# От авторов

Перед вами учебник по информатике углублённого уровня для 11 класса. Так же как и в учебнике для 10 класса, в некоторых главах мы рассмотрим вроде бы знакомые вопросы, но несколько с другой точки зрения, более глубоко. Например, в первой главе (она называется «Информация и информационные процессы») вы узнаете, как связано количество информации с теорией вероятностей, как работают упаковщики (программы для сжатия данных) и как можно построить код, позволяющий исправлять ошибки при передаче данных.

Большое внимание в учебнике уделено вопросам алгоритмизации и программирования. Этому посвящены три главы: «Элементы теории алгоритмов», «Алгоритмизация и программирование» и «Объектно-ориентированное программирование». В первой из них вы узнаете, как и зачем ввели строгое математическое понятие «алгоритм», как сравнивать качество алгоритмов и как доказывать правильность программ. Вторая глава знакомит с программированием структур данных, с некоторыми из них (например, со списками, графами и деревьями) вы уже работали в 10 классе. В последней главе объясняются основы объектного подхода к разработке программ, который применяется в крупных промышленных проектах. В практической части используется объектная версия языка Паскаль, реализованная в свободно распространяемом компиляторе *FreePascal*, и среда быстрой разработки *Lazarus*.

В конце каждого параграфа есть вопросы и задания, которые помогут понять, хорошо ли вы усвоили материал. В тексте нет прямых ответов на некоторые вопросы, но есть вся необходимая информация для ответа на них.

Задачи в конце параграфов помогут закрепить материал на практических работах. Самые сложные задачи (на взгляд авторов) отмечены звёздочкой (*).

Мы старались сделать так, чтобы содержание учебника как можно меньше зависело от программного обеспечения, установленного на ваших компьютерах. Весь курс можно успешно изучать, используя только свободное программное обеспечение (СПО) — операционную систему *Linux*, офисный пакет *OpenOffice.org* или его модификации (например, *LibreOffice*), компилятор *FreePascal*, графический редактор *Gimp*, программу трёхмерного моделирования *Blender* и др.

В заключение нам хочется поблагодарить наших коллег, которые взяли на себя труд прочитать предварительные версии отдельных глав учебника и высказать множество полезных замечаний, позволивших сделать учебник более точным, ясным и понятным:

- *А. П. Шестакова*, кандидата педагогических наук, заведующего кафедрой информатики и вычислительной техники Пермского государственного педагогического университета, который вдохновил авторов на написание этого учебника;
- *М. А. Ройтберга*, доктора физико-математических наук, заведующего лабораторией прикладной математики Института математических проблем биологии РАН, г. Пущино;
- *С. С. Михалковича*, кандидата физико-математических наук, доцента кафедры алгебры и дискретной математики Южного федерального университета, г. Ростов-на-Дону;
- *Е. В. Андрееву*, кандидата физико-математических наук, заведующую кафедрой информатики СУНЦ МГУ, г. Москва;
- *О. А. Тузову*, учителя информатики школы № 550, г. Санкт-Петербург;
- *А. Г. Тамаревскую*, учителя информатики лицея № 533, г. Санкт-Петербург;
- *Н. Д. Шумилину*, кандидата педагогических наук, учителя информатики МОУ «Тверская гимназия № 6», г. Тверь;
- *Л. Б. Кулагину*, учителя информатики ФМЛ № 239, г. Санкт-Петербург;
- *Ю. М. Розенфарба*, учителя информатики и ИКТ МОУ «Межозёрная средняя общеобразовательгая школа», Челябинская область;

- *Т. А. Мисаренкова*, учителя информатики школы № 163, г. Санкт-Петербург;
- *К. А. Малеванова*, технического директора компании «ПиН Телеком»;
- *А. Г. Архангельского*, генерального директора ЗАО «АЗет Дизайн»;
- *И. А. Васильевского*, дизайнера, специалиста по 3D-моделированию.

*Уважаемые ученики!*

В работе с книгой вам помогут навигационные значки:

— важное утверждение или определение.

