

С. М. Кашаев

Л. В. Шерстнева

Самостоятельная
подготовка к
по ИНФОРМАТИКЕ

ЕГЭ

Санкт-Петербург

«БХВ-Петербург»

2009

УДК 681.3.06
ББК 32.973.26-018.2
К31

Кашаев, С. М.

К31 Самостоятельная подготовка к ЕГЭ по информатике.
Необходимая теория и достаточная практика /
С. М. Кашаев, Л. В. Шерстнева. — СПб.: БХВ-Петербург,
2009. — 464 с.: ил. — (ИиИКТ)

ISBN 978-5-9775-0455-3

Книга предназначена для самостоятельной подготовки к ЕГЭ по информатике в общеобразовательных школах и содержит следующие разделы: "Информация и информационные процессы", "Системы счисления", "Логика", "Информационные модели и системы", "Компьютер, операционные системы и программные оболочки", "Обработка графической и звуковой информации", "Электронные таблицы", "Базы данных", "Телекоммуникационные технологии", "Алгоритмизация и программирование". Последней теме, с учетом преобладания этих вопросов в тестах ЕГЭ, посвящены три главы: алгоритмизация задач, программирование, алгоритмы без программирования.

Каждая глава построена следующим образом: сначала рассматриваются необходимые теоретические сведения, по мере изложения которых приводятся типовые примеры, в конце содержится отдельный раздел с примерами.

Книга может использоваться как при подготовке к ЕГЭ, так и в текущем учебном процессе учащимися и учителями школ и колледжей.

Для общеобразовательных школ

УДК 681.3.06
ББК 32.973.26-018.2

ISBN 978-5-9775-0455-3

© Кашаев С. М., Шерстнева Л. В., 2009
© Оформление, издательство "БХВ-Петербург", 2009

Оглавление

Введение.....	9
От авторов книги	13
Глава 1. Информация и информационные процессы.....	15
Количество информации	17
Закон аддитивности информации.....	19
Вычисление минимального количества вопросов	20
Вероятность и формула Шеннона	22
Процесс передачи информации	24
Кодирование текстовых данных.....	29
Ситуации с тремя элементарными событиями.....	30
Примеры заданий.....	31
Задания из ЕГЭ 2009.....	39
Глава 2. Системы счисления	43
Виды систем счисления.....	43
Перевод чисел из одной системы в другую.....	47
Дробные числа	51
Действия над числами	54
Представление отрицательных чисел	56
Интервал чисел на разрядной сетке	58
Ошибки при выполнении сложения и вычитания.....	58
Примеры заданий.....	60
Задания из ЕГЭ 2009.....	71
Глава 3. Математическая логика	76
Логические операции	78
Свойства логических операций	81

Примеры заданий.....	83
Задания из ЕГЭ 2009.....	107
Глава 4. Анализ и модели.....	115
Примеры заданий.....	115
Задания из ЕГЭ 2009.....	129
Глава 5. Компьютер и операционные системы	138
Архитектура вычислительных систем	139
Архитектура микропроцессора.....	142
Типы микропроцессоров.....	144
Операционные системы и файловая система	145
Примеры заданий.....	148
Задания из ЕГЭ 2009.....	152
Глава 6. Графика	154
Представление графических данных.....	154
Виды графики.....	155
Кодирование цветов	157
Примеры заданий.....	158
Глава 7. Интернет	161
Протоколы	163
Доменная система имен	165
Порядок в адресном пространстве	165
Службы Интернета	166
Электронная почта	166
Служба WWW	167
Служба передачи файлов FTP.....	168
Унифицированный указатель ресурсов	169
Язык гипертекстовой разметки HTML	169
Создание HTML-файла в Блокноте	171
Форматирование текста.....	176
Изображения.....	185
Гиперссылки.....	188
Списки.....	194
Таблицы	200
Внутренние гиперссылки	204
Тег <META>	205
Теги и <DIV>	206

Примеры заданий.....	208
Задания из ЕГЭ 2009.....	215

Глава 8. Электронные таблицы.....219

Запуск программы Microsoft Excel 2007.....	221
Создание книги Microsoft Excel.....	224
Создание формул и форматирование ячеек.....	226
Относительные и абсолютные ссылки.....	231
Ссылки на другие листы и другие книги	241
Доступ к встроенным функциям	242
Функция ЕСЛИ()	244
Функции И(), ИЛИ() и НЕ().....	245
Вложенные функции ЕСЛИ().....	246
Функция СЧЁТЕСЛИ()	246
Функции ОКРУГЛ(), ОКРУГЛВНИЗ() и ОКРУГЛВВЕРХ()	247
Функция ЦЕЛОЕ()	248
Функции дат и времени	248
Создание имен для ячеек и диапазонов	250
Массивы.....	251
Примеры автоматизации	252
Работа с диаграммами	255
Создание графика	259
Примеры заданий.....	261
Задания из ЕГЭ 2009.....	269

Глава 9. Базы данных273

Структура данных.....	274
Создание таблицы в базе данных	276
Запросы.....	277
Условия отбора	278
ЗАДАНИЕ несовпадающих значений	279
Числовые условия в отборе.....	279
Ввод нескольких условий в одном поле	280
Ввод условий отбора по нескольким полям	280
Примеры заданий.....	281
Задания из ЕГЭ 2009.....	283

