# **Лабораторная работа №2**

**Элементы теории информации. Параметры и характеристики дискретных информационных систем**

Цель: приобретение практических навыков расчета и анализа параметров и информативных характеристик дискретных ИС.

Задачи:

1. Закрепить теоретические знания по основам теории информации.

2. Разработать приложение для расчета и анализа параметров и информативных характеристик дискретных ИС.

3. Результаты выполнения лабораторной работы оформить в виде описания разработанного приложения, методики выполнения экспериментов с использованием приложения и результатов эксперимента.

**Теоретические сведения**

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



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:



Учитывая, что сумма р(1) + р(0) = 1, и выражая одну вероятность через другую (например, р(1) = 1 – р(0)), можно теоретически доказать информативность бинарного алфавита, решив дифференциальное уравнение [dH(A)/dp(1)] = 0.

Если вероятность ошибки в ДСК отлична от 0 (р > 0), переданное сообщение может содержать ошибки: Хk ≠ Yk. 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 Hе(A) алфавита или пропускной способностью канала:



где H(Y | X) – условная энтропия:



**Практическое задание**

1. рассчитать энтропию указанных преподавателем алфавитов: один – на латинице, другой – на кириллице (по формуле (2.1) перейти от частоты появления каждого символа алфавита к соответствующей вероятности); в качестве входного может быть принят произвольный электронный текстовый документ на основе соответствующего алфавита; частоты появления символов алфавитов оформить в виде гистограмм (можно воспользоваться приложением MS Excel).

**Код программы**

def Ubung1(inputStr = "", language = "Latin", p = 0, numbLineRead = 1):  
  
 if(inputStr != ""):  
 text = inputStr  
 else:  
 language = input("Введите язык(Latin or Kiril): ")  
 inp = int(input("Выберите метод ввода(1 - из файла, 2 - ввести ручками): "))  
 match inp:  
 case 1:  
 fileName = input("Введите название файла: ")  
 text = OpenFile(fileName, numbLineRead)  
 case 2:  
 text = input("Введите текст: ")  
  
 if(language == "Latin"):  
 alfLat = 65  
 endAlf = 123  
 else:  
 alfLat = 1040  
 endAlf = 1103  
  
 entropy = 0  
 helper = 0  
 a = []  
 b = []  
 aFreq = 0  
  
 while alfLat < endAlf:  
 for i in text:  
 if ord(i) == alfLat:  
 helper = helper + 1  
 if helper > 0:  
 entropy += (helper / len(text)) \* math.log2(helper / len(text))  
 if (alfLat < 91 or alfLat > 96):  
 a.insert(aFreq, (helper / len(text)))  
 b.insert(aFreq, chr(alfLat))  
 helper = 0  
 alfLat += 1  
 entropy = -entropy  
  
 if p != 0 and p != 1:  
 entropy -= (-p \* math.log2(p) - (1 - p) \* math.log2(1 - p))  
 elif p == 1:  
 entropy -= entropy  
 print("Текст: " + text)  
 print("Энтропия текста равна: " + str(entropy))  
 print("Количество информации: " + str(entropy \* len(text)))

**Результат выполнения программы**

**Исходный текст(латиница):**

AllinthegoldenafternoonFullleisurelyweglideForbothouroarswithlittleskillBylittlearmsarepliedWhilelittlehandsmakevainpretenceOurwanderingstoguide

**Исходный текст(кириллица):**

Коралина нашла дверь почти сразу после того, как они переехали.Это был очень старый дом с мансардой под самой крышей и подвалом. Его окружал заросший сад с огромными старыми деревьями.

