

7. О КОДИРОВАНИИ И ИЗМЕРЕНИИ ИНФОРМАЦИИ

Машина – не человек, и то, что удобно и хорошо для человека, может быть совершенно неудобно и нехорошо для машины. Как, например, разместить в памяти компьютера слово «галка»? В наших с вами головах это слово размещается как-то «само собой», и мы даже не задумываемся, как. Конечно, для читателей не будет открытием, что для работы с информацией в компьютере используются электрические сигналы. Вероятно, каждой букве русского алфавита можно сопоставить электрический сигнал определенной мощности, и тогда мы легко сможем зашифровать любое русское слово. Но ведь есть еще и латинские буквы, и цифры, и знаки препинания... Понадобится весьма много сигналов разной мощности. А, скажем, при сбоях в электропитании мощность сигнала может внезапно измениться, и в память «залезет» какая-нибудь «губка» вместо «галки». Это наводит нас на мысль о том, что неплохо было бы придумать такую систему кодирования информации при помощи электрических сигналов, чтобы эта система была устойчива к помехам при протекании электрического тока.

Такая система действительно была придумана. Она очень устойчива и очень проста. В ней используются всего два электрических сигнала, которые условно можно назвать «сигналом малой мощности» («тока нет») и «сигналом большой мощности» («ток есть»). Грубо говоря, электрическая сеть не настолько «плоха», чтобы «сигнал малой мощности» из-за помех превратился в сигнал «большой мощности», и наоборот. Условились для простоты и удобства один сигнал обозначать «0», а другой – «1» и назвали их БИТАМИ.

Как же можно с помощью всего двух сигналов закодировать все возможные буквы, цифры и знаки (правильно они называются СИМВОЛЫ)? Очень просто. Каждому символу надо поставить в соответствие конкретную последовательность из нескольких бит (какого-то конкретного их количества, для всех символов – одного и того же!). Тогда при записи текста в память компьютер должен будет сопоставлять каждому символу его «личную» последовательность бит, а при чтении текста, наоборот, отсчитывать в неразберихе из 0 и 1 по нужному количеству бит и определять, что за символ он прочитал, после чего уже можно будет предъявлять этот символ нам в «удобоваримом» виде, скажем, на экране монитора.

Какое же конкретное количество бит надо взять, чтобы с его помощью можно было закодировать все возможные символы? Наверное, вы уже знаете, что для кодирования информации используются последовательности из 8 бит, называемые БАЙТАМИ. 8 – это не много и не мало, а именно то, что нужно. С помощью последовательностей из 8 бит можно закодировать 256 разных символов, так как с помощью последовательностей из n бит можно закодировать 2^n разных символов. Докажем это методом математической индукции.¹

- С помощью последовательностей из 1 бита можно закодировать 2 разных символа, ибо таких последовательностей всего две: 0 и 1. Здесь $n=1$ и поэтому $2=2^n$.
- Пусть с помощью последовательностей из $n-1$ бит можно закодировать 2^{n-1} разных символов. Выясним, сколько существует разных последовательностей из n бит. На первых $n-1$ местах в таких последовательностях могут стоять 2^{n-1} разных комбинаций. С каждой из таких комбинаций в последовательностях из n бит может стоять 0 либо 1 – это увеличивает количество комбинаций из n бит по сравнению с количеством комбинаций из $n-1$ бит ровно в 2 раза: $2 \cdot 2^{n-1} = 2^n$. Нетрудно видеть, что, во-первых, таким путем мы перебираем **все возможные комбинации** (с **каждой** из **всех возможных** комбинаций для $n-1$ бит используется **каждая** из **всех возможных** комбинаций для n -го бита), а во-

¹ Вообще-то можно было бы и не доказывать, так как это элементарная комбинаторная задача, но я думаю, что не помешает.

вторых, все эти комбинации **разные** (с **каждой** из **разных** комбинаций из $n-1$ бит используется **каждая** из **разных** комбинаций для n -го бита).

- Так как доказательство справедливо для $n=1$ и доказано, что, если оно справедливо для $n-1$, то оно справедливо и для n , значит, оно справедливо для $n=2, n=3, n=4$ и т.д., то есть для любого n .

Доказательство можно проиллюстрировать достаточно наглядным рисунком: в нем каждая «ветка» «дерева» соответствует 1 биту, а каждый путь от «корня» до «вершины» - последовательности из стольких бит, сколько на «дереве» уровней «веток».

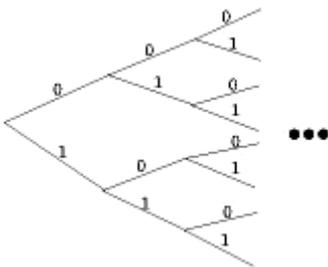


Рис.10. Дерево кодировок.

Байт – это основная единица измерения информации. Понятно, что слово «галка» содержит 5 байт информации, фраза «to be or not to be» - 18 байт (пробел тоже несет в себе информацию!). Существуют и более крупные единицы измерения информации и, соответственно, единицы измерения емкости памяти, в которую эту информацию можно поместить:

- 1 килобайт = 2^{10} = 1024 байта;
- 1 мегабайт = 2^{10} = 1024 килобайта;
- 1 гигабайт = 2^{10} = 1024 мегабайта;
- 1 терабайт = 2^{10} = 1024 гигабайта.

По мере развития научно-технического прогресса, совершенствования технологий крупные единицы измерения информации становятся все актуальнее.