

Глава 1

Погоня за релевантностью

Если ты за что-то не платишь, ты не клиент; ты продукт, который выставляется на продажу.

Эндрю Льюис под псевдонимом Blue_beetle на сайте MetaFilter

Как-то весной 1994 года Николас Негропонте в задумчивости сидел перед экраном компьютера. В его детище — лаборатории MIT Media Lab* — молодые разработчики чипов, мастера виртуальной реальности и гении роботехники неистово трудились над игрушками и инструментами будущего. Но Негропонте размышлял над более простой проблемой — той, с которой миллионы людей сталкиваются каждый день: что смотреть по телевизору.

К середине 90-х круглые сутки семь дней в неделю вещали уже сотни телеканалов. Большинство программ были ужасающими и скучными: реклама кухонной техники, клипы групп-однодневок, мультифильмы и новости из жизни знаменитостей. Любому отдельно взятому зрителю была интересна лишь малая часть из этого.

По мере увеличения числа каналов стандартный метод поиска нужной передачи становился все более безнадежным занятием. Одно дело искать подходящую программу на пяти каналах, другое —

* Лаборатория Массачусетского технологического института, где занимаются изучением новых медиа и технологий. Прим. пер.

на пяти сотнях. А когда их число доходит до пяти тысяч, старые методы оказываются просто бесполезными.

Но Негропонте не отчаялся. Еще не все было потеряно. Более того, решение было уже почти найдено. «Ключ к будущему телевидения, — писал он, — перестать думать о телевизоре как о телевизоре», а начать думать о нем как об устройстве со встроенным интеллектом. Потребителям был нужен пульт, который управляет сам собой: интеллектуальный автоматический помощник, способный запоминать, что смотрит каждый зритель, и подбирать нужные программы. «Сегодня телевизоры позволяют вам контролировать яркость, громкость и канал, — писал Негропонте. — Завтра они позволят вам корректировать уровень секса, насилия и политическую позицию»¹.

А зачем останавливаться на этом? Негропонте воображал себе будущее, кишащее интеллектуальными агентами, которые помогают решать проблемы вроде выбора телепрограммы. Они, как личный дворецкий, должны впускать в ваш дом лишь ваши любимые шоу и темы. «Представьте себе будущее, — писал Негропонте, — в котором ваш интерфейсный агент сможет читать любые ленты новостей, газеты, отслеживать все теле- и радиостанции на планете, а затем составлять персонализированный дайджест. Это газета, выходящая тиражом в один экземпляр... Назовем ее Daily Me — “Мои ведомости”»².

И чем больше он думал, тем больше смысла в этом видел. Решением проблемы информационной перегрузки в цифровую эпоху были разумные, персонализированные встроенные редакторы. И они вовсе не должны были ограничиваться телевидением. Как заметил Негропонте в беседе с редактором журнала Wired, «интеллектуальные агенты, несомненно, будущее компьютерной отрасли»³.

Джарон Ланир из Сан-Франциско был весьма встревожен этими рассуждениями. Ланир — один из создателей виртуальной реальности, и с 80-х годов прошлого века он пытался придумать, как сблизить людей и компьютеры. Но разговоры об «агентах» поразили его своим безумием. «Что в вас вселилось? — спрашивал он в послании ко всем «поклонникам стиля Wired», которое опубликовал на своем сайте. — Идея «интеллектуальных агентов» дурна и неправильна... Вопрос об агентах кажется решающим фактором в том, будет ли Сеть гораздо лучше, чем ТВ, или гораздо хуже»⁴.

Ланир был убежден, что агенты, не будучи людьми, станут склонять нас к неуклюжему и хаотическому взаимодействию с собой. «Модель ваших интересов, созданная агентом, будет мультишной, и глазами агента вы увидите только мультишный мир», — писал он.

Была и еще одна проблема. Идеальный агент, по логике, должен был блокировать всю или почти всю рекламу. Но поскольку онлайн-торговлю двигала вперед именно реклама, казалось маловероятным, что эти компании станут использовать агентов, наносящих такой удар по их прибыли. Куда вероятнее, как считал Ланир, что это будут «двойные агенты», подкупленные, «непонятно на кого работающие».

Это было ясное и четкое предостережение. Но хотя оно вызвало кое-какие споры на онлайн-форумах, компании-гиганты той ранней интернет-эпохи не сочли его убедительным. Им, наоборот, импонировала логика Негропонте: компания, сообразившая, как просеивать горы цифрового песка в поисках крупин золота, могла бы завладеть будущим. Они предвидели надвигающийся кризис внимания — информационный выбор каждого человека становился бесконечным. А значит, если вы хотели заработать, вам нужно было убедить людей «подключиться». В мире, где внимание дефицитно,

лучшим решением было предлагать контент, действительно близкий интересам, желаниям и нуждам каждого человека. В коридорах и информационных центрах Кремниевой долины распространился новый девиз: релевантность.

Все подряд рвались вывести на рынок «интеллектуальный» продукт. Microsoft выпустила Bob — целую операционную систему, построенную на концепции агентов. Ведущим в ней выступал странный мультишний персонаж, изрядно смахивавший на Билла Гейтса. Apple, почти за 10 лет до анонса iPhone, выпустила Newton — «личного настольного помощника», главным доводом в пользу покупки которого считали агента, почтительно затаившегося прямо под бежевой крышкой устройства.

