

# Оглавление

Предисловие научного редактора русского издания (В. И. Городецкий).....	5
Предисловие автора к русскому изданию .....	11
Благодарности.....	13
Предисловие.....	15
<b>Часть I. Концепты и мышление .....</b>	<b>18</b>
Глава 1. Эпоха науки о данных .....	19
1.1. Введение .....	19
1.2. Характерные черты эпохи данных .....	22
1.2.1. Некоторые базовые термины науки о данных.....	22
1.2.2. Дискуссии в эпоху данных .....	22
1.2.3. Знаковые черты и тенденции эпохи данных.....	23
1.3. Эволюция науки о данных .....	29
1.3.1. Новое поколение информационных продуктов и экономика .....	31
1.4. Как далеко простираются возможности данных .....	34
1.4.1. Потенциал данных .....	34
1.4.2. Движущие силы, мотивирующие использование данных .....	38
1.5. Новое X-поколение .....	39
1.5.1. X-сложность.....	39
1.5.2. X-интеллект.....	41
1.5.3. X-возможности .....	42
1.6. Тенденции интереса .....	43
1.7. Основные стратегии правительств относительно данных .....	44
1.7.1. Правительственные инициативы в области больших данных.....	44
1.7.2. Австралийские инициативы .....	46
1.7.3. Китайские инициативы .....	47
1.7.4. Европейские инициативы.....	47
1.7.5. Инициативы США .....	48
1.7.6. Другие правительственные инициативы.....	49
1.8. Научная повестка дня науки о данных .....	49
1.8.1. Правительственные инициативы .....	50
1.8.2. Исследовательские инициативы в науке о данных .....	50
1.9. Резюме .....	51
Глава 2. Что такое наука о данных .....	53
2.1. Введение .....	53
2.2. Датафикация и квантификация данных .....	53
2.3. Данные, информация, знания, интеллект и мудрость .....	55

2.4. ДНК данных.....	58
2.4.1. Что такое ДНК данных .....	58
2.4.2. Функции ДНК данных .....	59
2.5. Различные взгляды на науку о данных.....	59
2.5.1. Взгляд на науку о данных со стороны статистики .....	60
2.5.2. Мультидисциплинарный взгляд на науку о данных.....	60
2.5.3. Наука о данных: взгляд с позиций данных .....	61
2.6. Различные определения науки о данных .....	62
2.6.1. Высокоуровневое определение науки о данных.....	63
2.6.2. Трансдисциплинарное определение науки о данных .....	63
2.6.3. Процессное определение науки о данных.....	64
2.6.3.1. Образ мышления на уровне мудрости .....	66
2.6.3.2. Понимание предметной области .....	67
2.6.3.3. Управление данными .....	68
2.6.3.4. Вычисления с использованием данных.....	68
2.6.3.5. Обнаружение знаний.....	69
2.6.3.6. Коммуникации с заинтересованными сторонами .....	70
2.6.3.7. Доставка информационных продуктов .....	71
2.6.3.8. Действия на основании результатов .....	71
2.7. Открытая модель, открытые данные и открытая наука.....	72
2.7.1. Открытая модель .....	72
2.7.2. Открытые данные.....	73
2.7.3. Открытая наука .....	75
2.8. Информационные продукты .....	77
2.9. Мифы и заблуждения .....	78
2.9.1. Возможные негативные эффекты науки о данных.....	79
2.9.2. Концептуально ошибочные представления .....	80
2.9.3. Ошибочные представления, касающиеся объема данных .....	82
2.9.4. Ошибочные представления об инфраструктуре данных.....	83
2.9.5. Ошибочные представления об аналитике.....	84
2.9.6. Ошибочные представления о возможностях и роли науки о данных...	87
2.9.7. Другие вопросы .....	88
2.10. Резюме .....	91
Глава 3. Мышление в области науки о данных .....	92
3.1. Введение .....	92
3.2. Научное мышление .....	93
3.2.1. Научное мышление vs ненаучное мышление.....	93
3.2.2. Творческое мышление vs логическое мышление .....	96
3.2.2.1. Логическое мышление .....	96
3.2.2.2. Творческое мышление .....	97
3.2.2.3. Критическое мышление .....	98
3.2.2.4. Нестандартное мышление .....	99
3.3. Структура науки о данных.....	100
3.4. Наука о данных как сложная система.....	103
3.4.1. Систематический взгляд на проблемы науки о данных.....	103
3.4.2. Сложности в системах науки о данных .....	106
3.4.3. Основы образа мышления науки о данных .....	108
3.4.4. Мыслительное крыло науки о данных .....	109

