

Лекция №1 — Основные понятия теории информации

1 О курсе

- **Преподаватель:** Колесников Никита Сергеевич
<https://crypto-kantiana.com/nikita.kolesnikov>
e-mail: kn@tehnika39.ru
VK: <https://vk.com/id6424979>
- **Продолжительность:** 1 семестр
- **Отчетность:** Зачет

Список литературы

- [1] Н. Н. Поздниченко А. С. Гуменюк. *Прикладная теория информации*. изд-во ОмГТУ, 2015.
- [2] В.А. Фурсов. *Лекции по теории информации*. изд-во СГАУ, 2006.
- [3] С.И. Чечёта. *Введение в дискретную теорию информации и кодирования*. Пиксел, 2013.

2 Предмет теории информации

Термин «**информация**» происходит от латинского слова «*informatio*», что означает «разъяснения», и, по сути, предполагает наличие некоторого диалога между отправителями и получателями информации. Этот диалог можно представить себе (условно) в виде пяти составляющих компонент:

1. *физической* (источник и приемник сигналов, среда распространения сигнала);
2. *сигнальной* (способ кодирования и декодирования сигнала);
3. *синтаксический* (правила формирования сообщений);
4. *семантической* (смысловая нагрузка сообщения);
5. *прагматической* (мотивация передачи сообщения, его полезность).

В настоящем курсе термин «**информация**» понимается в узком смысле, принятом при описании так называемых информационных систем. К ним относятся телекоммуникационные и вычислительные сети, автоматизированные системы управления и контроля и т.п.

Информационные системы – это класс технических систем, предназначенных для хранения, передачи и преобразования информации. Соответственно **информация** – это сведения, являющиеся объектом хранения, передачи и преобразования.

Теория информации - раздел математики, в котором изучаются свойства информации на синтаксическом уровне. Она состоит из двух основных разделов:

1. **Теория информации** - рассматривает задачи о количественной оценке информации на основе частот встречаемости (вероятностей) различных знаков и слов, не затрагивая смысла (семантики) и ценности (прагматики) информации;
2. **Теория кодирования** - изучает методы преобразования информации для ее экономного представления или для обнаружения и исправления ошибок, возникающих при ее передаче и хранении.

Основополагающие работы по теории информации:

- Клод Шеннон 1948 г. "Математическая теория связи"
- работы Р. Хартли, В.А. Котельникова, А.Н. Колмогорова.

3 Измерение информации

Информацию можно измерить:

- По количеству знаков, которые используются для ее записи (если речь идет об информационных системах);
- По тому, насколько она важна в данной ситуации или какой в нее заложен смысл (субъективная оценка), на сколько снижается неопределенность.

Неопределенность - одно из базовых понятий теории информации, которому нельзя дать точного определения. Неопределенность нельзя измерить абсолютно, однако можно понять, уменьшается она или увеличивается, и во сколько раз.

Например, неопределенность можно снизить, если сообщить получателю информацию об объекте, которая ему нужна. Если информация увеличивает неопределенность, то она называется **недостовойной** (не будем рассматривать). Если информация не изменяет неопределенность, то она называется **бесполезной**.

Неопределенность информации можно эффективно описать средствами теории вероятности с помощью случайных величин, поэтому мы будем рассматривать

4 Представление информации

Будем рассматривать меры количественной оценки информации, порожденной дискретными источниками (ДИС). Для описания модели дискретного источника хорошо подходят дискретные случайные величины. В простейшем случае источник полностью характеризуется множеством сообщений и набором вероятности их появления, т.е. законом распределения соответствующей случайной величины:

a_1	a_2	$\dots$	a_m
p_1	p_2		p_m

Задача: Построить меру F , характеризующую неопределенность выбора очередного сообщения источника. При этом мера должна удовлетворять следующим естественным требованиям:

- **Монотонность.** Мера должна монотонно возрастать с ростом числа сообщений источника, но при этом невозможные сообщения источника должны игнорироваться;
- **Аддитивность.** Пусть имеется два независимых ДИС, генерирующих сообщения из множеств $A = (a_1, \dots, a_m)$ и $B = (b_1, \dots, b_n)$. Их можно рассматривать как один источник, генерирующий пары сообщений (a_i, b_j) . Естественное предположение: неопределенность объединенного источника должна равняться сумме неопределенностей отдельных источников, т.е. $F(mn) = F(m) + F(n)$.

Этим свойствам удовлетворяет логарифмическая мера, предложенная Р.Хартли в 1928г., **мера Хартли**:

$$F(m) = \log_b(m)$$

Замечание. Основание логарифма b , вообще говоря, значения не имеет, т.к. по определению количество неопределенности - величина относительная, а не абсолютная. Важно зафиксировать основание b и не менять его в рамках решения одной задачи. В зависимости от b единицу измерения неопределенности называют по-разному:

- $b = 2$ - бит (общепринято);
- $b = 10$ - дит;
- $b = e$ - нит.

Проблема: Мера Хартли не учитывает изменение неопределенности при изменении вероятностей появления тех или иных сообщений. Хотим, чтобы дополнительно выполнялись следующие свойства меры неопределенности:

- **Зависимость от закона распределения.** Наибольшее значение количества информации достигается при равномерном распределении, а наименьшее - при вырожденном распределении.

- **Симметричность.** Мера не должна зависеть от перенумерации сообщений источника.
- **Непрерывность.** При малом изменении распределения вероятностей $(p_1, \dots, p_m)$ значение неопределенности не должно сильно меняться.

Перечисленным свойствам удовлетворяет мера К.Шеннона, которая называется **энтропией** и вычисляется по формуле:

$$H(A) = - \sum_{i=1}^{\infty} p_i \log_2 p_i$$

(Полагаем по определению $\log_2 0 = 0$)

Заметим, что энтропия случайной величины не зависит от того, какие значения она принимает, а зависит только от того, с какой вероятностью она принимает эти значения.