Белорусский государственный технологический университет

Факультет информационных технологий

Кафедра программной инженерии

 Лабораторная работа 3

По дисциплине «Основы программной инженерии»

На тему «Кодирование информации»

 Выполнила:

Студентка1 курса 6 группы

Альшевская Алина Михайловна

Преподаватель: Наркевич А.С.

2023, Минск

Минимальная единица хранения информации – это бит. Он может принимать только два значения: 0 или 1.

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

Бит может принимать только два значения: 0 или 1.

В одном байте содержится 8 битов.

|  |  |  |
| --- | --- | --- |
| Число | Двоичная система | Шестнадца-теричная система |
| 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 |

A(0041)-a(0061);L(004C)-l(006C);S(0053)-s(0073);H(0048)-h(0068);E(0045)-e(0065)

**ASCII (American Standard Code for Information Interchange)** — американский стандартный код для обмена информацией. ASCII — 8-битная кодировка для представления десятичных цифр, латинского и национального алфавитов, знаков препинания и управляющих символов.

Юникод – 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).

Принято обозначение символа U+xxx, где xxx- число в шестнадцатеричном формате.

• UNICODE:

− UCS расположены в 17 плоскостях (0-16);

− в каждой плоскости 2 16 (65 536) символов;

− плоскость 0 – основная (основные символы);

− 1-14 – дополнительные;

− 15-16 – для частного использования

UNICODE: кодировка UTF-8 UTF-8 — представление Юникода, обеспечивающее совместимость со старыми системами, использовавшими 8-битные символы.

Алгоритм кодирования в UTF-8:

1) определить количество октетов (октет: 8 битов или 1 байт) – т.е. в какой диапазон значений попадает количество значащих символов (7, 11, 16, 21, 26, 31);

2) подготовить старшие биты первого октета:

a. 0xxxxxxx для одного октета;

b. 110xxxxx – для двух;

c. 1110xxxx – для трех и т.д.;

d. 10xxxxxx – для остальных октетов;

3) заполнить оставшиеся биты (выше обозначены как x) в октетах кодом символа Юникода в двоичном виде. Начать с младших битов, поставив их в младшие биты последнего октета кода. И так далее, пока все биты кода символа не будут перенесены в свободные биты октетов.

UNICODE: кодировка UTF-16 В UTF-16 символы кодируются двухбайтовыми словами (16 битов) с использованием всех возможных диапазонов значений (от 0 до FFFF16).

Маркер последовательности байтов UNICODE: **BOM** (Byte Order Mark)

Для определения формата представления Юникода в начало текстового файла записывается сигнатура (обозначение) — символ **U+FEFF** — **маркер последовательности байтов**

Шестнадцатеричное представление маркера последовательности байтов для кодировок:

UTF-8 -be **BB BF**

UTF-16 (BE) FE FF

UTF-16 (LE) FF FE

• Порядок следования байтов:

− LE (Little endian order, прямой порядок, от младшего к старшему);

− BE (Big endian order, обратный порядок, от старшего к младшему).