3.4.5. Надзорное крыло науки о данных .....	110
3.4.6. Исходный контекст науки о данных .....	110
3.4.7. Разработка механизмов для науки о данных .....	112
3.4.8. Уровень доставки результата потребителю.....	113
3.4.9. Гарантии науки о данных .....	113
3.5. Критическое мышление в науке о данных .....	114
3.5.1. Перспективы критического мышления .....	114
3.5.2. Мы не знаем, чего мы не знаем .....	115
3.5.3. Научные исследования, управляемые данными .....	119
3.5.3.1. Что такое исследование, управляемое данными? .....	119
3.5.3.2. Научные исследования, управляемые данными, и их сравнение с разработками на основе модели .....	120
3.5.4. Парадигма управления на основе данных и другие парадигмы .....	123
3.5.4.1. Различные гибридные парадигмы .....	123
3.5.4.2. Исследование на основе данных и предметной области.....	125
3.5.5. Основные вопросы, которые задают в науке о данных при решении проблем.....	128
3.6. Резюме.....	131
<b>Часть II. Вызовы и фундаментальные основы.....</b>	<b>132</b>
Глава 4. Вызовы науки о данных .....	133
4.1. Введение .....	133
4.2. X-сложности в науке о данных.....	134
4.2.1. Сложность данных .....	134
4.2.2. Поведенческая сложность .....	135
4.2.3. Сложность предметной области.....	135
4.2.4. Социальная сложность .....	136
4.2.5. Сложность окружающей среды .....	137
4.2.6. Сложность взаимодействия человека и компьютера .....	137
4.2.7. Сложность обучения.....	138
4.2.8. Сложность предоставления результатов .....	139
4.3. X-интеллект в науке о данных .....	140
4.3.1. Интеллект данных.....	140
4.3.2. Поведенческий интеллект .....	141
4.3.3. Интеллект предметной области.....	141
4.3.4. Человеческий интеллект .....	142
4.3.5. Сетевой интеллект .....	142
4.3.6. Интеллект организации .....	143
4.3.7. Социальный интеллект .....	143
4.3.8. Интеллект окружающей среды.....	144
4.4. Путь познания: от известного к неизвестному, от данных к возможностям и знаниям .....	145
4.4.1. Путь познания в науке о данных .....	145
4.4.2. Четыре пространства знаний в науке о данных .....	145
4.4.3. Эволюция науки о данных: от известного знания к неизвестному знанию.....	147
4.4.4. Перспективы и возможности для новых средств выявления знаний..	148
4.5. Проблема обработки зависимых и неодинаково распределенных данных .....	149

4.5.1. Независимые и одинаково распределенные данные vs зависимые и неодинаково распределенные данные .....	149
4.5.2. Задачи обработки поп-ИД данных.....	151
4.6. Революция человекоподобного машинного интеллекта .....	154
4.6.1. ИИ следующего поколения: человекоподобный ИИ .....	154
4.6.2. Человекоподобный ИИ в контексте науки о данных.....	156
4.7. Качество данных .....	158
4.7.1. Вопросы качества данных .....	158
4.7.2. Метрики качества данных.....	160
4.7.3. Обеспечение качества данных и его контроль .....	163
4.7.4. Аналитика качества данных.....	165
4.7.5. Контрольный список проверок качества данных .....	167
4.8. Вопросы, касающиеся социальных и этических аспектов данных .....	169
4.8.1. Социальные аспекты данных .....	170
4.8.2. Этические аспекты науки о данных .....	172
4.8.3. Гарантии соблюдения этики в отношении данных.....	174
4.9. Экстремальные проблемы науки о данных .....	176
4.10. Резюме .....	179
Глава 5. Наука о данных как научная дисциплина .....	180
5.1. Введение .....	180
5.2. Дисциплинарные пробелы в оценке возможностей данных.....	180
5.3. Методологии для сложных задач науки о данных .....	183
5.3.1. От редукционизма и холизма к систематическому подходу .....	183
5.3.1.1. Редукционизм — восходящая методология .....	184
5.3.1.2. Холистический подход — нисходящая методология .....	184
5.3.1.3. Интегральный систематический подход .....	185
5.3.1.4. Выбор подходящей методологии.....	186
5.3.2. Синтез Х-интеллекта .....	188
5.3.3. От качественного метасинтеза к количественному .....	189
5.4. Общая схема науки о данных как научной дисциплины .....	191
5.4.1. Междисциплинарное слияние в науке о данных .....	192
5.4.2. Дорожная карта исследований в науке о данных .....	194
5.4.3. Систематические подходы к исследованию.....	198
5.4.4. Словарь терминов науки о данных от А до Я .....	200
5.5. Некоторые существенные направления исследований в науке о данных..	201
5.5.1. Развитие мышления .....	202
5.5.1.1. Аспекты образа мышления, свойственного науке о данных .....	202
5.5.1.2. Задачи процесса развития образа мышления в науке о данных ...	203
5.5.2. Понимание свойств и сложностей данных .....	205
5.5.3. Формирование глубокого понимания поведения.....	207
5.5.3.1. Физический мир — мир данных — мир поведения.....	207
5.5.3.2. Поведенческая информатика для выявления новых знаний о поведении .....	209
5.5.4. Слияние науки о данных с общественными и управленческими науками .....	212
5.5.4.1. Переплетение науки о данных, общественных и управленческих наук .....	212
5.5.4.2. Междисциплинарные области и набор возможностей .....	213
5.5.4.3. Создание новых междисциплинарных областей и специалистов ..	214

