иркутск, e-mail: il07ya96@mail.ru

Authors

Gubanov Ilya Alecseevich - student of the ISM.1-18-1 group, Irkutsk State University of Railway Engineering, Irkutsk, e-mail: il07ya96@mail.ru

Для цитирования

Губанов И.А. Подход к имитации действий пользователя для сбора данных системы аналитики [Электронный ресурс] / В. А. Губанов // Молодая наука Сибири: электрон. науч. журн. – 2020. – №3(9). – Режим доступа: <http://mny.irkups.ru/toma/39-2020>, свободный. – Загл. с экрана. – Яз. рус., англ. (дата обращения: 02.10.2020)

For citation

Gubanov I.A. An approach to simulation of user actions for data collection of analytics system. *The electronic scientific journal "Young science of Siberia"*, 2020, no. 3(9). [Accessed 02/10/20] (in Russian)