**Лекция 1.Формальное представление знаний. Виды информации.**

**План лекции**

1. Теория информации – дочерняя наука кибернетики.
2. Формальное представление знаний.
3. Виды информации.
4. Принципы хранения, измерения, обработки и передачи информации

**Содержание**

**Ключевые слова:** [Теория информации](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=1#keyword1), [shannon](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=1#keyword3), [целый](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=1#keyword4), [статистика](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=1#keyword6), [алгебра](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=1#keyword7), [информация](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=1#keyword10), [таблицами кодировки](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=1#keyword11), [ASCII](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=1#keyword12), [standard](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=1#keyword13), [information](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=1#keyword14), [interchange](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=1#keyword15), [программа](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=1#keyword16), [Паскаль](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=1#keyword17), [значение](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=1#keyword23), [таблица кодировки](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=1#keyword24), [reader](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=1#keyword26), [дискретизация](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword32), [модем](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword33), [слово](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword34), [модуляция](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword35), [диск](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword37), [связь](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword38), [АЦП](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword39), [ADC](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword40), [analog](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword41), [digital](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword42), [A/D](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword43), [ЦАП](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword44), [DAC](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword45), [D/A](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword48), [DAT](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword49), [место](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword51), [битом](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword53), [BIT](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword54), [binary digit](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword55), [бит](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword56), [единица](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword59), [байт](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword60), [производные](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword61), [точность](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword62), [электрическая схема](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword63), [компонент](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword64), [скорость передачи](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword65), [baud](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword67), [bps](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword68), [последовательно](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword70), [параллельно](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#keyword72).

1. **Теория информации – дочерняя наука кибернетик****и**

*Теория информации*рассматривается как существенная часть кибернетики.

**Кибернетика**— это наука об общих законах получения, хранения, передачи и переработки информации. Ее основной предмет исследования — это так называемые кибернетические системы, рассматриваемые абстрактно, вне зависимости от их материальной природы. Примеры кибернетических систем: автоматические регуляторы в технике, ЭВМ, мозг человека или животных, биологическая популяция, социум. Часто кибернетику связывают с методами искусственного интеллекта, т.к. она разрабатывает общие принципы создания систем управления и систем для автоматизации умственного труда. Основными разделами (они фактически абсолютно самостоятельны и независимы) современной кибернетики считаются:*теория информации*, теория алгоритмов, теория автоматов, исследование операций, теория оптимального управления и теория распознавания образов.

Родоначальниками кибернетики (датой ее рождения считается 1948 год, год соответствующей публикации) считаются американские ученые Норберт Винер (Wiener, он - прежде всего) и Клод Шеннон (*Shannon*, он же основоположник теории информации).

Винер ввел основную категорию кибернетики – **управление**(основная категория кибернетики), показал существенные отличия этой категории от других, например, энергии, описал несколько задач, типичных для кибернетики, и привлек всеобщее внимание к особой роли вычислительных машин, считая их индикатором наступления новой научно-технической революции (НТР). Выделение категории управления позволило Винеру воспользоваться понятием информации, положив в основу кибернетики изучение законов передачи и преобразования информации.

Сущность принципа управления заключается в том, что движение и действие больших масс или передача и преобразование больших количеств энергии направляется и контролируется при помощи небольших количеств энергии, несущих информацию. Этот принцип управления лежит в основе организации и действия любых управляемых систем: автоматических устройств, живых организмов и т.п. Подобно тому, как введение понятия энергии позволило рассматривать все явления природы с единой точки зрения и отбросило*целый*ряд ложных теорий, так и введение понятия информации позволяет подойти с единой точки зрения к изучению самых различных процессов взаимодействия в природе.

В СССР значительный вклад в развитие кибернетики внесли академики Берг А.И. и Глушков В.М.

В нашей стране в 50-е годы кибернетика была объявлена лженаукой и была практически запрещена, что не мешало, однако, развиваться всем ее важным разделам (в том числе и теории информации) вне связи с обобщающим словом "кибернетика". Это было связано с тем, что сама по себе кибернетика представляет собой род философии, в кое-чем конфликтной с тогдашней официальной доктриной (марксистско-ленинской диалектикой).

