

Исследование связи между творчеством и технологиями...
восхитит читателей «Черного лебедя» Нассима Талеба.

КОД БЕСТСЕЛЛЕРА



*Джоди Арчер,
Мэтью Л. Джокерс*

«Арчер и Джокерс любят литературу и хотят, чтобы к хорошим книгам приходил успех». *Wired*

Мэттью Джокерс
Код бестселлера

«Азбука-Аттикус»

2016

УДК 808.1
ББК 81.1+81.2-5

Джокерс М. Л.

Код бестселлера / М. Л. Джокерс — «Азбука-Аттикус»,
2016

ISBN 978-5-389-13993-0

«Если верить расхожему мнению, любой бестселлер – аномалия. Удачная флуктуация на рынке. Интересный мутант. Черный лебедь. Но если это правда, то, единожды найдя писателя, способного выдавать бестселлеры, надо вкладывать все деньги в него? Зачем рисковать миллионами, ставя их на никому не известного двадцатилетнего юношу вместо Стивена Кинга? Используя компьютерную систему, которая читает книги, распознает в них определенные характеристики и просеивает тысячи таких характеристик в тысячах текстов, мы открыли, что существуют удивительные сочетания параметров, свойственные книгам, которые с наибольшей вероятностью будут пользоваться успехом на рынке. Наша система обрабатывала каждую книгу так, словно это новая, никому не известная рукопись, а затем не просто выносила один из двух вердиктов – “вероятно, бестселлер” или “скорее всего, нет”, – но и вычисляла индекс, указывающий, каковы шансы книги обрести популярность. Эти индексы сами по себе чрезвычайно интересны, и мы не только расскажем, как они устроены, но и объясним, что держит читателя как магнитом, заставляя переворачивать страницы одну за другой». Джоди Арчер, Мэттью Л. Джокерс

УДК 808.1
ББК 81.1+81.2-5

ISBN 978-5-389-13993-0

© Джокерс М. Л., 2016
© Азбука-Аттикус, 2016

Содержание

1. Бестселлерометр, или Почему анализ текста приводит к революции в книгоиздании	8
Белые лебеди	11
Из любви к книгам	15
Откуда мы взялись	17
Машинное чтение	20
Договор	24
1. Номер первый	25
2. Верьте нам, как мы доверились машине	26
3. Волшебного чая не существует	27
4. «Черный ящик»	28
2. Крестные родители, или Почему важно находить время друг для друга	29
Зачем мы читаем?	31
Секс, наркотики и рок-н-ролл	34
Проверка	41
Подходящие герои	43
Конец ознакомительного фрагмента.	46

Мэттью Джокерс, Джоди Арчер

Код бестселлера

Jodie Archer and Matthew L. Jockers

THE BESTSELLER CODE

Anatomy of the Blockbuster Novel

© Jodie Archer and Matthew L. Jockers, 2016

© Самсонова Т. П., перевод на русский язык, 2017

© Издание на русском языке, оформление. ООО «Издательская Группа «Азбука-Аттикус», 2017

КоЛибри®

* * *

Неожиданный, совершенно новый взгляд, основанный на солидных исследованиях».

GQ

«Читать эту книгу – чистое наслаждение. Рекомендуются и для развлечения, и для образования любому, кто интересуется настоящим и будущим книжной индустрии».

Digital Book World Daily

«Открывает неизведанные пути развития популярной литературы... Возможно, самым достойным результатом применения бестселлерометра будет демократизация книжного мира».

The Atlantic

«Этот интереснейший томик содержит мудрость, которой поделились с нами компьютеры – на случай, если и мы захотим написать бестселлер. А кто не хочет? Завораживающее чтение!»

The Times Review

«Начинающие романисты найдут здесь обильную пищу для ума... Книга открывает колоссальную возможность и для литературоведов».

Public Books

«Интригующее чтение».

Literary Review

«Авторы получили интереснейшие данные о “ДНК бестселлера” и создали книгу о том, как мы читаем, от которой невозможно оторваться. Замечательное достижение!»

Виктор Мейер-Шёнбергер, преподаватель Оксфорда

«Код бестселлера» привел меня в восторг, потряс и перевернул мои представления вверх дном. Арчер и Джокерс построили читающего робота, который может многому научить читателей, писателей и издателей в том, что

касается воздействия беллетристики на читателя. Это новаторская работа, прорыв в новой науке о повествовании».

Джонатан Готтшалл, профессор

*Посвящается Эндрю (отцу)
и Анджеле (жене)*

1. Бестселлерометр, или Почему анализ текста приводит к революции в книгоиздании

Одним прекрасным утром 2010 года литературный агент Стига Ларссона сидел и радовался жизни. 13 июня «Девушка, которая взрывала воздушные замки»¹ – третья книга в серии ранее неизвестного автора – дебютировала первой в списке бестселлеров в твердом переплете, составленном газетой New York Times. Спору нет, приятно увидеть такое за утренним кофе. «Воздушные замки» на первом месте, «Девушка с татуировкой дракона»² лидирует сразу в двух форматах среди книг в мягком переплете, и, в качестве приятного дополнения, «Девушка, которая играла с огнем»³ – на втором месте. И так – уже 49 недель в США и три года подряд в Европе. Есть чему радоваться.

Через месяц Amazon объявит, что Ларссон – первый в истории писатель, чьи творения разошлись в количестве миллиона экземпляров в формате для Kindle. За следующие два года совокупный тираж всех изданий Ларссона превысит 75 миллионов. Неплохо для никому не известного политического активиста из крохотной скандинавской страны, вдруг заделавшегося писателем. Особенно если учесть, что шведское издание вышло под довольно непривлекательным заглавием и что в книге содержатся откровенные описания насилия и пыток. «Мужчины, которые ненавидят женщин» – или «Девушка с татуировкой дракона», как называли английский перевод, – стала сенсацией года в тридцати с лишним странах.

Журналисты не могли объяснить этот успех. Крупные газеты заказывали специалистам большие статьи, желая понять, что, черт побери, происходит в книжном мире. Почему именно эта книга? Откуда такой ажиотаж? В чем секрет? Кто мог бы предсказать этот поворот событий?

Ответы не впечатляли. Литературные критики чесали в затылке. Они указывали на недостатки структуры, стиля, сюжета и характеров героев. Они ужасались качеству переводов. Они жаловались на глупость читающей публики. Но читатели продолжали расхватывать книги практически с печатного станка – в Великобритании, США, Японии, Германии; мужчины, женщины, старые, молодые, черные, белые, любой ориентации. Почти в любой точке земного шара у любого нашлись бы знакомые, прямо сейчас читающие трилогию о девушке с татуировкой дракона.

В книжном мире такое бывает не слишком часто. Феномен вроде Ларссона появляется раз в год, а может, и того реже. После Ларссона подобный оглушительный успех снискала Э. Л. Джеймс с книгой «Пятьдесят оттенков серого»⁴ – причем, в отличие от Ларссона, она была жива и могла принять участие в рекламном турне. Ларссон же умер, не дождавшись публикации. Его трилогия достигла такого огромного объема продаж вообще без участия автора. Непостижимо. Непредсказуемо. Просто аномалия какая-то.

Давайте посмотрим на цифры. В американском штате Делавэр есть компания Bowker – глобальный поставщик библиографической информации и единственный источник идентификационных номеров (ISBN) для книг, издающихся в США. По данным Bowker, каждый год в США выходит приблизительно 50–55 тысяч новых наименований художественной литературы. Учитывая, что все больше произведений публикуют сами авторы в электронном виде без получения ISBN, это число занижено. За год в список бестселлеров газеты New York

¹ Luftslottet som sprängdes, Stieg Larsson.

² Män som hatar kvinnor, Stieg Larsson.

³ Flickan som lekte med elden, Stieg Larsson.

⁴ Fifty Shades of Grey, E. L. James.

Times попадает примерно 200–220 книг. Даже если использовать заниженную общую цифру, это меньше чем 0,5 % от всех опубликованных. Из этой половины процента лишь немногие задерживаются в списке на несколько недель («двузначные», как называют их профессионалы книжного дела). И только малая доля авторов умудряется продержаться в списке десять недель или больше, а из их числа, вероятно, лишь у трех или четырех одна книга достигает миллионных тиражей в США в течение года. Но почему именно эта, а не другая?

Считается, что беллетристу, желающему завоевать читательскую аудиторию, нужно владеть определенными навыками: он должен правильно построить сюжет, убедительно обрисовать характеры и при этом грамотно писать. Очень успешные авторы пошли дальше: они зорко подмечают все состояния человеческого духа, умеют лихо закрутить сюжет не в ущерб правдоподобию и, наконец, редко, но метко пользуются точкой с запятой. Все это – признаки хорошего писателя, а хороший писатель, если он готов приложить достаточно усилий, в конце концов найдет своего читателя. Но мы говорим о всеобъемлющем успехе, когда сотни тысяч разом читают одно и то же – этот триллер, а не тот, этого кандидата на Пулитцеровскую премию, а не какого-либо другого. Если оставить в стороне вмешательство Опры Уинфри, это значит, что есть какой-то неучтенный фактор – будто на нашего автора просыпалась с неба щепотка звездной пыли, такой тонкой, что ее не распознать никакими приборами. Внезапный и необъяснимый успех таких книг, как трилогия о девушке с татуировкой дракона, «Пятьдесят оттенков серого», «Прислуга»⁵, «Исчезнувшая»⁶, «Код да Винчи»⁷, считается большой удачей автора, но при этом случайной, как выигрыш в лотерею.

Кстати сказать, слово «бестселлер» изначально было жаргонным словечком издательской отрасли. Оно относительно новое – впервые вошло в словари английского языка в конце XIX века, примерно тогда же, когда появились первые списки книг, лидирующих по количеству проданных экземпляров. Это слово по идее должно быть нейтральным, но за свою жизнь оно обросло определенными коннотациями, отчасти вводящими в заблуждение. Литературный журнал Bookman начал публиковать ежемесячные отчеты по продажам книг в 1891 году в Лондоне и в 1895-м – в Нью-Йорке, после появления Закона о международном авторском праве (1891), осложнившего распространение дешевых пиратских копий британских романов в США. До этого вести статистику по продажам было невозможно. Списки выходили во всех крупных городах и обычно включали в себя шесть книг, которые лучше всего продавались за истекший месяц. С самого начала эти списки выполняли две функции, невиданные ранее в книжном мире: во-первых, они составлялись исключительно на основе количества проданных экземпляров, а во-вторых, служили рекомендацией упомянутых книг для читателей. Эти рекомендации были основаны не на мнениях горстки критиков или рекламе издателей, а на выборе точно таких же читателей, обычных людей. Покупка книги была и остается единственным способом за нее проголосовать. Таким образом, термин «бестселлер» не несет никакой информации о качестве и тематике книги и не является синонимом выражений «жанровая проза» или «популярное чтение». Некоторые деятели литературного истеблишмента действительно употребляют это слово как оскорбление: они считают, что коллективный вкус читательской аудитории равняется только на плохую литературу, но объективные данные свидетельствуют об обратном. В числе бестселлеров были как книги, удостоенные Пулитцеровской премии, и великие американские романы, так и произведения, рассчитанные на массового читателя. Среди авторов можно найти не только Майкла Коннелли с Дебби Макомбер, но и Тони Моррисон с Маргарет Этвуд. Именно поэтому список бестселлеров – достойное внимания, интересное и динамичное культурное явление.

⁵ The Help, Kathryn Stockett.

⁶ Gone Girl, Gillian Flynn.

⁷ The Da Vinci Code, Dan Brown.

Ясно, что, написав такую книгу, автор внакладе не останется. Не останутся без прибыли и агенты, обнаружившие ее, и редакторы. И книготорговцы тоже – ведь они держатся на плаву и не уходят из бизнеса лишь благодаря немногим хорошо продающимся произведениям.

Конечно, мы сейчас говорим о материальной выгоде. Допустим, вы давно рассказывали друзьям, что у вас внутри сидит роман и просится наружу, – и вот вы в самом деле выплеснули его на бумагу и получили за это семи- или даже восьмизначный аванс. Такое бывает – с немногими авторами, но все же бывает. Можно окружать богемным флером фигуру бедного писателя, скрипящего карандашиком в блокнотике, но куда приятней думать о том, как твоя книга появится в каждой квартире – на полке или на тумбочке у кровати – и на iPad или Kindle у каждого пассажира в каждом вагоне метро! И все это – на многих языках, по всему миру!

Такой блеск, такое драматическое восхождение к славе характерны для немногих «книг года». В книжном мире они играют роль дорогих особняков, шикарных машин и бриллиантовых диадем. Если автор попал в список и умудрился там зацепиться, его ждут почет, уважение, зависть и злые слова. Автора могут пригласить в жюри литературного конкурса или давать ему на рецензию другие книги. Возможно, у него купят права на экранизацию. О нем будут говорить.

Правда, было бы здорово, если бы такой успех можно было предсказать?

Белые лебеди

Смелая гипотеза авторов сего труда состоит в том, что романы попадают в списки бестселлеров New York Times отнюдь не случайным образом и что пути книжного рынка вовсе не так неисповедимы, как принято считать. Бестселлеры, независимо от жанра, похожи друг на друга по большому количеству параметров (впрочем, не бросающихся в глаза), и, изучая эти параметры, можно лучше понять, что мы читаем и почему. Более того, можно разработать алгоритмы для выявления носителей «ДНК бестселлера» среди новых, еще не опубликованных книг.

У профессионалов книжной отрасли бытует мнение, что успех полностью определяется громким именем, затратами на рекламу или широкомасштабными мероприятиями по продвижению книги. Конечно, все это имеет значение, но наши результаты доказывают: далеко не все определяется шумихой, поднятой вокруг книги. Этот вывод должен обрадовать писателей, которые трудятся, оттачивая свое мастерство. Результаты нашей пятилетней работы дают основания предположить, что бестселлером книгу делают правильные слова, расположенные в правильном порядке. Это значит, что даже в самой интересной истории, связанной со списком бестселлеров New York Times, речь идет исключительно об авторской рукописи как таковой – только о черных чернилах на белой бумаге, и больше ни о чем.

