

ОГЛАВЛЕНИЕ

Предисловие	7
Благодарности	12
Глава 1. Что такое алгоритм?	13
Глава 2. Графы	38
Глава 3. Поиск	59
Глава 4. Сортировка	74
Глава 5. PageRank	97
Глава 6. Глубокое обучение	120
Послесловие	150
Глоссарий	158
Примечания	177
Предметный указатель	183
Библиография	185
Для дальнейшего чтения	190
Об авторе	191

ПРЕДИСЛОВИЕ

Я знаком с двумя подростками, которые обладают знаниями, невообразимыми для любых ученых, философов и исследователей прошлых веков. Это мои сыновья. Нет, я не из тех, кто любит похвастаться экстраординарными талантами своих детей. Но у этих двух парней есть карманные устройства, дающие им доступ к самому обширному хранилищу информации из когда-либо созданных. Теперь, когда они овладели искусством поиска по Интернету, нет такого фактологического вопроса, на который они не в состоянии ответить. Они переводят тексты с любого иностранного языка, не листая огромные словари (к слову, мы их не выбрасываем, чтобы наши дети знали, как все обстояло лишь несколько лет назад). Они мгновенно получают новости со всех уголков мира. Они общаются со своими сверстниками так, что вы даже не замечаете, независимо от того, какое расстояние их разделяет. Они детально планируют свои прогулки. С другой стороны, они могут тратить время на видеоигры или следовать модным тенденциям, которые меняются настолько быстро, что я даже не знаю, почему они так важны.

Все перечисленное выше стало возможным благодаря огромному прогрессу цифровых технологий. Устройства, которые мы сегодня носим в наших карманах, обладают большей вычислительной мощностью, чем те, с чьей помощью люди добрались до Луны. Как показывает пример с этими двумя подростками, наша жизнь претерпела огромные изменения; предсказания о будущем варьируются от утопий, в которых людям больше не придется работать, до антиутопий, где горстка счастливиц сможет наслаждаться полноценной жизнью, тогда как остальные окажутся обреченными на безысходность. К счастью, мы сами творцы своего будущего, и важную роль в этом играет то, насколько хорошо мы владеем технологиями, лежащими в основе потенциальных свершений и изменений. Мы живем в лучший период человеческой истории, хотя в суете повседневной жизни об этом можно забыть. Мы здоровее, чем когда-либо, и средняя продолжительность нашей жизни больше, чем у любого предыдущего поколения. Несмотря на несправедливость вопиющего неравенства, огромная часть человечества освободилась из оков нищеты. Мы еще никогда не были так близко друг к другу — как в буквальном, так и в виртуальном смысле. Можно сетовать на коммерциализацию массового глобального туризма, но дешевые путешествия позволяют нам знакомиться с другими культурами и посещать места, которыми раньше

мы любовались только в иллюстрированных журналах. Весь этот прогресс может и должен продолжаться.

Но, чтобы сделать свой вклад в этот прогресс, недостаточно просто использовать цифровые технологии. В них нужно разбираться. Во-первых, по сугубо практической причине: это открывает отличные карьерные возможности. Во-вторых, даже если вы не планируете работать в сфере технологий, вам нужно понимать их основополагающие принципы; это позволит как следует оценить их потенциал и поучаствовать в их развитии. Наряду с оборудованием (компонентами, из которых состоят компьютеры и цифровые устройства) не менее важной частью цифровых технологий является программное обеспечение — приложения, позволяющие выполнять на этом оборудовании ту или иную задачу. В основе приложений лежат реализуемые ими алгоритмы — наборы инструкций, которые описывают способы решения этих задач (не самое точное определение алгоритма, но не волнуйтесь: у нас есть целая книга, чтобы его уточнить). Без алгоритмов компьютеры были бы бесполезными, а все современные технологии попросту не существовали бы.

Необходимые знания меняются со временем. На протяжении значительной части истории человечества образование считалось необязательным. Большинство людей были неграмотными, но даже если их чему-то обучали, то только практическим навыкам и религиозным текстам. В начале девятнадцатого века более 80% мирового населения никогда не посещало школу; теперь же большинство людей обучались в ней хотя бы несколько лет. Ожидается, что к концу этого века число необразованных людей в мире упадет до нуля. Время, которое мы тратим на обучение, тоже выросло. Если в 1940 году менее 5% американцев имели степень магистра, то в 2015-м этот показатель почти достиг трети.