*Теория информации*тесно связана с такими разделами математики как теория вероятностей и математическая *статистика*, а также прикладная*алгебра*, которые предоставляют для нее математический фундамент. С другой стороны, теория информации исторически и практически представляет собой математический фундамент теории связи. Часто *теорию информации* вообще рассматривают как одну из ветвей теории вероятностей или как часть теории связи. Таким образом, предмет "*Теория информации*" весьма узок, т.к. зажат между "чистой" математикой и прикладными (техническими) аспектами теории связи.

**Теория информации** представляет собой математическую теорию, посвященную измерению информации, ее потока, "размеров" канала связи и т.п., особенно применительно к радио, телеграфии, телевидению и к другим средствам связи. Первоначально теория была посвящена каналу связи, определяемому длиной волны и частотой, реализация которого была связана с колебаниями воздуха или электромагнитным излучением. Обычно соответствующий процесс был непрерывным, но мог быть и дискретным, когда *информация* кодировалась, а затем декодировалась. Кроме того, теория информации изучает методы построения кодов, обладающих полезными свойствами.

**2. Формальное представление знаний.**

При формальном представлении знаний каждому описываемому объекту или понятию ставится в соответствие некоторый числовой код. Связи между кодируемыми сущностями также представляются кодами (адресами и указателями). Для такого перевода неформальных данных в формальный, цифровой вид должны использоваться специальные таблицы, сопоставляющие кодируемым сущностям их коды и называемые *таблицами кодировки*. Простейший пример такой таблицы – это *ASCII* (American *Standard* Code for *Information* *Interchange*), используемая повсеместно с вычислительной техникой. Она сопоставляет печатным и управляющим символам (управляющими являются, например, символы, отмечающие конец строки или страницы) числа от 0 до 127. Следующая *программа* на языке *Паскаль* выведет на экран все печатные символы этой таблицы и их коды:

\setbox\bzero=\vbox{\hsize=120pt{\prg
var i: byte;
begin
\  for i := 32 to 126 do
\    write(i:6, chr(i):2);
\  writeln
end\rm.}}
\centerline{\box\bzero}

На практике обычно используют не сам исходный *ASCII*, а так называемый расширенный *ASCII*(*ASCII*+), описывающий коды 256 символов (от 0 до 255). Первые 128 позиций расширенного *ASCII* совпадают со стандартом, а дополнительные 128 позиций определяются производителем оборудования или системного программного обеспечения. Кроме того, некоторым управляющим символам *ASCII* иногда назначают другое *значение*.

Хотя таблицы *кодировки* используются для формализации информации, сами они имеют неформальную природу, являясь мостом между реальными и формальными данными. Например, коду 65 в *ASCII* соответствует заглавная латинская буква A, но не конкретная, а любая. Этому коду будет соответствовать буква**A**, набранная жирным прямым шрифтом, и буква\sl A, набранная нежирным с наклоном вправо на9.5^\circшрифтом, и даже буква\frak
Aготического шрифта. Задача сопоставления реальной букве ее кода в выбранной таблице кодировки очень сложна и частично решается программами распознания символов (например, Fine*Reader*).

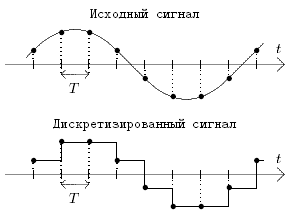
**Упражнение 1.** Каков код букв W и w в*ASCII*?

**3. Виды информации**

*Информация* в системах коммуникации и обработки данных может быть представлена в двух видах: **дискретная информация** и **непрерывная** (**аналоговая**). Дискретная *информация* характеризуется последовательными точными значениями некоторой величины, а непрерывная - непрерывным процессом изменения некоторой величины. Непрерывную информацию может, например, выдавать датчик атмосферного давления или датчик скорости автомашины. Дискретную информацию можно получить от любого цифрового индикатора: электронных часов, счетчика магнитофона и т.п.

Дискретная *информация* удобнее для обработки человеком, но непрерывная *информация* часто встречается в практической работе, поэтому необходимо уметь переводить непрерывную информацию в дискретную (*дискретизация*) и наоборот. *Модем* (это *слово* происходит от слов *модуляция* и демодуляция) представляет собой устройство для такого перевода: он переводит цифровые данные от компьютера в электромагнитные колебания-копии звука и наоборот. Этой теме посвящено практическое занятие 1.