5.5.5. Создание аналитических хранилищ и автономных систем данных ...	216
5.5.5.1. Тестовые задачи существующих аналитических систем.....	216
5.5.5.2. На пути к автономным аналитическим и обучающим системам	218
5.6. Резюме.....	220
Глава 6. Фундаментальные основы науки о данных .....	222
6.1. Введение .....	222
6.2. Когнитивистика в науке о данных .....	223
6.3. Статистика и наука о данных .....	226
6.3.1. Статистика для науки о данных .....	226
6.3.2. Наука о данных для статистики .....	228
6.4. На стыке информационной науки и науки о данных.....	230
6.4.1. Анализ и обработка .....	231
6.4.2. Информатика для науки о данных .....	232
6.4.3. Обычные информационные технологии.....	234
6.5. Наука об интеллекте и наука о данных .....	234
6.5.1. Распознавание образов, обнаружение связей и зависимостей в данных, аналитика и обучение .....	236
6.5.2. Вычислительный интеллект, инспирированный природой.....	237
6.5.3. Наука о данных: за пределами информационной науки и науки об интеллекте .....	238
6.6. На стыке вычислительной техники и науки о данных.....	242
6.6.1. Вычисления для науки о данных.....	242
6.6.2. Вклад науки о данных в вычисления.....	243
6.7. На стыке общественных наук и науки о данных.....	244
6.7.1. Вклад общественных наук в науку о данных .....	246
6.7.1.1. Использование социального мышления и методов .....	246
6.7.1.2. Создание социальной науки о данных для решения социальных проблем.....	247
6.7.1.3. Социальные черты в науке о данных и ее информационных продуктах.....	248
6.7.2. Вклад науки о данных в общественные науки.....	250
6.7.2.1. Изменение парадигмы и культуры исследований в общественных науках .....	251
6.7.2.2. Создание новых теорий и методов в общественных науках.....	252
6.7.2.3. Новые социальные, правовые и глобальные проблемы и их решение средствами науки о данных .....	254
6.7.3. Социальная наука о данных .....	257
6.8. На стыке менеджмента и науки о данных .....	260
6.8.1. Вклад менеджмента в науку о данных.....	260
6.8.1.1. Решение вопросов менеджмента в науке о данных .....	261
6.8.1.2. Построение компонентов менеджмента в науке о данных .....	263
6.8.2. Вклад науки о данных в менеджмент .....	264
6.8.3. Аналитика менеджмента и наука о данных .....	267
6.9. Коммуникационные исследования и наука о данных .....	268
6.10. Другие основы и факультативные дисциплины науки о данных .....	270
6.10.1. Сферы бизнеса, менеджмента и социальных наук .....	271
6.10.2. Предметные и экспертные знания .....	272
6.10.3. Изобретения, инновации и практика.....	273
6.11. Резюме.....	274

