

Игра слов



Erez Aiden and Jean-Baptiste Michel

Uncharted

Big Data as a Lens on Human Culture

Riverhead Books, New York, 2013, 288 pp., \$27.95 (cloth).

В этой книге много говорится о «не нанесенных на графики» (uncharted) областях. Но Эрез Айден и Жан-Батист Мишель при этом не подразумевают, что какие-то аспекты не охвачены анализом — на самом деле более уместно было бы назвать эту книгу «Все, что нанесено на графики».

В книге рассказывается о работе по сбору миллиардов слов, содержащихся во всех книгах мира, — слов, которые ранее терялись в смысле текста (то есть не отображались графически), но теперь их можно нанести на графики в свое удовольствие. Авторы надеются, что этот процесс позволит нам открыть для себя интересные аспекты нашей культуры, которые они называют «культуромикой». Айден и Мишель работали совместно с Google над созданием мощного веб-приложения, но их утверждения относительно его полезности, пожалуй, преувеличены.

При этом книге не откажешь в занимательности. Об этом можно догадаться с первых строк, в которых Айден благодарит своих трех детей и упоминает второе имя своей дочери: Банана. (По крайней мере, он последователен в своей эксцентричности: его сына зовут Галилео.)

Я совсем не против занимательности. Но Айден и Мишель проводят важную научную работу, и использование «занятных» приме-

ров не идет ей на пользу. Нет нужды просеивать большие массивы данных, чтобы убедить нас в том, что слово «чупакабра» (кровососущее животное, которое якобы видели в Пуэрто-Рико в 1995 году) встречается намного реже, чем «снежный человек» или «лохнесское чудовище». Представляется также несерьезным проследить на графике меняющийся характер использования междометий «argh» и «aargh» в книгах, опубликованных приблизительно с 1940-х годов по 2000 год (начальную дату трудно определить по воспроизводимому в книге графику). На суперобложке книги приводится цитата из журнала *Mother Jones*, в которой приложение Ngram Viewer названо «самым большим расхитителем времени в истории интернета». Издатель поступил довольно смело, процитировав этот отзыв.

Документировать историю культуры, используя роботов для считывания каждого слова в каждой когда-либо опубликованной книге — масштабная задача. Что же я имею в виду, говоря, что многое остается за рамками этой работы? Айден и Мишель признают, что они ведут поиск лишь в малой выборке всех слов, и хотя, по их словам, Google уже отсканировал примерно 30 млн книг (теперь уже, наверное, больше), все еще остается около 100 миллионов.

Кроме того, если признать, что употребление слова может служить ключом к истории нашей культуры, в этой книге не рассматриваются многие из возможных источников: статьи в газетах и журналах, письма, кинофильмы, интервью по телевидению и радио, расшифровки записей, лекции, то есть, по сути, все написанное или сказанное, но не изданное в виде книг. Кроме того, книги после написания нередко редактируются и корректируются для устранения грамматических и орфографических ошибок, не говоря уже о переводе на другие языки. Каждому автору известно, что редакторы вносят изменения в текст в соответствии со стилистическими особенностями, принятыми в данном издательстве. Я не уверен, что тексты книг, даже если их 30 миллионов, могут служить достоверным источником информации о меняющемся словоупотреблении.

Наверное, Айден и Мишель возразили бы, что им приходится работать только с теми книгами, которые отсканировал Google, но учитывая недостающие элементы, их утверждение от-

носительно «окна в культуру человечества» может показаться слишком смелым.

Что можно сказать о графиках: их много, и они приводятся в упрощенной форме, в черно-бело-серой гамме, с построением диаграммы непосредственно из данных. В простых графиках как таковых нет ничего плохого, но относительный недостаток меток и координатных линий, а также то, что сами линии графиче-

В книге рассказывается о работе по сбору миллиардов слов, содержащихся во всех книгах мира.

ков (число которых может достигать шести) показаны почти не различающимися оттенками серого, существенно затрудняет чтение. Авторы отсылают нас к веб-источникам, где все эти проблемы решены: линии показаны контрастными цветами, и щелкнув на них в любой точке, можно увидеть метку и дату. Это может служить примером разницы между графическими изображениями в печати и интернете.

Но давайте смотреть на вещи позитивно. Авторы проявляют мужество, не пытаясь уклониться от обсуждения вопросов авторского права, возникающих при сканировании книг, или, казалось бы, неэтичных «теневых» способов обходить эти вопросы. В книге приводится прелестный рассказ об эксперименте, поставленном в 2002 году Ларри Пейджем и Мариссой Майер, которые выяснили, сколько может потребоваться времени для сканирования всех книг в мире. По-видимому, это заняло бы «тысячелетия, даже целые эпохи». Так как же авторы справились с этой проблемой? Читайте книгу; это очень занимательно!

Найджел Холмс,
Ведущий сайта *Explanation Graphics*,
автор — в числе последних книг
Wordless Diagrams и *The Book of Everything*

Всеобщее мерило цен, но не истинной ценности



Diane Coyle

GDP

A Brief but Affectionate History

Princeton University Press, Princeton, New Jersey, 2014, 168 pp., \$19.95 (cloth).

Почему такой умница, как Аристотель, не придумал понятие валового внутреннего продукта (ВВП) 2000 лет назад — ведь слово «экономика» происходит от греческого *oikos* (домашнее хозяйство)? Дело в том, что Аристотель занимался изучением движущихся объектов, например, луны и планет.

