

1. Что такое алфавит источника сообщения?

Алфавит, A – это общее число знаков или символов (N), используемых для генерации или передачи сообщений.

2. Что такое мощность алфавита источника сообщения?

Мощность алфавита источника сообщения (N) — это количество символов в алфавите.

Минимальное число элементов алфавита $N_{\min} = 2$, $A = \{0, 1\}$ – двоичный код. Один дискретный знак представляет собой элементарное сообщение, последовательность знаков – сообщение.

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)

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

8. Что такое энтропия сообщения?

Среднее количество информации на один символ в сообщении, измеряется в битах.

Если про усл. энтропию:

Это энтропия, которая означает потерю инфы на каждый символ передаваемого сообщ.

8. От чего зависит энтропия алфавита?

От распределения вероятностей встречи того/иного символа.

!не зависит от алфавита и его мощности!

9. Записать формулу для вычисления энтропии.

Шенон:

$$H_S(A) = -\sum_{i=1}^N P(a_i) \cdot \log_2 P(a_i).$$

Хартли: (если все вероятности одинаковы и постоянны для всех символов алфавита)

$$H_{Ch}(A) = \log_2 N.$$

Двоичный: (Если принять, что $p(a_i = 1) = p(1)$ и $p(a_i = 0) = p(0)$)

$$H(A_2) = -p(0) \cdot \log_2(p(0)) - p(1) \cdot \log_2(p(1)).$$

Условная:

$$H(Y | X) = -p \cdot \log_2 p - q \cdot \log_2 q.$$

Эффективная: (Если вероятность ошибки в ДСК отлична от 0 ($p > 0$), переданное сообщение может содержать ошибки: $X_k \neq Y_k$. Количество информации в таком сообщении при его передаче по ДСК будет определяться не энтропией двоичного алфавита (в соответствии с выражением (2.3)), а эффективной энтропией $H_e(A)$ алфавита или пропускной способностью канала:)

$$H_e(A) = 1 - H(Y | X),$$

10. Что нужно знать для вычисления энтропии алфавита?

Вероятность встречи символа.

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

Шенон:

$$H_S(A) = -\sum_{i=1}^N P(a_i) \cdot \log_2 P(a_i).$$

Хартли: Частным случаем энтропии Шеннона считается энтропия Хартли.

Дополнительным условием при этом является то, что все вероятности одинаковы и постоянны для всех символов алфавита.

$$H_{Ch}(A) = \log_2 N.$$

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).

Логарифм чисел до 1 - отрицательный, а энтропия дб положительна.

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

Если подсчитать энтропию Хартли и энтропию Шеннона для одного и того же алфавита – они окажутся неравны, это неравенство и указывает на избыточность любого алфавита, при $N > 2$.

Избыточность алфавита и сообщений — это излишек символов или данных, который позволяет исправлять ошибки

Для снижения избыточности – сжатие.

Применяется в криптоанализе.

14. Расположить в порядке возрастания энтропии известные Вам алфавиты.

англ -> русск -> белорусский

15. Вычислить энтропию алфавита белорусского (русского) языка.

$$\log_2(33) = 5,044 - \text{Rus}$$

$$\log_2(32) = 5 - \text{Bel}$$

$$\log_2(26) = 4,7 - \text{Eng}$$

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 1,0; либо 0,5 и 0,5.

- 1) $H(A_2) = -p(0)\log_2(p(0)) - p(1)\log_2(p(1))$
- 2) $-0,25 \cdot \log_2(0,25) - 0,75 \cdot \log_2(0,75) = 0,811$
- 3) $-0 \cdot \log_2(0) - 1 \cdot \log_2(1) = 0$
- 4) $-0,5 \cdot \log_2(0,5) - 0,5 \cdot \log_2(0,5) = 1$

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

$$\log_2(1) = 0$$

$$\log_2(2) = 1$$

$$\log_2(8) = 3$$

! Чем больше мощность алфавита, тем более информативен каждый символ

Количество информации = энтропия * число символов в сообщ.

Физ.смысл.понятия информации в теории Шеннона: информацией является лишь то сообщение, которое снимает некую неопределенность, т.е.содержит новые для получателя данные.

Если известно, что сообщение будет состоять из набора одинаковых символов, то для получателя сообщения оно никакой неопределенности не содержит.

Физ.смысл усл.энтропии – показывает потери информации при передаче с вероятностью ошибки

- потерю информации на каждый символ передаваемого сообщения

Физ.смысл эф.энтропии – показывает пропускную способность канала

Виды избыточности:

- 1) временная (многократн.повторение 1 и того же)
- 2) информационная (введение доп.разрядов в слова)
- 3) комбинированная

Почему усл.энтропия одинакова, если менять местами вер-сть прав/ошиб.передачи?

Можем инвертировать биты сообщения

Св-ва 2сс канала:

2сс сим.канал – канал, в котором при передаче сообщ.с одинак.вер-стью появл.ошибка типа 0-1 и 1-0.

*дискретный – состоит из 2 символов (0,1)

*если вер-сть появления символа стремится к 0, то и энтропия также будет стремиться к 0.

*макс.энтропия – 1, при усл.равнвероятного появления каждого символа.

*вероятности искажения переданного 0 и переданной 1 равны, как и равны вероятности передачи 0 ($p(0)$) и 1 ($p(1)$)