**Home Work – 09-12-19**

**Name: N Nachimuthu Reg No: 17BCE0211**

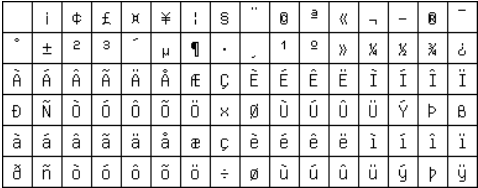
**Character encoding:**

Character encoding is the binary representation of characters that we use in various languages. For a digital representation of a character to be available, the character should have an encoding to which it can be mapped to.

**Various Character set encoding:**

**ISO 8859-1 Western Europe:**

Covers languages such as Albanian, Catalan, Danish, Dutch, English, Faeroese, Finnish, French, German, Galician, Irish, Icelandic, Italian, Norwegian, Portuguese, Spanish, and Swedish.



**ISO 8859-2 Western and Central Europe:**

Used mostly for Czech, German, Hungarian, Polish, Rumanian, Croatian, Slovak, Sloven.



**ISO 8859-3 Western Europe and South European (Turkish, Maltese plus Esperanto):**

Used for languages like Esperanto, Galician, Maltese, and Turkish.



**ISO 8859-4 Western Europe and Baltic countries (Lithuania, Estonia, Latvia and Lapp):**

Introduces letters for Estonian, Latvian, and Lithuanian.



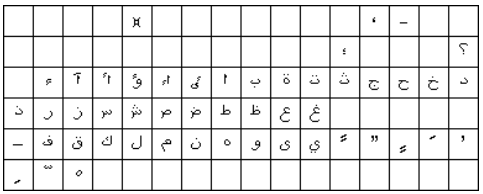
**ISO 8859-5 Cyrillic alphabet:**

Used for Bulgarian, Byelorussian, Macedonian, Russian, Serbian and Ukrainian. Lacks few Ukrainian letters.



**ISO 8859-6 Arabic:**

Includes support for Arabic letters. Arabic letters could be displayed with help of a display engine.



**ISO 8859-7 Greek:**

Used for modern Greek.



**ASCII:**

This is used to represent English alphabets.

**UTF-8:**

8-bit Unicode transformation format. It has variable length encoding. It is capable of encoding all 11,12,064 valid code points in Unicode. Current it is used in most of the documents.

**UTF-16:**

16-bit Unicode transformation format. It has variable length encoding. It is capable of encoding all 11,12,064 valid code points in Unicode. It is used internally by windows, java and javascript

**UTF-32:**

32-bit Unicode transformation format. It encodes using a constant 32 bit for all code points. Its main advantage is that it takes constant time to find a code point.