

УДК 004
ББК 32.973.2
К36

Understanding the Digital World:
What You Need to Know about Computers,
the Internet, Privacy, and Security, Second Edition
by Brian W. Kernighan.

Copyright © 2021 by Princeton University Press

All rights reserved.

No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without permission in writing from the Publisher.

Керниган, Брайан.

К36 Основы информационных технологий для неспециалистов: что происходит внутри машин / Брайан Керниган ; [перевод с английского Е. В. Жевлаковой]. — Москва : Эксмо, 2024. — 624 с. — (Библиотека ИТ. Главные книги о современных технологиях).

ISBN 978-5-04-184251-2

Компьютеры окружают нас повсюду, включая бытовую технику, автомобили, медицинское оборудование, транспортные системы, электросети и оружие. Однако большинство из них остаются невидимыми, собирая и иногда сливая наши личные данные. Это делает нас уязвимыми для правительств, компаний и преступников, которые могут использовать информацию в своих целях.

Второе издание популярной книги «Цифровой мир» Брайана Кернигана рассматривает принципы работы компьютерного оборудования, программного обеспечения и сетей. Новые разделы посвящены программированию на Python, большим данным, машинному обучению и многому другому.

УДК 004
ББК 32.973.2

ISBN 978-5-04-184251-2

© Жевлакова Е.В., перевод на русский язык, 2024
© Оформление. ООО «Издательство «Эксмо», 2024

Содержание

Предисловие	13
Введение	26
ЧАСТЬ I. АППАРАТНОЕ ОБЕСПЕЧЕНИЕ	39
1. Что такое компьютер?	46
1.1. Логическое построение	51
1.1.1. Процессор	52
1.1.2. Оперативная память	54
1.1.3. Внешняя память	57
1.1.4. Другие устройства	60
1.2. Механическая конструкция	62
1.3. Закон Мура	68
1.4. Краткие выводы	71
2. Биты, байты и формат записи информации	74
2.1. Различия аналоговой и цифровой форм	75
2.2. Аналого-цифровое преобразование	78
2.2.1. Оцифровка изображений	78
2.2.2. Оцифровка звука	80
2.2.3. Оцифровка фильмов	86
2.2.4. Оцифровка текста	87
2.3. Биты, байты и двоичная система исчисления	89
2.3.1. Биты	90
2.3.2. Степени чисел 2 и 10	92
2.3.3. Двоичные числа	93
2.3.4. Байты	97
2.4. Краткие выводы	101

3. Процессор изнутри	103
3.1. Компьютер-игрушка	104
3.1.1. Первая программа компьютера-игрушки	106
3.1.2. Вторая программа компьютера-игрушки	108
3.1.3. Инструкции ветвления	110
3.1.4. Представление в памяти	115
3.2. Настоящие процессоры	117
3.3. Кэширование	121
3.4. Другие виды вычислительных устройств	124
3.5. Краткие выводы	129
<i>Краткое заключение по аппаратному обеспечению</i>	131
ЧАСТЬ II. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ	135
4. Алгоритмы	141
4.1. Линейные алгоритмы	143
4.2. Двоичный поиск	147
4.3. Сортировка	151
4.4. Трудности и сложности	158
4.5. Краткие выводы	164
5. Программирование и языки программирования	167
5.1. Ассемблерный (сборочный) язык	169
5.2. Языки высокого уровня	171
5.3. Разработка программного обеспечения	185
5.3.1. Библиотеки, интерфейсы и средства разработки	187
5.3.2. Ошибки (баги)	190
5.4. Интеллектуальная собственность	195
5.4.1. Коммерческая тайна	196
5.4.2. Товарный знак	197
5.4.3. Авторское право	197
5.4.4. Патент	199
5.4.5. Лицензии	202
5.5. Стандарты	207
5.6. Программное обеспечение с открытым исходным кодом	209
5.7. Краткие выводы	214

