

Лабораторная работа 3 (4 часа)
Языки программирования

Кодировки ASCII, Windows-1251, UTF-8, UTF-16

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

1. Используйте при выполнении лабораторной работы материал лекции 2 [Л2].
2. Исследуйте таблицы кодировок **US-ASCII** и **Windows-1251** [Л2].
3. Создайте проект консольного C++-приложения с именем **LP_Lab03**.
4. Запишите в строках комментариев три строки:
 - 1) вашу фамилию, имя, год рождения на английском языке, например: **IvanovAlex1998**;
 - 2) фамилию, имя, год рождения на русском языке, например: **ИвановАлексей1998**;
 - 3) фамилию (на русском), год рождения, имя (на английском), например: **Иванов1998Alex**.
5. Используя таблицы кодировок в [Л2] запишите в комментариях шестнадцатеричное представление 3х строк записанных ранее в кодировке **Windows-1251**.
6. Используя таблицы кодировок в [Л2] запишите в комментариях шестнадцатеричное представление трех строк в кодировках **UTF-8** и **UTF-16**.
7. В программе проекта **LP_Lab03** объявите 6 строк: три записанные ранее в кодировке **Windows-1251** и эти же три строки в кодировке **UTF-16**.

```

#include "stdafx.h"
#include <cstdlib>

char lf[] = "IvanovAlex1998";
char rf[] = "ИвановАлексей1998";
char lr[] = "Иванов1998Alex";

wchar_t Llp[] = L"IvanovAlex1998";
wchar_t Lrp[] = L"ИвановАлексей1998";
wchar_t Llr[] = L"Иванов1998Alex";

int main()
{
    system("pause");
    return 0;
}

```

8. Запустите приложение и с помощью отладчика VS просмотрите области памяти, соответствующие шести объявленным строкам.
9. Убедитесь, что записанное в комментариях шестнадцатеричное представление строк в кодировках **Windows-1251** и **Unicode-16** (п.5, 6) идентично полученному с помощью отладчика.
10. Используя **HEX-Editor** (обратитесь к преподавателю) определите, в какой кодировке представлены входные файлы (*.cpp и *.h) транслятора C++.
11. С помощью VS разработайте три XML-файла, содержащих строки из пункта 4:
 - 1) **XML1251.xml** - в кодировке **Windows-1251**;
 - 2) **XMLUTF8.xml** - в кодировке **UTF-8**;
 - 3) **XMLUTF16.xml** в кодировке **UTF-16**.
12. Используя **HEX-Editor** проанализируйте XML-файлы (п.11) **XML1251.xml**, **XMLUTF8.xml**, **XMLUTF16.xml**.
13. Определите символы BOM (Byte order mark) в файлах **XML1251.xml**, **XMLUTF8.xml**, **XMLUTF16.xml**.
14. Определите в какой кодировке представлены входные файлы (*.cpp и *.h) транслятора C++.
15. Разработайте XML-файл **XMLUTF16BE.xml**, применяющий кодировку **UTF-16BE**.
16. Используя **HEX-Editor**, проанализируйте XML-файлы (п.15) **XMLUTF16BE.xml**, поясните отличие от файла **XMLUTF16.xml**.
17. Определите разницу значений кодов **Windows-1251** символов: **F** и **f**, **S** и **s**, **L** и **l**, **Б** и **б**, **Г** и **г**, **Э** и **э**.
18. В рамках проекта **LP_Lab03** разработайте функцию:

char* UpperW1251(char* destination, char* source)

функция принимает строку символов в кодировке **Windows-1251** (параметр **source**), формирует строку в кодировке **Windows-1251** (параметр **destination**), содержащей буквы только верхнего регистра и возвращает указатель на строку **destination**.

```
char* UpperW1251(char* destination, char* source);  
  
int main()  
{  
    char source[] = "12345sdfghjлйцгшщз";  
    char destination[] = "124567890123456789";  
  
    UpperW1251(destination, source);  
  
    std::cout<< destination<< std::endl;  
  
    system("pause");  
    return 0;  
}
```

Ответьте на следующие вопросы:

- что такое **таблица кодировки**?
- расшифруйте аббревиатуру **ASCII**;
- поясните структуру кодировки **Windows-1251**;
- что такое **UNICODE**?
- поясните структуру кодировки **UNICODE**;
- что такое **UTF-8** и **UTF-16**?
- в каких случаях целесообразно применять кодировку **UTF-8**?
- определите разницу значений кодов **UTF-16** символов: **F** и **f**, **S** и **s**, **L** и **l**, **Б** и **б**, **Г** и **г**, **Э** и **э**; каким образом получить из кода буквы нижнего регистра получить букву верхнего регистра.