#  Измерение информации

#### Подходы к измерению информации

Как измерить информацию? Часто мы говорим, что, прочитав статью в журнале или просмотрев новости, не получили никакой информации, или наоборот, краткое сообщение может оказаться для нас информативным. В то же время для другого человека та же самая статья может оказаться чрезвычайно иноформативной, а сообщение — нет. Информативными сообщения являются тогда, когда они новы, понятны, своевременны, полезны. Но то, что для одного понятно, для другого — нет. То, что для одного полезно, ново, для другого — нет. В этом проблема определения и измерения информации.

При всем многообразии подходов к определению понятия информации, с позиции измерения информации нас будут интересовать два из них: определение К. Шеннона, применяемое в **математической теории информации** **(содержательный подход)**, и определение А. Н. Колмогорова, применяемое в отраслях информатики, связанных с использованием компьютеров (**алфавитный подход)**.

#### 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 неопределеность 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 символо по формуле 2i = N, где N — мощность алфавита
2. Определить количество символов в сообщении (m)
3. Вычислить объем инофрмации по формуле: V = i \* m.

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

Как уже было сказано, **основная единица** измерения информации — бит. 8 бит составляют 1 **байт**.
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.

#### Вопросы для самоконтроля

1. Какие подходы к измерению информации вам известны?
2. Какова основная единица измерения информации?
3. Сколько байт содержит 1 Кб информации?
4. Приведите формулу подсчета количества информации при уменьшении неопределенности знания.
5. Как подсчитать количество информации, передаваемое в символьном сообщении?

В информатике используются различные подходы к измерению информации:

**Содержательный подход к измерению информации.** Сообщение – информативный поток, который в процессе передачи информации поступает к приемнику.  Сообщение несет информацию для человека, если содержащиеся в нем сведения являются для него новыми и понятными  Информация - знания человека ? сообщение должно быть информативно. Если сообщение не информативно, то количество информации с точки зрения человека = 0. (Пример: вузовский учебник по высшей математике содержит знания, но они не доступны 1-класснику)

**Алфавитный подход к измерению информации** не связывает кол-во информации с содержанием сообщения. Алфавитный подход - объективный подход к измерению информации. Он  удобен при использовании технических средств работы с информацией, т.к. не зависит от содержания сообщения. Кол-во информации зависит от объема текста и мощности алфавита. Ограничений на max мощность алфавита нет, но есть достаточный алфавит мощностью 256 символов. Этот алфавит используется для представления текстов в компьютере. Поскольку 256=28, то 1символ несет в тексте 8 бит информации.

**Вероятностный подход к измерения информации.** Все события происходят с различной вероятностью, но  зависимость между вероятностью событий и количеством информации, полученной при совершении того или иного события можно выразить формулой которую в 1948 году предложил  Шеннон.

**Количество информации** - это мера уменьшения неопределенности.

**1 БИТ** – такое кол-во информации, которое содержит сообщение, уменьшающее неопределенность знаний в два раза.  БИТ- это аименьшая единица измерения информации

**Единицы измерения информации**: 1байт = 8 бит

1Кб (килобайт) = 210 байт = 1024 байт

1Мб (мегабайт) = 210 Кб = 1024 Кб

1Гб (гигабайт) = 210 Мб = 1024 Мб

Измерение информации

В информатике, как правило, измерению подвергается информация, представленная дискретным сигналом. При этом различают следующие подходы:

1. структурный. Измеряет количество информации простым подсчетом  информационных элементов, составляющих сообщение. Применяется для оценки возможностей запоминающих устройств, объемов передаваемых сообщений, инструментов кодирования без учета статистических характеристик их эксплуатации.
2. статистический. Учитывает вероятность появления сообщений: более информативным считается то сообщение, которое менее вероятно, т.е. менее всего ожидалось. Применяется при оценке значимости получаемой информации.
3. семантический. Учитывает целесообразность и полезность информации. Применяется при оценке эффективности получаемой информации и ее соответствия реальности.

Структурный подход к измерению информации

В рамках структурного подхода выделяют три меры информации:

* геометрическая. Определяет максимально возможное количество информации в заданных объемах. Мера может быть использована для определения информационной емкости памяти компьютера;
* комбинаторная. Оценивает возможность представления информации при помощи различных комбинаций информационных элементов в заданном объеме. Комбинаторная мера может использоваться для оценки информационных возможностей некоторого системы кодирования;
* аддитивная, или мера Хартли.
* Геометрическая мера
* Определяет максимально возможное количество информации в заданных объемах. Единица измерения – информационный элемент. Мера может быть использована для определения информационной емкости памяти компьютера. В этом случае в качестве информационного элемента выступает минимальная единица хранения – бит. Список самых распространенных более крупных единиц и соотношение между ними приведено ниже:
* 8 бит = 1 байт (сокращенно б или Б),
* 1024 Б = 1 килобайт (сокращенно Кб или К),
* 1024 К = 1 мегабайт (сокращенно Мб или М),
* 1024 М = 1 гигабайт (сокращенно Гб или Г).
* Тогда, например, объем винчестера – 3 гигабайта; объем основной памяти компьютера – 32 мегабайта и т.д.
* Пример 1. Пусть сообщение
* 5555 6666 888888
* закодировано одним из специальных методов эффективного кодирования – кодирование повторений – и имеет вид:
* 5(4) 6(4) 8(6)        .
* Требуется измерить информацию в исходном и закодированном сообщениях геометрической мерой и оценить эффективность кодирования.
* В качестве информационного элемента зададимся символом сообщения. Тогда:
* I(исх.) = l(исх.) = 14 символов;
* I(закод.) = l(закод.) = 12 символов,
* где I(исх.), I(закод.) – количества информации, соответственно, в исходном и закодированном сообщениях;
* l(исх.), l(закод.) – длины (объемы) тех же сообщений, соответственно.
* Эффект кодирования определяется как разница между I(исх.)  и I(закод.) и составляет 2 символа.
* Очевидно, геометрическая мера не учитывает, какими символами заполнено сообщение. Так, одинаковыми по количеству информации, измеренной геометрической мерой, являются, например, сообщения «компьютер» и «программа»; а также 346 и 10В.