В итоге новые «интеллектуальные» продукты провалились. В чатах и электронных рассылках развернулась целая кампания издевательств в адрес Bob. Пользователи просто не выносили его. Журнал PC World окрестил его одним из 25 худших продуктов высоких технологий за всю историю⁵. Apple Newton добился немногим большего: хотя компания инвестировала более 100 миллионов долларов⁶ в его разработку, в первые 6 месяцев своего существования он продавался очень плохо. Взаимодействие с интеллектуальными агентами 90-х быстро давало понять: не так уж они и умны.

Сейчас, десять с лишним лет спустя, интеллектуальных агентов по-прежнему нет. Похоже, что революция, предсказанная Негропонте, не состоялась. По утрам нас не будит электронный дворецкий, и мы не уведомляем его о наших планах и пожеланиях на день.

Но это не значит, что таких агентов не существует вовсе. Просто они скрыты от нас. Персональные интеллектуальные агенты прячутся за фасадом всех сайтов, на которые мы заходим. Каждый день они становятся умнее и влиятельнее, накапливая все больше информации

о том, кто мы такие и что нас интересует. Как и предсказывал Ланир, агенты работают не только на нас: они также служат интернет-гигантам вроде Google, распределяя как контент, так и рекламу. Хотя они ничем не напоминают мультишного Боба, придуманного Microsoft, они управляют все большей долей наших действий в онлайне.

В 1995 году гонка за личную релевантность только начиналась. И она, возможно, больше всех остальных факторов повлияла на формирование Интернета и на то, каков он сегодня.

Проблема Джона Ирвинга

Джефф Безос, глава Amazon.com, одним из первых понял, что силу релевантности можно подчинить и заработать на этом миллиарды. Он начал свой бизнес в 1994 году, исходя из концепции, что онлайн-книготорговлю нужно вернуть «во времена мелких книготорговцев, которые знали вас очень хорошо и могли давать советы вроде: «Я знаю, вам нравится Джон Ирвинг, и представьте себе, вот новый автор, который, по-моему, очень похож на Джона Ирвинга»». Об этом Безос рассказывал своему биографу⁷. Но как же реализовать это в большом масштабе? Безос считал, что Amazon должен стать «чем-то вроде небольшой компании искусственного интеллекта»⁸, приводимой в действие алгоритмами, которые мгновенно подбирают для клиентов книги.

В 1994 году, когда Безос работал компьютерным инженером на Уолл-стрит, его нанял один венчурный капиталист. Он ждал бизнес-идей, применимых на зарождающемся интернет-рынке. Безос методично про-работал тему и составил список из 20 продуктов, которые его команда теоретически могла бы продавать в онлайне: музыка, одежда, электроника, — а потом начал вникать в особенности каждой отрасли. Книги

сначала стояли в самом низу списка, но, когда появились окончательные результаты, к изумлению Безоса, они поднялись на самый верх⁹.

По ряду причин книги оказались идеальным товаром. Прежде всего, отрасль была децентрализованной; крупнейший издатель, Random House, контролировал лишь 10 процентов рынка¹⁰. Если бы какой-то один издатель не захотел продавать книги онлайн-магазину, многие другие восполнили бы этот пробел. И покупателям не нужно было много времени, чтобы привыкнуть к приобретению книг в Интернете: большинство книжных продаж и так уже происходило вне традиционных книжных магазинов и, в отличие от одежды, книги не нужно примерять. Однако главной причиной привлекательности книг как товара стал тот простой факт, что их было очень много: в 1994 году в продаже было 3 миллиона наименований книг и только 300 тысяч наименований компакт-дисков¹¹. Обычный книжный магазин ни за что не вместил бы все эти книги, а онлайновый — запросто.

Когда Безос доложил об этом своему шефу, тот не очень-то заинтересовался. В информационную эпоху книжная отрасль казалась отсталой. Но Безос никак не мог забыть об этом. Не имея ограничений на число книг на складе, он мог выставлять на витрину в сотни, тысячи раз больше наименований, чем отраслевые гиганты вроде Borders или Barnes & Noble, и при этом обеспечить клиенту более душевные впечатления от покупки, чем крупные сети.

Он решил: Amazon должен усовершенствовать процесс открытия. Задачей персонализированного магазина было помочь читателям находить книги и представлять им новинки. Но как?

Безос начал думать, как научить этому компьютеры. Над этой серьезной проблемой бились немало инженеров и ученых в исследовательских учреждениях вроде МИТ и Калифорнийского университета

в Беркли начиная с 50-х. Эта область изысканий называлась «кибернетикой»¹². Слово было позаимствовано у Платона, который обозначал им саморегулирующуюся систему, например демократию. Для первых специалистов в этой области самым захватывающим занятием было создавать системы, способные подстраивать самих себя на основе обратной связи. За несколько десятилетий они заложили математические и теоретические основания, обеспечившие Amazon значительную часть его роста.

В 1990 году группа ученых в исследовательском центре Xerox в Пало-Альто (PARC) применила кибернетический подход к новой проблеме. PARC известен своими идеями, которые затем подхватывали и коммерциализировали другие компании — достаточно вспомнить графический пользовательский интерфейс и компьютерную мышь¹³. Исследователи из PARC, как и многие продвинутые компьютерные специалисты в то время, оказались в числе первых профессиональных пользователей электронной почты: они отправляли и получали сотни сообщений. Электронная почта оказалась прекрасным изобретением, но ее минусы тоже стали быстро очевидны. Когда послать сообщение любому количеству людей ничего не стоит, вы быстро тонете в потоке бесполезной информации.