При переводе непрерывной информации в дискретную важна так называемая **частота дискретизации**\nu, определяющая период (T=1/\nu) между измерениями значений непрерывной величины (См.[рис. 1.1](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#image.1.1)).



**Рис. 1.1.**

Чем выше частота дискретизации, тем точнее происходит перевод непрерывной информации в дискретную. Но с ростом этой частоты растет и размер дискретных данных, получаемых при таком переводе, и, следовательно, сложность их обработки, передачи и хранения. Однако для повышения точности дискретизации необязательно безграничное увеличение ее частоты. Эту частоту разумно увеличивать только до предела, определяемого теоремой о выборках, называемой также теоремой Котельникова или законом Найквиста (Nyquist).

Любая непрерывная величина описывается множеством наложенных друг на друга волновых процессов, называемых гармониками, определяемых функциями видаA\sin(\omega t+\varphi), гдеA- это амплитуда,\omega- частота,t- время и\varphi- фаза.

**Теорема о выборках** утверждает, что для точной дискретизации ее частота должна быть не менее чем в два раза выше наибольшей частоты гармоники, входящей в дискретизируемую величину.

Примером использования этой теоремы являются лазерные компакт-диски, звуковая *информация* на которых хранится в цифровой форме. Чем выше будет частота дискретизации, тем точнее будут воспроизводиться звуки и тем меньше их можно будет записать на один *диск*, но ухо обычного человека способно различать звуки с частотой до 20КГц, поэтому точно записывать звуки с большей частотой бессмысленно. Согласно теореме о выборках частоту дискретизации нужно выбрать не меньшей 40КГц (в промышленном стандарте на компакт-диске используется частота 44.1КГц).

При преобразовании дискретной информации в непрерывную, определяющей является скорость этого преобразования: чем она выше, с тем более высокочастотными гармониками получится непрерывная величина. Но чем большие частоты встречаются в этой величине, тем сложнее с ней работать. Например, обычные телефонные линии предназначены для передачи звуков частотой до 3КГц. *Связь* скорости передачи и наибольшей допустимой частоты подробнее будет рассмотрена далее.

Устройства для преобразования непрерывной информации в дискретную обобщающе называются *АЦП* (аналого-цифровой преобразователь) или *ADC* (*Analog*to*Digital*Convertor, *A/D*), а устройства для преобразования дискретной информации в аналоговую – *ЦАП* (цифро-аналоговый преобразователь) или *DAC* (*Digital*to*Analog* Convertor, *D/A*).

**Упражнение 2.** В цифровых магнитофонах *DAT* частота дискретизации - 48КГц. Какова максимальная частота звуковых волн, которые можно точно воспроизводить на таких магнитофонах?

**4. Принципы хранения, измерения, обработки и передачи информации**

Для хранения информации используются специальные устройства памяти. Дискретную информацию хранить гораздо проще непрерывной, т.к. она описывается последовательностью чисел. Если представить каждое число в двоичной системе счисления, то дискретная *информация* предстанет в виде последовательностей нулей и единиц. Присутствие или отсутствие какого-либо признака в некотором устройстве может описывать некоторую цифру в какой-нибудь из этих последовательностей. Например, позиция на дискете описывает *место* цифры, а полярность намагниченности – ее *значение*. Для записи дискретной информации можно использовать ряд переключателей, перфокарты, перфоленты, различные виды магнитных и лазерных дисков, электронные триггеры и т.п. Одна позиция для двоичной цифры в описании дискретной информации называется *битом* (*bit*,*binary digit*) .*Бит* служит для измерения информации. *Информация* размером в один *бит* содержится в ответе на вопрос, требующий ответа "да" или "нет". Непрерывную информацию тоже измеряют в битах.

