

Законы ЕС о куки и конфиденциальности и их последствия для сбора и анализа данных



Защита конфиденциальности — очень важная тема, особенно в цифровом мире. Вопрос о том, как в интернете идет сбор данных и что с ними происходит потом, обсуждается очень широко, и вполне заслуженно.

Или другой пример: люди не догадываются, что режим приватного просмотра, он же инкогнито или InPrivate (в разных браузерах он называется по-разному), еще не означает, что сайты, которые вы посещаете, не собирают о вас никакой информации. Все, что происходит в этих режимах, — ваши данные (куки, история посещений и т.п.), — не сохраняются на вашем же компьютере, когда вы закрываете браузер. Если ваш работодатель или интернет-провайдер следит за тем, как вы используете Сеть, они смогут следить за вами даже тогда, когда вы в режиме инкогнито или InPrivate. (Осторожнее! :))

Так что все сложно. Теперь вы понимаете, почему вокруг этого столько путаницы и споров.

В последнее время эти споры дважды разгорались с особой силой: в первый раз из-за того, что Facebook и Twitter отслеживали пользовательское поведение в Сети посредством сайтов, на которых установлены их социальные кнопки либо интегрирована система комментариев Facebook. Второй раз — из-за разрабатываемых Евросоюзом правил использования куки.

В этом посте я рассмотрю вторую проблему, а именно последствия некоторых правил ЕС, касающихся куки. И хотя в фокусе рассмотрения будет именно Европа, даже если вы живете где-то еще, вы многое сможете почерпнуть для себя о конфиденциальности в Сети, сборе данных, оптимальных инструментах и наилучших способах спланировать вашу стратегию работы с данными.

Процесс сбора цифровых данных также очень сложен. Возможно, это отражает саму природу интернета и то, как он, подобно взрыву, ворвался в нашу жизнь. Даже продвинутым пользователям не так просто узнать все об этом процессе, так что нет смысла ожидать от обычных людей понимания того, что происходит на самом деле.

Многие, например, очень удивляются, узнав, что запись их поведения на том или ином сайте ведется в веблоге сервера даже тогда, когда на этом сайте не установлены аналитические или рекламные инструменты. Сохраняется информация об IP-адресе, запрашиваемых страницах, метки времени, идентификаторы браузера и т.д. Затем можно провести базовую обработку этих данных с помощью готового программного обеспечения.



.TXT
АВИНАШ
КОШИК
Евангелист
Google

Поскольку мои публикации не посвящены юридическим вопросам (я не юрист!), я не буду фокусировать внимание на том, насколько замечательны эти законы, как вы можете их соблюдать или какие виды отслеживания поведения пользователей вам разрешены или нет. Если вам нужна помощь в этом, обратитесь к вашему местному юристу, специализирующемуся на этих технологиях. Пожалуйста, удостоверьтесь, что ваша работа соответствует вашим местным законам.

Я рассмотрю последствия положений, разрешающих или запрещающих использование куки. Надеюсь помочь вам понять, как эти решения отражаются на обработке собранных данных, какие параметры использовать нельзя, а какие можно, а также какие решения вы можете принимать с уверенностью, а какие — с долей сомнения.

СБОР ДАННЫХ В СЕТИ: КУКИ И ИНСТРУМЕНТЫ

Прежде чем продолжать, давайте определим контекст и разберемся с терминами, которые мы будем использовать в этой статье.

Куки

Куки — это небольшой текстовый файл, размещенный в вашем компьютере. В куки содержится некоторое количество анонимной информации, позволяющей сайту узнать, какие страницы вы посещали до этого, как вы отреагировали на ту или иную рекламную кампанию или что у вас в корзине лежит такое-то количество товаров (благодаря этому сайт может сохранять содержимое вашей корзины, когда вы на него возвращаетесь) и т.п.

Куки всегда устанавливаются от лица владельца сайта. Так, они эксплицитно запускают механизм отслеживания (например, SiteCatalyst, comScore), обычно через тэги JavaScript, или рекламное решение (такое, как DoubleClick или Kenshoo), или социальные кнопки/системы комментирования.

Исходные куки (*First-party cookies*)

Это куки, устанавливаемые от вашего имени на вашем собственном домене под вашим собственным доменным именем. Например, когда вы посещаете этот сайт, Google Analytics (а он использует только исходные куки) разместит маленький текстовый файл на вашем компьютере.

Этот текстовый файл (куки) может быть прочитан только в том же браузере, в который был установлен, и только на том же веб-сайте. Другими словами, он отслеживает ваше поведение лишь на этом сайте, и нигде больше.

Куки не бывают вечными, они могут исчезнуть по самым разным причинам. Но исходные куки — самые устойчивые, в том смысле, что их реже всего отклоняют, они дольше всего сохраняются и что при «чистке куки» их удаляется меньше, чем других.