Чтобы уследить за этим потоком, команда PARC принялась изобретать процесс, который они назвали «коллаборативной фильтрацией»^{*}

* Коллаборативная фильтрация, или совместная фильтрация, — это метод анализа информации и интересов пользователя, основанный на сборе данных о предпочтениях и действиях других пользователей. Он основан на идее, что люди, у которых в прошлом обнаружились общие интересы, могут иметь сходные интересы и в будущем. Такие системы могут прогнозировать, например, какая музыка понравится вам, исходя из того, какая музыка понравилась людям, чьи вкусы в прошлом пересекались с вашими. *Прим. пер.*

и воплотили в программе Tapestry¹⁴. Она отслеживала, как люди реагируют на получаемые электронные письма: какие сообщения открывают, на какие отвечают, какие удаляют, — и затем использовала эту информацию, чтобы более удобно организовывать входящие сообщения. Письма подобные тем, на которые люди реагируют активно, должны были перемещаться вверх списка, а такие, которые получатели часто удаляют или не открывают, сползали вниз. В сущности, это был инструмент экономии времени: вместо того чтобы процеживать кучу сообщений лично, вы могли положиться на помощь других в предварительной обработке полученных писем.

Естественно, эта система работала не только с электронной почтой. Tapestry, по словам ее создателей, была «разработана для управления любым потоком входящих электронных документов. Электронная почта лишь один из примеров такого потока; другие примеры — ленты информагентств и статьи онлайн-форумов»¹⁵.

Tapestry представила миру коллаборативную фильтрацию, но в 1990 году такая услуга была не очень интересна. Интернет насчитывал всего несколько миллионов пользователей; он оставался маленькой экосистемой, и информации, подлежащей сортировке, было не так уж много, а пропускная способность каналов — не столь велика. Так что много лет коллаборативная фильтрация оставалась уделом компьютерных исследователей и скучающих студентов. Если бы в 1994 году вы отправили на адрес ringo@media.mit.edu список альбомов, которые вам нравятся, то получили бы в ответ письмо с рекомендациями новой музыки и рецензиями. На сайте сообщалось, что «один раз в час сервер обрабатывает все входящие сообщения и отправляет ответы»¹⁶. Это был ранний предшественник

Pandora^{*} — персонализированный музыкальный сервис для эпохи, когда широкополосного Интернета еще не существовало.

Но когда в 1995 году стартовал Amazon, все изменилось. С самого начала этот сайт представлял собой книжный магазин со встроенной персонализацией. Изучая, какие книги люди покупают, и используя методы коллаборативной фильтрации, изобретенные в PARC, Amazon мог выдавать рекомендации мгновенно. (О, вы берете «Руководство для чайников по фехтованию»? Может, возьмете еще «Очнулся слепым: судебные иски в связи с травмой глаза»?) И, отслеживая покупки, через какое-то время Amazon мог выделять пользователей с похожими предпочтениями. («Другие люди, которым нравится то же, что и вам, приобрели новинку этой недели — “Ангард!”») Чем больше книг люди покупали на Amazon, тем точнее была персонализация.

В 1997 году Безос обслужил первый миллион покупателей. Спустя полгода — два миллиона. А в 2001 году компания впервые получила квартальную прибыль: это был один из первых бизнесов, доказавших, что в онлайне можно делать серьезные деньги.

И хотя Amazon не мог в полной мере передать атмосферу местного книжного магазинчика, система персонализации работала весьма неплохо. Топ-менеджеры держат язык за зубами и не признаются, какую долю выручки она обеспечивает, но часто называют этот механизм ключевым элементом успеха компании.

На Amazon идет нескончаемая погоня за пользовательскими данными: когда вы читаете книги на ридере Kindle, информация о фразах, которые вы выделяете, страницах, которые вы переворачиваете,

* Популярное интернет-радио, подсказывающее пользователям новую музыку исходя из их предпочтений. *Прим. пер.*

и о том, читаете ли вы внимательно или пролистываете, отправляется на серверы Amazon и используется, чтобы определить, какие книги могут вам еще понравиться. Если вы зайдете на сайт после дня на пляже с Kindle, Amazon может слегка изменить страницу, чтобы ее содержание соответствовало тому, что вы только что прочли. Если вы полдня читали новый роман Джеймса Паттерсона и лишь мельком заглянули в руководство по диете, то вы, возможно, увидите на первой странице сайта больше триллеров и куда меньше книг о здоровье¹⁷.

Пользователи Amazon так привыкли к персонализации, что сайт теперь использует обратный трюк, чтобы заработать еще больше денег. Издатели платят за размещение книг в традиционных магазинах, но не могут купить мнения продавцов. Однако, как и предсказывал Ланир, подкупить алгоритм нетрудно: заплатите Amazon достаточную сумму, и ваша книга будет продвигаться под видом «объективной» рекомендации собственного софта сайта¹⁸. Большинство клиентов не способны отличить одно от другого.

Amazon доказал, что релевантность может обеспечить доминирующие позиции в отрасли. Но затем на сцену вышли два аспиранта из Стэнфорда, которые применили принципы машинного обучения ко всему миру онлайн-информации.

КЛИК — ЭТО СИГНАЛ

Когда новая компания Джекфа Бозоса только начинала работу, основатели Google Ларри Пейдж и Сергей Брин занимались своими докторскими диссертациями в Стэнфорде. Они знали об успехе Amazon: в 1997 году «пузырь доткомов» раздувался вовсю, и Amazon — по крайней мере на бумаге — стоил миллиарды. Пейдж и Брин были

математическими гениями; первый, в частности, был одержим темой искусственного интеллекта. Но их интересовала другая проблема. Что если использовать компьютерные алгоритмы не для более эффективной продажи товара, а для сортировки сайтов?