Глава 7. Методы науки о данных.....	275
7.1. Введение .....	275
7.2. Задача аналитики и обучения .....	276
7.3. Концептуальная карта методов науки о данных.....	276
7.3.1. Фундаментальные методы аналитики и машинного обучения .....	279
7.3.2. Классические методы аналитики и машинного обучения.....	281
7.3.3. Современные (постклассические) методы аналитики и машинного обучения .....	284
7.3.4. Дополнительные методы, привлекаемые к решению задач науки о данных.....	290
7.3.4.1. Искусственный интеллект и интеллектуальные системы .....	290
7.3.4.2. Умное производство .....	292
7.3.4.3. Большие данные и методы облачных вычислений .....	292
7.3.4.4. Методы инженерии данных .....	294
7.3.4.5. Интернет вещей .....	295
7.3.4.6. Технологии обеспечения безопасности и конфиденциальности .....	295
7.3.4.7. Другие полезные методы .....	296
7.4. Аналитика и машинное обучение полного цикла: данные — знания — решения .....	297
7.4.1. Аналитика и обучение на данных прошлого.....	298
7.4.2. Аналитика и обучение на текущих данных.....	298
7.4.3. Аналитика и обучение на данных будущего.....	299
7.4.4. Поиск и формирование знаний в форме управленческих решений ..	299
7.5. От дескриптивной к предсказательной и далее к предписывающей аналитике.....	300
7.5.1. Этап 1. Дескриптивная аналитика и бизнес-отчеты .....	302
7.5.2. Этап 2. Предсказательная аналитика / предсказательное машинное обучение и бизнес-аналитика .....	303
7.5.3. Этап 3. Предписывающая аналитика и принятие решений.....	304
7.5.4. Смещение акцентов на разных этапах аналитики и обучения .....	305
7.5.5. Синергия дескриптивной, предсказательной и предписывающей аналитики .....	306
7.6. X-Аналитика.....	309
7.6.1. Спектр X-аналитики .....	309
7.6.2. Рабочий механизм X-аналитики .....	311
7.7. Резюме.....	313
<b>Часть III. Индустриализация и возможности.....</b>	<b>314</b>
Глава 8. Экономика данных и индустриализация .....	315
8.1. Введение .....	315
8.2. Экономика данных .....	315
8.2.1. Что такое экономика данных? .....	316
8.2.2. Примеры экономики данных: умные такси и электрические велосипеды общего пользования .....	320
8.2.3. Новая модель экономики данных.....	322
8.2.4. Отличительные свойства экономики данных .....	325
8.2.5. Интеллектуальная экономика и интеллектуальные вещи данных .....	328
8.2.6. Место реальной экономики .....	330

8.3. Индустрия данных .....	332
8.3.1. Виды индустрии данных .....	332
8.3.2. Новые индустрии данных .....	334
8.3.3. Преобразование традиционных отраслей индустрии .....	337
8.4. Сервисы данных .....	340
8.4.1. Модели сервисов данных .....	340
8.4.2. Сервисы аналитики данных .....	343
8.5. Резюме .....	346
Глава 9. Области применения науки о данных .....	347
9.1. Введение .....	347
9.2. Общее руководство по применению науки о данных .....	348
9.2.1. Сценарии использования науки о данных .....	348
9.2.2. Общие процессы в науке о данных .....	349
9.2.3. Сравнение алгоритмов общего назначения и предметно-ориентированных алгоритмов и решений, зависящих от производителя, с решениями, независимыми от него .....	349
9.2.4. Водопадная и гибкая модели управления проектом в науке о данных и их сравнение .....	351
9.2.5. Факторы успешности проектов в науке о данных .....	354
9.3. Реклама .....	355
9.4. Аэрокосмическая область и астрономия .....	356
9.5. Искусство, креативный дизайн и гуманитарные науки .....	357
9.6. Биоинформатика .....	357
9.7. Консалтинговые услуги .....	358
9.8. Экология и окружающая среда .....	358
9.9. Электронная и розничная торговля .....	359
9.10. Образование .....	361
9.11. Техника .....	362
9.12. Финансы и экономика .....	363
9.13. Игровая индустрия .....	364
9.14. Государственное управление .....	365
9.15. Здравоохранение и поликлиническая деятельность .....	365
9.16. Повседневная жизнь, спорт, развлечения и сопутствующие услуги .....	367
9.17. Менеджмент, операционная деятельность и планирование .....	367
9.18. Производство .....	368
9.19. Маркетинг и продажи .....	369
9.20. Медицина .....	371
9.21. Физические и виртуальные общества, сообщества, сети, рынки и группы людей .....	372
9.22. Издательская деятельность и средства массовой информации .....	375
9.23. Рекомендательные сервисы .....	375
9.24. Наука .....	376
9.25. Защищенность и безопасность .....	377
9.26. Общественные науки и социальные проблемы .....	379
9.27. Жизнеспособность .....	380
9.28. Телекоммуникации и мобильные сервисы .....	381
9.29. Туризм и путешествия .....	381
9.30. Транспорт .....	382
9.31. Резюме .....	384

