Правила выполнения лабораторных работ

Перед выполнением заданий лабораторной работы рекомендуется изучить теоретический материал по теме лабораторной работы и описание методов кодирования на псевдокоде, используя конспекты лекционных занятий и литературу из списка.

Задания лабораторных работ выполняются только на языке программирования С/С++, среда программирования по выбору студента.

Изучаемые методы кодирования данных рекомендуется программно реализовывать в виде отдельных функций (подпрограмм), данные должны передаваться в подпрограммы в качестве параметров. Заполнение массивов данными, вывод их на экран, вычисление вспомогательных величин и пр. необходимо также оформлять в виде отдельных подпрограмм.

При выполнении заданий следует обеспечить вывод на экран данных на всех шагах алгоритма. Программа должна иметь дружественный, интуитивно понятный интерфейс (меню пользователя, вывод подсказок, комментарии при вводе/выводе данных и т.д.).

Тестирование разработанной программы необходимо проводить для  текстовых файлов с различным частотным распределением символов. После тестирования необходимо проанализировать полученные результаты, т.е. проверить соответствие полученных экспериментальным путем величин теоретическим оценкам.

Для зачета по лабораторной работе студенту необходимо представить

·         Исходные тексты программ с подробными комментариями;

·         Исполняемые файлы и тестовые файлы;

·         Отчет по лабораторной работе.

Отчет должен включать в себя следующие разделы

·         Формулировку задания

·         Очень краткое описание алгоритмов, используемых в лабораторной работе;

·         Результаты работы программы (в виде файла или в виде скриншота);

·         Анализ и сравнение полученных результатов с теоретическими оценками.

Лабораторная работа №1

Вычисление энтропии Шеннона

Цель работы: Экспериментальное изучение свойств энтропии Шеннона.

Среда программирования: любая с С-подобным языком программирования.

Результат: программа, тестовые примеры, отчет.

Задание:

1. Для выполнения данной лабораторной 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.

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

Второй файл (F2) содержит последовательность символов с неравновероятным распределением.

2. Составить программу, определяющую несколько оценок энтропии созданных текстовых файлов. 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.

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

Лабораторная работа №2

Вычисление энтропии Шеннона

Цель работы: Экспериментальное изучение свойств энтропии Шеннона.

Среда программирования: любая с С-подобным языком программирования.

Результат: программа, тестовые примеры, отчет.

Задание:

1. Составить программу, определяющую несколько оценок энтропии текстового файла (размер не менее 10 Кб). 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.

Для художественных текстов (русский или английский языки) предполагается, что строчные и заглавные символы не отличаются, знаки препинания объединены в один символ, к алфавиту добавлен пробел, для русских текстов буквы «е» и «ё», «ь» и «ъ» совпадают. При использовании текста программы учитываются все символы, кроме знаков табуляции.

2. После тестирования программы необходимо заполнить таблицу для отчета и проанализировать полученные результаты. Сравнить полученные результаты с результатами лабораторной работы 1.

|  |  |  |  |
| --- | --- | --- | --- |
| Название текста | Максимально возможное значение энтропии | Оценка энтропии(одиночные символы) | Оценка энтропии(частоты пар символов) |
| Текст №1(фрагмент художественного произведения) |   |   |   |
| Текст №2 (фрагмент художественного произведения) |   |   |   |
| Текст написанной программы |   |   |   |

Лабораторные работы №3

Оптимальное побуквенное кодирование

Цель работы: Изучение метода оптимального кодирования Хаффмана.

Среда программирования: любая с С-подобным языком программирования.

Результат: программа, тестовые примеры, отчет.

1. Запрограммировать процедуру двоичного кодирования текстового файла  методом Хаффмана. Текстовые файлы использовать те же, что и в лабораторных работах №1,2. Для художественных текстов (русский или английский языки) предполагается, что строчные и заглавные символы не отличаются, знаки препинания объединены в один символ, к алфавиту добавлен пробел, для русских текстов буквы «е» и «ё», «ь» и «ъ» совпадают.