Пейдж изобрел новаторский подход к такой сортировке и с присущей компьютерным гикам склонностью к игре слов назвал его PageRank*. Большинство компаний, занимавшихся интернет-поиском в то время, сортировали страницы по ключевым словам и едва ли могли оценить, насколько первые соответствуют вторым. В статье, опубликованной в 1997 году, Брин и Пейдж сухо отмечали, что три из четырех крупнейших поисковых машин не могут найти сами себя. «Мы хотим, чтобы наша концепция “релевантности” распространялась только на лучшие документы, — писали они, — поскольку пользователю могут быть доступны десятки тысяч лишь слегка релевантных документов»¹⁹.

Пейдж понял, что в пронизанной ссылками структуре Сети кроется гораздо больше данных, чем могут использовать большинство поисковых машин. Тот факт, что одна страница содержит ссылку на другую, можно считать «голосом» в пользу второй. Пейдж наблюдал за тем, как стэнфордские профессора считают, сколько раз их статьи были процитированы, и составляют таким образом примерный рейтинг своей значимости. Он прикинул, что сайты, на которые часто ссылаются — например, главная страница Yahoo, — могут, подобно академическим статьям, считаться более значимыми, а те, за которые они «голосуют», тоже значат больше. Весь этот процесс, как утверждал Пейдж, «опирается на уникальную демократическую структуру Сети».

* Фамилия Пейдж (Page) в переводе с английского означает «страница». Прим. пер.

В те дни Google обитал на сайте google.stanford.edu, и Брин с Пейджем были убеждены, что сервис должен оставаться некоммерческим и свободным от рекламы. «Мы считаем, что поисковые машины, финансируемые за счет рекламы, будут неизбежно склоняться в сторону нужд рекламодателей, а не нужд потребителей, — писали они. — Чем лучше поисковая система, тем меньше сообщений понадобится потребителю, чтобы найти искомое... мы уверены, что вопрос о рекламе создает достаточно неоднозначные стимулы, и поэтому крайне важно иметь конкурентоспособную поисковую систему, которая прозрачна для пользователей и остается в академической сфере»²⁰.

Но когда они запустили бета-версию сайта на просторы Интернета, трафик зашкалил. Google действительно работал — внезапно он стал лучшей поисковой машиной Интернета. Вскоре искушение сделать на этом бизнес оказалось слишком сильным, и основатели Google, которым было по двадцать с небольшим, не смогли перед ним устоять.

По легенде, именно алгоритм PageRank вознес Google на вершину мирового господства. Я подозреваю, что компании нравится эта версия: это ясная, простая история, привязывающая успех поискового гиганта к одному гениальному прорыву, совершенному одним из основателей. Но с самого начала PageRank был лишь малой частью проекта Google. На самом деле Брин и Пейдж поняли вот что: ключ к релевантности, к сортировке массы данных в Интернете — это... еще больше данных.

Брину и Пейджу был важен не только сам факт, что страница ссылается на другую. Позиция ссылки, ее размеры, возраст страницы — все эти факторы имели значение. С годами Google стал называть эти путеводные нити, скрытые в данных, «сигналами».

С самого начала Пейдж и Брин понимали, что важнейшие сигналы будут поступать от самих пользователей поисковика. Скажем, если

кто-то ищет «Ларри Пейдж» и кликает на вторую ссылку в результатах, это тоже «голос»: он подсказывает, что вторая ссылка более релевантна, чем первая. Они назвали это «клик-сигналами» (click signal). «Очень интересными будут исследования, — писали Пейдж и Брин, — в основу которых лягут огромные объемы данных об использовании современных веб-систем... весьма трудно получить эту информацию, прежде всего потому, что она предположительно имеет коммерческую ценность»²¹. Вскоре в их руках оказалось одно из крупнейших в мире хранилищ таких данных.

По части информации Google был ненасытен. Брин и Пейдж намеревались сохранять все: каждую страницу, на которую когда-либо заходила поисковая машина, каждый клик каждого пользователя. Вскоре их серверы уже хранили копию большей части Интернета, обновляемую практически в режиме реального времени. Они были уверены, что, просеивая данные, найдут еще больше зацепок, еще больше сигналов, позволяющих уточнять результаты. Отдел качества поиска Google завоевал репутацию местного спецназа: минимум посетителей, абсолютная секретность — таковы были правила его работы²².

«Идеальная поисковая машина, — любил говорить Пейдж, — будет в точности понимать, что вы имеете в виду, и выдавать в точности то, что вы хотите»²³. Google должен был выдавать не тысячи ссылок в ответ на запрос, а одну, именно ту, которая нужна. Но идеальные ответы для разных пользователей — разные. Когда я ввожу в строку поиска «пантеры», вероятно, я имею в виду крупных диких кошек, а если это слово вводит футбольный фанат, то он, видимо, подразумевает футбольную команду из Южной Каролины. Чтобы добиться идеальной точности, нужно знать, чем конкретно интересуется каждый из нас. Нужно знать, что я ничего не понимаю в футболе; нужно знать, кто я такой.

Главной проблемой было получить достаточно данных и выяснить, что соответствует интересам каждого конкретного пользователя. Понять, чего хочет кто-то, — непростая задача, и, чтобы выполнить ее хорошо, нужно знать, как ведет себя человек в течение длительного времени.