Глава 10. Программирование.....285

Знакомство с языком Паскаль	287
Компиляция	291

Операторы для работы с целыми числами	292
Вещественные числа.....	296
Правила при вычислении выражений	299
Константы.....	301
Работа с символами	302
Логический тип данных.....	304
Оператор условия	305
Оператор выбора.....	311
Оператор цикла <i>for</i>	312
Цикл с предусловием.....	319
Цикл с постусловием	322
Массивы.....	323
Работа с символьными строками.....	331
Примеры заданий.....	334
Задания из ЕГЭ 2009.....	354

Глава 11. Алгоритмизация задач.....359

Анализ чисел и системы счисления	359
Подсчет суммы цифр в числе.....	360
Преобразование числа в двоичную систему.....	361
Анализ четности чисел	363
Построение треугольников из отрезков.....	365
Одномерные массивы.....	368
Нахождение среднего арифметического.....	368
Суммирование элементов массива с учетом условия	369
Проверка упорядоченности массива	370
Обмен значений массива	371
Суммирование соседних элементов массива.....	372
Подсчет элементов по условию	373
Перенос модулей значений в другой массив	375
Подсчет количества максимальных элементов	376
Двумерные массивы	379
Вычисление суммы максимальных элементов.....	379
Поиск столбца с максимальной суммой	381
Вычисление суммы при условии	383
Анализ функций	384
Расчет значений функции.....	384
Принадлежность точки определенной области.....	385
Принадлежность точки кольцу	388
Принадлежность точки квадрату	390
Принадлежность числа определенному интервалу.....	393

Решение уравнений.....	395
Квадратное уравнение	400
Решение неравенства	402
Обработка наборов данных.....	404
Тестирование учащихся	404
Отчет по олимпиаде	407
Сертификаты	409
Результаты экзамена	411
Полупроходной балл.....	413
Сортировка	415
Сортировка выбором	416
Сортировка обменом значений.....	421
Поиск решения на шахматной доске.....	422
Задания из ЕГЭ 2009.....	426
Глава 12. Алгоритмы без программирования	436
Калькулятор.....	436
ЗАДАНИЕ 1	436
ЗАДАНИЕ 2.....	437
ЗАДАНИЕ 3.....	438
ЗАДАНИЕ 4.....	439
ЗАДАНИЕ 5.....	439
Игровая стратегия.....	440
Игра 1	440
Игра 2	441
Игра 3	442
Игра 4	443
Робот	444
ЗАДАНИЕ 1	445
ЗАДАНИЕ 2.....	446
Черепашка	447
Задания из ЕГЭ 2009.....	448
Заключение	457
Литература	458
Предметный указатель	459

Введение

Итак, Единый государственный экзамен полноценно вошел в нашу жизнь. Эта книга адресована учащимся, которые собираются сдавать Единый государственный экзамен по дисциплине "Информатика и информационно-коммуникационные технологии". Назначение этой экзаменационной работы заключается в том, чтобы оценить подготовку выпускников общеобразовательных учреждений по данной дисциплине с целью проведения итоговой аттестации и конкурсного отбора абитуриентов в учреждения среднего и высшего профессионального образования.

Общее число заданий в экзаменационной работе — 32, а сама работа состоит из трех частей. Первая часть (А) содержит 18 заданий базового, повышенного и высокого уровней сложности. В этой части собраны задания с выбором ответа, подразумевающие указание одного правильного ответа из четырех предложенных. Задания выполняются на черновике, а ответы заносятся в специальный бланк для ответов части А.

Вторая часть (В) содержит 10 заданий базового, повышенного и высокого уровней сложности. В этой части собраны задания с краткой формой ответа, подразумевающие самостоятельное формулирование и ввод ответа в виде последовательности символов. Задания выполняются на черновике, а ответы заносятся в специальный бланк для ответов части В.

Третья часть (С) содержит 4 задания, первое из которых повышенного уровня сложности, а остальные три — высокого уровня

сложности. Задания этой части подразумевают запись в произвольной форме развернутого ответа на специальном бланке.

Содержание заданий разработано по основным темам курса информатики и информационно-коммуникационных технологий. При этом можно выделить такие тематические блоки: "Информация и ее кодирование", "Алгоритмизация и программирование", "Основы логики", "Моделирование и компьютерный эксперимент", "Программные средства информационных и коммуникационных технологий", "Технология обработки графической и звуковой информации", "Технология обработки информации в электронных таблицах", "Технология хранения, поиска и сортировки информации в базах данных", "Телекоммуникационные технологии". Содержанием экзаменационной работы охватываются основные разделы курса информатики и важнейшие его темы.

Наверняка, какие-то разделы школьной программы по информатике вы знаете лучше, а какие-то недостаточно хорошо для успешной сдачи экзамена. Книга построена таким образом, чтобы вы смогли самостоятельно рассмотреть как необходимую теоретическую информацию, так и на примерах заданий прошлых лет познакомиться с типовыми задачами, которые вас ожидают. Отдельными разделами в каждой главе выделены примеры, которые встретились в вариантах билетов Единого государственного экзамена 2009 года.

В целом книга состоит из двенадцати глав, содержание которых охватывает все разделы, выносимые на Единый государственный экзамен. Структура каждой главы построена по схожему сценарию. Сначала рассматриваются необходимые теоретические сведения, которые излагаются в понятном для учащихся стиле. При этом не требуется использование дополнительных источников информации. По мере изложения необходимых сведений приводятся типовые примеры, что позволяет читателям обратить внимание на ключевые моменты разделов. Кроме того, отдельный раздел в конце каждой главы наполнен исключительно разнообразными практическими примерами из Единого государственного экзамена 2009 года. Такая организация издания связана с удобством восприятия информации читателями.

