Белорусский государственный технологический университет

Факультет информационных технологий

Кафедра программной инженерии

Лабораторная работа 3

По дисциплине «Основы программной инженерии»

На тему «Кодирование информации»

Выполнил:

Студент 1 курса 7 группы

Ленкевич Павел Андреевич

2023, Минск

**Кодирование информации**

**Цель работы:** ознакомиться с основами кодирования информации; освоить кодировки ASCII, Windows-1251, UTF-8, UTF-16, ознакомиться с их различиями.

Определения:

Минимальные единицы хранения информации – это бит, означающий двоичное число.

Минимальная адресуемая единица хранения информации – байт. Байт состоит из 8 битов.

|  |  |  |
| --- | --- | --- |
| **Десятичная** | **Двоичная** | **Шестнадцатеричная** |
| 0 | 00000000 | 00 |
| 1 | 00000001 | 01 |
| 2 | 00000010 | 02 |
| 3 | 00000011 | 03 |
| 4 | 00000100 | 04 |
| 5 | 00000101 | 05 |
| 6 | 00000110 | 06 |
| 7 | 00000111 | 07 |
| 8 | 00001000 | 08 |
| 9 | 00001001 | 09 |
| 10 | 00001010 | 0A |
| 11 | 00001011 | 0B |
| 12 | 00001100 | 0C |
| 13 | 00001101 | 0D |
| 14 | 00001110 | 0E |
| 15 | 00001111 | 0F |
| 16 | 00010000 | 10 |
| 17 | 00010001 | 11 |
| 18 | 00010010 | 12 |
| 19 | 00010011 | 13 |
| 20 | 00010100 | 14 |

**Приложение:**

#include <iostream>

// LenkevichPavel2004

// Windows-1251: 4C-65-6E-6B-65-76-69-63-68 50-61-76-65-6C 32-30-30-34

// UTF-8: 4C-65-6E-6B-65-76-69-63-68 50-61-76-65-6C 32-30-30-34

// UTF-16: 4C-65-6E-6B-65-76-69-63-68 50-61-76-65-6C 32-30-30-34

// ЛенкевичПавелАндреевич2004

// Windows-1251: 41B-435-43D-43A-435-432-438-447 41F-430-432-435-43B 410-43D-434-440-435-435-432-438-447 32-30-30-34

// UTF-8: D09B-D0B5-D0BD-D0BA-D0B5-D0B2-D0B8-D187 D09F-D0B0-D0B2-D0B5-D0BB 32-30-30-34

// UTF-16: 41B-435-43D-43A-435-432-438-447 41F-430-432-435-43B 32-30-30-34

// Ленкевич2004Pavel

// Windows-1251: 41B-435-43D-43A-435-432-438-447 32-30-30-34 50-61-76-65-6C

// UTF-8: D09B-D0B5-D0BD-D0BA-D0B5-D0B2-D0B8-D187 32-30-30-34 50-61-76-65-6C

// UTF-16: 41B-435-43D-43A-435-432-438-447 32-30-30-34 50-61-76-65-6C

int main()

{

setlocale(LC\_CTYPE, "rus");

int number = 0x12345678;

char hello[] = "Hello, ";

char lfie[] = "LenkevichPavel2004";

char rfie[] = "ЛенкевичПавелАндреевич2004";

char lr[] = "Ленкевич2004Pavel";

wchar\_t Lfie[] = L"LenkevichPavel2004";

wchar\_t Rfie[] = L"ЛенкевичПавелАндреевич2004";

wchar\_t LR[] = L"Ленкевич2004Pavel";

std::cout << hello << rfie << std::endl;

return 0;

}

**Кодировка Windows 1251:**

1. **Цифры:** 0030-0039
2. **Латиница:**

**Прописные**: 0041-005A

**Строчные:** 0061-007A

1. **Кириллица:**

**Прописные:** 0410-042F

**Строчные:** 0430-044F

**Первые 5 букв из фамилии:**

**Прописными (ЛЕНКЕ):** 041B 0415 041D 041A 0415

**Строчными (ленке):** 043B 0435 043D 043A 0435

**Словесный алгоритм перевода прописной буквы в строчнуюWindows-1251:**К коду Прописной буквы прибавить 0х20 (32 в десятичной). В результате получится код строчной буквы в кодировке Windows-1251

**Контрольные вопросы:**

1. **Таблица кодировки –** это таблицы, где каждому символу, буквам, цифрам, а также специальным знакам присвоен уникальный номер — код символа.
2. **ASCII -** American Standard Code for Information Interchange **-** американский стандартный код для обмена информацией
3. **Структура Windows-1251 – стандартной 8 битной кодировки:**

-Первая половина таблицы соответствует таблице ASCII и является международным стандартом (от 0 до 127)

**-Цифры:** 0030-0039

**-Латиница:**

**Прописные**: 0041-005A

**Строчные:** 0061-007A

**-Кириллица:**

**Прописные:** 0410-042F

**Строчные:** 0430-044F

**4. Unicode -** стандарт кодирования символов,

позволяющий представить знаки почти всех

письменных языков, состоит из 2х разделов:

**-UCS** – universal character set (универсальный

набор символов);

**-UTF –** Unicode transformation format (семейство

кодировок).

**5. Структура Unicode:**

* 1. UCS расположены в 17 плоскостях (0-16);
  2. в каждой плоскости 2^16 (65 536) символов;
  3. плоскость 0 – основная (основные символы);
  4. 1-14 – дополнительные;
  5. 15-16 – для частного использования

**6. UTF-8 -** представление Юникода, обеспечивающее совместимость со старыми системами, использовавшими 8-битные символы.

**UTF-16 –** представление Юникода в 16 битных символах.

**7. Разница в символах в UTF-16:**

**F и f:** 46 и 66

**S и s:** 53 и 73

**L и l:** 4C и 6C

**Б и б:** 411 и 431

**Г и г:** 413 и 433

**Э и э:** 42D и 44D

Между Прописной и строчной разница в коде – 20. Чтоб получить из строчной прописную надо отнять 020, а чтоб из прописной получить строчную, надо прибавить 020.