Но как? В 2004 году компания выдвинула инновационную стратегию и начала предоставлять другие услуги — те, которые вынуждали пользователей регистрироваться и входить в Сеть. Одной из первых стала Gmail — чрезвычайно популярная электронная почта. Журналисты много писали о рекламе, выводимой в Gmail рядом с письмами, но маловероятно, что она была единственным мотивом запуска сервиса. Заставляя людей входить в Сеть под своим логином, Google заполучил в свое распоряжение колоссальные массивы данных: сотни миллионов писем, которые пользователи отправляют и получают каждый день. Компания теперь может сопоставлять сообщения и поведение каждого пользователя на сайте со ссылками, на которые он кликает в поисковой машине. Пакет онлайновых инструментов для работы с текстом и электронными таблицами Google Apps имел двойное назначение: с одной стороны, он подрывал позиции Microsoft, кровного врага Google, а с другой — стал еще одним крючком, заставляющим пользователей входить в Сеть и посыпать все новые клик-сигналы. Все эти данные позволили Google ускорить создание модели личности каждого пользователя — какими темами он интересуется, по каким ссылкам ходит.

К ноябрю 2008 года Google получил несколько патентов на алгоритмы персонализации — программный код, позволяющий выделить группы, к которым принадлежит индивид, и скорректировать поисковые результаты с учетом предпочтений. Категории Google оказались довольно узкими: для иллюстрации в патенте приводился пример «всех

лиц, интересующихся коллекционированием зубов древних акул», и «всех лиц, не интересующихся коллекционированием зубов древних акул»²⁴. При вводе слов «резцы большой белой акулы» люди из первой группы получили бы одни результаты, а из второй — другие.

Сегодня Google отслеживает любой сигнал от нас, который может заполучить. Важность этих данных трудно переоценить: если Google видит, что я вхожу в Сеть сперва из Нью-Йорка, потом из Сан-Франциско, потом снова из Нью-Йорка, то он понимает, что я регулярно летаю с одного побережья на другое, и может соответственно скорректировать выдаваемые результаты. Определив, какой браузер я использую, он может сделать некоторые выводы о моем возрасте и даже, возможно, о моих политических предпочтениях.

Время, проходящее от момента, когда вы вводите запрос, до момента, когда вы выбираете один из результатов, также проливает некоторый свет на вашу личность. И, конечно, сами поисковые запросы дают огромные объемы информации о вас.

Даже если вы не зашли в Сеть под своим логином, Google все равно выдает вам персонализированные результаты поиска. Ему доступна информация о районе — даже о квартале, — откуда вы зашли в Сеть, и это многое говорит о том, кто вы и чем интересуетесь. Слово «Sox» в поисковой строке, введенное на Уолл-стрит, вероятно, представляет собой сокращение от названия закона Сарбейнса — Оксли^{*}, тогда как, если запрос поступил из Стейтен-Айленда^{**}, речь, видимо, идет о названии бейсбольной команды.

* Закон 2002 года, существенно ужесточающий требования к финансовой отчетности компаний. *Прим. пер.*

** «Спальный» район Нью-Йорка, расположенный на одноименном острове. *Прим. пер.*

«Люди все время предполагают, будто с поиском мы уже разобрались, — говорил Пейдж в 2009 году. — Это очень далеко от истины. Возможно, мы прошли лишь пять процентов пути. Мы хотим создать идеальную поисковую машину, которая сможет понимать все... некоторые называют это искусственным интеллектом»²⁵.

В 2006 году на мероприятии Google Press Day гендиректор Google Эрик Шмидт изложил 5-летний план компании. Однажды, по его словам, Google сможет отвечать на вопросы вроде «В какой колледж мне стоит пойти?». «Пройдут годы, прежде чем мы сможем давать хотя бы частичные ответы на эти вопросы. Но в конечном итоге... Google сможет отвечать и на более гипотетические вопросы»²⁶.

Facebook повсюду

Алгоритмы Google не имеют себе равных, однако главная их задача — уговорить пользователей раскрыть свои вкусы и интересы. В феврале 2004 года в университетском общежитии Гарварда Марк Цукерберг придумал более простой подход. План, воплощенный в его творении Facebook, был таков: вместо того чтобы просеивать клик-сигналы в попытке понять, что же интересует людей, нужно просто спросить их.

С первого курса колледжа Цукерберг интересовался тем, что называл «социальным графом», — формальным описанием контактов и системы взаимоотношений каждого человека. Закачайте в компьютер эти данные, и он сможет делать довольно-таки интересные и полезные вещи: рассказывать вам, как дела у ваших друзей, где они побывали и чем интересуются. Это касалось и новостей: в самом первом своем воплощении, как внутренний гарвардский сайт,

Facebook автоматически проставлял на персональных страницах его членов ссылки на статьи газеты Crimson^{*}, где те упоминались.

Facebook был вовсе не первой социальной сетью: когда Цукерберг собирал в ночи свой сайт, небрежно сделанный музыкальный портал MySpace уже имел головокружительный успех. А до MySpace внимание технически подкованной аудитории на краткий миг смогла привлечь сеть Friendster. Но Цукерберг задумал совсем иной сайт — не службу знакомств для робких и застенчивых, какой был Friendster, и не приглашающую знакомых и не знакомых между собой людей пообщаться, как MySpace. Facebook стремился опираться на уже существующие в реальном мире социальные связи. В сравнении со своими предшественниками он выглядел весьма скромно: главный акцент был на информации, а не на кричащей графике или атмосфере. «Мы коммунальная служба», — говорил Цукерберг потом²⁷. Facebook больше походил не на ночной клуб, а на телефонную компанию; это была нейтральная платформа для общения и сотрудничества.

