**Билет 3.**

**Представление информации.**

**Количество и единицы измерения информации.**

**Что такое бит?**

**Дискретная и аналоговая формы представления информации (привести примеры**).

**Представление информации**

Информация может быть представлена в **аналоговой** или **дискретной** форме. Величина в аналоговой форме может принимать бесконечное множество значений. Примерами аналогового представления информации могут служить звук скрипки, картина художника, показатели температуры воздуха, уровня воды в реке.

Величина в дискретной форме может принимать только конечное множество значений. Примеры дискретного представления информации: цифровые показания часов или спидометра, текст в книге, изображение на экране монитора.

Величину в аналоговой форме представления информации можно преобразовать в величину в дискретной форме. Этот процесс называется дискретизацией.

## Единицы измерения информации

Единицы измерения информации служат для измерения объёма информации — величины, исчисляемой логарифмически. Это означает, что когда несколько объектов рассматриваются как один, количество возможных состояний перемножается, а количество информации — складывается. Не важно, идёт речь о случайных величинах в математике, регистрах цифровой памяти в технике или даже квантовых системах в физике.

Чаще всего измерение информации касается объёма компьютерной памяти и объёма данных, передаваемых по цифровым каналам связи.

Впервые объективный подход к измерению информации был предложен американским инженером Р. Хартли в 1928 году, затем в 1948 году обобщен американским учёным К. Шенноном. Хартли рассматривал процесс получения информации как выбор одного сообщения из конечного наперед заданного множества из N равновероятных сообщений, а количество информации I, содержащееся в выбранном сообщении, определял как двоичный логарифм N.

Вероятность - численная мера достоверности случайного события, которая при большом числе испытаний близка к отношению числа случаев, когда событие осуществилось с положительным исходом, к общему числу случаев. Два события называют равновероятными, если их вероятности совпадают.

Примеры равновероятных событий

1.при бросании монеты: «выпала решка», «выпал орел»; 2. на странице книги: «количество букв чётное», «количество букв нечётное»; 3. при бросании игральной кости: «выпала цифра 1»,«выпала цифра 2»,«выпала цифра 3»,«выпала цифра 4»,«выпала цифра 5»,«выпала цифра 6».

Неравновероятные события

Определим, являются ли равновероятными сообщения «первой из дверей здания выйдет женщина» и «первым из дверей здания выйдет мужчина». Однозначно ответить на этот вопрос нельзя. Во-первых, как известно количество мужчин и женщин неодинаково. Во-вторых, все зависит от того, о каком именно здании идет речь. Если это военная казарма, то для мужчины эта вероятность значительно выше, чем для женщины.

Логарифм числа a по основанию b (logba ) равен показателю степени, в которую надо возвести число b, чтобы получить число a. Широкое применение в информатике получили логарифмы по основанию два, которые называют двоичными логарифмами.

*Формула Хартли:*

**I = log2N**

Шеннон предложил другую формулу определения количества информации, учитывающую возможную неодинаковую вероятность сообщений в наборе.

*Формула Шеннона:*

**I=P1log21/P1+P2log21/P2+…+PNlog21/PN,**

где pi – вероятность i-го сообщения

Поскольку каждый регистр арифметического устройства и каждая ячейка памяти состоит из однородных элементов, а каждый элемент может находиться в одном из двух устойчивых состояний (которые можно отождествить с нулем и единицей), то К. Шенноном была введена единица измерения информации – бит.

Бит – слишком мелкая единица измерения. На практике чаще применяется более крупная единица – байт, равная восьми битам. Именно восемь битов требуется для того, чтобы закодировать любой из 256 символов алфавита клавиатуры компьютера (256=28).

Широко используются также еще более крупные производные единицы информации:

1 Килобайт (Кбайт) = 1024 байт,

1 Мегабайт (Мбайт) = 1024 Кбайт,

1 Гигабайт (Гбайт) = 1024 Мбайт.

В последнее время в связи с увеличением объемов обрабатываемой информации входят в употребление такие производные единицы, как:

+1 Терабайт (Тбайт) = 1024 Гбайт,

1 Петабайт (Пбайт) = 1024 Тбайт.

За единицу информации можно было бы выбрать количество информации, необходимое для различения, например, десяти равновероятных сообщений. Это будет не двоичная (бит), а десятичная (дит) единица информации.

**Что такое бит?**

***БИТ — наименьшая единица измерения информации, соответствующая одному разряду машинного двоичного кода.***

Двоичная кодировка (двоичная система счисления) имеет ряд преимуществ перед другими системами кодирования:

1. Для ее реализации нужны технически не сложные элементы с двумя возможными состояниями (есть ток — нет тока, намагничен — не намагничен и т.д.).
2. Представление информации посредством только двух состояний надежно и помехоустойчиво.
3. Возможно применение особой алгебры логики (булевой алгебры) для выполнения логических преобразований информации.
4. Двоичная арифметика намного проще десятичной. Двоичные таблицы сложения и умножения предельно просты.

* Обработка информации в компьютере основана на обмене электрическими сигналами между различными устройствами машины. Признак наличия сигнала можно обозначить цифрой 1, признак отсутствия — цифрой 0.

**ДВОИЧНОЕ КОДИРОВАНИЕ ТЕКСТА**

**Для представления текста в компьютере используется 256 различных знаков. Для кодирования 1 знака отводится 8 битов.**

**Кодирование** – присвоение каждому символу десятичного кода от 0 до 255 или соответствующего ему двоичного кода от 00000000 до 11111111

Присвоение символу определенного кода – это вопрос соглашения, которое фиксируется в кодовой таблице.

В качестве **международного стандарта** была принята кодовая **таблица ASCII** (American Standard Code for Information Interchange) :

***Коды с 0 по 32 (первые 33 кода)*** - коды операций (перевод строки, ввод пробела, т.е. соответствуют функциональным клавишам);

**Коды с 33 по 127** – интернациональные, соответствуют символам латинского алфавита, цифрам, знакам арифметических операций, знакам препинания;

**Коды с 128 по 255** – национальные, т.е. кодировка национального алфавита.

На **1 символ** отводится **1 байт** (8 бит), всего можно закодировать 28= 256 символов

С 1997 года появился новый международный стандарт **Unicode**, который отводит для кодировки одного символа **2 байта** (16 бит), и можно закодировать 65536 различных символов (Unicode включает в себя все существующие, вымершие и искусственно созданные алфавиты мира, множество математических, музыкальных, химических и прочих символов)

В настоящий момент существует пять кодировок кириллицы: КОИ-8, CP1251, CP866, ISO, Mac. Для преобразования текстовых документов из одной кодировки в другую существуют программы которые называются Конверторы

Чтобы подсчитать информационный объем текста необходимо количество информации, которое несет один символов, умножить на количество символов в тексте:

**I = i \* K**