Опыт прошедших лет показал, что традиционно наибольшие сложности вызывает последний раздел (С), который связан с практическим программированием. Бóльшая часть вопросов и заданий по разделам А и В требует в основном теоретической подготовки и не вызывает таких трудностей, как раздел С. Если рассматривать количество страниц, то примерно треть содержания книги отводится обучению разработке алгоритмов и реализации их в виде программ.

Поясним теперь, что же рассматривается в главах книги. Так, в *главе 1* читатели познакомятся с таким понятием, как информация. Мы рассмотрим различные подходы к вычислению количества информации, познакомимся с единицами измерения информации и разберем вопросы передачи данных. В целом, содержание главы позволяет читателям познакомиться с заданиями Единого государственного экзамена, которые связаны с количеством информации.

Традиционно базовый раздел информатики связан с рассмотрением различных систем счисления (двоичной, восьмеричной, шестнадцатеричной и десятичной). И этой теме посвящена *глава 2*. Здесь много внимания уделено технологии перевода чисел из одной системы в другую. Также в *главе 2* мы разберем представление отрицательных чисел в двоичном дополнительном коде. В целом представленный материал дает представление о хранении числовой информации в вычислительных системах и действиях над числовыми данными.

В *главе 3* приведены необходимые сведения по булевым функциям, логическим выражениям и их преобразованиям. Мы рассмотрим примеры, где будут использоваться таблицы истинности и законы алгебры логики.

Глава 4 касается проведения анализа, а также рассмотрения математических и информационных моделей.

Организации вычислительных систем отведена *глава 5*. Здесь приводится описание структуры микропроцессора (мы поговорим о регистрах, памяти, шине данных и шине адреса). Также рассмотрим необходимые для Единого государственного экзамена сведения по операционным системам и файловой системе.

В *главе 6* речь идет о представлении графической информации в вычислительных системах, приводятся необходимые сведения по растровой и векторной графике.

Глава 7 касается Интернета. Мы разберем технологию создания Web-страниц с помощью языка гипертекста HTML, а также вопросы, связанные с протоколами Интернета и доменной системой имен.

Электронные таблицы (Microsoft Excel) занимают емкую компоненту в школьной программе по информатике. И задания по этой теме присутствуют в каждом билете Единого государственного экзамена. Тематика этих заданий связана с использованием абсолютных и относительных ссылок, фильтрацией и отбором информации. *Глава 8* дает читателям необходимые сведения по данной теме.

Глава 9 касается работы с базами данных. Основное внимание будет уделено заданиям Единого государственного экзамена, которые ориентированы на составление запросов к таблицам.

В *главе 10* рассматривается процесс программирования. В качестве языка выбран Паскаль, знакомый многим по школьной программе. Основное содержание главы сводится к предоставлению учащимся начальных сведений по Паскалю, построению алгоритмов сложности, характерной для частей А и В экзаменационных билетов, и их реализации с помощью программ на Паскале. Большое внимание уделено вопросам организации циклов, ветвления в программах, вводу и выводу данных.

Тема программирования получила развитие в *главе 11*. Здесь основное внимание сконцентрировано на разработке алгоритмов, которые присутствуют в заданиях части С билетов Единого государственного экзамена. Фактически эта часть, судя по опыту ведения занятий на подготовительных курсах, вызывает наибольшие трудности у учащихся.

Часть типовых заданий билетов касается разработки и анализа алгоритмов, однако при этом не требуется написание программ на языке программирования. Большинство подобных заданий связано с разработкой и выбором выигрышных стратегий. Для

удобства читателей разбору таких примеров мы отвели отдельную — двенадцатую главу книги.

От авторов книги

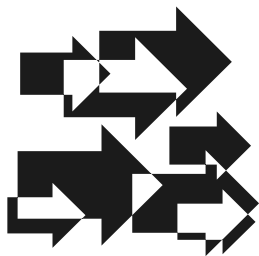
Нам бы хотелось выразить благодарность всем читателям, которые познакомились с нашей книгой. Что касается методики изложения информации, то она связана с работой на курсах по подготовке к Единому государственному экзамену в Нижегородском государственном техническом университете. Много работы в этом плане проводит кафедра "Прикладная математика". Заведующему кафедрой Митякову Сергею Николаевичу мы хотим выразить благодарность. Также большую помощь в работе нам оказал заместитель директора Института дистанционного обучения Воронков Юрий Васильевич, и мы ему очень признательны.

В заключение заметим, что материал, изложенный в книге, предназначается для самостоятельной подготовки учащихся к сдаче Единого государственного экзамена. Кроме того, книга может быть использована школьниками в качестве вспомогательного материала по информатике. Думаем, что книга может быть так же полезна и учителям информатики в плане организации занятий и подбора примеров.

Возможно, книга не лишена неточностей, все их мы относим исключительно на свой счет, и заранее приносим читателям за них свои извинения.

Связаться с авторами можно по электронной почте **mail@bhv.ru**, а также вы можете обратиться по адресу **www.bhv.ru**, чтобы выразить свое отношение к книге.

