Передача информации (данных) осуществляется между двумя абонентами, называемыми **источником сообщения (ИcС)** и **получателем сообщения (ПС)**. Третьим элементом информационной системы является **канал (среда) передачи**, связывающий ИсС и ПС.

Отображение сообщения обеспечивается **изменением какой-либо физической величины, характеризующей процесс** (например, амплитуда, частота, фаза). Эта величина является **информационным параметром сигнала** (в общем случае – информационной системы).

Сигналы, как и сообщения, могут быть ***непрерывными (аналоговые) и прерывными (дискретными)***. Информационный параметр непрерывного сигнала стечением времени может принимать любые мгновенные значенияв определенных пределах. Каналы и устройства, функционирующие на основе такого типа сигналов называют аналоговыми.

***Дискретный сигнал*** *(устройство или канал передачи) характеризуется конечным числом значений информационного параметра.*

*Дискретные сообщения состоят из последовательности дискретных знаков. Часто этот параметр принимает всего два значения (0 или 1). Сообщение или канал его передачи на основе этих двух значений сигнала называют двоичным или бинарным.*

*Построение сигнала по определенным правилам, обеспечивающим соответствие между сообщением и сигналом, называют* ***кодированием****.*

*Кодирование в широком смысле – преобразование сообщения в сигнал.*

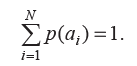
*Кодирование в узком смысле – представление исходных символов, в другом алфавите с меньшим числом знаков. Оно осуществляется с целью повышения надежности и преобразования сигналов к виду, удобному для передачи по каналам связи.*

Важнейшая характеристика источника, получателя или канала – алфавит.

*Алфавит*, *А* – это общее число знаков или символов (*N*), используемых для генерации или передачи сообщений. Символы алфавита будем обозначать через {*аi*}, где 1 ≤ *i* ≤ *N*; *N* – *мощность* *алфавита*.

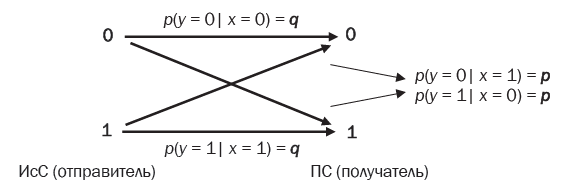
Минимальное число элементов алфавита *N*min = 2, ***А*** = {0, 1} – двоичный код. Один дискретный знак представляет собой *элементарное сообщение*, последовательность знаков – сообщение.

Набор элементов алфавита, создаваемых дискретным источником сообщений, заранее известен получателю. ИсС в каждый дискретный момент времени выдает один элемент алфавита. Понятно, что ПС заранее не известно, какой это элемент. Если обозначить вероятность выбора каждого элемента алфавита *p*(*аi*), то



Чтобы рассчитать вероятность того, что определенный символ (например, буква или цифра) появится в тексте, мы можем **анализировать частоту** его появления в большом документе (не менее нескольких десятков килобайт).

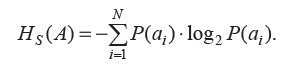
***Двоичный канал передачи информации***строится на основе двоичного алфавита: ***А*** = {0, 1}. При этом канал, в котором вероятности искажения переданного 0 и переданной 1 равны, как и равны вероятности передачи 0 и 1, называют **двоичным симметричным каналом (ДСК).**

****

**p + q = 1**

Информационной характеристикой алфавита (источника сообщений на основе этого алфавита) является ***энтропия***.

Энтропию алфавита ***А*** = {*ai*} по К. Шеннону рассчитывают по следующей формуле:



**С физической точки зрения энтропия алфавита показывает, какое количество информации приходится в среднем на один символ алфавита.**

Частным случаем энтропии Шеннона считается энтропия Хартли. Дополнительным условием при этом является то, что все вероятности одинаковы и постоянны для всех символов алфавита. С учетом этого формулу (2.1) можно преобразовать к виду:



Сообщение *Хk*, которое состоит из *k* символов, должно характеризоваться определенным *количеством информации I*(*Хk*):



Здесь *Н*(*А*) – энтропия алфавита с соответствующим распределением вероятностей *р*(*аi*).

Если принять, что *р*(*аi* = 1) = *р*(1) и *р*(*аi* = 0) = *р*(0), используя выражение (2.1), вычислим энтропию бинарного алфавита:



Если все символы равновероятны, то у нас максимальная неопределённость при выборе следующего символа. Вся информация распределена наиболее равномерно, и у нас нет предпочтений для какого-либо символа.