**Текст**: AllinthegoldenafternoonFullleisurelyweglideForbothouroarswithlittleskillBylittlearmsarepliedWhilelittlehandsmakevainpretenceOurwanderingstoguide

**Энтропия текста равна**: 4.040095046071584

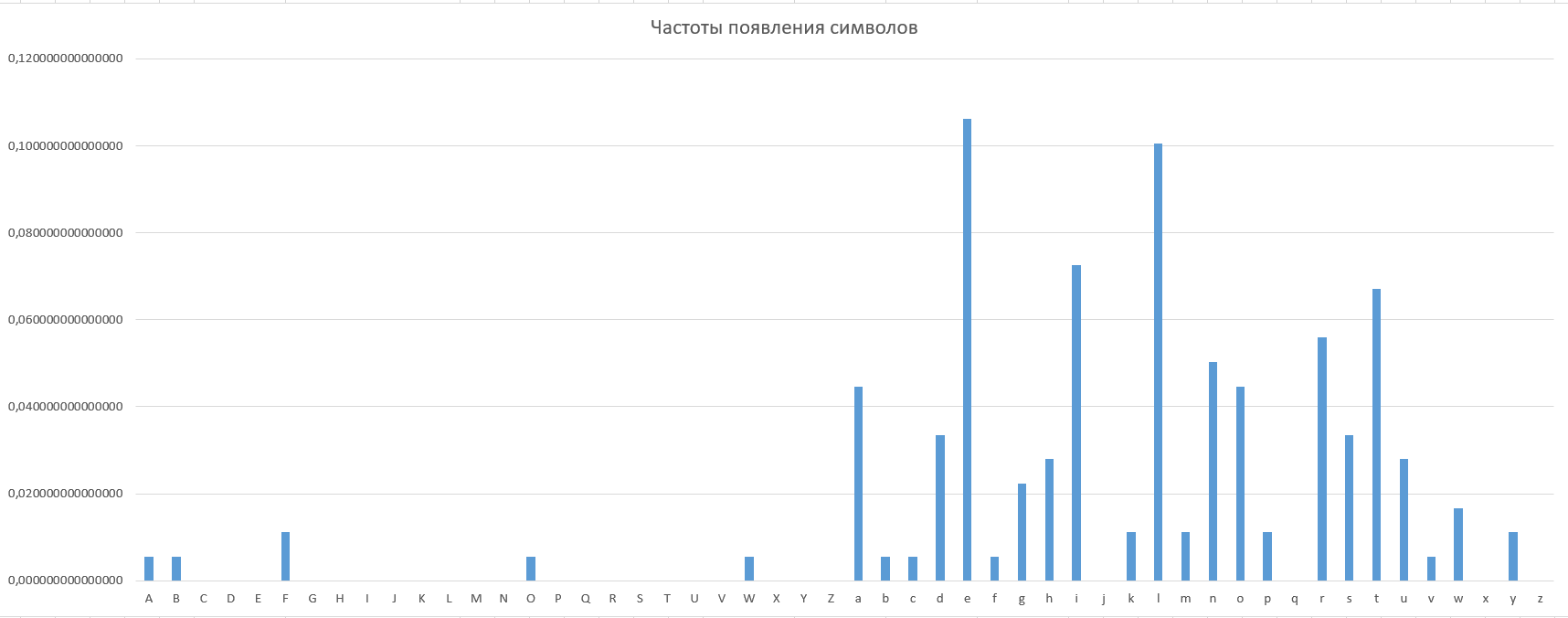
**Количество информации**: 581.7736866343081

**Текст**: Коралина нашла дверь почти сразу после того, как они переехали.Это был очень старый дом с мансардой под самой крышей и подвалом. Его окружал заросший сад с огромными старыми деревьями.

**Энтропия текста равна**: 3.7580396706294965

**Количество информации**: 691.4792993958274

**Частоты появления символов**

****

[Задания.xlsx](about:blank)

1. для входных документов, представленных в бинарных кодах, определить энтропию бинарного алфавита;

**Исходный текст в двоичном представлении**:

100000111011001101100110100111011101110100110100011001011100111110111111011001100100110010111011101100001110011011101001100101111001011011101101111110111111011101000110111010111011001101100110110011001011101001111001111101011110010110010111011001111001111011111001011100111110110011010011100100110010110001101101111111001011000101101111111010011010001101111111010111100101101111110000111100101110011111011111010011110100110100011011001101001111010011101001101100110010111100111101011110100111011001101100100001011110011101100110100111101001110100110110011001011100001111001011011011110011110000111100101100101111000011011001101001110010111001001010111110100011010011101100110010111011001101001111010011101001101100110010111010001100001110111011001001110011110110111000011101011110010111101101100001110100111011101110000111001011001011110100110010111011101100011110010110011111110101111001011101111100001110111011001001100101111001011010011101110110011111100111110100110111111001111110101110100111001001100101