Используя компьютерную систему, которая читает книги, распознает в них определенные характеристики и просеивает тысячи таких характеристик в тысячах текстов, мы открыли, что существуют удивительные сочетания параметров, свойственные книгам, которые с наибольшей вероятностью будут пользоваться успехом на рынке. Эти параметры могут многое поведать о нас, читателях, и о том, как мы читаем. На этих страницах мы расскажем, как и зачем мы построили такую модель и как она обнаружила, что в 80–90 % случаев очень легко различить бестселлер среди других книг в использованном нами корпусе текстов. Наши компьютеры выявили 80 % книг, побывавших в списке New York Times за последние 30 лет, и назвали их вероятными кандидатами в бестселлеры. Более того, наша система обрабатывала каждую книгу так, словно это новая, никому не известная рукопись, а затем не просто выносила один из двух вердиктов – «вероятно, бестселлер» или «скорее всего, нет», – но и вычисляла индекс, указывающий, каковы шансы книги обрести популярность. Эти индексы сами по себе чрезвычайно интересны, и мы не только расскажем, как они устроены, но и объясним, что держит читателя как магнитом, заставляя переворачивать страницы одну за другой.

Рассмотрим индексы, вычисленные нашей системой. Она предсказала, что «Инферно»⁸ Дэна Брауна окажется бестселлером с вероятностью 95,7 %. «“Линкольн” для адвоката»⁹ Майкла Коннелли – с вероятностью 99,2 %. Обе побывали на первом месте в списке New York Times – то есть, пожалуй, на самой престижной позиции, какую может завоевывать книга. Конечно, Браун и Коннелли – известнейшие авторы с прочным положением на литературном Олимпе. Но наша система не знает имен авторов и с той же уверенностью раздает оценки трудам неизвестных писателей. Например, «Вязание по пятницам»¹⁰, дебютный роман Кейт Джейкобс, получил оценку 98,9 %. «Счастливые девочки не умирают»¹¹, совсем другой по характеру дебютный роман Джессики Нолл¹², получил индекс успеха 99,9 % –

⁸ Inferno, Dan Brown.

⁹ The Lincoln Lawyer, Michael Connelly.

¹⁰ The Friday Night Knitting Club, Kate Jacobs.

¹¹ Luckiest Girl Alive, Jessica Knoll.

¹² В русском издании фамилия автора передана как «Кнолл». (примеч. перев.)

исключительно на основе текста. Книги Джейкобс и Нолл попали в список бестселлеров и задержались там на много недель. «Марсианин»¹³ (еще до того, как Мэтт Деймон пообещал сыграть главную роль в экранизации) получил 93,4 %. Можно привести примеры из самых разных жанров: «Телефонный звонок с небес»¹⁴ (произведение из разряда «литература о духовности») Митча Элбома – 99,2 %; «Искусство поля»¹⁵, литературный дебют Чеда Харбаха, – 93,3 %; «Обнаженная для тебя»¹⁶, эротический любовный роман Сильвии Дэй, – 91,2 %.

Эти индексы, измеряющие «бестселлерный потенциал» книги, приводят в восторг одних людей, возмущают других и у многих вызывают подозрение. Последних можно понять: эти индексы не укладываются в голове, они как будто ниспровергают существующий порядок вещей. Некоторым ветеранам книжной индустрии они кажутся абсурдными. Но они могут совершить переворот в книгоиздании, и уж точно благодаря им вы будете совсем по-другому воспринимать следующий бестселлер, который попадет вам в руки.

Здесь мы должны прояснить важный момент. Ни одна из книг, упоминающихся в данной работе, никогда никем не приобреталась на основе индексов, вычисленных нашей системой. Кроме того, за исключением данной публикации, мы никогда не доводили эти индексы до сведения какого-либо книжного агента или издательства. Следует также оговорить, что все полученные нами численные показатели специфичны для конкретного корпуса текстов, на котором проводилось исследование, – при составлении этого корпуса мы добивались максимального сходства с ассортиментом, который вы увидите, зайдя в любой крупный книжный магазин. Книжные агенты и издатели знают свое дело – они трудятся, наполняя полки магазинов книгами, и нам всегда есть что почитать. Иные деятели издательской нивы обладают чем-то вроде дара царя Мидаса – обращать своим прикосновением рукописи в золото. Но не забывайте, что средняя доля бестселлеров среди опубликованных книг – меньше половины процента. Иными словами, чтобы наконец выиграть, приходится долго рисковать. Заметьте также, что год за годом в списках фигурируют одни и те же чрезвычайно успешные авторы, десятилетиями возглавляющие рейтинги. Стивену Кингу шестьдесят восемь лет. Джеймсу Паттерсону шестьдесят восемь лет. Даниэле Стил шестьдесят восемь лет. Конечно, поклонники рады каждой новой книге любимого автора, но это значит, что книжный мир еще не открыл следующее поколение писателей, которые так же стабильно будут поставлять бестселлеры на рынок следующие тридцать или сорок лет. Кроме того, хотя ежегодно тысячи рукописей отвергаются и тысячи публикуются, в 2014 году издателям так и не удалось найти бешено успешный бестселлер («Татуировка дракона», «Оттенки серого» и «Исчезнувшая» были хитами предшествующих годов). В 2012 году ни одно из опубликованных литературных произведений не впечатлило комитет по присуждению Пулитцеровской премии. Почему?

Если верить расхожему мнению, любой бестселлер – аномалия. Удачная флуктуация на рынке. Интересный мутант. Черный лебедь. Но если это правда, то, единожды найдя писателя, способного выдавать бестселлеры, надо вкладывать все деньги в него? Зачем рисковать миллионами, ставя их на никому не известного двадцатилетнего юношу вместо Стивена Кинга? Ведь нет способа узнать, стоит ли новый автор всех тех денег, которые вы собираетесь в него вложить, и поможет ли он вам сорвать куш.

У жаргона книгоиздательской индустрии очень много общего с жаргоном игорного бизнеса, и это не случайно. На совещаниях по вопросам покупки рукописей часто употреб-

¹³ The Martian, Andy Weir.

¹⁴ The First Phone Call from Heaven, Mitch Albom.

¹⁵ The Art of Fielding, Chad Harbach.

¹⁶ Bared to You, Sylvia Day.

ляется выражение «поставить на» такого-то автора. Если два издательства наперебой торгуются за одну и ту же книгу, одно из них может «пойти ва-банк» и потратить на нее весь свой сезонный бюджет. Это чрезвычайно интересное занятие, и риск, на который идут издательства, безусловно, обоснован, но все же это – азартная игра, практически как в казино. Прежде чем рукопись «Гарри Поттера»¹⁷ приняли в издательстве Bloomsbury, Дж. К. Роулинг получила 12 отказов в других местах и совет «не бросать основную работу». Сегодня стоимость бренда «Гарри Поттер» оценивается в сумму около 15 миллиардов долларов. Джону Гришэму отказали не менее 16 разных издательств; с тех пор Гришэм создал более двенадцати бестселлеров года, опередивших по уровню продаж все остальные книги¹⁸. Джеймс Паттерсон несколько раз пытался опубликовать свои рукописи и всюду встречал отказы; в 2010 году три его книги разошлись общим тиражом 3,5 миллиона. Кэтрин Стокетт получила отказы от шестидесяти агентов, прежде чем нашла одного, который согласился представлять ее «Прислугу»; впоследствии этот роман пробыл в списке бестселлеров New York Times на протяжении ста недель. Без сомнения, множество подобных книг сейчас пылится в отвалах так называемого «самотёка» в конторах издателей, которыми усеяны Нью-Йорк и Лондон.

Если вы имеете хотя бы косвенное отношение к книжному миру, наверняка у вас есть знакомый или знакомый знакомого, который годами вставал в четыре часа утра, чтобы успеть написать хоть немножко до работы. У него был абсолютно беспримысленный сюжет, он чувствовал, что музы невидимо реют над его головой... А потом он разослал рукопись во все издательства на Манхэттене и не получил в ответ ничего, кроме стандартных отписок-отказов.

Эти знакомые знакомых вовсе не одиноки. Один издатель, прочитав рукопись «Шпиона, пришедшего с холода»¹⁹, заявил Джону ле Карре, что как писатель он безнадежен. «Повелитель мух»²⁰ Уильяма Голдинга получил 21 отказ. Ознакомившись с рукописью Джека Керуака «В дороге»²¹, ныне культовой книгой поколения, издатель заявил: «Я вообще не понял, о чем это». Урсуле ле Гуин сказали, что ее «невозможно читать». Впоследствии «совершенно нечитабельная» книга ле Гуин получила две крупные литературные награды. Даже «Скотный двор»²² Оруэлла сочли негодным к публикации – причем не кто иной, как Т. С. Элиот. Великий поэт решил, что эта выдающаяся политическая аллегория, актуальная для всех времен и народов, «недостаточно убедительна».

Публиковать или не публиковать? Вот в чем вопрос. Чтобы предсказать книге большой успех, нужно проанализировать душевные движения и чувствительные места сотен тысяч человек. Это непростая задача, и часто отказы выглядят вполне обоснованными. Например, мы опросили американских издателей, отвергнувших «Девушку с татуировкой дракона», и они сказали, что, по их мнению, американской публике неинтересны все эти детали шведской политики. Лисбет Саландер, на их взгляд, слишком агрессивна и непредсказуема для центрального женского персонажа. Они сочли, что широкие массы читателей отпугнет чудовищная сцена анального изнасилования и Лисбет-мстительница, вооруженная татуировочными иглами. Все эти соображения кажутся весьма разумными.

Поэтому неудивительно, что издатели в минуту откровенности заявляют: пророчить книге успех – все равно что определять направление ветра, поклонив палец, или смотреть в хрустальный шар, который, похоже, спрятан под столом у самых удачливых агентов

¹⁷ Harry Potter, J. K. Rowling.

¹⁸ Журнал книжной индустрии Publishers Weekly публикует список книг года – чемпионов по уровню продаж. (Примеч. авторов.)

¹⁹ The Spy Who Came in from the Cold, John le Carré.

²⁰ Lord of the Flies, William Golding.

²¹ On the Road, Jack Kerouac.

²² Animal Farm, George Orwell.

и издателей. За исключением случаев, когда автор уже популярен – например Нора Робертс или Джеймс Паттерсон, – делать такие прогнозы – все равно что стрелять в цель с завязанными глазами. Иногда помогают обстоятельства: например, если автор – голливудская звезда, а книга посвящена ее сексуальным похождениям. Но бывало и так, что дело казалось верным, за большим авансом следовали огромные тиражи – и их приходилось пускать под нож. Вкусы публики переменчивы.

Конечно, все агенты и издатели стараются понять, что движет коммерческим книжным рынком – в области чтения для масс, на которую ориентирована, к примеру, писательница Патрисия Корнуэлл, ветеран франшизы в литературе, или в более скромном, но все же удовлетворительном масштабе тиражей, характерных для успешных современных литераторов. Известен анекдот о том, как (ныне бывшего) генерального директора одного из крупнейших нью-йоркских издательств попросили предсказать заглавие книги, которая точно станет бестселлером. «Собака врача президента Линкольна», – ответил он. И впрямь, имя всеми уважаемого президента в сочетании со словами «врач» (наша одержимость собственным здоровьем) и «собака» (любимое домашнее животное американцев) – верный рецепт успеха.

Это, конечно, шутка, но, оказывается, под таким названием вышла даже не одна, а две книги! И обе провалились. Преподаватель литературы и писатель Джон Сазерленд, выпустивший два исследования на тему бестселлеров, завершил одно из них словами: «В конечном счете бестселлер отличается от всех остальных книг тем, что он лучше продается. И больше ничем». Далее он добавляет, что «искать какие-то значимые закономерности, тенденции, сходство [между книгами, завоевавшими популярность] бессмысленно, а результаты этого поиска повергают в недоумение». Вердикт Сазерленда кажется обоснованным, справедливым и окончательным. Точнее, казался таковым до появления компьютеров, которые научились читать тексты и вычислять секретный рецепт, позволяющий попасть в список бестселлеров New York Times.

Из любви к книгам

Вернемся к писателям, которые получили множество отказов, но в конце концов прославились. Наша система предсказала Дж. К. Роулинг успех с вероятностью 95 %. Джон Гришэму – с вероятностью 94 %. Паттерсону – 99,9 %. История позволила нам убедиться в точности этих предсказаний. А вот с книгой Кэтрин Стокетт «Прислуга» наша система ошиблась. «Прислуга» оказалась в числе примерно 15 % книг, которые повергли компьютер в недоумение. Он дал ей лишь 50 % вероятности стать бестселлером. В следующих главах мы расскажем о сложностях, с которыми столкнулись, моделируя работу редактора с помощью компьютера. Пока достаточно будет сказать, что система проводит чрезвычайно глубокий анализ; прочитав книгу Стокетт, она заключила, что стиль в целом хорош и понравится американским читателям, что тематика тоже в целом хороша, но использование эмоционально нагруженного языка и в особенности глаголов отличает «Прислугу» от книг, с большой вероятностью попадающих в список бестселлеров. Речь идет о книге, которая привлекла внимание рецензентов тем, что белая писательница активно использовала имитацию диалекта чернокожих персонажей. Мнения рецензентов о правомочности такого авторского решения разделились; и оказалось, что наша система в точности предсказала это разделение во всех источниках, от New York Times до сайта goodreads.com.

Вероятно, вы спросите: зачем пытаться с помощью компьютера смоделировать работу, которую и так выполняют хорошие редакторы? Возможно, будь наша система уже внедрена, творения Дж. К. Роулинг опубликовали бы раньше. Возможно, Джон Гришэм получил бы аванс побольше за свою первую книгу «Пора убивать»²³. Но в конце концов эти писатели добились славы. Редакторы были не уверены в успехе «Прислуги». Наша система – тоже. Какая тогда от нее польза?