Но даже в своем первом воплощении сайт рос как на дрожжах. После того как Facebook начал работать в нескольких избранных университетах «Лиги плюща», почтовый ящик Цукерberга переполнили запросы студентов из других вузов, умолявших открыть доступ и им. К маю 2005 года сайт активно работал в восьми с лишним сотнях университетов. Но в высшую лигу Facebook вывела лента новостей, введенная на сайте в сентябре 2006 года.

На Friendster и MySpace нужно было заходить на страницы друзей, чтобы узнать их новости. Алгоритм ленты новостей на Facebook вытягивал все эти обновления из огромной базы социальной сети

* Ежедневная газета для студентов и сотрудников Гарвардского университета.
Прим. пер.

и размещал их на первой странице, показываемой сразу после ввода логина и пароля. Facebook мгновенно превратился из сети связанных страниц в персонализированную газету о ваших друзьях, авторами которой были они сами. Трудно вообразить более чистый механизм поставки релевантного контента.

И это была просто золотая жила. В 2006 году пользователи Facebook публиковали уже миллиарды обновлений: философские цитаты, пикантные новости о тех, с кем они встречались, замечания о том, что ели на завтрак. Цукерберг и его команда постоянно подстрекали пользователей к этому: чем больше данных те вручали компании, тем приятнее был их опыт и тем чаще они возвращались на сайт. Уже на раннем этапе добавилась возможность загружать фото, и Facebook стал крупнейшей фотоколлекцией в мире. Компания поощряла пользователей публиковать ссылки с других сайтов, и на Facebook появились миллионы ссылок. В 2007 году Цукерберг хвастался: «Мы производим больше новостей в день для 19 миллионов наших пользователей, чем любое другое СМИ за всю свою историю»²⁸.

Поначалу лента новостей показывала почти все, чем занимались ваши друзья на сайте. Но по мере роста числа сообщений и друзей лента стала нечитаемой и неуправляемой. Даже если у вас была всего сотня друзей, все равно читать приходилось слишком много.

Facebook выдвинул решение: EdgeRank, алгоритм, который управляет главной страницей. Он выстраивает с точки зрения релевантности все случаи взаимодействия на сайте. Его формула сложна, но главная идея довольно проста и основывается на трех критериях²⁹. Первый — это близость. Чем более тесные дружеские отношения у вас завязались — это определяется длительностью времени, которое вы потратили на взаимодействие с пользователем и просмотр профиля, —

тем выше вероятность, что Facebook будет показывать вам обновления этого человека. Второй критерий — относительный вес данного типа контента. Скажем, обновлениям статуса отношений придается очень большой вес, поскольку всем интересно знать, с кем встречаются их друзья. (Многие эксперты полагают, что относительные веса тоже персонализированы, ведь разным людям важны разные виды контента.) Третий критерий — время: более свежие сообщения считаются важнее устаревших.

EdgeRank продемонстрировал парадокс, лежащий в основе гонки за релевантностью. Чтобы показывать релевантные результаты, алгоритмам персонализации нужны данные. Но чем больше данных, тем более изощренные фильтры требуются, чтобы эффективно организовать их. Это бесконечная спираль.

К 2009 году число пользователей Facebook достигло 300 миллионов³⁰ и росло на 10 миллионов в месяц. Цукерберг в 25 лет стал миллиардером, пусть и виртуальным. Но компания мечтила гораздо выше. Цукерберг намеревался сделать из всей информации в мире подобие ленты друзей. Хотя он никогда не говорил об этом напрямую, цель была ясна: опираясь на социальный граф и массу информации, предоставленной пользователями Facebook, внедрить свой алгоритм обработки новостей в центр всей Сети.

21 апреля 2010 года многих ждал большой сюрприз: читатели газеты Washington Post обнаружили*, что на главной странице ее сайта появились их друзья. В приметной рамке в верхнем правом углу — именно там, как подтвердят любой редактор, взгляд останавливается прежде всего — оказался раздел под названием Network

* Должен оговориться: в начале 2010 года я давал Washington Post небольшую консультацию относительно онлайн-сообществ и присутствия в Интернете.

News. Каждый посетитель видел в нем свою подборку контента — это были ссылки на статьи Washington Post, которыми его друзья поделились в Facebook. Газета позволила Facebook редактировать свой самый ценный онлайн-актив — главную страницу. Вскоре этому примеру последовала New York Times.

Новая функция была частью гораздо более крупного проекта, получившего название Facebook Everywhere («Facebook повсюду») и представленного на ежегодной конференции компании f8*. С тех пор как Стив Джобс предъявил миру Apple, объявив его «безумно великим», помпезность была одной из традиций Кремниевой долины. Слова Цукерберга, произнесенные 21 апреля 2010 года, звучали не менее убедительно. «Это самая радикальная трансформация, какую мы когда-либо предлагали Интернету», — заявил он³¹.

Цель Facebook Everywhere была проста: сделать весь Интернет «социальным» и внедрить персонализацию на миллионах сайтов. Хотите узнать, какую музыку слушают ваши друзья по Facebook? Pandora скажет вам. Хотите знать, какие рестораны нравятся вашим друзьям? У Yelp** теперь есть ответ. Новостные сайты — от Huffington Post до Washington Post — теперь персонализированы.

Facebook дал возможность поставить метку «Мне нравится» на любой объект в Интернете. В первые сутки после запуска нового сервиса таких отметок было поставлено около миллиарда, и все эти данные текли на серверы Facebook. Брет Тейлор, руководитель разработчиков платформы Facebook, объявил, что пользователи делятся друг с другом 25 миллиардами подобных объектов в месяц³². Google, прежде неоспоримый лидер в гонке за релевантностью, похоже, был встревожен

* F8 можно прочесть как *fate* — «судьба». Прим. пер.