**Код программы**

def Ubung2(inputStr = "", p = 0):  
 if (inputStr != ""):  
 text = inputStr  
 else:  
 inp = int(input("Выберите метод ввода( 1 - из файла, 2 - ввести ручками): "))  
 match inp:  
 case 1:  
 fileName = input("Введите название файла: ")  
 text = OpenFile(fileName, numbLineRead)  
 case 2:  
 text = input("Введите текст: ")  
  
 symb = '0'  
 counter = 0  
 entropy = 0  
 x = 0  
  
 while x < 2:  
 for i in text:  
 if (i == symb):  
 counter += 1  
 x += 1  
 symb = '1'  
 entropy += (counter / len(text)) \* math.log2(counter / len(text))  
 entropy = -entropy;  
 if p > 0 and p < 1 and p != 0.5:  
 entropy = 1 - (-p \* math.log2(p) - (1 - p) \* math.log2(1 - p))  
 elif p == 0:  
 entropy += 0  
 elif p == 0.5:  
 entropy = 0  
 elif p == 1:  
 entropy = 1  
  
 print("Текст: " + text)  
 print("Вероятность ошибочной передачи единичного бита: " + str(p))  
 print("Энтропия текста равна: " + str(entropy))  
 print("Количество информации: " + str(entropy \* len(text)))

**Результат выполнения программы**

**Текст:** 100000111011001101100110100111011101110100110100011001011100111110111111011001100100110010111011101100001110011011101001100101111001011011101101111110111111011101000110111010111011001101100110110011001011101001111001111101011110010110010111011001111001111011111001011100111110110011010011100100110010110001101101111111001011000101101111111010011010001101111111010111100101101111110000111100101110011111011111010011110100110100011011001101001111010011101001101100110010111100111101011110100111011001101100100001011110011101100110100111101001110100110110011001011100001111001011011011110011110000111100101100101111000011011001101001110010111001001010111110100011010011101100110010111011001101001111010011101001101100110010111010001100001110111011001001110011110110111000011101011110010111101101100001110100111011101110000111001011001011110100110010111011101100011110010110011111110101111001011101111100001110111011001001100101111001011010011101110110011111100111110100110111111001111110101110100111001001100101

**Вероятность ошибочной передачи единичного бита:** 0

**Энтропия текста равна:** 0.5285498406942574

**Количество информации:** 532.7782394198114

1. используя значения энтропии алфавитов, полученных в  
   пунктах (а) и (б), подсчитать количество информации в сообщении, состоящем из собственных фамилии, имени и отчества (на  
   основе исходного алфавита – (а) и в кодах ASCII – (б)); объяснить  
   полученный результат;

**Исходный текст:** LevshaMarkSergeevich

**Исходный текст в двоичном представлении:**

1001101110000111110101110101110101111001011110110110100111101001100011110100010000011101100110010111110001100101111100110100111100101111001011101101100101110010111101101101001111010011000111101000

**Результат выполнения программы**

**Текст**: LevshaMarkSergeevich

**Энтропия текста равна**: 3.5547828710380327

**Количество информации**: 99.53392038906492

**Текст:** 1001101110000111110101110101110101111001011110110110100111101001100011110100010000011101100110010111110001100101111100110100111100101111001011101101100101110010111101101101001111010011000111101000

**Вероятность ошибочной передачи единичного бита**: 0

**Энтропия текста равна:** 0.525953789993695

**Количество информации:** 103.08694283876423

1. выполнить задание пункта (в) при условии, что вероятность  
   ошибочной передачи единичного бита сообщения составляет: 0,1;  
   0,5; 1,0.