Дело в том, что мы ищем компоненты успеха не только для того, чтобы получить от этого материальную выгоду. Конечно, очень интересно, что наша система выделила из общей массы Дж. К. Роулинг, Лиану Мориарти (вероятность успеха 99,6 %) или Джонатана Франзена (98,5 %). Мы считаем, что публичное обсуждение сотрудничества человека и компьютера важно, особенно в том, что касается творческой деятельности. Но поиски новых подающих надежды рукописей в пошатнувшейся книгоиздательской отрасли проводятся с целью не только удержать эту отрасль на плаву, но и поддержать ее разнообразие. Безусловно, нами движет желание выявить и объяснить скрытые закономерности человеческой культуры. Но, если подходить к делу более практично, мы хотим помочь издательствам в поиске новых авторов, чтобы направить часть бюджета, отведенную Кингу, Паттерсону или Стил, на молодых писателей, которые придут им на смену. Мы хотим, чтобы писатели – как опытные, так и молодые – узнали больше о своем ремесле и получили информационную поддержку. Мы хотим, чтобы книги людей, не имеющих нужных контактов в Нью-Йорке, нашли своего читателя. Нашей системе все равно, публиковался ли этот автор раньше, есть ли у него диплом бакалавра искусств, мужчина это или женщина, латиноамериканского или китайского происхождения, красив автор или некрасив и сколько ему лет – двадцать пять или семьдесят. Поэтому наша работа еще и уравнивает шансы на успех. Может быть, в один прекрасный день знакомый вашего знакомого отправит рукопись в издательство, получит индекс 80 % и ему выдадут хороший аванс, так что он сможет наконец уволиться с работы, и ему больше не нужно будет вскакивать в четыре часа утра, чтобы выкроить время для писательских занятий.

²³ A Time to Kill, John Grisham.

Когда мы пишем о романах, фигурирующих в самом знаменитом и уважаемом списке – еженедельном списке бестселлеров New York Times, – мы тем самым кричим во все горло, призывая равнодушных читателей – как профессионалов книжного мира, так и любителей – поговорить о книгах для масс²⁴. Над бестселлерами чаще потешаются, чем рассматривают как серьезные литературные произведения или хотя бы плоды писательского мастерства. Однако, игнорируя их, мы упускаем большую часть современной культуры и истории чтения. Помимо того что авторы бестселлеров приносят в экономику миллионы долларов, их заслуга – в том, что они заставляют нас *читать*. Они заставляют нас воображать, обсуждать, думать, чувствовать, сопереживать другим. Они помогают фантазировать, подглядывать за чужой жизнью, бежать от действительности. Романисты из списка New York Times дают пищу для разговоров и споров о литературе по всей стране – в барах, в поездах, дома за ужином. Именно на этих писателей смотрят, желая узнать, куда движется наша литература. Именно к ним обращаются, чтобы лучше понять жизнь. К ним прибегают, чтобы развить вкус, выработать мнения и научиться их выражать. Мы будем только рады, если сможем рассказать читателям что-то новое об их любимых писателях.

Вероятно, вы уже поняли: авторы, то есть мы, до того страстно преданы книгам и чтению, что отдали в общей сложности пятьдесят лет изучению и преподаванию нарратива и еще несколько – закупкам и продажам книг для крупнейших издателей страны. Мы выступали в роли наставников при написании романов и отстаивали свое право любить и ненавидеть разные (а иногда – одни и те же) произведения. Мы расхваливали издателям книги самых разных жанров, стараясь добиться их публикации. Мы (иногда втайне) помогали своим лучшим студентам и своим друзьям, мечтающим о карьере литератора, писать письма их родителям, супругам и будущим издателям, объясняя, почему автор письма просто обязан был бросить устроенную жизнь, учебу на врача – и отправиться на поиски колдовского зелья, сулящего то экстаз, то депрессию, – на поиски жизни в словах и историях. Не будет преувеличением сказать, что мы «купились» на ту свободу, то высшее (высочайшее) образование, которое получаешь, читая и создавая книги. Мы в первую очередь читатели, а потом уже писатели. Учитывая наше преклонение перед книгами, нет ничего удивительного в том, что мы обратились за помощью к компьютеру.

²⁴ Списки бестселлеров в разных странах немного различаются, но мы будем говорить исключительно о списке NYT, поскольку он неизменно играет роль культурного стандарта и наиболее важен для книжной отрасли. В интернете легко найти критику методов, по которым составляются подобные списки, поскольку ни один из них не покрывает абсолютно все продажи из всех источников. Список NYT учитывает примерно 75 % всех продаж и строится на данных Nielsen BookScan. (Примеч. авторов.)

Откуда мы взяли

Вероятно, больше всего открытию «бестселлерометра» (прозвище нашей системы) удивились мы сами. Честно говоря, наши исследования начались лишь потому, что нас «нутром» потянуло в эту сторону. Четыре года ежедневного совместного труда принесли результаты, каких ни один из нас не ожидал, несмотря на наш разнообразный опыт: у Джоди – в книгоиздании и современной литературе, а у Мэтта – в литературе и модной нынче дисциплине, именуемой «цифровые гуманитарные науки».

Все началось с того, что Джоди ушла с должности редактора-рецензента в издательстве Penguin Books, чтобы пойти в аспирантуру по английскому языку в Стэнфордском университете. От работы в книжной индустрии у Джоди остался жгучий вопрос, на который она так и не получила ответа: отчего романы становятся бестселлерами? А также связанные с этим скрытые, еще более интересные вопросы: почему читатели читают? Какую роль выполняет чтение в современной культуре?

Когда Джоди только пришла работать в Penguin, ее определили в отдел продаж. Иногда в обеденный перерыв она заглядывала в ближайший книжный магазин, чтобы проверить, выполняет ли он свои обязательства по продвижению книг – действительно ли те размещены на бойких местах в соответствии с суммами, потраченными на маркетинг. Очень часто, и это отнюдь не секрет, издатели платят за то, чтобы их «хиты» выставлялись на виду. Руководство магазина берет деньги и обязуется расположить книгу, например, в первом ряду на первом столе или на полке так, чтобы обложка была полностью видна и смотрела прямо на покупателя. Считается, что такое стратегическое размещение способствует продажам. В то время по спискам бестселлеров шествовал победным маршем «Код да Винчи». Неделю за неделей Джоди видела большую синюю цифру «1» и убеждалась в том, что роман Дэна Брауна продолжает покорять мир.

Через несколько месяцев стало очевидно: сколько бы издатели ни платили за размещение клонов Дэна Брауна на видных местах, им не светило попасть в одну лигу с «Кодом да Винчи». Его феноменальный успех объяснялся не суммами, затраченными на рекламу и маркетинг, а чем-то другим. Никакие расходы не гарантируют столь длительного воздействия на умы читателей всего мира и тем более не обеспечат тиража 80 миллионов. Такой успех не может объясняться только рекламной шумихой. Значит, есть что-то кроме маркетинга, связанное именно с этими словами на этих страницах.

Конечно, глупо было бы утверждать, что маркетинг и реклама совсем не влияют на продажи. Разумеется, влияют. Должна быть некая корреляция, объясняющая тот факт, что 80 % бестселлеров принадлежат пяти крупнейшим издательствам, обладающим значительными средствами на рекламу и маркетинг. Но было бы так же глупо утверждать, что любой доллар, затраченный на маркетинг, всегда дает одинаковую отдачу: слишком много мы видели и огромных сумм, потраченных безрезультатно, и публикаций силами автора, имевших бешеный успех. «Пятьдесят оттенков серого» сначала появилась в электронном виде, а также распространялась по модели «печать по требованию» в мягкой обложке – издательством, у которого вообще не было ни гроша на рекламу. Чтобы напечатать свою книгу «Хижина»²⁵, Уильям Пол Янг воспользовался кредитными карточками, а на маркетинг потратил всего 300 долларов – во столько обошлась ему разработка веб-сайта. На сегодня эта книга разошлась десятиллионным тиражом. Вот примеры других бестселлеров, непохожих на «Хижину», но также пришедших к успеху и похвале критиков через нетрадиционные каналы распро-

²⁵ The Shack, William P. Young.

странения: экспериментальный онлайн-роман Марка Данилевского «Дом листьев»²⁶ и книга Криса Уэра «Джимми Корриган, самый умный ребенок на Земле»²⁷, первоначально опубликованная на средства автора, а ныне один из самых популярных «графических романов» новой волны. Примеров таких можно набрать много – достаточно, чтобы понять: на вопрос «почему одни книги читают миллионы, а другие в лучшем случае десятки людей?» ответить «маркетинг» – значит в лучшем случае ответить наугад.

Джоди пришла со своим исследовательским вопросом к Мэтту, который в то время был лектором в Стэнфордском университете и сооснователем Стэнфордской литературной мастерской. Объединив усилия, Джоди и Мэтт поняли, что существует другой ответ, не просто «маркетинг». В 2008 году Мэтт как раз закончил свою часть работы над исследованием авторского стиля священного текста «Книга мормона» (The Book of Mormon) с помощью компьютера. Эта работа вызвала много споров. Результаты компьютерного анализа авторского стиля «Книги» дали возможность предположить, что ее действительно писал не один человек, то есть выдвигавшиеся ранее гипотезы, возможно, были верны. Исследование также подкрепило одну определенную теорию происхождения «Книги», которую церковь мормонов отвергла как ложную. Результаты анализа не давали сделать однозначное заключение, но отклики на статью, в том числе весьма интересное опровержение, сделанное мормонскими богословами из Университета имени Бригама Янга, показали, каким революционным орудием может быть компьютерный анализ текста.

Эта работа над определением авторства и «стилеметрией» убедила Мэтта в том, что компьютеры видят многое и помогают нам обнаружить в тексте то, чего мы обычно не замечаем. Дальнейшие исследования позволили Мэтту создать компьютерную программу, которая с точностью 82 % определяла пол автора всего лишь по его или ее манере использовать самые простые слова английского языка, такие как *the*²⁸ и *of*²⁹. Мэтт не первым заметил, что у авторов-мужчин и авторов-женщин – разные стилевые привычки, но в своей работе он сосредоточился исключительно на романах XIX века. Затем он обнаружил, что, опираясь лишь на использование слова *the*, компьютер может достаточно уверенно определить, английский это писатель или американский.

Реакцию Джоди можно примерно описать словами «Ну и что?». Конечно, очень интересно, что компьютер может отличить британца от американца, но это – искусственная задача, решение которой никому не интересно. Чтобы убедиться в могуществе компьютерного метода, Джоди хотела увидеть, как с помощью машин решаются настоящие литературоведческие задачи. Мэтт же, в свою очередь, был не в восторге от того, что Джоди заиклена на современных бестселлерах. Он считал их одноразовыми поделками – прочел и забыл. Его надо было убедить, что в них скрыто золото, которое оправдает труды старателя.

Все это было несколько лет назад. С тех пор мы объединили свои силы и бросили их на доказательство следующей гипотезы: все бестселлеры характеризуются определенным набором скрытых сигналов, неким тайным кодом. Вместо того чтобы угадывать, какая книга окажется популярной, мы решили довериться тому, что читатели уже поняли, сами того не заметив. Список бестселлеров лишь на первый взгляд кажется случайной подборкой непохожих друг на друга книг – на самом деле это еженедельно обновляемый набор сигналов, на зов которых идет читатель, и этот набор постоянно уточняется посредством коллективного голосования. Учит ли чему-нибудь глас читающего народа? Смогут ли наши компьютеры

²⁶ House of Leaves, Mark Z. Danielewski.

²⁷ Jimmy Corrigan: The Smartest Kid on Earth, Chris Ware.

²⁸ Определенный артикль в английском языке.

²⁹ Предлог, в русском языке ему могут соответствовать разные предлоги, такие как «из», «от» и др. (Примеч. перев.)

выделить сигнал из этого шума? Есть ли у этих броских обложек – будь то высокоинтеллектуальные романы из университетской программы или чтиво для пляжа – что-нибудь общее?

Если ответ «да», то мы сможем узнать кое-что об анатомии успеха. Мы сможем даже опровергнуть распространенную в книжном мире теорию и создать алгоритм, позволяющий предсказать, станет ли та или иная книга бестселлером.

Итак, мы начали обучать компьютер чтению.

Машинное чтение

Конечно, компьютеры на самом деле не умеют читать, во всяком случае – в том смысле, в котором вы читаете эту страницу. Но можно также сказать, что они умеют читать в том смысле, в каком они умеют делать все остальное – они «считывают» текст, то есть принимают его в качестве входной информации. Затем они проводят разбор, то есть разбивают текст на единицы, которые мы, люди, считаем осмысленными: буквы, запятые, слова, предложения, главы и т. д. На этой стадии компьютер в значительной мере подражает человеку, и чем совершенней была обучающая программа, тем точнее подражание. Разница между читателем-человеком и читателем-машиной заключается в том, что человек понимает значение читаемого текста, а компьютер – нет. Ирония здесь, однако, в том, что компьютерный способ чтения позволяет нам взглянуть в текст пристальней любого самого опытного литературного критика. Все потому, что компьютеры – мастера различать детали в таком мелком масштабе, какой человеку и не снился.

Вспомним вопрос, с которого началось наше исследование. Можно ли предсказать заранее, что та или иная книга станет бестселлером? Чтобы предсказывать, нужно уметь обнаруживать закономерности. Если вы не умеете гадать на картах, то вся ваша надежда – на способность распознавать определенные характеристики текста. Обычно такими вещами занимаются литературные критики и ученые-литературоведы. Джозеф Кэмпбелл, великий специалист по мифологии, всю жизнь читал мифы и легенды разных народов мира и научился видеть в них схожие мотивы. Он был мастером по распознаванию повторяющихся узоров. Но даже если человек всецело предан своему делу, объем текстов, которые он может переработать, конечен, и столь же ограничено внимание, с которым он исследует каждую отдельную книгу. Следовательно, в обоих направлениях мы сталкиваемся с ограничениями масштаба и как будто вынуждены одним глазом смотреть в микроскоп, а другим – в телескоп.

