**Name – Patel Het Kiritkumar**

**Reg no. 17BCE0567**

Character encoding info systems

1. ASCII Encoding

* American Standard Code for Information Interchange
* 128 characters

2. Unicode Consortium and UTF encodings

* If we think about 8-bit fixed-width encoding, we have only 8 bits to represent a character. Hence the maximum of 256 characters can be represented by such an encoding.
* The Unicode Consortium also maintains the standard for UTF encodings. UTF (an acronym for Unicode Transformation Format) is a set of encoding schemes based on Unicode charset.

3. UTF Encodings

UTF-8

* UTF-8 is an 8-bit variable-length encoding scheme designed to be compatible with ASCII encoding.

UTF-16

* UTF-16 is 16-bit variable length encoding scheme and it uses the UTF character set for character code points. This means that a UTF-16 encoded character will have a 16-bit code unit.

4. UCS Encodings

* This ISO/IEC 10646 standard is maintained in conjunction with The Unicode Standard (“Unicode”), and they are code-for-code identical.

UCS-2

* UCS-2 is 16-bit fixed-width encoding (2 bytes)

UCS-4

* UCS-4 is 32-bit fixed-width encoding (4 bytes), which means 32 bits will be used to encode a character