Лабораторная работа №2

Кодирование текстовой информации в Unicode

Цель – понять принципы кодирования текстовой информации путем обработки содержимого текстовых файлов в кодах Unicode.

**Задание**

На «3» балла:

Перекодировать ASCII в Unicode (UTF-8).

Порядок выполнения:

1. Открыть из директории «5» файл в соответствии с вариантом при помощи текстового редактора и HEX-редактора. Текст содержит три слова на английском языке и их перевод на русский в кодировке windows-1251 (ASCII). Переносы строк и табуляторы отсутствуют.
2. При помощи HEX-редактора перекодировать переводы слов на русский из ASCII в UTF-8, добавить табуляторы между английским и русским словом, и переносы строк между словами.

На «4» балла:

«Восстановить» текстовый файл.

Порядок выполнения:

1. Открыть из директории «10» файл в соответствии с вариантом при помощи текстового редактора и HEX-редактора. Текст в файле закодирован в UTF-8 и соответствует отрывкам из «Евгения Онегина» (<http://lib.ru/LITRA/PUSHKIN/p4.txt>), однако содержимое заранее «испорчено».
2. «Восстановить» содержимое файла в кодах при помощи HEX-редактора.
3. Добавить в конце файла свои инициалы и номер группы.

На «5» баллов:

«Перекодировщик»

Написать программу «перекодировщик» из ASCII (кодировка по варианту) в Unicode (формат UTF-8) и обратно. Алгоритм перекодировки должен быть реализован самостоятельно и без использования сторонних библиотек