Еще один ученый, обладающий завидным трудолюбием, – Кристофер Букер – потратил тридцать лет на чтение сотен книг, чтобы выдвинуть следующую теорию: все литературные произведения и вообще все истории соответствуют семи основным сюжетам. Допустим, он трудился сорок лет и прочел тысячу книг. Допустим также, что из прочитанного он запомнил гораздо больше, чем средний человек. Но кластер компьютеров, обученных надлежащим образом, может прочитать несколько тысяч романов и проанализировать все эти тексты по тысяче параметров примерно за один день. При этом компьютеры обладают невероятной способностью замечать то, что мы, люди, принимаем как должное или вообще игнорируем.

Вот лишь один пример. Как читатели – особенно читатели, обученные пристально вглядываться в текст, – мы отмечаем прилагательные, которые использует писатель. Но, скорее всего, не обратим внимания на соотношение числа существительных и прилагательных, то есть на то, как часто автор использует прилагательное для описания существительного. А вот компьютер прекрасно приспособлен для того, чтобы замечать именно такие вещи – и они важны, поскольку определяют авторский стиль и манеру описаний. Компьютер может также вычислить это соотношение для данной книги и сравнить с тем же показателем у тысячи других книг. Если окажется, что у бестселлеров этот показатель в среднем выше (или ниже), значит, данный параметр имеет какое-то значение.

Предлагаю вам проделать небольшой эксперимент, когда будете в следующий раз выбирать книгу для чтения. Вместо того чтобы последовать рекомендации друга или выбрать книгу знакомого автора (или в любимом жанре), попробуйте взять недельный список бестселлеров NYT и прочитать их все по очереди. Можете это сделать вместе с книжным клубом, в котором состоите, или всей группой на курсе английской литературы. Если

вы будете читать внимательно, то скоро уподобитесь нашим компьютерам и начнете замечать неожиданные параллели между литературными шедеврами и массовым чтением, книгами «для мужчин» и «для женщин», Паттерсонами и пулитцеровскими лауреатами и т. п. Некоторые общие мотивы вас удивят. Например, вы зададитесь вопросом: почему так много книг, в которых главной героине двадцать восемь лет? Имеет ли это какое-то значение? Вы спросите себя, сознательно ли авторы помещают первую любовную сцену на страницу 200, если в романе 400 страниц, или на страницу 110, если страниц 220. Если да, то *почему*? Вы будете спорить с друзьями о том, определяется ли низкое качество романа, в остальном очень хорошего, отсутствием удовлетворительной развязки. Возможно, вы даже придете к выводу: бестселлеры всех категорий имеют столько скрытых общих черт, что фактически представляют собой отдельный жанр.

Интересно здесь то, как остро и притом незаметно для себя люди во время чтения реагируют на подобные характеристики книги. Специалисты из новой, стремительно развивающейся области «литературной неврологии» используют ЯМР-томографию для исследования мозга человека во время чтения. Эксперимент направлен на выявление того, что замечает читатель. Когнитивная психология подходит к вопросу «как читает читатель?» совершенно под иным углом, чем мы, но оба подхода базируются на одном постулате: отклик читателя на литературный текст зависит от того, из каких слов состоит этот текст, в каком порядке они выстроены и как организованы в предложения. И какую реакцию они вызывают.

Значит, использование компьютерного чтения никак не противоречит обычному подходу литературных критиков. Более того, рассмотрение характеристик текста «под лупой», их извлечение и последующий анализ очень часто используются в традиционных критических подходах и позволяют получить представление о тексте на уровне, который был невозможен раньше.

Компьютер можно научить читать текст и извлекать из него информацию множеством способов³⁰. Алгоритмы и программы, созданные нами для данного исследования, направлены на обработку текстов и сбор детальной информации (о неповторимом стиле каждой книги, ее тематике, эмоциональных подъемах и спадах в повествовании, персонажах, местах действия). Также эти программы обрабатывают огромное количество скучных на первый взгляд лингвистических данных, которые не так просто перевести на язык привычных концепций стиля и сюжета. При обработке художественных произведений большого объема, какие обычно изучаются на курсах писательского мастерства и в пособиях для начинающего романиста (тема, сюжет, стиль и т. д.), обычно анализируются сотни параметров. Например, чтобы дать формальное описание стиля, мы использовали несколько сотен переменных: как часто у автора попадаются слова типа *a*³¹, *the*, *in*³², *she*³³; как часто автор использует точки и восклицательные знаки; как часто – наречия и какие именно. Эти мелкие детали так много говорят читателю! Посмотрите, как важны местоимения для создания эффекта в знаменитой фразе из романа Шарлотты Бронте «Джейн Эйр»³⁴: «Читатель, я вышла за него замуж». Компьютер обращает внимание на слово «него», на то, как часто мы слышим про «него» и как часто это местоимение встречается рядом с самым главным местоимением – «я» рассказчицы. Компьютер замечает, что «я» и «он» попадают все ближе друг к другу во все большем количестве предложений, и их разделяет все меньше и меньше слов. Конечно, читатель тоже следит за этим. Разве главная развязка множества книг не заключается в том, чтобы

³⁰ Общее описание нашего метода содержится в послесловии. (Примеч. авторов.)

³¹ Неопределенный артикль в английском языке.

³² Предлог, в русском языке ему могут соответствовать разные предлоги, такие как «в», «через» и др. (Примеч. перев.)

³³ Она (англ.).

³⁴ Jane Eyre, Charlotte Brontë.

«я» и «он» пошли по страницам рука об руку, разделенные лишь немногими важнейшими словами, такими как «вышла за»? Очень часто мы, не в силах оторваться от книги, продолжаем листать ее исключительно ради этого.

Вопросительные и восклицательные знаки тоже многое могут поведать. Возможно, вы помните, как писали сочинения в школе и как учитель советовал вам не злоупотреблять восклицаниями. Если каждая фраза визжит от возбуждения («О боже!»), или командует («Руки вверх!»), или кричит от боли («Ой!»), или жутко грохочет в подвале среди ночи («Бу-бух!»), то у читателя, не ровен час, может случиться сердечный приступ. Если текст изобилует восклицательными знаками, это кое-что говорит как о его возможном содержании и уровне накала страстей, так и о профессионализме автора. Аналогичным образом обилие вопросительных знаков указывает на то, что этот текст – диалог, а бесконечные страницы текста без единого вопросительного знака могут замедлить динамику повествования и притупить интерес читателя. Эти черты индивидуального стиля мы обсуждаем в 4-й главе.

Мы начали с извлечения свыше 20 тысяч параметров – имейте в виду, что количество восклицательных знаков и частота употребления слова «него» это лишь два из них, – и изучили их все. Некоторые были приметами стиля, другие помогали что-то узнать о сюжете и месте действия, а третьи поясняли, о чем, собственно, книга. Оказалось, что не все эти параметры помогают отличить книгу, набравшую миллионы читателей, от той, которая, несмотря на все свои уникальные достоинства, провалилась. Например, выяснилось, что использование чисел в заглавии – «911», «1984», «867–5309», «\$ 1 000 000» – никак не стимулирует продажи. Аналогичным образом мы потратили много времени, обучая компьютер распознавать место действия (например, действие книги «Дьявол носит Prada»³⁵ происходит целиком в Нью-Йорке, а действие «Исчезнувшей» начинается в Нью-Йорке, а потом переносится в штат Миссури), но оказалось, что, за малыми исключениями, географическое положение героев совершенно не влияет на популярность у читателей. В Нью-Йорке происходит действие как бестселлеров, так и книг, не имевших никакого успеха. Литературные мегахиты, действие в которых разворачивается в этом городе, – взять хотя бы «Обнаженную для тебя» Сильвии Дэй, «Костры амбиций»³⁶ Тома Вулфа, «По-быстрому»³⁷ Джеймса Паттерсона, «Жутко громко и запредельно близко»³⁸ Джонатана Сафрана Фоера – явно (намеренно со стороны автора или случайно) обладают какими-то иными, более важными «генами бестселлера».

В конце концов мы отфильтровали свой список параметров, оставив от 20 тысяч примерно 2800, которые помогают отличить популярные у читателей книги от предназначенных для – будем честны – узкой читательской ниши. Сначала мы научили компьютеры читать книги и извлекать из них нужные данные, а затем проанализировали эти данные, используя другой набор компьютерных программ – для обнаружения и исследования скрытых закономерностей. В этой фазе анализа мы использовали процесс, весьма метко называемый машинным обучением. При анализе текстов бывает нужно сортировать или классифицировать их по сходству и различию. Например, мы хотим отличать спам от обычных сообщений в электронной почте. Поскольку у спамерских посланий много общего – искаженное написание слов, частое упоминание виагры и т. д., – можно написать программу, которая будет определять, с какой вероятностью данное сообщение окажется спамом. Сортировка романов на бестселлеры и небестселлеры очень похожа на действия, выполняемые спам-фильтром в электронной почте. Допустим, у нас есть новая книга, которую мы видим в первый раз, и

³⁵ The Devil Wears Prada, Lauren Weisberger.

³⁶ The Bonfire of the Vanities, Tom Wolfe.

³⁷ The Quickie, James Patterson.

³⁸ Extremely Loud & Incredibly Close, Jonathan Safran Foer.

мы хотим понять, может ли она стать бестселлером. Если у нас уже есть куча бестселлеров («не спам») и куча книг, которые бестселлерами не стали («спам»), мы можем ввести их все в компьютер и научить его отличать книги одной категории от другой по определенным параметрам. Именно так мы и поступили. Более того, мы проделали это тремя разными способами, усреднили результаты и обнаружили, что в 80 % случаев наша система правильно определяла, какой текст из нашего корпуса – бестселлер, а какой – нет³⁹.

Этот средний показатель 80 % означает, что, если взять наугад из недавно опубликованных книг 50 бестселлеров и 50 так и не ставших ими, наш компьютер правильно классифицирует 40 бестселлеров и 40 небестселлеров. Конечно, это значит также, что компьютер сочтет 10 бестселлеров провальными, а 10 небестселлерам напророчит блестящий успех. В проведенной нами серии тестов компьютер, в частности, был абсолютно уверен, что «Гордость и предубеждение и зомби»⁴⁰ не будет иметь успеха, – и ошибся. Конечно, «Гордость и предубеждение и зомби» вышла тогда, когда любая отсылка к Джейн Остин гарантировала книге внимание читателей, а в кинотеатрах пачками шли фильмы про зомби. Но в данном случае контекст оказал непропорционально большое влияние на уровень продаж.

Конечно, были еще и небестселлеры, которые наш компьютер превознес до небес, но это отдельная история.

³⁹ Первую попытку выявления бестселлеров мы предприняли в первом нашем эксперименте, в Стэнфорде, еще в 2008 г. Наш корпус текстов состоял из 20 000 романов, но параметров для анализа у нас было гораздо меньше, всего 505. Такой большой корпус текстов был полезен для определенных видов анализа, но при этом он был составлен с уклоном в пользу более ранних периодов – в нем содержалось слишком много книг XIX века, а среди более современной литературы преобладали научная фантастика, фэнтези и любовные романы. Несмотря на эти недостатки, мы получили хорошие результаты: точность определения бестселлера колебалась в пределах 70–80 %. При написании этой книги мы построили совершенно новый корпус текстов: более разнообразный и более современный. В нем содержалось почти 5000 произведений, в том числе электронные книги, не ставшие бестселлерами, романы, опубликованные традиционным путем, а также бестселлеры из списка NYT числом чуть больше 500. (*Примеч. авторов.*)

⁴⁰ *Pride and Prejudice and Zombies*, Seth Grahame-Smith.

Договор

Обсуждая новые романы, мы, авторы этой книги, часто говорим об отношениях читателя и автора так, будто существует некий неписанный договор – детали его туманны, но тем не менее он описывает эстетические, эмоциональные, интеллектуальные и даже этические причины, стоящие за выбором читателя. Обучая свою модель распознавать тему, сюжет, стиль и характеристики персонажей, мы много думали о том, чего ждет читатель от книги.

Этот неписанный договор содержит много статей. Если вы автор триллера, например, вы обязаны включить в него один-два трупа и сцену, в которой у читателя учащается пульс. Если вы пишете любовный роман, то он должен заканчиваться (но не начинаться) долгожданным воссоединением влюбленных. И независимо от жанра (если вы не новый литературный вундеркинд – им прощается двойная длина книги), у вас есть примерно 350 страниц, чтобы увести читателя куда-то далеко, а потом привести обратно. Таковы ваши основные обязательства, и вы наверняка видели, как брызжут желчью или роняют слезу читатели в отзывах на сайте goodreads.com, когда автор не оправдывает ожиданий.

Помня об этом, дорогой читатель, мы заключим с вами собственный договор, условия которого весьма прозрачны. Вот его статьи.

1. Номер первый

Одна из особенностей нашей культуры и книжного мира состоит в том, что мы обожаем ранжированные списки. Дело не ограничивается списком бестселлеров. В этом году в газетах и на сайтах крупных книжных сетей мелькали статьи с разнообразными заголовками вроде «Самые красивые места действия ваших любимых романов», «Десять самых влиятельных книг всех времен», «Выбери бойфренда среди литературных персонажей». Посетители сайта goodreads.com общими силами составили списки по самым разным темам: лучшие книги, действие которых происходит в космосе; лучшие японские издания; герои, оказавшие наибольшее влияние; книги, которые заставят вас всплакнуть. Этим списков тысячи. Книголюбам приятно составлять их, спорить о них и, конечно, сравнивать, с кем лучше пойти на свидание – с мистером Дарси или Кристианом Греем.