В течение приблизительно 50 000 лет ВВП оставался неподвижным. Во времена Аристотеля не было канала CNBC, который бы вызывал потрясения на рынках, сообщая о колебаниях ВВП. И никто не рассчитывал, что будет жить лучше, чем его родители. Конечно, людей беспокоили проблемы денег и долга. В первом десятилетии XVI века Генрих VIII приказал своим казначейм присматривать за его расходами в местном пабе, а также за его расходами на войны с Францией, но Генриху и в голову не пришло бы спрашивать, повысился ли ВВП на душу населения за прошедший год.

В умной и доходчивой новой книге Дианы Койл, *“GDP: A Brief but Affectionate History”* (ВВП: краткая история, рассказанная с любовью), идет речь об истории этой числовой конструкции двадцатого века, которая каждые три месяца грозит отставкой премьер-министрам. Койл, прежде всего, напоминает нам, что поставлено на карту, — не в древней Греции, а в современных Афинах, где руководитель статистического ведомства этой страны называет свою работу видом «боевых единоборств». Она делится рассказом своей подруги Паолы Субаччи из аналитического центра Chatham House, которая посетила это ведомство в Греции, ожидая увидеть суперкомпьютеры — или хотя

бы счеты. Вместо этого, поднявшись по лестнице жилого здания постройки 1950-х годов, она оказалась в «пыльной комнате, где находилось несколько человек», но совсем не было компьютеров.

Тем не менее, национальные статистические ведомства должны выдавать какой-то продукт, и они зачастую подтасовывают данные, стремясь продать облигации и уговорить других предоставить стране помощь. Койл высказывает мнение, что китайские чиновники иногда похваляются своим мощным ВВП, а иногда преуменьшают его, чтобы вписаться в критерии для получения помощи. После распада Советского Союза я посетил Санкт-Петербург в России. В моих старых учебниках по экономике говорилось, что СССР при коммунистической системе имел высокие темпы роста. Даже нобелевский лауреат Пол Самуэльсон публиковал такие сомнительные цифры. Но мне было достаточно вдохнуть затхлый воздух в вешающем музее Эрмитаж, чтобы понять: проблема коммунизма была не в его неспособности угнаться за Западом, а в том, что он не мог удержаться на уровне 1917 года!

Койл решает важную задачу, напоминая нам, что сама формула расчета ВВП ($C + I + G + [X - M]$), или потребление плюс инвестиции плюс государственные расходы плюс чистый экспорт, дает руководителем государства стимул тратить больше денег. Почему? Да потому что увеличение государственных расходов тавтологически повышает общую сумму. Все, что нужно сделать руководителю, это открыть кран расходов, и можно быть уверенным в том, что его счетоводы насчитают прибавку к ВВП. Кроме того, сумма государственных расходов исчисляется по заработной плате государственных служащих, а не стоимости их продукта. Один из моих студентов в Гарварде как-то высказал мнение, что, учитывая эту тавтологию, руководители, которые готовы хоть иногда подумать о введении строгой экономии бюджетных средств, заслуживают особых почестей.

Койл также прекрасно справляется с задачей анализа других недостатков ВВП, в том числе парадоксальных случаев, когда, например, вдовец, женясь на своей домработнице, снижает ВВП, поскольку теперь уже не платит ей жалованье. Особенно трудно правильно оценить услуги в информационной экономике. Последний пример из реальной жизни: я разработал новую числовую матрицу в качестве пособия для детей в изучении арифметики. Когда дети осваивают сложение при помощи этой матрицы, ко-

торая называется Math Arrow, они повышают свой учебный потенциал, скажем, на сто тысяч долларов. Но это приложение стоит 4,99 доллара. Возникает ли добавленная стоимость на сто тысяч долларов при каждом скачивании этой матрицы, или только в некоторых случаях?

После разбора недостатков ВВП Койл спрашивает, можем ли мы придумать что-то лучше, и приходит по списку конкурирующих показателей, включая индекс развития человеческого потенциала (ИРЧП), показатель экономического благосостояния и различные индексы счастья. Она с полным основанием дает им скептическую оценку, особенно показателям, распространяемым гуру, сулящими счастье, и демагогами. Уго Чавес называл ВВП «капиталистическим заговором». Но альтернативные показатели еще легче вывернуть в любую удобную сторону. Например, в 2009 году «Индекс счастливой планеты» вывел Коста-Рику на первое место среди всех стран; не намного от нее отстала Куба, занявшая седьмое место. Также, согласно этому индексу, жители Палестинской автономии счастливее и здоровее израильтян. Если бы представитель произраильских кругов стал доказывать, что палестинцы живут лучше, его бы подняли на смех или забросали камнями. И, кстати, США оказались в этом списке на 114-м месте. Странно, что мне не приходилось видеть лодок с эмигрантами, отплывающих из Майами на Кубу. Так что Койл правильно делает, что препарирует недостатки ВВП и в то же время выступает с предостережением относительно негативных последствий его возможной кончины.

Я заметил только одно упущение в ее в остальном кратком, но мастерски написанном трактате. Оценивая уровень жизни в той или иной стране, я часто задаю следующий простой вопрос: сколько часов нужно работать среднему работнику, чтобы купить курицу? В 1920-е годы президент Герберт Гувер в своей избирательной кампании обещал «по курице в каждой кастрюле». В те времена, чтобы заработать на курицу, требовалось два с половиной часа. Сегодня достаточно менее 15 минут. Я бы сказал, что это прогресс. Если, конечно, вы не курица.

Тодд Бухгольц,

бывший директор управления Белого дома по вопросам экономической политики, автор книги New Ideas from Dead Economists и исполнительный директор Sproglit, LLC (компания, разрабатывающей учебное программное обеспечение)