ГЛАВА 1



Информация и информационные процессы

Такое слово, как *информация*, достаточно часто встречается в нашей жизни. При этом практически всегда, даже интуитивно, оно понимается однозначно и правильно, несмотря на внутреннюю сложность самого понятия "*информация*". Например, объемное справочное руководство несет значительно больший объем информации для читателя, чем небольшая брошюра. Если мы будем с кем-то разговаривать на эту тему, то можем услышать что-то типа "книга содержит много информации" или "в брошюре информации нет". Похожая ситуация связана с компьютерами. В этом случае емкость диска для компьютера однозначно определяет объем данных (количество информации), который можно на нем разместить. Таким образом, если спросить у случайного прохожего, что такое информация, то мы можем услышать: объем или количество данных.

Однако существует возможность получить и другой вариант ответа на подобный вопрос, который будет с определенной точки зрения даже более точен. Он может звучать примерно так: "новизна, новости, неожиданность для слушателя". И чем более неожиданны для получателя сведения, тем больше в них информации. А если передаваемые получателю данные очевидны, или он уже все это знает, то сообщение не несет (или несет очень мало) информации. Важный момент здесь заключается том, что новизна данных рассматривается относительно получателя. Если для

него все и так известно, то и информации в сообщении нет (подтверждение уже имеющихся сведений).

Это общий взгляд на информацию, однако если попытаться дать строгое определение, то сразу возникнут сложности. Дело в том, что определить какое-либо понятие — значит выразить его через другие понятия, уже определенные ранее. Сложность ситуации, однако, заключается в том, что *информация* является одной из исходных категорий, и, следовательно, определение "информации вообще" невозможно свести к каким-то более простым, "исходным" терминам.

Информация необходима человеку для того, чтобы принять решение в определенных ситуациях, связанных с работой или повседневной жизнью. К ее свойствам относятся:

- ☐ полезность;
- ☐ понятность;
- ☐ актуальность;
- ☐ полнота;
- ☐ достоверность.

Особенность любой информации — это то, что она является категорией нематериальной. Следовательно, для существования и распространения в нашем материальном мире она должна быть обязательно связана с какой-либо материальной основой — без нее информация не может проявляться, передаваться, сохраняться, восприниматься и запоминаться нами. В качестве материальных носителей могут выступать акустические волны (касается человеческой речи), телефонные и компьютерные сети, почта и т. д.

В настоящее время принято выделять социальную информацию — многоуровневое знание, которое характеризует:

- ☐ общественные процессы в целом — экономические, политические, социальные, демографические, культурно-духовные;
- ☐ конкретные процессы, происходящие в различных ячейках общества, в организациях.

В самом общем смысле под социальной информацией понимают знания, сообщения и сведения о социальной форме движения материи.

Количество информации

Таким образом, если качественно суть информации понятна, то количественная ее оценка сопряжена с различными подходами. Исторически первые разработки в области информации и ее количественной оценки сделал американский инженер Р. Хартли в 1928 г. Именно он предложил формулу вычисления количества информации. Исходная ситуация такова: источник передает принимающей стороне информацию. При этом в качестве информации передается одно из K возможных сообщений. Например, это может быть K возможных слов, одно из которых человек может сказать другому человеку. Или из K команд одна передается от управляющего центра исполнительной системе.

Объединяющим моментом всех описанных ситуаций является то, что одно из K сообщений получатель может принять, но какое именно, он заранее не знает. Эта неизвестность и определяет *количество информации*, которое получает принимающая сторона. Сама формула достаточно простая и выглядит следующим образом:

$$I = \log_2 K, \quad (1.1)$$

где I — количество передаваемой информации, $\log_2$ — логарифм по основанию два. Количество информации измеряется в битах, и один бит представляет собой количество информации, передаваемое источником, генерирующим только два возможных сообщения (получателю может быть доставлено одно из двух сообщений). Например, от источника получателю доставляется одно из сообщений — "да" или "нет". Конечно, каждое из них может иметь для получателя очень большое значение, однако с точки зрения теории информации это, тем не менее, всего лишь 1 бит. В этом случае одно из двух возможных сообщений принято называть единицей (1), а второе нулем (0). Это хорошо согласуется с цифровыми системами, которые используются для хранения информации в компьютерах.

Если источник информации формирует одно из четырех сообщений ($K = 4$), то количество информации в соответствии с формулой (1.1) равно двум битам. И далее все просто — 8 сообщений приводят к трем битам, 16 сообщений к четырем и т. д.

Количество информации определяет минимальное количество двоичных разрядов (каждый двоичный разряд принимает значение 0 либо 1), которыми можно закодировать K сообщений. Например, если источник информации может выдавать одно из 32 сообщений, то в соответствии с формулой (1.1) количество информации равно 5-ти битам. Для того чтобы закодировать каждое сообщение набором из нулей и единиц, потребуется 5 разрядов. В этом случае каждому сообщению будет сопоставлен уникальный (для каждого сообщения свой особенный) набор из 5-ти двоичных разрядов (5 бит).

Понятно, что один бит — слишком малая величина информации, поэтому чаще используют другую — байт, которая представляет собой группу (совокупность) из восьми битов. Это уже существенная порция информации, и байтом можно без проблем закодировать буквы алфавита (практически во всех языках народов мира алфавит не такой уж большой), цифры и необходимые дополнительные символы. Если в формулу (1.1) подставить $I = 8$, то получим, что $K = 256$. Такого количества различных символов вполне достаточно для передачи букв текста.