Сторонние куки (*third-party cookies*)

Эти куки устанавливаются от вашего имени, но через другой домен. К примеру, если бы на этом сайте была использована система комментариев Facebook, ваше поведение на сайте отслеживалось бы при каждом посещении сайта. Но поскольку этот куки (и ваши анонимные данные в этом куки) не установлен на моем домене kaushik.net, куки позволяет отслеживать ваше поведение и на других сайтах. Так что если после моего сайта вы посетите Gawker или CatsThatMakeAnalystsLaugh.com, Facebook узнает и об этом.

Хранящиеся в куки данные анонимны, но если вы зайдете в раздел «политика конфиденциальности» на facebook.com, вы узнаете, как по правилам соцсети используются эти данные о вашем поведении в интернете и на самом Facebook.

Сторонние куки используют многие провайдеры. Пожалуй, чаще всего ими пользуются рекламные платформы (Yahoo!, DoubleClick, Microsoft и др.). Сторонние куки нужны, когда применяются инструменты поведенческого таргетинга, которым необходима сводная информация об особенностях поведения того или иного юзера на разных сайтах.

По описанным выше причинам сторонние куки — наименее устойчивые. Они чаще всего отклоняются дефолтными настройками браузера или по выбору пользователя, они легче всего вычищаются специальными программами и т.д.

Некоторые производители аналитических инструментов все еще пользуются сторонними куки, о чем их клиенты зачастую не знают. Это неразумно. Проверьте, каким типом куки пользуется ваш аналитический инструмент. Если это сторонние куки, то ваши данные настолько далеки от оптимальных, что лучше всего, как бы трудно это ни было, перейти с использования сторонних куки на исходные.

ИНСТРУМЕНТЫ ВЕБ-АНАЛИТИКИ:

Такие инструменты, как Google Analytics, CoreMetrics, Baidu Tongji, StatCounter и прочие, основная задача которых — измерять параметры поведения пользователя (вашего поведения) на сайте.

Инструменты рекламной аналитики

Это инструменты, которые вы применяете, скажем, с рекламными или социальными средствами на вашем сайте. Их основная цель — предоставить вам сервис (рекламные объявления, плагины для комментариев, социальные кнопки) и отслеживать пользовательское поведение на разных сайтах.

Возможно, информации слишком много и она сложная, но очень важно хорошо понимать основные термины, прежде чем включаться в дискуссию о конфиденциальности и сборе данных. Иначе невозможно будет понять, какое правило применимо в том или ином случае и каковы будут последствия применения этих правил.

Помните: куки очень важны для понимания пользовательского поведения на сайте или нескольких сайтах. Но даже если вы не используете куки и цифровые аналитические инструменты, данные о поведении пользователей на вашем сайте все равно записываются в веблог сервера.

ЕВРОПЕЙСКИЕ ПРАВИЛА КОНФИДЕНЦИАЛЬНОСТИ И ИХ ПРАВОВЫЕ ПОСЛЕДСТВИЯ

Я не адвокат и даже не играю роль адвоката на телевидении :), так что я не буду высказывать своего мнения о тех или иных законах. Лучше меня об этом вам расскажет юрист вашей местной юрисдикции.

В настоящий момент существует Директива ЕС по электронной конфиденциальности, которая в разных странах Евросоюза имеет разные интерпретации и практики применения. Директива требует получить согласие пользователя, прежде чем начнется отслеживание его активности в интернете. Один из разделов Директивы напрямую связан с использованием куки.

Поскольку в разных странах-членах ЕС этот документ трактуется немного по-разному, сайт из Великобритании может запросить у вас один вид разрешения, используя определенный метод, и разрешить отслеживание по таким-то параметрам, а сайт в

Германии, Нидерландах или Испании запросит другое разрешение, и совсем по-другому.

Сейчас в Европе можно обозначить четыре практики применения Директивы в части, касающейся использования куки. Давайте посмотрим на один параметр: правовые последствия применения каждого из четырех видов трактовки правил (принятых национальными правительствами), касающихся данных, которые вы собираете, и анализа этих данных, который вам разрешен.

1. Закон в части использования куки не претерпел изменений.

Жизнь прекрасна. Жизнь продолжает течь в нормальном русле. Вы можете сфокусироваться на сборе лучших метрик, вести анализ данных, трудиться денно и нощно, обеспечивая пользователям наилучшие впечатления от пользования Сетью и повышая прибыльность присутствия вашей компании в интернете. Поручите кому-то из сотрудников компании пристально следить за издаваемыми правительством правилами и рекомендациями. Убедитесь, что ваша политика конфиденциальности в части, относящейся к данным, которые вы собираете, и выборам ваших юзеров, прозрачна и соответствует текущим нормам, и у вас не будет проблем с законом.

