Ковкель Никита, ФИТ 3-4

Информационная безопасность

Отчет по лабораторной работе № 1.

Алфавит – это общее число знаков или символов (N), используемых для генерации или передачи сообщений. Информационной характеристикой алфавита (источника сообщений на основе этого алфавита) является энтропия. Ее физический смысл заключается в отображении количества информации в среднем на 1 символ алфавита.

По Шеннону энтропию высчитывают по следующей формуле:

Для выполнения данной лабораторной работы был разработан следующий класс с использованием языка Python.

Листинг 1. Реализация класса.

|  |
| --- |
| import collections  import math  import matplotlib.pyplot as plt    class FileEntropy:  def \_\_init\_\_(self, file\_name):  self.file\_name = file\_name  self.text = self.\_read\_file()  self.is\_binary = self.\_check\_if\_binary()  self.histogram = None  def \_read\_file(self):  try:  with open(self.file\_name, 'r') as f:  text = f.read()  return text.replace(' ', '')  except FileNotFoundError:  return None  def \_check\_if\_binary(self):  if self.text is None:  return False  return all(c in '01\n' for c in self.text)  def calculate\_entropy(self):  symbol\_counts = collections.Counter(self.text)  total\_symbols = len(self.text)  probabilities = {symbol: count / total\_symbols for symbol, count in symbol\_counts.items()}  self.histogram = probabilities  entropy = -sum(p \* math.log2(p) for p in probabilities.values())  return entropy  def draw\_histogram(self):  if self.histogram:  plt.bar(self.histogram.keys(), self.histogram.values())  plt.xlabel("Символ")  plt.ylabel("Вероятность")  plt.title("Вероятности символов")  plt.show()  def info\_amount(self):  if self.text:  return self.calculate\_entropy() \* len(self.text)  else:  return None  def effective\_entropy(self, p):  q = 1 - p  if self.is\_binary and (p == 0 or q == 0):  return 1  elif not self.is\_binary:  if p == 1:  return 0  else:  return self.calculate\_entropy() - (- p \* math.log2(p) - q \* math.log2(q))  return 1 - (- p \* math.log2(p) - q \* math.log2(q))  def info\_amount\_with\_errors(self, p):  if self.text:  return self.effective\_entropy(p) \* len(self.text)  else:  return None |

Для расчета энтропии были применены текстовые файлы на русском и английском языках. Кроме того, использовался бинарный алфавит, включающий 0 и 1, который был записан в текстовый документ.

Подставляя данные в формулу для расчета энтропии, получаем результаты, приведённые на рисунке 1.

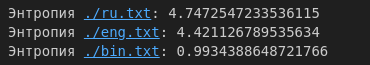


Рисунок 1 – Результаты подсчёта энтропии

Программная реализация представлена в листинге 1.

При известной энтропии алфавита можно посчитать количество информации по следующей формуле:

Также при передаче сообщения с некоторой вероятностью могут возникнуть ошибки, в таком случае используется формула условной энтропии:

,

где - условная энтропия:

.

Для бинарного алфавита формула имеет видоизменяется следующим образом:

Программная реализация функции вычисления эффективной энтропии и вычисления количества информации с учетом ошибок представлена в листинге 1.

Результат выполнения скрипта для файла содержащего ФИО на английском представлен на рисунке 1.

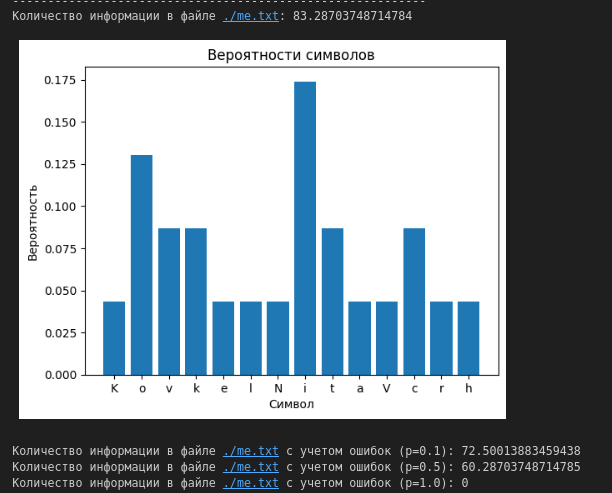


Рисунок 1 – Вывод количества информации (с учетом ошибок и без), гистограммы вероятностей символов для файла me.txt.

Результат выполнения скрипта для файла содержащего ФИО в бинарном представлен на рисунке 2.