Однако если мы говорим о существенной порции информации, то и байт мал, поэтому на практике используют более крупные единицы. Распространены и далее в книге нам встретятся:

- один килобайт (1 Кбайт) — составляет 2^{10} байтов или 1024 байта;
- один мегабайт (1 Мбайт) — составляет 2^{20} байтов или 1024 килобайта;
- один гигабайт (1 Гбайт) — составляет 2^{30} байтов или 1024 мегабайта.

Эти единицы часто используются, когда речь идет об информационной емкости памяти компьютеров и различных носителей, и вы с ними часто сталкиваетесь в повседневной жизни. Также на практике встречается и единица 1 Кбит, которая составляет 1024 бита.

Рассмотрим пару примеров на данную тему, которые встречались в вариантах Единого государственного экзамена прошлых лет.

ПРИМЕР

Вопрос: сколько различных последовательностей длиной 7 символов можно составить из цифр 0 и 1? Решение несложное, и здесь мы должны воспользоваться формулой (1.1). Учитывая исходные данные, запишем $7 = \log_2 K$, а далее в соответствии с определением логарифма по основанию два получим $K = 128$.

ПРИМЕР

Допустим, в корзине находится 8 шаров разного цвета (зеленый, красный, синий, белый и т. д.). Сколько информации несет сообщение о том, что извлечен шар конкретного цвета? Учитывая, что можно вытащить любой из восьми шаров, то в соответствии с формулой (1.1) получим 3 бита.

Закон аддитивности информации

Мы рассмотрели ситуации, когда источник информации формирует одно из K возможных сообщений. Как правило, в реальной ситуации передается несколько сообщений подряд. Например, передача текста представляет собой последовательную передачу буквы за буквой. Общая постановка в этом случае такова: источник информации передает сначала одно из K сообщений, затем еще одно из K сообщений и т. д. Каково же в этом случае (если передается подряд N букв) количество информации, доставляемое получателю? Здесь требуется просуммировать N одинаковых слагаемых

$$I = \log_2(K \cdot K \dots K) = \sum_{i=1}^N \log_2(K) = N \cdot \log_2 K, \quad (1.2)$$

где N — число сообщений, следующих друг за другом от источника информации к ее получателю. В записи (1.2) использовано свойство логарифмов и правило произведения. Что касается пер-

вого, то логарифмы вам хорошо знакомы из школьного курса математики, а второе относится к комбинаторике, и его следует пояснить.

Примечание

Допустим, объект A может быть выбран L способами, а для выбора объекта B существует M способов. В этом случае комбинация объектов A и B может быть выбрана $L \cdot M$ способами. Это и есть *правило произведения*, и оно распространяется и на большее число объектов.

Рассмотрим пример на данную тему, который встретился в одном из вариантов Единого государственного экзамена в прошлые годы.

ПРИМЕР

Будем считать, что каждый символ текстового сообщения кодируется одним байтом. Вопрос заключается в определении количества информации в сообщении "Не ошибается тот, кто ничего не делает!". Для начала необходимо подсчитать число символов в данной фразе, что мы и сделаем — получается 39. Каждый символ несет 8 бит (один байт) информации. После этого следует использовать соотношение (1.2), в котором в качестве N выступает 39, а $\log_2 K$ по условию задания равно 8 бит. В результате перемножения 39 на 8 получим 312 бит. Это и является ответом на вопрос о емкости рассматриваемого информационного сообщения.

Вычисление минимального количества вопросов

Взглянем теперь на количество информации с другой стороны. Дело в том, что количество информации можно трактовать как минимальное количество вопросов, на которые можно давать ответы только "да" или "нет", для того, чтобы определить одно сообщение из K возможных. Такой подход позволяет более понятно пояснить формулу (1.1) в случае, когда K не является

степенью двойки. Например, если $K=5$, то вычисление дает $\log_2 5 = 2.32\dots$ бита.

Сформулируем задачу следующим образом: сколько вопросов (ответы могут быть только "да" или "нет") необходимо задать для того, чтобы из пяти человек выбрать определенное лицо и при этом ограничиться минимальным числом вопросов? Первый вопрос может звучать так: необходимое лицо находится среди первых двух человек (из пяти)? Если ответ на этот вопрос положителен, то следующий вопрос однозначно определяет необходимое лицо. (Вариант вопроса: это первый человек из двух?) Таким образом, при подобном развитии событий за два вопроса мы получили ответ на интересующий нас вопрос. Однако на первый заданный вопрос мог быть получен и отрицательный ответ. Тогда необходимое лицо находится среди оставшихся трех человек. Поэтому следующий вопрос можно задать так: среди первых двух (из трех оставшихся) находится нужный нам человек? При отрицательном ответе необходимое лицо однозначно определено (это третий человек). При положительном ответе осталось с помощью последнего вопроса определить одного человека из двух. Таким образом, в рассмотренном примере две ситуации требуют двух вопросов для определения одного человека из пяти, а одна — трех вопросов. Это численно подтверждает вычисленное количество информации — $2.32\dots$ бита.

ПРИМЕР

Приз спрятан в одной из 16 имеющихся коробок. Необходимо определить количество битов информации в сообщении о том, в какой именно коробке находится приз. Также необходимо найти минимальное количество вопросов, которые следует задать для определения коробки с призом.

Количество информации в соответствии с формулой (1.1) равно четырем. И минимальное количество вопросов, которые необходимо задать для определения коробки с призом, также равно четырем. Первый вопрос может звучать так: в первой восьмерке коробок находится приз? Далее, в зависимости от ответа, следует спросить, в какой четверке коробок (скажем,

в первой ли четверке коробок) из указанной восьмерки размещается приз. Затем, после определения четверки, аналогичным образом выясняем в какой паре, а уже потом определяем одну коробку из двух.