Не думайте, что у нас, авторов, иммунитет к составлению списков. Мы знаем, что всех, кто имеет отношение к книжному делу, просят порекомендовать любимую книгу. Если об этом спросили вас, ответ «У меня нет любимой книги» звучит смертным приговором – он убивает на корню и светскую беседу, и ваш авторитет профессионального читателя. Эти пять слов гарантированно действуют на собеседника как холодный душ. Так что мы участвуем в рискованной игре, поскольку понимаем, что в нашем мире *номер первый* решает все. Номер первый – это *важно*. Первый номер в списке бестселлеров NYT значит нечто гораздо большее, чем номер десятый. Возможно, именно потому, что современный мир предоставляет бесконечные возможности выбора, мы испытываем психологическую и культурную потребность – определить одного победителя, короля, бога. Выбрать *что-нибудь одно*.

К концу этой книги мы приведем собственный список рекомендованного чтения и назовем победителя – произведение, которое наш компьютер выбрал в качестве образцового, эталонного бестселлера последних тридцати лет.

2. Верьте нам, как мы доверились машине

Второе обещание, которое мы даем читателям «Кода бестселлера», – не «исправлять» выбор модели. Мы с самого начала согласились с тем, что наша задача – не выбирать, а объяснить выбор. По правде сказать, мы не читали «книгу номер один» до того, как ее выбрал компьютер, хотя и были знакомы с другими произведениями этого автора. Конечно, мы тут же схватили ее с полки, прочитали вместе и вместе посмеялись над неожиданной иронией. Мы советуем вам не перескакивать сразу к названию романа-победителя – каждая глава нашей книги объясняет очередной кусочек головоломки. Но в то же время мы знаем, какое это искушение – прочитать первую страницу и потом сразу же заглянуть на последнюю.

3. Волшебного чая не существует

Мы не утверждаем, что после первого же (или даже после второго) прочтения нашей книги вы сможете писать бестселлеры. Это вовсе не пособие «для чайников», и мы не даем никаких гарантий. Вы обязательно найдете на этих страницах множество полезных советов, которым мы сами непременно последуем, если решим написать бестселлер. И конечно, прежде чем сдавать агенту свое новое произведение, мы обязательно проведем компьютерный анализ текста. Но красота этой книги заключается еще и в том, что она позволяет пересмотреть старую аксиому, которая гласит: «писательское мастерство – дар, которому невозможно научить». Нам интереснее ниспровергать, чем обучать.

Почти все известные нам пособия для писателей (нам больше всего понравились написанные авторами книг-блокбастеров, такими как Дин Кунц и Стивен Кинг) дают советы по разным аспектам прозы: стиль, характер, сюжет. Мы делаем то же самое. Мы надеемся, что поможем вам заглянуть в глубины ДНК бестселлера, куда не проникает человеческий глаз, и препарировать это невыразимое *je ne sais quoi*⁴¹, свойственное талантливым писателям. Но мы не дадим вам формулы, которую можно было бы применить. Эта книга поведает вам многое об анатомии бестселлера, но не поможет ее скопировать – как не помогут стать писателем отпечатки пальцев Адама Джонсона⁴², даже если вы срежете кожу у него с кончиков пальцев и приклеите себе.

Возможно, вы сочтете нас возмутительно старомодными, но мы твердо верим: тот, кто хочет написать бестселлер, должен сначала изучить и уметь отличать хорошую прозу, используя для этого все доступные средства. Если наша книга окажется полезной и вы напишете бестселлер – пожалуйста, дайте нам знать. Мы обязательно купим его и подвергнем компьютерному анализу. Но не жалуйтесь, что вы искали простую формулу, чтобы отхватить миллионный контракт на роман, а мы вам ее не дали. Тот, кто предложит вам такую формулу, – шарлатан, ничем не лучше тех, кто продает волшебный чай для похудения, гарантируя «радикальную потерю веса за одну ночь».

⁴¹ Не знаю что (*фр.*).

⁴² Джонсон Адам (р. 1967) – американский писатель, лауреат Пулитцеровской премии.

4. «Черный ящик»

Эта книга — не про алгоритмы. Мы расскажем об основных параметрах, которые научились вычислять, и опишем в общих чертах наш метод, но перед вами не пособие по машинному обучению, автоматическому поиску документов или обработке естественного языка. По этим темам существует множество хороших учебников, а мы пишем о книгах, в основном о бестселлерах⁴³. Мы хотим, чтобы вы задумались о себе как читателе или писателе, о задачах литературы, об авторах, которых вы любите или ненавидите, и даже об отношениях человека и машины. Мы расскажем вам о множестве результатов и интерпретаций того, как компьютеру удалось или не удалось определить бестселлеры и чему это нас научило, но мы будем говорить про «Щегла»⁴⁴ и «Исчезнувшую», а не про латентные размещения Дирихле или, к примеру, распознавание именованных сущностей. Эти (загадочные для непосвященных) методы были важны для проделанной нами работы, и без них мы не справились бы, но это лишь инструменты: живописец рисует кистью, но он не рисует саму кисть.

⁴³ Многие методы, упомянутые в данной книге, описаны в учебнике авторства Мэтта «Text Analysis with R for Students of Literature». Этот учебник представляет собой вводное пособие в анализ текста и извлечение данных и предназначен для читателей, не имеющих подготовки в области компьютерных наук. Другая книга Мэтта, «Macroanalysis: Digital Methods and Literary History», содержит подробное описание методов, использованных нами в данной работе, в том числе моделирования темы (обсуждается во 2-й главе) и стилиметрии (в 4-й главе). Для тех, кому нужно еще более подробное описание методов, лежащих в основе наших исследований, мы рекомендуем труд Christopher Manning, Prabhakar Raghavan и Hinrich Schütze «Introduction to Information Retrieval» издательства Cambridge University Press, а также Gareth James, Daniela Witten, Trevor Hastie и Robert Tibshirani «An Introduction to Statistical Learning» издательства Springer. Обе эти книги требуют некоторого знакомства с предметом, но вместе они дают полный и очень хорошо написанный обзор дисциплин анализа текста и машинного обучения. (*Примеч. авторов.*)

⁴⁴ The Goldfinch, Donna Tartt.

2. Крестные родители, или Почему важно находить время друг для друга

Входя в книжный магазин, первое, что видишь, – столы с новыми книгами. Теперь вы знаете: чаще всего это значит, что за многие из них, если не за все, кто-то хорошо заплатил, чтобы они первыми попались вам на глаза. Это произведения самых разных жанров – романы, автобиографии, кулинарные книги, триллеры, однако остальная часть магазина организована по категориям. Если вы любите читать беллетристику, то знаете, что в магазине обычно есть отдел прозы, где размещаются по алфавиту классические и современные авторы, а также отделы жанров с соответствующими табличками – например «Любовный роман» или «Научная фантастика». Мы так привыкли к подобному размещению книг в магазине, что, кажется, нашли бы нужную с закрытыми глазами.

Расположение книг как в магазинах сети Barnes & Noble, так и в любом онлайн-магазине продиктовано убеждением, что самая главная характеристика книги – это ответ на вопрос: «О чем она?» На этом построено все книжное дело. Каждой книге, опубликованной традиционным образом, присваивается один или несколько кодов тематики по классификации BISAC⁴⁵. Они устанавливаются Группой исследования книжного дела (BISG) – ассоциацией, ответственной за создание и поддержание стандартов в отрасли. Этих кодов тысячи; для одной беллетристики их существует 152, и они определяют, в какую категорию попадет книга, как она будет выставлена на полке и как будет продаваться. Коды очень подробно описывают содержание книги – например «исторический любовный роман с участием викингов».

Для других характеристик, тоже важных читателю, – например книги со счастливым концом, книги, выжимающие слезу, книги, действие которых разворачивается в Токио, книги с участием пожарных, принцесс или монахинь, – такой всеобъемлющей системы классификации не существует. Нет кода, который бы указывал, является ли стиль книги минималистским (как у Хемингуэя), или более сложным (как у Дэвида Фостера Уоллеса). По расположению томиков в определенном отделе магазина вы не узнаете, какого пола главный герой, стар он или молод, происходят ли описываемые события в Лондоне или в Гонконге.

В принципе, можно сказать, что самая важная информация о книге – это ответ на вопрос: «О чем она?» Если вы советуете подруге, что почитать, или если вы сами писатель и упомянули свою книгу в разговоре, вас тут же спросят: «О чем она?» Очень редко (если вы пишете исключительно биографии) у вас могут поинтересоваться, о ком она, или где происходит действие, или в какую эпоху. В первую очередь людей интересует тема. Поэтому возникает вопрос: существует ли беспроектная тема?

Наш компьютер считает, что такие темы есть, и авторы книг-блокбастеров с ними согласны. В своем труде «Как писать книги»⁴⁶ – одном из популярнейших среди вышедших за последние годы пособий по писательскому ремеслу – жанровый автор Стивен Кинг советует начинающим романистам взять хорошо знакомую им тему, а потом добавить «собственные знания о жизни, дружбе, отношениях, сексе и работе. Особенно о работе. Люди обожают читать о работе – бог знает почему». Это любопытное наблюдение насчет работы многое говорит о нашей культуре – и в данном случае наш компьютер более-менее согласен с ним. По поводу отношений Кинг тоже оказался прав. Впрочем, в том, что касается секса, он, как ни удивительно, ошибается (мы скоро вернемся к этой теме), и еще он играет дурную шутку

⁴⁵ Сокращение расшифровывается как «Стандарты и обмен информацией в книжном деле».

⁴⁶ On Writing: A Memoir of the Craft, Stephen King.

с начинающими писателями, сообщая им, что «сантехники в космосе» – неплохой сюжет, который привлечет читателя. Не верьте.

Конечно, эта работа Кинга написана увлекательно и бойко, а его слова согласуются с мнением других писателей по поводу тематики. Но в конечном счете его советы оказываются не очень полезными. Такая скупость в советах начинающим писателям со стороны одного из самых коммерчески успешных авторов последнего столетия, вероятно, объясняется его убеждениями, которые он также выражает в своей книге. Во-первых, он пишет, что «беллетристы, и я в том числе, сами не очень понимают, что делают, – они не знают, почему у них вышло хорошо, если книга оказалась хорошей, или плохо, если она оказалась плохой». Эти слова, вероятно, продиктованы скорее скромностью, чем стремлением к истине, но от писателя с таким живым воображением вряд ли можно ждать аналитического разбора того, что приходит к нему само собой (в данном случае – нужные темы, взятые в правильной пропорции, так что результат привлекает массового читателя со всех концов света). Во-вторых, Кинг пишет, что преднамеренно выбирать кассовую тему – «моральная нечистоплотность». Это заявление тоже, скорее всего, продиктовано более скромностью, нежели правдивостью. Вряд ли авторы-мультимиллионеры не думают о том, какие темы будут иметь спрос, – особенно когда знают, сколько экземпляров книги нужно продать, чтобы оправдать ожидания издателя. Кроме того, это нечестно со стороны Кинга – обзывать морально нечистоплотным писателя, который разбирается в читательском спросе и хочет работать именно с популярными темами.

Мы здесь не для того, чтобы хвалить или порицать выбор автора с точки зрения морали. А также не для того, чтобы объявлять один роман хорошим, а другой плохим или одну тему более достойной для разработки, чем другая. Мы предоставляем этический выбор в процессе творчества самим творцам, а обязанность их судить – критикам. Наша же задача как литературоведов – находить новые ответы на вопросы и выводить на свет скрытые истины, надеясь, что в результате нечто смутное и неуловимое станет более понятным. Чтобы таким образом прояснить роль темы в литературе, мы сначала должны провести четкую грань между темой книги и ее предметом и показать, как они взаимодействуют, творя произведения удивительного искусства – словесности.

Зачем мы читаем?

Вспомните себя в старших классах школы. Впервые в жизни перед вами стояла задача – прочитать книгу и увидеть ее глазами литературного критика. Скорее всего, это были «Повелитель мух» или «Убить пересмешника»⁴⁷. Вы помните, что вас просили определить разницу между темой и предметом. Вероятно, учитель искал добровольца, готового ответить на вопрос: о чем роман Уильяма Голдинга «Повелитель мух»? Нас, во всяком случае, об этом спрашивали – мы помним. Но вопрос был, разумеется, с подвохом. Тогдашние мы могли бы сказать, что эта книга – о группе мальчиков, оказавшихся на необитаемом тропическом острове; о том, как они самоорганизуются, выживают; как группа делится на фракции, одна из которых становится агрессивной и доходит до убийства. Темы романа – английские мальчики, необитаемые острова, охота, постройка хижин и т. д. Но наши учителя на этом не останавливались. Они хотели научить нас более глубокому пониманию прочитанного. Они хотели, чтобы мы заметили «большие вопросы» в романе, те, которые касаются движений человеческой души, – вероятно, именно в их трактовке и состоит послание Голдинга читателю. Двигаясь в этом направлении, мы могли бы заключить, что «Повелитель мух» на самом деле о споре, что важнее – природа или воспитание, или о борьбе добра и зла, или о дружбе, или о том, к чему склонны люди в изолированной группе – скатятся ли они к состоянию дикарей или останутся цивилизованными. Если вы были способны все это сформулировать на уроке литературы, то, скорее всего, получили пятерку. Возможно, даже пятерку с плюсом – если смогли показать, как с помощью раскрытия тем, в основном благодаря подбору существительных и их употреблению в определенных повторяющихся группах, Уильям Голдинг, согласно своему замыслу, наглядно демонстрирует читателю общечеловеческие истины.

Теперь подумайте немного о том, почему вы читаете именно то, что читаете. В книжном деле царит убеждение, что читатели выбирают книгу из-за темы. Это даже кажется очевидным. Если вас спросят, какие книги вы любите, вы ответите: «про преступления», «про войну», «про секс», «про рыбалку». Это особенно заметно в нехудожественной литературе: люди читают книги по определенным темам. Самые популярные из них – еда и бизнес. Но многие читатели художественной литературы на самом деле объясняют свой выбор совсем другими причинами. Если вы спросите читателя, по какому принципу он выбирал последний прочитанный им роман, то в ответ можете услышать:

«Я выбрал эту книгу, потому что она про холокост».

«Я выбрал эту книгу, потому что мне сказали, что на ней душа отдыхает».

«Я слышала, что это замечательный новый автор».

«Я же читаю все новые романы Стивена Кинга!»

Иногда читатели объясняют свой выбор регалиями книги: «Я хотел прочитать новый роман Адама Джонсона, которому дали Пулитцеровскую премию». Порой – тем, что они собираются в отпуск в Париж, а действие выбранной книги происходит именно там, или тем, что после разрыва с любимым человеком хотелось почитать что-нибудь сентиментальное. Ну и разумеется, выбирают роман, «потому что он был в списке бестселлеров NYT». Мы понимаем, что тема – не единственный стимул читательского интереса. Однако она является важнейшим фактором в книжном деле, к тому же в ней заключены возможности развития каждого конкретного сюжета. Поэтому в данной главе мы выделим вопрос темы и покажем, как с помощью компьютера можно определить ее роль в создании бестселлера.

⁴⁷ To Kill a Mockingbird, Harper Lee.

Ощущения при чтении романа вызываются исключительно словами, расставленными в определенном порядке. Слова – это «кирпичики», строительный материал языка. Многие из них – существительные, а подбор существительных в определенных пропорциях становится инструментом, с помощью которого авторы раскрывают определенные темы, а через них – предмет книги. Именно свойственное автору чувство меры и его умение соблюсти деликатный баланс в раскрытии темы мы можем исследовать с помощью компьютерного анализа. Компьютер помогает понять, как писатель использует инструменты, чтобы вызвать у читателя те или иные переживания.

Связь между предметом книги и переживаниями читателя описать легко. Возьмем, например, женщин, которые читают любовные романы – причем постоянно и исключительно их. Это не карикатура – такие читательницы действительно существуют, и их много. В прошлом году мы ездили на конференцию «Романтические времена», посвященную любовным романам. Это мероприятие продолжительностью в несколько дней кого угодно убедит, что любовь в самом деле правит миром. Мы встретили нескольких страстных поклонниц жанра, которые клялись, что читают по три-пять *сетен* любовных романов ежегодно. Как им это удастся? Мы не знаем. Но дело не в их сверхчеловеческих способностях. Дело в том, что, когда люди глотают книги одну за другой, это может быть как их свободным выбором, так и зависимостью сродни наркотической, заставляющей их выбирать один и тот же жанр, – но это не обязательно связано с интересом к определенным темам. Да, конечно, открывая любовный роман, мы ожидаем увидеть что-нибудь про любовь и отношения – это часть неписаного контракта между читателем и автором. Но кроме любви и отношений там могут присутствовать абсолютно разные темы. В BISAC есть коды для любовных романов с участием вампиров и шотландцев; любовных романов действие которых происходит в эпоху Тюдоров или в Средние века; любовных романов в спортивном мире; эротических любовных романов – и это лишь несколько примеров. Возможно, существуют читатели-привереды, которые выбирают *только* любовные романы про оборотней с экстрасенсорными способностями (да, такая категория в BISAC тоже есть). Но с тем же успехом можно предположить, что есть люди, которые читают любовные романы ради определенного переживания, вызванного широкой направленностью книги (тема любви), а не какой-то узкой темой (например, вестерном). Какая разница, кто главный герой – вампир или ветеринар? Главное, чтобы он был прекрасен телом и душой.

Связь между предметом и ощущениями читателя – важнейший вопрос в исследовании, цель которого – раскрыть механизм коммерческого успеха в литературе. Сравним любовные романы и триллеры. Это две важнейшие категории для любого исследователя современного книжного мира, поскольку они продаются лучше всего. Они правят рынком – но разными его сегментами. С одной стороны, триллеры, похоже, все еще занимают более сильные позиции в списке NYT. Это особенно верно для книг в твердом переплете. С другой стороны, стремительно развивающийся мир электронного «самиздата» растет в основном за счет любовных романов. Это золотая жила для исследований – хватило бы на целую отдельную монографию: пассажиры в метро по дороге на работу читают знойные истории любви со смартфонов, а детективы – в бумажных изданиях; при этом среди любителей детективов примерно поровну мужчин и женщин, а вот среди читателей любовных романов женщины преобладают. Однако важный момент, интересующий нас сейчас, заключается в том, что любовные романы и триллеры вызывают у читателя сходные переживания в интеллектуальной и эмоциональной сферах и в том, что касается работы воображения. Любитель триллеров, скорее всего, покупает книги, в которых описаны пытки, шпионаж, проверки алиби, – потому что стремится попасть в мир, где людям угрожают, преследуют их и убивают. Соответственно, триллеры, не создающие атмосферу нависшей угрозы, вряд ли будут иметь успех. Триллер, посвященный рыбалке и термопечати на футболках, скорее всего, не захватит воображение

читателя. Как читательница любовного романа жаждет читать о любви, так читатель триллера хочет читать в основном о преступлениях. Мораль ясна: если мы хотим понять, почему та или иная тема имеет успех у читателей, то нужно думать о том, как она действует на нас и совпадает ли это воздействие с нашими желаниями. Иными словами, совет Кинга «писать о том, что знаешь, плюс любовь, секс, работа и отношения» нуждается в существенном уточнении.

При построении компьютерной модели мы рассчитывали на то, что, исследуя темы подробней, чем это доступно обычному человеческому восприятию, мы сможем прийти к более глубокому пониманию тем, постоянно присутствующих в списках бестселлеров. И мы оказались правы. Поскольку в распоряжении писателя – все темы на свете, легко предположить, что бестселлеры – это книги про «секс, наркотики и рок-н-ролл». Но выяснилось, что это не так. Далеко не так.

Секс, наркотики и рок-н-ролл

Секс 0,001 %

Наркотики 0,003 %

Рок-н-ролл 0,001 %

Эти крохотные доли процента отражают присутствие каждой темы (в среднем) в современных романах из нашего корпуса текстов. Вероятно, эти показатели потрясли вас до глубины души. Сейчас мы объясним, почему они настолько низки, и, возможно, потрясем вас еще раз. Если взять набор из 500 тем и проследить их по всему корпусу из 5000 текстов, среди которых есть 500 бестселлеров, то в среднем описания секса в них будут занимать всего около одной тысячной доли процента. Если затем измерить тот же показатель (чуть позже мы объясним, как это делается) только среди бестселлеров, окажется, что доля секса в них еще ниже – 0,0009 %.

Удивительно. Кто бы подумал, что секс – не двигатель продаж? Когда мы об этом рассказываем, нам до сих пор не верят. Истина заключается в следующем: секс (точнее, эротическая литература) продается, причем в немалых количествах, но лишь на определенном, нишевом рынке. Книги этого жанра редко вырываются на основной рынок, который создает бестселлеры. Мы знаем, что вы сейчас подумали: «А как же “Пятьдесят оттенков серого”?» Этот роман (или эти романы, если говорить о всей трилогии) – чрезвычайно редкий пример эротического произведения, попавшего в бестселлеры. В следующей главе мы объясним причины этого успеха – и окажется, что они не имеют отношения к собственно сексу.

Вопреки видимости (учитывая навязчивое присутствие темы секса на ТВ, в кино и СМИ) читающая публика США в последние тридцать лет явно предпочитает другие темы. Сочетание тем, характерное для современного бестселлера, дает основания предположить, что современный читатель ждет от книги чего-то большего, нежели потакание самым низким инстинктам.

Но откуда мы это знаем?

В 1957 году лингвист Джон Руперт Фёрс сказал: лучший способ понять слово – это познакомиться с его соседями. Проще говоря, значение слова раскрывается через контекст, в котором оно употреблено. Слова «секс», «наркотики» и «рок-н-ролл», фигурирующие в названии этого раздела, можно было бы понять соответственно как сокращение слова «секстет», «обезболивающие средства» и «акробатические танцы» – но вы знаете, что в данном контексте имелось в виду не это, именно потому, что каждое из слов придает определенный оттенок смысла своим соседям. Возьмем другой пример – слово «три». Оно может означать числительное, а может – повелительное наклонение глагола⁴⁸. Чтобы различать эти два случая, компьютер нуждается в специальном обучении – по принципу, сформулированному Фёрсом семьдесят лет назад. Компьютер должен рассматривать каждое слово в контексте окружающих его слов в ближайших предложениях. Алгоритмический метод для формализации такого рассмотрения слов в контексте в больших объемах называется моделированием тем⁴⁹.

⁴⁸ В отдельных случаях примеры адаптированы для русского читателя. (Примеч. ред.)

⁴⁹ Мы использовали латентное размещение Дирихле – алгоритм моделирования тем, предложенный Дэвидом Блеем, сотрудником Колумбийского университета. Слово «латентный» в названии метода Блея отражает то, что некоторые темы могут присутствовать в тексте в неявном виде. (Примеч. авторов.)

Математический аппарат, задействованный в моделировании тем, весьма сложен, но его общие принципы достаточно просты. Каждый роман представляет собой комбинацию тем, а эти темы выражаются словами (в первую очередь – существительными). Например, в книге о финансах, скорее всего, встретятся слова «банки», «проценты», «деньги», «фидуциарный». В другой книге, посвященной домашнему консервированию, мы увидим слова «банки», «крышка», «огурцы» и «укроп». Слово «банки» попадает в обеих книгах, но, рассмотрев другие существительные, стоящие рядом с каждым вхождением этого слова, компьютерная модель регистрирует повторяющиеся закономерности и может понять – подобно читателю-человеку, – что в разных местах слово «банки» имеет разные значения⁵⁰. Очевидно, что слово «банки» само по себе еще не тема, но, когда оно постоянно попадает рядом с другими словами, которые (как нам уже известно) связаны с финансами, мы понимаем, что компьютер обнаружил в книге тему денег. Аналогично, если «банки» окружены «огурцами» и «пряностями», мы знаем: компьютер обнаружил тему домашних заготовок. При написании этой книги мы проводили моделирование тем, и одна из тем, обнаруженных компьютером, выглядела так, как показано на рис. 1.

⁵⁰ В 8-й главе своей книги «Macroanalysis: Digital Methods and Literary History» (UIUC Press, 2013) Мэтт более подробно описывает алгоритм моделирования тем. В 13-й главе своей книги «Text Analysis with R for Students of Literature» (Springer, 2014) Мэтт объясняет читателям, как проводить моделирование тем на алгоритмическом языке R (и как создавать словесные облака, которые вы увидите на этих страницах). Мэтт также опубликовал несколько объяснений для неспециалистов у себя в блоге (www.matthewjockers.org), в том числе: «The LDA Buffet is Now Open; or, Latent Dirichlet Allocation for English Majors» (29.09.2011) и «“Secret” Recipe for Topic Modeling Themes» (12.04.2013). (Примеч. авторов.)



Рис. 1

Эта тема посвящена *барам* – не единицам измерения, а заведениям, куда заходят выпить рюмочку. Значение этого слова вполне прозрачно. На нашей визуализации в виде облака слова изображены шрифтом разного размера: чем крупнее шрифт, тем чаще это слово попадает рядом со словом «бар». Это очень логично. Слова, непосредственно окружающие центральное слово «бар», – «бармен», «выпить», «виски», «пиво» – дают нам уверенность в том, что если компьютер нашел все эти слова в пределах одной-двух страниц, то действие разворачивается в обычном местном баре.

Не все темы, обнаруженные компьютером, так легко интерпретировать. Иногда требуются познания в литературе и дар рассуждения. Время от времени приходится распознавать необычный язык, сконструированный автором фантастической саги, или малоизвестный диалект. Если тема составлена из таких слов, как *аск*, *уолл на уолл*, *боллзы* и *амбрелла*, вы будете долго напрягать *брейнз*⁵¹, пока не догадаетесь, что все это – сленг из романа Энтони Берджесса «Заводной апельсин». Однако гораздо чаще тематическое облако выглядит примерно как на нашем рис. 2.

⁵¹ A Clockwork Orange, Anthony Burgess. Использован перевод «Заводного апельсина», сделанный Е. Г. Синельщи-
ковым.



Рис. 2

Такие существительные, как *глаза, рот, рука, голова*, говорят о том, что тема имеет какое-то отношение к телу. Однако просто *тело* – недостаточно конкретное название. *Тело* может фигурировать, например, в описании сцены убийства, но это явно не наш случай. Образующие контекст слова на самом краю облака – *поцелуи, наслаждение, улыбка, дыхание, постель, ритм, жар* – указывают на то, что эта группа существительных описывает любовную сцену. Но это тоже слишком общее выражение. Любовные сцены бывают разные. Большинство писателей утверждает, что правильно описать секс – нелегкая задача, за которую они берутся безо всякого удовольствия. Каждый автор решает ее по-своему, с большим или меньшим успехом. В данном случае нужно заметить, что выбор слов указывает на довольно сдержанный стиль писателя – он явно не злоупотребляет грубостью или откровенностью. Словесные облака, созданные из описаний любовных сцен, показывают, что эти описания сильно различаются: от таких, которые можно не смущаясь читать вслух собственной бабушке, – до таких, какие выдержит лишь закаленный читатель.

Благодаря алгоритму моделирования тем мы получаем два важных результата. Во-первых, компьютер сообщает, какие темы есть в нашем корпусе текстов (в том числе – какие слова составляют каждую из них; примеры мы только что видели в словесных облаках слов «бар» и «тело»). Во-вторых, компьютер определяет долю содержания каждой темы в каж-

дой книге. Мы задали список из пятисот возможных тем, так что вариации тематического состава могут быть огромны. Узнав эти пропорции, мы можем начать поиск закономерностей, типичных для бестселлеров. Это нечто вроде обратного конструирования. Возьмем в качестве метафоры тарелку супа. Компьютер сначала делит его на составляющие – мясо, бульон, капуста, лук, специи, – а затем тщательно измеряет количество каждого ингредиента.

