NAME: Saif Ullah

ID: 17352

SECTION:BS SE(6-A)

SUBJECT: WEB ENGINEERING

TEACHER: SIR AYUB KHAN

ASSIGNMENT NO: 05

ISO-10646-UCS-2

The Universal Coded Character Set is a standard set of characters defined by the international standard ISO/IEC 10646, Information technology — Universal Coded Character Set, which is the basis of many character encodings, improving as characters from previously unrepresented typing systems are added.

ISO-8859-1

ISO/IEC 8859-1:1998, Information technology — 8-bit single-byte coded graphic character sets — Part 1: Latin alphabet No. 1, is part of the ISO/IEC 8859 series of ASCII-based standard character encodings, first edition published in 1987. ISO/IEC 8859-1.

ISO-8859-9

ISO/IEC 8859-9:1999, Information technology — 8-bit single-byte coded graphic character sets — Part 9: Latin alphabet No. 5, is part of the ISO/IEC 8859 series of ASCII-based standard character encodings, first edition published in 1989.

UTF-8

UTF-8 is a variable-length character encoding standard used for electronic communication. Defined by the Unicode Standard, the name is derived from Unicode Transformation Format – 8-bit. UTF-8 is capable of encoding all 1,112,064 valid character code points in Unicode using one to four one-byte code units.

UTF-16

UTF-16 is a character encoding capable of encoding all 1,112,064 valid code points of Unicode. The encoding is variable-length, as code points are encoded with one or two 16-bit code units. UTF-16 is an encoding of Unicode in which each character is composed of either one or two 16-bit elements. Unicode was originally designed as a pure 16-bit encoding, aimed at representing all modern scripts.

ISO-2022-JP

ISO 2022 specifies a general structure which character encodings can conform to, dedicating particular ranges of bytes (0x00–1F and 0x7F–9F) to be used for non-printing control codes for formatting and in-band instructions (such as line breaks or formatting instructions for text terminals), rather than graphical characters. It also specifies a syntax for escape sequences, multiple-byte sequences beginning with the ESC control code, which can likewise be used for in-band instructions. Specific sets of control codes and escape sequences designed to be used with ISO 2022 include ISO/IEC 6429, portions of which are implemented by ANSI.SYS and terminal emulators.

SHIFT-JIS

Shift JIS is a character encoding for the Japanese language, originally developed by a Japanese company called ASCII Corporation in conjunction with Microsoft and standardized as JIS X 0208 Appendix 1. As of October 2022, 0.2% of all web pages used Shift JIS, a decline from 1.3% in July 2014.

EUC-JP

(EUC) is a multibyte character encoding system used primarily for Japanese, Korean, and simplified Chinese.

The most commonly used EUC codes are variable-length encodings with a character belonging to an ISO/IEC 646 compliant coded character set (such as ASCII) taking one byte, and a character belonging to a 94x94 coded character set (such as GB 2312) represented in two bytes. The EUC-CN form of GB 2312 and EUC-KR are examples of such two-byte EUC codes. EUC-JP includes characters represented by up to three bytes, including an initial shift code, whereas a single character in EUC-TW can take up to four bytes.