**На 26.03.2020**

**Курс:\_\_\_\_\_1\_\_\_\_\_\_, группа(ы)\_АМ-199**

**Дисциплина (МДК) Информатика**

**ФИО преподавателя\_Патроник Т.А.**

***Тема: Подходы к измерению информации***

При всем многообразии подходов к определению понятия информации, с позиций измерения информации нас интересуют два из них: определение К. Шеннона, применяемое в математической теории информации, и определение А. Н. Колмогорова, применяемое в отраслях информатики, связанных с использованием компьютеров.

В *содержательном подходе* возможна качественная оценка информации: новая, срочная, важная и т.д. Согласно Шеннону, информативность сообщения характеризуется содержащейся в нем полезной информацией - той частью сообщения, которая снимает полностью или уменьшает неопределенность какой-либо ситуации.

Неопределенность некоторого события - это количество возможных исходов данного события. Так, например, неопределенность погоды на завтра обычно заключается в диапазоне температуры воздуха и возможности выпадения осадков.

*Алфавитный подход* основан на том, что всякое сообщение можно закодировать с помощью конечной последовательности символов некоторого *алфавита*. Согласно Колмогорову, информативность последовательности символов не зависит от содержания сообщения, а определяется минимально необходимым количеством символов для ее кодирования.

***Содержательный подход***

Согласно Шеннону, информативность сообщения характеризуется содержащейся в нем полезной информацией — той частью сообщения, которая снимает полностью или уменьшает **неопределенность** какой-либо ситуации.

По Шеннону, информация — уменьшение неопределенности наших знаний.

Неопределенность некоторого события — это количество возможных исходов данного события.

Так, например, если из колоды карт наугад выбирают карту, то неопределенность равна количеству карт в колоде. При бросании монеты неопределенность равна 2.

Содержательный подход часто называют субъективным, так как разные люди (субъекты) информацию об одном и том же предмете оценивают по-разному.

Но если число исходов не зависит от суждений людей (случай бросания кубика или монеты), то информация о наступлении одного из возможных исходов является объективной.

Если сообщение уменьшило неопределенность знаний ровно в два раза, то говорят, что сообщение несет 1 бит информации.

**1 бит — объем информации такого сообщения, которое уменьшает неопределенность знания в два раза.**

Рассмотрим, как можно подсчитать количество информации в сообщении, используя содержательный подход.

Пусть в некотором сообщении содержатся сведения о том, что произошло одно из N равновероятных (равновозможных) событий. Тогда количество информации i, заключенное в этом сообщении, и число событий N связаны формулой: 2i = N. Эта формула носит название **формулы Хартли.** Получена она в 1928 г. американским инженером Р. Хартли.

Если N равно целой степени двойки (2, 4, 8, 16 и т.д.), то вычисления легко произвести "в уме". В противном случае количество информации становится нецелой величиной, и для решения задачи придется воспользоваться таблицей логарифмов либо определять значение логарифма приблизительно (ближайшее целое число, большее).

Например, если из 256 одинаковых, но разноцветных шаров наугад выбрали один, то сообщение о том, что выбрали красный шар несет 8 бит информации (28=256).
Для угадывания числа (наверняка) в диапазоне от 0 до 100, если разрешается задавать только двоичные вопросы (с ответом "да" или "нет"), нужно задать 7 вопросов, так как объем информации о загаданном числе больше 6 и меньше 7 (26<100>27)

***Алфавитный подход***

Алфавитный подход основан на том, что всякое сообщение можно закодировать с помощью конечной последовательности символов некоторого алфавита.

**Алфавит** — упорядоченный набор символов, используемый для кодирования сообщений на некотором языке.

**Мощность алфавита** — количество символов алфавита.
Двоичный алфавит содержит 2 символа, его мощность равна двум.
Сообщения, записанные с помощью символов ASCII, используют алфавит из 256 символов. Сообщения, записанные по системе UNICODE, используют алфавит из 65 536 символов.

С позиций computer science носителями информации являются любые последовательности символов, которые хранятся, передаются и обрабатываются с помощью компьютера. Согласно Колмогорову, информативность последовательности символов не зависит от содержания сообщения, алфавитный подход является объективным, т.е. он не зависит от субъекта, воспринимающего сообщение. Чтобы определить объем информации в сообщении при алфавитном подходе, нужно последовательно решить задачи:

1. Определить количество информации (i) в одном символов по формуле 2i = N, где N — мощность алфавита
2. Определить количество символов в сообщении (К)

Вычислить объем информации по формуле: *I = К\* i*.