P.S.: Кстати, если ваши аналитические инструменты используют сторонние куки, переходите на исходные куки, а если ваш инструмент к этому не приспособлен, то с ним пора распрощаться.

2. Закон не претерпел изменений в части использования исходных куки. Использование сторонних куки возможно только с согласия пользователей.

В этой группе оказались довольно многие страны ЕС. Каковы последствия?

Для исходных куки: никаких проблем с данными о пользовательском поведении, которые вы собираете с помощью аналитических инструментов на своем собственном сайте. Используйте эти данные, чтобы всем было хорошо. (Если ваши аналитические инструменты используют сторонние куки, тогда, к вам относится все написанное ниже.)

Для сторонних куки: когда пользователи заходят на ваш сайт, они должны видеть сообщение, в котором вы предлагаете им согласиться на отслеживание их действий при помощи сторонних куки.

Они могут согласиться или нет, а вы должны сделать так, чтобы ваша платформа запомнила их выбор и действовала в соответствии с ним.

Если пользователь согласен на куки, то для вас нет никаких проблем.

Если пользователь отказывается: в большинстве случаев рекламные объявления, системы комментирования в социальных сетях и прочие инструменты, использующие сторонние куки, будут продолжать работать. Реклама будет показываться, комментарии будут опрашиваться. Но на работе вашего инструмента анализа рекламы это скажется.

Данные, которые вы сможете собрать, будут хуже, чем прежде (хотя и прежде они были не всегда точны). Данные о количестве кликов, соотношении показов и нажатий, просмотрах и показах рекламы, о переходах и других параметрах, предоставляемые вашей рекламной платформой, станут еще менее точными.

Если вы используете решения для поведенческого таргетинга, ретаргетинга или ремаркетинга, то их эффективность также снизится — просто потому, что ретаргетинг и ремаркетинг ориентируются на сторонние куки при наблюдении за поведением пользователя на разных сайтах, выдавая на основе этого наблюдения оптимальный контент или объявления.

Если вы используете социальные инструменты для комментирования, предоставленные третьей стороной, ваш сайт не будет запоминать пользователя.

Если пользователи не согласны на использование сторонних куки, то на некоторых платформах вы не сможете отслеживать их поведение на других сайтах. На одной социальной платформе отслеживание ведется очень умным способом: Bonita видит ваш пост в социальной платформе, затем посещает сайт, упомянутый в вашем социальном посте, оставляет комментарий под этим постом и нажимает на кнопку «поделиться», снова публикуя ваш пост на социальной платформе. Этот метод работает в основном благодаря сторонним куки, а без них может не сработать, так что вы не сможете проанализировать это поведение.

Эффект будет в целом зависеть от инструментов анализа цифровой рекламы, которые вы используете. Свяжитесь со своим консультантом в компании, которая продала вам аналитический инструмент. Используя приведенную выше информацию, подробно его расспросите (от некоторых поставщиков бывает нелегко получить прямые ответы). Меняйте свою стратегию анализа данных в соответствии с полученными ответами.

3. *Согласие требуется для использования как сторонних, так и исходных куки.*

Пара европейских стран относятся к этой группе. Не совсем понятно, означает ли отказ от куки, что трекинг вообще невозможен или что куки просто нельзя сохранять. Предположим, что согласие спрашивается только на применение куки (то есть данные сохраняются в браузере пользователя).

Если пользователь согласен, то не происходит никаких изменений. Ваши инструменты веб-аналитики и рекламной аналитики будут безо всяких проблем предоставлять вам нужную информацию.

Если пользователь отвергает ваше предложение и не соглашается на установку куки в своем браузере, события будут развиваться по одному из двух сценариев.

3.1 Некоторые инструменты веб-аналитики будут продолжать собирать данные (о хитах). Затем они попытаются «сшить» сессию из кусков. Поскольку инструмент получает информацию о каждом хите (через тэг JavaScript), в него будет поступать довольно много информации, хотя это и не информация от куки. То, что будут выдавать аналитические инструменты, используя анонимный идентификатор браузера и IP-адрес, можно описать примерно так:

«Похоже, все эти хиты были сделаны из одного браузера, у них близкие метки времени, и, хотя у нас нет куки, хиты начались в момент «X» и закончились в момент «X+5», так что это, должно быть, было одно посещение!»

Все данные анонимны (но то же самое относится и к нормальным инструментам веб-аналитики, так как большинство из них не собирают информацию о персональных данных пользователей).

Последствия для параметра «посещения»: вы получаете наилучший возможный вариант по посещениям. Не идеальный, но почти.