В девятнадцатом веке ни в одной школе не преподавали молекулярную биологию, так как о ней еще никто не знал; ДНК открыли лишь ближе к середине двадцатого века. И теперь это часть любой учебной программы. Не что похожее можно сказать об алгоритмах: их придумали еще в древние времена, но до изобретения современных компьютеров ими интересовались лишь немногие. Автор этой книги убежден, что мы достигли уровня, на котором алгоритмы являются одним из ключевых аспектов того, что мы считаем базовыми знаниями. Тот, кто не знает, что они представляют собой и как работают, не может оценить их потенциал и понять, как они влияют на нашу жизнь, чего от них следует ожидать, какие у них ограничения и что требуется для их работы. Поскольку наше общество все больше полагается на алгоритмы, нам, как осведомленным гражданам, следует в них ориентироваться.

Наряду с оборудованием (компонентами, из которых состоят компьютеры и цифровые устройства) не менее важной частью цифровых технологий является программное обеспечение — приложения, позволяющие выполнять на этом оборудовании ту или иную задачу. В основе приложений лежат реализуемые ими алгоритмы.

Изучение алгоритмов может иметь и другие преимущества. Если математика знакомит нас с формальным видом мышления, знание алгоритмов позволяет рассуждать о вещах новым, алгоритмическим образом, направленным на практическое решение задач. Благодаря этому эффективные реализации алгоритмов в виде программ могут быстро выполняться на компьютерах. Подход, в котором акцент делается на практичных и эффективных процессах проектирования, может пригодиться не только профессиональным программистам.

Эта книга призвана познакомить с алгоритмами читателя, далекого от компьютерных наук, и объяснить ему, как они работают. Мы не стремимся показать, как алгоритмы влияют на наши жизни, — с этим отлично справляются другие книги, в которых речь идет о том, как прогресс в обработке больших данных, искусственный интеллект и повсеместное внедрение вычислительных устройств могут изменить человеческую природу. Нас в первую очередь интересует не то, *что* может произойти в будущем, а скорее, *как* это может случиться. Для этого мы представим здесь настоящие алгоритмы и покажем не только то, что они делают, но и принцип их работы. Вместо общих фраз вас ждут подробные объяснения.

Знание алгоритмов позволяет рассуждать о вещах новым, алгоритмическим образом, направленным на практическое решение задач. Благодаря этому эффективные реализации алгоритмов в виде программ могут быстро выполняться на компьютерах.

На вопрос «что такое алгоритм?» есть удивительно простой ответ. Это определенный способ решения задачи, который можно описать в виде простых шагов и затем выполнить на компьютере с поразительной скоростью и эффективностью. Однако в таких решениях нет ничего волшебного. Тот факт, что они состоят из элементарных шагов, означает, что в них способны разобраться большинство людей.

На самом деле для чтения этой книги достаточно знаний, полученных в средней школе. Конечно, в некоторых главах есть немного математики, так как невозможно рассуждать об алгоритмах *вообще* без формул и формальных обозначений. В тексте объясняются любые концепции, распространенные в мире алгоритмов, но малоизвестные за пределами компьютерных наук.

Вот что написал недавно скончавшийся физик во введении в свою самую популярную книгу, «Краткая история времени», изданную в 1988 году: «Мне сказали, что каждая включенная в книгу формула уменьшит число покупателей вдвое». Это звучит довольно угрожающе для настоящей книги, так как математика в ней встречается не раз. Но я решил не обращать на это внимание по двум причинам. Во-первых, для понимания физики Хокинга требуется знание математики университетского уровня; здесь все намного доступнее. Во-вторых, эта книга призвана показать не только назначение алгоритмов, но и то, как они работают внутри, поэтому читателю должны быть знакомы некоторые термины, используемые здесь, в том числе математические. Система обозначений не является прерогативой технической интеллигенции, и владение ею поможет развеять ореол загадочности, окружающий данную тему; в конце вы увидите, что это в основном сводится к умению рассуждать о вещах четко и поддающимся измерению способом.