Вероятность и формула Шеннона

Таким образом, мы рассмотрели один подход к определению количества информации. Однако только им теория информации не ограничивается. Далее разберем еще одну возможность определения количества информации. Однако, прежде чем переходить к необходимым соотношениям, рассмотрим один наглядный пример, поясняющий предмет разговора. Скажем, в очередном прогнозе погоды мы услышали, что завтра будет холодный день, например, температура составит -15°C . При этом если в предыдущие дни было также холодно, то подобное известие не несет особой новизны. Однако если мы услышим в прогнозе, что завтра ожидается резкое потепление, то информации данное сообщение принесет нам существенно больше, чем предыдущее.

Неожиданность или, наоборот, очевидность события (сообщения) для получателя количественно описывается с помощью понятия "*вероятность*", которое характеризует степень достоверности события. Вероятность может принимать любые значения в интервале от 0 до 1. Если некоторое событие должно произойти наверняка, то вероятность такого события равна единице. Если же, напротив, событие ни в коем случае не должно произойти, то его вероятность равна 0. Таким образом, вероятность события, которое может произойти либо не произойти, лежит в интервале между нулем и единицей.

Если некоторое событие имеет вероятность P , то количество информации, которое приносит получателю сообщение о данном событии, вычисляется по формуле:

$$I = \log_2(1/P). \quad (1.3)$$

Рассмотрим пару примеров из билетов Единого государственного экзамена прошлых лет на данную тему.

ПРИМЕР

Сколько информации несет сообщение о том, что тетраэдр при подбрасывании падает на синюю грань (для определенности отметим, что у тетраэдра четыре грани, а цвета выберем такие — синий, зеленый, красный и желтый)? Поскольку вероятности выпадения граней одинаковы, то вероятность (P) падения тетраэдра на синюю грань равна $1/4$. Подставив это значение в соотношение (1.3), получим 2 бита.

ПРИМЕР

Допустим, в корзине 16 шаров, которые пронумерованы числами (1, 2, ..., 16). Человек должен случайным образом извлечь один шар. Вероятности извлечения шаров одинаковы. Сколько информации несет сообщение о том, что был извлечен шар под номером 9? Здесь, учитывая, что $P = 1/16$ для каждого шара, то в соответствии с формулой (1.3) получим 4 бита.

Соотношение (1.3) определяет количество информации, доставляемое получателю конкретным сообщением. Однако существует еще одно соотношение, определяющее среднее количество информации, доставляемое получателю от источника информации. В 1948 г. американский математик К. Шеннон предложил формулу для вычисления *среднего количества информации* от источника для сообщений с различными вероятностями:

$$I_{\text{ср}} = \sum_{i=1}^K P_i \cdot \log_2(1/P_i), \quad (1.4)$$

где P_i — вероятность i -го сообщения, а суммирование производится по всем сообщениям (общее количество которых — K). Для $I_{\text{ср}}$ используется термин "энтропия источника сообщений".

ПРИМЕР

Допустим, что мы подбрасываем монету с дефектом, а именно орлом она выпадает чаще, чем решкой. Для определенности положим, что вероятность выпадения решки равна 0.3, а веро-

ятность выпадения орла — 0.7. Вопрос заключается в том, какое количество информации в среднем мы получаем после выполнения очередного броска?

Вычисление количества информации по формуле (1.4) дает $I_{\text{ср}} = 0.3 \cdot \log_2(1/3) + 0.7 \cdot \log_2(1/0.7)$. Далее после выполнения технических действий получаем 0.881 бита. Это меньше, чем в случае равновероятных событий, когда выпадение орла и решки одинаково ожидаемо. Понятно, что в случае дефектной монеты сообщение о том, что выпал орел несет меньше неожиданности (оно более предсказуемо), чем сообщение о том, что выпала решка.

Процесс передачи информации

Информация является таковой, если имеется ее получатель (человек или система, которым информация нужна или представляет для них интерес). И наиболее сложный технический момент связан с передачей информации. При этом можно выделить источник информации, линию связи (это может быть проводная, беспроводная, акустическая, визуальная связь) и приемник (получатель) информации.

Сообщение, подлежащее передаче, может представлять собой последовательность букв, образующих слова. Оно может быть также речевым сообщением или музыкальным произведением. Наконец, сообщением может быть изображение (фотография). В самом словосочетании "источник сообщений" подчеркивается то обстоятельство, что система связи позволяет передать сообщение, выбранное из множества возможных элементарных сообщений для данной системы. Это множество возможных элементарных сообщений называется *алфавитом* источника сообщений. А количество возможных элементарных сообщений представляет собой *мощность* источника сообщений. Так, если источник сообщений формирует сочетания букв русского языка, то совокупность букв является алфавитом источника, а мощность такого источника информации равна 32 (если не делать различий между е и ё, то в русском языке 32 буквы).

