

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В БИБЛИОТЕКАХ

УДК 026.06

А. А. Леонтьев

ГПНТБ СО РАН

Статистика посещаемости библиотечного сайта и квантовый принцип неопределённости

Рассмотрены особенности статистики посещаемости библиотечного сайта, в которой доля интернет-роботов составляет около 85%. Высказано предположение, что такая картина может быть интерпретирована с точки зрения квантового принципа неопределённости. Сделан вывод, что использование компьютерных технологий для предоставления доступа к информации и одновременно для учёта её востребованности пользователями интернета приводит к драматическому искажению результатов и создаёт возможность управлять этими результатами произвольно.

Ключевые слова: библиотечный сайт, статистика посещаемости, интернет-роботы, поисковые машины, группы влияния, принцип неопределённости.

UDC 026.06

Alexey Leontyev

Russian Academy of Sciences Siberian Branch, State Public Library for Science and Technology, Novosibirsk, Russia

Statistics of library website traffic and uncertainty principle

This article considers the features in statistics of library website traffic that includes a deal of bots of about 85%. It has been assumed that the features may be interpreted in terms of quantum uncertainty principle. A conclusion has been made that application of computer technologies to provide an access to information sources and, at the same time, to log the demand for the information on the Internet has a result that dramatic distortions and possible tweaks in the figures may occur at the discretion of somebody.

Key words: website traffic, Internet bots, search engines, lobby, uncertainty principle.

Последние исследования показывают, что доля интернет-роботов в статистике посещений библиотечного сайта составляет 85%*. Это означает, что на долю *живых* пользователей остаётся не более 15%. Между тем посещаемость интернет-сайтов представляет собой важнейший показатель, по которому судят об успешности работы того или иного учреждения, во всяком случае научного, включая и научные библиотеки. Поэтому становится злободневным вопрос о том, как корректно учесть реальную заинтересованность аудитории в контенте, представленном на страницах библиотечного сайта.

С одной стороны, традиционным является подход, когда в статистику посещений сайта включают все визиты со всех узлов интернета. Учитываются роботы, цифра посещаемости максимальна. Замечательно выглядят кривые роста количества визитов к сайту год от года; отчётность на высоте. Но динамика посещений по годам больше отражает прирост год от года активности роботов. Доля *живых*, реальных пользователей – конечных потребителей информации – относительно мала.

С другой стороны, можно фильтровать посетителей-роботов. Но тогда цифры посещаемости невыгодно снижаются, отчётность падает. Выводы об эффективности работы учреждения становятся угрожающими. Кроме того, полный запрет на посещения сайта роботами приведёт к невозможности его индексирования и, следовательно, к исчезновению сайта из результатов поисковых запросов. В конечном итоге сайт не будут посещать не только роботы, но и реальные пользователи. Стремясь наилучшим образом упорядочить статистическую картину, мы, образно говоря, достигнем нулевой температуры по Кельвину, когда прекращается хаотическое броуновское движение.

Ситуация усугубляется тем, что невозможно заранее определить, какой IP-адрес – машины-робота, а какой – конечного посетителя, заинтересованного в информации с веб-сайта. Лучшее, что можно сделать, – задним числом, т.е. после некоторого значительного числа посещений, констатировать, что посетитель с данного интернет-адреса, вероятно, является роботом. При этом каждый из вновь появляющихся роботов в течение некоторого времени будет оставаться «чёрным ящиком». В то же время реальные

* Канн С. К. Трафик ботов к библиотечному сайту (приблизительная оценка) // Наука и образование в жизни современного общества : сб. науч. тр. по материалам междунар. науч.-практ. конф. 30 апреля 2015 г. : в 14 т. Том 4 / М-во образования и науки РФ. – Тамбов : Консалтинговая компания Юком, 2015. – С. 63–64.

Kann S. K. Trafik botov k bibliotечnomu saytu (priblizitel'naya otsenka) // Nauka i obrazovanie v zhizni sovremennogo obshchestva : sb. nauch. tr. po materialam mezhdunarod. nauch.-prakt. konf. 30 aprelya 2015 g. : v 14 t. Tom 4 / M-vo obrazovaniya i nauki RF. – Tambov : Konsaltingovaya kompaniya Yukom, 2015. – S. 63–64.

пользователи из числа более опытных, напротив, могут предстать в образе роботов, отказавшись от ручного просмотра отдельных страниц и воспользовавшись автоматическими средствами для скачивания целого массива информации для последующего анализа и/или использования.

Итак, драматичность ситуации заключается в классической неопределённости – *«отфильтровать нельзя пропустить»*. Значит, мы должны принять тот факт, что в нашей тщательно анализируемой аудитории читателей/посетителей веб-сайта, которую мы подразделяем по десяткам разных признаков (пол, возраст, сфера деятельности, интересы или, например, платежеспособность), около 85% – совсем не те, кого бы мы хотели изучать. Забавно, что принятая форма подачи результатов статистических измерений включает десятые и сотые доли процента, притом что погрешность может доходить до 85!...

Как же быть с выявленными особенностями предмета нашего исследования? На помощь придут явления, знакомые физикам при изучении движения очень лёгких тел. Давно известно, что действия экспериментатора, интересующегося движением объектов, оказывают влияние на эти объекты. Посмотрите, как, к примеру, измеряют скорость самолёта. В его корпус монтируют датчик, который воспринимает давление встречного воздуха, и по этому давлению определяют скорость движения самолёта. Естественно, датчик оказывает дополнительное сопротивление полёту, вследствие чего скорость движения самолёта уменьшается. Но в случае реального самолёта такое воздействие на его полёт исчезающее мало.