Глава 10. Профессия в области данных .....	385
10.1. Введение .....	385
10.2. Формирование профессии в области данных .....	386
10.2.1. Дисциплинарный индикатор значимости .....	386
10.2.2. Значимые исследования в области науки о данных .....	387
10.2.3. Глобальные научные сообщества .....	388
10.2.4. Развитие профессиональных организаций .....	390
10.2.5. Социально-экономический эффект .....	391
10.3. Роли в науке о данных .....	392
10.3.1. Команда науки о данных .....	392
10.3.2. Должности в науке о данных .....	394
10.4. Основные знания и навыки профессионалов в науке о данных .....	396
10.4.1. Набор знаний и навыков для работы в науке о данных .....	396
10.4.2. Коммуникативные навыки в науке о данных .....	399
10.4.2.1. Общие коммуникативные навыки в сравнении со специальными навыками .....	399
10.4.2.2. Сравнение более и менее эффективных коммуникаций .....	401
10.5. Зрелость науки о данных .....	403
10.5.1. Модель зрелости способностей в науке о данных .....	403
10.5.2. Зрелость данных .....	405
10.5.3. Зрелость способностей .....	407
10.5.4. Организационная зрелость .....	409
10.6. Исследователи данных .....	411
10.6.1. Кто такие исследователи данных .....	412
10.6.2. Главный исследователь данных .....	412
10.6.3. Что делают исследователи данных .....	413
10.6.4. Квалификация исследователей данных .....	416
10.6.5. Исследователи данных в сравнении с бизнес-аналитиками .....	418
10.6.6. Обзор должностей специалистов по данным .....	419
10.7. Инженеры по данным .....	420
10.7.1. Кто такие инженеры по данным? .....	420
10.7.2. Чем занимаются инженеры по данным .....	422
10.8. Инструментарий профессионалов в области данных .....	424
10.9. Резюме .....	426
Глава 11. Образование в области науки о данных .....	429
11.1. Введение .....	429
11.2. Обзор курсов по науке о данных .....	430
11.2.1. Обзор существующих курсов .....	430
11.2.2. Курсы по дисциплинам .....	432
11.2.3. Массив знаний, предлагаемых в курсах .....	433
11.2.4. Организации, предлагающие образовательные курсы .....	433
11.2.5. Каналы предоставления курсов .....	434
11.2.6. Онлайн-курсы .....	434
11.2.7. Анализ недостатков существующих курсов .....	435
11.3. Структура образования в области науки о данных .....	439
11.3.1. Структура курсов по науке о данных .....	439
11.3.2. Наука о данных для бакалавров .....	441
11.3.2.1. Взгляд и цели .....	441



11.3.2.2. Знания и навыки.....	442
11.3.2.3. Обязательные и факультативные предметы .....	445
11.3.3. Магистр науки о данных.....	446
11.3.3.1. Взгляд на обучение и его цели .....	446
11.3.3.2. Знания и навыки.....	448
11.3.3.3. Обязательные и факультативные предметы .....	449
11.3.4. PhD в науке о данных.....	451
11.3.4.1. Взгляд и цели .....	451
11.3.4.2. Знания и навыки.....	451
11.4. Резюме.....	453
Глава 12. Перспективы и благоприятные возможности науки о данных .....	454
12.1. Введение .....	454
12.2. Четвертая революция = Данные + наука об интеллекте + технология + экономика .....	455
12.2.1. Наука о данных, технология и экономика как новая развивающаяся область .....	455
12.2.2. Четвертая научная, технологическая и экономическая революция ..	457
12.2.2.1. Коренные изменения в науке .....	458
12.2.2.2. Изменение приоритетов на множестве механизмов, управляющих развитием мира .....	458
12.2.2.3. Замещение профессий.....	459
12.2.2.4. Реформирование экономического роста и социального благополучия .....	461
12.3. Роль науки о данных в науке .....	462
12.4. Мозг данных.....	463
12.5. Машинный интеллект и мышление .....	465
12.6. Развитие науки о данных, технологии и экономики .....	467
12.7. Развитие образования и профессий в области данных .....	470
12.8. Резюме .....	471
Библиография .....	472
Толковый англо-русский словарь терминов науки о данных (В. И. Городецкий) .....	488