

```

import string
import math

f = open(r'E:\8_semestr\TI\lab2\res\1.txt', "r", encoding="windows-1251")
text = f.read()
text = text.lower()

def remove_chars_from_text(txt, chars):
    return ''.join([ch for ch in txt if ch not in chars])

spec_chars = string.punctuation + '«»-...1234567890-
qwertyuiopasdfghjklzxcvbnm\n\t\xa0©'
text = remove_chars_from_text(text, spec_chars)
text = text.replace('ë', 'e')
text = text.replace('Ъ', 'ь')

text_tokens = dict()

def shannon(txt):
    for d in range(1, 4):
        summa = 0
        for i in range(len(txt) - d + 1):
            if text_tokens.get(txt[i:i + d]) is None:
                text_tokens[txt[i:i + d]] = 1
                summa += 1
            else:
                text_tokens[txt[i:i + d]] = text_tokens.get(txt[i:i + d]) + 1
                summa += 1
        if d == 1:
            MaxH = math.log(len(text_tokens), 2)
            print("MaxH = ", MaxH)
        H = 0
        for value in text_tokens.values():
            H += (-1) * (value / summa) * math.log(value / summa, 2)
        H /= d
        print(f"H{d} = {H}")
        text_tokens.clear()

shannon(text)

```