Представим теперь, что масса условного самолёта и тяга его двигателей уменьшаются, но часть мощности по-прежнему расходуется для срабатывания датчика. При дальнейшем уменьшении массы условного самолёта наступает момент, когда фактически весь он превращается в один лишь датчик. Доля сопротивления датчика скорости в процессе полёта становится основной, например, 85%. И тогда мы изучаем уже не динамику полёта самолёта, а динамику полёта собственно датчика. Избежать этого принципиально невозможно. Для очень лёгких частиц невозможно выяснить скорость их движения, не нарушив сам характер движения. В этом, собственно, и состоит квантовый принцип неопределённости.

Вернёмся к исследованию движения информации в человеческом обществе. Во времена господства традиционных библиотек функция поиска необходимой информации осуществлялась посредством карточных каталогов. Учёт книгопользования проводился путём фиксирования статистики выдачи книг – отдельной книги отдельному читателю библиотеки.

При переходе информации и библиотек в интернет-пространство функция поиска выполняется поисковыми машинами. Интернет наполняется информационными агентами, т.е. средствами, осуществляющими запросы на получение информации, доступ к ней и её обработку или использование. При этом конечные цели информационных агентов могут варьироваться в очень широком диапазоне. В первом приближении можно разделить их на агентов, которые заинтересованы в получении информации как таковой, и агентов, которые считывают информацию из интернета только для того, чтобы затем организовать её упорядочение, или электронную каталогизацию, и предоставить к ней доступ конечному потребителю. Последний получает, таким образом, информацию уже из рук этих электронных каталогизаторов и в соответствии с их разумениями и предпочтениями.

Найденные величины 85 и 15 процентов определяют соотношение, которое в настоящий момент типично для библиотеки, – между долей индексаторов, т.е. посредников, при передаче информации и долей конечных потребителей информации. Это соотношение имеет фундаментальное значение: оно означает, что информация, подобно очень лёгким физическим телам, неиндифферентна к изучению своего движения.

Природа вещей, как показывает практика, такова, что для обеспечения эффективного доступа к информации в современном мире должно быть задействовано гораздо больше обращений к данным, чем для непосредственного их использования. Более того, существует внутренне присущее информационным агентам свойство, не позволяющее чётко разделить их на агентов-посредников и агентов-потребителей. Но это означает, что средства, необходимые для изучения движения информации, настолько сильно переплетены в электронном мире с непосредственным «полезным» её перемещением, что, по сути, не могут быть отделены друг от друга.

Законы движения информации представляются, таким образом, подобными законам движения лёгких тел: вмешательство исследователя драматическим образом меняет сам характер этого движения. Природа обмена электронной информацией такова, что она существенно ограничивает её изучения.

Побочные результаты изучения статистики посещаемости библиотечных сайтов и интернет-сайтов могут быть истолкованы и в другом контексте. Интернет-пространство можно рассматривать как арену столкновения различных групп влияния. В этом случае поисковые машины и их роботы предстают в качестве агентов, представляющих такие группы влияния.

Идеальная картина, из которой исходят исследователи, изучая статистику посещений, предположительно состоит из информации и её читателей. Владельцы сайтов поставляют информацию в интернет, а читатели, или

пользователи интернета, её востребуют. В зависимости от субъективных предпочтений степень востребованности информации варьируется. Одни страницы пользуются повышенным интересом читателей, другие, напротив, характеризуются малой посещаемостью. Популярность различных страниц отражается через статистические показатели сайта. Исследователь статистики посещений надеется сделать вывод как о качестве интернет-страниц, так и о свойствах субъектов, посещающих эти страницы.

Однако, в соответствии с найденным соотношением 85/15, в действительности дело обстоит иначе. Основную картину в статистике формируют агенты глобальных групп влияния – роботы. Хотим мы этого или нет, мировая аудитория интернета составляет лишь малую толику реального оборота информации. Доминирующая часть трафика обусловлена конкурентной борьбой за умы пользователей, поскольку поисковые машины – это не что иное, как инструмент для одновременного изучения интернет-аудитории и влияния на неё с точки зрения интересов определённых глобальных конкурирующих групп.

Таким образом, исходные посылки, выдвинутые в начале исследования, не получают своего подтверждения. Инструмент анализа общей статистики посещений даёт, прежде всего, картину столкновений и борьбы неких автоматических технологий на фронте компьютерных, т.е. тоже автоматических, информационных технологий. Это означает, что человек как родоначальник компьютерного прогресса оказывается, по сути дела, на периферии данной методологии. По-видимому, большой армии статистических исследователей следует отдавать себе отчёт в неизбежности этого явления.

Автоматическая генерация цифр и отчётов позволяет с относительной лёгкостью производить на свет впечатляющее количество данных и численных тенденций. Но остаётся очень большим вопросом, насколько эти отчётные «простыни» способны отразить не воспалённую фантазию так называемого компьютерного разума, а тонкие, трудноуловимые, ускользающие взгляды, вкусы, мотивы, разумения не смоделированного на скорую руку, а живого и необъятного человека.

Aleksey Leontyev – Cand. Sci. (Chemistry), researcher, State Public Library for Science and Technology of the Russian Academy of Sciences Siberian Branch;

alex@prometeus.nsc.ru

6 Lavrentyeva Prospekt, 630090, Novosibirsk, Russia