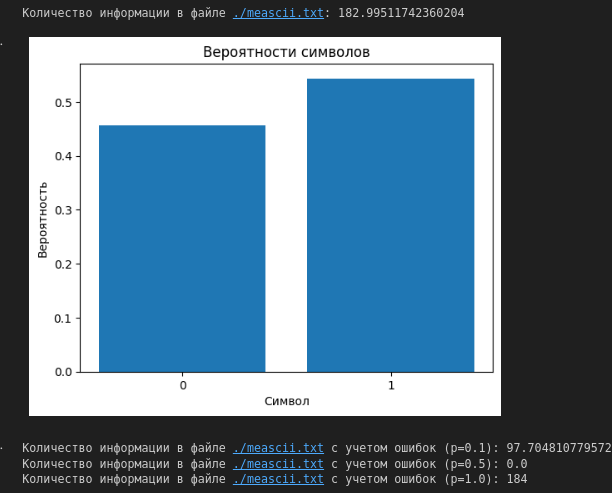


Рисунок 2 – Вывод количества информации (с учетом ошибок и без), гистограммы вероятностей символов для файла meascii.txt.

Результат выполнения скрипта для файла содержащего текст на английском представлен на рисунке 3.

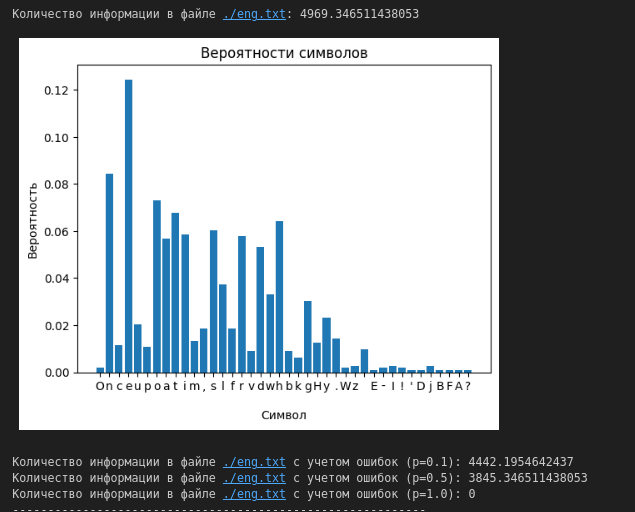


Рисунок 3 – Вывод количества информации (с учетом ошибок и без), гистограммы вероятностей символов для файла eng.txt.

Результат выполнения скрипта для файла содержащего текст на русском представлен на рисунке 4.

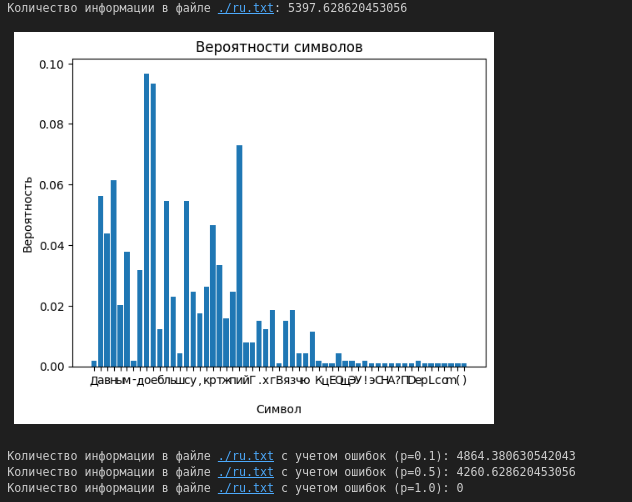


Рисунок 4 – Вывод количества информации (с учетом ошибок и без), гистограммы вероятностей символов для файла ru.txt.

Результат выполнения скрипта для файла содержащего текст в бинарном формате представлен на рисунке 5.

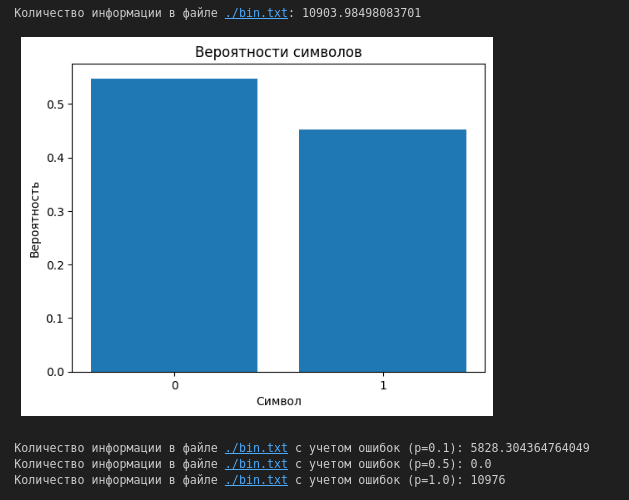


Рисунок 5 – Вывод количества информации (с учетом ошибок и без), гистограммы вероятностей символов для файла bin.txt.

Когда вероятность ошибки равна 1, количество информации в бинарном алфавите совпадает с количеством информации при вероятности ошибки равной 0. Это связано с тем, что все биты сообщения заменяются на противоположные. Однако в не бинарном алфавите при вероятности ошибки равной 1 такое невозможно, поскольку размер алфавита превышает 2, и каждый символ заменяется на другой случайный символ из этого алфавита. В результате количество информации в этом случае становится равным 0.