Комбинаторная мера

Оценивает возможность представления информации при помощи различных комбинаций информационных элементов в заданном объеме. Использует типы комбинаций элементов и соответствующие математические соотношения, которые приводятся в одном из разделов дискретной математики – [комбинаторике](http://www.klgtu.ru/ru/students/literature/inf_asu/330.html).

Комбинаторная мера может использоваться для оценки информационных возможностей некоторого автомата, который способен генерировать дискретные сигналы (сообщения) в соответствии с определенным правилом комбинаторики. Пусть, например, есть автомат, формирующий двузначные десятичные целые положительные числа (исходное множество информационных элементов {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}). В соответствии с положениями комбинаторики, данный автомат генерирует размещения (различаются числа, например, 34 и 43) из 10 элементов (используются 10 цифр) по 2 (по условию задачи, формируются двузначные числа) с повторениями (очевидно, возможны числа, состоящие из одинаковых цифр, например, 33). Тогда можно оценить, сколько различных сообщений (двузначных чисел) может сформировать автомат, иначе говоря, можно оценить информационную емкость данного устройства: Рп(102) = 102 = 100.

Комбинаторная мера используется для определения возможностей кодирующих систем, которые широко используются в информационной технике.

Пример 1. Определить емкость ASCII-кода, представленного в двоичной или шестнадцатеричной системе счисления.

ASCII-код – это сообщение, которое формируется как размещение с повторениями:

* для двоичного представления – из информационных элементов {0, 1}, сообщение длиной (объемом) 8 символов;
* для шестнадцатеричного представления – из информационных элементов {0, 1, 2, …., А, В, С, …. F}, сообщение длиной (объемом) 2 символа.

Тогда в соответствии с положениями комбинаторики:

I(двоичное) = РП(28) = 28 = 256;

I(шестнадцатеричное) = РП(162) = 162 = 256,

где I(двоичное), I(шестнадцатеричное) – количества информации, соответственно, для двоичного и шестнадцатеричного представления ASCII-кода.

Таким образом, емкость ASCII-кода для двоичного и шестнадцатеричного представления одинакова и равна 256.

Следует отметить, что все коды постоянной длины формируются по правилам комбинаторики или их комбинациям.

В случае, когда сообщения формируются как размещения с повторениями из элементов алфавита мощности h и известно количество сообщений М, можно определить требуемый объем сообщения (т.е. его длину l) для того, чтобы в этом объеме представить все сообщения: l = log h М .

Например, есть 4 сообщения – a, b, c, d. Выполняется двоичное кодирование этих сообщений кодом постоянной длины. Для этого требуются 2 двоичных разряда. В самом деле: l = log 2 4  = 2.

Очевидно, комбинаторная мера является развитием геометрической меры, так как помимо длины сообщения учитывает объем исходного алфавита и правила, по которым из его символов строятся сообщения.

Особенностью комбинаторной меры является то, что ею измеряется информация не конкретного сообщения, а всего множества сообщений, которые могут быть получены.

Единицей измерения информации в комбинаторной мере является число комбинаций информационных элементов.

Аддитивная мера

Эта мера предложена в 1928 году американским ученым Хартли, поэтому имеет второе название – мера Хартли. Хартли впервые ввел специальное обозначение для количества информации – I и предложил следующую логарифмическую зависимость между количеством информации и мощностью исходного алфавита:

I = l log h,

где I – количество информации, содержащейся в сообщении;

l – длина сообщения;

h – мощность исходного алфавита.

При исходном алфавите {0,1}; l = 1; h = 2 и основании логарифма, равном 2, имеем

I = 1\*log22 = 1.

Данная формула даёт аналитическое определение бита (BIT - BInary digiT) по Хартли: это количество информации, которое содержится в двоичной цифре.

Единицей измерения информации в аддитивной мере является бит.

Пример 1. Рассчитать количество информации, которое содержится в шестнадцатеричном и двоичном представлении ASCII-кода для числа 1.

В соответствии с таблицей ASCII-кодов имеем: шестнадцатеричное представление числа 1 – 31, двоичное представление числа 1 – 00110001.

Тогда по формуле Хартли получаем:

для шестнадцатеричного представления        I = 2log216 = 8 бит;

для двоичного представления                I = 8 log22 = 8 бит.

Таким образом, разные представления ASCII-кода для одного символа содержат одинаковое количество информации, измеренной аддитивной мерой.