— вопросы и задания к параграфу.

— дополнительное разъяснение.

— задания для подготовки к итоговой аттестации.

www — К каждой главе учебника рекомендуются:
1) электронный образовательный ресурс (ЭОР) с сайта Федерального центра образовательных ресурсов (ФЦИОР): http://fcior.edu.ru
Доступ к ЭОР из каталога ФЦИОР:
http://fcior.edu.ru/catalog/meta/4/mc/discipline%20OO/mi/4.06/p/page.html.
Ресурсы размещены в алфавитном порядке, согласно названиям учебных тем;
2) практические работы на методическом сайте издательства lbz.metodist.ru в авторской мастерской К. Ю. Полякова и Е. А. Еремина.

— Проектное или исследовательское задание.

В ходе выполнения проекта (исследования) вы можете:
- подготовить набор полезных ссылок с использованием веб-ресурсов;
- подготовить небольшое выступление с использованием презентации (5–7 мин.);
- оформить доклад и поместить его на сайт школьной конференции;
- подтвердить полученные результаты расчётами и графиками (диаграммами);
- подготовить видеоролик;
- разместить материалы проекта (исследования) в коллекции обучающих модулей по предмету на сайте школы.

# Глава 1
# Информация и информационные процесы

## § 1
## Количество информации

### Формула Хартли

Вы знаете, что при выборе из двух возможных вариантов количество полученной информации равно 1 биту. Если количество вариантов $N$ равно $2^I$, то количество информации при выборе одного из них равно $I$ битов. А как вычислить количество информации, если количество вариантов не равно степени числа 2?

Ответить на этот вопрос стало возможно только после того, как вы изучили логарифмы в курсе математики. Из формулы

$$N = 2^I$$

сразу следует, что $I$ — это степень, в которую нужно возвести 2, чтобы получить $N$, т. е. *логарифм*:

$$I = \log_2 N.$$

Эта формула называется **формулой Хартли** в честь американского инженера Ральфа Хартли, который предложил её в 1928 г.

**Ральф Хартли (1888–1979)**

Пусть, например, на лётном поле стоят 10 самолётов (с номерами от 1 до 10) и известно, что один из них летит в Санкт-Петербург. Сколько информации в сообщении «Самолёт № 2 летит в Санкт-Петербург»? У нас есть 10 вариантов, из которых выбирается один, поэтому по формуле Хартли количество информации равно

$$I = \log_2 10 \approx 3{,}322 \text{ бита.}$$

Обратите внимание, что для значений $N$, которые не равны целой степени числа 2, количество информации в битах — дробное число.

С помощью формулы Хартли можно вычислить теоретическое количество информации в сообщении. Предположим, что алфавит (полный набор допустимых символов) включает 50 символов (в этом случае говорят, что **мощность алфавита** равна 50). Тогда информация при получении каждого символа составляет

$$I = \log_2 50 \approx 5{,}644 \text{ бита.}$$

Если сообщение содержит 100 символов, его общий информационный объём примерно равен

$$5{,}644 \cdot 100 = 564{,}4 \text{ бита.}$$

В общем случае объём сообщения, использующего алфавит из $N$ символов и состоящего из $M$ символов, равен

$$I = M \cdot \log_2 N.$$

Такой **подход** к определению количества информации называют **алфавитным**. Конечно, на практике невозможно использовать для кодирования символа нецелое число битов, поэтому используют первое целое число, которое больше теоретически рассчитанного значения. Например, при использовании алфавита из 50 символов каждый символ будет закодирован с помощью 6 битов ($50 \leq 2^6 = 64$).

Сколько разных сообщений можно передать, если известен алфавит и длина сообщения? Предположим, что для кодирования сообщения используются 4 буквы, например «А», «Б», «В» и «Г», и сообщение состоит из двух символов. Поскольку каждый символ может быть выбран 4 разными способами, на каждый вариант выбора первого символа есть 4 варианта выбора второго. Поэтому общее число разных двухбуквенных сообщений вычисляется как $4 \cdot 4 = 4^2 = 16$. Если в сообщение добавить ещё один символ, то для каждой из 16 комбинаций первых двух символов третий можно выбрать четырьмя способами, так что число разных трёхсимвольных сообщений равно $4 \cdot 4 \cdot 4 = 4^3 = 64$.