Книга вроде этой не может полностью охватить тему алгоритмов, но мы попытаемся провести краткий обзор и познакомить читателя с алгоритмическим образом мышления. Основа будет заложена в первой главе, из нее вы узнаете, что такое алгоритмы и как измерить их эффективность. Сразу отметим, что алгоритм — это конечная последовательность шагов, которые можно выполнить с помощью ручки и листа бумаги, и это простое определение не так уж далеко от истины. Дальше в главе 1 будет исследована взаимосвязь между алгоритмами и математикой. Отличаются же они практичностью: при написании алгоритмов нас интересуют прикладные способы решения задач. Это означает, что нам нужно как-то оценивать практичность и эффективность наших алгоритмов. Вы увидите, что эти аспекты можно рассматривать в строго очерченных рамках вычислительной сложности; именно в этом контексте будет проходить обсуждение алгоритмов на страницах данной книги.

Следующие три главы посвящены трем наиболее важным сферам применения алгоритмов. В главе 2 рассматриваются алгоритмы для решения задач, связанных с сетями, известными как графы. В число этих задач входит поиск маршрута на дороге или последовательности звеньев, соединяющих вас с другими людьми в социальной сети. Сюда же отнесем задачи из других областей с не такими очевидными связями: секвенирование ДНК и планирование соревнований; это проиллюстрирует тот факт, что разного рода задачи могут быть эффективно решены с помощью одних и тех же инструментов.

В главах 3 и 4 вы узнаете, как искать и упорядочивать элементы. Тема может показаться прозаической, но она одна из самых важных с точки зрения

практического применения компьютеров. Компьютеры тратят много времени на сортировку и поиск, однако мы редко обращаем на это внимание, поскольку указанные операции — неотъемлемая и невидимая часть большинства приложений. Сортировка и поиск даже дают некоторое представление о важном аспекте алгоритмов. Многие задачи можно решать разными способами. Выбор подходящего алгоритма зависит от его характеристик; некоторые алгоритмы больше подходят для определенных задач, чем другие. Поэтому вам полезно увидеть, как разные алгоритмы с разными характеристиками справляются с решением одной и той же задачи.

В следующих двух главах представлены важные способы применения алгоритмов в широких масштабах. В главе 5 мы снова вернемся к графам, чтобы объяснить алгоритм, который можно применять для ранжирования веб-страниц в порядке их значимости. PageRank использовался компанией Google в момент ее основания. Успех этого алгоритма в ранжировании результатов поиска сыграл ключевую роль в феноменальном успехе Google как компании. К счастью, разобраться в его работе не так уж сложно. Это позволит увидеть, как алгоритмы справляются с задачей, которую, на первый взгляд, нельзя решить с помощью компьютеров: как оценить важность тех или иных вещей?

В главе 6 будет представлена одна из самых динамично развивающихся областей компьютерных наук: нейронные сети и глубокое обучение. Об успешном применении нейронных сетей можно узнать даже из популярных средств массовой информации. Наибольший интерес проявляется к сюжетам о системах, которые выполняют такие задачи, как анализ изображений, автоматический перевод и медицинская диагностика. Мы начнем с отдельных нейронов и постепенно перейдем к построению все больших и больших нейронных сетей, способных выполнять все более сложные задачи. Вы увидите, что все они основаны на общих фундаментальных принципах. Их эффективность объясняется взаимосвязанностью множества простых компонентов и применением алгоритма, который позволяет нейронным сетям учиться.

Вслед за демонстрацией возможностей алгоритмов идет послесловие, посвященное тому, как ограничены компьютерные вычисления. Мы знаем, что компьютеры способны на многое, и ожидаем от них еще большего, но есть вещи, которые им не под силу. Обсуждение пределов их возможностей позволит более точно объяснить природу алгоритмов и вычислений. Мы уже говорили, что алгоритм можно описать в виде конечной последовательности шагов, которые можно выполнить с помощью ручки и листа бумаги, но какими могут быть эти шаги? И насколько близка аналогия с ручкой и бумагой к настоящим алгоритмам?

БЛАГОДАРНОСТИ

Прежде всего, благодарю Мари Лафкин Ли из издательства MIT Press — ей принадлежит идея создания этой книги; Стефани Коэн, которая мягко подгоняла меня в процессе написания; Синди Милстрейн за ее тщательную редактуру и Вирджинию Кроссмэн за превосходное внимание к деталям и заботу обо всех аспектах этой книги. Книга об алгоритмах должна войти в цикл *Essential Knowledge* (необходимые знания), и я горжусь тем, что являюсь ее автором.

