|  |  |
| --- | --- |
| **Gerb-BMSTU_01** | **Министерство науки и высшего образования Российской Федерации**  Калужский филиал  федерального государственного бюджетного  образовательного учреждения высшего образования  ***«Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»***  ***(КФ МГТУ им. Н.Э. Баумана)*** |

**ФАКУЛЬТЕТ** ***ИУК «Информатика и управление»***

**КАФЕДРА** \_\_***ИУК4 «Программное обеспечение ЭВМ, информационные технологии»***

**ЛАБОРАТОРНАЯ РАБОТА**

**«Кодировки ASCII и Unicode»**

**ДИСЦИПЛИНА: «Теоретическая информатика»**

|  |  |  |
| --- | --- | --- |
| Выполнил: студент гр. ИУК4-12Б | | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ( Карельский М.К. )  (Подпись) |
| Проверил: | | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ( Амеличева К.А. )  (Подпись) |
| Дата сдачи (защиты):  Результаты сдачи (защиты): | | |
|  | - Балльная оценка:  - Оценка: | |

Калуга, 2020

**Цель:** формирование практических навыков обоснованного выбора подходящих форматов хранения текстовых данных исходя из требований технического задания.

**Вариант №3**

**Задание 1**

Покажите, как текст

Ше;нкGFяуюх3~

представляется с использованием 8 битных кодировок:

1. cp866
2. Windows‑1251
3. ISO 8859‑5

В ответе приведите шестнадцатеричные коды символов.

**Решение**

Коды символов “;”, “G”, “F”, “3”, “~” будут одинаковыми в данных кодировках.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Символ** | ; | G | F | 3 | ~ |
| **Код** | 3B | 47 | 46 | 33 | 7E |

Остальные символы в данных кодировках:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Символ** | Ш | е | н | к | я | у | ю | х |
| **cp866** | 98 | A5 | AD | AA | EF | E3 | EE | E5 |
| **Windows-1251** | D8 | E5 | ED | EA | FF | F3 | FE | F5 |
| **ISO 8859-5** | C8 | D5 | DD | DA | EF | E3 | EE | E5 |

Таким образом, текст будет представляться, как:

1. 98 A5 3B AD AA 47 46 EF E3 EE E5 33 7E
2. D8 E5 3B ED EA 47 46 FF F3 FE F5 33 7E
3. C8 D5 3B DD DA 47 46 EF E3 EE E5 33 7E

**Задание 2**

Некоторый текст с использованием кодировки cp866 представляется в виде следующей последовательности кодов:

31 20 A3 A8 A3 A0 A1 A0 A9 E2 20 3D 31 30 32 34 8C A1 A0 A9 E2 5A

Укажите текст, соответствующий приведенной последовательности, исправьте допущенные ошибки и приведите правильный текст. Какой объем памяти займет заданный текст, если для его представления использовать кодировку Windows-1251?

**Решение**

Декодирование текста:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Код** | 31 | 20 | A3 | A8 | A3 | A0 | A1 | A0 | A9 | E2 | 20 |
| **Символ** | 1 | [SP] | г | и | г | а | б | а | й | т | [sp] |
| **Код** | 3D | 31 | 30 | 32 | 34 | 8C | A1 | A0 | A9 | E2 | 5A |
| **Символ** | = | 1 | 0 | 2 | 4 | М | б | а | й | т | Z |

Декодированный текст: 1 гигабайт =1024МбайтZ

Исправленный текст: 1 гигабайт = 1024 Мбайт

Код текста в cp866:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Код** | 31 | 20 | A3 | A8 | A3 | A0 | A1 | A0 | A9 | E2 | 20 | 3D |
| **Символ** | 1 | [SP] | г | и | г | а | б | а | й | т | [sp] | = |
| **Код** | 20 | 31 | 30 | 32 | 34 | 20 | 8C | A1 | A0 | A9 | E2 |  |
| **Символ** | [sp] | 1 | 0 | 2 | 4 | [sp] | М | б | а | й | т |  |

Код исправленного текста: 31 20 A3 A8 A3 A0 A1 A0 A9 E2 20 3D 20 31 30 32 34 20 8C A1 A0 A9 E2

В исходном тексте 22 символа, каждый из них в кодировке Windows-1251 будет кодироваться 1 байтом, следовательно, текст будет занимать 22 байта. В исправленном тексте 23 символа, он будет занимать 23 байта.

**Задание 3**

Текстовый файл содержит следующую последовательность байт:

EF BB BF D0 92 D1 88 D0 BE D0 B3 3A 66 77 71 68 67 21

Зная, что текст был сохранен с использованием представления UTF-8, определите последовательность кодов Unicode символов текста (каждый код укажите в формате U+xxxx, где xxxx – шестнадцатеричное представление кода). Укажите текст, соответствующий полученной последовательности кодов Unicode.

Каков будет размер текстового файла, если этот текст сохранить с использованием представления UTF-8 и записать в начало файла маркер порядка байт (BOM)?

**Решение**

Коды Unicode:

EF = **1110** 1111; BB = **10**11 1011; BF = **10**11 1111

Символ: 1111 1110 1111 1111 = U+FEFF – маркер порядка байт

D0 = **110**1 0000; 92 = **10**01 0010

Символ: 0000 0100 0001 0010 = U+0412 – В

D1 = **110**1 0001; 88 = **10**00 1000

Символ: 0000 0100 0100 1000 = U+0448 – ш

D0 = **110**1 0000; BE = **10**11 1110

Символ: 0000 0100 0011 1110 = U+043E – о

D0 = **110**1 0000; B3 = **10**11 0011

Символ: 0000 0100 0011 0011 = U+0433 – г

3A = **0**011 1010

Символ: U+003A – :

66 = **0**110 0110

Символ: U+0066 – f

77 = **0**111 0111

Символ: U+0077 – w

71 = **0**111 0001

Символ: U+0071 – q

68 = **0**110 1000

Символ: U+0068 – h

67 = **0**110 0111

Символ: U+0067 – g

21 = **0**010 0001

Символ: U+0021 – !

Получившаяся последовательность кодов: U+FEFF U+0412 U+0448 U+043E U+0433 U+003A U+0066 U+0077 U+0071 U+0068 U+0067 U+0021

Полученный текст: Вшог:fwqhg!

4 символа “В”, “ш”, “о”, “г” будут кодироваться 2 байтами, остальные 7 символов – 1 байтом, маркер порядка байт кодируется 3 байтами. Размер текста в UTF-8 с BOM: 4 · 2 + 7 · 1 + 3 = 8 + 7 + 3 = 18 байт

**Вывод:** были получены практические навыки представления текста с использованием кодировок cpp866, Windows-1251, ISO8859-5, декодирования текста с использованием кодировок cp866, UTF-8, нахождения объема, занимаемого текстом в кодировках Windows-1251, UTF-8.