ENCODING SCHEMES

As we all know, computers do not understand the English alphabet, numbers except 0 and 1, or text symbols. We use encoding to convert these. So, encoding is the method or process of converting a series of characters, i.e. letters, numbers, punctuation, and symbols into a special or unique format for transmission or storage in computers. Data is represented in computers using ASCII, UTF8, UTF32, ISCII, and Unicode encoding schemes. All types of data, including numbers, text, photos, audio, and video files, can be handled by computers. For example, 65 is represented as A because all the characters, symbols, numbers are assigned some unique code by the standard encoding schemes. Some of the commonly used encoding schemes are described below:

ASCII Code:

It is a simple code assigned to a character to represent data. ASCII (pronounced as "askee") is a 7 – bit character code. It represents all the characters available for writing in the file. Every digit in ASCII code represents one byte. Let us have a look at the following commonly used characters and code range:

Characters	**ASCII Codes**
A to Z	65 to 90
a to z	97 to 122
0 to 9	48 to 57
Space	32

The computer generally converts these ASCII codes into an equivalent binary number.

ISCII Code:

When computers are used with English language ASCII codes enough to represent data. But as and when the use of computers broadly extended to countries like India, it's very important to represent data in Indian Languages. So for that in 1991, the Bureau of Indian Standards adopted the Indian Standard Code for Information Interchange has evolved. This code is capable of 8 bits. It is also known as the Indian Scripts Code for Information Interchange. It supports various Indian languages like Devnagri, Gurumukhi, Gujarati, Oriya, Bengali, Assamese, Telugu etc.

Unicode:

Unicode is used for a universal set of characters. As ISCII is used for Indian languages, Unicode is used accepted by the universal standards and represent data in different languages. It was designed for the purpose of representing almost all the languages in computers. The Unicode has different versions like UTF – 8, UTF – 16, UTF – 32.

In this, the numbers in UTF like 8 as 8 bits known as an octet, 16 as 2 – octets, and 32 as 4 – octets representation.