Я признателен Диомидису Спинеллису за его комментарии к разным частям книги. Отдельное спасибо Константиносу Маринакосу, который прочитал мой черновик, нашел ошибки, за которые мне теперь стыдно, и подсказал, как их исправить.

Наконец, моя огромная благодарность двум подросткам, Адрианосу и Эктору, чьи жизни в какой-то степени будут определяться темой данной книги, и их матери, Элени; эта работа проделана благодаря им.

ЧТО ТАКОЕ АЛГОРИТМ?

Алгоритмическая эра

Мы любим давать названия временным периодам — наверное, потому, что это позволяет нам совладать со скоротечностью времени. Поэтому мы начали говорить о настоящем как о новой *алгоритмической эре*, в которой алгоритмам отводится все более важная роль. Интересно, что речь больше не идет об *эпохе компьютеров* или *Интернета*. Их мы воспринимаем как само собой разумеющееся. Но именно алгоритмы дают нам ощущение качественного сдвига. «Вот он, всемогущий алгоритм, отрывок компьютерного кода, претендующий на высшую власть в нашу светскую эпоху, своего рода бог», — пишет Кристофер Лайдон, бывший журналист в *New York Time*, ведущий *Radio Open Source*. И в самом деле, алгоритмы воспринимают как некую вышестоящую инстанцию, которая участвует в организации политических кампаний, отслеживает нашу активность онлайн, помогает нам с покупками, показывает рекламу, подсказывает, с кем пойти на свидание, следит за нашим здоровьем.

Вокруг всего этого существует аура таинственности, которая, наверное, льстит служителям алгоритмов. Должность «программиста» или «специалиста в области информационных технологий» делает вас достойным человеком, хоть и технарем. Приятно принадлежать к племени, стоящему в шаге от того, чтобы перевернуть все с ног на голову, не так ли?

Алгоритмам, несомненно, свойственно нечто божественное. Как и боги, они в основном выходят сухими из воды; события случаются не из-за человеческой деятельности, а потому, что так решил алгоритм и с алгоритма спроса нет. Компьютеры, которые выполняют алгоритмы, превосходят человека во все большем количестве отраслей, поэтому кажется, что наша роль

уменьшается день ото дня; кто-то верит, что не за горами день, когда компьютеры превзойдут нас во всех аспектах умственной деятельности.

Но, если посмотреть на вещи с другой стороны, алгоритмы совсем не похожи на богов, и мы часто об этом забываем. Результаты, которые производят алгоритмы, не являются откровением свыше. Мы в точности знаем, каким правилам они подчиняются и какие шаги они выполняют. Каким бы чудесным ни был результат, его всегда можно свести к элементарным операциям. Люди, для которых алгоритмы в новинку, могут удивиться тому, насколько тривиальными они бывают внутри. Это вовсе не умаляет их значимость; просто понимание того, что происходит за кулисами, снимает налет таинственности. Но в то же время это позволяет по достоинству оценить, насколько элегантно написан алгоритм.

Идея этой книги в том, что в алгоритмах и в самом деле нет ничего таинственного. Это инструменты, позволяющие нам легко справляться с определенными задачами; они заточены на решение проблем. В этом смысле их можно считать «умственными» инструментами, включающими числа и арифметику. У нас ушли тысячелетия на выработку системы счисления, с помощью которой дети в школе могут решать математические задачи; без нее эти задачи остались бы нерешенными. Мы считаем этот навык само собой разумеющимся, но несколько поколений назад им владела лишь небольшая часть людей.

Точно так же знание алгоритмов не должно быть прерогативой крошечного элитного меньшинства; это умственные инструменты, которыми может овладеть кто угодно, а не только специалисты по компьютерам. Более того, этот навык *должен* быть доступен большому числу людей, так как он позволяет взглянуть на алгоритмы в более широком контексте: понять, что они делают, как они это делают и чего от них можно ожидать.

