Encoding Scheme

An encoding scheme is a system that is used to represent data, typically text or other forms of information, in a format that can be stored or transmitted using digital technology.

There are many different encoding schemes that have been developed over the years, each with their own advantages and limitations. Some common examples of encoding schemes include ASCII, Unicode, UTF-8, and ISO-8859.

ASCII (American Standard Code for Information Interchange) is one of the oldest and most widely used encoding schemes. It uses 7 bits to represent each character, allowing for a total of 128 possible characters.

Unicode is a more recent encoding scheme that was developed to support a wider range of characters from different languages and scripts. It uses 16 bits to represent each character, allowing for a much larger range of possible characters.

UTF-8 is a variable-length encoding scheme that is backwards-compatible with ASCII, meaning that it can represent all of the same characters as ASCII but can also represent additional characters from Unicode as needed.

ISO-8859 is a family of encoding schemes that are used to support specific languages or regions. For example, ISO-8859-1 is used to support Western European languages, while ISO-8859-5 is used to support Cyrillic scripts.

ISCII

ISCII stands for Indian Script Code for Information Interchange. It is an encoding scheme developed by the Indian government in the 1980s to support the representation of text in various Indian scripts, including Devanagari, Bengali, Gurmukhi, Gujarati, Oriya, Tamil, Telugu, Kannada, and Malayalam.

ISCII is an 8-bit encoding scheme, which means it can represent up to 256 characters. In ISCII, each character is represented by a unique code point, which is a number that corresponds to a particular character in the encoding. ISCII also supports a number of control codes that are used for formatting and other purposes.

ISCII has been widely used in India for many years, particularly in the government and academic sectors, but it has largely been superseded by Unicode in recent years. Unicode is a more

comprehensive encoding scheme that includes support for a much larger range of scripts and characters, and it is widely used

UTF32

UTF-32 is a fixed-length encoding scheme that is a part of the Unicode standard. In UTF-32, each Unicode code point is represented using 32 bits, which means that each character is represented by a single 32-bit value. This makes UTF-32 a straightforward encoding scheme that is easy to process and allows for constant-time random access to any character in a string.

Because UTF-32 uses a fixed-length encoding, it is immune to many of the problems that can occur with variable-length encodings, such as UTF-8. However, it also means that UTF-32 typically requires more storage space than variable-length encodings for the same amount of text.

UTF-32 is primarily used in situations where efficient random access to characters is important, such as in certain types of text processing and indexing applications. However, it is not as widely used as other encoding schemes like UTF-8 or UTF-16, which are more space-efficient and are better suited to many common text handling tasks.