

## Практическая работа №2

### Вычисление энтропии Шеннона

Цель работы: Экспериментальное изучение свойств энтропии Шеннона для текстов на естественном языке.

Язык программирования: C, C++, C#, Python

Результат: программа, тестовые примеры, отчет.

Задание:

1. Выбрать художественный текст на русском (английском) языке. Объем файла в формате txt более 10 Кб. Для алфавита текста предполагается, что строчные и заглавные символы не отличаются, знаки препинания опущены, к алфавиту добавлен пробел, для русских текстов буквы «е» и «ё», «ь» и «ъ» совпадают.
2. Составить программу, определяющую несколько оценок энтропии данного текстового файла. Оценки энтропии необходимо вычислить по формуле Шеннона двумя способами, т.е. используя частоты отдельных символов и используя частоты пар символов. По желанию можно продолжить процесс вычисления оценок с использованием частот троек, четверок символов и т.д.
3. После тестирования программы необходимо заполнить таблицу для отчета и проанализировать полученные результаты. Сравнить полученные результаты с результатами работы 1.

Название текста	Максимально возможное значение энтропии	Оценка энтропии (одиночные символы)	Оценка энтропии (частоты пар символов)

4. Оформить отчет, загрузить отчет и файл с исходным кодом в электронную среду. Отчет обязательно должен содержать заполненную таблицу и анализ полученных результатов. По желанию в отчет можно включить описание программной реализации. В отчет не нужно включать содержимое этого файла.