Основная цель этой книги — дать базовое понимание алгоритмов, которое позволит принять полноценное участие в дискуссиях, происходящих в алгоритмическую эру. Начало этой эре мы положили сами, используя ранее разработанные инструменты. Изучение этих инструментов и будет темой данной книги. Алгоритмы прекрасны, и представление о том, как они создаются и работают, может улучшить наш собственный образ мышления.

Но для начала развеим распространенное заблуждение о том, что алгоритмы неотделимы от компьютеров. Это так же неверно, как связывать числа с калькуляторами.

Способ решения задач

Головоломки, музыка, делимость чисел и ускорители нейтронов в физике элементарных частиц — вы увидите, что у всего этого есть общий алгоритм, который применяется в разных предметных областях, но при этом основан на тех же фундаментальных принципах. Как так?

Само слово «алгоритм» не раскрывает своего значения в полной мере. Оно происходит от имени Мухаммада ибн Мусы аль-Хорезми (ок. 783 — ок. 850), персидского ученого, который работал в областях математики, астрономии и географии. Аль-Хорезми внес большой вклад во многие сферы науки. Термин «алгебра» происходит от арабского названия его самой фундаментальной работы, «Краткая книга о восполнении и противопоставлении». Его вторая по известности работа, «Книга об индийском счете», была посвящена арифметике и переведена на латынь; именно из нее европейцы узнали об индо-арабской системе счисления. Имя Аль-Хорезми в латинском варианте выглядит как *Algorismus*; его стали использовать для обозначения численных методов с десятичными числами. Впоследствии термин *Algorismus* под влиянием греческого слова *arithmos* (ἀριθμός — число) превратился в *algorithm* (алгоритм) и все так же обозначал десятичную арифметику, пока в девятнадцатом веке не обзавелся современным значением.

Многие думают, что алгоритмы имеют какое-то отношение к компьютерам, но это не так. Они существовали задолго до появления компьютеров. Первый алгоритм, о котором нам известно, существовал еще во времена древнего Вавилона. Кроме того, алгоритмы решают не компьютерные проблемы. Они описывают, что и как нужно делать, в виде последовательности шагов. Звучит немного туманно. Вы можете спросить, о каких шагах идет речь? Мы могли бы внести ясность и дать четкое математическое определение самого алгоритма и того, что он делает (оно действительно существует), но это было бы излишним. Вам, скорее всего, будет достаточно знать, что алгоритм — это набор шагов, которые можно выполнить с помощью ручки и листа бумаги; несмотря на внешнюю простоту, это определение достаточно близко к тем, которые используют математики и специалисты в области информатики.

Начнем наше знакомство с алгоритмами с задачи, которую можно решить вручную. Представьте, что у нас есть два множества объектов и мы хотим как можно равномернее распределить объекты одного множества среди объектов другого. Объекты в первом множестве будут обозначаться как \times , а во втором — как \bullet . Мы хотим распределить крестики между точками.

Многие думают, что алгоритмы имеют какое-то отношение к компьютерам, но это не так. Они существовали задолго до появления компьютеров.

Если количество крестиков точно делится на количество точек, все просто: крестики между точками достаточно разместить так, как будто мы выполняем деление. Например, если всего у нас 12 объектов, из которых три — это крестики, а остальные девять — точки, мы рисуем один крестик, затем три точки, затем один крестик, затем три точки и, наконец, еще один крестик и три точки.

× • • • × • • • × • • •

Но что, если общее количество объектов нельзя точно разделить на число крестиков? Что, если крестиков у нас пять, а точек семь?

Вначале мы размещаем в один ряд все крестики и затем все точки, как показано ниже:

× × × × × • • • • • • •

После этого помещаем пять точек снизу от крестиков:

× × × × × • •
• • • • •

В результате справа остаются два столбца с одной точкой в каждом. Поместим их под двумя первыми столбцами, образуя третью строку:

× × × × ×
• • • • •
• •

Теперь можно заметить три оставшихся столбца. Два из них (крайние справа) помещаем под двумя первыми столбцами:

× × ×
• • •
• •
× ×
• •

Остается всего один лишний столбец, поэтому мы останавливаемся. Объединим столбцы, перемещаясь слева направо, и получим:

× • • × • × • • × • × •

Это наш результат. Мы распределили крестики между точками. Не так равномерно, как раньше, но это закономерно, ведь, как вы помните, пять