Учитывая, что информация часто передается на большие расстояния, возникают проблемы с качеством ее доставки. В этом случае любое сообщение должно быть представлено так, чтобы без ошибок (без искажений) дойти до получателя. Скажем, при передаче информации голосом от человека человеку часто возникают проблемы слышимости. В этом случае, если расслышать переданные слова не удастся, то можно попросить повторить фразу или подойти поближе. При телефонной связи также часто возникает проблема различимости исходного сообщения. Еще более сложная ситуация с современными цифровыми системами передачи данных. В них любая информация представляет собой набор битов (единиц и нулей), которые в совокупности несут полную информацию о тексте, звуке и видеоизображении. Каждый бит на аппаратном уровне в вычислительной системе представляет собой один из двух уровней напряжения, которые фиксируются в электронных схемах. И для того, чтобы, скажем, на экране монитора компьютера отобразилась информация о фрагменте текста, система должна выполнить преобразование — из кода (набора битов) определенного символа непосредственно в визуальное представление символа. Подчеркнем наиболее важное из сказанного — в вычислительных системах информация представляется в виде нулей и единиц и хранится достаточно надежно.

Однако для передачи этих данных на большие расстояния требуются дополнительные технические средства. Дело в том, что если мы хотим передать на расстояние набор битов (нулей и единиц), то необходимо использовать какой-нибудь физический процесс. Например, на рис. 1.1 показано соответствующее изменение напряжения во времени. Фактически в этом случае каждый бит передается по проводу в течение определенного интервала времени (T_0). Система, которая принимает информацию, воспринимает эти уровни напряжения и переводит результат в набор битов, которые уже могут быть однозначно интерпретированы. Однако использование ступенчатой функции, приведенной на рис. 1.1, при передаче на значительное расстояние приводит к искажениям. В этом случае принимающая сторона может увидеть заметно отличающуюся от первоначальной функцию (например, в виде, представленном на рис. 1.2). И в ряде случаев правильно восста-

новить переданный набор битов достаточно сложно. Все эти трудности связаны с действием шумов и многочисленных помех.

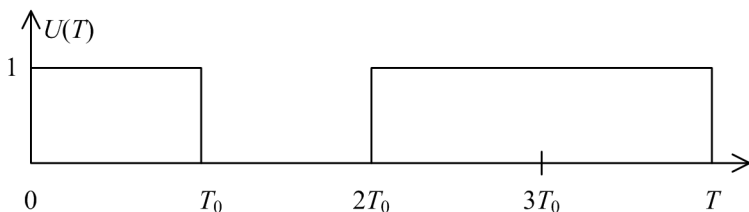


Рис. 1.1. Процесс передачи битов информации во времени

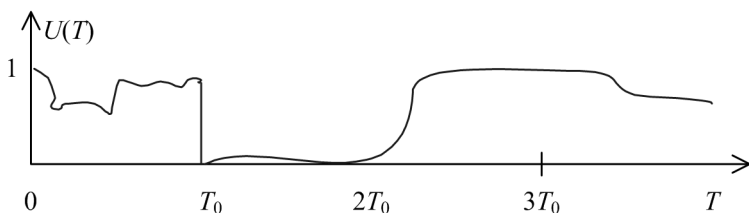


Рис. 1.2. Искажения в процессе передачи информации

Для устранения описанных проблем существует специальный компонент — модем (сокращенное название от "модулятор/демодулятор"). Он производит много полезных действий и самое важное из них заключается в том, что модем преобразует биты в сигналы, которые распространяются на далекие расстояния без искажений (точнее, без искажений, которые мешают качеству связи). Разумеется, модемы на приемной стороне производят и обратное преобразование — принятые сигналы в биты информации.

В модуляторе бит, соответствующий нулю, преобразуется в один сигнал, а бит, соответствующий единице, преобразуется в другой сигнал. Эти два сигнала должны друг от друга хорошо отличаться. В качестве сигналов используются гармонические функции (функции $\sin$ и $\cos$). При этом для обеспечения отличий одного сигнала от другого часто применяются разные частоты (частота одного гармонического колебания выбирается отличной от час-

тоты другого). Дело в том, что высокочастотные гармонические колебания хорошо передаются на удаленные расстояния без значительных искажений. На приемной стороне располагается другой модем, который делает обратное преобразование и каждый принятый сигнал заменяется соответствующим битом. При этом даже если принятый сигнал искажается, модем может правильно решить — какой сигнал был передан. Для этого в демодуляторе принимается решение — на какой из исходных сигналов принятая функция времени "больше похожа". Данная процедура реализуется с помощью согласованной фильтрации, но этот вопрос выходит за рамки рассматриваемого курса.

Модемы имеют свои технические характеристики, и наиболее важной является *информационная скорость* — какое количество битов в секунду передает модем. Реально модемы работают достаточно быстро, их скорость измеряется в килобитах в секунду (Кбит/с) и даже в мегабитах в секунду.

ПРИМЕР

Скорость передачи через модемное соединение равна 56 Кбит/с. Передача текстового файла через это соединение заняла 2 секунды. Требуется определить, сколько символов содержал переданный текст, если на каждый символ отводится 8 бит.

Известно, что 1 Кбит составляет 1024 бита. Таким образом, за 2 секунды передается: $56 \cdot 1024 \cdot 2$, что дает 114 688 бит. По условию для каждого символа отводится 1 байт или 8 бит. В результате требуется разделить 114 688 бит на 8 бит, что приводит к ответу 14 336 символов.

