МИНОБРНАУКИ РОССИИ

федеральное государственное бюджетное образовательное учреждение

высшего профессионального образования

«Череповецкий государственный университет»

**Лабораторная работа № 2**

**«Кодирование текстовой информации»**

**Выполнил:**

**студент гр.** 1ИВТпб-01-11оп

Климов А.Г. **Проверил: преподаватель**

Ганичева О.Г. **Отметка о зачете:**

Череповец

2016 год

**Тема: Кодирование тестовой информации**

**Цель: ознакомиться с принятыми системами кодирования информации**

**Краткие теоретические положения**

Имеются разные стандарты для представления, символов, которые отличаются лишь порядком нумерации символов. Наиболее распространён американский стандартный код для информационного обмена - ASCII [American Standard-Code for Information Interchange] введён в США в 1963г. В 1977 году в несколько модифицированном виде он был принят в качестве всемирного стандарта Международной организации стандартов [International Standards Organization -. ISO] под названием ISO-646. Согласно этому стандарту каждому символу поставлено в соответствие число от 0 до 255. Символы от 0 до 127 - латинские буквы, цифры и знаки препинания - составляют постоянную часть таблицы. Остальные символы используются для представления национальных алфавитов. Конкретный состав этих символов определяется кодовой страницей. В русской версии ОC Windows95 используется кодовая, страница 866. В ОС Linux для представления русских букв более употребительна кодировка КОИ-8. Недостатки такого способа кодировки национального, алфавита очевидны. Во-первых, невозможно одновременное представление русских и, например, французских букв. Во-вторых, такая кодировка совершенно непригодна для представления, китайских иероглифов. В 1991 году была создана некоммерческая организация Unicode, в которую входят представители ряда фирм (Borland. IBM, Noyell, Sun и др) и которая занимается развитием и внедрением нового стандарта. Кодировка Unicode использует 16 разрядов, и может содержать 65536 символов. Это символы большинства народов мира, элементы иероглифов, спецсимволы, 5000 – мест для частного использования, резерв из 30000 мест.

ASCII-код символа А= 6510 =4116= 010001112;

Unicode-код символа С= 6710=00000000011001112

**Ход Работы**

**Задание 1.**

Размер файлов, сохранённых в различных форматах разный. Файл в формате “.txt” занимает наименьший объём, а в “.doc” – наибольший.

|  |  |
| --- | --- |
| **Формат** | **Размер (байт)** |
| doc | 24064 |
| htm | 4305 |
| rtf | 6349 |
| txt | 96 |

DOC

Достоинства DOC:   
- сохраняет всю информацию о форматах и стилях текста   
- с большой вероятностью может быть прочитан на компьютере получателя

Недостатки:   
- проприетарный формат   
- занимает больше места на диске, чем простые текстовые файлы   
- с большой вероятностью может не читаться на компьютере получателя

HTML

Достоинства HTML:   
- сохраняет информацию о выбранных автором шрифтах, кодировках, стилях, цвете текста   
- читается на различных компьютерных платформах   
- многие программы (в их числе бесплатные) умеют с ним работать   
- это открытый стандарт

Недостатки:   
- графика хранится отдельно   
- часто представляет из себя не отдельный файл, а целую группу файлов   
- занимает больше места на диске, чем простые текстовые файлы

RTF

Достоинства RTF:   
- сохраняет информацию о выбранных автором шрифтах, кодировках, стилях, цвете текста   
- читается на различных компьютерных платформах   
- многие программы (в их числе бесплатные) умеют с ним работать   
- это открытый стандарт

Недостатки:   
- не поддерживает графику   
- занимает больше места на диске, чем простые текстовые файлы

TXT

Достоинства простого текстового формата:   
- простота   
- экономичность   
- это открытый стандарт

Недостатки:   
- при обмене данными между компьютерами различных платформ возможны конфликты кодировок   
- не поддерживаются стили текста и графика

**Задание 2.**

2.1

Имя, фамилия и отчество, закодированные при помощи таблицы KOI-8

* 410 43B 435 43A 441 430 43D 434 440;
* 41A 43B 438 43C 43E 432;
* 413 440 438 413 43E 440 44C 435 432 438 447.

2.2 Раскодированное ФИО соседа: Басистюк Антон Олегович.

2.3 Закодируйте следующие слова, используя таблицы ASCII-кодов: ИНФОРМАТИЗАЦИЯ, МИКРОПРОЦЕССОР, МОДЕЛИРОВАНИЕ

**10-ый код:**

**200 205 212 206 208 204 192 210 200 199 192 214 200 223**

**204 200 202 208 206 207 208 206 212 197 209 209 206 208**

**204 206 196 197 203 200 208 206 194 192 205 200 197**

2.4 Раскодируйте следующие слова, используя таблицы ASCII-кодов:

88 AD E4 AE E0 AC A0 E2 A8 AA A0

50 72 6F 67 72 61 6D

43 6F 6D 70 75 74 65 72 20 49 42 4D 20 50 43

**Информатика, Program, Computer IBM PC**

2.5 Открыть блокнот.

а) Используя клавишу Alt и малую цифровую клавиатуру, раскодировать фразу: 145 170 174 224 174 255 170 160 173 168 170 227 171 235;

Технология выполнения задания: При удерживаемой клавише Alt, набрать на малой цифровой клавиатуре указанные цифры. Отпустить клавишу Alt, после чего в тексте появится буква, закодированная набранным кодом.

