**Измерение информации: содержательный и алфавитный подходы. Единицы измерения информации.**

Определить понятие "количество информации" довольно сложно. В решении этой проблемы существует два основных подхода. Исторически они возникли почти одновременно. В конце 1940 г. один из основоположников кибернетики американский математик Клод Шенон развил **вероятностный подход** к измерению количества информации, а работы по созданию ЭВМ привели к "**объемному подходу**".

**Вероятностный подход**

За единицу количества информации принят 1 бит — количество информации, содержащееся в сообщении, уменьшающем неопределенность знаний в два раза. Далее рассмотрим 2 примера, иллюстрирующих этот факт.

Принята следующая система единиц измерения количества информации:

1 байт = 8 бит

1 Кбайт = 210 байт

1 Мбайт = 210 Кбайт = 220 байт

1 Гбайт = 210 Мбайт = 220 Кбайт = 230 байт

Информацию, которую получает человек, можно считать мерой уменьшения неопределенности знаний. Если некоторое сообщение приводит к уменьшению неопределенности наших знаний, то можно говорить, что такое сообщение содержит информацию. **Чем больше человек знает информации, тем меньше он не знает.**

            Сообщения обычно содержат информацию о каких-либо событиях. Количество информации для событий, которые ***равновероятны***, определяется по формуле:

N = 2I (\*)

 где I – количество информации,

N – количество возможных событий,

**Пример 1.** После экзамена по информатике, который сдавали ваши друзья, объявляются оценки («2», «3», «4» или «5»). Какое количество информации будет нести сообщение об оценке учащегося?

***Будем считать***, что для учащегося все четыре оценки (события) равновероятны и тогда количество информации, которое несет сообщение об оценке можно вычислить по формуле (\*): I = 2 бит (4=2I)

**Пример 2.** Какое количество вопросов достаточно задать вашему собеседнику, чтобы наверняка определить месяц, в котором он родился?

            Будем рассматривать 12 месяцев как 12 возможных событий. Если спрашивать о конкретном месяце рождения, то, возможно, придется задать 11 вопросов (если на 11 первых вопросов был получен отрицательный ответ, то 12-й задавать не обязательно, так как он и будет правильным).

            Правильно задавать «двоичные» (или «закрытые») вопросы, т.е. вопросы, на которые можно ответить только «Да» или «Нет». Например, «Вы родились во второй половине года?». Каждый такой вопрос разбивает множество вариантов на два подмножества: одно соответствует ответу «Да», а другое — ответу «Нет».

            Правильная стратегия состоит в том, что вопросы нужно задавать так, чтобы количество возможных вариантов каждый раз уменьшалось вдвое. В этом случае на каждом шаге ответ («Да» или «Нет») будет нести количество информации 1 бит.

             По формуле (\*) получаем:

12 = 2I, откуда I≈3,5 бит

            Количество полученных бит информации соответствует количеству заданных вопросов, однако количество вопросов не может быть нецелым числом. Округляем до большего целого числа и получаем ответ: при правильной стратегии необходимо задать не более 4 вопросов. *Если в лодку помещается 3,5 человека, то в данном случае необходимо округлить в меньшую сторону.*

**Алфавитный подход**

Согласно данному подходу, количеством информации сообщения считают его длину, то есть общее число знаков в сообщении. Но длина сообщения зависит не только от содержащейся в нем информации. На нее влияет мощность алфавита используемого языка (то есть количество используемых в алфавите знаков, например, мощность русского алфавита 33). Чем меньше знаков в используемом алфавите, тем длиннее сообщение. Так, например, в алфавите азбуки Морзе всего три знака (точка, тире, пауза), поэтому для кодирования каждой русской или латинской буквы нужно использовать несколько знаков, и текст, закодированный по Морзе, будет намного длиннее, чем при обычной записи.

Пример: Сигнал SOS: 3 знака в латинском алфавите;

11 знаков в алфавите Морзе: ··· пауза – – – пауза ···.

Для упорядочивания измерений информационный объем сообщений принято измерять в битах. Один бит соответствует одному знаку двоичного алфавита. Итак, чтобы измерить длину сообщения, его нужно представить в двоичном виде и подсчитать количество двоичных знаков – битов. При этом совсем не обязательно уметь интерпретировать сообщения.

Пример: Пусть сообщение в двоичном алфавите выглядит следующим образом: 000100010001. Мы не знаем, какая информация была заложена в этом сообщении, но можем легко подсчитать его длину – 12 двоичных знаков, следовательно, его информационный объем равен 12-ти битам.

Такой способ измерения количества информации называется алфавитным подходом. При этом измеряется не содержание информации с точки зрения его новизны и полезности, а размер несущего информацию сообщения. Мы уже убедились, что при алфавитном подходе к определению количества информации одни и те же сведения, закодированные по-разному, будут иметь различный информационный объем. Сообщения одинаковой длины могут нести совершенно как совершенно бесполезные сведения, так и нужную информацию.

Пример: применяя алфавитный подход, получаем, что информационный объем слов “фыырпбьощ” и “компьютер” совершенно одинаков, а слов “ученик” и “учащийся” – различен.

Алфавитный подход удобен при подсчете количества информации, хранимого, передаваемого и обрабатываемого техническими устройствами. Действительно, устройствам нет дела до содержательной стороны сообщений. Компьютеры, принтеры, модемы работают не с самой информацией, а с ее представлением в виде сообщений. Оценить информационные результаты их работы как полезные или бесполезные может только человек.

**Единицы измерения информации**

Для удобства, помимо бита используются более крупные единицы измерения количества информации. Вот соотношения между ними:

