

Подходы к измерению информации

На данный момент предложено два подхода к измерению информации: *содержательный* и *алфавитный* (объёмный).

Содержательный подход был предложен американским математиком Клодом Шенноном (1916 – 2001). Он предложил оценивать информационный объем сообщения от содержащейся в нём полезной информацией, т. е. той частью сообщения, которая снимает полностью или уменьшает существующую до её получения неопределённость какой-либо ситуации.

Информация — это снятая неопределённость. Величина неопределённости некоторого события — это количество возможных результатов (исходов) данного события.

Каково должно быть минимальное количество возможных вариантов, чтобы появилась неопределённость? Если перед человеком имеется только один вариант, значит он уже знает результат и неопределённости нет. Когда перед человеком стоит выбор между двумя вариантами, уже возникает неопределённость какой из них получится в итоге. Это и есть наименьшая неопределённость.

Пример 1. Допустим, вы подбрасываете монету, загадывая, что выпадет: «орёл» или «решка». Перед подбрасыванием монеты неопределённость знания о результате равна двум. Действительно, есть всего два возможных результата этого события (бросания монеты). Эти результаты мы считаем равновероятными, т. к. ни один из них не имеет преимущества перед другим.

После того как конкретный исход стал известен (например, подброшенная монета упала «орлом» вверх), неопределённость уменьшилась в 2 раза (было два возможных варианта, стал известен один из них, $2/x = 1$, $x = 2$).

Сообщение, уменьшающее неопределённость знания в 2 раза, несёт **1 бит** информации.

Рассмотрим пример с большей неопределённостью (с большим кол-вом исходных вариантов).

Пример 2. Предположим, в книжном шкафу восемь полок. Книга может быть поставлена на любую из них. Сколько бит информации несёт сообщение о том, что книга поставлена на третью полку?

Ответ на этот вопрос можно получить, если дополнить исходное сообщение ещё несколькими сообщениями так, чтобы каждое из них уменьшало неопределённость знания в 2 раза.

Итак, количество возможных результатов (исходов) события, состоящего в том, что книга поставлена в шкаф, равно восьми: **1, 2, 3, 4, 5, 6, 7 и 8.**

Сообщение **«Книга поставлена на полку не выше четвёртой»** уменьшает неопределённость знания о результате в два раза. Действительно, после такого сообщения остаётся всего четыре варианта: **1, 2, 3 и 4.** Получен один бит информации.

Сообщение **«Книга поставлена на полку выше второй»** уменьшает неопределённость знания о результате в два раза: после этого сообщения остаётся всего два варианта: **3 и 4.** Получен ещё один (второй) бит информации.

Сообщение **«Книга поставлена на третью полку»** также уменьшает неопределённость знания о результате в два раза. Получен третий бит информации.

Итак, мы построили цепочку сообщений, каждое из которых уменьшало неопределённость знания о результате в два раза, т. е. несло 1 бит информации. Всего было набрано 3 бита информации. Именно столько информации и содержится в сообщении «Книга поставлена на третью полку».

Вычислять количество информации с помощью метода половинного деления возможно, но затруднительно. Гораздо проще воспользоваться следующей закономерностью.

Количество информации **i**, содержащееся в сообщении об одном из **N** равновероятных результатов некоторого события, определяется из формулы:

$$2^i = N.$$

Воспользуемся данной закономерностью для «**Пример 2**» рассмотренного выше.

Предположим, в книжном шкафу восемь полок. Книга может быть поставлена на любую из них. Сколько бит информации несёт сообщение о том, что книга поставлена на третью полку?

Дано:

Решение:

$$N = 8 \quad 2^i = N \quad 2^i = 8 \quad 2^3 = 8 \text{ (чтобы получить 8 надо 2 возвести в третью степень)}$$

$$i = ? \quad i = 3$$

Ответ: $i = 3$ бита.

Алфавитный подход к измерению информации ввёл русский математик Андрей Николаевич Колмогоров (1903 – 1987). В нём предложено отвлечься от содержания информации и рассматривать её как последовательность символов (букв, цифр, кодов цвета точек изображения и т. д.) некоторого алфавита. Этот подход хорошо применять при хранении и передаче информации с помощью технических устройств.

Информация — последовательность символов (букв, цифр, кодов цвета точек изображения и т. д.) некоторого алфавита.

При алфавитном подходе информационный объём сообщения I , состоящего из K символов, вычисляется по формуле:

$$I = K \cdot i,$$

где i — информационный вес символа в битах, связанный с мощностью используемого алфавита N соотношением:

$$2^i = N.$$

Пример 3. Для регистрации на некотором сайте пользователю надо придумать пароль, состоящий из 10 символов. В качестве символов можно использовать десятичные цифры и шесть первых букв латинского алфавита, причём буквы используются только заглавные. Пароли кодируются посимвольно. Все символы кодируются одинаковым и минимально возможным количеством бит.

Необходимо выяснить, какой объём памяти потребуется для хранения 100 паролей.

Решение:

Для каждого пароля можно использовать 16 разных символов (10 цифр и 6 букв латинского алфавита), значит $N = 16$. Из формулы $2^i = N$ найдём информационный вес одного символа i , $2^i = 16 \Rightarrow 2^4 = 16 \Rightarrow i = 4$ бита. Пароль состоит из 10 символов, значит $K = 10$. По формуле $I = K \cdot i$ найдём информационный объём одного пароля. $I = 10 \cdot 4$ бита = 40 бит. По условию задачи требуется найти объём 100 паролей, соответственно информационный объём одного пароля (40 бит) надо умножить на 100. Получим $40 \text{ бит} \cdot 100 = 4000$ бит.

Ответ: 400 бит.

Пример оформления рассмотренной выше задачи:

Дано:

$$\begin{array}{l|l|l} N = 10 + 6 = 16 & I_{100} = 100 \cdot I & 16 = 2^i, i = 4 \text{ (бита на символ)} \\ K = 10 & I = K \cdot i & I = 10 \cdot 4 = 40 \text{ (бит)} = 5 \text{ (байт)} \\ \hline I_{100} = ? & N = 2^i & I_{100} = 100 \cdot 5 = 500 \text{ (байт)} \end{array}$$

Ответ: 500 байт.

Домашнее задание:

Решите задачу: Сообщение, записанное буквами 64-символьного алфавита, содержит 20 символов. Какой объём информации оно несёт?