**Скоро каникулы**

б) Используя ключ к кодированию, закодировать слово – зима;

Технология выполнения задания: Из предыдущего задания выяснить, каким кодом записана буква а. Учитывая, что буквы кодируются в алфавитном порядке, выяснить коды остальных букв.

**167 168 172 160**

2.6 Вывод: Ознакомился с принятыми системами кодирования информации.

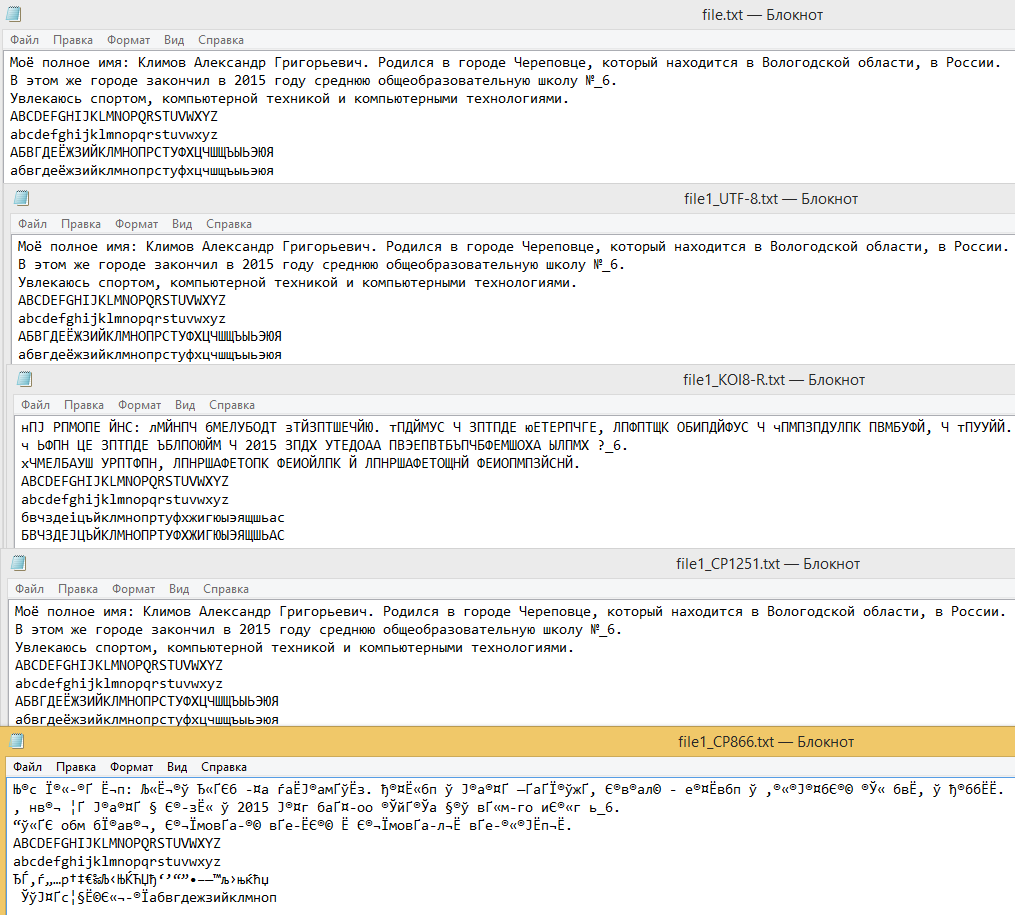
**Задание 3.**

Файл, созданный в Блокноте Windows

Размер файла - 400 байт:

1. 388 символов: буквы, цифры, пробелы и специальные символы (по 1-му байту на каждый);
2. 6 символов переноса (по 2 байта каждый).

Скриншот кодировок



Сравнительная таблица кодировок

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Кодировка | CP866 (MS-DOS) | CP1251 (Windows) | KOI8-R | UTF-8 (Юникод) |
| Вес символа | 1 байт | 1 байт | 1 байт | 2 байта |
| Вес файла | 402 байта | 402 байта | 402 байта | 693 байта |
| Поддержка русского алфавита | нет | есть | нет | есть |
| Поддержка латинского алфавита, цифр | есть | есть | есть | есть |
| Поддержка специальных символов (например, №) | нет | есть | нет | есть |

**Вывод**

Процедура кодирования необходима для передачи информации по каналам связи. Кодирование информации — процесс преобразования сигнала из формы, удобной для непосредственного использования информации, в форму, удобную для передачи, хранения или автоматической переработки. Для представления информации в компьютере используют кодировки. Набор символов (предшествовавший термин «кодировка») — таблица, задающая кодировку конечного множества символов алфавита. Такая таблица сопоставляет каждому символу последовательность длиной в один или несколько символов другого алфавита.

При работе с кодировками могут возникать различные проблемы (например, конфликт кодировок, вследствие чего происходит потеря информации). В основном, разные стандарты кодирования отличаются лишь порядком нумерации символов. Для решения данных проблем был введён в США в 1963 г. американский стандартный код для информационного обмена – ASCII [American Standard-Code for Information Interchange]. В 1977 году в несколько модифицированном виде он был принят в качестве всемирного стандарта Международной организации стандартов [International Standards Organization -. ISO]. Главными недостатками этой кодировки стали: отсутствие возможности одновременного представления нескольких алфавитов и непригодность для представления китайских иероглифов.

Для решения новых задач в 1991 году была создана кодировка Unicode, использующая 16 разрядов, и имеющая возможность содержать 65536 символов. Это символы большинства народов мира, элементы иероглифов, спецсимволы, 5000 – мест для частного использования, резерв из 30000 мест.