Измерение информации

На прошлом занятии мы с вами поговорили об основных действиях, которые мы можем совершать с информацией: получение, обработка, передача и хранение. Во всех этих четырёх случаях, которые называют ***основными информационными процессами***, нам нужно уметь информацию измерять. В случае хранения, чтобы быть уверенными, что объем хранилища и объем нашей информации соответствуют друг другу, в получении/передаче — чтобы знать, сколько по времени будет осуществляться получение/передача информации, в случае обработки, чтобы рассчитать время, за которое этот объем может быть обработан.

1 Содержательный подход

Предположим, что объем информации зависит от ее ***содержания***. Нам нужна информация, которая для нас нова и понятна, то есть та, которая приносит нам новые знания, решает наши задачи.

При содержательном подходе минимальным количеством информации будет ответ на наш вопрос «да» или «нет». Ответ на такой простой вопрос принесет нам минимум информации и уменьшит неопределенность в два раза. Было два варианта, мы выбрали один и получили минимум информации — 1 бит. Например, 1 бит информации получает человек, когда узнаёт выпал снег или нет, опаздывает работник или нет, правильно решено задание или нет и т.д.

Этот подход к измерению предложил Клод Шеннон.

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

1 бит – кол-во информации, уменьшающее неопределённость в 2 раза.

Количество информации *(i*), содержащееся в сообщении об одном из *N* равновероятных результатов некоторого события, определяется из решения уравнения (1):

$$N = 2^i \qquad (1)$$

Степени двойки:

2^0	2^1	2^2	2^3	2^4	2^5	2^6	2^7	2^8	2^9	2^{10}	2^{11}
1	2	4	8	16	32	64	128	256	512	1024	2048

Содержательный подход применяется для измерения информации, используемой человеком.

Пример 1. В корзине лежит 16 шаров разного цвета. Сколько информации несет сообщение, что достали белый шар?

Дано: $N = 16$ шаров ***Найти:*** i = ? (бит)	***Решение:*** Воспользуемся формулой (1): $N = 2^i$ $16 = 2^i \Rightarrow i = 4$ бита, т.к. $2^4 = 16$. Ответ: 4 бита.

2 Алфавитный подход

Разумно так же предположить, что текст, который для вас не понятен, понятен кому-то другому, то есть информация в нем все-таки есть. А ее объем зависит не от содержания текста, а от символов, которыми он написан. Назовем ***алфавитом*** все множество символов, используемых в языке, а их количество — ***мощностью алфавита***.

Каждый символ, выбранный из алфавита мощностью N, несет количество информации (i), вычисленное по формуле (2):

$$N = 2^i \quad (2)$$

Общее количество информации *(I)* во всем тексте можно посчитать по простой математической модели – формула (3):

$$I = k \cdot i \quad (3)$$

где k — количество символов в тексте, i – вес одного символа (бит).

Такой подход к измерению информации называют ***алфавитным***. Здесь объем информации зависит от используемого алфавита и количества символов в тексте.

Этот подход к измерению информации предложил советский ученый-математик А. Н. Колмогоров.

Алфавитный подход применяется для измерения информации, используемой компьютером. Так как компьютер не может работать со смыслом информации.

Пример 2. Определите кол-во информации (бит), который несет в себе один символ алфавита с мощностью 32 символа.

Дано: $N = 32$ символа ***Найти:*** i = ? (бит)	***Решение:*** Воспользуемся формулой (2): $N = 2^i$ $32 = 2^i \Rightarrow i = 5$ бит, т.к. $2^5 = 32$. Ответ: 5 бит.

Пример 3. Найдите информационный объём сообщения (в Кбайтах), состоящего из 8192 символов, если используется алфавит из 256 символов.

Дано: k = 8192 символа N = 256 символов ***Найти:*** I = ? (Кбайт)	***Решение:*** Для нахождения информационного объема I воспользуемся формулой (3): $I = k \cdot i$ Вес одного символа i мы не знаем, но знаем мощность алфавита N, тогда по формуле (2) находим: $N = 2^i$ $256 = 2^i \Rightarrow i = 8$ бит, т.к. $2^8 = 256$. $I = 8192 \cdot 8 = 65536$ бит = 8192 байт = 8 Кбайт. *Ответ:* 8 Кбайт.

Определив подходы и единицы измерения, перейдем к оценкам. Сколько информации содержит книга? Библиотека? Видеоролик? Много? Это конечно, можно посчитать по уже известным нам простым формулам, а вот оценить «много» или «мало» не удастся, потому что это не количественные категории.

Сегодня существует понятие «большие данные». Так называют социально-экономический феномен, связанный с появлением технологических возможностей анализировать огромные массивы данных. Эти технологические возможности стремительно развиваются и уже позволяют компьютерам узнавать нас на фото, советуют нам какую музыку слушать и какие книги читать. Такси безошибочно находит нас в большом городе и проходит тестирование беспилотный транспорт.

Объем данных, которым оперирует человечество, исчисляется единицами зеттабайт и с каждым годом растет в геометрической прогрессии. Данные становятся жизненно-важным активом, а их безопасность — критически важным вопросом.