В общем случае, если используется алфавит из $N$ символов и сообщение состоит из $M$ символов, то количество разных возможных сообщений равно

$$K = N^M.$$

## Задачи

1. В русском лото 99 бочонков. Какое количество информации содержится в сообщении «Первым вытащили бочонок с номером 16»?

2. В классе 25 учеников. Какое количество информации содержится в сообщении «Сегодня дежурит Василий Иванов»?
3. В чём состоит алфавитный подход к оценке количества информации?
4. Оцените теоретическое количество информации в сообщении «Приеду в четверг». Используемый алфавит состоит из заглавных и строчных русских букв и пробела.
5. Каков будет фактический объём сообщения «Приеду в четверг», если при передаче каждый символ кодируется минимально возможным целым числом битов?
6. В некоторой стране автомобильный номер длиной 7 символов составляется из заглавных букв (всего используется 26 букв) и десятичных цифр в любом порядке. Каждый символ кодируется одинаковым и минимально возможным количеством битов, а каждый номер — одинаковым и минимально возможным количеством байтов. Определите объём памяти, необходимый для хранения 20 автомобильных номеров.
7. Объём сообщения, содержащего 4096 символов, равен 1/512 части мегабайта. Какова мощность алфавита, с помощью которого записано это сообщение?
8. Какое наименьшее число символов должно быть в алфавите, чтобы с помощью всевозможных трёхбуквенных слов, состоящих из символов данного алфавита, можно было передать не менее 9 различных сообщений?

### Информация и вероятность

Всё, что написано в предыдущем пункте, верно только при одном уточнении: все события (символы алфавита) одинаково ожидаемы, т. е. нельзя заранее сказать, что какой-то символ встречается чаще, а какой-то — реже. В реальности это предположение не всегда верно. Например, в тексте на русском языке некоторые символы встречаются часто, а некоторые — очень редко. Числа во втором столбце табл. 1.1 означают относительную долю символа в больших текстах. Например, доля 0,175 для пробела означает, что примерно 17,5% всех символов в текстах на русском языке — пробелы.

*Таблица 1.1*

| | |
|---|---|
| **Пробел** | 0,175 |
| **О** | 0,090 |
| **Е, Ё** | 0,072 |
| **А** | 0,062 |
| **И** | 0,062 |
| **Т** | 0,053 |
| **Н** | 0,053 |
| **С** | 0,045 |
| **Р** | 0,040 |
| **В** | 0,038 |
| **Л** | 0,035 |
| **К** | 0,028 |
| **М** | 0,026 |
| **Д** | 0,025 |
| **П** | 0,023 |
| **У** | 0,021 |
| **Я** | 0,018 |
| **Ы** | 0,016 |
| **З** | 0,016 |
| **Ь, Ъ** | 0,014 |
| **Б** | 0,014 |
| **Г** | 0,013 |
| **Ч** | 0,012 |
| **Й** | 0,010 |
| **Х** | 0,009 |
| **Ж** | 0,007 |
| **Ю** | 0,006 |
| **Ш** | 0,005 |
| **Ц** | 0,004 |
| **Щ** | 0,003 |
| **Э** | 0,003 |
| **Ф** | 0,002 |

Иногда среди всех возможных событий есть ожидаемые и неожиданные. Например, на вопрос «Идёт ли сейчас снег?» летом мы ожидаем услышать ответ «Нет». При этом ответ «Да» будет очень неожиданным, и после его получения наши дальнейшие планы могут сильно измениться. Это значит, что при таком ответе мы получаем гораздо больше информации. Как её измерить точно?

Сначала нужно разобраться с тем, что значит «менее ожидаемое» событие и «более ожидаемое». Математики в этом случае используют понятие «**вероятность**»: если событие более ожидаемое, то его вероятность (точнее, вероятность того, что оно произойдёт) больше.

Вероятность — это число в интервале от 0 до 1. В математике вероятность принято обозначать буквой $p$ (от латинского *probabilis* — вероятный, возможный).

Сначала рассмотрим предельные случаи, когда вероятность равна 0 или 1. Пусть, например, $x$ — некоторое неизвестное вещественное число, которое задумал ведущий. Вы знаете, что для любого вещественного $x$ *всегда* $x^2 \geq 0$. В этом случае считают, что вероятность события $x^2 \geq 0$ равна 1 (событие $x^2 \geq 0$ обязательно произойдёт). Часто вероятность измеряют в процентах от 1, тогда вероятность события $x^2 \geq 0$ равна 100%. В то же время вероятность события $x^2 < 0$ равна нулю, это значит, что событие $x^2 < 0$ никогда не произойдёт.

Теперь предположим, что мы бросаем монету и смотрим, какой стороной она упала, «орлом» или «решкой». Если повторять этот опыт много раз, мы заметим, что количество «орлов» и «решек» примерно равно (конечно, если монета не имеет дефектов). При этом вероятность каждого из двух событий равна 0,5, или 50%. Скорее всего, вы слышали выражение «50 на 50», которое означает, что ни одному из двух вариантов нельзя отдать предпочтение — их вероятности равны.

Вероятность события можно определить с помощью большого количества испытаний. Если из $N$ испытаний нужное нам событие случилось $m$ раз, то вероятность такого события можно оценить как $\dfrac{m}{N}$. Например, классический игральный кубик имеет 6 граней; если кубик качественный, вероятность выпадения каждой грани равна 1/6. Вероятность выпадения чётного числа можно подсчитать так: среди чисел от 1 до 6 всего

3 чётных числа, поэтому при большом числе испытаний в половине случаев (в 3 из 6) будут выпадать чётные числа, т. е. вероятность равна 0,5. А вероятность выпадения числа, меньшего 3, равна $2/6 = 1/3 \approx 0,33$, потому что только 2 из 6 чисел (1 и 2) удовлетворяют условию.

Теперь можно переходить к главному вопросу: как вычислить количество информации, если в сообщении получен символ, вероятность появления которого равна $p$. Попробуем сначала определить, какими свойствами должна обладать эта величина, исходя из «здравого смысла».

Во-первых, чем меньше вероятность, тем более неожидан символ и тем больше информации мы получили. Если вероятность события близка к нулю, количество информации должно стремиться к бесконечности (получение такого символа очень неожиданно).

Во-вторых, представим себе, что мы получаем символы только одного вида, например только буквы «А». Тогда вероятность появления символа «А» равна 1 и никакой новой информации в этом символе для нас нет — мы всё заранее знали. Следовательно, при $p = 1$ информация должна быть равна нулю.

В-третьих, «здравый смысл» подсказывает, что когда мы бросаем игральный кубик два (три, 102, 1002) раза, мы получаем информации в два (три, 102, 1002) раза больше, чем при однократном бросании кубика. Это свойство называют принципом *аддитивности* (сложения).

Математики доказали, что этими свойствами обладает только логарифмическая функция вида $f(p) = -K \cdot \log_2 p$, где коэффициент $K$ можно выбирать произвольно, как удобно в конкретной задаче. Если взять $K = 1$, мы получим количество информации в битах.

Если событие имеет вероятность $p$, то количество информации в битах, полученное в сообщении об этом событии, равно

$$I = -\log_2 p = \log_2 \frac{1}{p}. \quad (1)$$

Поскольку вероятность $p$ не больше 1, количество информации не может быть меньше нуля. Легко проверить, что если вероятность равна 1, то количество полученной информации равно нулю. Если вероятность стремится к 0, величина $\log_2 p$ стремится к $-\infty$, а количество информации — к $\infty$.