6. Программные системы	216
6.1. Операционные системы	218
6.2. Как работает операционная система	229
6.2.1. Системные вызовы	231
6.2.2. Драйверы устройств	232
6.3. Другие операционные системы	234
6.4. Файловые системы	235
6.4.1. Файловые системы внешней памяти	238
6.4.2. Удаление файлов	241
6.4.3. Другие файловые системы	245
6.5. Приложения	248
6.6. Уровни программного обеспечения	252
6.7. Краткие выводы	256
7. Учимся программировать	258
7.1. Принципы языков программирования	261
7.2. Первая программа на JavaScript	263
7.3. Вторая программа на JavaScript	265
7.4. Циклы и условия	268
7.5. Библиотеки и интерфейсы JavaScript	273
7.6. Как работает JavaScript	275
7.7. Первая программа на Python	276
7.8. Вторая программа на Python	279
7.9. Библиотеки и интерфейсы Python	282
7.10. Как работает Python	286
7.11. Краткие выводы	287
<i>Краткое заключение по программному обеспечению</i>	289
ЧАСТЬ III. КОММУНИКАЦИИ	293
8. Сети	304
8.1. Телефоны и модемы	306
8.2. Кабель и DSL	307
8.3. Локальные сети и Ethernet	311
8.4. Беспроводные сети	315
8.5. Мобильные телефоны	319
8.6. Пропускная способность	326

8.7. Сжатие	328
8.8. Обнаружение и исправление ошибок	334
8.9. Краткие выводы	337
9. Интернет	340
9.1. Обзор интернета	343
9.2. Доменные имена и адреса	348
9.2.1. Система доменных имен	350
9.2.2. IP-адреса	352
9.2.3. Корневые серверы	353
9.2.4. Регистрация вашего собственного домена	355
9.3. Маршрутизация	357
9.4. Протоколы TCP/IP	361
9.4.1. Интернет-протокол (IP)	363
9.4.2. Протокол управления передачей (TCP)	365
9.5. Протоколы более высокого уровня	368
9.5.1. Telnet и SSH: удаленный вход в систему	369
9.5.2. SMTP: простой протокол передачи почты	371
9.5.3. Обмен файлами и одноранговые протоколы	375
9.6. Авторское право в интернете	377
9.7. Интернет вещей	381
9.8. Краткие выводы	384
10. Всемирная паутина	389
10.1. Как работает Всемирная паутина	392
10.2. HTML	394
10.3. Куки-файлы	398
10.4. Активный контент на веб-страницах	401
10.5. Активный контент в других местах	404
10.6. Вирусы, черви и троянские кони	406
10.7. Веб-безопасность	411
10.7.1. Атаки на клиентов	411
10.7.2. Атаки на серверы	421
10.7.3. Атаки на передаваемую информацию	424
10.8. Как защитить себя	427
10.9. Краткие выводы	433

ЧАСТЬ IV. ДАННЫЕ	437
11. Данные и информация	440
11.1. Поиск	442
11.2. Отслеживание	450
11.3. Социальные сети	464
11.4. Интеллектуальный анализ и агрегирование данных	472
11.5. Облачные вычисления	477
11.6. Краткие выводы	487
12. Искусственный интеллект и машинное обучение	490
12.1. Историческая справка	493
12.2. Классическое машинное обучение	496
12.3. Нейронные сети и глубокое обучение	501
12.4. Обработка естественного языка	505
12.5. Краткие выводы	511
13. Конфиденциальность и безопасность	516
13.1. Криптография	518
13.1.1. Криптография с секретным ключом	522
13.1.2. Криптография с открытым ключом	525
13.2. Анонимность	533
13.2.1. Конфиденциальность в Сети	535
13.2.2. Биткоин	540
13.3. Краткие выводы	544
14. Что дальше?	548
Примечания	559
Глоссарий	574
Алфавитный указатель	594

Посвящается Мэг

Предисловие

Курс под названием «Компьютеры в нашем мире» я преподаю в Принстоне почти каждую осень начиная с 1999 года. Название курса — размытое до неприличия, но дело в том, что однажды мне пришлось придумать его за пять минут, а поменять с тех пор стало довольно сложно. Впрочем, преподавать этот курс — самая увлекательная часть моей работы, которая и так почти всегда приносит радость.