Последствия для параметра «уникальные посетители»: поскольку куки не сохраняются, каждый раз, когда пользователь будет заходить на сайт (даже из того же самого браузера), он будет идентифицироваться как новый пользователь. Так что подсчет количества уникальных посетителей будет неточным, а уж насколько неточным — зависит от того, сколько пользователей будут возвращаться на сайт. Последствия для параметра «новые и повторные посетители»: см. выше. Цифры будут неточными (количество новых посетителей будет завышено). Очевидно, что по тем же причинам данные по недавним и постоянным посетителям также не будут отличаться точностью.

Узнайте у вашего продавца аналитических инструментов, не используют ли они этот метод. Насколько я знаю, его применяют очень немногие.

3.2 Большая часть аналитических инструментов распознают невозможность использования куки (после того, как пользователь ответит «нет» на вопрос о согласии использовать их) и не смогут получить никакой информации о его браузере и посещении. Их мотив — предоставить вам исключительно «чистые» данные (насколько это возможно), которые вы можете анализировать, будучи уверенным в их точности.

ТАК ПОСТУПАЕТ GOOGLE ANALYTICS

Таким образом, поведение юзеров, не согласившихся на использование куки, не подлежит изменению и не отражается в получаемых при помощи инструмента данных, так что данные будут отражать поведение лишь части аудитории.

Если исходные куки не принимаются, то количество пользователей, отказывающихся от куки, остается неизвестным, — помните, они не хотят, чтобы их поведение отслеживали.

Также стоит помнить о том, что я писал в начале: скорее всего, ваш браузер по-прежнему собирает информацию обо всех хитах и запросах. По умолчанию веб-серверы не устанавливают куки, и, соответственно, в их логах нет этой информации. Но другая информация — IP-адреса, идентификаторы пользовательского браузера, временные метки, адреса страниц и многое другое — сохраняется в логах. Данные этих логов можно анализировать, используя доступные программы-парсеры.

Хотя информация, полученная таким образом, будет сильно уступать в качестве данным SiteCatalyst или Yahoo! Web Analytics, вы все же сможете анализировать данные о многих параметрах пользовательского поведения, используя показания логов.

Убедитесь в том, что политика безопасности вашего сайта предупреждает пользователей об использовании логов, и проконсультируйтесь с экспертом в вашей стране и местности.

4. Куки разрешены, сбор информации об IP-адресах запрещен полностью или частично.

Такую политику сейчас применяет одна европейская страна, и возможно, что к ней присоединятся другие. Правительство этой страны решило, что IP-адрес должен быть отнесен к персональным данным, сбор

которых инструментами веб-аналитики недопустим.

Если вы находитесь в этой стране, то отчеты будут наименее точными в разделе географии и местоположения посетителей, независимо от того, какие куки используются. В Google Analytics эти отчеты находятся в разделе Аудитория>Демография>Местоположение.

Если ваш аналитический инструмент позволяет вам сообщать IP-адреса (в Google Analytics это невозможно) и сопоставлять их с адресами компаний и т.п., то в случае применения такой практики и это будет невозможно. Со всеми прочими данными все должно быть в порядке. Таковы четыре основных сценария, работающих в настоящее время. Я надеюсь, что написанное выше помогло вам лучше понять правовые последствия европейских правил для работы ваших инструментов веб-аналитики и рекламной аналитики.

ЗАКЛЮЧЕНИЕ: НЕ ПАНИКУЙТЕ ИЗ-ЗА НЕДОСТАТКА ДАННЫХ

Если вы вспомните о том, как мы производили измерения эффективности маркетинга (и как мы до сих пор измеряем ее на телевидении, радио, в журналах, газетах, билбордах и т.д.), вы придете к выводу, что то, что мы называем «измерением», по сути является актом веры.

Если вы задумаетесь о том, как мы измеряем пользовательский опыт, (используя наблюдения, лабораторные эксперименты и т.д.), то поймете, что мы экстраполировали поведение десяти, ста или тысячи человек на поведение миллионов людей.

Теперь если вы посмотрите на данные, которые собираете с помощью своего аналитического инструмента, то будете приятно поражены тем, как много у вас информации. Очень много. 80%? 90%? 60%? В любом случае, это значительно больше, чем информация по любому другому рекламному каналу.

Так что вы можете поплакать о тех данных, которых у вас нет и не будет. А можете радоваться, что у вас в 5 000 раз больше информации, чем было бы по любому другому каналу рекламы на планете, анализировать эту информацию и использовать результаты анализа. А если раньше, имея 100 % данных, вы принимали на их основании гигантское грандиозное большое решение, то теперь, с теми данными, которые у вас есть, вы можете принимать просто грандиозные большие решения.

Это ведь все же лучше, чем голая слепая вера, движущая рекламой в офлайне, не правда ли?

Пусть ваше стремление к совершенству не мешает вам принимать решения, основываясь на просто хороших результатах.