**Бит** - это очень маленькая *единица*, поэтому часто используется величина в 8 раз большая - **байт (byte)**, состоящая из двух 4-битных полубайт или тетрад. *Байт* обычно обозначают заглавной буквой B или Б. Как и для прочих стандартных единиц измерения для бита и байта существуют *производные* от них единицы, образуемые при помощи приставок кило (K), мега (M), гига (G или Г), тера (T), пета (P или П) и других. Но для битов и байтов они означают не степени 10, а степени двойки: кило -2^{10}=1024 \approx 10^3, мега -2^{20} \approx 10^6, гига -2^{30}\approx 10^{9}, тера -2^{40}\approx 10^{12}, пета -2^{50}\approx 10^{15}. Например, 1KB = 8Кbit = 1024B = 8192bit,1МБ = 1024КБ = 1048576Б = 8192Кбит.

Для обработки информации используют вычислительные машины, которые бывают двух видов: ЦВМ (цифровая вычислительная машина) - для обработки дискретной информации, АВМ (аналоговая вычислительная машина) - для обработки непрерывной информации. ЦВМ - универсальны, на них можно решать любые вычислительные задачи с любой точностью, но с ростом точности скорость их работы уменьшается. **ЦВМ**- это обычные компьютеры, или сокращенно ЭВМ.

Каждая АВМ предназначена только для узкого класса задач, например, интегрирования или дифференцирования. Если на вход такой АВМ подать сигнал, описываемый функциейf(t) то на ее выходе появится сигналF(t)илиf'(t). АВМ работают очень быстро, но их *точность* ограничена и не может быть увеличена без аппаратных переделок. **Программа для АВМ**- это *электрическая схема* из заданного набора электронных *компонент*, которую нужно физически собрать.

Бывают еще и гибридные вычислительные машины, сочетающие в себе элементы как ЦВМ, так и АВМ.

На[рис.1.2](https://intuit.ru/studies/courses/2256/140/lecture/3904?page=2#image.1.2)и рис. 1.3 изображена схема передачи информации в различных представлениях.



Рис. 1.2.

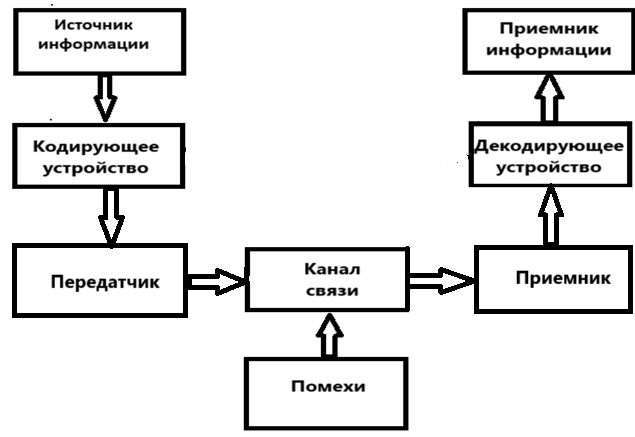


Рис.1.3.

Кодированием, например, является шифровка сообщения с целью обеспечения ее конфиденциальности или помехоустойчивое кодирование. Декодирование – обратный процесс кодированию, например его дешифровка. Процедуры кодирования и декодирования могут повторяться много раз. Ошибки при передаче информации происходят из-за шума (помех) в канале (атмосферные и технические помехи), а также при кодировании и декодировании. Теория информации изучает, в частности, способы минимизации количества таких ошибок.

*Скорость передачи* информации измеряется в количестве переданных за одну секунду *бит* или в бодах (*baud*): 1бод = 1бит/сек (*bps*). *Производные* единицы для бода такие же как и для бита и байта, например, 10Kbaud = 10240baud.

Информацию можно передавать *последовательно*, т.е. *бит* за битом, и *параллельно*, т.е. группами фиксированного количества *бит*. Параллельный способ быстрее, но он часто технически сложнее и дороже особенно при передаче данных на большие расстояния. Параллельный способ передачи используют, как правило, только на расстоянии не более 5 метров. Например, между определенными внешними устройствами ЭВМ и самой ЭВМ через параллельные порты и между основными внутренними устройствами самой ЭВМ.

Выполнить упражнения.

**Упражнение 3.**Сколько *бит* в одном килобайте?

**Упражнение 4.**Что больше 12,5Мбайт или 512000байт?

**Задание к ПЗ 1**.

1. Изучить материал, представленный в файле Приложение\_1.