Измерив таким образом содержание ингредиентов, мы начинаем задавать вопросы. Например, можно спросить, какие банки чаще попадают на страницах бестселлеров – те, что с деньгами, или те, что с вареньем. Или какого типа описания секса чаще встречаются в бестселлерах – совсем откровенные или более сдержанные. Кстати, если вам интересно: чтобы попасть в список бестселлеров NYT, лучше выбирать банки с деньгами, а любовные сцены (если уж без них никак нельзя) – те, которые рисуют душевную близость, а не насилие. Откровенная сцена поможет попасть в бестселлеры, если она продвигает вперед сюжет и взаимоотношения между персонажами. Если сцена добавлена автором для «оживляжа» и на самом деле не нужна для развития сюжета, она, скорее всего, не поможет книге продаваться и, следовательно, лишняя. Возможно, это объясняет, почему доля сексуальной темы в бестселлерах немного ниже по сравнению со всем корпусом текстов.

Конечно, нельзя создать книгу из одной темы. Возьмем, например, «Там, где течет река»⁵² Нормана Маклина. Можно сказать, что она – про рыбалку и про деньги, но она также про религию и про отношения между братьями. У каждого романа есть свой собственный тематический профиль, отражающий не только набор тем, направленный на создание определенного переживания у читателей, но и точную пропорцию, в которой эти темы появляются в романе. В плане тематики для успеха книги важно и то и другое. Можем ли мы, глядя на рукопись, понять, насколько ее тематический профиль обеспечит ей успех?

Легко предположить вслед за многими редакторами, что роман об убийствах, расследованиях и командном спорте скорее понравится массовому читателю, чем роман о коллекционировании рептилий, космических ракетах и университетской жизни. Но для настоящего прогноза нужно задаться вопросом: существуют ли темы, общие для большинства бестселлеров? И в то же время необходимо доказать, что эти темы реже встречаются в романах, которые в список бестселлеров не попадают. Это сложная задача. При поверхностном взгляде на любой недельный список бестселлеров кажется, что набор тем совершенно случаен и ничего общего в них нет.

Возьмем список за текущую неделю. Сейчас, когда мы пишем эти строки, на первом месте среди романов в твердом переплете стоит «Вне правил»⁵³ Джона Гришэма. В этом нет ничего удивительного – Гришэм и первое место в списке бестселлеров сочетаются так же органично, как картофельный салат и Четвертое июля. Вопрос, однако, в том, обусловлено ли первое место в списке темами, присутствующими в книге. Второе место занимает детектив Джеймса Паттерсона из серии про Алекса Кросса. На третьем месте – роман Тома Клэнси, написанный кем-то другим (Клэнси умер в 2013 году). Далее идет сборник рассказов Стивена Кинга, детектив Дэвида Балдаччи, любовный роман Николаса Спаркса и очередное творение Джанет Иванович об охотнице за головами. Все эти произведения можно так или иначе отнести к жанровой прозе. Поскольку книга Кинга не роман, а сборник рассказов, ей, вероятно, свойственно большее тематическое разнообразие, чем роману. А поскольку Кинг работает в жанре хоррора и саспенса, можно догадаться, что его темы будут отличаться от тех, что использует Джеймс Паттерсон. Далее идет удостоенный Пулитцеровской премии «Весь невидимый нам свет»⁵⁴ Энтони Дорра, потом новое творение Митча

⁵² A River Runs through It, Norman Maclean.

⁵³ Rogue Lawyer, John Grisham.

⁵⁴ All the Light We Cannot See, Anthony Doerr.

Элбома (как обычно, «про духовность»). Потом Пола Хокинс с «Девушкой в поезде»⁵⁵, которая держится в списке уже сорок седьмую неделю, потом – старый-новый роман Харпер Ли «Пойди поставь сторожа»⁵⁶, детектив Майкла Коннелли, «Соловей»⁵⁷ Кристин Ханна (его действие, как и действие книги Дорра, происходит во время Второй мировой войны) и наконец – любовный роман Даниэлы Стил и очередная часть саги Джорджа Р. Р. Мартина.

Посмотрев на список, можно ли сказать, «о чем» все эти бестселлеры в твердом переплете? Очевидно – ни о чем в частности и в то же время обо всем. Но в нем есть и закономерности. Подсказкой может служить то, что Джордж Р. Р. Мартин – своего рода аномалия. Его книга стала бестселлером в конце 2015 года, вероятно, в большой степени благодаря колоссальному успеху телесериала «Игра престолов». По крайней мере, про этот роман можно сразу сказать, что его попадание в бестселлеры было неожиданным. Тридцать лет назад – другое дело, тогда жанр фэнтези был популярен. Сейчас же в моду вошел современный реализм.

Мы пишем эту главу за неделю до Рождества – сейчас люди покупают подарки, и понятно, что на первое место в списке выходят самые популярные авторы. Поэтому задача распознавания тем слегка упрощается по сравнению с остальными неделями года. По-видимому, доминирующая тема – преступление. За ней идет война, которую можно назвать старшей сестрой преступления. В романах Спаркса и Стил, как и следовало ожидать, главная тема – любовь, но в аннотациях упоминаются отношения, грозящие героине опасностью, и смерть любимого человека. Так что и здесь не все безоблачно. Стивен Кинг (если не заболел) наверняка написал, как обычно, полный саспенса ужастик. В общем, несмотря на разнообразие (впрочем, не чрезмерное) тематики в списке этой недели, можно, вероятно, заключить, что основные темы, приводящие книгу в список бестселлеров, – насилие и страх. На этом мы могли бы завершить свой краткий анализ, сделав вывод, что удивительное культурное явление под названием «список бестселлеров NYT» отражает одержимость современного американского общества темой насилия (так что психиатры и социологи в этой стране еще долго не останутся без работы).

Надо сказать, что ни один из нас не собирается оспаривать этот вывод. Кроме того, ни один из нас не является дипломированным социологом. Раз уж мы пришли к этому выводу, мы его обнародуем. Но если вы решили срочно поменять профессию, пойти в литературные агенты и отбирать книги исключительно по наличию в них темы насилия, лучше погодите; давайте вместе рассмотрим менее очевидные закономерности, отраженные в этом списке.

При первом же взгляде на список бестселлеров в мягких обложках видно, что здесь ситуация несколько иная. Из десяти лидеров продаж преступлениям посвящена только одна книга. Это опять роман Джеймса Паттерсона об Алексе Кроссе, но уже другой: Алекс Кросс вынужден бросить семью за рождественским столом и бежать выручать заложников. Надо сказать, что в списке бестселлеров в мягкой обложке доминантная тема не просматривается вообще. На первом месте «Марсианин» – книга о покорении Марса, популярности которой очень способствовал успех фильма с Мэттом Деймоном. На втором – старая книга Филипа К. Дика (про Вторую мировую – но в альтернативной истории, где союзники войну проиграли). Два-три романа про жизнь женщин, любовь и дружбу. «Алхимик»⁵⁸ Пауло Коэльо, аллегорический роман о странствующем пастухе, находится в списке – потрясающе! – уже 383 недели. «Первому игроку приготовиться»⁵⁹ – книга Эрнеста Клайна о компьютерном игroke

⁵⁵ The Girl on the Train, Paula Hawkins.

⁵⁶ Go Set a Watchman, Harper Lee.

⁵⁷ The Nightingale, Kristin Hannah.

⁵⁸ O Alquimista, Paulo Coelho.

⁵⁹ Ready Player One, Ernest Cline.

в виртуальной реальности. Ее тоже сейчас экранизируют. В четырех романах – Филипа К. Дика, «Все, чего я не сказала»⁶⁰, «Поезд сирот»⁶¹ и «Бруклин»⁶² – действие происходит в XX веке, но в разные десятилетия. Может быть, ключ к мегауспеху – не насилие, а историческая тематика? Даже если это и так, ее все же маловато для обобщения: четыре книги – это не десять. Что же это все-таки за тема, общая для всего списка? Может быть – путешествие?

Пока ясно только одно: во всех книгах списка персонажам приходится нелегко. Но чтобы обнаружить закономерности, нужно в первую очередь наблюдать за фактами и проверять гипотезы. Давайте шутки ради скажем, что если издательство намерено выложить несколько миллионов долларов за новый бестселлер, то, судя по этому списку, он должен быть о путешествии, которое главный герой совершает против своей воли или в страхе. Пожалуй, на сегодняшнем рынке такая книга имеет хорошие шансы на успех. И конечно, это предоставит редактору большую свободу действий, поскольку книга, соответствующая этому описанию, может быть и любовным романом, и триллером, и фантастическим романом, и произведением серьезной литературы. Этот сюжет можно отдать на разработку разным писателям и получить множество непохожих книг, несущих читателю разные послания, – поэтому все бестселлеры кажутся различными, хотя, как мы скоро увидим, в каком-то смысле они сходны.

Вот важный вывод из этого блиц-анализа, который следует запомнить: тема шире границ жанра. Первое, что нужно сделать, если вы хотите писать, публиковать или распознавать бестселлеры, – вообще забыть о понятии жанра, несмотря на его роль в книжном деле. Свадьба и супружеская жизнь могут фигурировать в книге любого жанра. Любовь и преступление – тоже. Пропорции тем в разных жанрах могут отличаться, но важно, чтобы где-то в тексте эти ключевые темы были. Вся проделанная нами работа доказывает, что концепция жанра подобна смиренной рубашке. Забудьте о ней. Если вы готовы взять на вооружение подобный образ мыслей, значит, вы уже начали мыслить как наш компьютер, строящий прогностическую модель.

⁶⁰ Роман Селесты Инг. *Everything I Never Told You*, Celeste Ng.

⁶¹ Роман Кристины Бейкер Кляйн. *Orphan Train*, Christina Baker Kline.

⁶² Роман Колма Тойбина. *Brooklyn*, Colm Tóibín.

Проверка

Для каждой книги в нашей коллекции были вычислены процентные значения, указывающие на содержание в ней той или иной из 500 тем, попавших в наш список. Конечно, в некоторых книгах какие-то темы не появляются вообще – например, в «Коде да Винчи» нет ковбоев. Но для каждого текста мы получили диаграмму вроде той, что изображена на рис. 3. Эта диаграмма – для книги Джоди Пиколт «Последнее правило»⁶³: 23 % объема занимает тема, которую мы условно назвали «Дети и школа», 10 % – «Преступления», 7 % – «Судебные процессы и юридические вопросы», 6 % – «Домашние дела» и 2 % – «Близкие отношения».

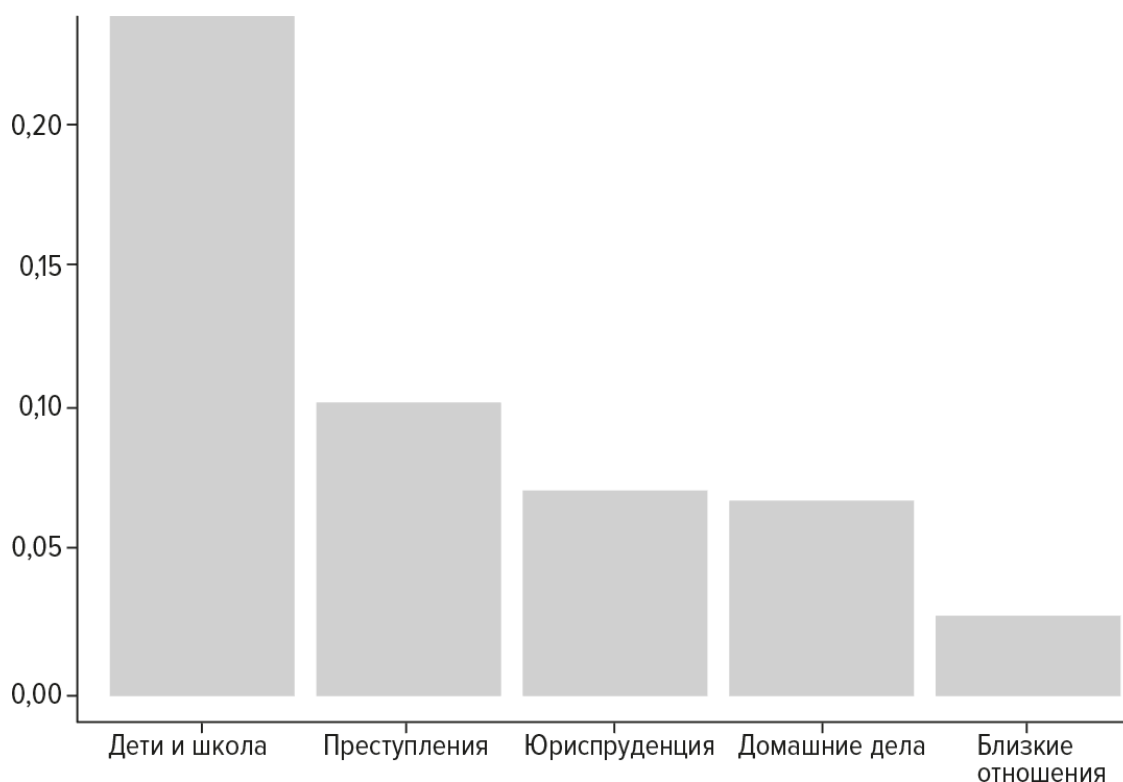


Рис. 3. Основные темы романа Джоди Пиколт «Последнее правило»

После того как наш компьютер выделил основные темы, каталогизировал их и определил пропорцию каждой темы в каждой книге, мы обработали случайно выбранное подмножество результатов с помощью алгоритма машинного обучения, которому было заранее известно, какие книги в коллекции являются бестселлерами, а какие – нет. Используя эту информацию, программа подсчитала, какие темы и в каких пропорциях наиболее вероятно встретить в случайно выбранном бестселлере. Более того, эта же программа может определить, в использовании каких тем наиболее ярко проявляется различие между бестселлерами и небестселлерами.

Хороший пример – тема секса, о которой мы уже говорили раньше. В среднем она возникает в небестселлерах вдвое чаще, чем в бестселлерах. Компьютер способен уловить эту разницу и использовать полученные данные, чтобы спрогнозировать вероятность успеха доселе неизвестной рукописи. Согласно нашей модели, книга, в которой секс фигурирует

⁶³ House Rules, Jodi Picoult.