** Сайт отзывов о ресторанах,очных клубах, гостиницах и т. д. Прим. пер.

действиями своего соперника, находящегося буквально в нескольких километрах.

Теперь два гиганта вступили в рукопашный бой: Facebook переманивает ключевых топ-менеджеров из Google; Google же усердно работает над социальной сетью, похожей на Facebook. Кажется неясным, почему двум великим мирам новых медиа обязательно воевать. Ведь суть Google в том, что он отвечает на вопросы, тогда как главная миссия Facebook — помогать людям общаться с друзьями.

Однако прибыль обоих бизнесов зависит от одного и того же: точно нацеленной, высокорелевантной рекламы. Контекстная реклама, которую Google размещает рядом с результатами поиска на вебстраницах, — это единственный значимый источник дохода компании. И хотя финансовая информация Facebook не раскрывается, инсайдеры отмечали, что реклама находится в центре его коммерческой модели. У Google и Facebook разные точки старта и разные стратегии — одна компания отталкивается от отношений между фрагментами информации, другая от отношений между людьми, — но в конечном итоге они конкурируют за одни и те же рекламные деньги.

С точки зрения рекламодателя вопрос прост: какая компания сможет обеспечить наибольшую прибыль? И вот здесь как никогда актуальна тема релевантности. Массивы данных, накопленных Facebook и Google, имеют две области применения. Пользователям они помогают получать нужные именно им новости и результаты. Рекламодателям же — искать потенциальных покупателей. Компания, у которой данных больше и которая применяет их с большей пользой, получит больше рекламных доходов.

Это подводит нас к патовой ситуации: в некий момент пользователи становятся настолько привязаны к технологии, что, даже если

конкуренты предложат более совершенный сервис, переход на него оправданным не будет. Если вы пользователь Facebook, только представьте, каково это: перейти в другую социальную сеть, даже если она предложит гораздо больше возможностей. Вероятно, потребуется много времени и усилий, и воссоздание профиля пользователя, загрузка всех картинок и усердный ввод имен друзей окажутся чрезвычайно нудным делом. Вы «подсели». Gmail, Google Talk, Google Voice, Google Docs и еще несколько продуктов — также часть тщательно выстроенной кампании, нацеленной на то, чтобы привязать к этой системе. Исход битвы между Google и Facebook зависит от того, кто из них сможет привязать к себе больше пользователей.

Динамику этого процесса описывает закон Меткалфа*, который гласит, что полезность Сети увеличивается нарастающими темпами по мере добавления каждого нового пользователя. Нет особого смысла в факте, если он у вас одного, но если у каждого, с кем вы работаете, есть такой аппарат, то не иметь его крайне невыгодно. Привязывание пользователя к технологии — оборотная сторона закона Меткалфа: Facebook во многом полезен именно потому, что им пользуются все. И понадобится немало управленческих ошибок, чтобы это ключевое обстоятельство изменилось.

Чем чаще пользователи посещают сайт, тем проще убедить их заходить на него под своим логином. А когда вы постоянно находитесь в системе, компания может отслеживать информацию о вас, даже если вы ушли с ее сайта. Если вы зашли в Gmail, а потом на сайт, использующий рекламную службу Doubleclick, принадлежащую Google, этот факт может быть «привязан» к вашему аккаунту Google.

* Боб Меткалф — изобретатель протокола Ethernet, позволяющего соединять разные компьютеры. *Прим. ред.*

И, отслеживая cookie-файлы, которые эти сервисы устанавливают на вашем компьютере, Facebook или Google могут показывать на сторонних сайтах рекламу с учетом вашей личной информации. Весь Интернет может стать платформой для Google или Facebook.

Причем Google с Facebook — далеко не единственные. Ежедневная битва за сферы влияния между ними занимает многочисленных бизнес-журналистов и выливается в гигабайты болтовни в блогах; однако в этой войне открывается мощный третий фронт. Хотя большинство участвующих в нем компаний находятся вне поля зрения публики, именно они в конечном итоге могут определить будущее персонализации.

РЫНОК ДАННЫХ

Охота за сообщниками террористов, устроивших «11 сентября», была одной из самых изматывающих в истории. Непосредственно после нападений масштаб заговора был неясен. Есть ли другие, еще не обнаруженные террористы? Насколько велика сеть, организовавшая атаки? ЦРУ, ФБР и масса других ведомств трое суток работали круглосуточно над обнаружением новых заговорщиков. Полеты самолетов по всей стране были остановлены, а аэропорты закрыты.

Помощь пришла, откуда не ждали. 14 сентября ФБР опубликовало имена террористов, захвативших самолеты, и просило — умоляло — любого, имевшего информацию о преступниках, поделиться ею. В тот же день позвонил Мэк Макларти, бывший чиновник Белого дома, который теперь входит в совет директоров малоизвестной, но чрезвычайно прибыльной компании под названием Axciom.

Как только имена террористов были опубликованы, Axciom стала искать их в своих колоссальных банках данных, которые занимают

два гектара площади в крошечном городе Конуэй, Арканзас. И компания обнаружила весьма интересную информацию. Более того, оказалось, что она знает об 11 из 19 заговорщиков больше, чем все правительственные ведомства США — включая их предыдущие и последние адреса и имена их соседей по квартире³³.