Факт, что компьютеры и вычислительная техника окружают нас повсюду. Некоторые технологии у всех на виду: у студентов есть ноутбуки, и они намного мощнее компьютера IBM 7094¹, который стоил несколько миллионов долларов, занимал большую комнату с кондиционером и обслуживал весь кампус Принстона, когда я пришел сюда аспирантом в 1964 году. Кроме того, каждый студент пользуется сотовым телефоном, вычислительная мощность которого намного выше, чем у той древней машины IBM. У всех студентов, как и у значительной доли населения земного шара, есть высокоскоростной доступ в интернет. Любой из них может искать информацию в браузере, совершать покупки онлайн, пользоваться электронной почтой, мессенджерами и социальными сетями для общения с друзьями и семьей.

Но это только вершина айсберга вычислительных технологий, большая часть скрыта в глубине. Мы обычно не думаем о компьютерах, которые незаметно работают

в бытовой технике, машинах, самолетах и вездесущих электронных устройствах вроде умных телевизоров, термостатов и дверных звонков, средств распознавания голоса, фитнес-трекеров, наушников, игрушек и игровых аксессуаров, — и мы воспринимаем их как данность. Мы мало задумываемся и о том, до какой степени от вычислительной техники зависит инфраструктура: телефонная сеть, кабельное телевидение, управление воздушным движением, энергосистема, банковские и финансовые услуги.

Большинство людей напрямую не создают эти системы, но ощущают их сильное влияние — а некоторым предстоит принимать важные решения по поводу их работы. Образованный человек должен знать по меньшей мере основы вычислительных технологий: что делают компьютеры и как, на что они не способны в принципе, а что просто чересчур сложно для них на данный момент, как они общаются между собой и что при этом происходит, какими многочисленными способами технологии и коммуникации воздействуют на мир вокруг нас.

То, что информационные технологии проникают повсюду, влияет на нас неожиданным образом. Хотя нам время от времени напоминают о распространении систем наблюдения, о вторжениях в нашу частную жизнь, угрозах кражи личных данных, мы не всегда осознаем, насколько это все обусловлено технологиями и коммуникациями.

В июле 2013 года Эдвард Сноуден, подрядчик Агентства национальной безопасности (АНБ) Соединенных Штатов Америки, предоставил журналистам 50 000 документов, в которых раскрывалось, что сотрудники АНБ регулярно отслеживали электронные средства связи: звонки, сообщения, письма, интернет — и собирали данные почти по каждому жителю мира. Что примечательно, они следили и за американцами, которые жили в своей стране и не представляли

никакой угрозы ее безопасности. Документы Сноудена показывали, что и другие государства тоже шпионили за своими гражданами. Удивительно, но после первоначального возмущения общественности все вернулось на круги своя, правительства стали еще более масштабно отслеживать, чем занимаются граждане, и шпионить за ними, а люди либо смирились с этим, либо не возражают по неведению.

Нашу деятельность в интернете и реальном мире отслеживают также и корпорации. Бизнес-модели многих компаний основаны на обширном сборе данных, возможностях прогнозировать наше поведение и влиять на него. Доступность громадных объемов данных позволила достичь серьезного прогресса в машинном распознавании речи и образов, а также автоматических переводах. Но она вредит неприкосновенности частной жизни, и теперь сложно сохранять анонимность.

Хакеры всех мастей стали особенно изощренными в своих атаках на хранилища данных. В государственных службах и на предприятиях почти ежедневно происходят электронные взломы; данные о клиентах и сотрудниках похищают в больших количествах и часто используют в целях мошенничества и кражи личной информации. Обычными стали атаки на отдельных лиц. Раньше считалось, что для защиты от онлайн-мошенников достаточно игнорировать письма от мнимых нигерийских принцев или их родственников. Но сейчас целевые атаки стали куда более хитроумными, и это один из наиболее распространенных способов взлома корпоративных компьютеров.

Социальные сети вроде Facebook*, Instagram*, X (Twitter), Reddit и других изменили способы общения между

* Принадлежат компании Meta, признанной экстремистской и запрещенной на территории РФ.

людьми. С одной стороны, это полезно: мы можем поддерживать связь с друзьями и семьей, просматривать новости и всевозможные развлекательные посты. Соцсети могут и помочь, как в середине 2020 года, когда движение Black Lives Matter привлекло всеобщее внимание благодаря «вирусным» видеороликам о жестокости полиции.