2. Проверить, что полученный код является префиксным.

3. После кодирования текстового файла вычислить оценки энтропии выходной последовательности, используя частоты отдельных символов, пар символов и троек символов.

4. Заполнить таблицу и проанализировать полученные результаты.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Методкодирования | Название текста | Оценкаизбыточности кодирования | Оценка энтропии выходной посл-ти (частоты символов) | Оценка энтропии выходной посл-ти (частоты пар символов) | Оценка энтропии выходной посл-ти (частоты троек символов) |
| Метод Хаффмана | Текст №1 |   |   |   |   |
| Текст №2 |   |   |   |   |

Избыточность кодирования определяется как , где *H*  – энтропия текста, *L*cp – средняя длина кодового слова.

Лабораторные работы №4

Методы почти оптимального кодирования

Цель работы: Изучение метода почти оптимального кодирования Фано.

Среда программирования: любая с С-подобным языком программирования.

Результат: программа, тестовые примеры, отчет.

1. Запрограммировать процедуры двоичного кодирования текстового файла  методом Фано. Текстовые файлы использовать те же, что и в лабораторной работе №1 и 2. Для художественных текстов (русский или английский языки) предполагается, что строчные и заглавные символы не отличаются, знаки препинания объединены в один символ, к алфавиту добавлен пробел, для русских текстов буквы «е» и «ё», «ь» и «ъ» совпадают.

2. Проверить, что полученный код является префиксным.

3 После кодирования текстового файла вычислить оценки энтропии выходной последовательности, используя частоты отдельных символов, пар символов и тройки символов.

4. После тестирования программы необходимо заполнить таблицу и проанализировать полученные результаты.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Методкодирования | Название текста | Оценкаизбыточности кодирования | Оценка энтропии выходной посл-ти (частоты символов) | Оценка энтропии выходной посл-ти (частоты пар символов) | Оценка энтропии выходной посл-ти (частоты троек символов) |
| Метод Хаффмана | Текст №1 |   |   |   |   |
| Текст №2 |   |   |   |   |
| Метод Фано | Текст №1 |   |   |   |   |
| Текст №2 |   |   |   |   |

Избыточность кодирования определяется как , где *H*  – энтропия текста, *L*cp – средняя длина кодового слова.

Лабораторные работы №5

Почти оптимальное кодирование

Цель работы: Изучение метода почти оптимального кодирования Шеннона.

Среда программирования: любая с С-подобным языком программирования.

Результат: программа, тестовые примеры, отчет.

1. Запрограммировать процедуру двоичного кодирования текстового файла  методом Шеннона. Текстовые файлы использовать те же, что и в лабораторной работе №1-4. Для художественных текстов (русский или английский языки) предполагается, что строчные и заглавные символы не отличаются, знаки препинания объединены в один символ, к алфавиту добавлен пробел, для русских текстов буквы «е» и «ё», «ь» и «ъ» совпадают.

2. Проверить, что полученный код является префиксным.

3. После кодирования текстового файла вычислить оценки энтропии выходной последовательности, используя частоты отдельных символов, пар символов и троек символов.

4. Заполнить таблицу и проанализировать полученные результаты.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Методкодирования | Название текста | Оценкаизбыточности кодирования | Оценка энтропии выходной посл-ти (частоты символов) | Оценка энтропии выходной посл-ти (частоты пар символов) | Оценка энтропии выходной посл-ти (частоты троек символов) |
| Метод Хаффмана | Текст №1 |   |   |   |   |
| Текст №2 |   |   |   |   |
| Метод Шеннона | Текст №1 |   |   |   |   |
| Текст №2 |   |   |   |   |
| Метод Фано | Текст №1 |   |   |   |   |
| Текст №2 |   |   |   |   |

Избыточность кодирования определяется как , где *H*  – энтропия текста, *L*cp – средняя длина кодового слова.