Возможно, мы так и не узнаем, что же было в файлах, которые Acxiom передала правительству (хотя один из ее топ-менеджеров сказал журналисту, что информация привела к депортациям и предъявлению обвинений). Но Acxiom достаточно знает о 96 процентах американских домохозяйств и о полумиллиарде людей за границей: имена членов семьи, предыдущие и нынешние адреса, как часто они оплачивают счет по своей кредитке, есть ли у них кошка или собака (и какой породы), левши они или правши, какие лекарства они используют (согласно аптечным базам рецептов)... В этом списке около 1500 пунктов.

Acxiom не высывается — может быть, компания не случайно получила такое название, которое не выговоришь. Но она обслуживает большинство крупнейших корпораций Америки — 9 из 10 крупных операторов кредитных карт и потребительские бренды от Microsoft до Blockbuster³⁴. «Считайте, что Acxiom — фабрика, — сказал журналисту один инженер компании, — только продукт, который мы выпускаем, — это данные»³⁵.

Чтобы прочувствовать представление Acxiom о будущем, возьмем туристические сайты Travelocity или Kayak. Вы когда-нибудь задумывались, на чем они зарабатывают? У Kayak два источника дохода. Один довольно простой — это наследие эры турагентов: когда вы покупаете авиабилеты, найденные на Kayak, авиакомпания платит сайту небольшую комиссию.

Другой менее очевиден. Когда вы подбираете себе билет, Kayak устанавливает на ваш компьютер маленький cookie-файл — по сути, наклейку на лоб со словами «расскажите мне о дешевых билетах с одного побережья на другое». Затем Kayak может продать эту информацию компании вроде Axciom или ее конкурента BlueKai, которая, в свою очередь, продает их тем, кто заплатит больше, — в данном случае, вероятно, крупной авиакомпании, скажем United³⁶. Как только United узнает, какие именно билеты вас интересуют, она сможет показать вам рекламу подходящих рейсов — и не только на сайте Kayak, а практически на любом сайте по всей Сети. И весь этот процесс — от сбора ваших данных до продажи их United — занимает меньше секунды³⁷.

Лидеры этого бизнеса называют такую схему «поведенческим ретаргетированием». Ретейлеры заметили, что 98 процентов посетителей онлайн-магазинов покидают их, не купив ничего³⁸. «Ретаргетирование» означает, что компании уже не принимают ответ «нет».

Допустим, вы посмотрели на сайте пару кроссовок, но ушли, не проявив к ним интереса. Если онлайн-магазин обуви, в котором вы их увидели, использует ретаргетирование, то их реклама — возможно, даже фотография тех самых кроссовок — будет преследовать вас по всему Интернету, появляясь рядом со счетом вечернего матча или сообщениями в вашем любимом блоге. А если вы наконец сдадитесь и купите кроссовки? Что ж, тогда онлайн-магазин сможет продать и эту информацию BlueKai, а та — перепродать ее, скажем, сайту, торгующему одеждой для спортсменов. И вскоре вы будете видеть на всех сайтах подряд рекламу впитывающих пот носков.

Эта настойчивая персонализированная реклама не обязательно привязана к вашему настольному компьютеру. Сайты типа Loopt

и Foursquare, позволяющие передавать информацию о местоположении пользователя с его мобильного телефона, дают рекламодателям возможность обращаться к потребителям с помощью таргетированной рекламы, даже когда те не сидят за компьютером. Loopt работает над рекламной платформой, обеспечивающей магазинам возможность присыпать информацию о специальных скидках и акциях постоянным покупателям на телефон — сразу, как только те входят в магазин. А если вы полетите рейсом Southwest Airlines, то реклама на экране в спинке кресла перед вами может отличаться от той, которую видят ваши соседи. Ведь Southwest знает ваше имя и род занятий, а сопоставляя эту личную информацию, к примеру, с базой данных Acxiom, может узнать о вас гораздо больше. Почему бы тогда не показывать вам вашу собственную рекламу или нацеленную специально на вас программу, рекламу в которой вы посмотрите с большей вероятностью?

TargusInfo, еще одна из новых компаний, занимающихся обработкой личной информации, хвастается, что «за год поставляет более 62 миллиардов атрибутов данных в реальном времени»³⁹. Это 62 миллиарда характеристик личностей покупателей, их действий и желаний. Другая организация со зловещим названием Rubicon Project утверждает, что в ее базе больше полутора миллиарда пользователей Интернета⁴⁰.

Сейчас ретаргетированием пользуются рекламодатели, но нет никаких оснований думать, что издатели и поставщики контента упустят такую возможность. Ведь если Los Angeles Times знает, что вы фанат блогера Переса Хилтона, редактор может поставить интервью с ним на главную страницу своего сайта в вашей персонализированной версии. Тогда вы с большей вероятностью останетесь на сайте газеты и кликнете на какой-нибудь еще материал.

Все это означает, что ваше поведение теперь — товар, крохотный элемент рынка, обеспечивающего платформу для персонализации всего Интернета. Мы привыкли думать об Интернете как о наборе двусторонних отношений: вы управляете своими взаимодействиями с Yahoo отдельно, а с вашим любимым блогом — тоже отдельно. Но за кадром Сеть становится все более интегрированной. Компании понимают, что делиться данными — это прибыльное дело. Благодаря Axciom и рынку данных сайты могут выводить на первый план самые подходящие для вас продукты и шептаться о вас между собой за вашей спиной.

Погоня за релевантностью обеспечила восхождение нынешних интернет-гигантов, и она мотивирует компании накапливать все больше данных о нас и незримо корректировать наши онлайн-впечатления. Она меняет саму ткань Сети. Но, как мы увидим, последствия персонализации в сфере потребления новостей, для принятия политических решений и даже нашего образа мышления будут еще более радикальными.