Кроме преобразования битов в сигналы модемы выполняют еще несколько полезных функций. Одна из них связана с уплотнением информации. Поясним это на примере. Допустим, мы передаем сообщение, составленное из букв некоего алфавита. При этом одни буквы встречаются в сообщении значительно чаще, чем другие. Для конкретности пусть в алфавите источника информации будут только три буквы — А, В и С. При этом статистика говорит о том, что частота появления буквы А в тексте составля-

ет 50% (для любой буквы конкретного алфавита такую статистику легко подсчитать — необходимо взять толстую книгу и проанализировать суммарное количество различных букв). Для букв В и С будем считать, что частоты появления составляют по 25%. Очевидно, что если для передачи букву А закодировать одним битом (скажем, единицей), а другие буквы двумя битами (В — 00, С — 01), то количество битов в большом сообщении в среднем будет меньше по сравнению с ситуацией, когда на каждую букву приходится одинаковое количество битов.

Так, в рассматриваемом случае при передаче сообщения из 1000 букв потребуется в среднем: 500 букв умножить на 1 бит плюс 500 букв умножить на 2 бита. В результате это составит 1500 бит (в среднем!). При выделении же на каждую букву целых двух бит понадобится 2000 бит. Понятно, что при больших передаваемых объемах информации подобный выигрыш существенен. И модемы данную процедуру, которая называется *эффективным кодированием*, выполняют.

Кроме того, как мы уже упоминали, при передаче информации возможны ошибки. Хотя сигналы, соответствующие нулю и единице, подбираются максимально отличающимися друг от друга, тем не менее, ошибки все-таки возможны. При возникновении подобных ситуаций в процессе человеческого общения мы можем попросить повторить плохо различимое слово или попросить произнести это слово по буквам. Тем самым мы вносим в исходное сообщение еще и избыточную информацию. Фактически внесение избыточной информации в сообщение позволяет на приемной стороне исключить ошибки. Аналогичное действие выполняет модем — вместе с полезной информацией он добавляет в сообщение избыточные биты, которые и помогают на приемной стороне устранить возможные ошибки.

Если попытаться пояснить идею избыточности, то самый простой вариант заключается в повторении передаваемых битов. Так, вместо единицы мы будем передавать пять единиц, а вместо нуля — пять нулей. Тогда, даже если один или два бита из пяти в процессе передачи исказятся, то все равно по большому числу (большому числу нулей или единиц из пятерки принятых битов)

мы сможем принять правильное решение в плане того — какой бит передавался.

Кодирование текстовых данных

Как мы уже выяснили, для передачи текста каждому символу алфавита необходимо сопоставить уникальный набор нулей и единиц. Это фактически представляет собой число в двоичной системе счисления (о системах счисления мы подробно поговорим в *главе 2*). С помощью восьми разрядов (байта) можно закодировать 256 различных символов. Этого вполне достаточно, чтобы выразить различными комбинациями из восьми битов все символы русского и английского алфавита, как строчные, так и прописные буквы. Также кодируются знаки препинания и обозначения арифметических действий.

Разумеется, вариантов кодирования можно придумать много (какую букву сопоставить с каким кодом), и для взаимного понимания участников информационного обмена необходим определенный стандарт. В этом случае все будут однозначно понимать — какой код какой букве соответствует. И такая деятельность началась с того, что Институт стандартизации США ввел в использование систему кодирования ASCII (American Standart Code for Information Interchange, стандартный код информационного обмена США). В системе ASCII закреплены две таблицы — базовая и расширенная. Базовая таблица определяет значения кодов от 0 до 127 (число 127 в десятичной системе счисления), а расширенная относится к символам с номерами от 128 до 255 (эти числа также представлены в десятичной системе счисления).

Первые 32 символа базовой таблицы, начиная с нулевого, отданы производителям вычислительных средств. В этой области таблицы размещаются так называемые управляющие коды, которым не соответствуют никакие символы языков. Эти символы не выводятся на экран и на устройства печати.

Начиная с 32-го символа и заканчивая 127-м, размещены коды символов английского алфавита, знаков препинания, цифр, арифметических действий и некоторых вспомогательных символов.

Национальные системы кодирования каждой страны располагаются в расширенной части с кодами от 128 до 255. В России наиболее распространена кодировка символов русского языка, известная как Windows-1251. Она была введена компанией Microsoft. Учитывая популярность операционных систем этой компании, кодировка получила широкое распространение. Она используется на большинстве компьютеров, работающих на платформе Windows.

Однако имеется и другая распространенная кодировка — КОИ-8 (код обмена информацией, восьмизначный); ее происхождение относится к временам действия Совета Экономической Взаимопомощи государств Восточной Европы. Сегодня кодировка КОИ-8 используется в компьютерных сетях на территории России и в российском секторе Интернета.

Имеется еще и международный стандарт, в котором предусмотрена кодировка символов русского алфавита. Она носит название кодировки ISO (International Standart Organization, Международный институт стандартизации). Однако на практике данная кодировка используется редко.

Если проанализировать организационные трудности, связанные с созданием единой системы кодирования текстовых данных, то можно прийти к выводу, что они вызваны ограниченным набором кодов (256). В то же время, очевидно, что если кодировать символы не восьмиразрядными кодами, а кодами с большей разрядностью, то и количество возможных символов увеличится. Такая система, основанная на 16-разрядном кодировании символов, получила название универсальной — Unicode. Шестнадцать разрядов позволяют обеспечить размещение 65 536 символов.

Ситуации с тремя элементарными событиями

Как мы увидели в этой главе, вычисление количества информации строится на двух базовых информационных символах — нуле и единице. Однако в ряде ситуаций положение не совсем такое. В заданиях Единого государственного экзамена есть примеры подобного типа. И для того чтобы было понятно, о чем идет речь, рассмотрим задачу.