Например, если текстовое сообщение, закодированное по системе ASCII, содержит 100 символов, то его информационный объем составляет 800 бит.

Для двоичного сообщения той же длины информационный объем составляет 100 бит. В компьютерной технике бит соответствует физическому состоянию носителя информации: намагничено — не намагничено, есть отверстие — нет отверстия. При этом одно состояние принято обозначать цифрой 0, а другое — цифрой 1.

**Единицы измерения информации**

Решая различные задачи, человек вынужден использовать информацию об окружающем нас мире. И чем более полно и подробно человеком изучены те или иные явления, тем подчас проще найти ответ на поставленный вопрос. Так, например, знание законов физики позволяет создавать сложные приборы, а для того, чтобы перевести текст на иностранный язык, нужно знать грамматические правила и помнить много слов

Часто приходится слышать, что сообщение или несет мало информации или, наоборот, содержит исчерпывающую информацию. При этом разные люди, получившие одно и то же сообщение (например, прочитав статью в газете), по-разному оценивают количество информации, содержащейся в нем. Это происходит оттого, что знания людей об этих событиях (явлениях) до получения сообщения были различными. Поэтому те, кто знал об этом мало, сочтут, что получили много информации, те же, кто знал больше, чем написано в статье, скажут, что информации не получили вовсе. Количество информации в сообщении, таким образом, зависит от того, насколько ново это сообщение для получателя.

Однако иногда возникает ситуация, когда людям сообщают много новых для них сведений (например, на лекции), а информации при этом они практически не получают (в этом нетрудно убедиться во время опроса или контрольной работы). Происходит это оттого, что сама тема в данный момент слушателям не представляется интересной

Итак, количество информации зависит от новизны сведений об интересном для получателя информации явлении. Иными словами, неопределенность (т.е. неполнота знания) по интересующему нас вопросу с получением информации уменьшается. Если в результате получения сообщения будет достигнута полная ясность в данном вопросе (т.е. неопределенность исчезнет), говорят, что была получена исчерпывающая информация. Это означает, что необходимости в получении дополнительной информации на эту тему нет. Напротив, если после получения сообщения неопределенность осталась прежней (сообщаемые сведения или уже были известны, или не относятся к делу), значит, информации получено не было (нулевая информация).

Если подбросить монету и проследить, какой стороной она упадет, то мы получим определенную информацию. Обе стороны монеты "равноправны", поэтому одинаково вероятно, что выпадет как одна, так и другая сторона. В таких случаях говорят, что событие несет информацию в 1 бит. Если положить в мешок два шарика разного цвета, то, вытащив вслепую один шар, мы также получим информацию о цвете шара в 1 бит. Единица измерения информации называется *бит* (bit) - что означает двоичная цифра.

В компьютерной технике бит соответствует физическому состоянию носителя информации: намагничено - не намагничено, есть отверстие - нет отверстия. При этом одно состояние принято обозначать цифрой 0, а другое - цифрой 1. Выбор одного из двух возможных вариантов позволяет также различать логические истину и ложь. Последовательностью битов можно закодировать текст, изображение, звук или какую-либо другую информацию. Такой метод представления информации называется двоичным кодированием

В информатике часто используется величина, называемая *байтом* (byte) и равная 8 битам. И если бит позволяет выбрать один вариант из двух возможных, то байт, соответственно, 1 из 256 (28). В большинстве современных ЭВМ при кодировании каждому символу соответствует своя последовательность из восьми нулей и единиц, т. е. байт.

 Байт – единица измерения информации, представляющая собой последовательность, состоящую из 8 бит: 1 байт = 23 бит = 8 бит.

Наряду с байтами для измерения количества информации используются более крупные единицы

1 Кбайт (один килобайт) = 210 байт = 1024 байта;

1 Мбайт (один мегабайт) = 210 Кбайт = 1024 Кбайта;

1 Гбайт (один гигабайт) = 210 Мбайт = 1024 Мбайта.

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

1 Терабайт (Тб) = 1024 Гбайта = 240 байта

1 Петабайт (Пб) = 1024 Тбайта = 250 байта.

Здесь мы рассмотрим только один, который называется алфавитным подходом

Для закрепления информации просмотрите видео.

<https://www.youtube.com/watch?reload=9&v=qTaaG1oyvrk>

Зайдите по ссылке и ответьте на вопросы теста

<https://forms.gle/dRfVdNstiRzRzdpM7>