почти в каждой главе, имеет мало шансов стать по-настоящему популярной. Конечно, из этого принципа есть исключения – трудно не заметить Сильвию Дэй или Э. Л. Джеймс, – но два автора не могут заметно повлиять на общие результаты исследования тысячи книг.

В конце концов наша компьютерная модель научилась на основании тематического профиля предсказывать с точностью 80 %, станет ли книга бестселлером⁶⁴. И когда мы спросили модель, каким авторам за последние 30 лет лучше всего удавалось (преднамеренно или инстинктивно) использовать нужные темы в правильной пропорции, компьютер назвал два имени: Джон Гришэм и Даниэла Стил. Это нас как громом поразило. Наша реакция была вызвана не столько мнением по поводу их творчества – на тот момент мы еще не слишком пристально изучили их тематический «геном», – сколько тем фактом, что наша модель смогла выявить двух самых успешных писателей в истории человечества. Из всех их произведений она выбрала несколько самых, на ее взгляд, перспективных. Среди книг Стил это оказались «Неожиданный роман»⁶⁵, «Благословение»⁶⁶ и «Жить дальше»⁶⁷. У Гришэма – «Противники»⁶⁸, «Юрист»⁶⁹ и «Джо из Калико»⁷⁰.

⁶⁴ Точность измеряется с помощью перекрестной проверки. В этом случае мы использовали два типа перекрестной проверки: в первом (скажем для читателей, которых интересуют подробности) сначала использовалось многократное обучение на случайной выборке 90 % книг, а затем проверка результатов на оставшихся 10 %. Во втором мы использовали принцип «все, кроме одной», заключающийся в том, что мы обучали модель на множестве всех книг, изымая оттуда лишь по одной книге единомоментно. В обоих экспериментах успехом считалось, когда машина правильно угадывала класс (бестселлер или небестселлер) изъятой книги (или книг, при первом подходе). Это длительный и кропотливый процесс, включающий в себя обучение и испытание множества моделей. При обучении этих моделей нам приходилось использовать уже изданные книги, про которые было известно, стали они бестселлерами или нет. Первая модель была построена в 2011 г. Добавив в корпус текстов все бестселлеры и некоторое количество небестселлеров, вышедшие за следующие пять лет, мы смогли снова испытать модель и в каком-то смысле проверить, работают ли наши наблюдения, сделанные в 2011-м, для 2015 г. Оказалось, что да. Подкрепив свою уверенность с помощью перекрестной проверки, мы можем исследовать прогноз и связанные с ним показатели вероятностей для каждой книги. По поводу некоторых книг модель не смогла прийти к окончательному выводу, и такие книги получали индекс 51 %. Поскольку случайная догадка имеет вероятность 50 на 50, это не очень высокая степень уверенности. В таком случае мы исследовали все остальные характеристики текста – например стиль или сюжет, – прежде чем сделать вывод о вероятности успеха. Но для других книг модель прогнозировала индекс успеха с большой уверенностью и присваивала им соответствующий рейтинг. (*Примеч. авторов.*)

⁶⁵ The House on Hope Street, Danielle Steel.

⁶⁶ Mixed Blessings, Danielle Steel.

⁶⁷ Accident, Danielle Steel.

⁶⁸ The Litigators, John Grisham.

⁶⁹ The Associate, John Grisham.

⁷⁰ Calico Joe, John Grisham.

Подходящие герои

Мы прозвали Даниэлу Стил и Джона Гришэма «крестными отцом и матерью» современного бестселлера – в частности, потому, что за последние несколько десятилетий их книги неизменно входили в список NYT. Их общий вклад огромен. Их преданность писательскому делу вдохновляет. Их работоспособность чрезвычайно высока. Их капиталов, заработанных писательским трудом, вероятно, хватило бы на раскрутку сотни или даже тысячи молодых литераторов. Еще мы зовем их «крестными» потому, что в книжном мире они представляют собой образцы для подражания, которые словно олицетворяют собой Великую Американскую Мечту.

Когда читаешь их биографии, кажется, что перед тобой герои увлекательного романа. Даниэла Фернандес Доминика Шулейн-Стил родилась в Нью-Йорке (единственный ребенок в семье). Отец происходил из богатой династии пивоваров; мать, ослепительная красавица, была дочерью португальского дипломата. Еще в детстве Даниэла наблюдала роскошную и разгульную жизнь великосветского общества и начала писать стихи и рассказы. Когда ей было всего семь лет, ее родители развелись, и она жила то в Париже, то в Нью-Йорке – воспитывал ее в основном отец, другие родственники и слуги. Мать она почти не видела. Даниэла поступила в школу дизайна, собираясь стать модельером, но от слишком напряженной учебы у нее открылась язва желудка. В восемнадцать лет Даниэла вышла замуж за первого из череды богатых и влиятельных мужчин, обеспечив себе роскошную жизнь. Но Даниэла хотела и начала работать в области пиара и рекламы. Один из клиентов подал ей идею писать книги. В начале 70-х она удалилась в свой шикарный особняк в Сан-Франциско и начала работу над первым романом. Вспоминая об этом, она рассказывает, как писала в комнате для стирки, на крышке стиральной машины, чтобы не разбудить семью.

Джон Рэй Гришэм-младший родился в 1955 году в Арканзасе. Отец его был строительным рабочим – он скитался по всему американскому югу, переезжая вместе с Джоном и его четырьмя братьями и сестрами туда, где была работа. Ни отец, ни мать Джона не учились в университете, и денег в семье не было. Когда Джону было двенадцать лет, семья осела в штате Миссисипи. Как и Стил, Гришэм мечтал о карьере, ничего общего не имеющей с писательским делом. В старших классах он был подающим надежды спортсменом, звездой бейсбола. Он надеялся пойти в профессиональный спорт, но позже передумал и начал изучать бухгалтерское дело и налоговое законодательство. Он стал «уличным адвокатом» – юристом, представляющим интересы людей, а не корпораций, – женился на школьной подруге (они до сих пор вместе) и родил с ней двоих детей. Он не думал о писательской карьере. Но потом он решил наблюдать за выступлениями дорогих адвокатов в суде и однажды оказался на заседании, где слушались показания десятилетней девочки: ее избили, изнасиловали и бросили, сочтя мертвой. Эта трагическая история произвела очень сильное впечатление на Гришэма, и он задумался о том, что было бы, если бы отец девочки убил насильника. Потом он стал думать: что, если бы дело происходило на американском юге и насильник бы был белый, а отец-мститель – чернокожий. Этот случай не шел у Гришэма из головы, и наконец он решил сесть и записать все свои мысли, просто чтобы посмотреть, не получится ли книга. Три года подряд он ежедневно просыпался в пять утра, чтобы поработать над книгой, прежде чем начнется обычный день адвоката – в общей сложности семьдесят рабочих часов в неделю.

Разве эти две биографии – не образцовые сюжеты великосветского любовного романа и блокбастера? Можно сказать, герои этих историй были обречены на успех. Впрочем, ни к кому из них успех не пришел за одну ночь. Десятки лет спустя оба писателя подчеркивали в интервью, как важно быть настойчивыми и упорно трудиться. В конечном счете это единственный путь к Великой Американской Мечте.

Первая книга Стил, «Возвращение»⁷¹, была опубликована в 1973 году и разошлась довольно быстро. Стил работала по восемнадцать-двадцать часов в день и вскоре написала еще пять книг, но никак не могла найти для них издателя. Писательница твердо решила не сдаваться и наконец нашла покупателя на свою седьмую книгу. Этот роман – ее вторая по счету публикация – вышел в 1977 году. С тех пор Стил написала еще почти сотню романов (не считая книг для детей и документальной литературы), и каждый из них становился бестселлером.

Книга Гришэма, написанная по материалам дела об изнасиловании, сначала называлась «Похоронный звон», а потом – «Время убивать». Ее отвергло множество агентов и редакторов, и наконец Дэвид Гернерт – последний из длинной вереницы агентов, к которым обращался Гришэм, – согласился работать с рукописью и обошел двадцать шесть издательств. В двадцать шестом ее взяли. С Гришэмом заключили контракт на 15 000 долларов. Книга вышла тиражом в пять тысяч и почти не продавалась (по слухам, сам Гришэм купил тысячу экземпляров). Но он продолжал писать. Его вторая книга называлась «Фирма»⁷². Сначала ему удалось продать права на экранизацию – за 600 000 долларов, а вскоре после этого и саму книгу за 200 000. С момента публикации в 1991 году «Фирма» около сотни раз фигурировала в различных списках бестселлеров.

И Стил, и Гришэм стали завсегдатаями этих списков с начала своей писательской карьеры и оставались в них год за годом. Статистические показатели несколько варьируются от одного источника к другому, но в любом случае у Гришэма и Стил тиражи больше, чем у любого другого из ныне живущих писателей. Стил часто называют четвертым по счету из самых популярных авторов всех времен и народов – после Шекспира, Агаты Кристи и Барбары Картленд. Около 130 ее книг выходили в общей сложности в 69 странах на 43 языках. В 1989 году она попала в Книгу рекордов Гиннеса за то, что ее роман оставался в списке бестселлеров NYT в течение 381 недели! По оценкам специалистов, общий тираж всех ее произведений составил 600 миллионов.

Список книг Гришэма гораздо скромней (их всего 35), но его литературная карьера развивалась беспрецедентным образом: в первые же годы он умудрился занять сразу четыре главные позиции в списках бестселлеров – номер первый в списке бестселлеров в твердом переплете и номера с первого по третий в списке бестселлеров в мягкой обложке. За это Гришэма прозвали самым быстропродающимся романистом в истории книжного дела. Он с самого начала практически застолбил первый номер в годовом списке бестселлеров Publishers Weekly. За 90-е годы его книги разошлись общим тиражом 61 миллион. Даниэле Стил за это десятилетие удалось продать 37,5 миллиона экземпляров своих книг.

Кто бы из них ни победил в этой гонке – думаю, вы согласитесь, что это не важно, – оба они добились успеха как литераторы. И какого успеха!

Но почему?

Ясно, что это очень важный момент: компьютер, прочитавший несколько тысяч современных романов и ничего не знающий о репутации писателей, выбирает именно Даниэлу Стил и Джона Гришэма как наиболее вероятных авторов бестселлера на основании того, как они понимают тему и предмет книги. Шансы каждого отдельного автора быть выбранным – исчезающе малы. Количество всех возможных комбинаций всех возможных тем в самых разных романах астрономически велико. Как же получилось, что наша модель выбрала двух из небольшой кучки потрясающе успешных авторов за последние тридцать лет?⁷³ Неужели

⁷¹ Going Home, Danielle Steel.

⁷² The Firm, John Grisham.

⁷³ Для тех, кого интересуют подробности, в послесловии объясняется, как мы вводили поправки при работе модели с

взрывной успех Стил и Гришэма был с самого начала заложен в «гены» их рукописей? Ответ, по всей вероятности, «да». Но тут возникает еще один вопрос: как вышло, что наш алгоритм объединил в одну группу двух таких непохожих писателей? Несмотря на то что один пишет триллеры, а другая – любовные романы, что-то – касающееся именно тематики – они должны делать одинаково.

Если вычислить содержание тем во всех книгах обоих авторов, то окажется, что Стил и Гришэм явно руководствовались формулой «пиши о том, что знаешь». В книгах писателя, который мечтал о бейсбольной карьере, а потом стал юристом, главная тема – «юристы и юриспруденция», а за ней следует «американский командный спорт». Стил, у которой на счету пять браков, девять выросших детей и один погибший, пишет в основном на темы «семейная жизнь», «любовь» и «материнство». Сама Стил утверждает, что в книгах затрагивает самые разные темы – мировые войны, Уолл-стрит, индустрию моды, болезни, взаимоотношения братьев и сестер, растрату, аборт, – но мы выяснили, что это разнообразие проявляется лишь в малых дозах. Исследуя тематический состав книг Гришэма и Стил, мы видим, что у каждого из «крестных родителей» есть одна тема, которая пронизывает все их творчество и прочно ассоциируется с каждым из них. Наша компьютерная модель обнаружила множество данных, подтверждающих, что это очень мудрый подход.

Мы уже говорили о неписаном договоре между читателем и писателем – о том, чего ждет читатель, открывая книгу любимого автора. Часть этого договора касается тем. И Гришэм, и Стил постоянно выполняют свои обязательства по договору, причем неизменно в одних и тех же пропорциях. Примерно треть абзацев во всех книгах Гришэма описывает работу юридической системы. Аналогичным образом почти ровно треть абзацев в книгах Стил посвящена домашней жизни, или, точнее, времени, которое человек проводит внутри дома. Здесь нужно сделать несколько важных замечаний. Эти авторы известны своей «фирменной» темой, и читатели именно ее от них и ждут. Если автор пишет по роману в год и главной теме отведена примерно треть объема, остальные две трети можно занять другими, благодаря чему каждая новая книга немного отличается от предыдущих. Эта формула позволяет работать бесконечно долго, и теперь мы понимаем скрытую в ней мудрость: все дело в пропорции. Треть – привычного и ожидаемого, две трети – нового.

Мы стали разрабатывать этот вопрос дальше и обнаружили интересную закономерность, характерную для всех бестселлеров, а не только для произведений Стил и Гришэма. Оказалось, что все успешные авторы отводят именно такой объем, 30 %, под одну или две темы, а прочие авторы пытаются втиснуть в книгу как можно больше идей. У автора-среднячка на треть романа приходится как минимум три темы (а зачастую и больше). На 40 % бестселлера возникает всего четыре темы, а на 40 % небестселлера – в среднем шесть тем. Вам кажется, что мы бомбардируем вас цифрами? На самом деле связность повествования очень важна для воздействия на читателя. Чем меньше тем понадобилось автору для раскрытия важнейшего момента в сюжете – тем проще и ему, и читателю сосредоточиться на главном. Это означает отсутствие ненужных отступлений и говорит о том, что автор лучше организован и точнее мыслит. Это – показатель опытности писателя.

Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.