Если хотя бы один символ встречается чаще, предсказуемость системы возрастает, а неопределённость (энтропия) уменьшается.

* **Максимальная энтропия = максимум неопределённости** (равномерное распределение вероятностей).
* **Нулевая энтропия = полная предсказуемость** (все сообщения одинаковые).
* **Чем меньше энтропия, тем выше предсказуемость**.

**1. Что такое алфавит источника сообщения?**

*Алфавит*, *А* – это общее число знаков или символов (*N*), используемых для генерации или передачи сообщений. Символы алфавита будем обозначать через {*аi*}, где 1 ≤ *i* ≤ *N*; *N* – *мощность* *алфавита*.

**2. Что такое мощность алфавита источника сообщения?**

общее число знаков или символ

**3. Какова мощность алфавита белорусского языка?**

32

**4. Какова мощность алфавита русского языка?**

33

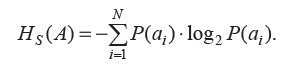
**5. Какова мощность алфавита «компьютерного» языка?**

2

**6. Что такое энтропия алфавита?**

Информационной характеристикой алфавита (источника сообщений на основе этого алфавита) является *энтропия*.

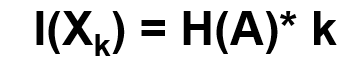
Энтропию алфавита *А* = {*ai*} по К. Шеннону рассчитывают по следующей формуле:



С физической точки зрения энтропия алфавита показывает, какое количество информации приходится в среднем на один символ алфавита.

**7. Что такое энтропия сообщения?**

Энтропия сообщения — это среднее количество информации, передаваемое в сообщении, составленном из символов алфавита.

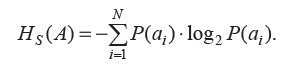


**8. От чего зависит энтропия алфавита?**

Размерности алфавита — чем больше различных символов в алфавите, тем потенциально выше энтропия.

Распределения вероятностей символов — максимальная энтропия достигается, когда все символы имеют равные вероятности (равномерное распределение).

**9. Записать формулу для вычисления энтропии.**



**10. Что нужно знать для вычисления энтропии алфавита?**

Состав алфавита — множество возможных символов (например, {A, B, C} или {0,1} в двоичной системе).

Вероятности появления каждого символа P(xi)P(x\_i)P(xi​) — с какой вероятностью каждый символ встречается в сообщении.

**11. Как рассчитываются энтропия Шеннона и энтропия Хартли? В чем принципиальное различие между этими характеристиками? Дайте толкование физического смысла энтропии.**

Частным случаем энтропии Шеннона считается энтропия Хартли. Дополнительным условием при этом является то, что все вероятности одинаковы и постоянны для всех символов алфавита. С учетом этого формулу (2.1) можно преобразовать к виду:



С физической точки зрения энтропия алфавита показывает, какое количество информации приходится в среднем на один символ алфавита.

**12. Пояснить назначение знака «минус» в формулах (2.1) и (2.4).**

Знак «минус» нужен для того, чтобы сделать энтропию неотрицательной.

**13. Что такое избыточность алфавита и избыточность сообщений, сформированных в компьютерных системах? Принцип действия каких систем основан на существовании данной избыточности?**

Избыточность в теории информации — это мера излишней информации, которая не несёт нового смысла, но может быть полезна для исправления ошибок или удобства восприятия.

Избыточность используется для повышения надёжности хранения и передачи информации.

**14. Расположить в порядке возрастания энтропии известные Вам алфавиты.**

Двоичный 1, английский 4.7, русский 5.1

**15. Вычислить энтропию алфавита белорусского (русского) языка.**

**16. Вычислить энтропию Шеннона бинарного алфавита, если вероятность появления в произвольном документе на основе этого алфавита одного из символов составляет 0,25, другого – 0,75; либо 0 и 1; либо 0,5 и 0,5.**

H = −(0.25 ​0.25 + 0.75 ​0.75) = − (0.25 × (−2) + 0.75 × (−0.415)) = − (− 0.5 − 0.311) = 0.811 бит

H = −(0 ​0 + 1 ​1) = 0 бит

H = −(0.5 ​0.5 + 0.5 ​0.5) = − 2 (0.5 × (−1)) = −2 (− 0.5) = 1 бит

17. Чему равна энтропия алфавита по Хартли, если мощность этого алфавита равна: а) 1 символ; б) 2 символа; в) 8 символов?

H = ​1 = 0 бит

H = ​2 = 1 бит

H = ​8 = 3 бита