**Текст**: 1001101110000111110101110101110101111001011110110110100111101001100011110100010000011101100110010111110001100101111100110100111100101111001011101101100101110010111101101101001111010011000111101000

**Вероятность** **ошибочной передачи единичного** **бита**: 0.1

**Энтропия текста равна**: 0.5310044064107188

**Количество информации**: 104.07686365650088

**Текст**: 1001101110000111110101110101110101111001011110110110100111101001100011110100010000011101100110010111110001100101111100110100111100101111001011101101100101110010111101101101001111010011000111101000

**Вероятность ошибочной передачи единичного бита**: 0.5

**Энтропия текста равна**: 0

**Количество информации:** 0

**Текст**: 1001101110000111110101110101110101111001011110110110100111101001100011110100010000011101100110010111110001100101111100110100111100101111001011101101100101110010111101101101001111010011000111101000

**Вероятность ошибочной передачи единичного бита**: 1

**Энтропия текста равна**: 1

**Количество информации:** 196

1. Что такое «алфавит источника сообщения»?

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

2. Что такое «мощность алфавита источника сообщения»?

Количество символов в алфавите.

3. Какова мощность алфавита белорусского языка?

34

4. Какова мощность алфавита русского языка?

33

5. Какова мощность алфавита «компьютерного» языка?

2

6. Что такое «энтропия алфавита»?

Энтропия – информационная характеристика алфавита (источника сообщений на основе этого алфавита), которая показывает, какое количество информации приходится в среднем на один символ алфавита (сообщения).

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

7. Что такое «энтропия сообщения»?

Энтропия – информационная характеристика алфавита (источника сообщений на основе этого алфавита), которая показывает, какое количество информации приходится в среднем на один символ алфавита (сообщения).

8. От чего зависит энтропия алфавита?

От частоты встречаемости символов и мощности алфавита.

9. Записать формулу для вычисления энтропии (рисунок 7).

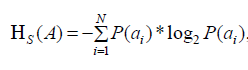


Рисунок 7

10. Что нужно знать для вычисления энтропии алфавита?

Частоту встречаемости каждого символа алфавита и мощность.

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

Энтропию алфавита А={*ai*} по К.Шеннону рассчитывают по следующей формуле, представленной на рисунке 8.

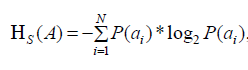


Рисунок 8

Частным случаем энтропии Шеннона является 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, представленному на рисунке 9.

https://lh6.googleusercontent.com/iVtaU6Me-5i_RKzvvF6QnO0j7MxiztNrSXZBNC9RRZ0wTtZ5d-vnXp1zOR3kojDdVF_vMibCAvpJgsK8HWeTepN2fzjOcZJJiHvyR3q9iTTTlsZtxnVGJYJrnwnO9AYDMZRk6DcC-Edbb2_YQL0vlxqPbEaW0BMUXNam5WQtmmbq9vyczvVzpOQPVJPCMC5bpjsE

Рисунок 9

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

Поскольку мы находим логарифмы от чисел меньших 1, то логарифм всегда будет отрицательным, поэтому мы добавляем минус перед логарифмом

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

Избыточностью алфавита называется уменьшение информационной нагрузки на один символ вследствие неравновероятности и взаимозависимости появления его символов.

Информационная избыточность характеризует относительную нагруженность алфавита.

14. Расположите в порядке возрастания энтропии известные вам алфавиты.

3,2 – белорусского

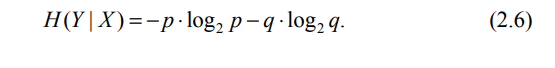
3,895 – французского

4,7 – английского

15. Вычислить энтропию алфавита белорусского (русского) языка.

  3.2 бит

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.



H(А) = 1-(0,25\*log20,25 + 0,75\*log20.75)

при p =0,25 и q = 0,75

H(А) = 1-(0,5\*log20,5 + 0,5\*log20.5) = 0

при p =0,5 и q = 0,5

H(А) = 1-(1\*log21+ 0\*log20) = 1

при p =1 и q = 0

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

a) -1\*log21 = 0

б) -(0.5\*log20.5 + 0.5\*log20.5 ) = 1

в) -1\*log2(1/8) = 3

Вывод: в ходе лабораторной работы были приобретены практические навыки расчета и анализа параметров и информативных характеристик дискретных ИС, высчитаны энтропии текстов на латинице кириллице и текста в бинарном виде. Если посмотреть на гистограмму, то станет заметно, что чаще всего встречаются гласные буквы.