Третье свойство (аддитивность, сложение вероятностей) проверим на примере. Пусть есть два мешка, в каждом из которых лежат 8 шариков разного цвета. Вычислим, какое количество информации мы получили из сообщения «Из первого мешка вытащили (наугад) красный шарик, а из второго — зелёный». Из каждого мешка можно вытащить один из восьми шариков, т. е. вероятность вытащить какой-то определённый шарик равна 1/8. Следовательно, количество информации в сообщении «Из первого мешка вытащили красный шарик» равно

$$I = \log_2 \frac{1}{\frac{1}{8}} = \log_2 8 = 3 \text{ бита}.$$

Точно такую же информацию, 3 бита, мы получаем из сообщения «Из второго мешка вытащили зелёный шарик».

Теперь посмотрим, какое количество информации несёт исходное полное сообщение «Из первого мешка вытащили (наугад) красный шарик, а из второго — зелёный». Какова вероятность именно этого исхода? Есть 8 способов вытащить шарик из каждого мешка, всего получаем $8 \cdot 8 = 64$ варианта, поэтому вероятность каждого из них равна 1/64. Тогда количество информации в сообщении равно

$$I = \log_2 \frac{1}{\frac{1}{64}} = \log_2 64 = 6 \text{ битов},$$

т. е. равно сумме количеств информации в двух отдельных сообщениях. Мы показали, что свойство аддитивности выполняется.

Возможно, вы уже заметили, что последний пример фактически свёлся к использованию формулы Хартли. Если все $N$ вариантов имеют равные вероятности, то вероятность каждого варианта равна $p = \frac{1}{N}$, следовательно, количество информации о любом из возможных событий вычисляется как

$$I = \log_2 \frac{1}{\frac{1}{N}} = \log_2 N.$$

Этот результат совпадает с формулой Хартли.

Однако формулу Хартли нельзя использовать, если вероятности событий разные. Пусть, например, детям раздают воздушные шарики разного цвета, причём 7 из каждых 10 шариков — зелёные. Какое количество информации содержится в сообщении

«Маша получила зелёный шарик»? Здесь формула Хартли неприменима, однако можно использовать формулу (1). Вероятность того, что Маше достался зелёный шарик, равна 7/10, а количество информации вычисляется как

$$I = \log_2 \frac{1}{\frac{7}{10}} = \log_2 \frac{10}{7} \text{ битов.}$$

Величина под знаком двоичного логарифма не является степенью двойки. Как её вычислить? Для этого используется свойство логарифма, позволяющее переходить к другому основанию, например, к десятичным или натуральным логарифмам, которые умеют вычислять калькуляторы:

$$I = \log_2 x = \frac{\lg x}{\lg 2} = \frac{\ln x}{\ln 2} \text{ битов.}$$

В данном случае получаем

$$I = \frac{\lg \frac{10}{7}}{\lg 2} = \frac{\ln \frac{10}{7}}{\ln 2} \approx 0{,}515 \text{ бита.}$$

## Вопросы и задания

1. Как вы понимаете термин «вероятность события»?
2. В каких случаях вероятность равна 1? Когда она равна 0?
3. Что неправильно в сообщении: «"Спартак" выиграет с вероятностью 200%»?

## Задачи

1. Вероятность появления символа @ в некотором тексте равна 0,125. Сколько битов информации несёт сообщение о том, что очередной символ текста — @?
2. В садке у рыбака сидят 2 окуня, 4 плотвы и 10 уклеек. Не смотря в садок, рыбак вытаскивает наугад одну рыбу. Какова вероятность того, что это будет плотва?
3. В пруду плавают 10 линей, 20 окуней и 70 карасей. Считаем, что они одинаково голодны и равномерно распределены по водоёму. Какова вероятность того, что первая рыба, пойманная рыбаком, будет линем? Окунем? Карасём?
4. Ученик задумал целое число от 1 до 100. Какова вероятность того, что это будет число в интервале от 21 до 30? От 31 до 55? Больше 25? Равно 25?

[. . .]