

## Практическая работа №1

### Вычисление энтропии Шеннона

Цель работы: Экспериментальное изучение свойств энтропии Шеннона.

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

Результат: программа, тестовые примеры, отчет.

Задание:

1. Для выполнения работы необходимо предварительно сгенерировать два файла. Каждый файл содержит последовательность символов, количество различных символов больше 2 (3,4 или 5). Объем файлов больше 10 Кб, формат txt. Символы **последовательно и независимо** генерируются с помощью датчика псевдослучайных чисел и записываются в файл.

Первый файл (назовем его F1) должен содержать последовательность символов с равномерным распределением, т.е. символы должны порождаться равновероятно и независимо.

Для генерации второго файла (F2) необходимо сначала задать набор вероятностей символов, а затем **последовательно и независимо** генерировать символы с соответствующей вероятностью и записывать их в файл.

2. Составить программу, определяющую несколько оценок энтропии созданных текстовых файлов. Вычисление значения по формуле Шеннона **настоятельно рекомендуется** оформить в виде отдельной функции, на вход которой подается массив (список) вероятностей, выходной параметр – значение, вычисленное по формуле Шеннона.

Оценки энтропии необходимо вычислить по формуле Шеннона двумя способами.

**Первый способ.** Сначала определить частоты отдельных символов файла, т.е. отношения количества отдельного символа к общему количеству символов в файле. Далее используя полученные частоты как оценки вероятностей, рассчитать оценку энтропии по формуле Шеннона.

**Второй способ.** Определить частоты всех последовательных пар символов в файле. Для того, чтобы правильно рассчитать оценку энтропии для двойных комбинаций символов пары символов нужно рассматривать следующим образом.

Пусть имеется такая последовательность  
фывафпро

Под парами понимаются пары соседних символов, т.е.  $xy$   $yx$   $af$   $fa$   $pr$   $rp$

Для подсчета энтропии подсчитайте частоту встречаемости для каждой пары. Полученное значение энтропии следует разделить на 2.

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. Для получения теоретических значений энтропии используйте наборы вероятностей, которые использовались при генерации файлов.

	Оценка энтропии (частоты отдельных символов)	Теоретическое значение энтропии (отдельные символы)	Оценка энтропии (частоты пар символов)	Теоретическое значение энтропии (для пар символов)
F1				
F2				

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.