Но через социальные сети также распространялось множество дурных идей. Расисты, участники групп ненависти, сторонники теорий заговора и другие безумцы, независимо от их убеждений и политических взглядов, могут легко находить друг друга в интернете, координировать свои действия и усиливать свое влияние. Сдерживать разгул ненависти и бреда мешают щекотливые вопросы свободы слова, а также технологические сложности модерирования контента.

В мире с обширным взаимодействием через интернет сложно решать вопросы юрисдикции. В 2018 году Европейский союз ввел в действие Общий регламент по защите персональных данных, который позволил жителям ЕС контролировать сбор и использование своих данных и запретил компаниям отправлять или хранить данные за пределами ЕС. Коллегия присяжных еще не определила, насколько Общий регламент оказался полезным для конфиденциальности. К тому же эти правила применяются только в ЕС, а в других частях мира действуют иные законы.

Еще одна сложность появилась после быстрого внедрения облачных вычислений, когда отдельные лица и компании стали хранить данные и обрабатывать их на серверах, принадлежащих Amazon, Google, Microsoft и т. д. Данные теперь хранятся не только у их владельца, но и у третьих лиц, у которых есть свои цели, обязанности, уязвимые места и неразрешенные проблемы с взаимноисключающими требованиями разных юрисдикций.

Быстро развивается «интернет вещей», в котором все виды устройств подключаются к сети. Смартфоны — это очевидный пример, но, кроме них, онлайн-доступ получают автомобили, камеры слежения, бытовая техника и средства управления домом, медицинское оборудование и большая часть инфраструктуры — например, управление воздушным движением и энергосети. Эта тенденция соединять все с интернетом будет продолжаться, потому что преимущества такого соединения неоспоримы. К сожалению, здесь есть и серьезные риски: некоторые устройства контролируют жизненно важные системы, а не только развлечения, и часто их защита намного слабее, чем у более зрелых технологий.

Криптография — одно из немногих эффективных средств защиты, поскольку она обеспечивает конфиденциальность и безопасность как коммуникации, так и хранения данных. Но надежная криптография подвергается постоянным атакам. Правительствам не нравится идея, что люди, компании или террористы получают возможность общаться по-настоящему тайно, поэтому они часто призывают внедрить в криптографические алгоритмы лазейки, которые позволили бы госслужбам взломать шифрование — разумеется, с «надлежащей страховкой» и только «в интересах национальной безопасности». Какими бы благими ни были их намерения, это очень плохая идея. Даже если вы верите, что правительство всегда будет вести себя достойно и не позволит утечь секретной информации (забудем о Сноудене), ослабленная криптография поможет не только вашим союзникам, но и противникам. А плохие парни все равно не будут ее применять.

Это лишь некоторые проблемы и вопросы, которые должны беспокоить обычных граждан на улицах,

вроде студентов моего курса или, как говорится, «людей просвещенных», независимо от их происхождения и образования.

Студенты моего курса — не технические специалисты. Они не инженеры, не физики, не математики. Они — бакалавры английского языка и политологии, историки, классицисты, экономисты, музыканты и артисты. Превосходная выборка по гуманитарным и общественным наукам. К окончанию курса эти одаренные люди должны научиться читать и понимать новостные статьи на компьютерную тематику, узнавать из них что-то новое и, возможно, находить неточности. В более широком смысле я хочу, чтобы мои студенты и читатели разумно и скептически относились к технологиям, понимали, что они полезны, но вовсе не панацея. И наоборот, понимали, что, хотя технологии иногда приводят к плохим последствиям, они вовсе не абсолютное зло.

Ричард Мюллер в своей прекрасной книге «Физика для будущих президентов»² пытается объяснить научно-техническую подоплеку, лежащую в основе проблем, с которыми сталкиваются лидеры: ядерная угроза, терроризм, энергетические кризисы, глобальное потепление и так далее. Эрудированные люди, хоть и не претендующие на пост президента, тоже должны разбираться в этих темах. Если опираться на подход Мюллера, то и мое направление работы можно определить схоже: «Компьютеризация для будущих президентов».

Что будущий президент должен знать о компьютеризации? Что должен знать об этом эрудированный человек — например, вы?³

Я выделяю четыре основные технические области: аппаратное обеспечение, программное обеспечение, коммуникации и данные.